

# CIKM '23: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management



[Full Citation in the ACM Digital Library](#)

## SESSION: Keynote Talks

### [Interpretable Natural Language Understanding](#)

Yulan He

In recent years, we have witnessed the shift of paradigms in Natural Language Processing (NLP) from fine-tuning large-scale pre-trained language models (PLMs) on task-specific data to prompt-based learning. In the latter, the task description is embedded into the PLM input, enabling the same model to handle multiple tasks. While both approaches have demonstrated impressive performance in various NLP tasks, their opaque nature makes comprehending their inner workings and decision-making processes challenging for humans.

In this talk, I will share the research undertaken in my group to address the interpretability concerns surrounding neural models in language understanding. This includes a hierarchical interpretable text classifier going beyond word-level interpretations, uncertainty interpretation of text classifiers built on PLMs, explainable recommender systems by harnessing information across diverse modalities, and explainable student answer scoring. I will conclude my talk by offering insights into potential future developments in interpretable language understanding.

### [Generative AI and the Future of Information Access](#)

Chirag Shah

The prominent model of retrieving, evaluating, and using relevant information from databases, collections, and the web is going through a significant transformation. This is largely due to wide-scale availability of various generative AI systems that can take in natural language inputs and generate highly customized natural language text, images, audio, and videos. This transformation in how people seek and access information will have profound impacts on users, developers, and policymakers. It is already changing many sectors including education, health, and commerce. But the hopes and hypes of generative AI are often not clear as we get swept up by either the current capabilities and limitations of this technology in the short term or fear from speculative future in the long term. Instead, I believe we need to approach this area pragmatically and with scientific curiosity, scholarly rigor, and societal responsibility. In this talk, I will highlight some of the opportunities and challenges for information access stemming from recent advancements in generative AI. For instance, there are new possibilities now for addressing accessibility, low-resource domains, and bias in training data using generative AI tools. On the other hand, there are new challenges concerning hallucination, toxicity, and information provenance. It is clear that we want to benefit from what AI systems are capable of, but how do we do that while curbing some of these problems? I will argue that the solution is multifaceted and complex -- some will require technical advancements and others will call for policy changes. We will need to not only build information systems with fairness, transparency, and accountability in mind, but also train a new generation of developers, policymakers, and of course the users. The goal here is to cut through both hype and fear and think pragmatically about the future of information access.

### [Knowledge Graphs for Knowing More and Knowing for Sure](#)

Steffen Staab

Knowledge graphs have been conceived to collect heterogeneous data and knowledge about large domains, e.g. medical or engineering domains, and to allow versatile access to such collections by means of querying and logical reasoning. A surge of methods has responded to additional requirements in recent years. (i) Knowledge graph embeddings use similarity and analogy of structures to speculatively add to the collected data and knowledge. (ii) Queries with shapes and schema information can be typed to provide certainty about results. We survey both developments and find that the development of techniques happens in disjoint communities that mostly do not understand each other, thus limiting the proper and most versatile use of knowledge graphs.

## SESSION: Full Papers

### [Optimizing for Member Value in an Edge Building Marketplace](#)

Ayan Acharya, Siyuan Gao, Ankan Saha, Borja Ocejjo, Kinjal Basu, Keerthi Selvaraj, Rahul Mazumder, Aman Gupta, Parag Agrawal, Ayan Acharya

Social networks are prosperous marketplaces where creators and consumers congregate to share and consume various content. In general, products that rank content for distribution (such as newsfeeds, stories, and notifications) and are related to edge recommendations (such as connect to members, follow celebrities or groups or hashtags) optimize the experience of active users. Typically, such users generate ample interaction data amenable to accurate model training and prediction. In contrast, we prioritize enhancing the experience of inactive members (IMs) who do not have a rich connection network. We formulate strategies for recommending superior edges to help members grow their connection network. Adapting the recommendations provides enormous value to the IMs and can significantly influence their future behaviour and engagement with the ecosystem. To that end, we propose a general and scalable multi-objective optimization (MOO) framework to provide more value to IMs as invitation recipients on LinkedIn, a professional network with over 900M members. To deal with the enormous scale, we formulate the problem as a massive constrained linear optimization involving billions of variables and millions of constraints and efficiently solve it using accelerated gradient descent, making this the largest deployment of LP-based recommender systems worldwide. Furthermore, the proposed MOO paradigm can solve the general problem of matching different types of entities in an m-sided marketplace. Finally, we discuss the challenges and benefits of implementing and ramping our method in production at scale at LinkedIn and report our findings about the core business metrics related to users' engagement and network health.

#### [Combining Inductive and Deductive Reasoning for Query Answering over Incomplete Knowledge Graphs](#)

Medina Andresel, Trung-Kien Tran, Csaba Domokos, Pasquale Minervini, Daria Stepanova

Current methods for embedding-based query answering over incomplete Knowledge Graphs (KGs) only focus on inductive reasoning, i.e., predicting answers by learning patterns from the data, and lack the complementary ability to do deductive reasoning, which requires the application of domain knowledge to infer further information. To address this shortcoming, we investigate the problem of incorporating ontologies into embedding-based query answering models by defining the task of embedding-based ontology-mediated query answering. We propose various integration strategies into prominent representatives of embedding models that involve (1) different ontology-driven data augmentation techniques and (2) adaptation of the loss function to enforce the ontology axioms. We design novel benchmarks for the considered task based on the LUBM and the NELL KGs and evaluate our methods on them. The achieved improvements in the setting that requires both inductive and deductive reasoning are from 20% to 55% in HITS@3.

#### [CLosER: Conversational Legal Longformer with Expertise-Aware Passage Response Ranker for Long Contexts](#)

Arian Askari, Mohammad Aliannejadi, Amin Abolghasemi, Evangelos Kanoulas, Suzan Verberne

In this paper, we investigate the task of response ranking in conversational legal search. We propose a novel method for conversational passage response retrieval (ConvPR) for long conversations in domains with mixed levels of expertise. Conversational legal search is challenging because the domain includes long, multi-participant dialogues with domain-specific language. Furthermore, as opposed to other domains, there typically is a large knowledge gap between the questioner (a layperson) and the responders (lawyers), participating in the same conversation. We collect and release a large-scale real-world dataset called LegalConv with nearly one million legal conversations from a legal community question answering (CQA) platform. We address the particular challenges of processing legal conversations, with our novel Conversational Legal Longformer with Expertise-Aware Response Ranker, called CLosER. The proposed method has two main innovations compared to state-of-the-art methods for ConvPR: (i) Expertise-Aware Post-Training; a learning objective that takes into account the knowledge gap difference between participants to the conversation; and (ii) a simple but effective strategy for re-ordering the context utterances in long conversations to overcome the limitations of the sparse attention mechanism of the Longformer architecture. Evaluation on LegalConv shows that our proposed method substantially and significantly outperforms existing state-of-the-art models on the response selection task. Our analysis indicates that our Expertise-Aware PostTraining, i.e., continued pre-training or domain/task adaptation, plays an important role in the achieved effectiveness. Our proposed method is generalizable to other tasks with domain-specific challenges and can facilitate future research on conversational search in other domains.

#### [GripRank: Bridging the Gap between Retrieval and Generation via the Generative Knowledge Improved Passage Ranking](#)

Jiaqi Bai, Hongcheng Guo, Jiaheng Liu, Jian Yang, Xinnian Liang, Zhao Yan, Zhoujun Li

Retrieval-enhanced text generation has shown remarkable progress on knowledge-intensive language tasks, such as open-domain question answering and knowledge-enhanced dialogue generation, by leveraging passages retrieved from a large passage corpus for delivering a proper answer given the input query. However, the retrieved passages are not ideal for guiding answer generation because of the discrepancy between retrieval and generation, i.e., the candidate passages are all treated equally during the retrieval procedure without considering their potential to generate a proper answer. This discrepancy makes a passage retriever deliver a sub-optimal collection of candidate passages to generate the answer. In this paper, we propose the GeneRative Knowledge Improved Passage Ranking (GripRank) approach, addressing the above challenge by distilling knowledge from a generative passage estimator (GPE) to a passage ranker, where the GPE is a generative language model used to measure how likely the candidate passages can generate the proper answer. We realize the distillation procedure by teaching the passage ranker learning to rank the passages ordered by the GPE. Furthermore, we improve the distillation quality by devising a curriculum knowledge distillation mechanism, which allows the knowledge provided by the GPE can be progressively distilled to the ranker through an easy-to-hard curriculum, enabling the passage ranker to correctly recognize the provenance of the answer from many plausible candidates. We conduct extensive experiments on four datasets across three knowledge-intensive language tasks. Experimental results show advantages over the state-of-the-art methods for both passage ranking and answer generation on the KILT benchmark.

#### **MPMRC-MNER: A Unified MRC framework for Multimodal Named Entity Recognition based Multimodal Prompt**

Xigang Bao, Mengyuan Tian, Zhiyuan Zha, Biao Qin

Multimodal named entity recognition (MNER) is a vision-language task, which aims to detect entity spans and classify them to corresponding entity types given a sentence-image pair. Existing methods often regard an image as a set of visual objects, trying to explicitly capture the relations between visual objects and entities. However, since visual objects are often not identical to entities in quantity and type, they may suffer the bias introduced by visual objects rather than aid. Inspired by the success of textual prompt-based fine-tuning (PF) approaches in many methods, in this paper, we propose a Multimodal Prompt-based Machine Reading Comprehension based framework to implicit alignment between text and image for improving MNER, namely MPMRC-MNER. Specifically, we transform text-only query in MRC into multimodal prompt containing image tokens and text tokens. To better integrate image tokens and text tokens, we design a prompt-aware attention mechanism for better cross-modal fusion. At last, contrastive learning with two types of contrastive losses is designed to learn more consistent representation of two modalities and reduce noise. Extensive experiments and analyses on two public MNER datasets, Twitter2015 and Twitter2017, demonstrate the better performance of our model against the state-of-the-art methods.

#### **Deep Integrated Explanations**

Oren Barkan, Yehonatan Elisha, Jonathan Weill, Yuval Asher, Amit Eshel, Noam Koenigstein

This paper presents Deep Integrated Explanations (DIX) - a universal method for explaining vision models. DIX generates explanation maps by integrating information from the intermediate representations of the model, coupled with their corresponding gradients. Through an extensive array of both objective and subjective evaluations spanning diverse tasks, datasets, and model configurations, we showcase the efficacy of DIX in generating faithful and accurate explanation maps, while surpassing current state-of-the-art methods. Our code is available at: <https://github.com/dix-cikm23/dix>

#### **GraphERT-- Transformers-based Temporal Dynamic Graph Embedding**

Moran Beladev, Gilad Katz, Lior Rokach, Uriel Singer, Kira Radinsky

Dynamic temporal graphs evolve over time, adding and removing nodes and edges between time snapshots. The tasks performed on such graphs are diverse and include detecting temporal trends, finding graph-to-graph similarities, and graph visualization and clustering. For all these tasks, it is necessary to embed the entire graph in a low-dimensional space by using graph-level representations instead of the more common node-level representations. This embedding requires handling the appearance of new nodes over time as well as capturing temporal patterns of the entire graph. Most existing methods perform temporal node embeddings and focus on different methods of aggregating them for a graph-based representation. In this work, we propose an end-to-end architecture that captures both the node embeddings and their influence in a structural context during a specific time period of the graph. We present GraphERT (Graph Embedding Representation using Transformers), a novel approach to temporal graph-level embeddings. Our method pioneers the use of Transformers to seamlessly integrate graph structure learning with temporal analysis. By employing a masked language model on sequences of graph random walks, together with a novel temporal classification task, our model not only comprehends the intricate graph dynamics but also unravels the temporal significance of each node and path. This novel training paradigm empowers

GraphERT to capture the essence of both the structural and temporal aspects of graphs, surpassing state-of-the-art approaches across multiple tasks on real-world datasets.

### **Faster Approximation Algorithms for Parameterized Graph Clustering and Edge Labeling**

Vedangi Bengali, Nate Veldt

Graph clustering is a fundamental task in network analysis where the goal is to detect sets of nodes that are well-connected to each other but sparsely connected to the rest of the graph. We present faster approximation algorithms for an NP-hard parameterized clustering framework called LambdaCC, which is governed by a tunable resolution parameter and generalizes many other clustering objectives such as modularity, sparsest cut, and cluster deletion. Previous LambdaCC algorithms are either heuristics with no approximation guarantees, or computationally expensive approximation algorithms. We provide fast new approximation algorithms that can be made purely combinatorial. These rely on a new parameterized edge labeling problem we introduce that generalizes previous edge labeling problems that are based on the principle of strong triadic closure and are of independent interest in social network analysis. Our methods are orders of magnitude more scalable than previous approximation algorithms and our lower bounds allow us to obtain a posteriori approximation guarantees for previous heuristics that have no approximation guarantees of their own.

### **Relevance-based Infilling for Natural Language Counterfactuals**

Lorenzo Betti, Carlo Abrate, Francesco Bonchi, Andreas Kaltenbrunner

Counterfactual explanations are a natural way for humans to gain understanding and trust in the outcomes of complex machine learning algorithms. In the context of natural language processing, generating counterfactuals is particularly challenging as it requires the generated text to be fluent, grammatically correct, and meaningful. In this study, we improve the current state of the art for the generation of such counterfactual explanations for text classifiers. Our approach, named RELITC (Relevance-based Infilling for Textual Counterfactuals), builds on the idea of masking a fraction of text tokens based on their importance in a given prediction task and employs a novel strategy, based on the entropy of their associated probability distributions, to determine the infilling order of these tokens. Our method uses less time than competing methods to generate counterfactuals that require less changes, are closer to the original text and preserve its content better, while being competitive in terms of fluency. We demonstrate the effectiveness of the method on four different datasets and show the quality of its outcomes in a comparison with human generated counterfactuals.

### **Bridged-GNN: Knowledge Bridge Learning for Effective Knowledge Transfer**

Wendong Bi, Xueqi Cheng, Bingbing Xu, Xiaoqian Sun, Li Xu, Huawei Shen

The data-hungry problem, characterized by insufficiency and low-quality of data, poses obstacles for deep learning models. Transfer learning has been a feasible way to transfer knowledge from high-quality external data of source domains to limited data of target domains, which follows a domain-level knowledge transfer to learn a shared posterior distribution. However, they are usually built on strong assumptions, e.g., the domain invariant posterior distribution, which is usually unsatisfied and may introduce noises, resulting in poor generalization ability on target domains. Inspired by Graph Neural Networks (GNNs) that aggregate information from neighboring nodes, we redefine the paradigm as learning a knowledge-enhanced posterior distribution for target domains, namely Knowledge Bridge Learning (KBL). KBL first learns the scope of knowledge transfer by constructing a Bridged-Graph that connects knowledgeable samples to each target sample and then performs sample-wise knowledge transfer via GNNs. KBL is free from strong assumptions and is robust to noises in the source data. Guided by KBL, we propose the Bridged-GNN including an Adaptive Knowledge Retrieval module to build Bridged-Graph and a Graph Knowledge Transfer module. Comprehensive experiments on both un-relational and relational data-hungry scenarios demonstrate the significant improvements of Bridged-GNN compared with SOTA methods

### **Multi-modal Mixture of Experts Representation Learning for Sequential Recommendation**

Shuqing Bian, Xingyu Pan, Wayne Xin Zhao, Jinpeng Wang, Chuyuan Wang, Ji-Rong Wen

Within online platforms, it is critical to capture the dynamic user preference from the sequential interaction behaviors for making accurate recommendation over time. Recently, significant progress has been made in sequential recommendation with deep learning. However, existing neural sequential recommender often suffer from the data sparsity issue in real-world applications.

To tackle this problem, we propose a Multi-Modal Mixture of experts model for Sequential Recommendation, named M3SRec, which leverage rich multi-modal interaction data for improving sequential recommendation. Different from existing

multi-modal recommendation models, our approach jointly considers reducing the semantic gap across modalities and adapts multi-modal semantics to fit recommender systems. For this purpose, we make two important technical contributions in architecture and training. Firstly, we design a novel multi-modal mixture-of-experts (MoE) fusion network, which can deeply fuse the across-modal semantics and largely enhance the modeling capacity of complex user intents. For training, we design specific pre-training tasks that can mimic the goal of the recommendation, which help model learn the semantic relatedness between the multi-modal sequential context and the target item. Extensive experiments conducted on both public and industry datasets demonstrate the superiority of our proposed method over existing state-of-the-art methods, especially when only limited training data is available.

#### **CPMR: Context-Aware Incremental Sequential Recommendation with Pseudo-Multi-Task Learning**

Qingtian Bian, Jiaxing Xu, Hui Fang, Yiping Ke

The motivations of users to make interactions can be divided into static preference and dynamic interest. To accurately model user representations over time, recent studies in sequential recommendation utilize information propagation and evolution to mine from batches of arriving interactions. However, they ignore the fact that people are easily influenced by the recent actions of other users in the contextual scenario, and applying evolution across all historical interactions dilutes the importance of recent ones, thus failing to model the evolution of dynamic interest accurately. To address this issue, we propose a Context-Aware Pseudo-Multi-Task Recommender System (CPMR) to model the evolution in both historical and contextual scenarios by creating three representations for each user and item under different dynamics: static embedding, historical temporal states, and contextual temporal states. To dually improve the performance of temporal states evolution and incremental recommendation, we design a Pseudo-Multi-Task Learning (PMTL) paradigm by stacking the incremental single-target recommendations into one multi-target task for joint optimization. Within the PMTL paradigm, CPMR employs a shared-bottom network to conduct the evolution of temporal states across historical and contextual scenarios, as well as the fusion of them at the user-item level. In addition, CPMR incorporates one real tower for incremental predictions, and two pseudo towers dedicated to updating the respective temporal states based on new batches of interactions. Experimental results on four benchmark recommendation datasets show that CPMR consistently outperforms state-of-the-art baselines and achieves significant gains on three of them. The source code is available at <https://github.com/DiMarzioBian/CPMR>.

#### **Enabling Health Data Sharing with Fine-Grained Privacy**

Luca Bonomi, Sepand Gousheh, Liyue Fan

Sharing health data is vital in advancing medical research and transforming knowledge into clinical practice. Meanwhile, protecting the privacy of data contributors is of paramount importance. To that end, several privacy approaches have been proposed to protect individual data contributors in data sharing, including data anonymization and data synthesis techniques. These approaches have shown promising results in providing privacy protection at the dataset level. In this work, we study the privacy challenges in enabling fine-grained privacy in health data sharing. Our work is motivated by recent research findings, in which patients and healthcare providers may have different privacy preferences and policies that need to be addressed. Specifically, we propose a novel and effective privacy solution that enables data curators (e.g., healthcare providers) to protect sensitive data elements while preserving data usefulness. Our solution builds on randomized techniques to provide rigorous privacy protection for sensitive elements and leverages graphical models to mitigate privacy leakage due to dependent elements. To enhance the usefulness of the shared data, our randomized mechanism incorporates domain knowledge to preserve semantic similarity and adopts a block-structured design to minimize utility loss. Evaluations with real-world health data demonstrate the effectiveness of our approach and the usefulness of the shared data for health applications.

#### **Beyond Trading Data: The Hidden Influence of Public Awareness and Interest on Cryptocurrency Volatility**

Zeyd Boukhers, Azeddine Bouabdallah, Cong Yang, Jan Jürjens

Since Bitcoin first appeared on the scene in 2009, cryptocurrencies have become a worldwide phenomenon as important decentralized financial assets. Their decentralized nature, however, leads to notable volatility against traditional fiat currencies, making the task of accurately forecasting the crypto-fiat exchange rate complex. In this study, we examine the various independent factors that affect the Bitcoin-Dollar exchange rate's volatility. To this end, we propose CoMForE, a multimodal AdaBoost-LSTM ensemble model, which not only utilizes historical trading data but also incorporates public sentiments from related tweets, public interest demonstrated by search volumes, and blockchain hash-rate data. Our developed model goes a step further by predicting fluctuations in the overall cryptocurrency value distribution, thus increasing its value for investment decision-making. We have subjected this method to extensive testing via comprehensive



experiments, thereby validating the importance of multimodal combination over exclusive reliance on trading data. Further experiments show that our method significantly surpasses existing forecasting tools and methodologies, demonstrating a 19.29% improvement. This result underscores the influence of external independent factors on cryptocurrency volatility.

### **Fair&Share: Fast and Fair Multi-Criteria Selections**

Kathleen Cachel, Elke Rundensteiner

Traditional multi-criteria selection methods are the leading approach for selecting a set of candidates when multiple criteria determine selection relevancy. For instance, hiring platforms combine candidates' proximity, skills, and years of experience to build shortlists for recruiters. While these methods succeed in efficiently selecting candidates, their chosen set may unfairly affect marginalized candidate groups (e.g., race or gender). Bridging the gap between traditional fairness-unaware multi-criteria selection and contemporary fairness interventions, we characterize the open problem of fair multi-criteria selection. We design Fair&Share the first efficient fairness-tunable multi-criteria selection method. Fair&Share supports several fair representation notions. The key to Fair&Share is the design of its group-aware utility objective. Fair&Share uses a novel fairness calibration component to provide a user-friendly tuning mechanism for controlling the balance between selection relevancy (utility) and representation fairness. Our fairness-focused selection policy iteratively builds the result set by prioritizing candidates as aiding either the fair representation or the shared overall utility goals. We prove the optimality of Fair&Share, meaning that Fair&Share selects the best possible candidates such that the desired fair representation is achieved. Our experimental study demonstrates that Fair&Share achieves the best fairness and utility performance of state-of-the-art alternatives adapted to this new problem while taking a fraction of the time.

### **Generalization Bound for Estimating Causal Effects from Observational Network Data**

Ruichu Cai, Zeqin Yang, Weilin Chen, Yuguang Yan, Zhifeng Hao

Estimating causal effects from observational network data is a significant but challenging problem. Existing works in causal inference for observational network data lack an analysis of the generalization bound, which can theoretically provide support for alleviating the complex confounding bias and practically guide the design of learning objectives in a principled manner. To fill this gap, we derive a generalization bound for causal effect estimation in network scenarios by exploiting 1) the reweighting schema based on joint propensity score and 2) the representation learning schema based on Integral Probability Metric (IPM). We provide two perspectives on the generalization bound in terms of reweighting and representation learning, respectively. Motivated by the analysis of the bound, we propose a weighting regression method based on the joint propensity score augmented with representation learning. Extensive experimental studies on two real-world networks with semi-synthetic data demonstrate the effectiveness of our algorithm.

### **How Expressive are Graph Neural Networks in Recommendation?**

Xuheng Cai, Lianghao Xia, Xubin Ren, Chao Huang

Graph Neural Networks (GNNs) have demonstrated superior performance in various graph learning tasks, including recommendation, where they explore user-item collaborative filtering signals within graphs. However, despite their empirical effectiveness in state-of-the-art recommender models, theoretical formulations of their capability are scarce. Recently, researchers have explored the expressiveness of GNNs, demonstrating that message passing GNNs are at most as powerful as the Weisfeiler-Lehman test, and that GNNs combined with random node initialization are universal. Nevertheless, the concept of "expressiveness" for GNNs remains vaguely defined. Most existing works adopt the graph isomorphism test as the metric of expressiveness, but this graph-level task may not effectively assess a model's ability in recommendation, where the objective is to distinguish nodes of different closeness. In this paper, we provide a comprehensive theoretical analysis of the expressiveness of GNNs in recommendation, considering three levels of expressiveness metrics: graph isomorphism (graph-level), node automorphism (node-level), and topological closeness (link-level). We propose the topological closeness metric to evaluate GNNs' ability to capture the structural distance between nodes, which closely aligns with the recommendation objective. To validate the effectiveness of this new metric in evaluating recommendation performance, we introduce a learning-less GNN algorithm that is optimal on the new metric and can be optimal on the node-level metric with suitable modification. We conduct extensive experiments comparing the proposed algorithm against various types of state-of-the-art GNN models to explore the effectiveness of the new metric in the recommendation task. For the sake of reproducibility, implementation codes are available at <https://github.com/HKUDS/GTE>.

### **L2R: Lifelong Learning for First-stage Retrieval with Backward-Compatible Representations**

Yinqiong Cai, Keping Bi, Yixing Fan, Jiafeng Guo, Wei Chen, Xueqi Cheng

First-stage retrieval is a critical task that aims to retrieve relevant document candidates from a large-scale collection. While existing retrieval models have achieved impressive performance, they are mostly studied on static data sets, ignoring that in the real-world, the data on the Web is continuously growing with potential distribution drift. Consequently, retrievers trained on static old data may not suit new-coming data well and inevitably produce sub-optimal results. In this work, we study lifelong learning for first-stage retrieval, especially focusing on the setting where the emerging documents are unlabeled since relevance annotation is expensive and may not keep up with data emergence. Under this setting, we aim to develop model updating with two goals: (1) to effectively adapt to the evolving distribution with the unlabeled new-coming data, and (2) to avoid re-inferring all embeddings of old documents to efficiently update the index each time the model is updated.

We first formalize the task and then propose a novel Lifelong Learning method for the first-stage Retrieval, namely L2R. L2R adopts the typical memory mechanism for lifelong learning, and incorporates two crucial components: (1) selecting diverse support negatives for model training and memory updating for effective model adaptation, and (2) a ranking alignment objective to ensure the backward-compatibility of representations to save the cost of index rebuilding without hurting the model performance. For evaluation, we construct two new benchmarks from LoTTE and Multi-CPR datasets to simulate the document distribution drift in realistic retrieval scenarios. Extensive experiments show that L<sup>2</sup>R significantly outperforms competitive lifelong learning baselines.

### **MemDA: Forecasting Urban Time Series with Memory-based Drift Adaptation**

Zekun Cai, Renhe Jiang, Xinyu Yang, Zhaonan Wang, Diansheng Guo, Hill Hiroki Kobayashi, Xuan Song, Ryosuke Shibasaki

Urban time series data forecasting featuring significant contributions to sustainable development is widely studied as an essential task of the smart city. However, with the dramatic and rapid changes in the world environment, the assumption that data obey Independent Identically Distribution is undermined by the subsequent changes in data distribution, known as concept drift, leading to weak replicability and transferability of the model over unseen data. To address the issue, previous approaches typically retrain the model, forcing it to fit the most recent observed data. However, retraining is problematic in that it leads to model lag, consumption of resources, and model re-invalidation, causing the drift problem to be not well solved in realistic scenarios. In this study, we propose a new urban time series prediction model for the concept drift problem, which encodes the drift by considering the periodicity in the data and makes on-the-fly adjustments to the model based on the drift using a meta-dynamic network. Experiments on real-world datasets show that our design significantly outperforms state-of-the-art methods and can be well generalized to existing prediction backbones by reducing their sensitivity to distribution changes.

### **Causality and Independence Enhancement for Biased Node Classification**

Guoxin Chen, Yongqing Wang, Fangda Guo, Qinglang Guo, Jiangli Shao, Huawei Shen, Xueqi Cheng

Most existing methods that address out-of-distribution (OOD) generalization for node classification on graphs primarily focus on a specific type of data biases, such as label selection bias or structural bias. However, anticipating the type of bias in advance is extremely challenging, and designing models solely for one specific type may not necessarily improve overall generalization performance. Moreover, limited research has focused on the impact of mixed biases, which are more prevalent and demanding in real-world scenarios. To address these limitations, we propose a novel Causality and Independence Enhancement (CIE) framework, applicable to various graph neural networks (GNNs). Our approach estimates causal and spurious features at the node representation level and mitigates the influence of spurious correlations through the backdoor adjustment. Meanwhile, independence constraint is introduced to improve the discriminability and stability of causal and spurious features in complex biased environments. Essentially, CIE eliminates different types of data biases from a unified perspective, without the need to design separate methods for each bias as before. To evaluate the performance under specific types of data biases, mixed biases, and low-resource scenarios, we conducted comprehensive experiments on five publicly available datasets. Experimental results demonstrate that our approach CIE not only significantly enhances the performance of GNNs but outperforms state-of-the-art debiased node classification methods.

### **Inducing Causal Structure for Abstractive Text Summarization**

Lu Chen, Ruqing Zhang, Wei Huang, Wei Chen, Jiafeng Guo, Xueqi Cheng

The mainstream of data-driven abstractive summarization models tends to explore the correlations rather than the causal relationships. Among such correlations, there can be spurious ones which suffer from the language prior learned from the training corpus and therefore undermine the overall effectiveness of the learned model. To tackle this issue, we introduce a Structural Causal Model (SCM) to induce the underlying causal structure of the summarization data. We assume several

latent causal factors and non-causal factors, representing the content and style of the document and summary. Theoretically, we prove that the latent factors in our SCM can be identified by fitting the observed training data under certain conditions. On the basis of this, we propose a Causality Inspired Sequence-to-Sequence model (CI-Seq2Seq) to learn the causal representations that can mimic the causal factors, guiding us to pursue causal information for summary generation. The key idea is to reformulate the Variational Auto-encoder (VAE) to fit the joint distribution of the document and summary variables from the training corpus. Experimental results on two widely used text summarization datasets demonstrate the advantages of our approach.

### [Region Profile Enhanced Urban Spatio-Temporal Prediction via Adaptive Meta-Learning](#)

Jie Chen, Tong Liu, Ruiyuan Li

Urban spatio-temporal (ST) prediction plays a crucial role in smart city construction. Due to the high cost of ST data collection, improving ST prediction in a lack of data is significant. For this purpose, existing meta-learning methods have been demonstrated powerful by learning an initial network from training tasks and adjusting to target tasks with limited data. However, such shared knowledge from a set of tasks may contain irrelevant noise due to the gap of region-varying ST dynamics, resulting in the negative transfer issue. As a revelation of regional functional patterns, region profiles give rise to the diversity of ST dynamics. Thus, we design a novel adaptive meta-optimized model MetaRSTP, which conducts the initial prediction model in a finer-granularity of region level with region profiles as semantic evidence. To enhance the expressiveness of profiles, we firstly build a semantic alignment space to explore the inter-view co-semantics. Fusing it with view-specific uniqueness, the multi-view region profiles can be better applied in urban tasks. Then, a regional bias generator derives non-shared parameters in terms of profiles, which alleviates the divergence among regions. We set a new meta-learning strategy as initialize the network with fixed generalizable parameters and region-adaptive bias, thus enhancing the personalized prediction performance even in few-shot scenarios. Extensive experiments on real-world datasets illustrate the effectiveness of our MetaRSTP and our learned region profiles.

### [Knowledge-inspired Subdomain Adaptation for Cross-Domain Knowledge Transfer](#)

Liyue Chen, Linian Wang, Jinyu Xu, Shuai Chen, Weiqiang Wang, Wenbiao Zhao, Qiyu Li, Leye Wang

Most state-of-the-art deep domain adaptation techniques align source and target samples in a global fashion. That is, after alignment, each source sample is expected to become similar to any target sample. However, global alignment may not always be optimal or necessary in practice. For example, consider cross-domain fraud detection, where there are two types of transactions: credit and non-credit. Aligning credit and non-credit transactions separately may yield better performance than global alignment, as credit transactions are unlikely to exhibit patterns similar to non-credit transactions. To enable such fine-grained domain adaption, we propose a novel Knowledge-Inspired Subdomain Adaptation (KISA) framework. In particular, (1) We provide the theoretical insight that KISA minimizes the shared expected loss which is the premise for the success of domain adaptation methods. (2) We propose the knowledge-inspired subdomain division problem that plays a crucial role in fine-grained domain adaption. (3) We design a knowledge fusion network to exploit diverse domain knowledge. Extensive experiments demonstrate that KISA achieves remarkable results on fraud detection and traffic demand prediction tasks.

### [Hallucination Detection: Robustly Discerning Reliable Answers in Large Language Models](#)

Yuyan Chen, Qiang Fu, Yichen Yuan, Zhihao Wen, Ge Fan, Dayiheng Liu, Dongmei Zhang, Zhixu Li, Yanghua Xiao

Large language models (LLMs) have gained widespread adoption in various natural language processing tasks, including question answering and dialogue systems. However, a major drawback of LLMs is the issue of hallucination, where they generate unfaithful or inconsistent content that deviates from the input source, leading to severe consequences. In this paper, we propose a robust discriminator named ReID to effectively detect hallucination in LLMs' generated answers. ReID is trained on the constructed ReQA, a bilingual question-answering dialogue dataset along with answers generated by LLMs and a comprehensive set of metrics. Our experimental results demonstrate that the proposed ReID successfully detects hallucination in the answers generated by diverse LLMs. Moreover, it performs well in distinguishing hallucination in LLMs' generated answers from both in-distribution and out-of-distribution datasets. Additionally, we also conduct a thorough analysis of the types of hallucinations that occur and present valuable insights. This research significantly contributes to the detection of reliable answers generated by LLMs and holds noteworthy implications for mitigating hallucination in the future work.

### [Learning Pair-Centric Representation for Link Sign Prediction with Subgraph](#)



Jushuo Chen, Feifei Dai, Xiaoyan Gu, Haihui Fan, Jiang Zhou, Bo Li, Weiping Wang

Signed graphs are prevalent data structures containing both positive and negative links. Recently, the fundamental network analysis task on signed graphs, namely link sign prediction, has received careful attention. Existing methods learn two target node representations independently, and the sign between these two nodes is predicted based on similarity. However, such a paradigm is node-centric that cannot distinguish node pairs with distinct contexts, thus lowering the prediction performance. Learning pair-centric representation is therefore a rewarding way to be aware of differences between pairs. There is no study yet on how to build such an appropriate representation that can effectively infer the sign between the target node pair. In this paper, we provide a new perspective to conduct link sign prediction within the paradigm of subgraph classification and propose a novel Subgraph-based link Sign Prediction (SSP) model. Technically, SSP uses importance-based sampling to extract an informative subgraph around each target node pair. For each subgraph, an innovative node labeling scheme is designed to encode its structural and signed information for representation learning. To further utilize the subgraph representation for imbalanced sign classification, SSP employs self-pruning contrastive learning to gain balanced representations. Extensive experiments on real-world datasets demonstrate that SSP consistently and significantly outperforms all the state-of-the-art baselines.

#### [Meta-Transfer-Learning for Time Series Data with Extreme Events: An Application to Water Temperature Prediction](#)

Shengyu Chen, Nasrin Kalanat, Simon Topp, Jeffrey Sadler, Yiqun Xie, Zhe Jiang, Xiaowei Jia

This paper proposes a meta-transfer-learning method for predicting daily maximum water temperature in stream networks with explicit modeling of extreme events. Accurate prediction of these extreme events is challenging because of their sparsity in the training data and their distinct responses to external drivers when compared to non-extreme observations. To overcome these challenges, we propose a sample reweighting strategy to escalate the importance of extreme events in the training process while preserving the predictive performance in normal time periods. The sample weight for each training data point is estimated as the similarity with the target test data point using contextual information and physical simulation. The obtained sample weight values are then used to fine-tune the initial model to transfer it to the test data. This method is further enhanced by an extreme value theory-based loss function to enforce the distribution of extreme data points and accelerated by a clustering algorithm based on the estimated similarities. Additionally, we introduce an online learning strategy to further refine the predictive model using newly collected observed data. The experimental results using real stream data from the Delaware River Basin over the past 36 years demonstrate that our meta-transfer-learning method produces more accurate predictions in both normal and extreme time periods when compared to baselines without the sample re-weighting scheme. The similarity learning method can reveal meaningful relationships amongst data points. We also show that the clustering algorithm can be used to accelerate the prediction while not compromising the predictive performance. The online learning strategy is shown to further improve predictive performance using recently observed data.

#### [Hadamard Adapter: An Extreme Parameter-Efficient Adapter Tuning Method for Pre-trained Language Models](#)

Yuyan Chen, Qiang Fu, Ge Fan, Lun Du, Jian-Guang Lou, Shi Han, Dongmei Zhang, Zhixu Li, Yanghua Xiao

Recent years, Pre-trained Language models (PLMs) have swept into various fields of artificial intelligence and achieved great success. However, most PLMs, such as T5 and GPT3, have a huge amount of parameters, fine-tuning them is often expensive and time consuming, and storing them takes up a lot of space. Therefore, it is necessary to adopt a parameter-efficient approach to reduce parameters of PLMs in fine-tuning without compromising their performance in downstream tasks. In this paper, we design a novel adapter which only acts on self-attention outputs in PLMs. This adapter adopts element-wise linear transformation using Hadamard product, hence named as Hadamard adapter, requires the fewest parameters compared to previous parameter-efficient adapters. In addition, we also summarize some tuning patterns for Hadamard adapter shared by various downstream tasks, expecting to provide some guidance for further parameter reduction with shared adapters in future studies. The experiments conducted on the widely-used GLUE benchmark with several SOTA PLMs prove that the Hadamard adapter achieves competitive performance with only 0.033% parameters compared with full fine-tuning, and it has the fewest parameters compared with other adapters. Moreover, we further find that there is also some redundant layers in the Hadamard adapter which can be removed to achieve more parameter efficiency with only 0.022% parameters.

#### [Incorporating Constituent Syntax into Grammatical Error Correction with Multi-Task Learning](#)

Chen Chen, Bo He, Jing Yuan, Chunyan Hou, Xiaojie Yuan

Grammatical Error Correction (GEC) is usually considered as a translation task where an erroneous sentence is treated as the source language and the corrected sentence as the target language. The state-of-the-art GEC models often adopt transformer-based sequence-to-sequence architecture of machine translation. However, most of these approaches ignore the syntactic information because the syntax of an erroneous sentence is also full of errors and not beneficial to GEC. In this paper, we propose a novel Error-Correction Constituent Parsing (ECCP) task which uses the constituent parsing of corrected sentences to avoid the harmful effect of the erroneous sentence. We also propose an architecture that includes one encoder and two decoders. There are millions of parameters in transformer-based GEC models, and the labeled training data is substantially less than synthetic pre-training data. Therefore, adapter layers are added to the proposed architecture, and adapter tuning is used for fine-tuning our model to alleviate the low-resource issue. We conduct experiments on CoNLL-2014, BEA-2019, and JFLEG test datasets in unsupervised and supervised settings. Experimental results show that our method outperforms the-state-of-art baselines and achieves superior performance on all datasets.

### **HEProto: A Hierarchical Enhancing ProtoNet based on Multi-Task Learning for Few-shot Named Entity Recognition**

Wei Chen, Lili Zhao, Pengfei Luo, Tong Xu, Yi Zheng, Enhong Chen

Few-shot Named Entity Recognition (NER) task, which aims to identify and classify entities from different domains with limited training samples, has long been treated as a basic step for knowledge graph (KG) construction. Great efforts have been made on this task with competitive performance, however, they usually treat the two subtasks, namely span detection and type classification, as mutually independent, and the integrity and correlation between subtasks have been largely ignored. Moreover, prior arts may fail to absorb the coarse-grained features of entities, resulting in a semantic-insufficient representation of entity types. To that end, in this paper, we propose a Hierarchical Enhancing ProtoNet (HEProto) based on multi-task learning, which is utilized to jointly learn these two subtasks and model their correlation. Specifically, we adopt contrastive learning to enhance the span boundary information and the type semantic representations in these two subtasks. Then, the hierarchical prototypical network is designed to leverage the coarse-grained information of entities in the type classification stage, which could help the model to better learn the fine-grained semantic representations. Along this line, we construct a similarity margin loss to reduce the similarity between fine-grained entities and other irrelevant coarse-grained prototypes. Finally, extensive experiments on the Few-NERD dataset prove that our solution outperforms competitive baseline methods. The source code of HEProto is available at <https://github.com/fanshu6hao/HEProto>

### **Continual Learning for Generative Retrieval over Dynamic Corpora**

Jiangui Chen, Ruqing Zhang, Jiafeng Guo, Maarten de Rijke, Wei Chen, Yixing Fan, Xueqi Cheng

Generative retrieval (GR) directly predicts the identifiers of relevant documents (i.e., docids) based on a parametric model. It has achieved solid performance on many ad-hoc retrieval tasks. So far, these tasks have assumed a static document collection. In many practical scenarios, however, document collections are dynamic, where new documents are continuously added to the corpus. The ability to incrementally index new documents while preserving the ability to answer queries with both previously and newly indexed relevant documents is vital to applying GR models. In this paper, we address this practical continual learning problem for GR. We put forward a novel Continual-Learner for generative Retrieval (CLEVER) model and make two major contributions to continual learning for GR: (i) To encode new documents into docids with low computational cost, we present Incremental Product Quantization, which updates a partial quantization codebook according to two adaptive thresholds; and (ii) To memorize new documents for querying without forgetting previous knowledge, we propose a memory-augmented learning mechanism, to form meaningful connections between old and new documents. Empirical results demonstrate the effectiveness and efficiency of the proposed model.

### **Deep Generative Imputation Model for Missing Not At Random Data**

Jialei Chen, Yuanbo Xu, Pengyang Wang, Yongjian Yang

Data analysis usually suffers from the Missing Not At Random (MNAR) problem, where the cause of the value missing is not fully observed. Compared to the naive Missing Completely At Random (MCAR) problem, it is more in line with the realistic scenario whereas more complex and challenging. Existing statistical methods model the MNAR mechanism by different decomposition of the joint distribution of the complete data and the missing mask. But we empirically find that directly incorporating these statistical methods into deep generative models is sub-optimal. Specifically, it would neglect the confidence of the reconstructed mask during the MNAR imputation process, which leads to insufficient information extraction and less-guaranteed imputation quality. In this paper, we revisit the MNAR problem from a novel perspective that the complete data and missing mask are two modalities of incomplete data on an equal footing. Along with this line, we put

forward a generative-model-specific joint probability decomposition method, conjunction model, to represent the distributions of two modalities in parallel and extract sufficient information from both complete data and missing mask. Taking a step further, we exploit a deep generative imputation model, namely GNR, to process the real-world missing mechanism in the latent space and concurrently impute the incomplete data and reconstruct the missing mask. The experimental results show that our GNR surpasses state-of-the-art MNAR baselines with significant margins (averagely improved from 9.9% to 18.8% in RMSE) and always gives a better mask reconstruction accuracy which makes the imputation more principle.

#### [\*\*Towards Spoken Language Understanding via Multi-level Multi-grained Contrastive Learning\*\*](#)

Xuxin Cheng, Wanshi Xu, Zhihong Zhu, Hongxiang Li, Yuexian Zou

Spoken language understanding (SLU) is a core task in task-oriented dialogue systems, which aims at understanding user's current goal through constructing semantic frames. SLU usually consists of two subtasks, including intent detection and slot filling. Although there are some SLU frameworks joint modeling the two subtasks and achieve the high performance, most of them still overlook the inherent relationships between intents and slots, and fail to achieve mutual guidance between the two subtasks. To solve the problem, we propose a multi-level multi-grained SLU framework MMCL to apply contrastive learning at three levels, including utterance level, slot level, and word level to enable intent and slot to mutually guide each other. For the utterance level, our framework implements coarse granularity contrastive learning and fine granularity contrastive learning simultaneously. Besides, we also apply the self-distillation method to improve the robustness of the model. Experimental results and further analysis demonstrate that our proposed model achieves new state-of-the-art results on two public multi-intent SLU datasets, obtaining a 2.6 overall accuracy improvement on MixATIS dataset compared to previous best models.

#### [\*\*DAS-CL: Towards Multimodal Machine Translation via Dual-Level Asymmetric Contrastive Learning\*\*](#)

Xuxin Cheng, Zhihong Zhu, Yaowei Li, Hongxiang Li, Yuexian Zou

Multimodal machine translation (MMT) aims to exploit visual information to improve neural machine translation (NMT). It has been demonstrated that image captioning and object detection can further improve MMT. In this paper, to leverage image captioning and object detection more effectively, we propose a Dual-level ASymmetric Contrastive Learning (DAS-CL) framework. Specifically, we leverage image captioning and object detection to generate more pairs of visual inputs and textual inputs. At the utterance level, we introduce an image captioning model to generate more coarse-grained pairs. At the word level, we introduce an object detection model to generate more fine-grained pairs. To mitigate the negative impact of noise in generated pairs, we apply asymmetric contrastive learning at these two levels. Experiments on the Multi30K dataset of three translation directions demonstrate that DAS-CL significantly outperforms existing MMT frameworks and achieves new state-of-the-art performance. More encouragingly, further analysis displays that DAS-CL is more robust to irrelevant visual information.

#### [\*\*PCT-CycleGAN: Paired Complementary Temporal Cycle-Consistent Adversarial Networks for Radar-Based Precipitation Nowcasting\*\*](#)

Jaeho Choi, Yura Kim, Kwang-Ho Kim, Sung-Hwa Jung, Ikhyun Cho

The precipitation nowcasting methods have been elaborated over the centuries because rain has a crucial impact on human life. Not only quantitative precipitation forecast (QPF) models and convolutional long short-term memory (ConvLSTM), but also various sophisticated methods such as the latest MetNet-2 are emerging. In this paper, we propose a paired complementary temporal cycle-consistent adversarial networks (PCT-CycleGAN) for radar-based precipitation nowcasting, inspired by cycle-consistent adversarial networks (CycleGAN), which shows strong performance in image-to-image translation. PCT-CycleGAN generates temporal causality using two generator networks with forward and backward temporal dynamics in paired complementary cycles. Each generator network learns a huge number of one-to-one mappings about time-dependent radar-based precipitation data to approximate a mapping function representing the temporal dynamics in each direction. To create robust temporal causality between paired complementary cycles, novel connection loss is proposed. And torrential loss to cover exceptional heavy rain events is also proposed. The generator network learning forward temporal dynamics in PCT-CycleGAN generates radar-based precipitation data 10 minutes from the current time. Also, it provides a reliable prediction of up to 2 hours with iterative forecasting. The superiority of PCT-CycleGAN is demonstrated through qualitative and quantitative comparisons with several previous methods.

#### [\*\*Delivery Optimized Discovery in Behavioral User Segmentation under Budget Constraint\*\*](#)

Harshita Chopra, Atanu R. Sinha, Sunav Choudhary, Ryan A. Rossi, Paavan Kumar Indela, Veda Pranav Parwatala, Srinjayee Paul, Aurngha Maiti

Users' behavioral footprints online enable firms to discover behavior-based user segments (or, segments) and deliver segment specific messages to users. Following the discovery of segments, delivery of messages to users through preferred media channels like Facebook and Google can be challenging, as only a portion of users in a behavior segment find match in a medium, and only a fraction of those matched actually see the message (exposure). Even high quality discovery becomes futile when delivery fails. Many sophisticated algorithms exist for discovering behavioral segments; however, these ignore the delivery component. The problem is compounded because (i) the discovery is performed on the behavior data space in firms' data (e.g., user clicks), while the delivery is predicated on the static data space (e.g., geo, age) as defined by media; and (ii) firms work under budget constraint. We introduce a stochastic optimization based algorithm for delivery optimized discovery of behavioral user segments and offer new metrics to address the joint optimization. We leverage optimization under a budget constraint for delivery combined with a learning-based component for discovery. Extensive experiments on a public dataset from Google and a proprietary dataset show the effectiveness of our approach by simultaneously improving delivery metrics, reducing budget spend and achieving strong predictive performance in discovery.

### **Rebalancing Social Feed to Minimize Polarization and Disagreement**

Federico Cinus, Aristides Gionis, Francesco Bonchi

Social media have great potential for enabling public discourse on important societal issues. However, adverse effects, such as polarization and echo chambers, greatly impact the benefits of social media and call for algorithms that mitigate these effects. In this paper, we propose a novel problem formulation aimed at slightly nudging users' social feeds in order to strike a balance between relevance and diversity, thus mitigating the emergence of polarization, without lowering the quality of the feed. Our approach is based on re-weighting the relative importance of the accounts that a user follows, so as to calibrate the frequency with which the content produced by various accounts is shown to the user.

We analyze the convexity properties of the problem, demonstrating the non-matrix convexity of the objective function and the convexity of the feasible set. To efficiently address the problem, we develop a scalable algorithm based on projected gradient descent. We also prove that our problem statement is a proper generalization of the undirected-case problem so that our method can also be adopted for undirected social networks. As a baseline for comparison in the undirected case, we develop a semidefinite programming approach, which provides the optimal solution. Through extensive experiments on synthetic and real-world datasets, we validate the effectiveness of our approach, which outperforms non-trivial baselines, underscoring its ability to foster healthier and more cohesive online communities.

### **Can Knowledge Graphs Simplify Text?**

Anthony Colas, Haodi Ma, Xuanli He, Yang Bai, Daisy Zhe Wang

Knowledge Graph (KG)-to-Text Generation has seen recent improvements in generating fluent and informative sentences which describe a given KG. As KGs are widespread across multiple domains and contain important entity-relation information, and as text simplification aims to reduce the complexity of a text while preserving the meaning of the original text, we propose KGSimple, a novel approach to unsupervised text simplification which infuses KG-established techniques in order to construct a simplified KG path and generate a concise text which preserves the original input's meaning. Through an iterative and sampling KG-first approach, our model is capable of simplifying text when starting from a KG by learning to keep important information while harnessing KG-to-text generation to output fluent and descriptive sentences. We evaluate various settings of the KGSimple model on currently-available KG-to-text datasets, demonstrating its effectiveness compared to unsupervised text simplification models which start with a given complex text. Our code is available on GitHub.

### **Dually Enhanced Delayed Feedback Modeling for Streaming Conversion Rate Prediction**

Sunhao Dai, Yuqi Zhou, Jun Xu, Ji-Rong Wen

In online industrial advertising systems, conversion actions (e.g., purchases or downloads) often occur significantly delayed, even up to several days or weeks after the user clicks. This phenomenon leads to the crucial challenge called delayed feedback problem in streaming CVR prediction, that is, the online systems cannot receive the true label of conversions immediately for continuous training. To mitigate the delayed feedback problem, recent state-of-the-art methods often apply sample duplicate mechanisms to introduce early certain conversion information. Nevertheless, these works have overlooked a crucial issue of rapid shifts in data distribution and considered both the newly observed data and duplicated early data together, resulting in biases in both distributions. In this work, we propose a Dually enhanced Delayed Feedback Model (DDFM), which tackles the above issues by treating the newly observed data and duplicated early data separately. DDFM consists of dual unbiased CVR estimators that share the same form but utilize different latent variables as weights: one for

the newly observed data and the other for the duplicated early data. To avoid high variance, we adopt an addition-only formula for these latent variables, eliminating multiplication or division operations. Furthermore, we design a shared-bottom network that efficiently and jointly estimates the latent variables in DDFM. Theoretical analysis demonstrates the unbiasedness and convergence properties of DDFM. Extensive experiments on both public and industrial large-scale real-world datasets exhibit that our proposed DDFM consistently outperforms existing state-of-the-art methods.

#### **NeoMaPy: A Parametric Framework for Reasoning with MAP Inference on Temporal Markov Logic Networks**

Victor David, Raphaël Fournier-S'niehotta, Nicolas Travers

Reasoning on inconsistent and uncertain data is challenging, especially for Knowledge-Graphs (KG) to abide temporal consistency. Our goal is to enhance inference with more general time interval semantics that specify their validity, as regularly found in historical sciences. We propose a new Temporal Markov Logic Networks (TMLN) model which extends the Markov Logic Networks (MLN) model with uncertain temporal facts and rules. Total and partial temporal (in)consistency relations between sets of temporal formulae are examined. We then propose a new Temporal Parametric Semantics (TPS) which allows combining several sub-functions leading to different assessment strategies. Finally, we present the new NeoMaPy tool, to compute the MAP inference on MLNs and TMLNs with several TPS. We compare our performances with state-of-the-art inference tools and exhibit faster and higher quality results.

#### **TPUF: Enhancing Cross-domain Sequential Recommendation via Transferring Pre-trained User Features**

Yujia Ding, Huan Li, Ke Chen, Lidan Shou

Sequential recommendation has long been challenged by data sparsity issues. Most recently, cross-domain sequential recommendation (CDSR) techniques have been proposed to leverage sequential interaction data from other domains. However, accessing raw data from source domains is often restricted due to privacy concerns. To tackle this issue, we introduce TPUF, a novel CDSR model that transfers pre-trained latent user features from the source domain (UFS) instead of the original interaction data. By doing so, TPUF improves recommendation effectiveness while maintaining practicality. TPUF has three functional characteristics: (1) It is a feature mapping-and-aggregation framework that does not impose specific constraints on the nature of pre-trained UFS. (2) It incorporates a temporal feature mapping unit to effectively extract domain-shared information from UFS with temporal information recovered. (3) It additionally employs an adversarial feature alignment unit to align features across domains to combat feature transfer bias. Experimental results on real-world datasets demonstrate that TPUF outperforms other state-of-the-art cross-domain recommendation models and is compatible with multiple UFS types.

#### **Cross-heterogeneity Graph Few-shot Learning**

Pengfei Ding, Yan Wang, Guanfeng Liu

In recent years, heterogeneous graph few-shot learning has been proposed to address the label sparsity issue in heterogeneous graphs (HGs), which contain various types of nodes and edges. The existing methods have achieved good performance by transferring generalized knowledge extracted from rich-labeled classes in source HG(s) to few-labeled classes in a target HG. However, these methods only consider the single-heterogeneity scenario where the source and target HGs share a fixed set of node/edge types, ignoring the more general scenario of cross-heterogeneity, where each HG can have a different and non-fixed set of node/edge types. To this end, we focus on the unexplored cross-heterogeneity scenario and propose a novel model for Cross-heterogeneity Graph Few-shot Learning, namely CGFL. In CGFL, we first extract meta-patterns to capture heterogeneous information and propose a multi-view heterogeneous graph neural network (MHGN) to learn meta-patterns across HGs. Then, we propose a score module to measure the informativeness of labeled samples and determine the transferability of each source HG. Finally, by integrating MHGN and the score module into a meta-learning mechanism, CGFL can effectively transfer generalized knowledge to predict new classes with few-labeled data. Extensive experiments on four real-world datasets have demonstrated the superior performance of CGFL over the state-of-the-art methods.

#### **A Multi-Task Semantic Decomposition Framework with Task-specific Pre-training for Few-Shot NER**

Guanting Dong, Zechen Wang, Jinxu Zhao, Gang Zhao, Daichi Guo, Dayuan Fu, Tingfeng Hui, Chen Zeng, Keqing He, Xuefeng Li, Liwen Wang, Xinyue Cui, Weiran Xu

The objective of few-shot named entity recognition is to identify named entities with limited labeled instances. Previous works have primarily focused on optimizing the traditional token-wise classification framework, while neglecting the exploration of information based on NER data characteristics. To address this issue, we propose a Multi-Task Semantic Decomposition



Framework via Joint Task-specific Pre-training (MSDP) for few-shot NER. Drawing inspiration from demonstration-based and contrastive learning, we introduce two novel pre-training tasks: Demonstration-based Masked Language Modeling (MLM) and Class Contrastive Discrimination. These tasks effectively incorporate entity boundary information and enhance entity representation in Pre-trained Language Models (PLMs). In the downstream main task, we introduce a multi-task joint optimization framework with the semantic decomposing method, which facilitates the model to integrate two different semantic information for entity classification. Experimental results of two few-shot NER benchmarks demonstrate that MSDP consistently outperforms strong baselines by a large margin. Extensive analyses validate the effectiveness and generalization of MSDP.

### **I3 Retriever: Incorporating Implicit Interaction in Pre-trained Language Models for Passage Retrieval**

Qian Dong, Yiding Liu, Qingyao Ai, Haitao Li, Shuaiqiang Wang, Yiqun Liu, Dawei Yin, Shaoping Ma

Passage retrieval is a fundamental task in many information systems, such as web search and question answering, where both efficiency and effectiveness are critical concerns. In recent years, neural retrievers based on pre-trained language models (PLM), such as dual-encoders, have achieved huge success. Yet, studies have found that the performance of dual-encoders are often limited due to the neglecting of the interaction information between queries and candidate passages. Therefore, various interaction paradigms have been proposed to improve the performance of vanilla dual-encoders. Particularly, recent state-of-the-art methods often introduce late-interaction during the model inference process. However, such late-interaction based methods usually bring extensive computation and storage cost on large corpus. Despite their effectiveness, the concern of efficiency and space footprint is still an important factor that limits the application of interaction-based neural retrieval models. To tackle this issue, we Incorporate Implicit Interaction into dual-encoders, and propose I3 retriever. In particular, our implicit interaction paradigm leverages generated pseudo-queries to simulate query-passage interaction, which jointly optimizes with query and passage encoders in an end-to-end manner. It can be fully pre-computed and cached, and its inference process only involves simple dot product operation of the query vector and passage vector, which makes it as efficient as the vanilla dual encoders. We conduct comprehensive experiments on MSMARCO and TREC2019 Deep Learning Datasets, demonstrating the I3 retriever's superiority in terms of both effectiveness and efficiency. Moreover, the proposed implicit interaction is compatible with special pre-training and knowledge distillation for passage retrieval, which brings a new state-of-the-art performance. The codes are available at <https://github.com/Deriq-Qian-Dong/I3-Retriever>.

### **Reveal the Unknown: Out-of-Knowledge-Base Mention Discovery with Entity Linking**

Hang Dong, Jiaoyan Chen, Yuan He, Yinan Liu, Ian Horrocks

Discovering entity mentions that are out of a Knowledge Base (KB) from texts plays a critical role in KB maintenance, but has not yet been fully explored. The current methods are mostly limited to the simple threshold-based approach and feature-based classification, and the datasets for evaluation are relatively rare. We propose BLINKout, a new BERT-based Entity Linking (EL) method which can identify mentions that do not have corresponding KB entities by matching them to a special NIL entity. To better utilize BERT, we propose new techniques including NIL entity representation and classification, with synonym enhancement. We also apply KB Pruning and Versioning strategies to automatically construct out-of-KB datasets from common in-KB EL datasets. Results on five datasets of clinical notes, biomedical publications, and Wikipedia articles in various domains show the advantages of BLINKout over existing methods to identify out-of-KB mentions for the medical ontologies, UMLS, SNOMED CT, and the general KB, WikiData.

### **Optimal Linear Subspace Search: Learning to Construct Fast and High-Quality Schedulers for Diffusion Models**

Zhongjie Duan, Chengyu Wang, Cen Chen, Jun Huang, Weining Qian

In recent years, diffusion models have become the most popular and powerful methods in the field of image synthesis, even rivaling human artists in artistic creativity. However, the key issue currently limiting the application of diffusion models is its extremely slow generation process. Although several methods were proposed to speed up the generation process, there still exists a trade-off between efficiency and quality. In this paper, we first provide a detailed theoretical and empirical analysis of the generation process of the diffusion models based on schedulers. We transform the designing problem of schedulers into the determination of several parameters, and further transform the accelerated generation process into an expansion process of the linear subspace. Based on these analyses, we consequently propose a novel method called Optimal Linear Subspace Search (OLSS), which accelerates the generation process by searching for the optimal approximation process of the complete generation process in the linear subspaces spanned by latent variables. OLSS is able to generate high-quality images with a very small number of steps. To demonstrate the effectiveness of our method, we conduct extensive

comparative experiments on open-source diffusion models. Experimental results show that with a given number of steps, OLSS can significantly improve the quality of generated images. Using an NVIDIA A100 GPU, we make it possible to generate a high-quality image by Stable Diffusion within only one second without other optimization techniques.

### **CLSPRec: Contrastive Learning of Long and Short-term Preferences for Next POI Recommendation**

Chenghua Duan, Wei Fan, Wei Zhou, Hu Liu, Junhao Wen

Next point-of-interest (POI) recommendation optimizes user travel experiences and enhances platform revenues by providing users with potentially appealing next location choices. In recent research, scholars have successfully mined users' general tastes and varying interests by modeling long-term and short-term check-in sequences. However, conventional methods for long and short-term modeling predominantly employ distinct encoders to process long and short-term interaction data independently, with disparities in encoders and data limiting the ultimate performance of these models. Instead, we propose a shared trajectory encoder and a novel Contrastive learning of Long and Short-term Preferences for next POI Recommendation (CLSPRec) model to better utilize the preference similarity among the same users and distinguish different users' travel preferences for more accurate next POI prediction. CLSPRec adopts a masking strategy in long-term sequences to enhance model robustness and further strengthens user representation through short-term sequences. Extensive experiments on three real-world datasets validate the superiority of our model. Our code is publicly available at <https://github.com/Wonderdch/CLSPRec>.

### **Zero-shot Item-based Recommendation via Multi-task Product Knowledge Graph Pre-Training**

Ziwei Fan, Zhiwei Liu, Shelby Heinecke, Jianguo Zhang, Huan Wang, Caiming Xiong, Philip S. Yu

Existing recommender systems face difficulties with zero-shot items, i.e. items that have no historical interactions with users during the training stage. Though recent works extract universal item representation via pre-trained language models (PLMs), they ignore the crucial item relationships. This paper presents a novel paradigm for the Zero-Shot Item-based Recommendation (ZSIR) task, which pre-trains a model on product knowledge graph (PKG) to refine the item features from PLMs. We identify three challenges for pre-training PKG, which are multi-type relations in PKG, semantic divergence between item generic information and relations and domain discrepancy from PKG to downstream ZSIR task. We address the challenges by proposing four pre-training tasks and novel task-oriented adaptation (ToA) layers. Moreover, this paper discusses how to fine-tune the model on new recommendation task such that the ToA layers are adapted to ZSIR task. Comprehensive experiments on 18 markets dataset are conducted to verify the effectiveness of the proposed MPKG model.

### **Batch-Mix Negative Sampling for Learning Recommendation Retrievers**

Yongfu Fan, Jin Chen, Yongquan Jiang, Defu Lian, Fangda Guo, Kai Zheng

Recommendation retrievers commonly retrieve user potentially preferred items from numerous items, where the query and item representation are learned according to the dual encoders with the log-softmax loss. Under real scenarios, the number of items becomes considerably large, making it exceedingly difficult to calculate the partition function with the whole item corpus. Negative sampling, which samples a subset from the item corpus, is widely used to accelerate the model training. Among different samplers, the in-batch sampling is commonly adopted for online recommendation retrievers, which regards the other items within the mini-batch as the negative samples for the given query, owing to its time and memory efficiency. However, the sample selection bias occurs due to the skewed feedback, harming the retrieval quality. In this paper, we propose a negative sampling approach named Batch-Mix Negative Sampling (BMNS), which adopts batch mixing operation to generate additional negatives for model training. Concretely, BMNS first generates new negative items with the sampled mix coefficient from the Beta distribution, after which a tailored correct strategy guided by frequency is designed to match the sampled softmax loss. In this way, the effort of re-encoding items out of the mini-batch is reduced while also improving the representation space of the negative set. The empirical experiments on four real-world datasets demonstrate BMNS is superior to the competitive negative inbatch sampling method.

### **Spatial-Temporal Graph Boosting Networks: Enhancing Spatial-Temporal Graph Neural Networks via Gradient Boosting**

Yujie Fan, Chin-Chia Michael Yeh, Huiyuan Chen, Yan Zheng, Liang Wang, Junpeng Wang, Xin Dai, Zhongfang Zhuang, Wei Zhang

Spatial-temporal graph neural networks (STGNNs) are promising in solving real-world spatial-temporal forecasting problems. Recognizing the inherent sequential relationship of spatial-temporal data, it is natural to explore the integration of boosting training mechanism to further enhance the performance of STGNNs. However, few studies have touched this research area. To bridge this gap, in this work, we propose spatial-temporal graph boosting networks, namely STGBN, which to the best of

our knowledge is the first attempt to leverage gradient boosting for enhancing STGNNs. STGBN follows the general training procedure of conventional gradient boosting, but incorporates two distinctive designs to improve its efficiency in training on spatial-temporal graphs. Specifically, we design an incremental learning strategy that progressively includes spatial-temporal data into training. Additionally, we enforce an identical architecture for the base learner in all boosting iterations with each base learner inheriting from the one in the previous iteration. These designs facilitate rapid convergence of the base learner and expedite the overall training process. The base learner in STGBN is designed as a Transformer sandwich, which consists of two temporal Transformers on the top and bottom and a spatial Transformer in the middle. Structuring them in such a way helps the model capture long-range temporal dynamics, global spatial dependencies, and deep spatial-temporal interactions. We perform extensive spatial-temporal forecasting experiments on four spatial-temporal graph benchmarks. Promising results demonstrate the outstanding performance of STGBN against a wide range of state-of-the-art baseline models.

#### **BOMGraph: Boosting Multi-scenario E-commerce Search with a Unified Graph Neural Network**

Shuai Fan, Jinping Gou, Yang Li, Jiaxing Bai, Chen Lin, Wanxian Guan, Xubin Li, Hongbo Deng, Jian Xu, Bo Zheng

Mobile Taobao Application delivers search services on multiple scenarios that take textual, visual, or product queries. This paper aims to propose a unified graph neural network for these search scenarios to leverage data from multiple scenarios and jointly optimize search performances with less training and maintenance costs. Towards this end, this paper proposes BOMGraph, BOosting Multi-scenario E-commerce Search with a unified Graph neural network. BOMGraph is embodied with several components to address challenges in multi-scenario search. It captures heterogeneous information flow across scenarios by inter-scenario and intra-scenario metapaths. It learns robust item representations by disentangling specific characteristics for different scenarios and encoding common knowledge across scenarios. It alleviates label scarcity and long-tail problems in scenarios with low traffic by contrastive learning with cross-scenario augmentation. BOMGraph has been deployed in production by Alibaba's E-commerce search advertising platform. Both offline evaluations and online A/B tests demonstrate the effectiveness of BOMGraph.

#### **Cognitive-inspired Graph Redundancy Networks for Multi-source Information Fusion**

Yao Fu, Junhong Wan, Junlan Yu, Weihao Jiang, Shiliang Pu

The recent developments in technologies bring not only increasing amount of information but also multiple information sources for Graph Representation Learning. With the success of Graph Neural Networks (GNN), there have been increasing attempts to learn representation of multi-source information leveraging its graph structures. However, existing graph methods basically combine multi-source information with different contribution scores and over-simplify the graph structures based on prior knowledge, which fail to unify complex and conflicting multi-source information. Multisensory Processing theory in cognitive neuroscience reveals human mechanism of learning multi-source information by identifying the redundancy and complementarity. Inspired by that, we propose Graph Redundancy Network (GRN) that: 1). learns a suitable representation space that maximizes multi-source interactions; 2). encodes the redundant and complementary information according to Graph Intersection and Difference of their graph structures; 3). further reinforces and explores the redundant and complementary information through low-pass and high-pass graph filters. The empirical study shows that GRN outperforms existing methods on various tasks.

#### **Cross-Scenario Maneuver Decision with Adaptive Perception for Autonomous Driving**

Yuan Fu, Shuncheng Liu, Yuyang Xia, Fangda Guo, Kai Zheng

Autonomous driving is a rapidly advancing field that promises to revolutionize the transportation industry through an intelligent perception-and-decision paradigm. Despite decades of research, existing methods are limited in adapting to complex scenarios or expanding to unseen situations, which pose significant challenges to the development of autonomous driving. Inspired by the process of human learning to drive, autonomous vehicles can prioritize developing driving capabilities in basic scenarios and then extending the atomic abilities to more complex scenarios. To this end, we proposed a perception-and-decision framework, called ATEND, which consists of an adaptive perception module and a maneuver decision module. Specifically, the perception module based on Variational Autoencoder is proposed to map perceptual data of complex scenarios into basic scenarios. Then the reinforcement learning-based decision module can make high-level decisions in transformed scenarios. Once ATEND learns to drive in basic scenarios, it can achieve safe and efficient driving in real scenarios without additional training. Extensive experiments in different traffic scenarios evidence that the proposed framework advances the state of the art in terms of both macroscopic and microscopic effectiveness.

### [Semantic-aware Node Synthesis for Imbalanced Heterogeneous Information Networks](#)

Xinyi Gao, Wentao Zhang, Tong Chen, Junliang Yu, Hung Quoc Viet Nguyen, Hongzhi Yin

Heterogeneous graph neural networks (HGNNs) have exhibited exceptional efficacy in modeling the complex heterogeneity in heterogeneous information networks (HINs). The critical advantage of HGNNs is their ability to handle diverse node and edge types in HINs by extracting and utilizing the abundant semantic information for effective representation learning. As a widespread phenomenon in many real-world scenarios, the class-imbalance distribution in HINs creates a performance bottleneck for existing HGNNs. Apart from the node imbalance in quantity, the more crucial and distinctive challenge in HINs is semantic imbalance. Minority classes in HINs often lack diverse and sufficient neighbor nodes, resulting in biased and incomplete semantic information. This semantic imbalance further compounds the difficulty of accurately classifying minority nodes, leading to the performance degradation of HGNNs. However, existing remedies are either tailored for non-graph data or designed specifically for homogeneous graphs, thus overlooking the inherent semantic imbalance in HINs. To tackle the imbalance of minority classes and supplement their inadequate semantics, we present the first method for the semantic imbalance problem in imbalanced HINs named Semantic-aware Node Synthesis (SNS). By assessing the influence on minority classes, SNS adaptively selects the heterogeneous neighbor nodes and augments the network with synthetic nodes while preserving the minority semantics. In addition, we introduce two dedicated regularization approaches for HGNNs that explore the inter-type and intra-type information and constrain the representation of synthetic nodes from both semantic and class perspectives to effectively suppress the potential noises from synthetic nodes, facilitating more expressive embeddings for classification. The comprehensive experimental study demonstrates that SNS consistently outperforms existing methods in different benchmark datasets.

### [AI in the Gray: Exploring Moderation Policies in Dialogic Large Language Models vs. Human Answers in Controversial Topics](#)

Vahid Ghafouri, Vibhor Agarwal, Yong Zhang, Nishanth Sastry, Jose Such, Guillermo Suarez-Tangil

The introduction of ChatGPT and the subsequent improvement of Large Language Models (LLMs) have prompted more and more individuals to turn to the use of ChatBots, both for information and assistance with decision-making. However, the information the user is after is often not formulated by these ChatBots objectively enough to be provided with a definite, globally accepted answer.

Controversial topics, such as "religion", "gender identity", "freedom of speech", and "equality", among others, can be a source of conflict as partisan or biased answers can reinforce preconceived notions or promote disinformation. By exposing ChatGPT to such debatable questions, we aim to understand its level of awareness and if existing models are subject to socio-political and/or economic biases. We also aim to explore how AI-generated answers compare to human ones. For exploring this, we use a dataset of a social media platform created for the purpose of debating human-generated claims on polemic subjects among users, dubbed Kialo.

Our results show that while previous versions of ChatGPT have had important issues with controversial topics, more recent versions of ChatGPT (gpt-3.5-turbo) are no longer manifesting significant explicit biases in several knowledge areas. In particular, it is well-moderated regarding economic aspects. However, it still maintains degrees of implicit libertarian leaning toward right-winged ideals which suggest the need for increased moderation from the socio-political point of view. In terms of domain knowledge on controversial topics, with the exception of the "Philosophical" category, ChatGPT is performing well in keeping up with the collective human level of knowledge. Finally, we see that sources of Bing AI have slightly more tendency to the center when compared to human answers. All the analyses we make are generalizable to other types of biases and domains.

### [On the Trade-off between Over-smoothing and Over-squashing in Deep Graph Neural Networks](#)

Jhony H. Giraldo, Konstantinos Skianis, Thierry Bouwmans, Fragkiskos D. Malliaros

Graph Neural Networks (GNNs) have succeeded in various computer science applications, yet deep GNNs underperform their shallow counterparts despite deep learning's success in other domains. Over-smoothing and over-squashing are key challenges when stacking graph convolutional layers, hindering deep representation learning and information propagation from distant nodes. Our work reveals that over-smoothing and over-squashing are intrinsically related to the spectral gap of the graph Laplacian, resulting in an inevitable trade-off between these two issues, as they cannot be alleviated simultaneously. To achieve a suitable compromise, we propose adding and removing edges as a viable approach. We introduce the Stochastic Jost and Liu Curvature Rewiring (SJLR) algorithm, which is computationally efficient and preserves

fundamental properties compared to previous curvature-based methods. Unlike existing approaches, SJLR performs edge addition and removal during GNN training while maintaining the graph unchanged during testing. Comprehensive comparisons demonstrate SJLR's competitive performance in addressing over-smoothing and over-squashing.

### **Homophily-enhanced Structure Learning for Graph Clustering**

Ming Gu, Gaoming Yang, Sheng Zhou, Ning Ma, Jiawei Chen, Qiaoyu Tan, Meihan Liu, Jiajun Bu

Graph clustering is a fundamental task in graph analysis, and recent advances in utilizing graph neural networks (GNNs) have shown impressive results. Despite the success of existing GNN-based graph clustering methods, they often overlook the quality of graph structure, which is inherent in real-world graphs due to their sparse and multifarious nature, leading to subpar performance. Graph structure learning allows refining the input graph by adding missing links and removing spurious connections. However, previous endeavors in graph structure learning have predominantly centered around supervised settings, and cannot be directly applied to our specific clustering tasks due to the absence of ground-truth labels. To bridge the gap, we propose a novel method called homophily-enhanced structure learning for graph clustering (HoLe). Our motivation stems from the observation that subtly enhancing the degree of homophily within the graph structure can significantly improve GNNs and clustering outcomes. To realize this objective, we develop two clustering-oriented structure learning modules, i.e., hierarchical correlation estimation and cluster-aware sparsification. The former module enables a more accurate estimation of pairwise node relationships by leveraging guidance from latent and clustering spaces, while the latter one generates a sparsified structure based on the similarity matrix and clustering assignments. Additionally, we devise a joint optimization approach alternating between training the homophily-enhanced structure learning and GNN-based clustering, thereby enforcing their reciprocal effects. Extensive experiments on seven benchmark datasets of various types and scales, across a range of clustering metrics, demonstrate the superiority of HoLe against state-of-the-art baselines.

### **Hierarchical Meta-Learning with Hyper-Tasks for Few-Shot Learning**

Yunchuan Guan, Yu Liu, Ke Zhou, Junyuan Huang

Meta-learning excels in few-shot learning by extracting shared knowledge from the observed tasks. However, it needs the tasks to adhere to the i.i.d. constraint, which is challenging to achieve due to complex task relationships between data content. Current methods that create tasks in a one-dimensional structure and use meta-learning to learn all tasks flatly struggle with extracting shared knowledge from tasks with overlapping concepts. To address this issue, we propose further constructing tasks from the same environment into hyper-tasks. Since the distributions of hyper-tasks and tasks in a hyper-task can both be approximated as i.i.d. due to further summarization, the meta-learning algorithm can capture shared knowledge more efficiently. Based on the hyper-task, we propose a hierarchical meta-learning paradigm to meta-learn the meta-learning algorithm. The paradigm builds a customized meta-learner for each hyper-task, which makes meta-learners more flexible and expressive. We apply the paradigm to three classic meta-learning algorithms and conduct extensive experiments on public datasets, which confirm the superiority of hierarchical meta-learning in the few-shot learning setting. The code is released at <https://github.com/tuantuange/H-meta-learning>.

### **KG4Ex: An Explainable Knowledge Graph-Based Approach for Exercise Recommendation**

Quanlong Guan, Fang Xiao, Xinghe Cheng, Liangda Fang, Ziliang Chen, Guanliang Chen, Weiqi Luo

Effective exercise recommendation is crucial for guiding students' learning trajectories and fostering their interest in the subject matter. However, the vast exercise resource and the varying learning abilities of individual students pose a significant challenge in selecting appropriate exercise questions. Collaborative filtering-based methods often struggle with recommending suitable exercises, while deep learning-based methods lack explanation, limiting their practical adoption. To address these limitations, this paper proposes KG4Ex, a knowledge graph-based exercise recommendation method. KG4Ex facilitates the matching of diverse students with suitable exercises while providing recommendation reasons. Specifically, we introduce a feature extraction module to represent students' learning states and construct a knowledge graph for exercise recommendation. This knowledge graph comprises three key entities (knowledge concepts, students, and exercises) and their interrelationships, and can be used to recommend suitable exercises. Extensive experiments on three real-world datasets and expert interviews demonstrate the superiority of KG4Ex over existing baseline methods and highlight its strong explainability.

### **Attacking Neural Networks with Neural Networks: Towards Deep Synchronization for Backdoor Attacks**

Zihan Guan, Lichao Sun, Mengnan Du, Ninghao Liu



Backdoor attacks inject poisoned samples into training data, where backdoor triggers are embedded into the model trained on the mixture of poisoned and clean samples. An interesting phenomenon can be observed in the training process: the loss of poisoned samples tends to drop significantly faster than that of clean samples, which we call the early-fitting phenomenon. Early-fitting provides a simple but effective evidence to defend against backdoor attacks, where the poisoned samples can be detected by selecting the samples with the lowest loss values in the early training epochs. Then, two questions naturally arise: (1) What characteristics of poisoned samples cause early-fitting? (2) Does a stronger attack exist which could circumvent the defense methods? To answer the first question, we find that early-fitting could be attributed to a unique property among poisoned samples called synchronization, which depicts the similarity between two samples at different layers of a model. Meanwhile, the degree of synchronization could be controlled based on whether it is captured by shallow or deep layers of the model. Then, we give an affirmative answer to the second question by proposing a new backdoor attack method, Deep Backdoor Attack (DBA), which utilizes deep synchronization to reverse engineer trigger patterns by activating neurons in the deep layer of a base neural network. Experimental results validate our propositions and the effectiveness of DBA. Our code is available at <https://github.com/GuanZihan/Deep-Backdoor-Attack>.

#### **RoCourseNet: Robust Training of a Prediction Aware Recourse Model**

Hangzhi Guo, Feiran Jia, Jinghui Chen, Anna Squicciarini, Amulya Yadav

Counterfactual (CF) explanations for machine learning (ML) models are preferred by end-users, as they explain the predictions of ML models by providing a recourse (or contrastive) case to individuals who are adversely impacted by predicted outcomes. Existing CF explanation methods generate recourses under the assumption that the underlying target ML model remains stationary over time. However, due to commonly occurring distributional shifts in training data, ML models constantly get updated in practice, which might render previously generated recourses invalid and diminish end-users trust in our algorithmic framework. To address this problem, we propose RoCourseNet, a training framework that jointly optimizes predictions and recourses that are robust to future data shifts. This work contains four key contributions: (1) We formulate the robust recourse generation problem as a tri-level optimization problem which consists of two sub-problems: (i) a bi-level problem that finds the worst-case adversarial shift in the training data, and (ii) an outer minimization problem to generate robust recourses against this worst-case shift. (2) We leverage adversarial training to solve this tri-level optimization problem by: (i) proposing a novel virtual data shift (VDS) algorithm to find worst-case shifted ML models via explicitly considering the worst-case data shift in the training dataset, and (ii) a block-wise coordinate descent procedure to optimize for prediction and corresponding robust recourses. (3) We evaluate RoCourseNet's performance on three real-world datasets, and show that RoCourseNet consistently achieves more than 96% robust validity and outperforms state-of-the-art baselines by at least 10% in generating robust CF explanations. (4) Finally, we generalize the RoCourseNet framework to accommodate any parametric post-hoc methods for improving robust validity.

#### **Query-dominant User Interest Network for Large-Scale Search Ranking**

Tong Guo, Xuanping Li, Haitao Yang, Xiao Liang, Yong Yuan, Jingyou Hou, Bingqing Ke, Chao Zhang, Junlin He, Shunyu Zhang, Enyun Yu, Wenwu Ou

Historical behaviors have shown great effect and potential in various prediction tasks, including recommendation and information retrieval. The overall historical behaviors are various but noisy while search behaviors are always sparse. Most existing approaches in personalized search ranking adopt the sparse search behaviors to learn representation with bottleneck, which do not sufficiently exploit the crucial long-term interest. In fact, there is no doubt that user long-term interest is various but noisy for instant search, and how to exploit it well still remains an open problem.

To tackle this problem, in this work, we propose a novel model named Query-dominant user Interest Network (QIN), including two cascade units to filter the raw user behaviors and reweigh the behavior subsequences. Specifically, we propose a relevance search unit (RSU), which aims to search a subsequence relevant to the query first and then search the sub-sequences relevant to the target item. These items are then fed into an attention unit called Fused Attention Unit (FAU). It should be able to calculate attention scores from the ID field and attribute field separately, and then adaptively fuse the item embedding and content embedding based on the user engagement of past period. Extensive experiments and ablation studies on real-world datasets demonstrate the superiority of our model over state-of-the-art methods. The QIN now has been successfully deployed on Kuaishou search, an online video search platform, and obtained 7.6% improvement on CTR.

#### **MGICL: Multi-Grained Interaction Contrastive Learning for Multimodal Named Entity Recognition**

Aibo Guo, Xiang Zhao, Zhen Tan, Weidong Xiao

Multimodal Named Entity Recognition (MNER) aims to combine data from different modalities (e.g. text, images, videos, etc.) for recognition and classification of named entities, which is crucial for constructing Multimodal Knowledge Graphs (MMKGs). However, existing researches suffer from two prominent issues: over-reliance on textual features while neglecting visual features, and the lack of effective reduction of the feature space discrepancy of multimodal data. To overcome these challenges, this paper proposes a Multi-Grained Interaction Contrastive Learning framework for MNER task, namely MGICL. MGICL slices data into different granularities, i.e., sentence level/word token level for text, and image level/object level for image. By utilizing multimodal features with different granularities, the framework enables cross-contrast and narrows down the feature space discrepancy between modalities. Moreover, it facilitates the acquisition of valuable visual features by the text. Additionally, a visual gate control mechanism is introduced to dynamically select relevant visual information, thereby reducing the impact of visual noise. Experimental results demonstrate that the proposed MGICL framework satisfactorily tackles the challenges of MNER through enhancing information interaction of multimodal data and reducing the effect of noise, and hence, effectively improves the performance of MNER.

#### [Targeted Shilling Attacks on GNN-based Recommender Systems](#)

Sihan Guo, Ting Bai, Weihong Deng

GNN-based recommender systems have shown their vulnerability to shilling attacks in recent studies. By conducting shilling attacks on recommender systems, the attackers aim to have homogeneous impacts on all users. However, such indiscriminate attacks suffer from a waste of resources because even if the target item is promoted to users who are not interested, they are unlikely to click on them. In this paper, we conduct targeted shilling attacks in GNN-based recommender systems. By automatically constructing the features and edges of the fake users, our proposed framework AutoAttack achieves accurate attacks on a specific group of users while minimizing the impact on non-target users. Specifically, the features of fake users are generated based on a similarity function, which is optimized according to the features of target users. The structure of fake users is learned by conducting spectral clustering on the target users based on their graph Laplacian matrix, which contains the degree and adjacency information that provides guidance to the edge generation of fake users. We conduct extensive experiments on four real-world datasets in different GNN-based RS and evaluate the performance of our method on the shilling attack and recommendation tasks comprehensively, showing the effectiveness and flexibility of our framework.

#### [Interpretable Fake News Detection with Graph Evidence](#)

Hao Guo, Weixin Zeng, Jiuyang Tang, Xiang Zhao

Automatic detection of fake news has received widespread attentions over recent years. A pile of efforts has been put forward to address the problem with high accuracy, while most of them lack convincing explanations, making it difficult to curb the continued spread of false news in real-life cases. Although some models leverage external resources to provide preliminary interpretability, such external signals are not always available. To fill in this gap, in this work, we put forward an interpretable fake news detection model IKA by making use of the historical evidence in the form of graphs. Specifically, we establish both positive and negative evidence graphs by collecting the signals from the historical news, i.e., training data. Then, given a piece of news to be detected, in addition to the common features used for detecting false news, we compare the news and evidence graphs to generate both the matching vector and the related graph evidence for explaining the prediction. We conduct extensive experiments on both Chinese and English datasets. The experiment results show that the detection accuracy of IKA exceeds the state-of-the-art approaches and IKA can provide useful explanations for the prediction results. Besides, IKA is general and can be applied on other models to improve their interpretability.

#### [Towards Fair Graph Neural Networks via Graph Counterfactual](#)

Zhimeng Guo, Jialiang Li, Teng Xiao, Yao Ma, Suhang Wang

Graph neural networks have shown great ability in representation (GNNs) learning on graphs, facilitating various tasks. Despite their great performance in modeling graphs, recent works show that GNNs tend to inherit and amplify the bias from training data, causing concerns of the adoption of GNNs in high-stake scenarios. Hence, many efforts have been taken for fairness-aware GNNs. However, most existing fair GNNs learn fair node representations by adopting statistical fairness notions, which may fail to alleviate bias in the presence of statistical anomalies. Motivated by causal theory, there are several attempts utilizing graph counterfactual fairness to mitigate root causes of unfairness. However, these methods suffer from non-realistic counterfactuals obtained by perturbation or generation. In this paper, we take a causal view on fair graph learning problem. Guided by the casual analysis, we propose a novel framework CAF, which can select counterfactuals from training data to avoid non-realistic counterfactuals and adopt selected counterfactuals to learn fair node representations for

node classification task. Extensive experiments on synthetic and real-world datasets show the effectiveness of CAF. Our code is available at <https://github.com/TimeLovercc/CAF-GNN>.

### **James ate 5 oranges = Steve bought 5 pencils: Structure-Aware Denoising for Paraphrasing Word Problems**

Rishabh Gupta, Venktesh V, Mukesh Mohania, Vikram Goyal

We propose SCANING, an unsupervised framework for paraphrasing via controlled noise injection. We focus on the novel task of paraphrasing algebraic word problems having practical applications in online pedagogy as a means to reduce plagiarism as well as evoke reasoning capabilities on the part of the student instead of rote memorization. This task is more complex than paraphrasing general-domain corpora due to the difficulty in preserving critical information for solution consistency of the paraphrased word problem, managing the increased length of the text and ensuring diversity in the generated paraphrase. Existing approaches fail to demonstrate adequate performance on at least one, if not all, of these facets, necessitating the need for a more comprehensive solution. To this end, we model the noising search space as a composition of contextual and syntactic aspects to sample noising functions. This allows for learning a denoising function, that operates over both aspects and produces semantically equivalent and syntactically diverse outputs through grounded noise injection. The denoising function serves as a foundation for training a paraphrasing function, which operates solely in the input-paraphrase space without carrying any direct dependency on noise. We demonstrate that SCANING improves performance in terms of producing semantically equivalent and syntactically diverse paraphrases by 35% through extensive automated and human evaluation across 4 datasets.

### **Enhancing Spatio-temporal Traffic Prediction through Urban Human Activity Analysis**

Sumin Han, Youngjun Park, Minji Lee, Jisun An, Dongman Lee

Traffic prediction is one of the key elements to ensure the safety and convenience of citizens. Existing traffic prediction models primarily focus on deep learning architectures to capture spatial and temporal correlation. They often overlook the underlying nature of traffic. Specifically, the sensor networks in most traffic datasets do not accurately represent the actual road network exploited by vehicles, failing to provide insights into the traffic patterns in urban activities. To overcome these limitations, we propose an improved traffic prediction method based on graph convolution deep learning algorithms. We leverage human activity frequency data from National Household Travel Survey to enhance the inference capability of a causal relationship between activity and traffic patterns. Despite making minimal modifications to the conventional graph convolutional recurrent networks and graph convolutional transformer architectures, our approach achieves state-of-the-art performance without introducing excessive computational overhead.

### **On Root Cause Localization and Anomaly Mitigation through Causal Inference**

Xiao Han, Lu Zhang, Yongkai Wu, Shuhan Yuan

Due to a wide spectrum of applications in the real world, such as security, financial surveillance, and health risk, various deep anomaly detection models have been proposed and achieved state-of-the-art performance. However, besides being effective, in practice, the practitioners would further like to know what causes the abnormal outcome and how to further fix it. In this work, we propose RootCLAM, which aims to achieve Root Cause Localization and Anomaly Mitigation from a causal perspective. Especially, we formulate anomalies caused by external interventions on the normal causal mechanism and aim to locate the abnormal features with external interventions as root causes. After that, we further propose an anomaly mitigation approach that aims to recommend mitigation actions on abnormal features to revert the abnormal outcomes such that the counterfactuals guided by the causal mechanism are normal. Experiments on three datasets show that our approach can locate the root causes and further flip the abnormal labels.

### **Robust Basket Recommendation via Noise-tolerated Graph Contrastive Learning**

Xinrui He, Tianxin Wei, Jingrui He

The growth of e-commerce has seen a surge in popularity of platforms like Amazon, eBay, and Taobao. This has given rise to a unique shopping behavior involving baskets - sets of items purchased together. As a less studied interaction mode in the community, the question of how should shopping basket complement personalized recommendation systems remains under-explored. While previous attempts focused on jointly modeling user purchases and baskets, the distinct semantic nature of these elements can introduce noise when directly integrated. This noise negatively impacts the model's performance, further exacerbated by significant noise (e.g., a user is misled to click an item or recognizes it as uninteresting after consuming it) within both user and basket behaviors. In order to cope with the above difficulties, we propose a novel Basket recommendation framework via Noise-tolerated Contrastive Learning, named BNCL, to handle the noise existing in the

cross-behavior integration and within-behavior modeling. First, we represent the basket-item interactions as the hypergraph to model the complex basket behavior, where all items appearing in the same basket are treated as a single hyperedge. Second, cross-behavior contrastive learning is designed to suppress the noise during the fusion of diverse behaviors. Next, to further inhibit the within-behavior noise of the user and basket interactions, we propose to exploit invariant properties of the recommenders w.r.t augmentations through within-behavior contrastive learning. A novel consistency-aware augmentation approach is further designed to better identify the noisy interactions with the consideration of the above two types of interactions. Our framework BNCL offers a generic training paradigm that is applicable to different backbones. Extensive experiments on three shopping transaction datasets verify the effectiveness of our proposed method.

### [Large Language Models as Zero-Shot Conversational Recommenders](#)

Zhankui He, Zhouhang Xie, Rahul Jha, Harald Steck, Dawen Liang, Yesu Feng, Bodhisattwa Prasad Majumder, Nathan Kallus, Julian McAuley

In this paper, we present empirical studies on conversational recommendation tasks using representative large language models in a zero-shot setting with three primary contributions. (1) Data: To gain insights into model behavior in "in-the-wild" conversational recommendation scenarios, we construct a new dataset of recommendation-related conversations by scraping a popular discussion website. This is the largest public real-world conversational recommendation dataset to date. (2) Evaluation: On the new dataset and two existing conversational recommendation datasets, we observe that even without fine-tuning, large language models can outperform existing fine-tuned conversational recommendation models. (3) Analysis: We propose various probing tasks to investigate the mechanisms behind the remarkable performance of large language models in conversational recommendation. We analyze both the large language models' behaviors and the characteristics of the datasets, providing a holistic understanding of the models' effectiveness, limitations and suggesting directions for the design of future conversational recommenders.

### [Understanding User Immersion in Online Short Video Interaction](#)

Zhiyu He, Shaorun Zhang, Peijie Sun, Jiayu Li, Xiaohui Xie, Min Zhang, Yiqun Liu

Short video~(SV) online streaming has been one of the most popular Internet applications in recent years. When browsing SVs, users gradually immerse themselves and derive relaxation or knowledge. Whereas prolonged browsing will lead to a decline in positive feelings, users continue due to inertia, resulting in decreased satisfaction. Immersion is shown to be an essential factor for users' positive experience and highly related to users' interactions in film, games, and virtual reality. However, immersion in SV interaction is still unexplored, which differs from the previously studied scenarios essentially because SV delivery is fragmented, discrete, and with limited time for each video.

In this paper, we aim to make an extensive understanding of user immersion in online short video interaction, include related factors, detecting possibility, and satisfaction representation. We conduct a three-step user study on real SV browsing, including an online survey, a field study, and a lab study with EEG signals. The user study reveals that immersion is a common feeling in SV interaction, and it is related to video features, personalization of recommendations, user mood, and interaction behaviors. Specifically, prolonged browsing leads to a significant decrease in immersion. Furthermore, analyses of EEG signals demonstrate that the prefrontal lobe and parietal lobe of the gamma band are associated with immersion. Besides, immersion prediction experiments achieve encouraging results, showing that user immersion status is predictable and EEG signals do help improve prediction performance. Moreover, correlation analysis indicates that the predicted immersion is more representative of user satisfaction than user behaviors, revealing the potential of immersion as an indicator of satisfaction in the recommender system. To the best of our knowledge, it is the first study on user immersion in real online SV interaction scenarios, and our findings are enlightening for SV users and recommender system designers.

### [Dynamic Embedding Size Search with Minimum Regret for Streaming Recommender System](#)

Bowei He, Xu He, Renrui Zhang, Yingxue Zhang, Ruiming Tang, Chen Ma

With the continuous increase of users and items, conventional recommender systems trained on static datasets can hardly adapt to changing environments. The high-throughput data requires the model to be updated in a timely manner for capturing the user interest dynamics, which leads to the emergence of streaming recommender systems. Due to the prevalence of deep learning-based recommender systems, the embedding layer is widely adopted to represent the characteristics of users, items, and other features in low-dimensional vectors. However, it has been proved that setting an identical and static embedding size is sub-optimal in terms of recommendation performance and memory cost, especially for streaming recommendations. To tackle this problem, we first rethink the streaming model update process and model the dynamic

embedding size search as a bandit problem. Then, we analyze and quantify the factors that influence the optimal embedding sizes from the statistics perspective. Based on this, we propose the Dynamic Embedding Size Search (DESS) method to minimize the embedding size selection regret on both user and item sides in a non-stationary manner. Theoretically, we obtain a sublinear regret upper bound superior to previous methods. Empirical results across two recommendation tasks on four public datasets also demonstrate that our approach can achieve better streaming recommendation performance with lower memory cost and higher time efficiency.

### **Designing and Evaluating Presentation Strategies for Fact-Checked Content**

Danula Hettiachchi, Kaixin Ji, Jenny Kennedy, Anthony McCosker, Flora D. Salim, Mark Sanderson, Falk Scholer, Damiano Spina

With the rapid growth of online misinformation, it is crucial to have reliable fact-checking methods. Recent research on finding check-worthy claims and automated fact-checking have made significant advancements. However, limited guidance exists regarding the presentation of fact-checked content to effectively convey verified information to users. We address this research gap by exploring the critical design elements in fact-checking reports and investigating whether credibility and presentation-based design improvements can enhance users' ability to interpret the report accurately. We co-developed potential content presentation strategies through a workshop involving fact-checking professionals, communication experts, and researchers. The workshop examined the significance and utility of elements such as veracity indicators and explored the feasibility of incorporating interactive components for enhanced information disclosure. Building on the workshop outcomes, we conducted an online experiment involving 76 crowd workers to assess the efficacy of different design strategies. The results indicate that proposed strategies significantly improve users' ability to accurately interpret the verdict of fact-checking articles. Our findings underscore the critical role of effective presentation of fact reports in addressing the spread of misinformation. By adopting appropriate design enhancements, the effectiveness of fact-checking reports can be maximized, enabling users to make informed judgments.

### **Predictive Uncertainty-based Bias Mitigation in Ranking**

Maria Heuss, Daniel Cohen, Masoud Mansoury, Maarten de Rijke, Carsten Eickhoff

Societal biases that are contained in retrieved documents have received increased interest. Such biases, which are often prevalent in the training data and learned by the model, can cause societal harms, by misrepresenting certain groups, and by enforcing stereotypes. Mitigating such biases demands algorithms that balance the trade-off between maximized utility for the user with fairness objectives, which incentivize unbiased rankings. Prior work on bias mitigation often assumes that ranking scores, which correspond to the utility that a document holds for a user, can be accurately determined. In reality, there is always a degree of uncertainty in the estimate of expected document utility. This uncertainty can be approximated by viewing ranking models through a Bayesian perspective, where the standard deterministic score becomes a distribution.

In this work, we investigate whether uncertainty estimates can be used to decrease the amount of bias in the ranked results, while minimizing loss in measured utility. We introduce a simple method that uses the uncertainty of the ranking scores for an uncertainty-aware, post hoc approach to bias mitigation. We compare our proposed method with existing baselines for bias mitigation with respect to the utility-fairness trade-off, the controllability of methods, and computational costs. We show that an uncertainty-based approach can provide an intuitive and flexible trade-off that outperforms all baselines without additional training requirements, allowing for the post hoc use of this approach on top of arbitrary retrieval models.

### **Search-Efficient Computerized Adaptive Testing**

Yuting Hong, Shiwei Tong, Wei Huang, Yan Zhuang, Qi Liu, Enhong Chen, Xin Li, Yuanjing He

Computerized Adaptive Testing (CAT) arises as a promising personalized test mode in online education, targeting at revealing students' latent knowledge state by selecting test items adaptively. The item selection strategy is the core component of CAT, which searches for the best suitable test item based on students' current estimated ability at each test step. However, existing selection strategies behave in a brute-force manner, which results in the time complexity being linear to the number of items ( $N$ ) in the item pool, i.e.,  $O(N)$ . Thus, in reality, the search latency becomes the bottleneck for CAT with a large-scale item pool. To this end, we propose a Search-Efficient Computerized Adaptive Testing framework (SECAT), which aims at enhancing CAT with an efficient selection strategy. Specifically, SECAT contains two main phases: item pool indexing and item search. In the item pool indexing phase, we apply a student-aware spatial partition method on the item pool to divide the test items into many sub-spaces, considering the adaptability of test items. In the item search phase, we optimize the traditional single-round search strategy with the asymptotic theory and propose a multi-round search strategy



that can further improve the time efficiency. Compared with existing strategies, the time complexity of SECAT decreases from  $O(N)$  to  $O(\log N)$ . Across two real-world datasets, SECAT achieves over 200x speed up with negligible accuracy degradation.

### **SANN: Programming Code Representation Using Attention Neural Network with Optimized Subtree Extraction**

Muntasir Hoq, Sushanth Reddy Chilla, Melika Ahmadi Ranjbar, Peter Brusilovsky, Bitu Akram

Automated analysis of programming data using code representation methods offers valuable services for programmers, from code completion to clone detection to bug detection. Recent studies show the effectiveness of Abstract Syntax Trees (AST), pre-trained Transformer-based models, and graph-based embeddings in programming code representation. However, pre-trained large language models lack interpretability, while other embedding-based approaches struggle with extracting important information from large ASTs. This study proposes a novel Subtree-based Attention Neural Network (SANN) to address these gaps by integrating different components: an optimized sequential subtree extraction process using Genetic algorithm optimization, a two-way embedding approach, and an attention network. We investigate the effectiveness of SANN by applying it to two different tasks: program correctness prediction and algorithm detection on two educational datasets containing both small and large-scale code snippets written in Java and C, respectively. The experimental results show SANN's competitive performance against baseline models from the literature, including code2vec, ASTNN, TBCNN, CodeBERT, GPT-2, and MVG, regarding accurate predictive power. Finally, a case study is presented to show the interpretability of our model prediction and its application for an important human-centered computing application, student modeling. Our results indicate the effectiveness of the SANN model in capturing important syntactic and semantic information from students' code, allowing the construction of accurate student models, which serve as the foundation for generating adaptive instructional support such as individualized hints and feedback.

### **Celebrity-aware Graph Contrastive Learning Framework for Social Recommendation**

Zheng Hu, Satoshi Nakagawa, Liang Luo, Yu Gu, Fuji Ren

Social networks exhibit a distinct "celebrity effect" whereby influential individuals have a more significant impact on others compared to ordinary individuals, unlike other network structures such as citation networks and knowledge graphs. Despite its common occurrence in social networks, the celebrity effect is frequently overlooked by existing social recommendation methods when modeling social relationships, thereby hindering the full exploitation of social networks to mine similarities between users. In this paper, we fill this gap and propose a Celebrity-aware Graph Contrastive Learning Framework for Social Recommendation (CGCL), which explicitly models the celebrity effect in the social domain. Technically, we measure the different influences of celebrity and ordinary nodes by mining social network structure features, such as closeness centrality. To model the celebrity effect in social networks, we design a novel user-user impact-aware aggregation method, which incorporates the celebrity-aware influence information into the message propagation process. Additionally, we design a graph neural network-based framework which incorporates social semantics into the user-item interaction modeling with contrastive learning-enhanced data augmentation. The experimental results on three real-world datasets show the effectiveness of the proposed framework. We conduct ablation experiments to prove that the key components of our model benefit the recommendation performance improvement.

### **HyperFormer: Enhancing Entity and Relation Interaction for Hyper-Relational Knowledge Graph Completion**

Zhiwei Hu, Víctor Gutiérrez-Basulto, Zhiliang Xiang, Ru Li, Jeff Z. Pan

Hyper-relational knowledge graphs (HKGs) extend standard knowledge graphs by associating attribute-value qualifiers to triples, which effectively represent additional fine-grained information about its associated triple. Hyper-relational knowledge graph completion (HKGC) aims at inferring unknown triples while considering its qualifiers. Most existing approaches to HKGC exploit a global-level graph structure to encode hyper-relational knowledge into the graph convolution message passing process. However, the addition of multi-hop information might bring noise into the triple prediction process. To address this problem, we propose HyperFormer, a model that considers local-level sequential information, which encodes the content of the entities, relations and qualifiers of a triple. More precisely, HyperFormer is composed of three different modules: an entity neighbor aggregator module allowing to integrate the information of the neighbors of an entity to capture different perspectives of it; a relation qualifier aggregator module to integrate hyper-relational knowledge into the corresponding relation to refine the representation of relational content; a convolution-based bidirectional interaction module based on a convolutional operation, capturing pairwise bidirectional interactions of entity-relation, entity-qualifier, and relation-qualifier. Furthermore, we introduce a Mixture-of-Experts strategy into the feed-forward layers of HyperFormer to strengthen its representation capabilities while reducing the amount of model parameters and computation. Extensive experiments on

three well-known datasets with four different conditions demonstrate HyperFormer's effectiveness. Datasets and code are available at <https://github.com/zhiwei1103/HKGC-HyperFormer>.

### **Enhanced Template-Free Reaction Prediction with Molecular Graphs and Sequence-based Data Augmentation**

Haozhe Hu, Yongquan Jiang, Yan Yang, Jim X. Chen

Retrosynthesis and forward synthesis prediction are fundamental challenges in organic synthesis, computer-aided synthesis planning (CASP), and computer-aided drug design (CADD). The objective is to predict plausible reactants for a given target product and its corresponding inverse task. With the rapid development of deep learning, numerous approaches have been proposed to solve this problem from various perspectives. The methods based on molecular graphs benefit from their rich features embedded inside but face difficulties in applying existing sequence-based data augmentations due to the permutation invariance of graph structures. In this work, we propose SeqAGraph, a template-free approach that annotates input graphs with its root atom index to ensure compatibility with sequence-based data augmentation. The matrix product for global attention in graph encoders is implemented by indexing, elementwise product, and aggregation to fuse global attention with local message passing without graph padding. Experiments demonstrate that SeqAGraph fully benefits from molecular graphs and sequence-based data augmentation and achieves state-of-the-art accuracy in template-free approaches.

### **Independent Distribution Regularization for Private Graph Embedding**

Qi Hu, Yangqiu Song

Learning graph embeddings is a crucial task in graph mining tasks. An effective graph embedding model can learn low-dimensional representations from graph-structured data for data publishing benefiting various downstream applications such as node classification, link prediction, etc. However, recent studies have revealed that graph embeddings are susceptible to attribute inference attacks, which allow attackers to infer private node attributes from the learned graph embeddings. To address these concerns, privacy-preserving graph embedding methods have emerged, aiming to simultaneously consider primary learning and privacy protection through adversarial learning. However, most existing methods assume that representation models have access to all sensitive attributes in advance during the training stage, which is not always the case due to diverse privacy preferences. Furthermore, the commonly used adversarial learning technique in privacy-preserving representation learning suffers from unstable training issues. In this paper, we propose a novel approach called Private Variational Graph AutoEncoders (PVGAE) with the aid of independent distribution penalty as a regularization term. Specifically, we split the original variational graph autoencoder (VGAE) to learn sensitive and non-sensitive latent representations using two sets of encoders. Additionally, we introduce a novel regularization to enforce the independence of the encoders. We prove the theoretical effectiveness of regularization from the perspective of mutual information. Experimental results on three real-world datasets demonstrate that PVGAE outperforms other baselines in private embedding learning regarding utility performance and privacy protection.

### **Liberate Pseudo Labels from Over-Dependence: Label Information Migration on Sparsely Labeled Graphs**

Zhihui Hu, Yao Fu, Hong Zhao, Xiaoyu Cai, Weihao Jiang, Shiliang Pu

Graph Convolutional Networks (GCNs) have made outstanding achievements in many tasks on graphs in recent years, but their success relies on sufficient training data. In practice, sparsely labeled graphs widely exist in the real world so self-training methods have become popular approaches by adding pseudo labeled nodes to enhance the performance of GCNs. However, we observe that most selected high-confidence pseudo labeled nodes by the existing methods would surround the true labeled nodes. It is what we called pseudo label over-dependence, which could lead to the non-uniform pseudo label distribution. Furthermore, a thorough experiment shows that the classification accuracy changes significantly under different label densities and the label-sparse regions show great potential improvement in the model performance. Based on the above findings, we theoretically analyze the constraint factors in the label-sparse regions and further propose reducing the feature distribution difference between the label-dense regions and label-sparse regions can effectively decrease the classification error. Thus, in this paper, we propose a novel Graph Label Information Migration framework (GLIM) to liberate pseudo labels from over-dependence. Specifically, we first propose a training dynamics module (TDM) that uses abundant training process information to find more reliable node labels and improve the model robustness against label noise. Then we propose a label migration module (LMM) that migrates label information from label-dense regions to label-sparse regions by a spectral based graph matching algorithm. These migrated labels are like the glimmers in the darkness, providing the supervision signals for the unlabeled nodes in label-sparse regions. Finally, we conduct extensive experiments to demonstrate the effectiveness of the proposed GLIM.

### **Adaptive Multi-Modalities Fusion in Sequential Recommendation Systems**

Hengchang Hu, Wei Guo, Yong Liu, Min-Yen Kan

In sequential recommendation, multi-modal information (e.g., text or image) can provide a more comprehensive view of an item's profile. The optimal stage (early or late) to fuse modality features into item representations is still debated. We propose a graph-based approach (named MMSR) to fuse modality features in an adaptive order, enabling each modality to prioritize either its inherent sequential nature or its interplay with other modalities. MMSR represents each user's history as a graph, where the modality features of each item in a user's history sequence are denoted by cross-linked nodes. The edges between homogeneous nodes represent intra-modality sequential relationships, and the ones between heterogeneous nodes represent inter-modality interdependence relationships. During graph propagation, MMSR incorporates dual attention, differentiating homogeneous and heterogeneous neighbors. To adaptively assign nodes with distinct fusion orders, MMSR allows each node's representation to be asynchronously updated through an update gate. In scenarios where modalities exhibit stronger sequential relationships, the update gate prioritizes updates among homogeneous nodes. Conversely, when the interdependent relationships between modalities are more pronounced, the update gate prioritizes updates among heterogeneous nodes. Consequently, MMSR establishes a fusion order that spans a spectrum from early to late modality fusion. In experiments across six datasets, MMSR consistently outperforms state-of-the-art models, and our graph propagation methods surpass other graph neural networks. Additionally, MMSR naturally manages missing modalities. The code is available at: <https://github.com/HoldenHu/MMSR>.

### **STAMINA (Spatial-Temporal Aligned Meteorological INformation Attention) and FPL (Focal Precip Loss): Advancements in Precipitation Nowcasting for Heavy Rainfall Events**

Ping-Chia Huang, Yueh-Li Chen, Yi-Syuan Liou, Bing-Chen Tsai, Chun-Chieh Wu, Winston H. Hsu

Precipitation nowcasting is crucial for weather-dependent decision-making in various sectors, providing accurate and high-resolution predictions of precipitation within a typical two-hour timeframe. Deep learning techniques have shown promise in improving nowcasting accuracy by leveraging large radar datasets. However, accurately predicting heavy rainfall events remains challenging due to several persistent problems in previous work. These include spatial-temporal misalignment between meteorological information and precipitation data, as well as the performance gap between different rainfall levels. To address these challenges, we propose two innovative modules: Spatial-Temporal Aligned Meteorological INformation Attention (STAMINA) and Focal Precip Loss (FPL). STAMINA integrates meteorological information using spatial-temporal embedding and pixelwise linear attention mechanisms to overcome spatial-temporal misalignment. FPL addresses event imbalance through event weighting and a penalty mechanism. Through extensive experiments, we demonstrate significant performance improvements achieved by STAMINA and FPL, with an 8% improvement in predicting light rainfall and, more significantly, a 30% improvement in heavy rainfall compared to the state-of-the-art DGMR model. These modules offer practical and effective solutions for enhancing nowcasting accuracy, with a specific focus on improving predictions for heavy rainfall events. By tackling the persistent problems in previous work, our proposed approach represents a significant advancement in the field of precipitation nowcasting.

### **Single-User Injection for Invisible Shilling Attack against Recommender Systems**

Chengzhi Huang, Hui Li

Recommendation systems (RS) are crucial for alleviating the information overload problem. Due to its pivotal role in guiding users to make decisions, unscrupulous parties are lured to launch attacks against RS to affect the decisions of normal users and gain illegal profits. Among various types of attacks, shilling attack is one of the most subsistent and profitable attacks. In shilling attack, an adversarial party injects a number of well-designed fake user profiles into the system to mislead RS so that the attack goal can be achieved. Although existing shilling attack methods have achieved promising results, they all adopt the attack paradigm of multi-user injection, where some fake user profiles are required. This paper provides the first study of shilling attack in an extremely limited scenario: only one fake user profile is injected into the victim RS to launch shilling attacks (i.e., single-user injection). We propose a novel single-user injection method SUI-Attack for invisible shilling attack. SUI-Attack is a graph based attack method that models shilling attack as a node generation task over the user-item bipartite graph of the victim RS, and it constructs the fake user profile by generating user features and edges that link the fake user to items. Extensive experiments demonstrate that SUI-Attack can achieve promising attack results in single-user injection. In addition to its attack power, SUI-Attack increases the stealthiness of shilling attack and reduces the risk of being detected. We provide our implementation at: <https://github.com/KDEGroup/SUI-Attack>.

### **Spans, Not Tokens: A Span-Centric Model for Multi-Span Reading Comprehension**

Zixian Huang, Jiaying Zhou, Chenxu Niu, Gong Cheng

Many questions should be answered by not a single answer but a set of multiple answers. This emerging Multi-Span Reading Comprehension (MSRC) task requires extracting multiple non-contiguous spans from a given context to answer a question. Existing methods extend conventional single-span models to predict the positions of the start and end tokens of answer spans, or predict the beginning-inside-outside tag of each token. Such token-centric paradigms can hardly capture dependencies among span-level answers which are critical to MSRC. In this paper, we propose SpanQualifier, a span-centric scheme where spans, as opposed to tokens, are directly represented and scored to qualify as answers. Explicit span representations enable their interaction which exploits their dependencies to enhance representations. Experiments on three MSRC datasets demonstrate the effectiveness of our span-centric scheme and show that SpanQualifier achieves state-of-the-art results.

#### **Safe-NORA: Safe Reinforcement Learning-based Mobile Network Resource Allocation for Diverse User Demands**

Wenzhen Huang, Tong Li, Yuting Cao, Zhe Lyu, Yanping Liang, Li Yu, Depeng Jin, Junge Zhang, Yong Li

As mobile communication technologies advance, mobile networks become increasingly complex, and user requirements become increasingly diverse. To satisfy the diverse demands of users while improving the overall performance of the network system, the limited wireless network resources should be efficiently and dynamically allocated to them based on the magnitude of their demands and their relative location to the base stations. We separated the problem into four constrained subproblems, which we then solved using a safe reinforcement learning method. In addition, we design a reward mechanism to encourage agent cooperation in distributed training environments. We test our methodology in a simulated scenario with thousands of users and hundreds of base stations. According to experimental findings, our method guarantees that over 95% of user demands are satisfied while also maximizing the overall system throughput.

#### **Deep Variational Bayesian Modeling of Haze Degradation Process**

Eun Woo Im, Junsung Shin, Sungyong Baik, Tae Hyun Kim

Relying on the representation power of neural networks, most recent works have often neglected several factors involved in haze degradation, such as transmission (the amount of light reaching an observer from a scene over distance) and atmospheric light. These factors are generally unknown, making dehazing problems ill-posed and creating inherent uncertainties. To account for such uncertainties and factors involved in haze degradation, we introduce a variational Bayesian framework for single image dehazing. We propose to take not only a clean image and but also transmission map as latent variables, the posterior distributions of which are parameterized by corresponding neural networks: dehazing and transmission networks, respectively. Based on a physical model for haze degradation, our variational Bayesian framework leads to a new objective function that encourages the cooperation between them, facilitating the joint training of and thereby boosting the performance of each other. In our framework, a dehazing network can estimate a clean image independently of a transmission map estimation during inference, introducing no overhead. Furthermore, our model-agnostic framework can be seamlessly incorporated with other existing dehazing networks, greatly enhancing the performance consistently across datasets and models.

#### **SAGE: A Storage-Based Approach for Scalable and Efficient Sparse Generalized Matrix-Matrix Multiplication**

Myung-Hwan Jang, Yunyong Ko, Hyuck-Moo Gwon, Ikhyeon Jo, Yongjun Park, Sang-Wook Kim

Sparse generalized matrix-matrix multiplication (SpGEMM) is a fundamental operation for real-world network analysis. With the increasing size of real-world networks, the single-machine-based SpGEMM approach cannot perform SpGEMM on large-scale networks, exceeding the size of main memory (i.e., not scalable). Although the distributed-system-based approach could handle large-scale SpGEMM based on multiple machines, it suffers from severe inter-machine communication overhead to aggregate results of multiple machines (i.e., not efficient). To address this dilemma, in this paper, we propose a novel storage-based SpGEMM approach (SAGE) that stores given networks in storage (e.g., SSD) and loads only the necessary parts of the networks into main memory when they are required for processing via a 3-layer architecture. Furthermore, we point out three challenges that could degrade the overall performance of SAGE and propose three effective strategies to address them: (1) block-based workload allocation for balancing workloads across threads, (2) in-memory partial aggregation for reducing the amount of unnecessarily generated storage-memory I/Os, and (3) distribution-aware memory allocation for preventing unexpected buffer overflows in main memory. Via extensive evaluation, we verify the superiority of SAGE over existing SpGEMM methods in terms of scalability and efficiency.

#### **Relevant Entity Selection: Knowledge Graph Bootstrapping via Zero-Shot Analogical Pruning**

Lucas Jarnac, Miguel Couceiro, Pierre Monnin

Knowledge Graph Construction (KGC) can be seen as an iterative process starting from a high quality nucleus that is refined by knowledge extraction approaches in a virtuous loop. Such a nucleus can be obtained from knowledge existing in an open KG like Wikidata. However, due to the size of such generic KGs, integrating them as a whole may entail irrelevant content and scalability issues. We propose an analogy-based approach that starts from seed entities of interest in a generic KG, and keeps or prunes their neighboring entities. We evaluate our approach on Wikidata through two manually labeled datasets that contain either domain-homogeneous or -heterogeneous seed entities. We empirically show that our analogy-based approach outperforms LSTM, Random Forest, SVM, and MLP, with a drastically lower number of parameters. We also evaluate its generalization potential in a transfer learning setting. These results advocate for the further integration of analogy-based inference in tasks related to the KG lifecycle.

### **Self-supervised Contrastive Enhancement with Symmetric Few-shot Learning Towers for Cold-start News**

#### **Recommendation**

Hao Jiang, Chuanzhen Li, Juanjuan Cai, Runyu Tian, Jingling Wang

Nowadays, news spreads faster than it is consumed. This, alongside the rapid news cycle and delayed updates, has led to a challenging news cold-start issue. Likewise, the user cold-start problem, due to limited user engagement, has long hindered recommendations. To tackle both of them, we introduce the Symmetric Few-shot Learning framework for Cold-start News Recommendation (SFCNR), built upon self-supervised contrastive enhancement. Our approach employs symmetric few-shot learning towers (SFTs) to transform warm user/news attributes into their behavior/content features during training. We design two innovative feature alignment strategies to enhance towers training. Subsequently, this tower generates virtual features for cold users/news during inference, leveraging tower-stored prior knowledge through a personalized gating network. We assess the SFCNR on four quality news recommendation models, conducting comprehensive experiments on two kinds of News dataset. Results showcase significant performance boosts for both warm and cold-start scenarios compared to baseline models.

### **PriSHAP: Prior-guided Shapley Value Explanations for Correlated Features**

Guanyu Jiang, Fuzhen Zhuang, Bowen Song, Tianyi Zhang, Deqing Wang

Among numerous explainable AI (XAI) methods proposed in recent years, model explanations based on Shapley values are widely accepted for their solid theoretical support from game theory. However, most existing methods approximate Shapley values based on a feature independence assumption considering the complexity of calculating exact Shapley values. This assumption could bring some counterfactual problems when interpreted features are highly correlated and result in explanations contrary to human intuition. In this paper, we propose PriSHAP to explicitly model the dependency relationship between correlated features and provide reasonable explanations for tabular data. Feature dependencies are analyzed and taken as prior information to guide the process of estimating Shapley values. Additionally, PriSHAP is free to be applied in popular Shapley value-based explainers to address counterfactual problems while providing more faithful explanations. A pipeline is given to apply PriSHAP in existing explainers with simple adjustments. Extensive experiments on both public datasets and artificial datasets are provided to demonstrate the effectiveness of our method.

### **Knowledge-Aware Cross-Semantic Alignment for Domain-Level Zero-Shot Recommendation**

Junji Jiang, Hongke Zhao, Ming He, Likang Wu, Kai Zhang, Jianping Fan

Recommendation systems have attracted attention from academia and industry due to their wide range of application scenarios. However, cold start remains a challenging problem limited by sparse user interactions. Some scholars propose to transfer the dense information from the source domain to the target domain through cross-domain recommendation, but most of the work assumes that there is a small amount of historical interaction in the target domain. However, this approach essentially presupposes the existence of at least some historical interaction within the target domain. In this paper, we focus on the domain-level zero-shot recommendation (DZSR) problem. To address the above challenges, we propose a knowledge-aware cross-semantic alignment (K-CSA) framework to learn transferable source domain semantic information. The motivation is to establish stable alignments of interests in different domains through class semantic descriptions (CSDs). Specifically, due to the lack of effective information in the target domain, we learn semantic representations of source and target domain items based on knowledge graphs. Moreover, we conduct multi-view K-means to extract item CSDs from the learned semantic representations. Further, K-CSA learns universal user CSDs through the designed multi-head self-attention. To facilitate the transference of user interest from the source domain to the target domain, we devise a cross-semantic



contrastive learning strategy, grounded in the prototype distribution matrix. We conduct extensive experiments on several real-world cross-domain datasets, and the experimental results clearly demonstrate the superiority of our proposed K-CSA compared with other baselines.

#### **AdaMCT: Adaptive Mixture of CNN-Transformer for Sequential Recommendation**

Juyong Jiang, Peiyan Zhang, Yingtao Luo, Chaozhao Li, Jae Boum Kim, Kai Zhang, Senzhang Wang, Xing Xie, Sunghun Kim

Sequential recommendation (SR) aims to model users' dynamic preferences from a series of interactions. A pivotal challenge in user modeling for SR lies in the inherent variability of user preferences. An effective SR model is expected to capture both the long-term and short-term preferences exhibited by users, wherein the former can offer a comprehensive understanding of stable interests that impact the latter. To more effectively capture such information, we incorporate locality inductive bias into the Transformer by amalgamating its global attention mechanism with a local convolutional filter, and adaptively ascertain the mixing importance on a personalized basis through layer-aware adaptive mixture units, termed as AdaMCT. Moreover, as users may repeatedly browse potential purchases, it is expected to consider multiple relevant items concurrently in long-/short-term preferences modeling. Given that softmax-based attention may promote unimodal activation, we propose the Squeeze-Excitation Attention (with sigmoid activation) into SR models to capture multiple pertinent items (keys) simultaneously. Extensive experiments on three widely employed benchmarks substantiate the effectiveness and efficiency of our proposed approach. Source code is available at <https://github.com/juyongjiang/AdaMCT>.

#### **Enhancing the Robustness via Adversarial Learning and Joint Spatial-Temporal Embeddings in Traffic Forecasting**

Juyong Jiang, Binqing Wu, Ling Chen, Kai Zhang, Sunghun Kim

Traffic forecasting is an essential problem in urban planning and computing. The complex dynamic spatial-temporal dependencies among traffic objects (e.g., sensors and road segments) have been calling for highly flexible models; unfortunately, sophisticated models may suffer from poor robustness especially in capturing the trend of the time series (1st-order derivatives with time), leading to unrealistic forecasts. To address the challenge of balancing dynamics and robustness, we propose TrendGCN, a new scheme that extends the flexibility of GCNs and the distribution-preserving capacity of generative and adversarial loss for handling sequential data with inherent statistical correlations. On the one hand, our model simultaneously incorporates spatial (node-wise) embeddings and temporal (time-wise) embeddings to account for heterogeneous space-and-time convolutions; on the other hand, it uses GAN structure to systematically evaluate statistical consistencies between the real and the predicted time series in terms of both the temporal trending and the complex spatial-temporal dependencies. Compared with traditional approaches that handle step-wise predictive errors independently, our approach can produce more realistic and robust forecasts. Experiments on six benchmark traffic forecasting datasets and theoretical analysis both demonstrate the superiority and the state-of-the-art performance of TrendGCN. Source code is available at <https://github.com/juyongjiang/TrendGCN>.

#### **A Momentum Loss Reweighting Method for Improving Recall**

Chenzhi Jiang, Yin Jin, Ningtao Wang, Ruofan Wu, Xing Fu, Weiqiang Wang

In many practical binary classification applications, such as financial fraud detection or medical diagnosis, it is crucial to optimize a model's performance on high-confidence samples whose scores are higher than a specific threshold, which is calculated by a given false positive rate according to practical requirements. However, the proportion of high-confidence samples is typically extremely small, especially in long-tailed datasets, which can lead to poor recall results and an alignment bias between realistic goals and loss. To address this challenge, we propose a novel loss reweighting framework called Momentum Threshold-Oriented Loss (MTOL) for binary classification tasks and propose two instantiated losses of it. Given a limited FPR range, MTOL aims to improve the recall of binary classification models at that FPR range by incorporating a batch memory queue and momentum estimation mechanisms. The MTOL adaptively estimates thresholds of FPR during the model training iterations and up-weights the loss of samples in the threshold range, with little consumption of storage and computation. Our experimental results on various datasets, including CIFAR-10, CIFAR-100, Tiny-ImageNet, demonstrate the significant effect of MTOL in improving the recall at low FPR especially in class imbalance settings. These results suggest that MTOL is a promising approach in scenarios where the model's performance in the low FPR range is of utmost importance.

#### **Replace Scoring with Arrangement: A Contextual Set-to-Arrangement Framework for Learning-to-Rank**

Jiarui Jin, Xianyu Chen, Weinan Zhang, Mengyue Yang, Yang Wang, Yali Du, Yong Yu, Jun Wang

Learning-to-rank is a core technique in the top-N recommendation task, where an ideal ranker would be a mapping from an item set to an arrangement (a.k.a. permutation). Most existing solutions fall in the paradigm of probabilistic ranking principle (PRP), i.e., first score each item in the candidate set and then perform a sort operation to generate the top ranking list. However, these approaches neglect the contextual dependence among candidate items during individual scoring, and the sort operation is non-differentiable. To bypass the above issues, we propose Set-To-Arrangement Ranking (STARank), a new framework directly generates the permutations of the candidate items without the need for individually scoring and sort operations; and is end-to-end differentiable. As a result, STARank can operate when only the ground-truth permutations are accessible without requiring access to the ground-truth relevance scores for items. For this purpose, STARank first reads the candidate items in the context of the user browsing history, whose representations are fed into a Plackett-Luce module to arrange the given items into a list. To effectively utilize the given ground-truth permutations for supervising STARank, we leverage the internal consistency property of Plackett-Luce models to derive a computationally efficient list-wise loss. Experimental comparisons against 9 the state-of-the-art methods on 2 learning-to-rank benchmark datasets and 3 top-N real-world recommendation datasets demonstrate the superiority of STARank in terms of conventional ranking metrics. Notice that these ranking metrics do not consider the effects of the contextual dependence among the items in the list, we design a new family of simulation-based ranking metrics, where existing metrics can be regarded as special cases. STARank can consistently achieve better performance in terms of PBM and UBM simulation-based metrics.

### **Capturing Popularity Trends: A Simplistic Non-Personalized Approach for Enhanced Item Recommendation**

Jiazheng Jing, Yinan Zhang, Xin Zhou, Zhiqi Shen

Recommender systems have been gaining increasing research attention over the years. Most existing recommendation methods focus on capturing users' personalized preferences through historical user-item interactions, which may potentially violate user privacy. Additionally, these approaches often overlook the significance of the temporal fluctuation in item popularity that can sway users' decision-making. To bridge this gap, we propose Popularity-Aware Recommender (PARE), which makes non-personalized recommendations by predicting the items that will attain the highest popularity. PARE consists of four modules, each focusing on a different aspect: popularity history, temporal impact, periodic impact, and side information. Finally, an attention layer is leveraged to fuse the outputs of four modules. To our knowledge, this is the first work to explicitly model item popularity in recommendation systems. Extensive experiments show that PARE performs on par or even better than sophisticated state-of-the-art recommendation methods. Since PARE prioritizes item popularity over personalized user preferences, it can enhance existing recommendation methods as a complementary component. Our experiments demonstrate that integrating PARE with existing recommendation methods significantly surpasses the performance of standalone models, highlighting PARE's potential as a complement to existing recommendation methods. Furthermore, the simplicity of PARE makes it immensely practical for industrial applications and a valuable baseline for future research.

### **Hierarchical Multi-Label Classification with Partial Labels and Unknown Hierarchy**

Suhyeon Jo, DongHyeok Shin, Byeonghu Na, JoonHo Jang, Il-Chul Moon

Hierarchical multi-label classification aims at learning a multi-label classifier from a dataset whose labels are organized into a hierarchical structure. To the best of our knowledge, we propose for the first time the problem of finding a multi-label classifier given a partially labeled hierarchical multi-label dataset. We also assume the situation where the classifier cannot access hierarchical information during training. This work proposes an iterative framework for learning both multi-labels and a hierarchical structure of classes. When training a multi-label classifier from partial labels, our model extracts a class hierarchy from the classifier output using our hierarchy extraction algorithm. Then, our proposed loss exploits the extracted hierarchy to train the classifier. Theoretically, we show that our hierarchy extraction algorithm correctly finds the unknown hierarchy under a mild condition, and we prove that our loss function of multi-label classification with such hierarchy becomes an unbiased estimator of true multi-label classification risk. Our experiments show that our model obtains a class hierarchy close to the ground-truth dataset hierarchy, and simultaneously, our method outperforms previous methods for hierarchical multi-label classification and multi-label classification from partial labels.

### **Robust Graph Clustering via Meta Weighting for Noisy Graphs**

Hyeonsoo Jo, Fanchen Bu, Kijung Shin

How can we find meaningful clusters in a graph robustly against noise edges? Graph clustering (i.e., dividing nodes into groups of similar ones) is a fundamental problem in graph analysis with applications in various fields. Recent studies have demonstrated that graph neural network (GNN) based approaches yield promising results for graph clustering. However, we

observe that their performance degenerates significantly on graphs with noise edges, which are prevalent in practice. In this work, we propose MetaGC for robust GNN-based graph clustering. MetaGC employs a decomposable clustering loss function, which can be rephrased as a sum of losses over node pairs. We add a learnable weight to each node pair, and MetaGC adaptively adjusts the weights of node pairs using meta-weighting so that the weights of meaningful node pairs increase and the weights of less-meaningful ones (e.g., noise edges) decrease. We show empirically that MetaGC learns weights as intended and consequently outperforms the state-of-the-art GNN-based competitors, even when they are equipped with separate denoising schemes, on five real-world graphs under varying levels of noise. Our code and datasets are available at <https://github.com/HyeonsooJo/MetaGC>.

### **Real-time Emotion Pre-Recognition in Conversations with Contrastive Multi-modal Dialogue Pre-training**

Xincheng Ju, Dong Zhang, Suyang Zhu, Junhui Li, Shoushan Li, Guodong Zhou

This paper presents our pioneering effort in addressing a new and realistic scenario in multi-modal dialogue systems called Multi-modal Real-time Emotion Pre-recognition in Conversations (MREPC). The objective is to predict the emotion of a forthcoming target utterance that is highly likely to occur. We believe that this task can enhance the dialogue system's understanding of the interlocutor's state of mind, enabling it to prepare an appropriate response in advance. However, addressing MREPC poses the following challenges: 1) Previous studies on emotion elicitation typically focus on textual modality and perform sentiment forecasting within a fixed contextual scenario. 2) Previous studies on multi-modal emotion recognition aim to predict the emotion of existing utterances, making it difficult to extend these approaches to MREPC due to the absence of the target utterance. To tackle these challenges, we construct two benchmark multi-modal datasets for MREPC and propose a task-specific multi-modal contrastive pre-training approach. This approach leverages large-scale unlabeled multi-modal dialogues to facilitate emotion pre-recognition for potential utterances of specific target speakers. Through detailed experiments and extensive analysis, we demonstrate that our proposed multi-modal contrastive pre-training architecture effectively enhances the performance of multi-modal real-time emotion pre-recognition in conversations.

### **CFOM: Lead Optimization For Drug Discovery With Limited Data**

Natan Kaminsky, Uriel Singer, Kira Radinsky

Drug development is a long and costly process consisting of several stages that can take many years to complete. One of the early stage's goals is to optimize a novel chemical compound to be active against a target protein associated with the disease. Often machine learning techniques are used to improve the procedure of discovering and optimizing potential drug candidates. The goal of molecule optimization is, given an input molecule, to produce a new molecule that is chemically similar to the input molecule but with an improved property. We present a novel algorithm that during optimization divides a molecule into two disjoint substructures that we call: the molecule chains and the molecule core. Our approach is inspired by expert design of chemical compounds that employ a fundamental molecular template and add to it chemical functional groups to generate compounds with desired properties. We train a model to generate the molecule chains with the desired properties for optimization, which are then attached to the molecule core to construct a novel molecule with high similarity to the input molecule. This is achieved by selective masking of pairs of input molecules' chains and cores during training. Additionally, we demonstrate the extension of this approach to data-scarce tasks, like targeting a drug to a novel protein. We first evaluate our method on standard molecule optimization tasks such as inhibition against glycogen synthase kinase-3 beta (GSK3 $\beta$ ). We then empirically compared the model performance with the state-of-the-art algorithms over 21 novel proteins and show superior performance.

### **Nudging Neural Click Prediction Models to Pay Attention to Position**

Efi Karra Taniskidou, Wenjie Zhao, Iain Murray, Roberto Pellegrini

Predicting the click-through rate (CTR) of an item is a fundamental task in online advertising and recommender systems. CTR prediction models are typically trained on user click data from traffic logs. However, users are more likely to interact with items that were shown prominently on a website. CTR models often over-estimate the value of such items and show them more often, at the expense of items of higher quality that were previously shown at less prominent positions. This self-reinforcing position bias effect reduces both the immediate and long-term quality of recommendations for users. In this paper, we revisit position bias in a family of state-of-the-art neural models for CTR prediction, and use synthetic data to demonstrate the difficulty of controlling for position. We propose an approach that encourages neural networks to use position (or other confounding variables) as much as possible to explain the training data, and a metric that can directly measure bias. Experiments on two real-world datasets demonstrate the effectiveness of our approach in correcting for position-like features in 2 state-of-the-art CTR prediction models.

### [Meta-Learning with Adaptive Weighted Loss for Imbalanced Cold-Start Recommendation](#)

Minchang Kim, Yongjin Yang, Jung Hyun Ryu, Taesup Kim

Sequential recommenders have made great strides in capturing a user's preferences. Nevertheless, the cold-start recommendation remains a fundamental challenge as they typically involve limited user-item interactions for personalization. Recently, gradient-based meta-learning approaches have emerged in the sequential recommendation field due to their fast adaptation and easy-to-integrate abilities. The meta-learning algorithms formulate the cold-start recommendation as a few-shot learning problem, where each user is represented as a task to be adapted. While meta-learning algorithms generally assume that task-wise samples are evenly distributed over classes or values, user-item interactions in real-world applications do not conform to such a distribution (e.g., watching favorite videos multiple times, leaving only positive ratings without any negative ones). Consequently, imbalanced user feedback, which accounts for the majority of task training data, may dominate the user adaptation process and prevent meta-learning algorithms from learning meaningful meta-knowledge for personalized recommendations. To alleviate this limitation, we propose a novel sequential recommendation framework based on gradient-based meta-learning that captures the imbalanced rating distribution of each user and computes adaptive loss for user-specific learning. Our work is the first to tackle the impact of imbalanced ratings in cold-start sequential recommendation scenarios. Through extensive experiments conducted on real-world datasets, we demonstrate the effectiveness of our framework.

### [Diffusion Variational Autoencoder for Tackling Stochasticity in Multi-Step Regression Stock Price Prediction](#)

Kelvin J.L. Koa, Yunshan Ma, Ritchie Ng, Tat-Seng Chua

Multi-step stock price prediction over a long-term horizon is crucial for forecasting its volatility, allowing financial institutions to price and hedge derivatives, and banks to quantify the risk in their trading books. Additionally, most financial regulators also require a liquidity horizon of several days for institutional investors to exit their risky assets, in order to not materially affect market prices. However, the task of multi-step stock price prediction is challenging, given the highly stochastic nature of stock data. Current solutions to tackle this problem are mostly designed for single-step, classification-based predictions, and are limited to low representation expressiveness. The problem also gets progressively harder with the introduction of the target price sequence, which also contains stochastic noise and reduces generalizability at test-time.

To tackle these issues, we combine a deep hierarchical variational-autoencoder (VAE) and diffusion probabilistic techniques to do seq2seq stock prediction through a stochastic generative process. The hierarchical VAE allows us to learn the complex and low-level latent variables for stock prediction, while the diffusion probabilistic model trains the predictor to handle stock price stochasticity by progressively adding random noise to the stock data. To deal with the additional stochasticity in the target price sequence, we also augment the target series with noise via a coupled diffusion process. We then perform a denoising process to "clean" the prediction outputs that were trained on the stochastic target sequence data, which increases the generalizability of the model at test-time. Our Diffusion-VAE (D-Va) model is shown to outperform state-of-the-art solutions in terms of its prediction accuracy and variance. Through an ablation study, we also show how each of the components introduced helps to improve overall prediction accuracy by reducing the data noise. Most importantly, the multi-step outputs can also allow us to form a stock portfolio over the prediction length. We demonstrate the effectiveness of our model outputs in the portfolio investment task through the Sharpe ratio metric and highlight the importance of dealing with different types of prediction uncertainties. Our code can be accessed through <https://github.com/koa-fin/dva>.

### [Modeling Sequential Collaborative User Behaviors For Seller-Aware Next Basket Recommendation](#)

Ziyi Kou, Saurav Manchanda, Shih-Ting Lin, Min Xie, Haixun Wang, Xiangliang Zhang

Next Basket Recommendation (NBR) aims to recommend a set of products as a basket to users based on their historical shopping behavior. In this paper, we investigate the problem of NBR in online marketplaces (e.g., Instacart, Uber Eats) that connect users with multiple sellers. In such scenarios, effective NBR can significantly enhance the shopping experience of users by recommending diversified and completed products based on specific sellers, especially when a user purchases from a seller they have not visited before. However, conventional NBR approaches assume that all considered products are from the same sellers, which overlooks the complex relationships between users, sellers, and products. To address such limitations, we develop SecGT, a sequential collaborative graph transformer framework that recommends users with baskets from specific sellers based on seller-aware user preference representations that are generated by collaboratively modeling the joint user-seller-product interactions and sequentially exploring the user-agnostic basket transitions in an interactive way. We evaluate the performance of SecGT on users from a leading online marketplace at multiple cities with various involved

sellers. The results show that SecGT outperforms existing NBR and also traditional product recommendation approaches on recommending baskets from cold sellers for different types of users across all cities.

### [\*\*A Model-Agnostic Method to Interpret Link Prediction Evaluation of Knowledge Graph Embeddings\*\*](#)

Narayanan Asuri Krishnan, Carlos R. Rivero

In link prediction evaluation, an embedding model assigns plausibility scores to unseen triples in a knowledge graph using an input partial triple. Performance metrics like mean rank are useful to compare models side by side, but do not shed light on their behavior. Interpreting link prediction evaluation and comparing models based on such interpretation are appealing. Current interpretation methods have mainly focused on single predictions or other tasks different from link prediction. Since knowledge graph embedding methods are diverse, interpretation methods that are applicable only to certain machine learning approaches cannot be used. In this paper, we propose a model-agnostic method for interpreting link prediction evaluation as a whole. The interpretation consists of Horn rules mined from the knowledge graph containing the triples a model deems plausible. We combine precision and recall measurements of mined rules using F $\beta$  score to quantify interpretation accuracy. To maximize interpretation accuracy when comparing models, we study two approximations to the hard problem of merging rules. Our quantitative study shows that interpretation accuracy serves to compare diverse models side by side, and that these comparisons are different from those using ranks. Our qualitative study shows that several models globally capture expected semantics, and that models make a common set of predictions despite of redundancy reduction.

### [\*\*Diving into a Sea of Opinions: Multi-modal Abstractive Summarization with Comment Sensitivity\*\*](#)

Raghvendra Kumar, Ratul Chakraborty, Abhishek Tiwari, Sriparna Saha, Naveen Saini

In the modern era, the rapid expansion of social media and the proliferation of the internet community has led to a multi-fold increase in the richness and range of views and outlooks expressed by readers and viewers. To obtain valuable insights from this vast sea of opinions, we present an inventive and holistic procedure for multi-modal abstractive summarization with comment sensitivity. Our proposed model utilizes both textual and visual modalities and examines the remarks provided by the readers to produce summaries that apprehend the significant points and opinions made by them. Our model features a transformer-based encoder that seamlessly processes both news articles and comments, merging them before transmitting the amalgamated information to the decoder. Additionally, the core segment of our architecture consists of an attention-based merging technique which is trained adversarially by means of a generator and discriminator to bridge the semantic gap between comments and articles. We have used a Bi-LSTM-based branch for image pointer generation. We assess our model on the reader-aware multi-document summarization (RA-MDS) dataset which contains news articles, their summaries, and related comments. We have extended the dataset by adding images pertaining to news articles in the corpus to increase the richness and diversity of the dataset. Our comprehensive experiments reveal that our model outperforms similar pre-trained models and baselines across two of the four evaluated metrics, showcasing its superior performance.

### [\*\*Prompting Strategies for Citation Classification\*\*](#)

Suchetha N. Kunnath, David Pride, Petr Knuth

Citation classification aims to identify the purpose of the cited article in the citing article. Previous citation classification methods rely largely on supervised approaches. The models are trained on datasets with citing sentences or citation contexts annotated for a citation's purpose or function or intent. Recent advancements in Large Language Models (LLMs) have dramatically improved the ability of NLP systems to achieve state-of-the-art performances under zero or few-shot settings. This makes LLMs particularly suitable for tasks where sufficiently large labelled datasets are not yet available, which remains to be the case for citation classification. This paper systematically investigates the effectiveness of different prompting strategies for citation classification and compares them to promptless strategies as a baseline. Specifically, we evaluate the following four strategies, two of which we introduce for the first time, which involve updating Language Model (LM) parameters while training the model: (1) Promptless fine-tuning, (2) Fixed-prompt LM tuning, (3) Dynamic Context-prompt LM tuning (proposed), (4) Prompt + LM fine-tuning (proposed). Additionally, we test the zero-shot performance of LLMs, GPT3.5, and (5) Tuning-free prompting strategy that involves no parameter updating. Our results show that prompting methods based on LM parameter updating significantly improve citation classification performances on both domain-specific and multi-disciplinary citation classifications. Moreover, our Dynamic Context-prompting method achieves top scores both for the ACL-ARC and ACT2 citation classification datasets, surpassing the highest-performing system in the 3C shared task benchmark. Interestingly, we observe zero-shot GPT3.5 to perform well on ACT2 but poorly on the ACL-ARC dataset.



**Non-Compliant Bandits**

Branislav Kveton, Yi Liu, Johan Matteo Kruijssen, Yisu Nie

Bandit algorithms arose as a standard approach to learning better models online. As they become more popular, they are increasingly deployed in complex machine learning pipelines, where their actions can be overwritten. For example, in ranking problems, a list of recommended items can be modified by a downstream algorithm to increase diversity. This may break the classic bandit algorithms and lead to linear regret. Specifically, if the proposed action is not taken, uncertainty in its estimated mean reward may not get reduced. In this work, we study this setting and call it non-compliant bandits; as the agent tries to learn rewarding actions that comply with a downstream task. We propose two algorithms, compliant contextual UCB (CompUCB) and Thompson sampling (CompTS), which learn separate reward and compliance models. The compliance model allows the agent to avoid non-compliant actions. We derive a sublinear regret bound for CompUCB. We also conduct experiments that compare our algorithms to classic bandit baselines. The experiments show failures of the baselines and that we mitigate them by learning compliance models.

**VFedAD: A Defense Method Based on the Information Mechanism Behind the Vertical Federated Data Poisoning Attack**

Jinrong Lai, Tong Wang, Chuan Chen, Yihao Li, Zibin Zheng

In recent years, federated learning has achieved remarkable results in the medical and financial fields, but various attacks have always plagued federated learning. Data poisoning attack and defense research in horizontal federated learning are sufficient, yet vertical federated data poisoning attack and defense remains an open area due to two challenges: (1) Complex data distributions lead to immense attack possibilities, and (2) defense methods are insufficient for complex data distributions. We have discovered that from the perspective of information theory, the above challenges can be addressed elegantly and succinctly with a solution. We first reveal the information-theoretic mechanisms underlying vertical federated data poisoning attacks and then propose an unsupervised vertical federated data poisoning defense method (VFedAD) based on information theory. VFedAD learns semantic-rich client data representations through contrastive learning task and cross-client prediction task to identify anomalies. Experiments show VFedAD effectively detects vertical federated anomalies, protecting subsequent algorithms from vertical federated data poisoning attacks.

**Learning the Co-evolution Process on Live Stream Platforms with Dual Self-attention for Next-topic Recommendations**

Hsu-Chao Lai, Philip S. Yu, Jiun-Long Huang

Live stream platforms have gained popularity in light of emerging social media platforms. Unlike traditional on-demand video platforms, viewers and streamers on the live stream platforms are able to interact in real-time, and this makes viewer interests and live stream topics mutually affect each other on the fly, which is the unique co-evolution phenomenon on live stream platforms. In this paper, we make the first attempt to introduce a novel next-topic recommendation problem for the streamers, LSNR, which incorporates the co-evolution phenomenon. A novel framework CENTR introducing the Co-evolutionary Sequence Embedding Structure that captures the temporal relations of viewer interests and live stream topic sequences with two stacks of self-attention layers is proposed. Instead of learning the sequences individually, a novel dual self-attention mechanism is designed to model interactions between the sequences. The dual self-attention includes two modules, LCA and LVA, to leverage viewer loyalty to improve efficiency and flexibility. Finally, to facilitate cold-start recommendations for new streamers, a collaborative diffusion mechanism is implemented to improve a meta learner. Through the experiments in real datasets, CENTR outperforms state-of-the-art recommender systems in both regular and cold-start scenarios.

**A Re-evaluation of Deep Learning Methods for Attributed Graph Clustering**

Xinying Lai, Dingming Wu, Christian S. Jensen, Kezhong Lu

Attributed graph clustering aims to partition the nodes in a graph into groups such that the nodes in the same group are close in terms of graph proximity and also have similar attribute values. Recently, deep learning methods have achieved state-of-the-art clustering performance. However, the effectiveness of existing methods remains unclear due to two reasons. First, the datasets used for evaluation do not support fully the goal of attributed graph clustering. The category labels of nodes are only relevant to node attributes, and nodes with the same category label are often distant in the graph. Second, existing methods for the attributed graph clustering are complex and consist of several components. There is lack of comparisons of methods composed of different components from existing methods. This study proposes six benchmark datasets that support better the goal of attributed graph clustering and reports the performance of existing representative methods. Given that existing methods leave room for improvement on the proposed benchmark datasets, we systematically analyze five aspects of

existing methods: encoded information, training networks, fusion mechanisms, loss functions, and clustering result generation. Based on these aspects, we decompose existing methods into modules and evaluate the performance of reconfigured methods based on these modules. According to the experimental results on the proposed benchmark datasets, we identify two promising configurations: (i) taking the attribute matrix as input to a graph convolutional network and (ii) layer-wise linear fusing deep neural network and graph attention network. And we also find that complex loss function fails to improve the clustering performance.

### **Tackling Diverse Minorities in Imbalanced Classification**

Kwei-Herng Lai, Daochen Zha, Huiyuan Chen, Mangesh Bendre, Yuzhong Chen, Mashweta Das, Hao Yang, Xia Hu

Imbalanced datasets are commonly observed in various real-world applications, presenting significant challenges in training classifiers. When working with large datasets, the imbalanced issue can be further exacerbated, making it exceptionally difficult to train classifiers effectively. To address the problem, over-sampling techniques have been developed to linearly interpolating data instances between minorities and their neighbors. However, in many real-world scenarios such as anomaly detection, minority instances are often dispersed diversely in the feature space rather than clustered together. Inspired by domain-agnostic data mix-up, we propose generating synthetic samples iteratively by mixing data samples from both minority and majority classes. It is non-trivial to develop such a framework, the challenges include source sample selection, mix-up strategy selection, and the coordination between the underlying model and mix-up strategies. To tackle these challenges, we formulate the problem of iterative data mix-up as a Markov decision process (MDP) that maps data attributes onto an augmentation strategy. To solve the MDP, we employ an actor-critic framework to adapt the discrete-continuous decision space. This framework is utilized to train a data augmentation policy and design a reward signal that explores classifier uncertainty and encourages performance improvement, irrespective of the classifier's convergence. We demonstrate the effectiveness of our proposed framework through extensive experiments conducted on seven publicly available benchmark datasets using three different types of classifiers. The results of these experiments showcase the potential and promise of our framework in addressing imbalanced datasets with diverse minorities.

### **DuoGAT: Dual Time-oriented Graph Attention Networks for Accurate, Efficient and Explainable Anomaly Detection on**

#### **Time-series**

Jongsoo Lee, Byeongtae Park, Dong-Kyu Chae

Recently, Graph Neural Networks (GNNs) have achieved state-of-the-art performance on the multivariate time-series anomaly detection task by learning relationships between variables (sensors). However, they show limitations in capturing temporal dependencies due to lack of sufficient consideration on the characteristics of time to their graph structure. Several studies constructed a time-oriented graph, where each node represents a timestamp within a certain sliding window, to model temporal dependencies, but they failed to learn the trend of changes in time-series. This paper proposes Dual time-oriented Graph ATtention networks (DuoGAT) that resolves the aforementioned problems. Unlike previous work that uses the simple complete undirected structure for time-oriented graphs, our work models directed graphs with weighted edges that only connect from prior events to posterior events, and the edges that connect nearby events are given higher weights. In addition, another time-oriented graph is used to model time series stationary via differencing, which especially focuses on capturing the series of changes. Empirically, our method outperformed the existing state-of-the-art work with the highest F1-score for the four real-world dataset while maintaining low training cost. We also proposed a novel explanation method for anomaly detection using DuoGAT, which provides time-oriented reasoning via hierarchically tracking time points critical in a specific anomaly detection. Our code is available at: <https://github.com/ByeongtaePark/DuoGAT>

### **GUARD: Graph Universal Adversarial Defense**

Jintang Li, Jie Liao, Ruofan Wu, Liang Chen, Zibin Zheng, Jiawang Dan, Changhua Meng, Weiqiang Wang

Graph convolutional networks (GCNs) have been shown to be vulnerable to small adversarial perturbations, which becomes a severe threat and largely limits their applications in security-critical scenarios. To mitigate such a threat, considerable research efforts have been devoted to increasing the robustness of GCNs against adversarial attacks. However, current defense approaches are typically designed to prevent GCNs from untargeted adversarial attacks and focus on overall performance, making it challenging to protect important local nodes from more powerful targeted adversarial attacks. Additionally, a trade-off between robustness and performance is often made in existing research. Such limitations highlight the need for developing an effective and efficient approach that can defend local nodes against targeted attacks, without compromising the overall performance of GCNs. In this work, we present a simple yet effective method, named Graph Universal AdveRsarial Defense (GUARD). Unlike previous works, GUARD protects each individual node from attacks with a

universal defensive patch, which is generated once and can be applied to any node (node-agnostic) in a graph. GUARD is fast, straightforward to implement without any change to network architecture nor any additional parameters, and is broadly applicable to any GCNs. Extensive experiments on four benchmark datasets demonstrate that GUARD significantly improves robustness for several established GCNs against multiple adversarial attacks and outperforms state-of-the-art defense methods by large margins.

#### **ST-MoE: Spatio-Temporal Mixture-of-Experts for Debiasing in Traffic Prediction**

Shuhao Li, Yue Cui, Yan Zhao, Weidong Yang, Ruiyuan Zhang, Xiaofang Zhou

The pervasiveness of GPS-enabled devices and wireless communication technologies results in a proliferation of traffic data in intelligent transportation systems, where traffic prediction is often essential to enable reliability and safety. Many recent studies target traffic prediction using deep learning techniques. They model spatio-temporal dependencies among traffic states by deep learning and achieve good overall performance. However, existing studies ignore the bias on traffic prediction models, which refers to non-uniformed performance distribution across road segments, especially the significantly poor prediction results on certain road segments. To solve this issue, we propose a framework named spatio-temporal mixture-of-experts (ST-MoE) that aims to eliminate the bias on traffic prediction. In general, we refer to any traffic prediction model as the based model, and adopt the proposed ST-MoE framework as a plug-in to debias. ST-MoE uses stacked convolution-based networks to learn spatio-temporal representations of individual patterns of road segments and then adaptively assigns appropriate expert layers (sub-networks) to different patterns through a spatio-temporal gating network. To this end, the patterns can be distinguished, and biased performance among road segments can be eliminated by experts tailored for specific patterns, which also further improves the overall prediction accuracy of the base model. Extensive experimental results on various base models and real-world datasets prove the effectiveness of ST-MoE.

#### **Class-Specific Word Sense Aware Topic Modeling via Soft Orthogonalized Topics**

Wenbo Li, Yao Yang, Einoshin Suzuki

We propose a word sense aware topic model for document classification based on soft orthogonalized topics. An essential problem for this task is to capture word senses related to classes, i.e., class-specific word senses. Traditional models mainly introduce semantic information of knowledge libraries for word sense discovery. However, this information may not align with the classification targets, because these targets are often subjective and task-related. We aim to model the class-specific word senses in topic space. The challenge is to optimize the class separability of the senses, i.e., obtaining sense vectors with (a) high intra-class and (b) low inter-class similarities. Most existing models predefine specific topics for each class to specify the class-specific sense vectors. We call them hard orthogonalization based methods. These methods can hardly achieve both (a) and (b) since they assume the conditional independence of topics to classes and inevitably lose topic information. To this problem, we propose soft orthogonalization for topics. Specifically, we reserve all the topics and introduce a group of class-specific weights for each word to handle the importance of topic dimensions to class separability. Besides, we detect and use highly class-specific words in each document to guide sense estimation. Our experiments on two standard datasets show that our proposal outperforms other state-of-the-art models in terms of accuracy of sense estimation, document classification, and topic modeling. In addition, our joint learning experiments with the pre-trained language model BERT showcased the best complementarity of our model in most cases compared to other topic models.

#### **AutoMRM: A Model Retrieval Method Based on Multimodal Query and Meta-learning**

Zhaotian Li, Binhang Qi, Hailong Sun, Xiang Gao

With more and more Deep Neural Network (DNN) models are publicly available on model sharing platforms (e.g., HuggingFace), model reuse has become a promising way in practice to improve the efficiency of DNN model construction by avoiding the costs of model training. To that end, a pivotal step for model reuse is model retrieval, which facilitates discovering suitable models from a model hub that match the requirements of users. However, the existing model retrieval methods have inadequate performance and efficiency, since they focus on matching user requirements with the model names, and thus cannot work well for high-dimensional data such as images. In this paper, we propose a user-task-centric multimodal model retrieval method named AutoMRM. AutoMRM can retrieve DNN models suitable for the user's task according to both the dataset and description of the task. Moreover, AutoMRM utilizes meta-learning to retrieve models for previously unseen task queries. Specifically, given a task, AutoMRM extracts the latent meta-features from the dataset and description for training meta-learners offline and obtaining the representation of user task queries online. Experimental results demonstrate that AutoMRM outperforms existing model retrieval methods including the state-of-the-art method in both effectiveness and efficiency.

### **Towards Automatic ICD Coding via Knowledge Enhanced Multi-Task Learning**

Xinhang Li, Xiangyu Zhao, Yong Zhang, Chunxiao Xing

The aim of ICD coding is to assign International Classification of Diseases (ICD) codes to unstructured clinical notes or discharge summaries. Numerous methods have been proposed for automatic ICD coding in an effort to reduce human labor and errors. However, existing works disregard the data imbalance problem of clinical notes. In addition, the noisy clinical note issue has not been thoroughly investigated. To address such issues, we propose a knowledge enhanced Graph Attention Network (GAT) under multi-task learning setting. Specifically, multi-level information transitions and interactions have been implemented. On the one hand, a large heterogeneous text graph is constructed to capture both intra- and inter-note correlations between various semantic concepts, thereby alleviating the data imbalance issue. On the other hand, two auxiliary healthcare tasks have been proposed to facilitate the sharing of information across tasks. Moreover, to tackle the issue of noisy clinical notes, we propose to utilize the rich structured knowledge facts and information provided by medical domain knowledge, thereby encouraging the model to focus on the clinical notes' noteworthy portion and valuable information. The experimental results on the widely-used medical dataset, MIMIC-III, demonstrate the advantages of our proposed framework.

### **Relation-Aware Diffusion Model for Controllable Poster Layout Generation**

Fengheng Li, An Liu, Wei Feng, Honghe Zhu, Yaoyu Li, Zheng Zhang, Jingjing Lv, Xin Zhu, Junjie Shen, Zhangang Lin, Jingping Shao

Poster layout is a crucial aspect of poster design. Prior methods primarily focus on the correlation between visual content and graphic elements. However, a pleasant layout should also consider the relationship between visual and textual contents and the relationship between elements. In this study, we introduce a relation-aware diffusion model for poster layout generation that incorporates these two relationships in the generation process. Firstly, we devise a visual-textual relation-aware module that aligns the visual and textual representations across modalities, thereby enhancing the layout's efficacy in conveying textual information. Subsequently, we propose a geometry relation-aware module that learns the geometry relationship between elements by comprehensively considering contextual information. Additionally, the proposed method can generate diverse layouts based on user constraints. To advance research in this field, we have constructed a poster layout dataset named CGL-Dataset V2. Our proposed method outperforms state-of-the-art methods on CGL-Dataset V2. The data and code will be available at <https://github.com/liuan0803/RADM>.

### **ACGAN-GNNExplainer: Auxiliary Conditional Generative Explainer for Graph Neural Networks**

Yiqiao Li, Jianlong Zhou, Yifei Dong, Niusha Shafiabady, Fang Chen

Graph neural networks (GNNs) have proven their efficacy in a variety of real-world applications, but their underlying mechanisms remain a mystery. To address this challenge and enable reliable decision-making, many GNN explainers have been proposed in recent years. However, these methods often encounter limitations, including their dependence on specific instances, lack of generalizability to unseen graphs, producing potentially invalid explanations, and yielding inadequate fidelity. To overcome these limitations, we, in this paper, introduce the Auxiliary Classifier Generative Adversarial Network (ACGAN) into the field of GNN explanation and propose a new GNN explainer dubbed ACGAN-GNNExplainer. Our approach leverages a generator to produce explanations for the original input graphs while incorporating a discriminator to oversee the generation process, ensuring explanation fidelity and improving accuracy. Experimental evaluations conducted on both synthetic and real-world graph datasets demonstrate the superiority of our proposed method compared to other existing GNN explainers.

### **HAMUR: Hyper Adapter for Multi-Domain Recommendation**

Xiaopeng Li, Fan Yan, Xiangyu Zhao, Yichao Wang, Bo Chen, Huifeng Guo, Ruiming Tang

Multi-Domain Recommendation (MDR) has gained significant attention in recent years, which leverages data from multiple domains to enhance their performance concurrently. However, current MDR models are confronted with two limitations. Firstly, the majority of these models adopt an approach that explicitly shares parameters between domains, leading to mutual interference among them. Secondly, due to the distribution differences among domains, the utilization of static parameters in existing methods limits their flexibility to adapt to diverse domains. To address these challenges, we propose a novel model HAMUR. Specifically, HAMUR consists of two components: (1). Domain-specific adapter, designed as a pluggable module that can be seamlessly integrated into various existing multi-domain backbone models, and (2). Domain-shared hyper-network, which implicitly captures shared information among domains and dynamically generates the parameters for the

adapter. We conduct extensive experiments on two public datasets using various backbone networks. The experimental results validate the effectiveness and scalability of the proposed model.

### [\*\*REST: Drug-Drug Interaction Prediction via Reinforced Student-Teacher Curriculum Learning\*\*](#)

Xinhang Li, Zhaopeng Qiu, Xiangyu Zhao, Yong Zhang, Chunxiao Xing, Xian Wu

Accurate prediction of drug-drug interaction (DDI) is crucial to achieving effective decision-making in medical treatment for both doctors and patients. Recently, many deep learning based methods have been proposed to learn from drug-related features and conduct DDI prediction. These works have achieved promising results. However, the extreme imbalance of medical data poses a serious problem to DDI prediction, where a small fraction of DDI types occupy the majority training data. A straightforward way is to develop an appropriate policy to sample the data. Due to the high complexity and speciality of medical science, a dynamic learnable policy is required instead of a heuristic, uniform or static one. Therefore, we propose a REinforced Student-Teacher curriculum learning model (REST) for effective sampling to tackle this imbalance problem. Specifically, REST consists of two interactive parts, which are a heterogeneous graph neural network as the student and a reinforced sampler as the teacher. In each interaction, the teacher model takes action to sample an appropriate batch to train the student model according to the student model state while the cumulated improvement in performance of the student model is treated as the reward for policy gradient of the teacher model. The experimental results on two benchmarking datasets have demonstrated the significant effectiveness of our proposed model in DDI prediction, especially for the DDI types with low frequency.

### [\*\*Simplifying Temporal Heterogeneous Network for Continuous-Time Link prediction\*\*](#)

Ce Li, Rongpei Hong, Xovee Xu, Goce Trajcevski, Fan Zhou

Temporal heterogeneous networks (THNs) investigate the structural interactions and their evolution over time in graphs with multiple types of nodes or edges. Existing THNs describe evolving networks as a sequence of graph snapshots and adopt mechanisms from static heterogeneous networks to capture the spatial-temporal correlation. However, these works are confined to the discrete-time setting and the implementation of stacked mechanisms often introduces a high level of complexity, both conceptually and computationally. Here, we conduct comprehensive examinations and propose STHN, a simplifying THN for continuous-time link prediction. Concretely, to integrate continuous dynamics, we maintain a historical interaction memory for each node. A link encoder that incorporates two components - type encoding and relative time encoding - is introduced to encapsulate implicit heterogeneous characteristics of interaction and extract the most informative temporal information. We further propose to use a patching technique that assists with Transformer feature extractor to support the interaction sequence with long histories. Extensive experiments on three real-world datasets empirically demonstrate that STHN outperforms state-of-the-art methods with competitive task accuracy and predictive efficiency on both transductive and inductive settings.

### [\*\*Heterogeneous Temporal Graph Neural Network Explainer\*\*](#)

Jiazheng Li, Chunhui Zhang, Chuxu Zhang

Graph Neural Networks (GNNs) have been a prominent research area and have been widely deployed in various high-stakes applications in recent years, leading to a growing demand for explanations. While existing explainer methods focus on explaining homogeneous and static GNNs, none of them have attempted to explain heterogeneous temporal GNNs. However, in practice, many real-world databases should be represented as heterogeneous temporal graphs (HTGs), which serve as the fundamental data structure for GNN backbone models in applications. To address this gap, in this paper, we propose HTGExplainer, a novel method for explaining heterogeneous temporal GNNs by considering temporal dependencies and preserving heterogeneity when generating subgraphs as explanations. HTGExplainer employs a deep neural network to re-parameterize the generation process of explanations and incorporates effective heterogeneous and temporal edge embeddings to capture informative semantics used for generating explanatory subgraphs. Extensive experiments are conducted on multiple HTG datasets constructed from real-world scenarios, and the results demonstrate the superior performance of HTGExplainer compared to state-of-the-art baselines.

### [\*\*Harnessing the Power of Pre-trained Vision-Language Models for Efficient Medical Report Generation\*\*](#)

Qi Li

Medical images are commonly used in clinical practice. But the need for diagnosis and reporting from image-based examinations far exceeds the current medical capacity. Automatic Medical Report Generation (MRG) can help to ease the burden of radiologists. Vision-Language Pre-training (VLP) has received tremendous success on various tasks, therefore it is



naturally expected that MRG can harvest from this rapid advancement. However, directly applying existing VLP models in the medical domain is impracticable due to their data-hungry nature, the need for aligning different modalities, prohibitive training time, exorbitant hardware barrier, and the challenge of open-ended text generation. To address these problems, we propose MedEPT, a parameter-efficient approach for MRG that can utilize ever-ignored image-only datasets. It employs parameter-efficient tuning (PET) for VLP adaption to mitigate inefficiency in fine-tuning time and hardware. MedEPT also employs MRGPID to augment and expand adaption datasets by synthesizing meaningful text for image-only datasets. We perform a systematic evaluation of our method. Empirical results show that we obtain a better performance than the state-of-the-art method while using less than 10% trainable parameters and not more than 30% training time than ever before.

#### [Graph Enhanced Hierarchical Reinforcement Learning for Goal-oriented Learning Path Recommendation](#)

Qingyao Li, Wei Xia, Li'ang Yin, Jian Shen, Renting Rui, Weinan Zhang, Xianyu Chen, Ruiming Tang, Yong Yu

Goal-oriented Learning path recommendation aims to recommend learning items (concepts or exercises) step-by-step to a learner to promote the mastery level of her specific learning goals. By formulating this task as a Markov decision process, reinforcement learning (RL) methods have demonstrated great power. Although extensive research efforts have been made, previous methods still fail to recommend effective goal-oriented paths due to the under-utilizing of goals. Specifically, it is mainly reflected in two aspects: (1)The lack of goal planning. When learners have multiple goals with different difficulties, the previous methods can't fully utilize the difficulties and dependencies between goal learning items to plan the sequence of achieving these goals, making the path chaotic and inefficient; (2)The lack of efficiency in goal achieving. When pursuing a single goal, the path may contain learning items unrelated to the goal, which makes realizing a certain goal inefficient. To address these challenges, we present a novel Graph Enhanced Hierarchical Reinforcement Learning (GEHRL) framework for goal-oriented learning path recommendation. The framework divides learning path recommendation into two parts: sub-goal selection(planning) and sub-goal achieving(learning item recommendation). Specifically, we employ a high-level agent as a sub-goal selector to select sub-goals for the low-level agent to achieve. The low-level agent in the framework is to recommend learning items to the learner. To make the path only contain goal-related learning items to improve the efficiency of achieving the goal, we develop a graph-based candidate selector to constrain the action space of the low-level agent based on the sub-goal and knowledge graph. We also develop test-based internal reward for low-level training so that the sparsity problem of external reward can be alleviated. Extensive experiments on three different simulators demonstrate our framework achieves state-of-the-art performance.

#### [Tight-Sketch: A High-Performance Sketch for Heavy Item-Oriented Data Stream Mining with Limited Memory Size](#)

Weihe Li, Paul Patras

Accurate and fast data stream mining is critical and fundamental to many tasks, including time series database handling, big data management and machine learning. Different heavy-based detection tasks, such as heavy hitter, heavy changer, persistent item and significant item detection, have drawn much attention from both the industry and academia. Unfortunately, due to the growing data stream speeds and limited memory (L1 cache) available for real-time processing, existing schemes face challenges in simultaneously achieving high detection accuracy, high memory efficiency, and fast update throughput, as we reveal. To tackle this conundrum, we propose a versatile and elegant sketch framework named Tight-Sketch, which supports a spectrum of heavy-based detection tasks. Considering that most items are cold (non-heavy/persistent/significant) in practice, we employ different eviction treatments for different types of items to discard these potentially cold ones as soon as possible, and offer more protection to those that are hot (heavy/persistent/significant). In addition, we propose an eviction method that follows a stochastic decay strategy, enabling Tight-Sketch to only bear small one-sided errors (no overestimation). We present a theoretical analysis of the error bounds and conduct extensive experiments on diverse detection tasks to demonstrate that Tight-Sketch significantly outperforms existing methods in terms of accuracy and update speed. Lastly, we accelerate Tight-Sketch's update throughput by up to 36% with Single Instruction Multiple Data (SIMD) instructions.

#### [Contrastive Representation Learning Based on Multiple Node-centered Subgraphs](#)

Dong Li, Wenjun Wang, Minglai Shao, Chen Zhao

As the basic element of graph-structured data, node has been recognized as the main object of study in graph representation learning. A single node intuitively has multiple node-centered subgraphs from the whole graph (e.g., one person in a social network has multiple social circles based on his different relationships). We study this intuition under the framework of graph contrastive learning, and propose a multiple node-centered subgraphs contrastive representation learning method to learn node representation on graphs in a self-supervised way. Specifically, we carefully design a series of node-centered regional

subgraphs of the central node. Then, the mutual information between different subgraphs of the same node is maximized by contrastive loss. Experiments on various real-world datasets and different downstream tasks demonstrate that our model has achieved state-of-the-art results.

### **Prompt Distillation for Efficient LLM-based Recommendation**

Lei Li, Yongfeng Zhang, Li Chen

Large language models (LLM) have manifested unparalleled modeling capability on various tasks, e.g., multi-step reasoning, but the input to these models is mostly limited to plain text, which could be very long and contain noisy information. Long text could take long time to process, and thus may not be efficient enough for recommender systems that require immediate response. In LLM-based recommendation models, user and item IDs are usually filled in a template (i.e., discrete prompt) to allow the models to understand a given task, but the models usually need extensive fine-tuning to bridge the user/item IDs and the template words and to unleash the power of LLM for recommendation. To address the problems, we propose to distill the discrete prompt for a specific task to a set of continuous prompt vectors so as to bridge IDs and words and to reduce the inference time. We also design a training strategy with an attempt to improve the efficiency of training these models. Experimental results on three real-world datasets demonstrate the effectiveness of our PrOmpt Distillation (POD) approach on both sequential recommendation and top-N recommendation tasks. Although the training efficiency can be significantly improved, the improvement of inference efficiency is limited. This finding may inspire researchers in the community to further improve the inference efficiency of LLM-based recommendation models.

### **Multi-Order Relations Hyperbolic Fusion for Heterogeneous Graphs**

Junlin Li, Yueheng Sun, Minglai Shao

Heterogeneous graphs with multiple node and edge types are prevalent in real-world scenarios. However, most methods use meta-paths on the original graph structure to learn information in heterogeneous graphs, and these methods only consider pairwise relations and rely on meta-paths. In this paper, we use simplicial complexes to extract higher-order relations containing multiple nodes from heterogeneous graphs. We also discover power-law structures in both the heterogeneous graph and the extracted simplicial complex. Thus, we propose the Simplicial Hyperbolic Attention Network (SHAN), a graph neural network for heterogeneous graphs. SHAN extracts simplicial complexes and the original graph structure from the heterogeneous graph to represent multi-order relations between nodes. Next, SHAN uses hyperbolic multi-perspective attention to learn the importance of different neighbors and relations in hyperbolic space. Finally, SHAN integrates multi-order relations to obtain a more comprehensive node representation. We conducted extensive experiments to verify the effectiveness of SHAN and the results of node classification experiments on three publicly available heterogeneous graph datasets demonstrate that SHAN outperforms representative baseline models.

### **THGNN: An Embedding-based Model for Anomaly Detection in Dynamic Heterogeneous Social Networks**

Yilin Li, Jiaqi Zhu, Congcong Zhang, Yi Yang, Jiawen Zhang, Ying Qiao, Hongan Wang

Anomaly detection, particularly the detection of anomalous behaviors in dynamic and heterogeneous social networks, is becoming more and more crucial in real life. Traditional rule-based and feature-based methods cannot well capture the structural and temporal patterns of ever-changing user behaviors. Moreover, most of the existing works based on network embedding either rely on discretized snapshots, which have ignored accurate temporal relations among user behaviors and weakened the impact of new edges, or fail to utilize dynamic and heterogeneous information simultaneously to distinguish varying effects of new edges on existing nodes. In this paper, we propose an end-to-end continuous-time model, named Temporal Heterogeneous Graph Neural Network (THGNN), to detect anomalous behaviors (edges) in dynamic heterogeneous social networks. Specifically, the model constantly updates node embeddings by propagating the information of a new edge to its source and target nodes as well as their neighbors. In this process, heterogeneous encoders are employed to handle different types of nodes and edges. What is more, a novel dual-level distributive attention mechanism is designed to allocate the influence degree of a currently interacting node to its multiple neighbors, considering the combined effect of edge type and time interval information. That can be regarded as an extension of the classical aggregative attention mechanism in the opposite direction. Extensive experiments on four real-world datasets demonstrate that THGNN outperforms all the baselines on the task of anomalous edge detection, achieving an average AUC gain of 6% across all datasets.

### **Retrieving GNN Architecture for Collaborative Filtering**

Fengqi Liang, Huan Zhao, Zhenyi Wang, Wei Fang, Chuan Shi

Graph Neural Networks (GNNs) have been widely used in Collaborative Filtering (CF). However, when given a new recommendation scenario, the current options are either selecting from existing GNN architectures or employing Neural Architecture Search (NAS) to obtain a well-performing GNN model, both of which are expensive in terms of human expertise or computational resources. To address the problem, in this work, we propose a novel neural retrieval approach, dubbed RGCF, to search a well-performing architecture for GNN-based CF rapidly when handling new scenarios. Specifically, we design the neural retrieval approach based on meta-learning by developing two-level meta-features, ranking loss, and task-level data augmentation, and in a retrieval paradigm, RGCF can directly return a well-performing architecture given a new dataset (query), thus being efficient inherently. Experimental results on two mainstream tasks, i.e., rating prediction and item ranking, show that RGCF outperforms all models either by human-designed or NAS on two new datasets in terms of effectiveness and efficiency. Particularly, the efficiency improvement is significant, taking as an example that RGCF is 61.7-206.3x faster than a typical reinforcement learning based NAS approach on the two new datasets. Code and data are available at <https://github.com/BUPT-GAMMA/RGCF>.

### **SAILOR: Structural Augmentation Based Tail Node Representation Learning**

Jie Liao, Jintang Li, Liang Chen, Bingzhe Wu, Yatao Bian, Zibin Zheng

Graph neural networks (GNNs) have achieved state-of-the-art performance in representation learning for graphs recently. However, the effectiveness of GNNs, which capitalize on the key operation of message propagation, highly depends on the quality of the topology structure. Most of the graphs in real-world scenarios follow a long-tailed distribution on their node degrees, that is, a vast majority of the nodes in the graph are tail nodes with only a few connected edges. GNNs produce inferior node representations for tail nodes due to the lack of sufficient structural information. In the pursuit of promoting the performance of GNNs for tail nodes, we explore how the deficiency of structural information deteriorates the performance of tail nodes and propose a general structural augmentation based tail node representation learning framework, dubbed as ours, which can jointly learn to augment the graph structure and extract more informative representations for tail nodes. Extensive experiments on six public benchmark datasets demonstrate that ours outperforms the state-of-the-art methods for tail node representation learning.

### **Bias Invariant Approaches for Improving Word Embedding Fairness**

Siyu Liao, Ringting Zhang, Barbara Poblete, Vanessa Murdock

Many public pre-trained word embeddings have been shown to encode different types of biases. Embeddings are often obtained from training on large pre-existing corpora, and therefore resulting biases can be a reflection of unfair representations in the original data. Bias, in this scenario, is a challenging problem since current mitigation techniques require knowing and understanding existing biases in the embedding, which is not always possible. In this work, we propose to improve word embedding fairness by borrowing methods from the field of data privacy. The idea behind this approach is to treat bias as if it were a special type of training data leakage. This has the unique advantage of not requiring prior knowledge of potential biases in word embeddings. We investigated two types of privacy algorithms, and measured their effect on bias using four different metrics. To investigate techniques from differential privacy, we applied Gaussian perturbation to public pre-trained word embeddings. To investigate noiseless privacy, we applied vector quantization during training. Experiments show that both approaches improve fairness for commonly used embeddings, and additionally, noiseless privacy techniques reduce the size of the resulting embedding representation.

### **MadSGM: Multivariate Anomaly Detection with Score-based Generative Models**

Haksoo Lim, Sewon Park, Minjung Kim, Jaehoon Lee, Seonkyu Lim, Noseong Park

The time-series anomaly detection is one of the most fundamental tasks for time-series. Unlike the time-series forecasting and classification, the time-series anomaly detection typically requires unsupervised (or self-supervised) training since collecting and labeling anomalous observations are difficult. In addition, most existing methods resort to limited forms of anomaly measurements and therefore, it is not clear whether they are optimal in all circumstances. To this end, we present a multivariate time-series anomaly detector based on score-based generative models, called MadSGM, which considers the broadest ever set of anomaly measurement factors: i) reconstruction-based, ii) density-based, and iii) gradient-based anomaly measurements. We also design a conditional score network and its denoising score matching loss for the time-series anomaly detection. Experiments on five real-world benchmark datasets illustrate that MadSGM achieves the most robust and accurate predictions.

### **Adaptation Speed Analysis for Fairness-aware Causal Models**

Yujie Lin, Chen Zhao, Minglai Shao, Xujiang Zhao, Haifeng Chen

For example, in machine translation tasks, to achieve bidirectional translation between two languages, the source corpus is often used as the target corpus, which involves the training of two models with opposite directions. The question of which one can adapt most quickly to a domain shift is of significant importance in many fields. Specifically, consider an original distribution  $p$  that changes due to an unknown intervention, resulting in a modified distribution  $p^*$ . In aligning  $p$  with  $p^*$ , several factors can affect the adaptation rate, including the causal dependencies between variables in  $p$ . In real-life scenarios, however, we have to consider the fairness of the training process, and it is particularly crucial to involve a sensitive variable (bias) present between a cause and an effect variable. To explore this scenario, we examine a simple structural causal model (SCM) with a cause-bias-effect structure, where variable  $A$  acts as a sensitive variable between cause ( $X$ ) and effect ( $Y$ ). The two models respectively exhibit consistent and contrary cause-effect directions in the cause-bias-effect SCM. After conducting unknown interventions on variables within the SCM, we can simulate some kinds of domain shifts for analysis. We then compare the adaptation speeds of two models across four shift scenarios. Additionally, we prove the connection between the adaptation speeds of the two models across all interventions.

#### [printf: Preference Modeling Based on User Reviews with Item Images and Textual Information via Graph Learning](#)

Hao-Lun Lin, Jyun-Yu Jiang, Ming-Hao Juan, Pu-Jen Cheng

Nowadays, modern recommender systems usually leverage textual and visual contents as auxiliary information to predict user preference. For textual information, review texts are one of the most popular contents to model user behaviors. Nevertheless, reviews usually lose their shine when it comes to top- $N$  recommender systems because those that solely utilize textual reviews as features struggle to adequately capture the interaction relationships between users and items. For visual one, it is usually modeled with naive convolutional networks and also hard to capture high-order relationships between users and items. Moreover, previous works did not collaboratively use both texts and images in a proper way. In this paper, we propose printf, preference modeling based on user reviews with item images and textual information via graph learning, to address the above challenges. Specifically, the dimension-based attention mechanism directs relations between user reviews and interacted items, allowing each dimension to contribute different importance weights to derive user representations. Extensive experiments are conducted on three publicly available datasets. The experimental results demonstrate that our proposed printf consistently outperforms baseline methods with the relative improvements for NDCG@5 of 26.80%, 48.65%, and 25.74% on Amazon-Grocery, Amazon-Tools, and Amazon-Electronics datasets, respectively. The in-depth analysis also indicates the dimensions of review representations definitely have different topics and aspects, assisting the validity of our model design.

#### [On the Thresholding Strategy for Infrequent Labels in Multi-label Classification](#)

Yu-Jen Lin, Chih-Jen Lin

In multi-label classification, the imbalance between labels is often a concern. For a label that seldom occurs, the default threshold used to generate binarized predictions of that label is usually sub-optimal. However, directly tuning the threshold to optimize F-measure has been observed to overfit easily. In this work, we explain why this overfitting occurs. Then, we analyze the FBR heuristic, a previous technique proposed to address the overfitting issue. We explain its success but also point out some problems unobserved before. Then, we first propose a variant of the FBR heuristic that not only fixes the problems but is also more justifiable. Second, we propose a new technique based on smoothing the F-measure when tuning the threshold. We theoretically prove that, with proper parameters, smoothing results in desirable properties of the tuned threshold. Based on the idea of smoothing, we then propose jointly optimizing micro-F and macro-F as a lightweight alternative free from extra hyperparameters. Our methods are empirically evaluated on text and node classification datasets. The results show that our methods consistently outperform the FBR heuristic.

#### [Cross-city Few-Shot Traffic Forecasting via Traffic Pattern Bank](#)

Zhanyu Liu, Guanjie Zheng, Yanwei Yu

Traffic forecasting is a critical service in Intelligent Transportation Systems (ITS). Utilizing deep models to tackle this task relies heavily on data from traffic sensors or vehicle devices, while some cities might lack device support and thus have few available data. So, it is necessary to learn from data-rich cities and transfer the knowledge to data-scarce cities in order to improve the performance of traffic forecasting. To address this problem, we propose a cross-city few-shot traffic forecasting framework via Traffic Pattern Bank (TPB) due to that the traffic patterns are similar across cities. TPB utilizes a pre-trained traffic patch encoder to project raw traffic data from data-rich cities into high-dimensional space, from which a traffic pattern

bank is generated through clustering. Then, the traffic data of the data-scarce city could query the traffic pattern bank and explicit relations between them are constructed. The metaknowledge is aggregated based on these relations and an adjacency matrix is constructed to guide a downstream spatial-temporal model in forecasting future traffic. The frequently used meta-training framework Reptile is adapted to find a better initial parameter for the learnable modules. Experiments on real-world traffic datasets show that TPB outperforms existing methods and demonstrates the effectiveness of our approach in cross-city few-shot traffic forecasting.

#### **GranCATs: Cross-Lingual Enhancement through Granularity-Specific Contrastive Adapters**

Meizhen Liu, Jiakai He, Xu Guo, Jianye Chen, Siu Cheung Hui, Fengyu Zhou

Multilingual language models (MLLMs) have demonstrated remarkable success in various cross-lingual downstream tasks, facilitating the transfer of knowledge across numerous languages, whereas this transfer is not universally effective. Our study reveals that while existing MLLMs like mBERT can capture phrase-level alignments across the language families, they struggle to effectively capture sentence-level and paragraph-level alignments. To address this limitation, we propose GranCATs, Granularity-specific Contrastive Adapters. We collect a new dataset that observes each sample at three distinct levels of granularity and employ contrastive learning as a pre-training task to train GranCATs on this dataset. Our objective is to enhance MLLMs' adaptation to a broader range of cross-lingual tasks by equipping them with improved capabilities to capture global information at different levels of granularity. Extensive experiments show that MLLMs with GranCATs yield significant performance advancements across various language tasks with different text granularities, including entity alignment, relation extraction, sentence classification and retrieval, and question-answering. These results validate the effectiveness of our proposed GranCATs in enhancing cross-lingual alignments across various text granularities and effectively transferring this knowledge to downstream tasks.

#### **UniTE: A Unified Treatment Effect Estimation Method for One-sided and Two-sided Marketing**

Runshi Liu, Zhipeng Hou

Many internet platforms are two-sided markets that involve two types of participants. Examples include e-commerce platforms like Taobao (retailers and consumers) and ride-hailing platforms like Uber (drivers and passengers). Participants of different types in the two-sided market have relationships (i.e., supply and demand) that provide externalities and network benefits. On two-sided platforms, marketing campaigns are designed by subsidizing supply or demand. Uplift models built in this scenario usually consider the treatment assignment for only one of the two sides. However, ignoring the interaction of treatments between two sides or treating them as noises may result in incomplete models and inaccurate predictions. As far as we know, there is not much work related to modeling the combinational treatment effects in the two-sided market. In this paper, we first introduce the two-sided treatment effects estimation problem and then propose a Unified Treatment effect Estimation (UniTE) method for one-sided and two-sided marketing. We extend the Robinson Decomposition to two-sided, in which the relationship of the three involved tasks, namely the outcome, the propensity, and the treatment effect, is theoretically derived. Based on the decomposition result, a multi-task-based neural network model is proposed to integrate the three tasks and learn the inter-task-related common information, which prompts the model to estimate the treatment effects better. We also propose a unified synthetic data generation method that adapts to one/two-sided situations to verify the treatment effects estimation performance. Extensive and comprehensive experimental results show that our method outperforms the other methods.

#### **PopDCL: Popularity-aware Debiased Contrastive Loss for Collaborative Filtering**

Zhuang Liu, Haoxuan Li, Guanming Chen, Yuanxin Ouyang, Wenge Rong, Zhang Xiong

Collaborative filtering (CF) is the basic method for recommendation with implicit feedback. Recently, various state-of-the-art CF integrates graph neural networks. However, they often suffer from popularity bias, causing recommendations to deviate from users' genuine preferences. Additionally, several contrastive learning methods based on the in-batch sample strategy have been proposed to train the CF model effectively, but they are prone to suffering from sample bias. To address this problem, debiased contrastive loss has been employed in the recommendation, but instead of personalized debiasing, it treats each user equally. In this paper, we propose a popularity-aware debiased contrastive loss for CF, which can adaptively correct the positive and negative scores based on the popularity of users and items. Our approach aims to reduce the negative impact of popularity and sample bias simultaneously. We theoretically analyze the effectiveness of the proposed method and reveal the relationship between popularity and gradient, which justifies the correction strategy. We extensively evaluate our method on three public benchmarks over balanced and imbalanced settings. The results demonstrate its



superiority over the existing debiased strategies, not only on the entire datasets but also when segmenting the datasets based on item popularity.

### **AutoSeqRec: Autoencoder for Efficient Sequential Recommendation**

Sijia Liu, Jiahao Liu, Hansu Gu, Dongsheng Li, Tun Lu, Peng Zhang, Ning Gu

Sequential recommendation demonstrates the capability to recommend items by modeling the sequential behavior of users. Traditional methods typically treat users as sequences of items, overlooking the collaborative relationships among them. Graph-based methods incorporate collaborative information by utilizing the user-item interaction graph. However, these methods sometimes face challenges in terms of time complexity and computational efficiency. To address these limitations, this paper presents AutoSeqRec, an incremental recommendation model specifically designed for sequential recommendation tasks. AutoSeqRec is based on autoencoders and consists of an encoder and three decoders within the autoencoder architecture. These components consider both the user-item interaction matrix and the rows and columns of the item transition matrix. The reconstruction of the user-item interaction matrix captures user long-term preferences through collaborative filtering. In addition, the rows and columns of the item transition matrix represent the item out-degree and in-degree hopping behavior, which allows for modeling the user's short-term interests. When making incremental recommendations, only the input matrices need to be updated, without the need to update parameters, which makes AutoSeqRec very efficient. Comprehensive evaluations demonstrate that AutoSeqRec outperforms existing methods in terms of accuracy, while showcasing its robustness and efficiency.

### **MATA\*: Combining Learnable Node Matching with A\* Algorithm for Approximate Graph Edit Distance Computation**

Junfeng Liu, Min Zhou, Shuai Ma, Lujia Pan

Graph Edit Distance (GED) is a general and domain-agnostic metric to measure graph similarity, widely used in graph search or retrieving tasks. However, the exact GED computation is known to be NP-complete. For instance, the widely used A\* algorithms explore the entire search space to find the optimal solution which inevitably suffers scalability issues. Learning-based methods apply graph representation techniques to learn the GED by formulating a regression task, which can not recover the edit path and lead to inaccurate GED approximation (i.e., the predicted GED is smaller than the exact). To this end, in this work, we present a data-driven hybrid approach MATA\* for approximate GED computation based on Graph Neural Networks (GNNs) and A\* algorithms, which models from the perspective of learning to match nodes instead of directly regressing GED. Specifically, aware of the structure-dominant operations (i.e., node and edge insertion/deletion) property in GED computation, a structure-enhanced GNN is firstly designed to jointly learn local and high-order structural information for node embeddings for node matchings. Second, top-k candidate nodes are produced via a differentiable top-k operation to enable the training for node matchings, which is adhering to another property of GED, i.e., multiple optimal node matchings. Third, benefiting from the candidate nodes, MATA\* only performs on the promising search directions, reaching the solution efficiently. Finally, extensive experiments show the superiority of MATA\* as it significantly outperforms the combinatorial search-based, learning-based and hybrid methods and scales well to large-size graphs.

### **Leveraging Event Schema to Ask Clarifying Questions for Conversational Legal Case Retrieval**

Bulou Liu, Yiran Hu, Qingyao Ai, Yiqun Liu, Yueyue Wu, Chenliang Li, Weixing Shen

Legal case retrieval is a special IR task aiming to retrieve supporting cases for a given query case. Existing works have shown that conversational search paradigm can improve users' search experience in legal case retrieval. One of the keys to a practical conversational search system is how to ask high-quality clarifying questions to initiate conversations with users and understand their search intents. Recently, Large Language Models, such as ChatGPT and GPT-4, have shown superior ability in both open-domain QA and conversations with human. Thus it is natural to believe that they could be applied to legal conversational search as well. However, our preliminary study has shown that generating clarifying questions in legal conversational search with SOTA LLMs (e.g., GPT-4) often suffers from several problems such as duplication and low-utility contents. To address these problems, we propose LeClari, which leverages legal event schema as external knowledge to instruct LLMs to generate effective clarifying questions for legal conversational search. LeClari is constructed with a prompt module and a novel legal event selection module. The former defines a prompt with legal events for clarifying question generation and the latter selects potential event types by modeling the relationships of legal event types, conversational context, and candidate cases. We also propose ranking-oriented rewards and employ the reward augmented maximum likelihood (RAML) method to optimize LeClari directly based on the final retrieval performance of the conversational legal search system. Empirical results over two widely adopted legal case retrieval datasets demonstrate the effectiveness of our approach as compared with the state-of-the-art baselines.

### **ForeSeer: Product Aspect Forecasting Using Temporal Graph Embedding**

Zixuan Liu, Gaurush Hiranandani, Kun Qian, Edward W. Huang, Yi Xu, Belinda Zeng, Karthik Subbian, Sheng Wang

Developing text mining approaches to mine aspects from customer reviews has been well-studied due to its importance in understanding customer needs and product attributes. In contrast, it remains unclear how to predict the future emerging aspects of a new product that currently has little review information. This task, which we named product aspect forecasting, is critical for recommending new products, but also challenging because of the missing reviews. Here, we propose ForeSeer, a novel textual mining and product embedding approach progressively trained on temporal product graphs for this novel product aspect forecasting task. ForeSeer transfers reviews from similar products on a large product graph and exploits these reviews to predict aspects that might emerge in future reviews. A key novelty of our method is to jointly provide review, product, and aspect embeddings that are both time-sensitive and less affected by extremely imbalanced aspect frequencies. We evaluated ForeSeer on a real-world product review system containing 11,536,382 reviews and 11,000 products over 3 years. We observe that ForeSeer substantially outperformed existing approaches with at least 49.1% AUPRC improvement under the real setting where aspect associations are not given. ForeSeer further improves future link prediction on the product graph and the review aspect association prediction. Collectively, ForeSeer offers a novel framework for review forecasting by effectively integrating review text, product network, and temporal information, opening up new avenues for online shopping recommendation and e-commerce applications.

### **Text Matching Improves Sequential Recommendation by Reducing Popularity Biases**

Zhenghao Liu, Sen Mei, Chenyan Xiong, Xiaohua Li, Shi Yu, Zhiyuan Liu, Yu Gu, Ge Yu

This paper proposes Text mAtching based SequenTial rEcommenda-tion model (TASTE), which maps items and users in an embedding space and recommends items by matching their text representations. TASTE verbalizes items and user-item interactions using identifiers and attributes of items. To better characterize user behaviors, TASTE additionally proposes an attention sparsity method, which enables TASTE to model longer user-item interactions by reducing the self-attention computations during encoding. Our experiments show that TASTE outperforms the state-of-the-art methods on widely used sequential recommendation datasets. TASTE alleviates the cold start problem by representing long-tail items using full-text modeling and bringing the benefits of pretrained language models to recommendation systems. Our further analyses illustrate that TASTE significantly improves the recommendation accuracy by reducing the popularity bias of previous item id based recommendation models and returning more appropriate and text-relevant items to satisfy users. All codes are available at <https://github.com/OpenMatch/TASTE>.

### **Neural Personalized Topic Modeling for Mining User Preferences on Social Media**

Luyang Liu, Qunyang Lin, Haonan Tong, Hongyin Zhu, Ke Liu, Min Wang, Chuang Zhang

With the rapid development of web services, social media has been a prevalent and readily way for people to express themselves and share their daily lives. Consequently, numerous user-generated content is accumulated on social media platforms. These data usually contain rich information and knowledge for users, which is a viable source for user data mining. As one of the prevalent techniques in user data mining, mining personalized topics and discovering user preferences from social media data attract much interest in academic and industrial communities. The emerging Neural Topic Models (NTMs) have recently shown leading performance and scalability by employing neural networks. However, most existing NTMs usually model topics simply from observed document token information and do not explicitly take user preferences into the generative process, which inevitably fails to model personalized topics. To address this issue, we introduce Neural Personalized Topic Model (NPTM), a novel NTM that can discover personalized topics and user preferences. NPTM introduces a novel hybrid generative process for combining user preferences and contextualized document codes in modeling personalized topics. A transformer-based document encoder to obtain contextualized document codes. For user preference modeling, NPTM regards user-related information as trainable user embeddings, further determining user preferences over the topics. Following the proposed hybrid generative process, we present a module-wise asynchronous optimization strategy to get coherent topics and user preferences. Then, we apply our model to two challenging real-world social media post collections and compare them against several baseline methods to verify our contributions. The experimental results demonstrate the effectiveness of the proposed method.

### **Hierarchical Prompt Tuning for Few-Shot Multi-Task Learning**

Jingping Liu, Tao Chen, Zujie Liang, Haiyun Jiang, Yanghua Xiao, Feng Wei, Yuxi Qian, Zhenghong Hao, Bing Han

Prompt tuning has enhanced the performance of Pre-trained Language Models for multi-task learning in few-shot scenarios. However, existing studies fail to consider that the prompts among different layers in Transformer are different due to the diverse information learned at each layer. In general, the bottom layers in the model tend to capture low-level semantic or structural information, while the upper layers primarily acquire task-specific knowledge. Hence, we propose a novel hierarchical prompt tuning model for few-shot multi-task learning to capture this regularity. The designed model mainly consists of three types of prompts: shared prompts, auto-adaptive prompts, and task-specific prompts. Shared prompts facilitate the sharing of general information across all tasks. Auto-adaptive prompts dynamically select and integrate relevant prompt information from all tasks into the current task. Task-specific prompts concentrate on learning task-specific knowledge. To enhance the model's adaptability to diverse inputs, we introduce deep instance-aware language prompts as the foundation for constructing the above prompts. To evaluate the effectiveness of our proposed method, we conduct extensive experiments on multiple widely-used datasets. The experimental results demonstrate that the proposed method achieves state-of-the-art performance for multi-task learning in few-shot settings and outperforms ChatGPT in the full-data setting.

#### **SMEF: Social-aware Multi-dimensional Edge Features-based Graph Representation Learning for Recommendation**

Xiao Liu, Shunmei Meng, Qianmu Li, Lianying Qi, Xiaolong Xu, Wanchun Dou, Xuyun Zhang

Exploring user-item interaction cues is crucial for the performance of recommender systems. Explicit investigation of interaction cues is made possible by using graph-based models, where each user-item relationship is described by an edge, and the introduction of user-user social network. While existing graph-based recommendation methods use only a single-value edge to define the relationship between a pair of user and item, which limits the ability to represent complex user-item interactions. Furthermore, some social recommendation methods overlook the heterogeneous user behavior patterns in social and interaction relationships, resulting in the suboptimal performance of existing systems. In this paper, we propose a novel Social-aware Multi-dimensional Edge Feature-based Graph Representation Learning method, called SMEF. It represents all users and items as a graph and deep learns a multi-dimensional edge feature to explicitly describe the task-specific relationships of each user-item pair. Specifically, the proposed SMEF focuses on two distinct user behavior patterns toward social friends and interactive items, which explore the underlying heterogeneous relationship cues within them. This way, the learned multi-dimensional edge features encode user information from both social and interaction aspects. The proposed SMEF is a plug-and-play module that can be combined with different recommendation frameworks and Graph Neural Networks (GNNs) backbones to generate high quality user representations. The experimental results achieved on three publicly accessible datasets show that our SMEF-based method outperforms strong baselines.

#### **Diffusion Augmentation for Sequential Recommendation**

Qidong Liu, Fan Yan, Xiangyu Zhao, Zhaocheng Du, Huifeng Guo, Ruiming Tang, Feng Tian

Sequential recommendation (SRS) has become the technical foundation in many applications recently, which aims to recommend the next item based on the user's historical interactions. However, sequential recommendation often faces the problem of data sparsity, which widely exists in recommender systems. Besides, most users only interact with a few items, but existing SRS models often underperform these users. Such a problem, named the long-tail user problem, is still to be resolved. Data augmentation is a distinct way to alleviate these two problems, but they often need fabricated training strategies or are hindered by poor-quality generated interactions. To address these problems, we propose a Diffusion Augmentation for Sequential Recommendation (DiffuASR) for a higher quality generation. The augmented dataset by DiffuASR can be used to train the sequential recommendation models directly, free from complex training procedures. To make the best of the generation ability of the diffusion model, we first propose a diffusion-based pseudo sequence generation framework to fill the gap between image and sequence generation. Then, a sequential U-Net is designed to adapt the diffusion noise prediction model U-Net to the discrete sequence generation task. At last, we develop two guide strategies to assimilate the preference between generated and origin sequences. To validate the proposed DiffuASR, we conduct extensive experiments on three real-world datasets with three sequential recommendation models. The experimental results illustrate the effectiveness of DiffuASR. As far as we know, DiffuASR is one pioneer that introduce the diffusion model to the recommendation. The implementation code is available online.

#### **Weak Regression Enhanced Lifelong Learning for Improved Performance and Reduced Training Data**

Tong Liu, Xulong Wang, He Huang, Po Yang

As an emerging learning paradigm, lifelong learning intends to solve multiple consecutive tasks over long-time scales upon previously accumulated knowledge. When facing with a new task, existing lifelong learning approaches need first gather

sufficient training data to identify task relationships before knowledge transfer can succeed. However, annotating large number of training data persistently for every coming task is time-consuming, which can be prohibitive for real-world lifelong regression problems. To reduce this burden, we propose to incorporate weak regression into lifelong learning so as to enhance training data and improve predictive performance. Specifically, the weak prediction is first produced by single-task predictor, which is encoded as feature vectors that contain essential prior output information. This weak regression is further linked with task model via coupled dictionary learning. The integration of weak regression and task model can facilitate both cross-task and inter-task knowledge transfer, thus improving the overall performance. More critically, the weak regression can backup the task model especially when there is insufficient training data to construct an accurate model. Three real-world datasets are used to evaluate the effectiveness of our proposed method. Results show that our method outperforms existing lifelong models and single-task models even if training data is minimal.

#### **Online Efficient Secure Logistic Regression based on Function Secret Sharing**

Jing Liu, Jamie Cui, Cen Chen

Logistic regression is an algorithm widely used for binary classification in various real-world applications such as fraud detection, medical diagnosis, and recommendation systems. However, training a logistic regression model with data from different parties raises privacy concerns. Secure Multi-Party Computation (MPC) is a cryptographic tool that allows multiple parties to train a logistic regression model jointly without compromising privacy. The efficiency of the online training phase becomes crucial when dealing with large-scale data in practice. In this paper, we propose an online efficient protocol for privacy-preserving logistic regression based on Function Secret Sharing (FSS). Our protocols are designed in the two non-colluding servers setting and assume the existence of a third-party dealer who only poses correlated randomness to the computing parties. During the online phase, two servers jointly train a logistic regression model on their private data by utilizing pre-generated correlated randomness. Furthermore, we propose accurate and MPC-friendly alternatives to the sigmoid function and encapsulate the logistic regression training process into a function secret sharing gate. The online communication overhead significantly decreases compared with the traditional secure logistic regression training based on secret sharing. We provide both theoretical and experimental analyses to demonstrate the efficiency and effectiveness of our method.

#### **Khronos: A Real-Time Indexing Framework for Time Series Databases on Large-Scale Performance Monitoring Systems**

Xinyu Liu, Zijing Wei, Wenqing Yu, Shaozhi Liu, Gang Wang, Xiaoguang Liu, Yusen Li

Time series databases play a critical role in large-scale performance monitoring systems. Metrics are required to be observable immediately after being generated to support real-time analysis. However, the commonly used Log-Structured Merge-Tree structure suffers from periodically visible delay spikes when a new segment is created due to the instantaneous index construction pressure.

In this paper, we present Khronos, an asynchronous indexing framework tailored for high-cardinal monitoring data, aiming at reducing the visible delay. Firstly, we analyze the temporal locality nature of time series and propose a complementary index construction algorithm by only indexing series not reported before to relieve indexing workload. Secondly, we design index structures based on Minimum Excluded value function to effectively reuse indexes of previous segments. Thirdly, we take advantage of the non-repetitive feature of complementary indexes and further develop an intermediate query results reusing approach for deduplicating index traversal among segments. Moreover, we propose an index dependency management strategy that cuts off the previous reusing dependency before persistence to avoid extended dependency overhead.

Experimental results show that our framework significantly reduces the visible delay from minutes to milliseconds. Khronos outperforms the state-of-the-art databases InfluxDB and TimeScaleDB with at least 4 times higher write throughput, hundreds of times lower visible delay, and 6 times lower query latency. Khronos has been deployed in production since 2020 and has become the largest performance monitoring database in Alibaba.

#### **Self-Supervised Dynamic Hypergraph Recommendation based on Hyper-Relational Knowledge Graph**

Yi Liu, Hongrui Xuan, Bohan Li, Meng Wang, Tong Chen, Hongzhi Yin

Knowledge graphs (KGs) are commonly used as side information to enhance collaborative signals and improve recommendation quality. In the context of knowledge-aware recommendation (KGR), graph neural networks (GNNs) have emerged as promising solutions for modeling factual and semantic information in KGs. However, the long-tail distribution of entities leads to sparsity in supervision signals, which weakens the quality of item representation when utilizing KG enhancement. Additionally, the binary relation representation of KGs simplifies hyper-relational facts, making it challenging to

model complex real-world information. Furthermore, the over-smoothing phenomenon results in indistinguishable representations and information loss.

To address these challenges, we propose the SDK (Self-Supervised Dynamic Hypergraph Recommendation based on Hyper-Relational Knowledge Graph) framework. This framework establishes a cross-view hypergraph self-supervised learning mechanism for KG enhancement. Specifically, we model hyper-relational facts in KGs to capture interdependencies between entities under complete semantic conditions. With the refined representation, a hypergraph is dynamically constructed to preserve features in the deep vector space, thereby alleviating the over-smoothing problem. Furthermore, we mine external supervision signals from both the global perspective of the hypergraph and the local perspective of collaborative filtering (CF) to guide the model prediction process.

Extensive experiments conducted on different datasets demonstrate the superiority of the SDK framework over state-of-the-art models. The results showcase its ability to alleviate the effects of over-smoothing and supervision signal sparsity.

#### **Quantifying the Effectiveness of Advertising: A Bootstrap Proportion Test for Brand Lift Testing**

Wanjun Liu, Xiufan Yu, Jialiang Mao, Xiaoxu Wu, Justin Dyer

Brand Lift test is a widely deployed statistical tool for measuring the effectiveness of online advertisements on brand perception such as ad recall, brand familiarity and favorability. By formulating the problem of interest into a two-sample test on the binomial proportions from the control group ( $p_0$ ) and the treatment group ( $p_1$ ), Brand Lift test evaluates ads impact based on the statistical significance of test results. Traditional approaches construct the test statistics based on the absolute difference between the two observed proportions, a.k.a, absolute lift. In this work, we propose a new bootstrap test based on the percentage difference between the two observed proportions, i.e., relative lift. We provide rigorous theoretical guarantees on the asymptotic validity of the proposed relative-lift-based test. Our numerical studies suggest that the relative-lift-based test requires less stringent conditions than the absolute-lift-based test for controlling the type-I error rate. Interestingly, we also prove that the relative-lift-based test is more powerful than the absolute-lift-based test when the alternative is positive (i.e.,  $p_1 - p_0 > 0$ ), but less powerful when the alternative is negative (i.e.,  $p_1 - p_0 < 0$ ). The empirical performance of the proposed test is demonstrated by extensive simulation studies, an application to a publicly available A/B testing dataset from advertising, and real datasets collected from the Brand Lift Testing platform at LinkedIn.

#### **Deep Task-specific Bottom Representation Network for Multi-Task Recommendation**

Qi Liu, Zhilong Zhou, Gangwei Jiang, Tiezheng Ge, Defu Lian

Neural-based multi-task learning (MTL) has gained significant improvement, and it has been successfully applied to recommendation system (RS). Recent deep MTL methods for RS (e.g. MMoE, PLE) focus on designing soft gating-based parameter-sharing networks that implicitly learn a generalized representation for each task. However, MTL methods may suffer from performance degeneration when dealing with conflicting tasks, as negative transfer effects can occur on the task-shared bottom representation. This can result in a reduced capacity for MTL methods to capture task-specific characteristics, ultimately impeding their effectiveness and hindering the ability to generalize well on all tasks. In this paper, we focus on the bottom representation learning of MTL in RS and propose the Deep Task-specific Bottom Representation Network (DTRN) to alleviate the negative transfer problem. DTRN obtains task-specific bottom representation explicitly by making each task have its own representation learning network in the bottom representation modeling stage. Specifically, it extracts the user's interests from multiple types of behavior sequences for each task through the parameter-efficient hypernetwork. To further obtain the dedicated representation for each task, DTRN refines the representation of each feature by employing a SENet-like network for each task. The two proposed modules can achieve the purpose of getting task-specific bottom representation to relieve tasks' mutual interference. Moreover, the proposed DTRN is flexible to combine with existing MTL methods. Experiments on one public dataset and one industrial dataset demonstrate the effectiveness of the proposed DTRN.

#### **Black-box Adversarial Attacks against Dense Retrieval Models: A Multi-view Contrastive Learning Method**

Yu-An Liu, Ruqing Zhang, Jiafeng Guo, Maarten de Rijke, Wei Chen, Yixing Fan, Xueqi Cheng

Neural ranking models (NRMs) and dense retrieval (DR) models have given rise to substantial improvements in overall retrieval performance. In addition to their effectiveness, and motivated by the proven lack of robustness of deep learning-based approaches in other areas, there is growing interest in the robustness of deep learning-based approaches to the core retrieval problem. Adversarial attack methods that have so far been developed mainly focus on attacking NRMs, with very little attention being paid to the robustness of DR models.



In this paper, we introduce the adversarial retrieval attack (AREA) task. The AREA task is meant to trick DR models into retrieving a target document that is outside the initial set of candidate documents retrieved by the DR model in response to a query. We consider the decision-based black-box adversarial setting, which is realistic in real-world search engines. To address the AREA task, we first employ existing adversarial attack methods designed for NRMs. We find that the promising results that have previously been reported on attacking NRMs, do not generalize to DR models: these methods underperform a simple term spamming method. We attribute the observed lack of generalizability to the interaction-focused architecture of NRMs, which emphasizes fine-grained relevance matching. DR models follow a different representation-focused architecture that prioritizes coarse-grained representations. We propose to formalize attacks on DR models as a contrastive learning problem in a multi-view representation space. The core idea is to encourage the consistency between each view representation of the target document and its corresponding viewer via view-wise supervision signals. Experimental results demonstrate that the proposed method can significantly outperform existing attack strategies in misleading the DR model with small indiscernible text perturbations.

### **BRep-BERT: Pre-training Boundary Representation BERT with Sub-graph Node Contrastive Learning**

Yunzhong Lou, Xueyang Li, Haotian Chen, Xiangdong Zhou

Obtaining effective entity feature representations is crucial in the field of Boundary Representation (B-Rep), a key parametric representation method in Computer-Aided Design (CAD). However, the lack of labeled large-scale database and the scarcity of task-specific label sets pose significant challenges. To address these problems, we propose an innovative unsupervised neural network approach called BRep-BERT, which extends the concept of BERT to the B-Rep domain. Specifically, we utilize Graph Neural Network (GNN) Tokenizer to generate discrete entity labels with geometric and structural semantic information. We construct new entity representation sequences based on the structural relationships and pre-train the model through the Masked Entity Modeling (MEM) task. To address the attention sparsity issue in large-scale geometric models, we incorporate graph structure information and learnable relative position encoding into the attention module to optimize feature updates. Additionally, we employ geometric sub-graphs and multi-level contrastive learning techniques to enhance the model's ability to learn regional features. Comparisons with previous methods demonstrate that BRep-BERT achieves the state-of-the-art performance on both full-data training and few-shot learning tasks across multiple B-Rep datasets. Particularly, BRep-BERT outperforms previous methods significantly in the few-shot learning scenarios. Comprehensive experiments demonstrate the substantial advantages and potential of BRep-BERT in handling B-Rep data representation. Code will be released at [https://github.com/louyz1026/Brep\\_Bert](https://github.com/louyz1026/Brep_Bert).

### **Exploring Low-Dimensional Manifolds of Deep Neural Network Parameters for Improved Model Optimization**

Ke Lu, Xiaotong He, Ze Qin, Xinyao Li, Zhekai Du

Manifold learning techniques have significantly enhanced the comprehension of massive data by exploring the geometric properties of the data manifold in low-dimensional subspaces. However, existing research on manifold learning primarily focuses on understanding the intricate data, overlooking the explosive growth of the scale and complexity of deep neural networks (DNNs), which presents a significant challenge for model optimization. In this work, we propose to explore the intrinsic low-dimensional manifold of network parameters for efficient model optimization. Specifically, we analyze parameter distributions in a deep model and perform sampling to map them onto a low-dimensional parameter manifold using the local tangent space alignment (LTSa). Since our focus is on studying parameter manifolds to guide model optimization, we therefore select dynamic optimal training trajectories for sampling and approximate tangent spaces to obtain low-dimensional representations of DNNs. By applying manifold learning techniques and employing a two-step alternate optimization method, we achieve a fixed subspace that reduces training time and resource costs for commonly used deep networks. The trained low-dimensional network can be mapped back to the original parameter space for further use. We demonstrate the benefits of learning low-dimensional parameterization of DNNs on both noisy label learning and federated learning tasks. Extensive experimental results on various benchmarks show the effectiveness of our method concerning both superior accuracy and reduced resource consumption.

### **Selecting Walk Schemes for Database Embedding**

Yuval Lev Lubarsky, Jan Tönshof, Martin Grohe, Benny Kimelfeld

Machinery for data analysis often requires a numeric representation of the input. Towards that, a common practice is to embed components of structured data into a high-dimensional vector space. We study the embedding of the tuples of a relational database, where existing techniques are often based on optimization tasks over a collection of random walks from the database. The focus of this paper is on the recent ForWaRD algorithm that is designed for dynamic databases, where

walks are sampled by following foreign keys between tuples. Importantly, different walks have different schemas, or "walk schemes," that are derived by listing the relations and attributes along the walk. Also importantly, different walk schemes describe relationships of different natures in the database.

We show that by focusing on a few informative walk schemes, we can obtain tuple embedding significantly faster, while retaining the quality. We define the problem of scheme selection for tuple embedding, devise several approaches and strategies for scheme selection, and conduct a thorough empirical study of the performance over a collection of downstream tasks. Our results confirm that with effective strategies for scheme selection, we can obtain high-quality embeddings considerably (e.g., three times) faster, preserve the extensibility to newly inserted tuples, and even achieve an increase in the precision of some tasks.

#### [Forward Creation, Reverse Selection: Achieving Highly Pertinent Multimodal Responses in Dialogue Contexts](#)

Ge Luo, Manman Zhang, Yuchen Ma, Sheng Li, Zhenxing Qian, Xinpeng Zhang

Multimodal Dialogue agents are often required to respond to conversation history using both textual and visual content. Even though current dialogue studies predominantly strive to generate natural texts or images, they fall short in considering the relevance of multimodal responses within a dialogue context, consequently confining agents from making prudent choices based on multiple alternatives and their associated relevance scores for decision-making. In this paper, we present a bidirectional multimodal dialogue framework that skillfully combines the forward generation of multiple text and image response candidates with reverse selection guided by relevance scores evaluated on dialogue context, facilitating agents in selecting the most suitable multimodal responses. Specifically, the forward generation aspect of our framework leverages a stage-wise approach, first producing textual replies and composite visual descriptions from the dialogue context, followed by the generation of visual responses aligned with the descriptions. In the reverse selection process, visual responses are translated into tangible descriptive texts that, in conjunction with textual responses, are inversely tied back to the dialogue context for relevance assessment, assigning a reference score to each multimodal response candidate to assist the intelligent agent in making informed decisions. Experimental outcomes demonstrate that our proposed bidirectional dialogue response framework markedly elevates performance in both automatic and human evaluations, yielding a range of contextually fitting multimodal responses for selection.

#### [Timestamps as Prompts for Geography-Aware Location Recommendation](#)

Yan Luo, Haoyi Duan, Ye Liu, Fu-Lai Chung

Location recommendation plays a vital role in improving users' travel experience. The timestamp of the POI to be predicted is of great significance, since a user will go to different places at different times. However, most existing methods either do not use this kind of temporal information, or just implicitly fuse it with other contextual information. In this paper, we revisit the problem of location recommendation and point out that explicitly modeling temporal information is a great help when the model needs to predict not only the next location but also further locations. In addition, state-of-the-art methods do not make effective use of geographic information and suffer from the hard boundary problem when encoding geographic information by gridding. To this end, a Temporal Prompt-based and Geography-aware (TPG) framework is proposed. The temporal prompt is firstly designed to incorporate temporal information of any further check-in. A shifted window mechanism is then devised to augment geographic data for addressing the hard boundary problem. Via extensive comparisons with existing methods and ablation studies on five real-world datasets, we demonstrate the effectiveness and superiority of the proposed method under various settings. Most importantly, our proposed model has the superior ability of interval prediction. In particular, the model can predict the location that a user wants to go to at a certain time while the most recent check-in behavioral data is masked, or it can predict specific future check-in (not just the next one) at a given timestamp.

#### [Improving Long-Tail Item Recommendation with Graph Augmentation](#)

Sichun Luo, Chen Ma, Yuanzhang Xiao, Linqi Song

The ubiquitous long-tail distribution of inherent user behaviors results in worse recommendation performance for the items with fewer user records (i.e., tail items) than those with richer ones (i.e., head items). Graph-based recommendation methods (e.g., using graph neural networks) have recently emerged as a powerful tool for recommender systems, often outperforming traditional methods. However, existing techniques for alleviating the long-tail problem mainly focus on traditional methods. There is a lack of graph-based methods that can efficiently deal with the long-tail problem.

In this paper, we propose a novel approach, Graph Augmentation for Long-tail Recommendation (GALORE), which can be plugged into any graph-based recommendation models to improve the performance for tail items. GALORE incorporates an

edge addition module that enriches the graph's connectivity for tail items by injecting additional item-to-item edges. To further balance the graph structure, GALORE utilizes a degree-aware edge dropping strategy, preserving the more valuable edges from the tail items while selectively discarding less informative edges from the head items. Beyond structural augmentation, we synthesize new data samples, thereby addressing the data scarcity issue for tail items. We further introduce a two-stage training strategy to facilitate the learning for both head and tail items. Comprehensive empirical studies conducted on four datasets show that GALORE outperforms existing methods in terms of the performance for tail items as well as the overall performance.

#### [Context-Aware Prompt for Generation-based Event Argument Extraction with Diffusion Models](#)

Lei Luo, Yajing Xu

Event argument extraction (EAE) has attracted increasing attention via generation-based methods. However, most existing works tend to independently extract arguments for each role, ignoring the correlation between different arguments, especially in long contexts. Motivated by these observations and the high-quality generation results of recent diffusion models, we propose an effective model called PGAD (Context-Aware Prompt for Generation-based EAE with Diffusion models) for both sentence-level and document-level EAE. In PGAD, a text diffusion model is designed to generate diverse context-aware prompt representations in conjunction with a series of random Gaussian noise. Firstly, cross-attention is employed between the designed prompt and input context within the text diffusion model in order to generate the context-aware prompt. Through this interaction, the context-aware prompt is able to capture multiple role-specific argument span queriers. Secondly, the context-aware prompt is aligned with the context to generate event arguments by joint optimization. Extensive experiments on three publicly available EAE datasets demonstrate the superiority of our proposed PGAD model over existing approaches.

#### [Integrating Priors into Domain Adaptation Based on Evidence Theory](#)

Ying Lv, Jianpeng Ma, Yiqiu Zhang, Gang Xu

Domain adaptation aims to build up a learning model for target domain by leveraging transferable knowledge from different but related source domains. Existing domain adaptation methods generally transfer the knowledge from source domain to target domain through measuring the consistency between the different domains. Under this strategy, if the data of source domain is not sufficient to guarantee the consistency, the transferable knowledge will be very limited. On the other hand, we often have priors about target domain which facilitate knowledge transfer but are neglected in the extant domain adaptation methods. To tackle the problems, we integrate the priors of target domain into transfer process and propose a domain adaptation method based on evidence theory. We represent the priors with evidential belief function and reformulate the domain adaptation objective based on likelihood principle, in which the priors are used to adjust transferred knowledge to suit for target domain. Based on this, we propose an improved coordinate ascent algorithm to optimize likelihood objective of domain adaption. Experimental results on both text and image datasets validate that the proposed method is effective to improve the knowledge transferability in domain adaptation, especially when the source domain is limited.

#### [Multi-scale Graph Pooling Approach with Adaptive Key Subgraph for Graph Representations](#)

Yiqin Lv, Zhiliang Tian, Zheng Xie, Yiping Song

The recent progress in graph representation learning boosts the development of many graph classification tasks, such as protein classification and social network classification. One of the mainstream approaches for graph representation learning is the hierarchical pooling method. It learns the graph representation by gradually reducing the scale of the graph, so it can be easily adapted to large-scale graphs. However, existing graph pooling methods discard the original graph structure during downsizing the graph, resulting in a lack of graph topological structure. In this paper, we propose a multi-scale graph neural network (MSGNN) model that not only retains the topological information of the graph but also maintains the key-subgraph for better interpretability. MSGNN gradually discards the unimportant nodes and retains the important subgraph structure during the iteration. The key subgraphs are first chosen by experience and then adaptively evolved to tailor specific graph structures for downstream tasks. The extensive experiments on seven datasets show that MSGNN improves the SOTA performance on graph classification and better retains key subgraphs.

#### [A Principled Decomposition of Pointwise Mutual Information for Intention Template Discovery](#)

Denghao Ma, Kevin Chen-Chuan Chang, Yueguo Chen, Xueqiang Lv, Liang Shen

With the rise of Artificial Intelligence (AI), question answering systems have become common for users to interact with computers, e.g., ChatGPT and Siri. These systems require a substantial amount of labeled data to train their models.

However, the labeled data is scarce and challenging to be constructed. The construction process typically involves two stages: discovering potential sample candidates and manually labeling these candidates. To discover high-quality candidate samples, we study the intention paraphrase template discovery task: Given some seed questions or templates of an intention, discover new paraphrase templates that describe the intention and are diverse to the seeds enough in text. As the first exploration of the task, we identify the new quality requirements, i.e., relevance, divergence and popularity, and identify the new challenges, i.e., the paradox of divergent yet relevant paraphrases, and the conflict of popular yet relevant paraphrases. To untangle the paradox of divergent yet relevant paraphrases, in which the traditional bag of words falls short, we develop usage-centric modeling, which represents a question/template/answer as a bag of usages that users engaged (e.g., up-votes), and uses a usage-flow graph to interrelate templates, questions and answers. To balance the conflict of popular yet relevant paraphrases, we propose a new and principled decomposition for the well-known Pointwise Mutual Information from the usage perspective (usage-PMI), and then develop a Bayesian inference framework over the usage-flow graph to estimate the usage-PMI. Extensive experiments over three large CQA corpora show strong performance advantage over the baselines adopted from paraphrase identification task. We release 885,000 paraphrase templates of high quality discovered by our proposed PMI decomposition model, and the data is available in site [https://github.com/Para-Questions/Intention\\\_template\\\_discovery](https://github.com/Para-Questions/Intention\_template\_discovery).

### **Rethinking Sensors Modeling: Hierarchical Information Enhanced Traffic Forecasting**

Qian Ma, Zijian Zhang, Xiangyu Zhao, Haoliang Li, Hongwei Zhao, Yiqi Wang, Zitao Liu, Wanyu Wang

With the acceleration of urbanization, traffic forecasting has become an essential role in smart city construction. In the context of spatio-temporal prediction, the key lies in how to model the dependencies of sensors. However, existing works basically only consider the micro relationships between sensors, where the sensors are treated equally, and their macroscopic dependencies are neglected. In this paper, we argue to rethink the sensor's dependency modeling from two hierarchies: regional and global perspectives. Particularly, we merge original sensors with high intra-region correlation as a region node to preserve the inter-region dependency. Then, we generate representative and common spatio-temporal patterns as global nodes to reflect a global dependency between sensors and provide auxiliary information for spatio-temporal dependency learning. In pursuit of the generality and reality of node representations, we incorporate a Meta GCN to calibrate the regional and global nodes in the physical data space. Furthermore, we devise the cross-hierarchy graph convolution to propagate information from different hierarchies. In a nutshell, we propose a Hierarchical Information Enhanced Spatio-Temporal prediction method, HIENT, to create and utilize the regional dependency and common spatio-temporal patterns. Extensive experiments have verified the leading performance of our HIENT against state-of-the-art baselines. We publicize the code to ease reproducibility<sup>1</sup>.

### **MultiCAD: Contrastive Representation Learning for Multi-modal 3D Computer-Aided Design Models**

Weijian Ma, Minyang Xu, Xueyang Li, Xiangdong Zhou

CAD models are multimodal data where information and knowledge contained in construction sequences and shapes are complementary to each other and representation learning methods should consider both of them. Such traits have been neglected in previous methods learning unimodal representations. To leverage the information from both modalities, we develop a multimodal contrastive learning strategy where features from different modalities interact via contrastive learning paradigm, driven by a novel multimodal contrastive loss. Two pretext tasks on both geometry and sequence domains are designed along with a two-stage training strategy to make the representation focus on encoding geometric details and decoding representations into construction sequences, thus being more applicable to downstream tasks such as multimodal retrieval and CAD sequence reconstruction. Experimental results show that the performance of our multimodal representation learning scheme has surpassed the baselines and unimodal methods significantly.

### **LambdaRank Gradients are Incoherent**

Federico Marcuzzi, Claudio Lucchese, Salvatore Orlando

In Information Retrieval (IR), the Learning-to-Rank (LTR) task requires building a ranking model that optimises a specific IR metric. One of the most effective approaches to do so is the well-known LambdaRank algorithm. LambdaRank uses gradient descent optimisation, and at its core, it defines approximate gradients, the so-called lambdas, for a non-differentiable IR metric. Intuitively, each lambda describes how much a document's score should be "pushed" up/down to reduce the ranking error.

In this work, we show that lambdas may be incoherent w.r.t. the metric being optimised: e.g., a document with high relevance in the ground truth may receive a smaller gradient push than a document with lower relevance. This behaviour goes far beyond the expected degree of approximation. We analyse such behaviour of LambdaRank gradients and we introduce some strategies to reduce their incoherencies. We demonstrate through extensive experiments, conducted using publicly available datasets, that the proposed approach reduces the frequency of the incoherencies in LambdaRank and derivatives, and leads to models that achieve statistically significant improvements in the NDCG metric, without compromising the training efficiency.

#### [A Graph Neural Network Model for Concept Prerequisite Relation Extraction](#)

Debjani Mazumder, Jiaul H. Paik, Anupam Basu

In recent years, with the emergence of online learning platforms and e-learning resources, many documents are available for a particular topic. For a better learning experience, the learner often needs to know and learn first the prerequisite concepts for a given concept. Traditionally, the identification of such prerequisite concepts is done manually by subject experts, which in turn, often limits self-paced learning. Recently, machine learning models have found encouraging success for the task, obviating manual effort. In this paper, we propose a graph neural network based approach that leverages node attention over a heterogeneous graph to extract the prerequisite concepts for a given concept. Experiments on a set of benchmark data show that the proposed model outperforms the existing models by large margins almost always, making the model a new state-of-the-art for the task.

#### [Parallel Knowledge Enhancement based Framework for Multi-behavior Recommendation](#)

Chang Meng, Chenhao Zhai, Yu Yang, Hengyu Zhang, Xiu Li

Multi-behavior recommendation algorithms aim to leverage the multiplex interactions between users and items to learn users' latent preferences. Recent multi-behavior recommendation frameworks contain two steps: fusion and prediction. In the fusion step, advanced neural networks are used to model the hierarchical correlations between user behaviors. In the prediction step, multiple signals are utilized to jointly optimize the model with a multi-task learning (MTL) paradigm. However, recent approaches have not addressed the issue caused by imbalanced data distribution in the fusion step, resulting in the learned relationships being dominated by high-frequency behaviors. In the prediction step, the existing methods use a gate mechanism to directly aggregate expert information generated by coupling input, leading to negative information transfer. To tackle these issues, we propose a Parallel Knowledge Enhancement Framework (PKEF) for multi-behavior recommendation. Specifically, we enhance the hierarchical information propagation in the fusion step using parallel knowledge (PKF). Meanwhile, in the prediction step, we decouple the representations to generate expert information and introduce a projection mechanism during aggregation to eliminate gradient conflicts and alleviate negative transfer (PME). We conduct comprehensive experiments on three real-world datasets to validate the effectiveness of our model. The results further demonstrate the rationality and effectiveness of the designed PKF and PME modules. The source code and datasets are available at <https://github.com/MC-CV/PKEF>.

#### [System Initiative Prediction for Multi-turn Conversational Information Seeking](#)

Chuan Meng, Mohammad Aliannejadi, Maarten de Rijke

Identifying the right moment for a system to take the initiative is essential to conversational information seeking (CIS). Existing studies have extensively studied the clarification need prediction task, i.e., predicting when to ask a clarifying question, however, it only covers one specific system-initiative action. We define the system initiative prediction (SIP) task as predicting whether a CIS system should take the initiative at the next turn. Our analysis reveals that for effective modeling of SIP, it is crucial to capture dependencies between adjacent user/system initiative-taking decisions. We propose to model SIP by CRFs. Due to their graphical nature, CRFs are effective in capturing such dependencies and have greater transparency than more complex methods, e.g., LLMs. Applying CRFs to SIP comes with two challenges: (i) CRFs need to be given the unobservable system utterance at the next turn, and (ii) they do not explicitly model multi-turn features. We model SIP as an input-incomplete sequence labeling problem and propose a multi-turn system initiative predictor (MuSIC) that has (i) prior-posterior inter-utterance encoders to eliminate the need to be given the unobservable system utterance, and (ii) a multi-turn feature-aware CRF layer to incorporate multi-turn features into the dependencies between adjacent initiative-taking decisions. Experiments show that MuSIC outperforms LLM-based baselines including LLaMA, achieving state-of-the-art results on SIP. We also show the benefits of SIP on clarification need prediction and action prediction.

#### [Disparity, Inequality, and Accuracy Tradeoffs in Graph Neural Networks for Node Classification](#)



Arpit Merchant, Carlos Castillo

Graph neural networks (GNNs) are increasingly used in critical human applications for predicting node labels in attributed graphs. Their ability to aggregate features from nodes' neighbors for accurate classification also has the capacity to exacerbate existing biases in data or to introduce new ones towards members from protected demographic groups. Thus, it is imperative to quantify how GNNs may be biased and to what extent their harmful effects may be mitigated. To this end, we propose two new GNN-agnostic interventions namely, (i) PFR-AX which decreases the separability between nodes in protected and non-protected groups, and (ii) PostProcess which updates model predictions based on a blackbox policy to minimize differences between error rates across demographic groups. Through a large set of experiments on four datasets, we frame the efficacies of our approaches (and three variants) in terms of their algorithmic fairness-accuracy tradeoff and benchmark our results against three strong baseline interventions on three state-of-the-art GNN models. Our results show that no single intervention offers a universally optimal tradeoff, but PFR-AX and PostProcess provide granular control and improve model confidence when correctly predicting positive outcomes for nodes in protected groups.

### **DebCSE: Rethinking Unsupervised Contrastive Sentence Embedding Learning in the Debiasing Perspective**

Pu Miao, Zeyao Du, Junlin Zhang

Several prior studies have suggested that word frequency biases can cause the Bert model to learn indistinguishable sentence embeddings. Contrastive learning schemes such as SimCSE and ConSERT have already been adopted successfully in unsupervised sentence embedding to improve the quality of embeddings by reducing this bias. However, these methods still introduce new biases such as sentence length bias and false negative sample bias, that hinders model's ability to learn more fine-grained semantics. In this paper, we reexamine the challenges of contrastive sentence embedding learning from a debiasing perspective and argue that effectively eliminating the influence of various biases is crucial for learning high-quality sentence embeddings. We think all those biases are introduced by simple rules for constructing training data in contrastive learning and the key for contrastive learning sentence embedding is to "mimic" the distribution of training data in supervised machine learning in unsupervised way. We propose a novel contrastive framework for sentence embedding, termed DebCSE, which can eliminate the impact of these biases by an inverse propensity weighted sampling method to select high-quality positive and negative pairs according to both the surface and semantic similarity between sentences. Extensive experiments on semantic textual similarity (STS) benchmarks reveal that DebCSE significantly outperforms the latest state-of-the-art models with an average Spearman's correlation coefficient of 80.33% on BERTbase.

### **Hybrid Contrastive Constraints for Multi-Scenario Ad Ranking**

Shanlei Mu, Penghui Wei, Wayne Xin Zhao, Shaoguo Liu, Liang Wang, Bo Zheng

Multi-scenario ad ranking aims at leveraging the data from multiple domains or channels for training a unified ranking model for improving the performance at each individual scenario. Although the research on this task has made important progress, it still lacks the consideration of cross-scenario relations, thus leading to limitation in learning capability and difficulty in interrelation modeling.

In this paper, we propose a Hybrid Contrastive Constrained approach (HC2) for multi-scenario ad ranking. To enhance the modeling of data interrelation, we elaborately design a hybrid contrastive learning approach to capturing commonalities and differences among multiple scenarios. The core of our approach consists of two elaborated contrastive losses, namely generalized and individual contrastive loss, which aim at capturing common knowledge and scenario-specific knowledge, respectively. To adapt contrastive learning to the complex multi-scenario setting, we propose a series of important improvements. For generalized contrastive loss, we enhance contrastive learning by extending the contrastive samples (label-aware and diffusion noise enhanced contrastive samples) and reweighting the contrastive samples (reciprocal similarity weighting). For individual contrastive loss, we use the strategies of dropout-based augmentation and cross-scenario encoding for generating meaningful positive and negative contrastive samples, respectively. Extensive experiments on both offline evaluation and online test have demonstrated the effectiveness of the proposed HC<sup>2</sup> by comparing it with a number of competitive baselines.

### **Unlocking the Potential of Non-PSD Kernel Matrices: A Polar Decomposition-based Transformation for Improved Prediction Models**

Maximilian Münch, Manuel Röder, Frank-Michael Schleif

Kernel functions are a key element in many machine learning methods to capture the similarity between data points. However, a considerable number of these functions do not meet all mathematical requirements to be a valid positive semi-

definite kernel, a crucial precondition for kernel-based classifiers such as Support Vector Machines or Kernel Fisher Discriminant classifiers. In this paper, we propose a novel strategy employing a polar decomposition to effectively transform invalid kernel matrices to positive semi-definite matrices, while preserving the topological structure inherent to the data points. Utilizing polar decomposition allows the effective transformation of indefinite kernel matrices from Krein space to positive semi-definite matrices in Hilbert space, thereby providing an efficient out-of-sample extension for new unseen data and enhancing kernel method applicability across diverse classification tasks. We evaluate our approach on a variety of benchmark datasets and demonstrate its superiority over competitive methods.

#### [Joint Link Prediction Via Inference from a Model](#)

Parmis Naddaf, Erfaneh Mahmoudzaheh Ahmadi Nejad, Kiarash Zahirnia, Manfred Jaeger, Oliver Schulte

A Joint Link Prediction Query (JLPQ) specifies a set of links to be predicted, given another set of links as well as node attributes as evidence. While single link prediction has been well studied in literature on deep graph learning, predicting multiple links together has gained little attention. This paper presents a novel framework for computing JLPQs using a probabilistic deep Graph Generative Model. Specifically, we develop inference procedures for an inductively trained Variational Graph Auto-Encoder (VGAE) that estimates the joint link probability for any input JLPQ, without retraining. For evaluation, we apply inference to a range of joint link prediction queries on six benchmark datasets. We find that for most datasets and query types, joint link prediction via inference from a model achieves good predictive performance, better than the independent link prediction baselines (by 0.02-0.4 AUC points depending on the dataset).

#### [Non-Uniform Adversarial Perturbations for Discrete Tabular Datasets](#)

Jay Nandy, Jatin Chauhan, Rishi Saket, Aravindan Raghuvier

We study the problem of adversarial attack and robustness on tabular datasets with discrete features. The discrete features of a tabular dataset represent high-level meaningful concepts, with different sets of vocabularies, leading to requiring non-uniform robustness. Further, the notion of distance between tabular input instances is not well defined, making the problem of producing adversarial examples with minor perturbations qualitatively more challenging compared to existing methods. Towards this, our paper defines the notion of distance through the lens of feature embeddings, learnt to represent the discrete features. We then formulate the task of generating adversarial examples as a binary set selection problem under non-uniform feature importance. Next, we propose an efficient approximate gradient-descent based algorithm, called Discrete Non-uniform Approximation (DNA) attack, by reformulating the problem into a continuous domain to solve the original optimization problem for generating adversarial examples. We demonstrate the effectiveness of our proposed DNA attack using two large real-world discrete tabular datasets from e-commerce domains for binary classification, where the datasets are heavily biased for one-class. We also analyze challenges for existing adversarial training frameworks for such datasets under our DNA attack.

#### [Contrastive Learning of Temporal Distinctiveness for Survival Analysis in Electronic Health Records](#)

Mohsen Nayebi Kerdabadi, Arya Hadizadeh Moghaddam, Bin Liu, Mei Liu, Zijun Yao

Survival analysis plays a crucial role in many healthcare decisions, where the risk prediction for the events of interest can support an informative outlook for a patient's medical journey. Given the existence of data censoring, an effective way of survival analysis is to enforce the pairwise temporal concordance between censored and observed data, aiming to utilize the time interval before censoring as partially observed time-to-event labels for supervised learning. Although existing studies mostly employed ranking methods to pursue an ordering objective, contrastive methods which learn a discriminative embedding by having data contrast against each other, have not been explored thoroughly for survival analysis. Therefore, in this paper, we propose a novel Ontology-aware Temporality-based Contrastive Survival (OTCSurv) analysis framework that utilizes survival durations from both censored and observed data to define temporal distinctiveness and construct negative sample pairs with adjustable hardness for contrastive learning. Specifically, we first use an ontological encoder and a sequential self-attention encoder to represent the longitudinal EHR data with rich contexts. Second, we design a temporal contrastive loss to capture varying survival durations in a supervised setting through a hardness-aware negative sampling mechanism. Last, we incorporate the contrastive task into the time-to-event predictive task with multiple loss components. We conduct extensive experiments using a large EHR dataset to forecast the risk of hospitalized patients who are in danger of developing acute kidney injury (AKI), a critical and urgent medical condition. The effectiveness and explainability of the proposed model are validated through comprehensive quantitative and qualitative studies.

#### [Measuring the Effect of Causal Disentanglement on the Adversarial Robustness of Neural Network Models](#)

Preben M. Ness, Dusica Marijan, Sunanda Bose

Causal Neural Network models have shown high levels of robustness to adversarial attacks as well as an increased capacity for generalisation tasks such as few-shot learning and rare-context classification compared to traditional Neural Networks. This robustness is argued to stem from the disentanglement of causal and confounder input signals. However, no quantitative study has yet measured the level of disentanglement achieved by these types of causal models or assessed how this relates to their adversarial robustness.

Existing causal disentanglement metrics are not applicable to deterministic models trained on real-world datasets. We, therefore, utilise metrics of content/style disentanglement from the field of Computer Vision to measure different aspects of the causal disentanglement for four state-of-the-art causal Neural Network models. By re-implementing these models with a common ResNet18 architecture we are able to fairly measure their adversarial robustness on three standard image classification benchmarking datasets under seven common white-box attacks. We find a strong association ( $r=0.820$ ,  $p=0.001$ ) between the degree to which models decorrelate causal and confounder signals and their adversarial robustness. Additionally, we find a moderate negative association between the pixel-level information content of the confounder signal and adversarial robustness ( $r=-0.597$ ,  $p=0.040$ ).

### **Multi-domain Recommendation with Embedding Disentangling and Domain Alignment**

Wentao Ning, Xiao Yan, Weiwen Liu, Reynold Cheng, Rui Zhang, Bo Tang

Multi-domain recommendation (MDR) aims to provide recommendations for different domains (e.g., types of products) with overlapping users/items and is common for platforms such as Amazon, Facebook, and LinkedIn that host multiple services. Existing MDR models face two challenges: First, it is difficult to disentangle knowledge that generalizes across domains (e.g., a user likes cheap items) and knowledge specific to a single domain (e.g., a user likes blue clothing but not blue cars). Second, they have limited ability to transfer knowledge across domains with small overlaps. We propose a new MDR method named EDDA with two key components, i.e., embedding disentangling recommender and domain alignment, to tackle the two challenges respectively. In particular, the embedding disentangling recommender separates both the model and embedding for the inter-domain part and the intra-domain part, while most existing MDR methods only focus on model-level disentanglement. The domain alignment leverages random walks from graph processing to identify similar user/item pairs from different domains and encourages similar user/item pairs to have similar embeddings, enhancing knowledge transfer. We compare EDDA with 12 state-of-the-art baselines on 3 real datasets. The results show that EDDA consistently outperforms the baselines on all datasets and domains. All datasets and codes are available at <https://github.com/Stevenn9981/EDDA>.

### **MUSE: Music Recommender System with Shuffle Play Recommendation Enhancement**

Yunhak Oh, Sukwon Yun, Dongmin Hyun, Sein Kim, Chanyoung Park

Recommender systems have become indispensable in music streaming services, enhancing user experiences by personalizing playlists and facilitating the serendipitous discovery of new music. However, the existing recommender systems overlook the unique challenges inherent in the music domain, specifically shuffle play, which provides subsequent tracks in a random sequence. Based on our observation that the shuffle play sessions hinder the overall training process of music recommender systems mainly due to the high unique transition rates of shuffle play sessions, we propose a Music Recommender System with Shuffle Play Recommendation Enhancement (MUSE). MUSE employs the self-supervised learning framework that maximizes the agreement between the original session and the augmented session, which is augmented by our novel session augmentation method, called transition-based augmentation. To further facilitate the alignment of the representations between the two views, we devise two fine-grained matching strategies, i.e., item- and similarity-based matching strategies. Through rigorous experiments conducted across diverse environments, we demonstrate MUSE's efficacy over 12 baseline models on a large-scale Music Streaming Sessions Dataset (MSSD) from Spotify. The source code of MUSE is available at <https://github.com/yunhak0/MUSE>.

### **Dual-Oriented Contrast for Recommendation with A Stop-Gradient Operation**

Byungkook Oh, Yul Kim, Bumky Min

Recently, contrastive loss is adopted as a main objective of recommender systems. InfoNCE-like losses penalize hard negative items more and control the strength of penalties with a temperature, called hardness-aware sensitivity. However, since they leverage user  $\rightarrow$  item patterns in a non-symmetric way, negative items are pushed away from anchor users and attract semantically-similar items to each other, focusing on the distribution of item embeddings. We point out that user embeddings also have inherent semantic structures that can be captured from item  $\rightarrow$  user patterns. This paper presents Dual-

oriented Contrast(DuCo), a novel symmetric learning objective for recommendation to learn more comprehensive representations from user-item patterns. DuCo controls user-/item-centric hardness-aware sensitivities and simultaneously optimizes the score distributions over sampled items (user-oriented contrast) and users (item-oriented contrast). This aims to explore ideal user and item distributions that are locally clustered and globally uniform. However, since user-/item-side temperatures are interdependent, naive control over temperatures may break the underlying semantic structures of the other side. To this end, we employ a stop-gradient operation to preserve the individual characteristics of user/item embedding distributions. Furthermore, we balance user-/item-oriented contrasts during learning to maintain consistent high-rank performance (e.g., recall@1). Empirical results show that DuCo contributes to the top-k user and item prediction simultaneously, and outperforms state-of-the-art learning objectives across different backbones from ID-based to neighbor-based encoders.

#### **Quad-Tier Entity Fusion Contrastive Representation Learning for Knowledge Aware Recommendation System**

Rongqing Kenneth Ong, Wei Qiu, Andy W.H. Khong

Knowledge graph (KG) has recently emerged as a powerful source of auxiliary information in the realm of knowledge-aware recommendation (KGR) systems. However, due to the lack of supervision signals caused by the sparse nature of user-item interactions, existing supervised graph neural network (GNN) models suffer from performance degradation. Moreover, the over-smoothing issue further limits the number of GNN layers or hops required to propagate messages - these models ignore the non-local information concealed deep within the knowledge graph. We propose the Quad-Tier Entity Fusion Contrastive Representation Learning (QTEF-CRL) knowledge-aware framework to achieve learning of deep user preferences from four perspectives: the collaborative, semantic, preference, and structural view. Unlike existing methods, the proposed tri-local and single-global quad-tier architecture exploits the knowledge graph holistically to achieve effective self-supervised representation learning. The newly-introduced preference view constructed from the collaborative knowledge graph (CKG) comprises a preference graph and preference-guided GNN that are specifically designed to capture non-local information explicitly. Experiments conducted on three datasets highlight the efficacy of our proposed model.

#### **How Discriminative Are Your Qrels? How To Study the Statistical Significance of Document Adjudication Methods**

David Otero, Javier Parapar, Nicola Ferro

Creating test collections for offline retrieval evaluation requires human effort to judge documents' relevance. This expensive activity motivated much work in developing methods for constructing benchmarks with fewer assessment costs. In this respect, adjudication methods actively decide both which documents and the order in which experts review them, in order to better exploit the assessment budget or to lower it. Researchers evaluate the quality of those methods by measuring the correlation between the known gold ranking of systems under the full collection and the observed ranking of systems under the lower-cost one. This traditional analysis ignores whether and how the low-cost judgements impact on the statistically significant differences among systems with respect to the full collection. We fill this void by proposing a novel methodology to evaluate how the low-cost adjudication methods preserve the pairwise significant differences between systems as the full collection. In other terms, while traditional approaches look for stability in answering the question "is system A better than system B?", our proposed approach looks for stability in answering the question "is system A significantly better than system B?", which is the ultimate questions researchers need to answer to guarantee the generalisability of their results. Among other results, we found that the best methods in terms of ranking of systems correlation do not always match those preserving statistical significance.

#### **Rule-based Knowledge Graph Completion with Canonical Models**

Simon Ott, Patrick Betz, Daria Stepanova, Mohamed H. Gad-Elrab, Christian Meilicke, Heiner Stuckenschmidt

Rule-based approaches have proven to be an efficient and explainable method for knowledge base completion. Their predictive quality is on par with classic knowledge graph embedding models such as TransE or ComplEx, however, they cannot achieve the results of neural models proposed recently. The performance of a rule-based approach depends crucially on the solution of the rule aggregation problem, which is concerned with the computation of a score for a prediction that is generated by several rules. Within this paper, we propose a supervised approach to learn a reweighted confidence value for each rule to get an optimal explanation for the training set given a specific aggregation function. In particular, we apply our approach to two aggregation functions: We learn weights for a noisy-or multiplication and apply logistic regression, which computes the score of a prediction as a sum of these weights. Due to the simplicity of both models the final score is fully explainable. Our experimental results show that we can significantly improve the predictive quality of a rule-based approach. We compare our method with current state-of-the-art latent models that lack explainability, and achieve promising results.

### [Unsupervised Aspect Term Extraction by Integrating Sentence-level Curriculum Learning with Token-level Self-paced Learning](#)

Jihong Ouyang, Zhiyao Yang, Chang Xuan, Bing Wang, Yiyuan Wang, Ximing Li

Aspect Term Extraction (ATE), a key sub-task of aspect-based sentiment analysis, aims to extract aspect terms from review sentences on which users express opinions. Existing studies mainly treat ATE as a sequence labeling problem, and the aspect terms of training data are annotated at the token level, such as "BIO" tagging. However, such fine-grained annotations are often too costly to collect in many real applications, giving rise to the urgent demand for the challenging Unsupervised ATE (UATE). This paper suggests a novel UATE method by integrating sentence-level curriculum learning with token-level self-paced learning, namely UATE-SCTS. We design a set of hand-crafted rules to generate pseudo-labels but with noise. To combat this issue, our key idea is to train the ATE model from easier samples to harder samples to achieve a more robust model with more precise predictions at the early training epochs. This enables better refining of the noisy pseudo-labels. At the sentence level, we propose a frequency-induced pseudo-label cardinality to measure the learning difficulty of the review sentence and train the model in a curriculum-learning manner. At the token level, we formulate a self-paced learning objective that can adaptively select easier samples for training. We compare UATE-SCTS with baseline methods on benchmark collections of reviews from different domains. Empirical results demonstrate that UATE-SCTS can outperform existing UATE baselines.

### [A Retrieve-and-Read Framework for Knowledge Graph Link Prediction](#)

Vardaan Pahuja, Boshi Wang, Hugo Latapie, Jayanth Srinivasa, Yu Su

Knowledge graph (KG) link prediction aims to infer new facts based on existing facts in the KG. Recent studies have shown that using the graph neighborhood of a node via graph neural networks (GNNs) provides more useful information compared to just using the query information. Conventional GNNs for KG link prediction follow the standard message-passing paradigm on the entire KG, which leads to superfluous computation, over-smoothing of node representations, and also limits their expressive power. On a large scale, it becomes computationally expensive to aggregate useful information from the entire KG for inference. To address the limitations of existing KG link prediction frameworks, we propose a novel retrieve-and-read framework, which first retrieves a relevant subgraph context for the query and then jointly reasons over the context and the query with a high-capacity reader. As part of our exemplar instantiation for the new framework, we propose a novel Transformer-based GNN as the reader, which incorporates graph-based attention structure and cross-attention between query and context for deep fusion. This simple yet effective design enables the model to focus on salient context information relevant to the query. Empirical results on two standard KG link prediction datasets demonstrate the competitive performance of the proposed method. Furthermore, our analysis yields valuable insights for designing improved retrievers within the framework.

### [IUI: Intent-Enhanced User Interest Modeling for Click-Through Rate Prediction](#)

Mao Pan, Tao Yu, Kun Zhou, Zheng Li, Dongyue Wang, Zhuoye Ding, Xiwei Zhao, Sulong Xu

Click-Through Rate (CTR) prediction is becoming increasingly vital in many industrial applications, such as recommendations and online advertising. How to precisely capture users' dynamic and evolving interests from previous interactions (e.g., clicks, purchases, etc.) is a challenging task in CTR prediction. Mainstream approaches focus on disentangling user interests in a heuristic way or modeling user interests into a static representation. However, these approaches overlook the importance of users' current intent and the complex interactions between their current intent and global interests. To address these concerns, in this paper, we propose a novel intent-enhanced user interest modeling for click-through rate prediction in large-scale e-commerce recommendations, abbreviated as IUI. Methodologically, different from existing works, we consider users' recent interactions to be inspired by their implicit intent and then leverage an intent-aware network to model their current local interests in a more precise and fine-grained manner. In addition, to obtain a more stable co-dependent global and local interest representation, we employ a co-attention network capable of activating the corresponding interest in global-level interactions and capturing the dynamic interactions between global- and local-level interaction behaviors. Finally, we incorporate self-supervised learning into the model training by maximizing the mutual information between the global and local representations obtained via the above two networks to enhance the CTR prediction performance. Compared with existing methods, IUI benefits from the different granularity of user interest to generate a more accurate and comprehensive preference representation. Experimental results demonstrate that the proposed model outperforms previous state-of-the-art methods in various metrics on three real-world datasets. In addition, an online A/B test deployed on the JD recommendation platforms shows a promising improvement across multiple evaluation metrics.



### **Post-hoc Selection of Pareto-Optimal Solutions in Search and Recommendation**

Vincenzo Paparella, Vito Walter Anelli, Franco Maria Nardini, Raffaele Perego, Tommaso Di Noia

Information Retrieval (IR) and Recommender Systems (RSs) tasks are moving from computing a ranking of final results based on a single metric to multi-objective problems. Solving these problems leads to a set of Pareto-optimal solutions, known as Pareto frontier, in which no objective can be further improved without hurting the others. In principle, all the points on the Pareto frontier are potential candidates to represent the best model selected with respect to the combination of two, or more, metrics. To our knowledge, there are no well-recognized strategies to decide which point should be selected on the frontier in IR and RSs. In this paper, we propose a novel, post-hoc, theoretically-justified technique, named "Population Distance from Utopia" (PDU), to identify and select the one-best Pareto-optimal solution. PDU considers fine-grained utopia points, and measures how far each point is from its utopia point, allowing to select solutions tailored to user preferences, a novel feature we call "calibration". We compare PDU against state-of-the-art strategies through extensive experiments on tasks from both IR and RS, showing that PDU combined with calibration notably impacts the solution selection.

### **Cracking the Code of Negative Transfer: A Cooperative Game Theoretic Approach for Cross-Domain Sequential Recommendation**

Chung Park, Taesan Kim, Taekyoon Choi, Junui Hong, Yelim Yu, Mincheol Cho, Kyunam Lee, Sungil Ryu, Hyungjun Yoon, Minsung Choi, Jaegul Choo

This paper investigates Cross-Domain Sequential Recommendation (CDSR), a promising method that uses information from multiple domains (more than three) to generate accurate and diverse recommendations, and takes into account the sequential nature of user interactions. The effectiveness of these systems often depends on the complex interplay among the multiple domains. In this dynamic landscape, the problem of negative transfer arises, where heterogeneous knowledge between dissimilar domains leads to performance degradation due to differences in user preferences across these domains. As a remedy, we propose a new CDSR framework that addresses the problem of negative transfer by assessing the extent of negative transfer from one domain to another and adaptively assigning low weight values to the corresponding prediction losses. To this end, the amount of negative transfer is estimated by measuring the marginal contribution of each domain to model performance based on a cooperative game theory. In addition, a hierarchical contrastive learning approach that incorporates information from the sequence of coarse-level categories into that of fine-level categories (e.g., item level) when implementing contrastive learning was developed to mitigate negative transfer. Despite the potentially low relevance between domains at the fine-level, there may be higher relevance at the category level due to its generalised and broader preferences. We show that our model is superior to prior works in terms of model performance on two real-world datasets across ten different domains.

### **Toward a Better Understanding of Loss Functions for Collaborative Filtering**

Seongmin Park, Mincheol Yoon, Jae-woong Lee, Hogun Park, Jongwuk Lee

Collaborative filtering (CF) is a pivotal technique in modern recommender systems. The learning process of CF models typically consists of three components: interaction encoder, loss function, and negative sampling. Although many existing studies have proposed various CF models to design sophisticated interaction encoders, recent work shows that simply reformulating the loss functions can achieve significant performance gains. This paper delves into analyzing the relationship among existing loss functions. Our mathematical analysis reveals that the previous loss functions can be interpreted as alignment and uniformity functions: (i) the alignment matches user and item representations, and (ii) the uniformity disperses user and item distributions. Inspired by this analysis, we propose a novel loss function that improves the design of alignment and uniformity considering the unique patterns of datasets called Margin-aware Alignment and Weighted Uniformity (MAWU). The key novelty of MAWU is two-fold: (i) margin-aware alignment (MA) mitigates user/item-specific popularity biases, and (ii) weighted uniformity (WU) adjusts the significance between user and item uniformities to reflect the inherent characteristics of datasets. Extensive experimental results show that MF and LightGCN equipped with MAWU are comparable or superior to state-of-the-art CF models with various loss functions on three public datasets.

### **Concept Evolution in Deep Learning Training: A Unified Interpretation Framework and Discoveries**

Haekyu Park, Seongmin Lee, Benjamin Hoover, Austin P. Wright, Omar Shaikh, Rahul Duggal, Nilaksh Das, Kevin Li, Judy Hoffman, Duen Horng Chau

We present ConceptEvo, a unified interpretation framework for deep neural networks (DNNs) that reveals the inception and evolution of learned concepts during training. Our work addresses a critical gap in DNN interpretation research, as existing

methods primarily focus on post-training interpretation. ConceptEvo introduces two novel technical contributions: (1) an algorithm that generates a unified semantic space, enabling side-by-side comparison of different models during training, and (2) an algorithm that discovers and quantifies important concept evolutions for class predictions. Through a large-scale human evaluation and quantitative experiments, we demonstrate that ConceptEvo successfully identifies concept evolutions across different models, which are not only comprehensible to humans but also crucial for class predictions. ConceptEvo is applicable to both modern DNN architectures, such as ConvNeXt, and classic DNNs, such as VGGs and InceptionV3.

### [Evaluating and Optimizing the Effectiveness of Neural Machine Translation in Supporting Code Retrieval Models: A Study on the CAT Benchmark](#)

Hung Phan, Ali Jannesari

Neural Machine Translation (NMT) is widely applied in software engineering tasks. The effectiveness of NMT for code retrieval relies on the ability to learn from the sequence of tokens in the source language to the sequence of tokens in the target language. While NMT performs well in pseudocode-to-code translation[17], it might have challenges in learning to translate from natural language query to source code in newly curated real-world code documentation/ implementation datasets. In this work, we analyze the performance of NMT in natural language-to-code translation in the newly curated CAT benchmark[31] that includes the optimized versions of three Java datasets TLCodeSum, CodeSearchNet, Funcom, and a Python dataset PCSD. Our evaluation shows that NMT has low accuracy, measured by CrystalBLEU[10] and Meteor[9] metrics in this task. To alleviate the duty of NMT in learning complex representation of source code, we propose ASTTrans Representation, a tailored representation of an Abstract Syntax Tree (AST) using a subset of non-terminal nodes. We show that the classical approach NMT performs significantly better in learning ASTTrans Representation over code tokens with up to 36% improvement on Meteor score. Moreover, we leverage ASTTrans Representation to conduct combined code search processes from the state-of-the-art code search processes using GraphCodeBERT[13], and UniXcoder[12]. Our NMT models of learning ASTTrans Representation can boost the Mean Reciprocal Rank of these state-of-the-art code search processes by up to 3.08% and improve 23.08% of queries' results over the CAT benchmark.

### [RotDiff: A Hyperbolic Rotation Representation Model for Information Diffusion Prediction](#)

Hongliang Qiao, Shanshan Feng, Xutao Li, Huiwei Lin, Han Hu, Wei Wei, Yunming Ye

The massive amounts of online user behavior data on social networks allow for the investigation of information diffusion prediction, which is essential to comprehend how information propagates among users. The main difficulty in diffusion prediction problem is to effectively model the complex social factors in social networks and diffusion cascades. However, existing methods are mainly based on Euclidean space, which cannot well preserve the underlying hierarchical structures that could better reflect the strength of user influence. Meanwhile, existing methods cannot accurately model the obvious asymmetric features of the diffusion process. To alleviate these limitations, we utilize rotation transformation in the hyperbolic to model complex diffusion patterns. The modulus of representations in the hyperbolic space could effectively describe the strength of the user's influence. Rotation transformations could represent a variety of complex asymmetric features. Further, rotation transformation could model various social factors without changing the strength of influence. In this paper, we propose a novel hyperbolic rotation representation model RotDiff for the diffusion prediction problem. Specifically, we first map each social user to a Lorentzian vector and use two groups of transformations to encode global social factors in the social graph and the diffusion graph. Then, we combine attention mechanism in the hyperbolic space with extra rotation transformations to capture local diffusion dependencies within a given cascade. Experimental results on five real-world datasets demonstrate that the proposed model RotDiff outperforms various state-of-the-art diffusion prediction models.

### [Bi-channel Multiple Sparse Graph Attention Networks for Session-based Recommendation](#)

Shutong Qiao, Wei Zhou, Junhao Wen, Hongyu Zhang, Min Gao

Session-based Recommendation (SBR) has recently received significant attention due to its ability to provide personalized recommendations based on the interaction sequences of anonymous session users. The challenges facing SBR consist mainly of how to utilize information other than the current session and how to reduce the negative impact of irrelevant information in the session data on the prediction. To address these challenges, we propose a novel graph attention network-based model called Multiple Sparse Graph Attention Networks (MSGAT). MSGAT leverages two parallel channels to model intra-session and inter-session information. In the intra-session channel, we utilize a gated graph neural network to perform initial encoding, followed by a self-attention mechanism to generate the target representation. The global representation is then noise-reduced based on the target representation. Additionally, the target representation is used as a medium to connect the two channels. In the inter-session channel, the noise-reduced relation representation is generated using the

global attention mechanism of target perception. Moreover, MSGAT fully considers session similarity from the intent perspective by integrating valid information from both channels. Finally, the intent neighbor collaboration module effectively combines relevant information to enhance the current session representation. Extensive experiments on five datasets demonstrate that simultaneous modeling of intra-session and inter-session data can effectively enhance the performance of the SBR model.

### **MERIT: A Merchant Incentive Ranking Model for Hotel Search & Ranking**

Shigang Quan, Hailong Tan, Shui Liu, Zhenzhe Zheng, Ruihao Zhu, Liangyue Li, Quan Lu, Fan Wu

Online Travel Platforms (OTPs) have been working on improving their hotel Search & Ranking (S&R) systems that facilitate efficient matching between consumers and hotels. Existing OTPs focus on improving platform revenue. In this work, we take a first step in incorporating hotel merchants' objectives into the design of hotel S&R systems to achieve an incentive loop: the OTP tilts impressions and better-ranked positions to merchants with high service quality, and in return, the merchants provide better service to consumers. Three critical design challenges need to be resolved to achieve this incentive loop: Matthew Effect in the consumer feedback-loop, unclear relation between hotel service quality and performance, and conflicts between platform revenue and consumer experience.

To address these challenges, we propose MERIT, a MERchant InceTive ranking model, which can simultaneously take the interests of merchants and consumers into account. We introduce information about the hotel service quality at the input-output level. At the input level, we incorporate factors of hotel service quality as features (as the underlying reasons for service quality), while at the output level, we introduce the metric Hotel Rating Score (HRS) as a label (as the evaluated outcome of service quality). Also, we design a monotonic structure for Merchant Tower to provide a clear relation between hotel quality and performance. Finally, we propose a Multi-objective Stratified Pairwise Loss, which can mitigate the conflicts between OTP's revenue and consumer experience. To demonstrate the effectiveness of MERIT, we compare our method with several state-of-the-art benchmarks. The offline experiment results indicate that MERIT outperforms these methods in optimizing the demands of consumers and merchants. Furthermore, we conduct an online A/B test and obtain an improvement of 3.02% for the HRS score. Based on these results, we have deployed MERIT online on Fliggy, one of the most popular OTPs in China, to serve tens of millions of consumers and hundreds of thousands of hotel merchants. To address these challenges, we propose MERIT, a MERchant InceTive ranking model, which can simultaneously take the interests of merchants and consumers into account. We introduce information about the hotel service quality at the input-output level. At the input level, we incorporate factors of hotel service quality as features (as the underlying reasons for service quality), while at the output level, we introduce the metric Hotel Rating Score (HRS) as a label (as the evaluated outcome of service quality). Also, we design a monotonic structure for Merchant Tower to provide a clear relation between hotel quality and performance. Finally, we propose a Multi-objective Stratified Pairwise Loss, which can mitigate the conflicts between OTP's revenue and consumer experience. To demonstrate the effectiveness of MERIT, we compare our method with several state-of-the-art benchmarks. The offline experiment results indicate that MERIT outperforms these methods in optimizing the demands of consumers and merchants. Furthermore, we conduct an online A/B test and obtain an improvement of 3.02% for the HRS score. Based on these results, we have deployed MERIT online on Fliggy, one of the most popular OTPs in China, to serve tens of millions of consumers and hundreds of thousands of hotel merchants.

### **Enhancing Repeat-Aware Recommendation from a Temporal-Sequential Perspective**

Shigang Quan, Shui Liu, Zhenzhe Zheng, Fan Wu

Repeat consumption, such as re-purchasing items and re-listening songs, is a common scenario in daily life. To model repeat consumption, the repeat-aware recommendation has been proposed to predict which item will be re-interacted based on the user-item interactions. In this paper, we investigate various inherent characteristics to enhance the performance of repeat-aware recommendation. Specifically, we explore these characteristics from two aspects: one is from the temporal aspect where we consider the time interval relationship in user behavior sequence; the other is from the sequential aspect where we consider the sequential-level relationship. Our intuition is that both the temporal pattern and sequential pattern reflect users' intentions of repeat consumption.

By utilizing these two patterns, a novel model called Temporal and Sequential repeat-aware Recommendation (TSRec for short) is proposed to enhance repeat-aware recommendation. TSRec has three main components: 1) User-specific Temporal Representation Module (UTRM), which encodes and extracts user historical repeat temporal information. 2) Item-specific Temporal Representation Module (ITRM), which incorporates item time interval information as side information to alleviate the data sparsity problem of user repeat behavior sequence. 3) Sequential Repeat-Aware Module (SRAM), which represents

the similarity between user's current and the last repeat sequences. Extensive experimental results on three public benchmarks demonstrate the superiority of TSRec over state-of-the-art methods. The code is released online.

### **Federated Competing Risk Analysis**

Md Mahmudur Rahman, Sanjay Purushotham

Conducting survival analysis on distributed healthcare data is an important research problem, as privacy laws and emerging data-sharing regulations prohibit the sharing of sensitive patient data across multiple institutions. The distributed healthcare survival data often exhibit heterogeneity, non-uniform censoring and involve patients with multiple health conditions (competing risks), which can result in biased and unreliable risk predictions. To address these challenges, we propose employing federated learning (FL) for survival analysis with competing risks. In this work, we present two main contributions. Firstly, we propose a simple algorithm for estimating consistent federated pseudo values (FPV) for survival analysis with competing risks and censoring. Secondly, we introduce a novel and flexible FPV-based deep learning framework named Fedora, which jointly trains our proposed transformer-based model, TransPseudo, specific to the participating institutions (clients) within the Fedora framework without accessing clients' data, thus, preserving data privacy. We conducted extensive experiments on both real-world distributed healthcare datasets characterized by non-IID and non-uniform censoring properties, as well as synthetic data with various censoring settings. Our results demonstrate that our Fedora framework with the TransPseudo model performs better than the federated learning frameworks employing state-of-the-art survival models for competing risk analysis.

### **ELTRA: An Embedding Method based on Learning-to-Rank to Preserve Asymmetric Information in Directed Graphs**

Masoud Rehyani Hamedani, Jin-Su Ryu, Sang-Wook Kim

Double-vector embedding methods capture the asymmetric information in directed graphs first, and then preserve them in the embedding space by providing two latent vectors, i.e., source and target, per node. Although these methods are known to be superior to the single-vector ones (i.e., providing a single latent vector per node), we point out their three drawbacks as inability to preserve asymmetry on NU-paths, inability to preserve global nodes similarity, and impairing in/out-degree distributions. To address these, we first propose CRW, a novel similarity measure for graphs that considers contributions of both in-links and out-links in similarity computation, without ignoring their directions. Then, we propose ELTRA, an effective double-vector embedding method to preserve asymmetric information in directed graphs. ELTRA computes asymmetry preserving proximity scores (AP-scores) by employing CRW in which the contribution of out-links and in-links in similarity computation is upgraded and downgraded, respectively. Then, for every node  $u$ , ELTRA selects its top- $t$  closest nodes based on AP-scores and conforms their ranks of their corresponding target vectors w.r.t  $u$ 's source vector in the embedding space to their original ranks. Our extensive experimental results with seven real-world datasets and sixteen embedding methods show that (1) CRW significantly outperforms Katz and RWR in computing nodes similarity in graphs, (2) ELTRA outperforms the existing state-of-the-art methods in graph reconstruction, link prediction, and node classification tasks.

### **Dual-Process Graph Neural Network for Diversified Recommendation**

Yuanyi Ren, Hang Ni, Yingxue Zhang, Xi Wang, Guojie Song, Dong Li, Jianye Hao

The recommender system is one of the most fundamental information services. A significant effort has been devoted to improving prediction accuracy, inevitably leading to the potential degradation of recommendation diversity. Moreover, individuals have different needs for diversity. To address these problems, diversity-enhanced approaches are proposed to modify the recommender models. However, these methods fail to break free from the relevance-oriented paradigm and are mostly haunted by sharply-declined accuracy and high computational costs. To tackle these challenges, we propose the Dual-Process Graph Neural Network (DPGNN), an efficient diversity-enhanced recommender system, resonating with the dual-process model of human cognition and the arousal theory of human interest. The first stage reduces the risk of suboptimal output during the training procedure, which helps to find a solution outside the relevance-oriented paradigm. Moreover, the second stage utilizes user-specific rating adjustments, boosting the recommendation diversity and accommodating users' distinctive needs with minimum computational costs. Extensive experiments on real-world datasets verify the effectiveness of our method in improving diversity, while maintaining accuracy with low computational costs.

### **Incremental Graph Classification by Class Prototype Construction and Augmentation**

Yixin Ren, Li Ke, Dong Li, Hui Xue, Zhao Li, Shuigeng Zhou

Graph neural networks (GNNs) are prone to catastrophic forgetting of past experience in continuous learning scenarios. In this work, we propose a novel method for class-incremental graph learning (CGL) by class prototype construction and

augmentation, which can effectively overcome catastrophic forgetting and requires no storage of exemplars (i.e., data-free). Concretely, on the one hand, we construct class prototypes in the embedding space that contain rich topological information of nodes or graphs to represent past data, which are then used for future learning. On the other hand, to boost the adaptability of the model to new classes, we employ class prototype augmentation (PA) to create virtual classes by combining current prototypes. Theoretically, we show that PA can promote the model's adaptation to new data and reduce the inconsistency of old prototypes in the embedding space, therefore further mitigate catastrophic forgetting. Extensive experiments on both node and graph classification datasets show that our method significantly outperforms the existing methods in reducing catastrophic forgetting, and beats the existing methods in most cases in terms of classification accuracy.

### [Dual-view Contrastive Learning for Auction Recommendation](#)

Dan Ni Ren, Leong Hou U, Wei Liu

Recommendation systems in auction platforms like eBay function differently in comparison to those found in traditional trading platforms. The bidding process involves multiple users competing for a product, with the highest bidder winning the item. As a result, each transaction is independent and characterized by varying transaction prices. The individual nature of auction items means that users cannot purchase identical items, adding to the uniqueness of the purchasing history. Bidders in auction systems rely on their judgment to determine the value of a product, as bidding prices reflect preferences rather than cost-free actions like clicking or collecting. Conventional methodologies that heavily rely on user-item purchase history are ill-suited to handle these unique and extreme product features. Unfortunately, prior recommendation approaches have failed to give due attention to the contextual intricacies of auction items, thereby missing out on the full potential of the invaluable bidding record at hand.

This paper introduces a novel contrastive learning approach for auction recommendation, addressing the challenges of data sparsity and uniqueness in auction recommendation. Our method focuses on capturing multiple behavior relations and item context through contrastive pairs construction, contrastive embedding, and contrastive optimization techniques from both user and item perspectives. By overcoming the limitations of previous approaches, our method delivers promising results on two auction datasets, highlighting the practicality and effectiveness of our model.

### [Good Intentions: Adaptive Parameter Management via Intent Signaling](#)

Alexander Renz-Wieland, Andreas Kieslinger, Robert Gericke, Rainer Gemulla, Zoi Kaoudi, Volker Markl

Model parameter management is essential for distributed training of large machine learning (ML) tasks. Some ML tasks are hard to distribute because common approaches to parameter management can be highly inefficient. Advanced parameter management approaches---such as selective replication or dynamic parameter allocation---can improve efficiency, but they typically need to be integrated manually into each task's implementation and they require expensive upfront experimentation to tune correctly. In this work, we explore whether these two problems can be avoided. We first propose a novel intent signaling mechanism that integrates naturally into existing ML stacks and provides the parameter manager with crucial information about parameter accesses. We then describe AdaPM, a fully adaptive, zero-tuning parameter manager based on this mechanism. In contrast to prior parameter managers, our approach decouples how access information is provided (simple) from how and when it is exploited (hard). In our experimental evaluation, AdaPM matched or outperformed state-of-the-art parameter managers out of the box, suggesting that automatic parameter management is possible.

### [Seq-HyGAN: Sequence Classification via Hypergraph Attention Network](#)

Khaled Mohammed Saifuddin, Corey May, Farhan Tanvir, Muhammad Ifte Khairul Islam, Esra Akbas

Extracting meaningful features from sequences and devising effective similarity measures are vital for sequence data mining tasks, particularly sequence classification. While neural network models are commonly used to automatically learn sequence features, they are limited to capturing adjacent structural connection information and ignoring global, higher-order information between the sequences. To address these challenges, we propose a novel Hypergraph Attention Network model, namely Seq-HyGAN for sequence classification problems. To capture the complex structural similarity between sequence data, we create a novel hypergraph model by defining higher-order relations between subsequences extracted from sequences. Subsequently, we introduce a Sequence Hypergraph Attention Network that learns sequence features by considering the significance of subsequences and sequences to one another. Through extensive experiments, we demonstrate the effectiveness of our proposed Seq-HyGAN model in accurately classifying sequence data, outperforming several state-of-the-art methods by a significant margin.

### [PaperLM: A Pre-trained Model for Hierarchical Examination Paper Representation Learning](#)



Minghui Shan, Yixiao Ma, Shulan Ruan, Zhi Cao, Shiwei Tong, Qi Liu, Yu Su, Shijin Wang

Representation learning of examination papers is significantly crucial for online education systems, as it benefits various applications such as estimating paper difficulty and examination paper retrieval. Previous works mainly explore the representation learning of individual questions in an examination paper, with limited attention given to the examination paper as a whole. In fact, the structure of examination papers is strongly correlated with paper properties such as paper difficulty, which existing paper representation methods fail to capture adequately. To this end, we propose a pre-trained model namely PaperLM to learn the representation of examination papers. Our model integrates both the text content and hierarchical structure of examination papers within a single framework by converting the path of the Examination Organization Tree (EOT) into embedding. Furthermore, we specially design three pre-training objectives for PaperLM, namely EOT Node Relationship Prediction (ENRP), Question Type Prediction (QTP) and Paper Contrastive Learning (PCL), aiming to capture features from text and structure effectively. We pre-train our model on a real-world examination paper dataset, and then evaluate the model with three down-stream tasks: paper difficulty estimation, examination paper retrieval, and paper clustering. The experimental results demonstrate the effectiveness of our method.

### [\*\*Transferable Structure-based Adversarial Attack of Heterogeneous Graph Neural Network\*\*](#)

Yu Shang, Yudong Zhang, Jiansheng Chen, Depeng Jin, Yong Li

Heterogeneous graph neural networks (HGNNs) have achieved remarkable development recently and exhibited superior performance in various tasks. However, recently HGNNs have been shown to have robustness weakness towards adversarial perturbations, which brings critical pitfalls for real applications, e.g. node classification and recommender systems. In particular, the transfer-based black-box attack is the most practical method to attack unknown models and poses a great threat to the reliability of HGNNs. In this work, we take the first step to explore the transferability of adversarial examples of HGNNs. Due to the overfitting of the source model, the adversarial perturbations generated by traditional methods usually exhibit unpromising transferability. To address this problem and boost adversarial transferability, we expect to seek common vulnerable directions of different models to attack. Inspired by the observation of the notable commonality of edge attention distribution between different HGNNs, we propose to guide the perturbation generation toward disrupting edge attention distribution. This edge attention-guided attack prioritizes the perturbation on edges that are more likely to be given common attention by different models, which benefits the transferability of adversarial perturbations. Finally, we develop two edge attention-guided attack methods towards heterogeneous relations tailored for HGNNs, called EA-FGSM and EA-PGD. Extensive experiments on six representative models and two datasets verify the effectiveness of our methods and form an unprecedented transfer robustness benchmark for HGNNs.

### [\*\*Automatic and Precise Data Validation for Machine Learning\*\*](#)

Shreya Shankar, Labib Fawaz, Karl Gyllstrom, Aditya Parameswaran

Machine learning (ML) models in production pipelines are frequently retrained on the latest partitions of large, continually-growing datasets. Due to engineering bugs, partitions in such datasets almost always have some corrupted features; thus, it's critical to find data issues and block retraining before downstream ML accuracy decreases. However, current ML data validation methods are difficult to operationalize: they yield too many false positive alerts, require manual tuning, or are infeasible at scale. In this paper, we present an automatic, precise, and scalable data validation system for ML pipelines, employing a simple idea that we call a Partition Summarization (PS) approach to data validation: each timestamp-based partition of data is summarized with data quality metrics, and summaries are compared to detect corrupted partitions. We demonstrate how to adapt PS for any data validation method in a robust manner and evaluate several adaptations-which by themselves provide limited precision. Finally, we present gate, our data validation method that leverages these adaptations, giving a  $2.1\times$  average improvement in precision over the baseline from prior work on a case study within our large tech company.

### [\*\*TOAK: A Topology-oriented Attack Strategy for Degrading User Identity Linkage in Cross-network Learning\*\*](#)

Jiangli Shao, Yongqing Wang, Fangda Guo, Boshen Shi, Huawei Shen, Xueqi Cheng

Privacy concerns on social networks have received extensive attention in recent years. The task of user identity linkage (UIL), which aims to identify corresponding users across different social networks, poses a threat to privacy if applied unethically. Sensitive user information would be inferred with cross-network identity linkages. A feasible solution to this issue is to design an adversarial strategy that degrades the matching performance of UIL models. Nevertheless, most of the current adversarial attacks on graphs are tailored towards models working within a single network, failing to account for the

challenges presented by cross-network learning tasks such as UIL. Also, in real-world scenarios, the adversarial strategy against UIL has more constraints as service providers can only add perturbations to their own networks. To tackle these challenges, this paper proposes a novel poisoning strategy to prevent nodes in a target network from being linked to other networks by UIL algorithms. Specifically, the UIL problem is formalized in the kernelized topology consistency perspective, and the objective is formulated as maximizing the structural variations in the target network before and after modifications. To achieve this, a novel graph kernel is defined based on earth mover's distance (EMD) in the edge-embedding space. In terms of efficiency, a fast attack strategy is proposed using greedy searching and a lower bound approximation of EMD. Results on three real-world datasets demonstrate that the proposed method outperforms six baselines and reaches a balance between effectiveness and imperceptibility while being efficient.

#### **CANA: Causal-enhanced Social Network Alignment**

Jiangli Shao, Yongqing Wang, Fangda Guo, Boshen Shi, Huawei Shen, Xueqi Cheng

Social network alignment is widely applied in web applications for identifying corresponding nodes across different networks, such as linking users across two social networks. Existing methods for social network alignment primarily rely on alignment consistency, assuming that nodes with similar attributes and neighbors are more likely to be aligned. However, distributional discrepancies in node attributes and neighbors across different networks would bring biases in alignment consistency, leading to inferior alignment performance. To address this issue, we conduct a causal analysis of alignment consistency. Based on this analysis, we propose a novel model called CANA that uses causal inference approaches to mitigate biases and enhance social network alignment. Firstly, we disentangle observed node attributes into endogenous features and exogenous features with multi-task learning. Only endogenous features are retained to overcome node attribute discrepancies. To eliminate biases caused by neighbors discrepancies, we propose causal-aware attention mechanisms and integrate them in graph neural network to reweight contributions of different neighbors in alignment consistency comparison. Additionally, backdoor adjustment is applied to reduce confounding effects and estimate unbiased alignment probability. Through experimental evaluation on four real-world datasets, the proposed method demonstrates superior performance in terms of alignment accuracy and top-k hits precision.

#### **Representation Learning in Continuous-Time Dynamic Signed Networks**

Kartik Sharma, Mohit Raghavendra, Yeon-Chang Lee, Anand Kumar M, Srijan Kumar

Signed networks allow us to model conflicting relationships and interactions, such as friend/enemy and support/oppose. These signed interactions happen in real-time. Modeling such dynamics of signed networks is crucial to understanding the evolution of polarization in the network and enabling effective prediction of the signed structure (i.e., link signs) in the future. However, existing works have modeled either (static) signed networks or dynamic (unsigned) networks but not dynamic signed networks. Since both sign and dynamics inform the graph structure in different ways, it is non-trivial to model how to combine the two features. In this work, we propose a new Graph Neural Network (GNN)-based approach to model dynamic signed networks, named SEMBA: Signed link's Evolution using Memory modules and Balanced Aggregation. Here, the idea is to incorporate the signs of temporal interactions using separate modules guided by balance theory and to evolve the embeddings from a higher-order neighborhood. Experiments on 4 real-world datasets and 3 different tasks demonstrate that SEMBA consistently and significantly outperforms the baselines by up to 80% on the tasks of predicting signs of future links while matching the state-of-the-art performance on predicting existence of these links in the future. We find that this improvement is due specifically to superior performance of SEMBA on the minority negative class. Code is made available at <https://github.com/claws-lab/semba>.

#### **HyperBandit: Contextual Bandit with Hypernetwork for Time-Varying User Preferences in Streaming Recommendation**

Chenglei Shen, Xiao Zhang, Wei Wei, Jun Xu

In real-world streaming recommender systems, user preferences often dynamically change over time (e.g., a user may have different preferences during weekdays and weekends). Existing bandit-based streaming recommendation models only consider time as a timestamp, without explicitly modeling the relationship between time variables and time-varying user preferences. This leads to recommendation models that cannot quickly adapt to dynamic scenarios. To address this issue, we propose a contextual bandit approach using hypernetwork, called HyperBandit, which takes time features as input and dynamically adjusts the recommendation model for time-varying user preferences. Specifically, HyperBandit maintains a neural network capable of generating the parameters for estimating time-varying rewards, taking into account the correlation between time features and user preferences. Using the estimated time-varying rewards, a bandit policy is employed to make online recommendations by learning the latent item contexts. To meet the real-time requirements in streaming

recommendation scenarios, we have verified the existence of a low-rank structure in the parameter matrix and utilize low-rank factorization for efficient training. Theoretically, we demonstrate a sublinear regret upper bound against the best policy. Extensive experiments on real-world datasets show that the proposed HyperBandit consistently outperforms the state-of-the-art baselines in terms of accumulated rewards.

#### **Improving Graph Domain Adaptation with Network Hierarchy**

Boshen Shi, Yongqing Wang, Fangda Guo, Jiangli Shao, Huawei Shen, Xueqi Cheng

Graph domain adaptation models have become instrumental in addressing cross-network learning problems due to their ability to transfer abundant label and structural knowledge from source graphs to target graphs. A crucial step in transfer involves measuring domain discrepancy, which refers to distribution shifts between graphs from source and target domains. While conventional models simply provide a node-level measurement, exploiting information from different levels of network hierarchy is intuitive. As each hierarchical level characterizes distinct and meaningful properties or functionalities of the original graph, integrating domain discrepancy based on such hierarchies should contribute to a more precise domain discrepancy measurement. Moreover, class conditional distribution shift is often overlooked in node classification tasks, which could potentially lead to sub-optimal performance. To address the above limitations, we propose a new graph domain adaptation model and apply it to cross-network node classification tasks. Specifically, a hierarchical pooling model to extract meaningful and adaptive hierarchical structures is designed, where both marginal and class conditional distribution shifts on each hierarchical level are jointly minimized. The effectiveness is demonstrated through theoretical analysis and experimental studies across various datasets.

#### **GiGaMAE: Generalizable Graph Masked Autoencoder via Collaborative Latent Space Reconstruction**

Yucheng Shi, Yushun Dong, Qiaoyu Tan, Jundong Li, Ninghao Liu

Self-supervised learning with masked autoencoders has recently gained popularity for its ability to produce effective image or textual representations, which can be applied to various downstream tasks without retraining. However, we observe that the current masked autoencoder models lack good generalization ability on graph data. To tackle this issue, we propose a novel graph masked autoencoder framework called GiGaMAE. Different from existing masked autoencoders that learn node presentations by explicitly reconstructing the original graph components (e.g., features or edges), in this paper, we propose to collaboratively reconstruct informative and integrated latent embeddings. By considering embeddings encompassing graph topology and attribute information as reconstruction targets, our model could capture more generalized and comprehensive knowledge. Furthermore, we introduce a mutual information based reconstruction loss that enables the effective reconstruction of multiple targets. This learning objective allows us to differentiate between the exclusive knowledge learned from a single target and common knowledge shared by multiple targets. We evaluate our method on three downstream tasks with seven datasets as benchmarks. Extensive experiments demonstrate the superiority of GiGaMAE against state-of-the-art baselines. We hope our results will shed light on the design of foundation models on graph-structured data. Our code is available at: <https://github.com/sycny/GiGaMAE>.

#### **Calibrate Graph Neural Networks under Out-of-Distribution Nodes via Deep Q-learning**

Weili Shi, Xueying Yang, Xujiang Zhao, Haifeng Chen, Zhiqiang Tao, Sheng Li

Graph neural networks (GNNs) have achieved great success in dealing with graph-structured data that are prevalent in the real world. The core of graph neural networks is the message passing mechanism that aims to generate the embeddings of nodes by aggregating the neighboring node information. However, recent work suggests that GNNs also suffer the trustworthiness issues. Our empirical study shows that the calibration error of the in-distribution (ID) nodes would be exacerbated if a graph is mixed with out-of-distribution (OOD) nodes, and we assume that the noisy information from OOD nodes is the root for the worsened calibration error. Both previous study and our empirical study suggest that adjusting the weights of edges could be a promising way to reduce the adverse impact from the OOD nodes. However, how to precisely select the desired edges and modify the corresponding weights is not trivial, since the distribution of OOD nodes is unknown to us. To tackle this problem, we propose a Graph Edge Re-weighting via Deep Q-learning (GERDQ) framework to calibrate the graph neural networks. Our framework aims to explore the potential influence of the change of the edge weights on target ID nodes by sampling and traversing the edges in the graph, and we formulate this process as a Markov Decision Process (MDP). Many existing GNNs could be seamlessly incorporated into our framework. Experimental results show that when wrapped with our method, the existing GNN models can yield lower calibration error under OOD nodes as well as comparable accuracy compared to the original ones and other strong baselines. The source code is available at:

### **EmFore: Online Learning of Email Folder Classification Rules**

Mukul Singh, José Cambronero, Sumit Gulwani, Vu Le, Gust Verbruggen

Modern email clients support predicate-based folder assignment rules that can automatically organize emails. Unfortunately, users still need to write these rules manually. Prior machine learning approaches have framed automatically assigning email to folders as a classification task and do not produce symbolic rules. Prior inductive logic programming (ILP) approaches, which generate symbolic rules, fail to learn efficiently in the online environment needed for email management. To close this gap, we present EmFORE, an online system that learns symbolic rules for email classification from observations. Our key insights to do this successfully are: (1) learning rules over a folder abstraction that supports quickly determining candidate predicates to add or replace terms in a rule, (2) ensuring that rules remain consistent with historical assignments, (3) ranking rule updates based on existing predicate and folder name similarity, and (4) building a rule suppression model to avoid surfacing low-confidence folder predictions while keeping the rule for future use. We evaluate on two popular public email corpora and compare to 13 baselines, including state-of-the-art folder assignment systems, incremental machine learning, ILP and transformer-based approaches. We find that EmFORE performs significantly better, updates four orders of magnitude faster, and is more robust than existing methods and baselines.

### **Investigating the Impact of Multimodality and External Knowledge in Aspect-level Complaint and Sentiment Analysis**

Apoorva Singh, Apoorv Verma, Raghav Jain, Sriparna Saha

Automated complaint analysis is vital for generating critical insights, which in turn enhance customer satisfaction, product quality, and overall business performance. Nevertheless, conventional methods frequently fail to capture the nuances of aspect-level complaints and inadequately utilize external knowledge, thus creating a gap in effective complaint detection and analysis. In response to this issue, we proactively explore the role of external knowledge and multimodality in this domain. This leads to the development of MGasD (Multimodal Generative framework for aspect-based complaint and sentiment Detection), a multimodal knowledge-infused unified framework. MGasD diverges from traditional methods by reframing the complaint detection problem as a multimodal text-to-text generation task. Significantly, our research includes the development of a novel aspect-level dataset. Annotated for both complaint and sentiment categories across diverse domains such as books, electronics, edibles, fashion, and miscellaneous, this dataset provides a comprehensive platform for the concurrent study of complaints and sentiment. This resource facilitates a more robust understanding of consumer feedback. Our proposed methodology establishes a benchmark performance in the novel aspect-based complaint and sentiment detection tasks based on extensive evaluation. We also demonstrate that our model consistently outperforms all other baselines and state-of-the-art models in both full and few-shot settings (The dataset and code are available at: <https://github.com/appy1608/CIKM2023>).

### **The Role of Unattributed Behavior Logs in Predictive User Segmentation**

Atanu R. Sinha, Harshita Chopra, Aurghya Maiti, Atishay Ganesh, Sarthak Kapoor, Saili Myana, Saurabh Mahapatra

Online browsing on firms' sites generates user behavior logs (or, logs). These logs are mainstays that drive several user modeling tasks. The logs that inform user modeling are the ones that are attributed to each user, termed Attributed Behaviors (AB). But, a lot more logs are anonymous, upwards of 85%. For example, many users do not sign in while browsing. These logs are not attributed to users, termed Unattributed Behaviors (UB), and are not recognized in user modeling. We examine whether and how UB can benefit user modeling. We focus on a common task, that of user segmentation, for which the prior art uses only AB. We demonstrate that information from UBs, although unattributed to any individual, when used along with ABs, enriches performance of machine learning model for user segmentation. We perform predictive segmentation, whereby predicted outcomes for each segment are evaluated against actual outcomes. Multiple evaluations on two datasets, one of which is public, relative to state of the art baseline, show strong performance of our model in predicting outcomes and in reducing user segmentation error.

### **Follow the Will of the Market: A Context-Informed Drift-Aware Method for Stock Prediction**

Chen-Hui Song, Xi Xiao, Bin Zhang, Shu-Tao Xia

The dynamic nature of stock market styles, referred to as concept drift, poses a formidable challenge when applying deep learning to stock prediction. Models trained on historical data often struggle to adapt to the latest market styles, as the patterns they have learned may no longer hold true over time. To alleviate this issue, the recently popularized concept of In-Context learning has provided us with valuable insights. In this approach, large language models (LLMs) are exposed to multiple examples of input-label pairs, also known as demonstrations, as part of the prompt before performing a task on an

unseen example. By thoroughly analyzing these demonstrations, LLMs can uncover potential patterns and effectively adapt to new tasks. Building upon this concept, we propose a Context-Informed drift-aware method for Stock Prediction (CISP), which continually adjusts to the latest market styles and offers more accurate predictions. Our proposed method consists of two key parts. Firstly, we introduce a straightforward and efficient technique for designing demonstrations that aggregate current market information, thereby indicating the prevailing stock market style. Secondly, we incorporate a prediction module with dynamic parameters, allowing it to appropriately adjust its model parameters based on the market patterns embedded in the aforementioned demonstrations. Through extensive experiments conducted on real-world stock market datasets, our approach consistently outperforms the most advanced existing methods for stock prediction.

#### **CDR: Conservative Doubly Robust Learning for Debiased Recommendation**

Zijie Song, Jiawei Chen, Sheng Zhou, Qihao Shi, Yan Feng, Chun Chen, Can Wang

In recommendation systems (RS), user behavior data is observational rather than experimental, resulting in widespread bias in the data. Consequently, tackling bias has emerged as a major challenge in the field of recommendation systems. Recently, Doubly Robust Learning (DR) has gained significant attention due to its remarkable performance and robust properties. However, our experimental findings indicate that existing DR methods are severely impacted by the presence of so-called Poisonous Imputation, where the imputation significantly deviates from the truth and becomes counterproductive.

To address this issue, this work proposes Conservative Doubly Robust strategy (CDR) which filters imputations by scrutinizing their mean and variance. Theoretical analyses show that CDR offers reduced variance and improved tail bounds. In addition, our experimental investigations illustrate that CDR significantly enhances performance and can indeed reduce the frequency of poisonous imputation.

#### **Towards Fair Financial Services for All: A Temporal GNN Approach for Individual Fairness on Transaction Networks**

Zixing Song, Yuji Zhang, Irwin King

Discrimination against minority groups within the banking sector has long resulted in unequal treatment in financial services. Recent works in the general machine learning domain can promote group fairness for predictions on static tabular data, but their direct application in finance often proves ineffective. Financial losses of banks may arise from inaccurate predictions due to the overlooked dynamic nature of data, and illegal discrimination against some individual clients could still occur since fairness is promoted on the subgroup level. Therefore, we model the data as a dynamic or temporal transaction network for better utility and investigate individual fairness on this dynamic graph for the loan approval task. We define two novel individual fairness properties on temporal graphs with a theoretical analysis of their respective regret. Using these notions, we design a temporally fair graph neural network (TF-GNN) approach under a new real-time evaluation scheme for dynamic transaction networks. Experiments on real-world datasets demonstrate the superiority of the proposed method for both utility improvement in accuracy and fairness promotion in NDCG@k.

#### **SAND: Semantic Annotation of Numeric Data in Web Tables**

Yuchen Su, Davood Rafiei, Behrad Khorram Nazari

A large portion of quantitative information about entities is expressed as Web tables, and these tables often lack proper schema and annotation, which introduces challenges for the purpose of querying and analysis. In this paper, we introduce SAND, a novel approach for annotating numeric columns of Web tables by linking them to properties in a knowledge graph. Our approach relies only on the semantic information readily available in knowledge graphs and not on contextual information that can be missing or labelled data which may be difficult to obtain. We show that our approach can reliably detect both semantic types (e.g., height) and unit labels (e.g., Centimeter) when the semantic type is present in the knowledge graph. Our evaluation on real-world web tables shows that our method outperforms by a large margin, in terms of accuracy, some of the state-of-the-art approaches on semantic labeling and unit detection.

#### **Treatment Effect Estimation across Domains**

Yi-Xuan Sun, Ya-Lin Zhang, Wei Wang, Longfei Li, Jun Zhou

Treatment effect estimation is essential in the causal inference literature, which has attracted increasing attention in recent years. Most previous methods assume that the training and test data are drawn from the same distribution, which may not hold in practice since the effect estimators may need to be deployed across domains. Meanwhile, in real-world applications, little or no targeted treatments may be conducted in the new domain. Therefore, we focus on a more realistic scenario in this paper, where treatments and outcomes can be observed in the source domain, but the target domain only contains some



unlabeled data, i.e., only features are available. In this scenario, the distribution shift exists not only in the source data due to the selection bias between the control and treated groups, but also between the source and target data. We propose a novel direct learning framework along with the distribution adaptation and reliable scoring modules. In the distribution adaptation module, we design three specialized density ratio estimators to aid the issue of complex distribution shifts. Even so, we may face the challenge of unreliable pseudo-effects in this framework. To address that, we also design the uncertainty-based reliable scoring module as a vital support, which makes the method more reliable. The experiments are conducted on synthetic data and benchmark datasets, which demonstrate the superiority of our method.

### [Topic-Aware Contrastive Learning and K-Nearest Neighbor Mechanism for Stance Detection](#)

Yepeng Sun, Jicang Lu, Ling Wang, Shunhang Li, Ningbo Huang

The goal of stance detection is to automatically recognize the author's expressed attitude in text towards a given target. However, social media users often express themselves briefly and implicitly, which leads to a significant number of comments lacking explicit reference information to the target, posing a challenge for stance detection. To address the missing relationship between text and target, existing studies primarily focus on incorporating external knowledge, which inevitably introduces noise information. In contrast to their work, we are dedicated to mining implicit relational information within data. Typically, users tend to emphasize their attitudes towards a relevant topic or aspect of the target while concealing others when expressing opinions. Motivated by this phenomenon, we suggest that the potential correlation between text and target can be learned from instances with similar topics. Therefore, we design a pretext task to mine the topic associations between samples and model this topic association as a dynamic weight introduced into contrastive learning. In this way, we can selectively cluster samples that have similar topics and consistent stances, while enlarging the gap between samples with different stances in the feature space. Additionally, we propose a nearest-neighbor prediction mechanism for stance classification to better utilize the features we constructed. Our experiments on two datasets demonstrate the advanced and generalization ability of our method, yielding the state-of-the-art results.

### [Fairness through Aleatoric Uncertainty](#)

Anique Tahir, Lu Cheng, Huan Liu

We propose a simple yet effective solution to tackle the often-competing goals of fairness and utility in classification tasks. While fairness ensures that the model's predictions are unbiased and do not discriminate against any particular group or individual, utility focuses on maximizing the model's predictive performance. This work introduces the idea of leveraging aleatoric uncertainty (e.g., data ambiguity) to improve the fairness-utility trade-off. Our central hypothesis is that aleatoric uncertainty is a key factor for algorithmic fairness and samples with low aleatoric uncertainty are modeled more accurately and fairly than those with high aleatoric uncertainty. We then propose a principled model to improve fairness when aleatoric uncertainty is high and improve utility elsewhere. Our approach first intervenes in the data distribution to better decouple aleatoric uncertainty and epistemic uncertainty. It then introduces a fairness-utility bi-objective loss defined based on the estimated aleatoric uncertainty. Our approach is theoretically guaranteed to improve the fairness-utility trade-off. Experimental results on both tabular and image datasets show that the proposed approach outperforms state-of-the-art methods w.r.t. the fairness-utility trade-off and w.r.t. both group and individual fairness metrics. This work presents a fresh perspective on the trade-off between utility and algorithmic fairness and opens a key avenue for the potential of using prediction uncertainty in fair machine learning.

### [Graph Inference via the Energy-efficient Dynamic Precision Matrix Estimation with One-bit Data](#)

Xiao Tan, Yangyang Shen, Meng Wang, Beilun Wang

Graph knowledge discovery from graph-structured data is a fascinating data mining topic in various domains, especially in the Internet of Things, where inferring the graph structure from such informative data can benefit many downstream tasks. Deep neural networks are typically used to perform such predictions, but they produce unreliable results without sufficient high-quality data. Therefore, researchers introduce lightweight statistical precision matrix learning to infer the graph structure in many IoT scenarios with limited communication and resolution of sensors. However, these methods still suffer from low-resolution data or the omission of hidden information in time-series data. To address the challenges, we propose a novel approach for Energy-efficient Dynamic Sparse Graph Structure Estimation with one-bit data, EDGE. Our method proposes a novel estimator to estimate the covariance matrix from one-bit data, and then utilize the covariance matrices to capture the dynamic structure. We theoretically demonstrate the effectiveness of the estimators by deriving two non-asymptotic estimation error bounds for the estimated covariance matrix and precision matrix, respectively. The theoretical results show that our method can achieve a consistent result of the precision matrix at the rate  $O(\log p/n)$ . On multiple synthetic and real-

world datasets, the experimental results demonstrate that our proposed estimator is able to obtain a relatively high detection rate using one-bit data, which exceeds the baseline by 35%, and identify potentially perturbed nodes in real-time dynamic network inference.

### [Joint Rebalancing and Charging for Shared Electric Micromobility Vehicles with Energy-informed Demand](#)

Heng Tan, Yukun Yuan, Shuxin Zhong, Yu Yang

Shared electric micromobility (e.g., shared electric bikes and electric scooters), as an emerging way of urban transportation, has been increasingly popular in recent years. However, managing thousands of micromobility vehicles in a city, such as rebalancing and charging vehicles to meet spatial-temporally varied demand, is challenging. Existing management frameworks generally consider demand as the number of requests without the energy consumption of these requests, which can lead to less effective management. To address this limitation, we design RECOMMEND, a rebalancing and charging framework for shared electric micromobility vehicles with energy-informed demand to improve the system revenue. Specifically, we first re-define the demand from the perspective of energy consumption and predict the future energy-informed demand based on the state-of-the-art spatial-temporal prediction method. Then we fuse the predicted energy-informed demand into different components of a rebalancing and charging framework based on reinforcement learning. We evaluate the RECOMMEND system with 2-month real-world electric micromobility system operation data. Experimental results show that our method can be easily integrated into a general RL framework and outperform state-of-the-art baselines by at least 26.89% in terms of net revenue.

### [EAGLE: Enhance Target-Oriented Dialogs by Global Planning and Topic Flow Integration](#)

Zee Hen Tang, Mi-Yen Yeh

In this study, we propose a novel model EAGLE for target-oriented dialogue generation. Without relying on any knowledge graphs, our method integrates the global planning strategy in both topic path generation and response generation given the initial and target topics. EAGLE comprises three components: a topic path sampling strategy, a topic flow generator, and a global planner. Our approach confers a number of advantages: EAGLE is robust to the target that has never appeared in the training data set and able to plan the topic flow globally. The topic path sampling strategy samples topic paths based on two predefined rules and use the sampled paths to train the topic path generator. The topic flow generator then applies a non-autoregressive method to generate intermediate topics that link the initial and target topics smoothly. In addition, the global planner is a response generator that generates a response based on the future topic sequence and conversation history, enabling it to plan how to transition to future topics smoothly. Our experimental results demonstrate that EAGLE produces more coherent responses and smoother transitions than state-of-the-art baselines, with an overall success rate improvement of approximately 25% and an average smoothness score improvement of 10% in both offline and human evaluations.

### [Spatio-Temporal Meta Contrastive Learning](#)

Jiabin Tang, Lianghao Xia, Jie Hu, Chao Huang

Spatio-temporal prediction is crucial in numerous real-world applications, including traffic forecasting and crime prediction, which aim to improve public transportation and safety management. Many state-of-the-art models demonstrate the strong capability of spatio-temporal graph neural networks (STGNN) to capture complex spatio-temporal correlations. However, despite their effectiveness, existing approaches do not adequately address several key challenges. Data quality issues, such as data scarcity and sparsity, lead to data noise and a lack of supervised signals, which significantly limit the performance of STGNN. Although recent STGNN models with contrastive learning aim to address these challenges, most of them use pre-defined augmentation strategies that heavily depend on manual design and cannot be customized for different Spatio-Temporal Graph (STG) scenarios. To tackle these challenges, we propose a new spatio-temporal contrastive learning (CL4ST) framework to encode robust and generalizable STG representations via the STG augmentation paradigm. Specifically, we design the meta view generator to automatically construct node and edge augmentation views for each disentangled spatial and temporal graph in a data-driven manner. The meta view generator employs meta networks with parameterized generative model to customize the augmentations for each input. This personalizes the augmentation strategies for every STG and endows the learning framework with spatio-temporal-aware information. Additionally, we integrate a unified spatio-temporal graph attention network with the proposed meta view generator and two-branch graph contrastive learning paradigms. Extensive experiments demonstrate that our CL4ST significantly improves performance over various state-of-the-art baselines in traffic and crime prediction. Our model implementation is available at the link: <https://github.com/HKUDS/CL4ST>.

**Single-Cell Multimodal Prediction via Transformers**

Wenzhuo Tang, Hongzhi Wen, Renming Liu, Jiayuan Ding, Wei Jin, Yuying Xie, Hui Liu, Jiliang Tang

The recent development of multimodal single-cell technology has made the possibility of acquiring multiple omics data from individual cells, thereby enabling a deeper understanding of cellular states and dynamics. Nevertheless, the proliferation of multimodal single-cell data also introduces tremendous challenges in modeling the complex interactions among different modalities. The recently advanced methods focus on constructing static interaction graphs and applying graph neural networks (GNNs) to learn from multimodal data. However, such static graphs can be suboptimal as they do not take advantage of the downstream task information; meanwhile GNNs also have some inherent limitations when deeply stacking GNN layers. To tackle these issues, in this work, we investigate how to leverage transformers for multimodal single-cell data in an end-to-end manner while exploiting downstream task information. In particular, we propose a scMoFormer framework which can readily incorporate external domain knowledge and model the interactions within each modality and cross modalities. Extensive experiments demonstrate that scMoFormer achieves superior performance on various benchmark datasets. Remarkably, scMoFormer won a Kaggle silver medal with the rank of 24/1221 (Top 2%) without ensemble in a NeurIPS 2022 competition<sup>1</sup>. Our implementation is publicly available at Github<sup>2</sup>.

**Explainable Spatio-Temporal Graph Neural Networks**

Jiabin Tang, Lianghao Xia, Chao Huang

Spatio-temporal graph neural networks (STGNNs) have gained popularity as a powerful tool for effectively modeling spatio-temporal dependencies in diverse real-world urban applications, including intelligent transportation and public safety. However, the black-box nature of STGNNs limits their interpretability, hindering their application in scenarios related to urban resource allocation and policy formulation. To bridge this gap, we propose an Explainable Spatio-Temporal Graph Neural Networks (STExplainer) framework that enhances STGNNs with inherent explainability, enabling them to provide accurate predictions and faithful explanations simultaneously. Our framework integrates a unified spatio-temporal graph attention network with a positional information fusion layer as the STG encoder and decoder, respectively. Furthermore, we propose a structure distillation approach based on the Graph Information Bottleneck (GIB) principle with an explainable objective, which is instantiated by the STG encoder and decoder. Through extensive experiments, we demonstrate that our STExplainer outperforms state-of-the-art baselines in terms of predictive accuracy and explainability metrics (i.e., sparsity and fidelity) on traffic and crime prediction tasks. Furthermore, our model exhibits superior representation ability in alleviating data missing and sparsity issues. The implementation code is available at: <https://github.com/HKUDS/STExplainer>.

**Periodicity May Be Emanative: Hierarchical Contrastive Learning for Sequential Recommendation**

Changxin Tian, Binbin Hu, Wayne Xin Zhao, Zhiqiang Zhang, Jun Zhou

Nowadays, contrastive self-supervised learning has been widely incorporated into sequential recommender systems. However, most existing contrastive sequential recommender systems simply emphasize the overall information of interaction sequences, thereby neglecting the special periodic patterns of user behavior. In this study, we propose that users exhibit emanative periodicity towards a group of correlated items, i.e., user behavior follow a certain periodic pattern while their interests may shift from one item to other related items over time. In light of this observation, we present a hierarchical contrastive learning framework to model EmAnative periodicity for SEquential Recommendation (referred to as EASE). Specifically, we design dual-channel contrastive strategy from the perspective of correlation and periodicity to capture emanative periodic patterns. Furthermore, we extend the traditional binary contrastive loss with hierarchical constraint to handle hierarchical contrastive samples, thus preserving the inherent hierarchical information of correlation and periodicity. Comprehensive experiments conducted on five datasets substantiate the effectiveness of our proposed EASE in improving sequential recommendation.

**Experience and Evidence are the eyes of an excellent summarizer! Towards Knowledge Infused Multi-modal Clinical Conversation Summarization**

Abhisek Tiwari, Anisha Saha, Sriparna Saha, Pushpak Bhattacharyya, Minakshi Dhar

With the advancement of telemedicine, both researchers and medical practitioners are working hand-in-hand to develop various techniques to automate various medical operations, such as diagnosis report generation. In this paper, we first present a multi-modal clinical conversation summary generation task that takes a clinician-patient interaction (both textual and visual information) and generates a succinct synopsis of the conversation. We propose a knowledge-infused, multi-modal, multi-tasking medical domain identification and clinical conversation summary generation (MM-CliConSummation)

framework. It leverages an adapter to infuse knowledge and visual features and unify the fused feature vector using a gated mechanism. Furthermore, we developed a multi-modal, multi-intent clinical conversation summarization corpus annotated with intent, symptom, and summary. The extensive set of experiments, both quantitatively and qualitatively, led to the following findings: (a) critical significance of visuals, (b) more precise and medical entity preserving summary with additional knowledge infusion, and (c) a correlation between medical department identification and clinical synopsis generation. Furthermore, the dataset and source code are available at <https://github.com/NLP-RL/MM-CliConSummation>

#### **Multi-Representation Variational Autoencoder via Iterative Latent Attention and Implicit Differentiation**

Nhu-Thuat Tran, Hady W. Lauw

Variational Autoencoder (VAE) offers a non-linear probabilistic modeling of user's preferences. While it has achieved remarkable performance at collaborative filtering, it typically samples a single vector for representing user's preferences, which may be insufficient to capture the user's diverse interests. Existing solutions extend VAE to model multiple interests of users by resorting a variant of self-attentive method, i.e., employing prototypes to group items into clusters, each capturing one topic of user's interests. Despite showing improvements, the current design could be more effective since prototypes are randomly initialized and shared across users, resulting in uninformative and non-personalized clusters.

To fill the gap, firstly, we introduce iterative latent attention for personalized item grouping into VAE framework to infer multiple interests of users. Secondly, we propose to incorporate implicit differentiation to improve training of our iterative refinement model. Thirdly, we study the self-attention to refine cluster prototypes for item grouping, which is largely ignored by existing works. Extensive experiments on three real-world datasets demonstrate stronger performance of our method over those of baselines.

#### **Citation Intent Classification and Its Supporting Evidence Extraction for Citation Graph Construction**

Hong-Jin Tsai, An-Zi Yen, Hen-Hsen Huang, Hsin-Hsi Chen

As the significant growth of scientific publications in recent years, an efficient way to extract scholarly knowledge and organize the relationship among literature is necessitated. Previous works constructed scientific knowledge graph with authors, papers, citations, and scientific entities. To assist researchers to grasp the research context comprehensively, this paper constructs a fine-grained citation graph in which citation intents and their supporting evidence are labeled between citing and cited papers instead. We propose a model with a Transformer encoder to encode the long-lengthy paper. To capture the coreference relations of words and sentences in a paper, a coreference graph is created by utilizing Gated Graph Convolution Network (GGCN). We further propose a graph modification mechanism to dynamically update the coreference links. Experimental results show that our model achieves promising results on identifying multiple citation intents in sentences.

#### **Disentangled Interest importance aware Knowledge Graph Neural Network for Fund Recommendation**

Ke Tu, Wei Qu, Zhengwei Wu, Zhiqiang Zhang, Zhongyi Liu, Yiming Zhao, Le Wu, Jun Zhou, Guannan Zhang

At present, people are gradually becoming aware of financial management and thus fund recommendation attracts more and more attention to help them find suitable funds quickly. As a user usually takes many factors (e.g., fund theme, fund manager) into account when investing a fund and the fund usually consists of a substantial collection of investments, effectively modeling multi-interest representations is more crucial for personalized fund recommendation than the traditional goods recommendation. However, existing multi-interest methods are largely sub-optimal for fund recommendation, since they ignore financial domain knowledge and diverse fund investment intentions. In this work, we propose a Disentangled Interest importance aware Knowledge Graph Neural Network (DIKGNN) for personalized fund recommendation on FinTech platforms. In particular, we restrict the multiple intent spaces by introducing the attribute nodes from the fund knowledge graph as the minimum intent modeling unit to utilize financial domain knowledge and provide interpretability. In the intent space, we define disentangled intent representations, equipped with intent importance distributions to describe the diverse fund investment intentions. Then we design a new neighbor aggregation mechanism with the learned intent importance distribution upon the interaction graph and knowledge graph to collect multi-intent information. Furthermore, we leverage micro independence and macro balance constraints on the representations and distributions respectively to encourage intent independence and diversity. The extensive experiments on public recommendation benchmarks demonstrate that DIKGNN can achieve substantial improvement over state-of-the-art methods. Our proposed model is also evaluated over one real-world industrial fund dataset from a FinTech platform and has been deployed online.

#### **PSLF: Defending Against Label Leakage in Split Learning**

Xinwei Wan, Jiankai Sun, Shengjie Wang, Lei Chen, Zhenzhe Zheng, Fan Wu, Guihai Chen

With increasing concern over data privacy, split learning has become a widely used distributed machine learning paradigm in practice, where two participants (namely the non-label party and the label party) own raw features and raw labels respectively, and jointly train a model. Although no raw data is communicated between the two parties during model training, several works have demonstrated that data privacy, especially label privacy, is still vulnerable in split learning, and have proposed several defense algorithms against label attacks. However, the theoretical guarantee on the privacy preservation of these algorithms is limited. In this work, we propose a novel Private Split Learning Framework (PSLF). In PSLF, the label party shares only the gradients computed by flipped labels with the non-label party, which improves privacy preservation on raw labels, and meanwhile, we further design an extra sub-model from true labels to improve prediction accuracy. We also design a Flipped Multi-Label Generation mechanism (FMLG) based on randomized response for the label party to generate flipped labels. FMLG is proven differentially private and the label party could make a trade-off between privacy and utility by setting the DP budget. In addition, we design an upsampling method to further protect the labels against some existing attacks. We have evaluated PSLF over real-world datasets to demonstrate its effectiveness in protecting label privacy and achieving promising prediction accuracy.

#### **GraphFADE: Field-aware Decorrelation Neural Network for Graphs with Tabular Features**

Junhong Wan, Yao Fu, Junlan Yu, Weihao Jiang, Shiliang Pu, Ruiheng Yang

Graph Neural Networks (GNNs) have achieved great success in recent years for their remarkable ability to extract effective representations from both node features and graph structures. Most of GNNs only focus on graphs with homogeneous features that correspond to one single feature field. For tabular features that are heterogeneous with multiple feature fields, GNNs often perform less favorably compared to machine learning methods such as boosted trees. In this work, we propose a new perspective to uncover the problem of GNNs on graphs with tabular features through both empirical study and theoretical analysis. The assumption of GNNs that connected nodes exhibit similar patterns can barely hold true for tabular features since multiple feature fields already exhibit different patterns. And propagation on such mismatched graph causes propagated features overcorrelated on graphs, which leads to the reduction of feature diversity and the increase of information redundancy. Therefore, we propose Field-aware Decorrelation Neural Network for graphs with tabular features (GraphFADE), a novel framework that directly optimizes the overcorrelation problem for graphs with tabular features. We first hierarchically partition the dataset into subsets with minimal correlation and then according to the decorrelation clustering results assemble the optimal matched graphs for each feature dimension to propagate on. The empirical study shows that our method achieves superior performance on multiple graphs with tabular features, demonstrating the effectiveness of our model.

#### **MPerformer: An SE(3) Transformer-based Molecular Perceptron**

Fanmeng Wang, Hongteng Xu, Xi Chen, Shuqi Lu, Yuqing Deng, Wenbing Huang

Molecular perception aims to construct 3D molecules from 3D atom clouds (i.e., atom types and corresponding 3D coordinates), determining bond connections, bond orders, and other molecular attributes within molecules. It is essential for realizing many applications in cheminformatics and bioinformatics, such as modeling quantum chemistry-derived molecular structures in protein-ligand complexes. Additionally, many molecular generation methods can only generate molecular 3D atom clouds, requiring molecular perception as a necessary post-processing. However, existing molecular perception methods mainly rely on predefined chemical rules and fail to leverage 3D geometric information, whose performance is sub-optimal fully. In this study, we propose MPerformer, an SE(3) Transformer-based molecular perceptron exhibiting SE(3)-invariance, to construct 3D molecules from 3D atom clouds efficiently. Besides, we propose a multi-task pretraining-and-finetuning paradigm to learn this model. In the pretraining phase, we jointly minimize an attribute prediction loss and an atom cloud reconstruction loss, mitigating the data imbalance issue of molecular attributes and enhancing the robustness and generalizability of the model. Experiments show that MPerformer significantly outperforms state-of-the-art molecular perception methods in precision and robustness, benefiting various molecular generation scenarios.

#### **Towards Deeper, Lighter and Interpretable Cross Network for CTR Prediction**

Fangye Wang, Hansu Gu, Dongsheng Li, Tun Lu, Peng Zhang, Ning Gu

Click Through Rate (CTR) prediction plays an essential role in recommender systems and online advertising. It is crucial to effectively model feature interactions to improve the prediction performance of CTR models. However, existing methods face three significant challenges. First, while most methods can automatically capture high-order feature interactions, their



performance tends to diminish as the order of feature interactions increases. Second, existing methods lack the ability to provide convincing interpretations of the prediction results, especially for high-order feature interactions, which limits the trustworthiness of their predictions. Third, many methods suffer from the presence of redundant parameters, particularly in the embedding layer. This paper proposes a novel method called Gated Deep Cross Network (GDCN) and a Field-level Dimension Optimization (FDO) approach to address these challenges. As the core structure of GDCN, Gated Cross Network (GCN) captures explicit high-order feature interactions and dynamically filters important interactions with an information gate in each order. Additionally, we use the FDO approach to learn condensed dimensions for each field based on their importance. Comprehensive experiments on five datasets demonstrate the effectiveness, superiority and interpretability of GDCN. Moreover, we verify the effectiveness of FDO in learning various dimensions and reducing model parameters. The code is available on <https://github.com/anonctr/GDCN>.

### **Iteratively Learning Representations for Unseen Entities with Inter-Rule Correlations**

Zihan Wang, Kai Zhao, Yongquan He, Zhumin Chen, Pengjie Ren, Maarten de Rijke, Zhaochun Ren

Recent work on knowledge graph completion (KGC) focuses on acquiring embeddings of entities and relations in knowledge graphs. These embedding methods necessitate that all test entities be present during the training phase, resulting in a time-consuming retraining process for out-of-knowledge-graph (OOKG) entities. To tackle this predicament, current inductive methods employ graph neural networks (GNNs) to represent unseen entities by aggregating information of the known neighbors, and enhance the performance with additional information, such as attention mechanisms or logic rules. Nonetheless, Two key challenges continue to persist: (i) identifying inter-rule correlations to further facilitate the inference process, and (ii) capturing interactions among rule mining, rule inference, and embedding to enhance both rule and embedding learning.

In this paper, we propose a virtual neighbor network with inter-rule correlations (VNC) to address the above challenges. VNC consists of three main components: (i) rule mining, (ii) rule inference, and (iii) embedding. To identify useful complex patterns in knowledge graphs, both logic rules and inter-rule correlations are extracted from knowledge graphs based on operations over relation embeddings. To reduce data sparsity, virtual networks for OOKG entities are predicted and assigned soft labels by optimizing a rule-constrained problem. We also devise an iterative framework to capture the underlying interactions between rule and embedding learning. Experimental results on both link prediction and triple classification tasks show that the proposed VNC framework achieves state-of-the-art performance on four widely-used knowledge graphs.

### **CLOCK: Online Temporal Hierarchical Framework for Multi-scale Multi-granularity Forecasting of User Impression**

XiaoYu Wang, YongHui Guo, Xiaoyang Ma, Dongbo Huang, Lan Xu, Haisheng Tan, Hao Zhou, Xiang-Yang Li

User impression forecasting underpins various commercial activities, from long-term strategic decisions to short-term automated operations. As a representative that involves both kinds, the highly profitable Guaranteed Delivery (GD) advertising focuses mainly on promoting brand effect by allowing advertisers to order target impressions weeks in advance and get allocated online at the scheduled time. Such a business mode naturally incurs three issues making existing solutions inferior: 1) Timescale-granularity dilemma of coherently supporting the sales of day-level impressions of the distant future and the corresponding fine-grained allocation in real-time. 2) High dimensionality due to the Cartesian product of user attribute combinations. 3) Stability-plasticity dilemma of instant adaptation to emerging patterns of temporal dependency without catastrophic forgetting of repeated ones facing the non-stationary traffic.

To overcome the obstacles, we propose an online temporal hierarchical framework that functions analogously to a CLOCK and hence its name. Long-timescale, coarse-grained temporal data (e.g., the daily impression of one quarter) and short-timescale but fine-grained ones are handled separately by dedicated models, just like the hour/minute/second hands. Each tier in the hierarchy is triggered for forecasting and updating by need at different frequencies, thus saving the maintenance overhead. Furthermore, we devise a reconciliation mechanism to coordinate tiers by aggregating the separately learned local variance and global trends tier by tier. CLOCK solves the dimensionality dilemma by subsuming the autoencoder design to achieve an end-to-end, nonlinear factorization of streaming data into a low-dimension latent space, where a neural predictor produces predictions for the decoder to project them back to the high dimension. Lastly, we regulate the CLOCK's continual refinement by combining the complementary Experience Replay (ER) and Knowledge Distillation (KD) techniques to consolidate and recall previously learned temporal patterns. We conduct extensive evaluations on three public datasets and the real-life user impression log from the Tencent advertising system, and the results demonstrate CLOCK's efficacy.

### **FAMC-Net: Frequency Domain Parity Correction Attention and Multi-Scale Dilated Convolution for Time Series Forecasting**

Min Wang, Hua Wang, Fan Zhang

In recent years, time series forecasting models based on the Transformer framework have shown great potential, but they suffer from the inherent drawback of high computational complexity and only focus on global modeling. Inspired by trend-seasonality decomposition, we propose a method that combines global modeling with local feature extraction within the seasonal cycle. It aims at capturing the global view while fully exploring the potential features within each seasonal cycle and better expressing the long-term and periodic characteristics of time series. We introduce a frequency domain parity correction block to compute global attention and utilize multi-scale dilated convolution to extract local correlations within each cycle. Additionally, we adopt a dual-branch structure to separately model the seasonality and trend based on their intrinsic features, improving prediction performance and enhancing model interpretability. This model is implemented on a completely single-layer decoder architecture, breaking through the traditional encoder-decoder architecture paradigm and reducing computational complexity to a certain extent. We conducted sufficient experimental validation on eight benchmark datasets, and the results demonstrate its superior performance compared to existing methods in both univariate and multivariate forecasting.

### **Diversity-aware Deep Ranking Network for Recommendation**

Zihong Wang, Yingxia Shao, Jiyuan He, Jinbao Liu, Shitao Xiao, Tao Feng, Ming Liu

Diversity is a vital factor in recommendation systems. Improving the diversity in recommendations helps broaden users' horizons, bring good user experience and promote the enterprises' sales. In the past years, many efforts have been devoted to optimizing the diversity in the matching stage and the re-ranking stage of the recommendation system, but few in the ranking stage. The ranking stage is the intermediate stage of the recommendation system. Improving the diversity of the ranking stage can preserve the diversity of the matching stage, and provide a more diversified list for the re-ranking stage. Besides, the ranking models are able to achieve a better balance between accuracy and diversity. In this paper, we aim to improve the diversity in the ranking stage. To address the diversity challenges posed by the pointwise ranking model and biased user interaction history, we propose a Diversity-aware Deep Ranking Network by carefully designing two diversity-aware components that are diversity-aware listwise information fusion and balanced weighting loss. We conduct both offline and online experiments, and the results demonstrate that our proposed model effectively improves the recommendation diversity in the ranking stage while maintaining the accuracy. Moreover, the new model achieves 1.27%, 2.30% and 1.98% improvements in VBR, GMV and Coverage in Meituan, one of the world's largest E-commerce platforms.

### **UrbanFloodKG: An Urban Flood Knowledge Graph System for Risk Assessment**

Yu Wang, Feng Ye, Binqun Li, Gaoyang Jin, Dong Xu, Fengsheng Li

Increasing numbers of people live in flood-prone areas worldwide. With continued development, urban flood will become more frequent, which has caused casualties and property damage. Researchers have been dedicating to urban flood risk assessments in recent years. However, current research is still facing the challenges of multi-modal data fusion and knowledge representation of urban flood events. Therefore, in this paper, we propose an Urban Flood Knowledge Graph (UrbanFloodKG) system that enables KG to support urban flood risk assessment. The system consists of data layer, graph layer, algorithm layer, and application layer, which implements knowledge extraction and storage functions, integrates knowledge representation learning models and graph neural network models to support link prediction and node classification tasks. We conduct model comparison experiments on link prediction and node classification tasks based on urban flood event data from Guangzhou, and demonstrate the effectiveness of the models used. Our experiments prove that the accuracy of risk assessment can reach 91% when using GEN, which provides a promising research direction for urban flood risk assessment.

### **Optimizing Upstream Representations for Out-of-Domain Detection with Supervised Contrastive Learning**

Bo Wang, Tsunenori Mine

Out-of-Domain (OOD) text detection has attracted significant research interest. However, conventional approaches primarily employ Cross-Entropy loss during upstream encoder training and seldom focus on optimizing discriminative In-Domain (IND) and OOD representations. To fill this gap, we introduce a novel method that applies supervised contrastive learning (SCL) to IND data for upstream representation optimization. This effectively brings the embeddings of semantically similar texts together while pushing dissimilar ones further apart, leading to more compact and distinct IND representations. This

optimization subsequently improves the differentiation between IND and OOD representations, thereby enhancing the detection effect in downstream tasks. To further strengthen the ability of SCL to consolidate IND embedding clusters, and to improve the generalizability of the encoder, we propose a method that generates two different variations of the same text as "views". This is achieved by applying a twice "dropped-out" on the embeddings before performing SCL. Extensive experiments indicate that our method not only outperforms state-of-the-art approaches, but also reduces the requirement for training a large 354M-parameter model down to a more efficient 110M-parameter model, highlighting its superiority in both effectiveness and computational economy.

#### **Flexible and Robust Counterfactual Explanations with Minimal Satisfiable Perturbations**

Yongjie Wang, Hangwei Qian, Yongjie Liu, Wei Guo, Chunyan Miao

Counterfactual explanations (CFEs) exemplify how to minimally modify a feature vector to achieve a different prediction for an instance. CFEs can enhance informational fairness and trustworthiness, and provide suggestions for users who receive adverse predictions. However, recent research has shown that multiple CFEs can be offered for the same instance or instances with slight differences. Multiple CFEs provide flexible choices and cover diverse desiderata for user selection. However, individual fairness and model reliability will be damaged if unstable CFEs with different costs are returned. Existing methods fail to exploit flexibility and address the concerns of non-robustness simultaneously. To address these issues, we propose a conceptually simple yet effective solution named Counterfactual Explanations with Minimal Satisfiable Perturbations (CEMSP). Specifically, CEMSP constrains changing values of abnormal features with the help of their semantically meaningful normal ranges. For efficiency, we model the problem as a Boolean satisfiability problem to modify as few features as possible. Additionally, CEMSP is a general framework and can easily accommodate more practical requirements, e.g., causality and actionability. Compared to existing methods, we conduct comprehensive experiments on both synthetic and real-world datasets to demonstrate that our method provides more robust explanations while preserving flexibility.

#### **AFRF: Angle Feature Retrieval Based Popularity Forecasting**

Haoyu Wang, Zongxia Xie, Meiyao Liu, Canhua Guan

Social media popularity forecasting has become a hot research topic in recent years. It is of great significance in assisting public opinion monitoring and advertising placement. Time series prediction is one of the simple and commonly used methods for popularity forecasting, which takes the popularity of the first few time steps in the observed data as inputs. However, the complete popularity trend of each social media is known in the training dataset, while the historical time series information except for the first few time steps is neglected in the existing models. In order to utilize the complete historical information from the observed data, a retrieval method is introduced in this paper. Therefore, how to retrieve similar social media based on the first few steps time series and how to integrate the similar historical information have become two challenges. A two-stage prediction method named Angle Feature Retrieval based Forecasting (AFRF) is proposed in this paper to solve the upper two problems. In the first stage, based on the angle features of series, we retrieve K similar series from the historical posts and concatenate them with the target series as the model's input. In the second stage, an attention mechanism is used to learn the temporal relationships among the series and generate future popularity forecasts. We evaluated the multi-step and single-point forecasting performance of AFRF on three real-world datasets and compared it with state-of-the-art popularity forecasting methods, such as temporal feature-based and cascade-based methods, verifying the effectiveness of AFRF.

#### **Continuous Personalized Knowledge Tracing: Modeling Long-Term Learning in Online Environments**

Chunpai Wang, Shaghayegh Sahebi

With the advance of online education systems, accessibility to learning materials has increased. In these systems, students can practice independently and learn from different learning materials over long periods of time. As a result, it is essential to trace students' knowledge states over long learning sequences while maintaining a personalized model of each individual student's progress. However, the existing deep learning-based knowledge tracing models are either not personalized or not tailored for handling long sequences. Handling long sequences are especially essential in the online education environments, in where models are preferred to be updated with the newly collected user data in a timely manner as students could acquire knowledge on each learning activity. In this paper, we propose a knowledge tracing model, Continuous Personalized Knowledge Tracing (CPKT), that can mimic the real-world long-term continuous learning scenario by incorporating a novel online model training paradigm that is suitable for the knowledge tracing problem. To achieve personalized knowledge tracing, we propose two model components: 1) personalized memory slots to maintain learner's knowledge in a lifelong

manner, and 2) personalized user embeddings that help to accurately predict the individual responses, correctly detect the personalized knowledge acquisition and forgetting patterns, and better interpret and analyze the learner's progress. Additionally, we propose transition-aware stochastic shared embedding according to the learning transition matrix to regularize the online model training. Extensive experiments on four real-world datasets showcase the effectiveness and superiority of CPKT, especially for students with longer sequences.

### **Low-bit Quantization for Deep Graph Neural Networks with Smoothness-aware Message Propagation**

Shuang Wang, Bahaeddin Eravci, Rustam Guliyev, Hakan Ferhatosmanoglu

Graph Neural Network (GNN) training and inference involve significant challenges of scalability with respect to both model sizes and number of layers, resulting in degradation of efficiency and accuracy for large and deep GNNs. We present an end-to-end solution that aims to address these challenges for efficient GNNs in resource constrained environments while avoiding the oversmoothing problem in deep GNNs. We introduce a quantization based approach for all stages of GNNs, from message passing in training to node classification, compressing the model and enabling efficient processing. The proposed GNN quantizer learns quantization ranges and reduces the model size with comparable accuracy even under low-bit quantization. To scale with the number of layers, we devise a message propagation mechanism in training that controls layer-wise changes of similarities between neighboring nodes. This objective is incorporated into a Lagrangian function with constraints and a differential multiplier method is utilized to iteratively find optimal embeddings. This mitigates oversmoothing and suppresses the quantization error to a bound. Significant improvements are demonstrated over state-of-the-art quantization methods and deep GNN approaches in both full-precision and quantized models. The proposed quantizer demonstrates superior performance in INT2 configurations across all stages of GNN, achieving a notable level of accuracy. In contrast, existing quantization approaches fail to generate satisfactory accuracy levels. Finally, the inference with INT2 and INT4 representations exhibits a speedup of  $5.11 \times$  and  $4.70 \times$  compared to full precision counterparts, respectively.

### **A Mix-up Strategy to Enhance Adversarial Training with Imbalanced Data**

Wentao Wang, Harry Shomer, Yuxuan Wan, Yaxin Li, Jiangtao Huang, Hui Liu

Adversarial training has been proven to be one of the most effective techniques to defend against adversarial examples. The majority of existing adversarial training methods assume that every class in the training data is equally distributed. However, in reality, some classes often have a large number of training data while others only have a very limited amount. Recent studies have shown that the performance of adversarial training will degrade drastically if the training data is imbalanced. In this paper, we propose a simple yet effective framework to enhance the robustness of DNN models under imbalanced scenarios. Our framework, Imb-Mix, first augments the training dataset by generating multiple adversarial examples for samples in the minority classes. This is done by first adding random noise to the original adversarial examples created by one specific adversarial attack method. It then constructs Mixup-mimic mixed examples upon the augmented dataset used by adversarial training. In addition, we theoretically prove the regularization effect of our Mixup-mimic mixed examples generation technique in Imb-Mix. Extensive experiments on various imbalanced datasets verify the effectiveness of the proposed framework.

### **Node-dependent Semantic Search over Heterogeneous Graph Neural Networks**

Zhenyi Wang, Huan Zhao, Fengqi Liang, Chuan Shi

In recent years, Heterogeneous Graph Neural Networks (HGNNs) have been the state-of-the-art approaches for various tasks on Heterogeneous Graphs (HGs), e.g., recommendation and social network analysis. Despite the success of existing HGNNs, the utilization of the intricate semantic information in HGs is still insufficient. In this work, we study the problem of how to design powerful HGNNs under the guidance of node-dependent semantics. Specifically, to perform semantic search over HGNNs, we propose to develop semantic structures in terms of relation selection and connection selection, which could guide a task-relevant message flow. Furthermore, to better capture the diversified property of different node samples in HGs, we design predictors to adaptively decide the semantic structures per node. Extensive experiments on seven benchmarking datasets across different downstream tasks, i.e., node classification and recommendation, show that our method can consistently outperform various state-of-the-art baselines with shorter inference latency, which justifies its effectiveness and efficiency. The code and data are available at <https://github.com/BUPT-GAMMA/NDS>.

### **NOVO: Learnable and Interpretable Document Identifiers for Model-Based IR**

Zihan Wang, Yujia Zhou, Yiteng Tu, Zhicheng Dou

Model-based Information Retrieval (Model-based IR) has gained attention due to advancements in generative language models. Unlike traditional dense retrieval methods relying on dense vector representations of documents, model-based IR leverages language models to retrieve documents by generating their unique discrete identifiers (docids). This approach effectively reduces the requirements to store separate document representations in an index. Most existing model-based IR approaches utilize pre-defined static docids, i.e., these docids are fixed and are not learnable by training on the retrieval tasks. However, these docids are not specifically optimized for retrieval tasks, which makes it difficult to learn semantics and relationships between documents and achieve satisfactory retrieval performance. To address the above limitations, we propose Neural Optimized VOcabularial (NOVO) docids. NOVO docids are unique n-gram sets identifying each document. They can be generated in any order to retrieve the corresponding document and can be optimized through training to better learn semantics and relationships between documents. We propose to optimize NOVO docids through query denoising modeling and retrieval tasks, allowing for optimizing both semantic and token representations for such docids. Experiments on two datasets under the normal and zero-shot settings show that NOVO exhibits strong performance in more effective and interpretable model-based IR.

#### **WOT-Class: Weakly Supervised Open-world Text Classification**

Tianle Wang, Zihan Wang, Weitang Liu, Jingbo Shang

State-of-the-art weakly supervised text classification methods, while significantly reduced the required human supervision, still requires the supervision to cover all the classes of interest. This is never easy to meet in practice when human explore new, large corpora without complete pictures. In this paper, we work on a novel yet important problem of weakly supervised open-world text classification, where supervision is only needed for a few examples from a few known classes and the machine should handle both known and unknown classes in test time. General open-world classification has been studied mostly using image classification; however, existing methods typically assume the availability of sufficient known-class supervision and strong unknown-class prior knowledge (e.g., the number and/or data distribution). We propose a novel framework *our* that lifts those strong assumptions. Specifically, it follows an iterative process of (a) clustering text to new classes, (b) mining and ranking indicative words for each class, and (c) merging redundant classes by using the overlapped indicative words as a bridge. Extensive experiments on 7 popular text classification datasets demonstrate that *our* outperforms strong baselines consistently with a large margin, attaining 23.33% greater average absolute macro-F1 over existing approaches across all datasets. Such competent accuracy illuminates the practical potential of further reducing human effort for text classification.

#### **MultiPLe: Multilingual Prompt Learning for Relieving Semantic Confusions in Few-shot Event Detection**

Siyuan Wang, Jianming Zheng, Wanyu Chen, Fei Cai, Xueshan Luo

Event detection (ED) is a challenging task in the field of information extraction. Due to the monolingual text and rampant confusing triggers, traditional ED models suffer from semantic confusions in terms of polysemy and synonym, leading to severe detection mistakes. Such semantic confusions can be further exacerbated in a practical situation where scarce labeled data cannot provide sufficient semantic clues. To mitigate such bottleneck, we propose a multilingual prompt learning (MultiPLe) framework for few-shot event detection (FSED), including three components, i.e., a multilingual prompt, a hierarchical prototype and a quadruplet contrastive learning module. In detail, to ease the polysemy confusion, the multilingual prompt module develops the in-context semantics of triggers via the multilingual disambiguation and prior knowledge in pretrained language models. Then, the hierarchical prototype module is adopted to diminish the synonym confusion by connecting the captured inmost semantics of fuzzy triggers with labels at a fine granularity. Finally, we employ the quadruplet contrastive learning module to tackle the insufficient label representation and potential noise. Experiments on two public datasets show that MultiPLe outperforms the state-of-the-art baselines in weighted F1-score, presenting a maximum improvement of 13.63% for FSED.

#### **Selecting Top-k Data Science Models by Example Dataset**

Mengying Wang, Sheng Guan, Hanchao Ma, Yiyang Bian, Haolai Che, Abhishek Daundkar, Alp Sehrioglu, Yinghui Wu

Data analytical pipelines routinely involve various domain-specific data science models. Such models require expensive manual or training effort and often incur expensive validation costs (e.g., via scientific simulation analysis). Meanwhile, high-value models remain to be ad-hocly created, isolated, and underutilized for a broad community. Searching and accessing proper models for data analysis pipelines is desirable yet challenging for users without domain knowledge. This paper introduces ModsNet, a novel MODel Selection framework that only requires an Example daTaset. (1) We investigate the following problem: Given a library of pre-trained models, a limited amount of historical observations of their performance, and



an "example" dataset as a query, return  $k$  models that are expected to perform the best over the query dataset. (2) We formulate a regression problem and introduce a knowledge-enhanced framework using a model-data interaction graph. Unlike traditional methods, (1) ModsNet uses a dynamic, cost-bounded "probe-and-select" strategy to incrementally identify promising pre-trained models in a strict cold-start scenario (when a new dataset without any interaction with existing models is given). (2) To reduce the learning cost, we develop a clustering-based sparsification strategy to prune unpromising models and their interactions. (3) We showcase of ModsNet built on top of a crowdsourced materials knowledge base platform. Our experiments verified its effectiveness, efficiency, and applications over real-world analytical pipelines.

#### [A Multi-Modality Framework for Drug-Drug Interaction Prediction by Harnessing Multi-source Data](#)

Qianlong Wen, Jiazheng Li, Chuxu Zhang, Yanfang Ye

Drug-drug interaction (DDI), as a possible result of drug combination treatment, could lead to adverse physiological reactions and increasing mortality rates of patients. Therefore, predicting potential DDI has always been an important and challenging issue in medical health applications. Owing to the extensive pharmacological research, we can get access to various drug-related features for DDI predictions; however, most of the existing works on DDI prediction do not incorporate comprehensive features to analyze the DDI patterns. Despite the high performance that the existing works have achieved, the incomplete and noisy information generated from limited sources usually leads to sub-optimal performance and poor generalization ability on the unknown DDI pairs. In this work, we propose a holistic framework, namely Multi-modality Feature Optimal Fusion for Drug-Drug Interaction Prediction (MOF-DDI), that incorporates the features from multiple data sources to resolve the DDI predictions. Specifically, the proposed model jointly considers DDIs literature descriptions, biomedical knowledge graphs, and drug molecular structures to make the prediction. To overcome the issue induced by directly aggregating features in different modalities, we bring a new insight by mapping the representations learned from different sources to a unified hidden space before the combination. The empirical results show that MOF-DDI achieves a large performance gain on different DDI datasets compared with multiple state-of-the-art baselines, especially under the inductive setting.

#### [Modeling Preference as Weighted Distribution over Functions for User Cold-start Recommendation](#)

Jingxuan Wen, Huafeng Liu, Liping Jing

User cold-start recommendation is a well-known challenge in current recommender systems. The cause is that the number of user interactions is too few to accurately estimate user preferences. Furthermore, the uncertainty of user interactions intensifies along with the number of user interactions decreasing. Although existing meta-learning based models with globally sharing knowledge show good performance in most cold-start scenarios, the ability of handling challenges on intention importance and prediction uncertainty is missing: (1) Intra-user uncertainty. When estimating user preferences (reflected in the user's latent representation), each of user interactions is independently considered in the form of user-item pair, which cannot capture the correlation between user interactions, as well as considering the global intent under user interactions. (2) Inter-user importance. During the model training, all users are treated as equally important, which cannot distinguish the contribution of users in the model training process. Assigning the same weight to all users may lead to users with high uncertainty incorrectly guiding the model learning in the early stage of training. To tackle the above challenges, in this paper, we focus on modeling user preference as a weighted distribution over functions (WDoF) for user cold-start recommendation, which not only models the intra-user uncertainty through neural processes with Multinomial likelihood but also considers the importance of different users with curriculum learning during the model training process. Furthermore, we provide a theoretical explanation that why the proposed model performs better than regular neural processes based recommendation methods. Experiments on four real-world datasets demonstrate the effectiveness of the proposed model over several state-of-the-art cold-start recommendation methods.

#### [Dual Intents Graph Modeling for User-centric Group Discovery](#)

Xixi Wu, Yun Xiong, Yao Zhang, Yizhu Jiao, Jiawei Zhang

Online groups have become increasingly prevalent, providing users with space to share experiences and explore interests. Therefore, user-centric group discovery task, i.e., recommending groups to users can help both users' online experiences and platforms' long-term developments. Existing recommender methods can not deal with this task as modeling user-group participation into a bipartite graph overlooks their item-side interests. Although there exist a few works attempting to address this task, they still fall short in fully preserving the social context and ensuring effective interest representation learning.

In this paper, we focus on exploring the intents that motivate users to participate in groups, which can be categorized into different types, like the social-intent and the personal interest-intent. The former refers to users joining a group affected by

their social links, while the latter relates to users joining groups with like-minded people for interest-oriented self-enjoyment. To comprehend different intents, we propose a novel model, DiRec, that first models each intent separately and then fuses them together for predictions. Specifically, for social-intent, we introduce the hypergraph structure to model the relationship between groups and members. This allows for more comprehensive group information preservation, leading to a richer understanding of the social context. As for interest-intent, we employ novel structural refinement on the interactive graph to uncover more intricate user behaviors, item characteristics, and group interests, realizing better representation learning of interests. Furthermore, we also observe the intent overlapping in real-world scenarios and devise a novel self-supervised learning loss that encourages such alignment for final recommendations. Extensive experiments on three public datasets show the significant improvement of DiRec over the state-of-the-art methods.

#### **Prompt-and-Align: Prompt-Based Social Alignment for Few-Shot Fake News Detection**

Jiaying Wu, Shen Li, Ailin Deng, Miao Xiong, Bryan Hooi

Despite considerable advances in automated fake news detection, due to the timely nature of news, it remains a critical open question how to effectively predict the veracity of news articles based on limited fact-checks. Existing approaches typically follow a "Train-from-Scratch" paradigm, which is fundamentally bounded by the availability of large-scale annotated data. While expressive pre-trained language models (PLMs) have been adapted in a "Pre-Train-and-Fine-Tune" manner, the inconsistency between pre-training and downstream objectives also requires costly task-specific supervision. In this paper, we propose "Prompt-and-Align" (P&A), a novel prompt-based paradigm for few-shot fake news detection that jointly leverages the pre-trained knowledge in PLMs and the social context topology. Our approach mitigates label scarcity by wrapping the news article in a task-related textual prompt, which is then processed by the PLM to directly elicit task-specific knowledge. To supplement the PLM with social context without inducing additional training overheads, motivated by empirical observation on user veracity consistency (i.e., social users tend to consume news of the same veracity type), we further construct a news proximity graph among news articles to capture the veracity-consistent signals in shared readerships, and align the prompting predictions along the graph edges in a confidence-informed manner. Extensive experiments on three real-world benchmarks demonstrate that P&A sets new states-of-the-art for few-shot fake news detection performance by significant margins.

#### **SplitGNN: Spectral Graph Neural Network for Fraud Detection against Heterophily**

Bin Wu, Xinyu Yao, Boyan Zhang, Kuo-Ming Chao, Yinsheng Li

Fraudsters in the real world frequently add more legitimate links while concealing their direct ones with other fraudsters, leading to heterophily in fraud graphs, which is a problem that most GNN-based techniques are not built to solve. Several works have been proposed to tackle the issue from the spatial domain. However, researches on addressing the heterophily problem in the spectral domain are still limited due to a lack of understanding of spectral energy distribution in graphs with heterophily. In this paper, we analyze the spectral distribution with different heterophily degrees and observe that the heterophily of fraud nodes leads to the spectral energy moving from low-frequency to high-frequency. Further, we verify that splitting graphs using heterophilic and homophilic edges can obtain more significant expressions of signals in different frequency bands. The observation drives us to propose the spectral graph neural network, SplitGNN, to capture signals for fraud detection against heterophily. SplitGNN uses an edge classifier to split the original graph and adopts flexible band-pass graph filters to learn representations. Extensive experiments on real-world datasets demonstrate the effectiveness of our proposed method. The code and data are available at <https://github.com/Split-GNN/SplitGNN>.

#### **Community-Based Hierarchical Positive-Unlabeled (PU) Model Fusion for Chronic Disease Prediction**

Yang Wu, Xurui Li, Xuhong Zhang, Yangyang Kang, Changlong Sun, Xiaozhong Liu

Positive-Unlabeled (PU) Learning is a challenge presented by binary classification problems where there is an abundance of unlabeled data along with a small number of positive data instances, which can be used to address chronic disease screening problem. State-of-the-art PU learning methods have resulted in the development of various risk estimators, yet they neglect the differences among distinct populations. To address this issue, we present a novel Positive-Unlabeled Learning Tree (PUtree) algorithm. PUtree is designed to take into account communities such as different age or income brackets, in tasks of chronic disease prediction. We propose a novel approach for binary decision-making, which hierarchically builds community-based PU models and then aggregates their deliverables. Our method can explicate each PU model on the tree for the optimized non-leaf PU node splitting. Furthermore, a mask-recovery data augmentation strategy enables sufficient training of the model in individual communities. Additionally, the proposed approach includes an adversarial PU risk estimator to capture hierarchical PU-relationships, and a model fusion network that integrates data from each tree

path, resulting in robust binary classification results. We demonstrate the superior performance of PUtree as well as its variants on two benchmarks and a new diabetes-prediction dataset.

### **Rethinking Sentiment Analysis under Uncertainty**

Yuefei Wu, Bin Shi, Jiarun Chen, Yuhang Liu, Bo Dong, Qinghua Zheng, Hua Wei

Sentiment Analysis (SA) is a fundamental task in natural language processing, which is widely used in public decision-making. Recently, deep learning have demonstrated great potential to deal with this task. However, prior works have mostly treated SA as a deterministic classification problem, and meanwhile, without quantifying the predictive uncertainty. This presents a serious problem in the SA, different annotator, due to the differences in beliefs, values, and experiences, may have different perspectives on how to label the text sentiment. Such situation will lead to inevitable data uncertainty and make the deterministic classification models feel puzzle to make decision. To address this issue, we propose a new SA paradigm with the consideration of uncertainty and conduct an expensive empirical study. Specifically, we treat SA as the regression task and introduce uncertainty quantification to obtain confidence intervals for predictions, which enables the risk assessment ability of the model and can improve the credibility of SA-aids decision-making. Experiments on five datasets show that our proposed new paradigm effectively quantifies uncertainty in SA while remaining competitive performance to point estimation, in addition to being capable of Out-Of-Distribution~(OOD) detection.

### **Dimension Independent Mixup for Hard Negative Sample in Collaborative Filtering**

Xi Wu, Liangwei Yang, Jibing Gong, Chao Zhou, Tianyu Lin, Xiaolong Liu, Philip S. Yu

Collaborative filtering (CF) is a widely employed technique that predicts user preferences based on past interactions. Negative sampling plays a vital role in training CF-based models with implicit feedback. In this paper, we propose a novel perspective based on the sampling area to revisit existing sampling methods. We point out that current sampling methods mainly focus on Point-wise or Line-wise sampling, lacking flexibility and leaving a significant portion of the hard sampling area un-explored. To address this limitation, we propose Dimension Independent Mixup for Hard Negative Sampling (DINS), which is the first Area-wise sampling method for training CF-based models. DINS comprises three modules: Hard Boundary Definition, Dimension Independent Mixup, and Multi-hop Pooling. Experiments with real-world datasets on both matrix factorization and graph-based models demonstrate that DINS outperforms other negative sampling methods, establishing its effectiveness and superiority. Our work contributes a new perspective, introduces Area-wise sampling, and presents DINS as a novel approach that achieves state-of-the-art performance for negative sampling. Our implementations are available in PyTorch.

### **Towards Communication-Efficient Model Updating for On-Device Session-Based Recommendation**

Xin Xia, Junliang Yu, Guandong Xu, Hongzhi Yin

On-device recommender systems recently have garnered increasing attention due to their advantages of providing prompt response and securing privacy. To stay current with evolving user interests, cloud-based recommender systems are periodically updated with new interaction data. However, on-device models struggle to retrain themselves because of limited onboard computing resources. As a solution, we consider the scenario where the model retraining occurs on the server side and then the updated parameters are transferred to edge devices via network communication. While this eliminates the need for local retraining, it incurs a regular transfer of parameters that significantly taxes network bandwidth. To mitigate this issue, we develop an efficient approach based on compositional codes to compress the model update. This approach ensures the on-device model is updated flexibly with minimal additional parameters whilst utilizing previous knowledge. The extensive experiments conducted on multiple session-based recommendation models with distinctive architectures demonstrate that the on-device model can achieve comparable accuracy to the retrained server-side counterpart through transferring an update 60x smaller in size. The codes are available at <https://github.com/xiaxin1998/ODUpdate>.

### **HiPo: Detecting Fake News via Historical and Multi-Modal Analyses of Social Media Posts**

Tianshu Xiao, Sichang Guo, Jingcheng Huang, Riccardo Spolaor, Xiuzhen Cheng

In recent years, fake news has been a primary concern as it plays a significant role in influencing the political, economic, and social spheres. The scientific community has proposed several solutions to detect such fraudulent information. However, such solutions are unsuitable for social media posts since they cannot extract sufficient information from one-line textual and graphical content or are highly dependent on prior knowledge, which may be unavailable in the case of unprecedented events (e.g., breaking news).

This paper tackles this issue by proposing HiPo, a novel multi-modal historical post-based fake news detection method. By combining the features extracted from the graphical and textual content, HiPo assesses the truthfulness of a social media post by building its historical context from prior off-label posts with high similarity, therefore, achieving online detection without maintaining a context or knowledge database. We evaluate the performance of HiPo via an exhaustive set of experiments involving four real-world datasets. Our method achieves a detection accuracy higher than 84%, outperforming the state-of-the-art methods in most experimental instances.

### [An Efficient Selective Ensemble Learning with Rejection Approach for Classification](#)

Hao Xu, Chiranjeev Chetia

Recent studies found that selective ensemble learning (e.g., dynamic ensemble selection) shows better predictive performance for classification tasks, compared to traditional static ensemble. However, there are some limitations of the available methods, such as high computational cost and multiple restrictions in base model ranking and aggregation (especially for class-imbalanced data modeling). Besides, the current methods make predictions for all data without measuring the credibility regarding different data patterns. This paper proposes a selective ensemble learning with rejection approach that aggregates base models from a different perspective. The approach introduces rejection measures to quantify base model credibility, and learns how to use the models according to their credibility on different sample patterns. It avoids the complexity in base model ranking and therefore is computationally more efficient than current methods. Any common evaluation metrics can be adopted in the selective ensemble strategy, which allows the developed model to handle class-imbalanced data properly. Also, a global rejection region is developed which indicates whether the ensemble model can provide reliable predictions for the targets. We implement the approach in the modeling of 12 datasets, including both class-imbalanced and class-balanced cases. Results show that the approach significantly reduces the inference time while showing promising performance, compared to 8 dynamic ensemble selection methods in the literature. Feature contributions and impacts of different rejection ratios on performance are also investigated to better demonstrate the approach.

### [Sentiment-aware Review Summarization with Personalized Multi-task Fine-tuning](#)

Hongyan Xu, Hongtao Liu, Zhepeng Lv, Qing Yang, Wenjun Wang

Personalized review summarization is a challenging task in recommender systems, which aims to generate condensed and readable summaries for product reviews. Recently, some methods propose to adopt the sentiment signals of reviews to enhance the review summarization. However, most previous works only share the semantic features of reviews via preliminary multi-task learning, while ignoring the rich personalized information of users and products, which is crucial to both sentiment identification and comprehensive review summarization. In this paper, we propose a sentiment-aware review summarization method with an elaborately designed multi-task fine-tuning framework to make full use of personalized information of users and products effectively based on Pretrained Language Models (PLMs). We first denote two types of personalized information including IDs and historical summaries to indicate their identification and semantics information respectively. Subsequently, we propose to incorporate the IDs of the user/product into the PLMs-based encoder to learn the personalized representations of input reviews and their historical summaries in a fine-tuning way. Based on this, an auxiliary context-aware review sentiment classification task and a further sentiment-guided personalized review summarization task are jointly learned. Specifically, the sentiment representation of input review is used to identify relevant historical summaries, which are then treated as additional semantic context features to enhance the summary generation process. Extensive experimental results show our approach could generate sentiment-consistent summaries and outperforms many competitive baselines on both review summarization and sentiment classification tasks.

### [Density-Aware Temporal Attentive Step-wise Diffusion Model For Medical Time Series Imputation](#)

Jingwen Xu, Fei Lyu, Pong C. Yuen

Medical time series have been widely employed for disease prediction. Missing data hinders accurate prediction. While existing imputation methods partially solve the problem, there are two challenges for medical time series: (1) High dimensionality: Existing imputation methods suffer from the trade-off between accuracy and computational efficiency. (2) Irregularity: Medical time series exhibit the dynamic temporal relationship that changes over varying sampling densities. However, existing methods mainly take the stationary mechanism, which struggles with capturing the dynamic temporal relationships. To overcome the above deficiencies, we propose a Density-Aware Temporal Attentive Step-wise Diffusion Model (DA-TASWDM), which imputes each time step based on a non-iterative diffusion model and captures inter-step dependency with the density-aware time similarity. Specifically, DA-TASWDM exploits two novel modules: (1) Density-Aware Temporal Attention (DA-TA): It correlates inter-step values from the time embedding similarity adjusted with varying

sampling densities. (2) Non-Iterative Step-wise Diffusion Imputer (NI-SWDI): It directly recovers the missing values at each time step from noise without diffusion iteration. Compared with the existing methods, DA-TASWDM can achieve promising accuracy without sacrificing computational efficiency. Extensive experimental results on three real-world datasets demonstrate that our method can significantly outperform state-of-the-art methods in both imputation and post-imputation performance.

### [DPGN: Denoising Periodic Graph Network for Life Service Recommendation](#)

Hao Xu, Huixuan Chi, Danyang Liu, Sheng Zhou, Mengdi Zhang

Different from traditional e-commerce platforms, life service recommender systems provide hundreds of millions of users with daily necessities services such as nearby food ordering. In this scenario, users have instant intentions and living habits, which exhibit a periodic tendency to click or buy products with similar intentions. This can be summarized as the intentional periodicity problem, which was not well-studied in previous works. Existing periodic-related recommenders exploit time-sensitive functions to capture the evolution of user preferences. However, these methods are easily affected by the real noisy signal in life service platforms, wherein the recent noisy signals can mislead the instant intention and living habits modeling. We summarize it as the noise issue. Although there are some denoising recommenders, these methods cannot effectively solve the noise issue for intentional periodicity modeling.

To alleviate the issues, we propose a novel Denoising Periodic Graph Network (DPGN) for life service recommendation. First, to alleviate the noisy signals and model the instant intention accurately, we propose (i) temporal pooling (TP) to encode the most representative information shared by recent behaviors; (ii) temporal encoding (TE) to encode the relative time intervals. Second, to capture the user's living habits accurately, we propose the memory mechanism to maintain a series of instant intentions in different time periods. Third, to further capture the intentional periodicity, we propose the temporal graph transformer (TGT) layer to aggregate temporal information. Last, the denoising task is further proposed to alleviate the noisy signals. Extensive experiments on both real-world public and industrial datasets validate the state-of-the-art performance of DPGN. Code is available in <https://github.com/ytchx1999/DPGN>

### [Minimizing Polarization in Noisy Leader-Follower Opinion Dynamics](#)

Wanyue Xu, Zhongzhi Zhang

The operation of creating edges has been widely applied to optimize relevant quantities of opinion dynamics. In this paper, we consider a problem of polarization optimization for the leader-follower opinion dynamics in a noisy social network with  $n$  nodes and  $m$  edges, where a group  $Q$  of  $q$  nodes are leaders, and the remaining  $n-q$  nodes are followers. We adopt the popular leader-follower DeGroot model, where the opinion of every leader is identical and remains unchanged, while the opinion of every follower is subject to white noise. The polarization is defined as the steady-state variance of the deviation of each node's opinion from leaders' opinion, which equals one half of the effective resistance  $R_Q$  between the node group  $Q$  and all other nodes. Concretely, we propose and study the problem of minimizing  $R_Q$  by adding  $k$  new edges with each incident to a node in  $Q$ . We show that the objective function is monotone and supermodular. We then propose a simple greedy algorithm with an approximation factor  $1-1/e$  that approximately solves the problem in  $O((n-q)^3)$  time. To speed up the computation, we also provide a fast algorithm to compute  $(1-1/e-\epsilon)$ -approximate effective resistance  $R_Q$ , the running time of which is  $O(mk\epsilon^{-2})$  for any  $\epsilon > 0$ , where the  $O(\cdot)$  notation suppresses the  $\text{poly}(\log n)$  factors. Extensive experiment results show that our second algorithm is both effective and efficient.

### [Time-series Shapelets with Learnable Lengths](#)

Akihiro Yamaguchi, Ken Ueno, Hisashi Kashima

Shapelets are subsequences that are effective for classifying time-series instances. Learning shapelets by a continuous optimization has recently been studied to improve computational efficiency and classification performance. However, existing methods have employed predefined and fixed shapelet lengths during the continuous optimization, despite the fact that shapelets and their lengths are inherently interdependent and thus should be jointly optimized. To efficiently explore shapelets of high quality in terms of interpretability and inter-class separability, this study makes the shapelet lengths continuous and learnable. The proposed formulation jointly optimizes not only a binary classifier and shapelets but also shapelet lengths. The derived SGD optimization can be theoretically interpreted as improving the quality of shapelets in terms of shapelet closeness to the time series for target / off-target classes. We demonstrate improvements in area under the curve, total training time, and shapelet interpretability on UCR binary datasets.

### [Identifying Regional Driving Risks via Transductive Cross-City Transfer Learning Under Negative Transfer](#)



Hua Yan, Hao Wang, Desheng Zhang, Yu Yang

Identifying regional driving risks is important for real-world applications such as driving safety warning applications, public safety management, and insurance company premium pricing. Previous approaches are either based on traffic accident reports or vehicular sensor data. They either fail to identify potential risks, such as near-miss collisions, which would need other important measurements (e.g., hard break, acceleration, etc.), or fail to generalize to cities without vehicular sensor data, severely limiting their practicality. In this work, we address these two challenges and successfully identify regional driving risks in a target city without vehicular sensor data via cross-city transfer learning. Specifically, we design a novel framework RiskTrans by optimizing both the predictor and the relationship between cities to achieve transfer learning. We advance the existing works from two aspects: (i) we achieve it in a transductive manner without accessing labeled data in the target cities; (ii) we identify and address the problem of negative transfer in cross-city transfer learning, a prominent issue that is often (surprisingly) neglected in previous works. Finally, we conduct extensive experiments based on data collected from 175 thousand vehicles in six cities. The results show RiskTrans outperforms baselines by at least 50.2% and reduces negative transfer by 49.4%.

#### [Few-Shot Learning via Task-Aware Discriminant Local Descriptors Network](#)

Leilei Yan, Fanzhang Li, Xiaohan Zheng, Li Zhang

Few-shot learning for image classification task aims to classify images from several novel classes with limited number of samples. Recent studies have shown that the deep local descriptors have better representation ability than image-level features, and achieve great success. However, most of these methods often use all local descriptors or over-screening local descriptors for classification. The former contains some task-irrelevant descriptors, which may misguide the final classification result. The latter is likely to lose some key descriptors. In this paper, we propose a novel Task-Aware Discriminant local descriptors Network (TADNet) to address these issues, which can adaptively select the discriminative query descriptors and eliminate the task-irrelevant query descriptors among the entire task. Specifically, TADNet assigns a value to each query descriptor by comparing its similarity to all support classes to represent its discriminant power for classification. Then the discriminative query descriptors can be preserved via a task-aware attention map. Extensive experiments on both fine-grained and generalized datasets demonstrate that the proposed TADNet outperforms the existing state-of-the-art methods.

#### [MoCLIM: Towards Accurate Cancer Subtyping via Multi-Omics Contrastive Learning with Omics-Inference Modeling](#)

Ziwei Yang, Zheng Chen, Yasuko Matsubara, Yasushi Sakurai

Precision medicine fundamentally aims to establish causality between dysregulated biochemical mechanisms and cancer subtypes. Omics-based cancer subtyping has emerged as a revolutionary approach, as different level of omics records the biochemical products of multistep processes in cancers. This paper focuses on fully exploiting the potential of multi-omics data to improve cancer subtyping outcomes, and hence developed MoCLIM, a representation learning framework. MoCLIM independently extracts the informative features from distinct omics modalities. Using a unified representation informed by contrastive learning of different omics modalities, we can well-cluster the subtypes, given cancer, into a lower latent space. This contrast can be interpreted as a projection of inter-omics inference observed in biological networks. Experimental results on six cancer datasets demonstrate that our approach significantly improves data fit and subtyping performance in fewer high-dimensional cancer instances. Moreover, our framework incorporates various medical evaluations as the final component, providing high interpretability in medical analysis.

#### [FARA: Future-aware Ranking Algorithm for Fairness Optimization](#)

Tao Yang, Zhichao Xu, Zhenduo Wang, Qingyao Ai

Ranking systems are the key components of modern Information Retrieval (IR) applications, such as search engines and recommender systems. Besides the ranking relevance to users, the exposure fairness to item providers has also been considered an important factor in ranking optimization. Many fair ranking algorithms have been proposed to jointly optimize both ranking relevance and fairness. However, we find that most existing fair ranking methods adopt greedy algorithms that only optimize rankings for the next immediate session or request. As shown in this paper, such a myopic paradigm could limit the upper bound of ranking optimization and lead to suboptimal performance in the long term.

To this end, we propose FARA, a novel Future-Aware Ranking Algorithm for ranking relevance and fairness optimization. Instead of greedily optimizing rankings for the next immediate session, FARA plans ahead by jointly optimizing multiple ranklists together and saving them for future sessions. Specifically, FARA first uses the Taylor expansion to investigate how

future ranklists will influence the overall fairness of the system. Then, based on the analysis of the Taylor expansion, FARA adopts a two-phase optimization algorithm where we first solve an optimal future exposure planning problem and then construct the optimal ranklists according to the optimal future exposure planning. Theoretically, we show that FARA is optimal for ranking relevance and fairness joint optimization. Empirically, our extensive experiments on three semi-synthesized datasets show that FARA is efficient, effective, and can deliver significantly better ranking performance compared to state-of-the-art fair ranking methods. We make our implementation public at [https://github.com/Taosheng-ty/QP\\_fairness/](https://github.com/Taosheng-ty/QP_fairness/).

### [\*\*A Bipartite Graph is All We Need for Enhancing Emotional Reasoning with Commonsense Knowledge\*\*](#)

Kailai Yang, Tianlin Zhang, Shaoxiong Ji, Sophia Ananiadou

The context-aware emotional reasoning ability of AI systems, especially in conversations, is of vital importance in applications such as online opinion mining from social media and empathetic dialogue systems. Due to the implicit nature of conveying emotions in many scenarios, commonsense knowledge is widely utilized to enrich utterance semantics and enhance conversation modeling. However, most previous knowledge infusion methods perform empirical knowledge filtering and design highly customized architectures for knowledge interaction with the utterances, which can discard useful knowledge aspects and limit their generalizability to different knowledge sources. Based on these observations, we propose a Bipartite Heterogeneous Graph (BHG) method for enhancing emotional reasoning with commonsense knowledge. In BHG, the extracted context-aware utterance representations and knowledge representations are modeled as heterogeneous nodes. Two more knowledge aggregation node types are proposed to perform automatic knowledge filtering and interaction. BHG-based knowledge infusion can be directly generalized to multi-type and multi-grained knowledge sources. In addition, we propose a Multi-dimensional Heterogeneous Graph Transformer (MHGT) to perform graph reasoning, which can retain unchanged feature spaces and unequal dimensions for heterogeneous node types during inference to prevent unnecessary loss of information. Experiments show that BHG-based methods significantly outperform state-of-the-art knowledge infusion methods and show generalized knowledge infusion ability with higher efficiency. Further analysis proves that previous empirical knowledge filtering methods do not guarantee to provide the most useful knowledge information. Our code is available at: <https://github.com/SteveKGYang/BHG>.

### [\*\*A Two-tier Shared Embedding Method for Review-based Recommender Systems\*\*](#)

Zhen Yang, Junrui Liu, Tong Li, Di Wu, Shiqiu Yang, Huan Liu

Reviews are valuable resources that have been widely researched and used to improve the quality of recommendation services. Recent methods use multiple full embedding layers to model various levels of individual preferences, increasing the risk of the data sparsity issue. Although it is a potential way to deal with this issue that models homophily among users who have similar behaviors, the existing approaches are implemented in a coarse-grained way. They calculate user similarities by considering the homophily in their global behaviors but ignore their local behaviors under a specific context. In this paper, we propose a two-tier shared embedding model (TSE), which fuses coarse- and fine-grained ways of modeling homophily. It considers global behaviors to model homophily in a coarse-grained way, and the high-level feature in the process of each user-item interaction to model homophily in a fine-grained way. TSE designs a whole-to-part principle-based process to fuse these ways in the review-based recommendation. Experiments on five real-world datasets demonstrate that TSE significantly outperforms state-of-the-art models. It outperforms the best baseline by 20.50% on the root-mean-square error (RMSE) and 23.96% on the mean absolute error (MAE), respectively. The source code is available at <https://github.com/dianziliu/TSE.git>.

### [\*\*CARPG: Cross-City Knowledge Transfer for Traffic Accident Prediction via Attentive Region-Level Parameter Generation\*\*](#)

Guang Yang, Yuequn Zhang, Jinqun Hang, Xinyue Feng, Zejun Xie, Desheng Zhang, Yu Yang

Traffic accident prediction is a crucial problem for public safety, emergency treatment, and urban management. Existing works leverage extensive data collected from city infrastructures to achieve encouraging performance based on various machine learning techniques but cannot achieve a good performance in situations with limited data (i.e., data scarcity). Recent developments in transfer learning bring a new opportunity to solve the data scarcity problem. In this paper, we design a novel cross-city transfer learning framework named CARPG for predicting traffic accidents in data-scarce cities. We address the unique challenge of predicting traffic accidents caused by its two fundamental characteristics, i.e., spatial heterogeneity and inherent rareness, which result in the biased performance of the state-of-the-art transfer learning methods. Specifically, we build cross-city region connections by jointly learning the spatial region representations for both source and target cities with an inter-city global graph knowledge transfer process. Further, we design an efficient attention-based

parameter-generating mechanism to learn region-specific traffic accident patterns, while controlling the total number of parameters. Built upon that, we ensure that only relevant patterns are transferred to each target region during the knowledge transfer process and further to be fine-tuned. We conduct extensive experiments on three real-world datasets, and the evaluation results demonstrate the superiority of our framework compared with state-of-the-art baseline models.

#### [Co-guided Random Walk for Polarized Communities Search](#)

Fanyi Yang, Huifang Ma, Cairui Yan, Zhixin Li, Liang Chang

Polarized Communities Search (PCS) aims to identify query-dependent communities where positive links predominantly connect nodes within each community, while negative links primarily connect nodes across different communities. Existing solutions primarily focus on modeling network topology, disregarding the crucial factor of node attributes. However, it is non-trivial to incorporate node attributes into PCS. In this paper, we propose a novel method called CO-guided RAndom walk in attributed signed networks (CORA) for PCS. Our approach involves constructing an attribute-based signed network to represent the auxiliary relations between nodes. We introduce a weight assignment mechanism to assess the reliability of edges in the signed network. Then, we design a co-guided random walk scheme that operates on two signed networks to model the connections between network topology and node attributes, thereby enhancing the search outcomes. Finally, we identify polarized communities using the Rayleigh quotient in the signed network. Extensive experiments conducted on three public datasets demonstrate the superior performance of CORA compared to state-of-the-art baselines for polarized communities search.

#### [Multimodal Optimal Transport Knowledge Distillation for Cross-domain Recommendation](#)

Wei Yang, Jie Yang, Yuan Liu

Recommendation systems have been widely used in e-commerce, news media, and short video platforms. With the abundance of images, text, and audio information, users often engage in personalized interactions based on their multimodal preferences. With the continuous expansion of application scenarios, cross domain recommendation issues have become important, such as recommendations in both the public and private domains of e-commerce. The current cross domain recommendation methods have achieved certain results through methods such as shared encoders and contrastive learning. However, few studies have focused on the effective extraction and utilization of multimodal information in cross domain recommendations. Furthermore, due to the existence of distribution drift issues, directly constructing feature alignment between source domain and target domain representations is not an effective way. Therefore, we propose a Multimodal Optimal Transport Knowledge Distillation (MOTKD) method for cross domain recommendation. Specifically, we propose a multimodal graph attention network to model the multimodal preference representation of users. Then, we introduce a proxy distribution space as a bridge between the source and target domains. Based on the common proxy distribution, we utilize the optimal transport method to achieve cross domain knowledge transfer. Further, in order to improve the auxiliary training effect of source domain supervised signals on target domain, we design a multi-level cross domain knowledge distillation module. We conducted extensive experiments on two pairs of cross domain datasets composed of four datasets. The experimental results indicate that our proposed MOTKD method outperforms other state-of-the-art models.

#### [Group Identification via Transitional Hypergraph Convolution with Cross-view Self-supervised Learning](#)

Mingdai Yang, Zhiwei Liu, Liangwei Yang, Xiaolong Liu, Chen Wang, Hao Peng, Philip S. Yu

With the proliferation of social media, a growing number of users search for and join group activities in their daily life. This develops a need for the study on the group identification (GI) task, i.e., recommending groups to users. The major challenge in this task is how to predict users' preferences for groups based on not only previous group participation of users but also users' interests in items. Although recent developments in Graph Neural Networks (GNNs) accomplish embedding multiple types of objects in graph-based recommender systems, they, however, fail to address this GI problem comprehensively. In this paper, we propose a novel framework named Group Identification via Transitional Hypergraph Convolution with Graph Self-supervised Learning (GTGS). We devise a novel transitional hypergraph convolution layer to leverage users' preferences for items as prior knowledge when seeking their group preferences. To construct comprehensive user/group representations for GI task, we design the cross-view self-supervised learning to encourage the intrinsic consistency between item and group preferences for each user, and the group-based regularization to enhance the distinction among group embeddings. Experimental results on three benchmark datasets verify the superiority of GTGS. Additional detailed investigations are conducted to demonstrate the effectiveness of the proposed framework.

#### [Mulco: Recognizing Chinese Nested Named Entities through Multiple Scopes](#)

Jiuding Yang, Jinwen Luo, Weidong Guo, Jerry Chen, Di Niu, Yu Xu

Nested Named Entity Recognition (NNER), as a subarea of Named Entity Recognition, has presented longstanding challenges to researchers. In NNER, one entity may be part of a larger entity, which can occur at multiple levels. These nested structures prevent traditional sequence labeling methods from properly recognizing all entities. While recent research has focused on designing better recognition methods for NNER in various languages, Chinese Nested Named Entity Recognition (CNNER) is still underdeveloped, largely due to a lack of freely available CNNER benchmarks. To support CNNER research, in this paper, we introduce ChiNesE, a CNNER dataset comprising 20,000 sentences from online passages in multiple domains and containing 117,284 entities that fall into 10 categories, of which 43.8% are nested named entities. Based on ChiNesE, we propose Mulco, a novel method that can recognize named entities in nested structures through multiple scopes. Each scope uses a scope-based sequence labeling method that predicts an anchor and the length of a named entity to recognize it. Experimental results show that Mulco outperforms state-of-the-art baseline methods with different recognition schemes on ChiNesE and ACE 2005 Chinese corpus.

### [Explore Epistemic Uncertainty in Domain Adaptive Semantic Segmentation](#)

Kai Yao, Zixian Su, Xi Yang, Jie Sun, Kaizhu Huang

In domain adaptive segmentation, domain shift may cause erroneous high-confidence predictions on the target domain, resulting in poor self-training. To alleviate the potential error, most previous works mainly consider aleatoric uncertainty arising from the inherit data noise. This may however lead to overconfidence in incorrect predictions and thus limit the performance. In this paper, we take advantage of Deterministic Uncertainty Methods (DUM) to explore the epistemic uncertainty, which reflects accurately the domain gap depending on the model choice and parameter fitting trained on source domain. The epistemic uncertainty on target domain is evaluated on-the-fly to facilitate online reweighting and correction in the self-training process. Meanwhile, to tackle the class-wise quantity and learning difficulty imbalance problem, we introduce a novel data resampling strategy to promote simultaneous convergence across different categories. This strategy prevents the class-level over-fitting in source domain and further boosts the adaptation performance by better quantifying the uncertainty in target domain. We illustrate the superiority of our method compared with the state-of-the-art methods.

### [Improving Query Correction Using Pre-train Language Model In Search Engines](#)

Dezhi Ye, Bowen Tian, Jiabin Fan, Jie Liu, Tianhua Zhou, Xiang Chen, Mingming Li, Jin Ma

Query correction is a task that automatically detects and corrects errors in what users type into a search engine. Misspelled queries can lead to user dissatisfaction and churn. However, correcting a user query accurately is not an easy task. One major challenge is that a correction model must be capable of high-level language comprehension. Recently, pre-trained language models (PLMs) have been successfully applied to text correction tasks, but few works have been done on query correction. However, it is nontrivial to directly apply these PLMs to query correction in large-scale search systems due to the following challenging issues: 1) Expensive deployment. Deploying such a model requires expensive computations. 2) Lacking domain knowledge. A neural correction model needs massive training data to activate its power.

To this end, we introduce KSTEM, a Knowledge-based Sequence To Edit Model for Chinese query correction. KSTEM transforms the sequence generation task into sequence tagging by mapping errors into five categories: KEEP, REPLACE, SWAP, DELETE, and INSERT, reducing computational complexity. Additionally, KSTEM adopts 2D position encoding, which is composed of the internal and external order of the words. Meanwhile, to compensate for the lack of domain knowledge, we propose a task-specific training paradigm for query correction, including edit strategy-based pre-training, user click-based post pre-train, and human label-based fine-tuning. Finally, we apply KSTEM to the industrial search system. Extensive offline and online experiments show that KSTEM significantly improves query correction performance. We hope that our experience will benefit frontier researchers.

### [APGL4SR: A Generic Framework with Adaptive and Personalized Global Collaborative Information in Sequential Recommendation](#)

Mingjia Yin, Hao Wang, Xiang Xu, Likang Wu, Sirui Zhao, Wei Guo, Yong Liu, Ruiming Tang, Defu Lian, Enhong Chen

The sequential recommendation system has been widely studied for its promising effectiveness in capturing dynamic preferences buried in users' sequential behaviors. Despite the considerable achievements, existing methods usually focus on intra-sequence modeling while overlooking exploiting global collaborative information by inter-sequence modeling, resulting in inferior recommendation performance. Therefore, previous works attempt to tackle this problem with a global collaborative item graph constructed by pre-defined rules. However, these methods neglect two crucial properties when capturing global

collaborative information, i.e., adaptiveness and personalization, yielding sub-optimal user representations. To this end, we propose a graph-driven framework, named Adaptive and Personalized Graph Learning for Sequential Recommendation (APGL4SR), that incorporates adaptive and personalized global collaborative information into sequential recommendation systems. Specifically, we first learn an adaptive global graph among all items and capture global collaborative information with it in a self-supervised fashion, whose computational burden can be further alleviated by the proposed SVD-based accelerator. Furthermore, based on the graph, we propose to extract and utilize personalized item correlations in the form of relative positional encoding, which is a highly compatible manner of personalizing the utilization of global collaborative information. Finally, the entire framework is optimized in a multi-task learning paradigm, thus each part of APGL4SR can be mutually reinforced. As a generic framework, APGL4SR can not only outperform other baselines with significant margins, but also exhibit promising versatility, the ability to learn a meaningful global collaborative graph, and the ability to alleviate the dimensional collapse issue of item embeddings.

#### **FINRule: Feature Interactive Neural Rule Learning**

Lu Yu, Meng Li, Ya-Lin Zhang, Longfei Li, Jun Zhou

Though neural networks have achieved impressive prediction performance, it's still hard for people to understand what neural networks have learned from the data. The black-box property of neural networks already becomes one of the main obstacles preventing from being applied to many high-stakes applications, such as finance and medicine that have critical requirement on the model transparency and interpretability. In order to enhance the explainability of neural networks, we propose a neural rule learning method-Feature Interactive Neural Rule Learning (FINRule) to incorporate the expressivity of neural networks and the interpretability of rule-based systems. Specifically, we conduct rule learning as differential discrete combination encoded by a feedforward neural network, in which each layer acts as a logical operator of explainable decision conditions. The first hidden layer can act as sharable atomic conditions which are connected to next hidden layer for formulating decision rules. Moreover, we propose to represent both atomic condition and rules with contextual embeddings, with aim to enrich the expressivity power by capturing high-order feature interactions. We conduct comprehensive experiments on real-world datasets to validate both effectiveness and explainability of the proposed method.

#### **iHAS: Instance-wise Hierarchical Architecture Search for Deep Learning Recommendation Models**

Yakun Yu, Shi-ang Qi, Jiuding Yang, Liyao Jiang, Di Niu

Current recommender systems employ large-sized embedding tables with uniform dimensions for all features, leading to overfitting, high computational cost, and suboptimal generalizing performance. Many techniques aim to solve this issue by feature selection or embedding dimension search. However, these techniques typically select a fixed subset of features or embedding dimensions for all instances and feed all instances into one recommender model without considering heterogeneity between items or users. This paper proposes a novel instance-wise Hierarchical Architecture Search framework, iHAS, which automates neural architecture search at the instance level. Specifically, iHAS incorporates three stages: searching, clustering, and retraining. The searching stage identifies optimal instance-wise embedding dimensions across different field features via carefully designed Bernoulli gates with stochastic selection and regularizers. After obtaining these dimensions, the clustering stage divides samples into distinct groups via a deterministic selection approach of Bernoulli gates. The retraining stage then constructs different recommender models, each one designed with optimal dimensions for the corresponding group. We conduct extensive experiments to evaluate the proposed iHAS on two public benchmark datasets from a real-world recommender system. The experimental results demonstrate the effectiveness of iHAS and its outstanding transferability to widely-used deep recommendation models.

#### **Search Result Diversification Using Query Aspects as Bottlenecks**

Puxuan Yu, Razieh Rahimi, Zhiqi Huang, James Allan

We address some of the limitations of coverage-based search result diversification models, which often consist of separate components and rely on external systems for query aspects. To overcome these challenges, we introduce an end-to-end learning framework called DUB. Our approach preserves the intrinsic interpretability of coverage-based methods while enhancing diversification performance. Drawing inspiration from the information bottleneck method, we propose an aspect extractor that generates query aspect embeddings optimized as information bottlenecks for the task of diversified document re-ranking. Experimental results demonstrate that DUB outperforms state-of-the-art diversification models.

#### **Understanding and Modeling Collision Avoidance Behavior for Realistic Crowd Simulation**

Zihan Yu, Guozhen Zhang, Yong Li, Depeng Jin



For walking pedestrians, when they are blocked by obstacles or other pedestrians, they adjust their speeds and directions to avoid colliding with them, which is called collision avoidance behavior. This behavior is the most complex part of pedestrians' walking processes and its modeling and simulation are the keys to realistic crowd simulation, which serves as the foundation for various applications. However, most existing methods either lack the representation power to accurately model the complex collision behavior or do not model it explicitly, which leads to a poor level of realism of the simulation. To realize realistic crowd simulation, we propose to analyze, understand, and model the collision avoidance behavior in a data-driven way. First, to automatically detect collision avoidance behavior for further analysis, we propose a domain transformation algorithm that detects it by transforming the trajectories in the spatial domain into a new domain where the behavior is much more apparent and is thus easier to detect. The new domain also provides a new perspective for understanding collision avoidance behavior. Second, since there are no mature metrics to evaluate the level of realism, we propose a new evaluation metric based on the least-effort theory, which evaluates the realism of collision avoidance behavior by its physical and mental consumption. This evaluation metric also provides the foundation of modeling. Third, for realistic crowd simulation, we design a reinforcement learning model. It trains agents with our proposed reward function that models pedestrians' intrinsic needs of "reducing effort consumption" and thus can guide agents to behave realistically when avoiding collisions. Extensive experiments show our model is 55.9% and 52.5% more realistic in collision avoidance behavior than the best baselines on two real-world datasets. We release our codes at <https://github.com/tsinghua-fib-lab/TECRL>.

### **DSformer: A Double Sampling Transformer for Multivariate Time Series Long-term Prediction**

Chengqing Yu, Fei Wang, Zezhi Shao, Tao Sun, Lin Wu, Yongjun Xu

Multivariate time series long-term prediction, which aims to predict the change of data in a long time, can provide references for decision-making. Although transformer-based models have made progress in this field, they usually do not make full use of three features of multivariate time series: global information, local information, and variables correlation. To effectively mine the above three features and establish a high-precision prediction model, we propose a double sampling transformer (DSformer), which consists of the double sampling (DS) block and the temporal variable attention (TVA) block. Firstly, the DS block employs down sampling and piecewise sampling to transform the original series into feature vectors that focus on global information and local information respectively. Then, TVA block uses temporal attention and variable attention to mine these feature vectors from different dimensions and extract key information. Finally, based on a parallel structure, DSformer uses multiple TVA blocks to mine and integrate different features obtained from DS blocks respectively. The integrated feature information is passed to the generative decoder based on a multi-layer perceptron to realize multivariate time series long-term prediction. Experimental results on nine real-world datasets show that DSformer can outperform eight existing baselines.

### **Federated News Recommendation with Fine-grained Interpolation and Dynamic Clustering**

Sanshi Lei Yu, Qi Liu, Fei Wang, Yang Yu, Enhong Chen

Researchers have successfully adapted the privacy-preserving Federated Learning (FL) to news recommendation tasks to better protect users' privacy, although typically at the cost of performance degradation due to the data heterogeneity issue. To address this issue, Personalized Federated Learning (PFL) has emerged, among which model interpolation is a promising approach that interpolates the local personalized models with the global model. However, the existing model interpolation method may not work well for news recommendation tasks for some reasons. First, it neglects the fine-grained personalization needs at both the temporal and spatial levels in news recommendation tasks. Second, due to the cold-user problem in real-world news recommendation tasks, the local personalized models may perform poorly, thus limiting the performance gain from model interpolation. To this end, we propose FINDING (Federated News Recommendation with Fine-grained Interpolation and Dynamic Clustering), a novel personalized federated learning framework based on model interpolation. Specifically, we first propose the fine-grained model interpolation strategy which interpolates the local personalized models with the global model in a time-aware and layer-aware way. Then, to address the cold-user problem in news recommendation tasks, we adopt the group-level personalization approach where users are dynamically clustered into groups and the group-level personalized models are used for interpolation. Extensive experiments on two real-world datasets show that our method can effectively handle the above limitations of the current model interpolation method and alleviate the heterogeneity issue faced by traditional FL.

### **Causality-guided Graph Learning for Session-based Recommendation**

Dianer Yu, Qian Li, Hongzhi Yin, Guandong Xu

Session-based recommendation systems (SBRs) aim to capture user preferences over time by taking into account the sequential order of interactions within sessions. One promising approach within this domain is session graph-based recommendation, which leverages graph-based models to represent and analyze user sessions. However, current graph-based methods for SBRs mainly rely on attention or pooling mechanisms that are prone to exploiting shortcut paths and thus lead to suboptimal recommendations.

To address this issue, we propose Causality-guided Graph Learning for Session-based Recommendation (CGSR) that is capable of blocking shortcut paths on the session graph and exploring robust causal connections capturing users' true preferences. Specifically, by employing back-door adjustment of causality, we can generate a distilled causal session graph capturing causal relations among items. CGSR then performs high-order aggregation on the distilled graph, incorporating information from various edge types, to estimate the session preference of the user. This enables us to provide more accurate recommendations grounded in causality while offering fine-grained interaction explanations by highlighting influential items in the graph. Extensive experiments on three datasets show the superior performance of CGSR compared to state-of-the-art SBRs.

#### **MUSE: Multi-view Contrastive Learning for Heterophilic Graphs via Information Reconstruction**

Mengyi Yuan, Minjie Chen, Xiang Li

In recent years, self-supervised learning has emerged as a promising approach in addressing the issues of label dependency and poor generalization performance in traditional GNNs. However, existing self-supervised methods have limited effectiveness on heterophilic graphs, due to the homophily assumption that results in similar node representations for connected nodes. In this work, we propose a multi-view contrastive learning model for heterophilic graphs, namely, MUSE. Specifically, we construct two views to capture the information of the ego node and its neighborhood by GNNs enhanced with contrastive learning, respectively. Then we integrate the information from these two views to fuse the node representations. Fusion contrast is utilized to enhance the effectiveness of fused node representations. Further, considering that the influence of neighboring contextual information on information fusion may vary across different ego nodes, we employ an information fusion controller to model the diversity of node-neighborhood similarity at both the local and global levels. Finally, an alternating training scheme is adopted to ensure that unsupervised node representation learning and information fusion controller can mutually reinforce each other. We conduct extensive experiments to evaluate the performance of MUSE on 9 benchmark datasets. Our results show the effectiveness of MUSE on both node classification and clustering tasks. We provide our data and codes at <https://github.com/dcxr969/MUSE>.

#### **VILE: Block-Aware Visual Enhanced Document Retrieval**

Huaying Yuan, Zhicheng Dou, Yujia Zhou, Yu Guo, Ji-Rong Wen

Document retrieval has always been a crucial problem in Web search. Recent works leverage pre-trained language models to represent documents in dense vectors. However, these works focus on the textual content but ignore the appearance of web pages (e.g., the visual style, the layout, and the images), which are actually essential for information delivery. To alleviate this problem, we propose a new dense retrieval model, namely VILE, to incorporate visual features into document representations. However, because a web page is usually very large and contains diverse information, simply concatenating its textual and visual features may result in a cluttered multi-modal representation that lacks focus on the important parts of the page. We observe that web pages often have a structured content organization, comprising multiple blocks that convey different information. Motivated by the observation, we propose building a multi-modal document representation by aggregating the fine-grained multi-modal block representations, to enable a more comprehensive understanding of the page. Specifically, we first segment a web page into multiple blocks, then create multi-modal features for each block, allowing for more effective capture of its content and visual information. The representations of all blocks are then integrated into the final multi-modal page representation. VILE can better model the importance of different content regions, leading to a high-quality multi-modal representation. We collect screenshots and the corresponding layout information of some web pages in the MS MARCO Document Ranking dataset, resulting in a new multi-modal document retrieval dataset. Experimental results conducted on this dataset demonstrate that our model exhibits significant improvements over existing document retrieval models. Our code is available at <https://github.com/yhy-2000/VILE>.

#### **Manipulating Out-Domain Uncertainty Estimation in Deep Neural Networks via Targeted Clean-Label Poisoning**

Huimin Zeng, Zhenrui Yue, Yang Zhang, Lanyu Shang, Dong Wang

Robust out-domain uncertainty estimation has gained growing attention for its capacity of providing adversary-resistant uncertainty estimates on out-domain samples. However, existing work on robust uncertainty estimation mainly focuses on evasion attacks that happen during test time. The threat of poisoning attacks against uncertainty models is largely unexplored. Compared to evasion attacks, poisoning attacks do not necessarily modify test data, and therefore, would be more practical in real-world applications. In this work, we systematically investigate the robustness of state-of-the-art uncertainty estimation algorithms against data poisoning attacks, with the ultimate objective of developing robust uncertainty training methods. In particular, we focus on attacking the out-domain uncertainty estimation. Under the proposed attack, the training process of models is affected. A fake high-confidence region is established around the targeted out-domain sample, which originally would have been rejected by the model due to low confidence. More fatally, our attack is clean-label and targeted: it leaves the poisoned data with clean labels and attacks a specific targeted test sample without degrading the overall model performance. We evaluate the proposed attack on several image benchmark datasets and a real-world application of COVID-19 misinformation detection. The extensive experimental results on different tasks suggest that the state-of-the-art uncertainty estimation methods could be extremely vulnerable and easily corrupted by our proposed attack.

#### **Target-Oriented Maneuver Decision for Autonomous Vehicle: A Rule-Aided Reinforcement Learning Framework**

Ximu Zeng, Quanlin Yu, Shuncheng Liu, Yuyang Xia, Han Su, Kai Zheng

Autonomous driving systems (ADSs) have the potential to revolutionize transportation by improving traffic safety and efficiency. As the core component of ADSs, maneuver decision aims to make tactical decisions to accomplish road following, obstacle avoidance, and efficient driving. In this work, we consider a typical but rarely studied task, called Target-Lane-Entering (TLE), where an autonomous vehicle should enter a target lane before reaching an intersection to ensure a smooth transition to another road. For navigation-assisted autonomous driving, a maneuver decision module chooses the optimal timing to enter the target lane in each road section, thus avoiding rerouting and reducing travel time. To achieve the TLE task, we propose a rule-aided reinforcement learning framework, called LINE, which combines the advantages of RL-based policy and rule-based strategy, allowing the autonomous vehicle to make target-oriented maneuver decisions. Specifically, an RL-based policy with a hybrid reward function is able to make safe, efficient, and comfortable decisions while considering the factors of target lanes. Then a strategy of rule revision aims to help the policy learn from intervention and block the risk of missing target lanes. Extensive experiments based on the SUMO simulator confirm the effectiveness of our framework. The results show that LINE achieves state-of-the-art driving performance with over 95% task success rate.

#### **AKE-GNN: Effective Graph Learning with Adaptive Knowledge Exchange**

Liang Zeng, Jin Xu, Zijun Yao, Yanqiao Zhu, Jian Li

Graph Neural Networks (GNNs) have already been widely used in various graph mining tasks. However, recent works reveal that the learned weights (channels) in well-trained GNNs are highly redundant, which inevitably limits the performance of GNNs. Instead of removing these redundant channels for efficiency consideration, we aim to reactivate them to enlarge the representation capacity of GNNs for effective graph learning. In this paper, we propose to substitute these redundant channels with other informative channels to achieve this goal. We introduce a novel GNN learning framework named AKE-GNN, which performs the Adaptive Knowledge Exchange strategy among multiple graph views generated by graph augmentations. AKE-GNN first trains multiple GNNs each corresponding to one graph view to obtain informative channels. Then, AKE-GNN iteratively exchanges redundant channels in the weight parameter matrix of one GNN with informative channels of another GNN in a layer-wise manner. Additionally, existing GNNs can be seamlessly incorporated into our framework. AKE-GNN achieves superior performance compared with various baselines across a suite of experiments on node classification, link prediction, and graph classification. In particular, we conduct a series of experiments on 15 public benchmark datasets, 8 popular GNN models, and 3 graph tasks and show that AKE-GNN consistently outperforms existing popular GNN models and even their ensembles. Extensive ablation studies and analyses on knowledge exchange methods validate the effectiveness of AKE-GNN.

#### **DYANE: DYnamic Attributed Node roIEs Generative Model**

Giselle Zeno, Jennifer Neville

Recent work has shown that modeling higher-order structures, such as motifs or graphlets, can capture the complex network structure and dynamics in a variety of graph domains (e.g., social sciences, biology, chemistry). However, many dynamic networks are not only rich in structure, but also in content information. For example, an academic citation network has content such as the title and abstracts of the papers. Currently, there is a lack of generative models for dynamic networks that also generate content. To address this gap, in this work we propose DYnamic Attributed Node roIEs (DYANE)-a generative

model that (i) captures network structure dynamics through temporal motifs, and (ii) extends the structural roles of nodes in motifs (e.g., a node acting as a hub in a wedge) to roles that generate content embeddings. We evaluate DYANE on real-world networks against other dynamic graph generative model baselines. DYANE outperforms the baselines in graph structure and node behavior, improving the KS score for graph metrics by 21-31% and node metrics by 17-27% on average, and produces content embeddings similar to the observed network. We also derive a methodology to evaluate the content embeddings generated by nodes, taking into account keywords extracted from the content (as topic representations), and using distance metrics.

### **MemoNet: Memorizing All Cross Features' Representations Efficiently via Multi-Hash Codebook Network for CTR Prediction**

Pengtao Zhang, Junlin Zhang

New findings in natural language processing (NLP) demonstrate that the strong memorization capability contributes a lot to the success of Large Language Models (LLM). This inspires us to explicitly bring an independent memory mechanism into CTR ranking model to learn and memorize cross features' representations. In this paper, we propose multi-Hash Codebook Network (HCNet) as the memory mechanism for efficiently learning and memorizing representations of cross features in CTR tasks. HCNet uses a multi-hash codebook as the main memory place and the whole memory procedure consists of three phases: multi-hash addressing, memory restoring, and feature shrinking. We also propose a new CTR model named MemoNet which combines HCNet with a DNN backbone. Extensive experimental results on three public datasets and online test show that MemoNet reaches superior performance over state-of-the-art approaches. Besides, MemoNet shows scaling law of large language model in NLP, which means we can enlarge the size of the codebook in HCNet to sustainably obtain performance gains. Our work demonstrates the importance and feasibility of learning and memorizing representations of cross features, which sheds light on a new promising research direction. The source code is in <https://github.com/ptzhangAlg/RecAlg>.

### **TriD-MAE: A Generic Pre-trained Model for Multivariate Time Series with Missing Values**

Kai Zhang, Chao Li, Qinmin Yang

Multivariate time series(MTS) is a universal data type related to various real-world applications. Data imputation methods are widely used in MTS applications to deal with the frequent data missing problem. However, these methods inevitably introduce biased imputation and training-redundancy problems in downstream training. To address these challenges, we propose TriD-MAE, a generic pre-trained model for MTS data with missing values. Firstly, we introduce TriD-TCN, an end-to-end module based on TCN that effectively extracts temporal features by integrating dynamic kernel mechanisms and a time-flipping trick. Building upon that, we designed an MAE-based pre-trained model as the precursor of specialized downstream models. Our model cooperates with a dynamic positional embedding mechanism to represent the missing information and generate transferable representation through our proposed encoder units. The overall mixed data feed-in strategy and weighted loss function are established to ensure adequate training of the whole model. Comparative experiment results in time series prediction and classification manifest that our TriD-MAE model outperforms the other state-of-the-art methods within six real-world datasets. Moreover, ablation and interpretability experiments are delivered to verify the validity of TriD-MAE's

### **RDGSL: Dynamic Graph Representation Learning with Structure Learning**

Siwei Zhang, Yun Xiong, Yao Zhang, Yiheng Sun, Xi Chen, Yizhu Jiao, Yangyong Zhu

Temporal Graph Networks (TGNs) have shown remarkable performance in learning representation for continuous-time dynamic graphs. However, real-world dynamic graphs typically contain diverse and intricate noise. Noise can significantly degrade the quality of representation generation, impeding the effectiveness of TGNs in downstream tasks. Though structure learning is widely applied to mitigate noise in static graphs, its adaptation to dynamic graph settings poses two significant challenges. i) Noise dynamics. Existing structure learning methods are ill-equipped to address the temporal aspect of noise, hampering their effectiveness in such dynamic and ever-changing noise patterns. ii) More severe noise. Noise may be introduced along with multiple interactions between two nodes, leading to the re-pollution of these nodes and consequently causing more severe noise compared to static graphs.

In this paper, we present RDGSL, a representation learning method in continuous-time dynamic graphs. Meanwhile, we propose dynamic graph structure learning, a novel supervisory signal that empowers RDGSL with the ability to effectively combat noise in dynamic graphs. To address the noise dynamics issue, we introduce the Dynamic Graph Filter, where we innovatively propose a dynamic noise function that dynamically captures both current and historical noise, enabling us to

assess the temporal aspect of noise and generate a denoised graph. We further propose the Temporal Embedding Learner to tackle the challenge of more severe noise, which utilizes an attention mechanism to selectively turn a blind eye to noisy edges and hence focus on normal edges, enhancing the expressiveness for representation generation that remains resilient to noise. Our method demonstrates robustness towards downstream tasks, resulting in up to 5.1% absolute AUC improvement in evolving classification versus the second-best baseline.

#### **CoSaR: Combating Label Noise Using Collaborative Sample Selection and Adversarial Regularization**

Xiaobo Zhang, Yutao Liu, Hao Wang, Wei Wang, Panpan Ni, Ji Zhang

Learning with noisy labels is nontrivial for deep learning models. Sample selection is a widely investigated research topic for handling noisy labels. However, most existing methods face challenges such as imprecise selection, a lack of global selection capabilities, and the need for tedious hyperparameter tuning. In this paper, we propose CoSaR (Collaborative Selection and adversarial Regularization), a twin-networks based model that performs globally adaptive sample selection to tackle label noise. Specifically, the collaborative selection estimates the average distribution distances between predictions and generation labels through the collaboration of two networks to address the bias of the average distribution distances and the manual tuning of hyperparameters. Adversarial regularization is integrated into CoSaR to restrict the network's tendency to fit and memorize noisy labels, thereby enhancing its collaborative selection capability. In addition, we employ a label smoothing regularization and two types of data augmentation to enhance the robustness of the model further. Extensive experiments on both synthetic and real-world noisy datasets demonstrate that the proposed model outperforms baseline methods remarkably, with an accuracy improvement ranging between +0.56% and +15.14%.

#### **PromptST: Prompt-Enhanced Spatio-Temporal Multi-Attribute Prediction**

Zijian Zhang, Xiangyu Zhao, Qidong Liu, Chunxu Zhang, Qian Ma, Wanyu Wang, Hongwei Zhao, Yiqi Wang, Zitao Liu

In the era of information explosion, spatio-temporal data mining serves as a critical part of urban management. Considering the various fields demanding attention, e.g., traffic state, human activity, and social event, predicting multiple spatio-temporal attributes simultaneously can alleviate regulatory pressure and foster smart city construction. However, current research can not handle the spatio-temporal multi-attribute prediction well due to the complex relationships between diverse attributes. The key challenge lies in how to address the common spatio-temporal patterns while tackling their distinctions. In this paper, we propose an effective solution for spatio-temporal multi-attribute prediction, PromptST. We devise a spatio-temporal transformer and a parameter-sharing training scheme to address the common knowledge among different spatio-temporal attributes. Then, we elaborate a spatio-temporal prompt tuning strategy to fit the specific attributes in a lightweight manner. Through the pretrain and prompt tuning phases, our PromptST is able to enhance the specific spatio-temporal characteristic capture by prompting the backbone model to fit the specific target attribute while maintaining the learned common knowledge. Extensive experiments on real-world datasets verify that our PromptST attains state-of-the-art performance. Furthermore, we also prove PromptST owns good transferability on unseen spatio-temporal attributes, which brings promising application potential in urban computing. The implementation code is available to ease reproducibility.

#### **Out of the Box Thinking: Improving Customer Lifetime Value Modelling via Expert Routing and Game Whale Detection**

Shijie Zhang, Xin Yan, Xuejiao Yang, Binfeng Jia, Shuangyang Wang

Customer lifetime value (LTV) prediction is essential for mobile game publishers trying to optimize the advertising investment for each user acquisition based on the estimated worth. In mobile games, deploying microtransactions is a simple yet effective monetization strategy, which attracts a tiny group of game whales who splurge on in-game purchases. The presence of such game whales may impede the practicality of existing LTV prediction models, since game whales' purchase behaviours always exhibit varied distribution from general users. Consequently, identifying game whales can open up new opportunities to improve the accuracy of LTV prediction models. However, little attention has been paid to applying game whale detection in LTV prediction, and existing works are mainly specialized for the long-term LTV prediction with the assumption that the high-quality user features are available, which is not applicable in the UA stage. In this paper, we propose ExpLTV, a novel multi-task framework to perform LTV prediction and game whale detection in a unified way. In ExpLTV, we first innovatively design a deep neural network-based game whale detector that can not only infer the intrinsic order in accordance with monetary value, but also precisely identify high spenders (i.e., game whales) and low spenders. Then, by treating the game whale detector as a gating network to decide the different mixture patterns of LTV experts assembling, we can thoroughly leverage the shared information and scenario-specific information (i.e., game whales modelling and low spenders modelling). Finally, instead of separately designing a purchase rate estimator for two tasks, we



design a shared estimator that can preserve the inner task relationships. The superiority of ExpLTV in terms of its LTV prediction and game whale detection effectiveness is further validated via extensive experiments on three industrial datasets.

### **iLoRE: Dynamic Graph Representation with Instant Long-term Modeling and Re-occurrence Preservation**

Siwei Zhang, Yun Xiong, Yao Zhang, Xixi Wu, Yiheng Sun, Jiawei Zhang

Continuous-time dynamic graph modeling is a crucial task for many real-world applications, such as financial risk management and fraud detection. Though existing dynamic graph modeling methods have achieved satisfactory results, they still suffer from three key limitations, hindering their scalability and further applicability. i) Indiscriminate updating. For incoming edges, existing methods would indiscriminately deal with them, which may lead to more time consumption and unexpected noisy information. ii) Ineffective node-wise long-term modeling. They heavily rely on recurrent neural networks (RNNs) as a backbone, which has been demonstrated to be incapable of fully capturing node-wise long-term dependencies in event sequences. iii) Neglect of re-occurrence patterns. Dynamic graphs involve the repeated occurrence of neighbors that indicates their importance, which is disappointingly neglected by existing methods.

In this paper, we present iLoRE, a novel dynamic graph modeling method with instant node-wise Long-term modeling and Re-occurrence preservation. To overcome the indiscriminate updating issue, we introduce the Adaptive Short-term Updater module that will automatically discard the useless or noisy edges, ensuring iLoRE's effectiveness and instant ability. We further propose the Long-term Updater to realize more effective node-wise long-term modeling, where we innovatively propose the Identity Attention mechanism to empower a Transformer-based updater, bypassing the limited effectiveness of typical RNN-dominated designs. Finally, the crucial re-occurrence patterns are also encoded into a graph module for informative representation learning, which will further improve the expressiveness of our method. Our experimental results on real-world datasets demonstrate the effectiveness of our iLoRE for dynamic graph modeling.

### **No Length Left Behind: Enhancing Knowledge Tracing for Modeling Sequences of Excessive or Insufficient Lengths**

Moyu Zhang, Xinning Zhu, Chunhong Zhang, Feng Pan, Wenchen Qian, Hui Zhao

Knowledge tracing (KT) aims to predict students' responses to practices based on their historical question-answering behaviors. However, most current KT methods focus on improving overall AUC, leaving ample room for optimization in modeling sequences of excessive or insufficient lengths. As sequences get longer, computational costs will increase exponentially. Therefore, KT methods usually truncate sequences to an acceptable length, which makes it difficult for models on online service systems to capture complete historical practice behaviors of students with too long sequences. Conversely, modeling students with short practice sequences using most KT methods may result in overfitting due to limited observation samples. To address the above limitations, we propose a model called Sequence-Flexible Knowledge Tracing (SFKT). Specifically, to flexibly handle long sequences, SFKT introduces a total-term encoder to effectively model complete historical practice behaviors of students at an affordable computational cost. Additionally, to improve the prediction accuracy of students with short practice sequences, we introduce a contrastive learning task and data augmentation schema to improve the generality of modeling short sequences by constructing more learning objectives. Extensive experimental results show that SFKT achieves significant improvements over multiple benchmarks, demonstrating the value of exploring the modeling of sequences of excessive or insufficient lengths. Our code is available at <https://github.com/zmy-9/SFKT>.

### **Counterfactual Monotonic Knowledge Tracing for Assessing Students' Dynamic Mastery of Knowledge Concepts**

Moyu Zhang, Xinning Zhu, Chunhong Zhang, Wenchen Qian, Feng Pan, Hui Zhao

As the core of the Knowledge Tracking (KT) task, assessing students' dynamic mastery of knowledge concepts is crucial for both offline teaching and online educational applications. Since students' mastery of knowledge concepts is often unlabeled, existing KT methods focus on predicting students' responses to practices. However, purely predicting student responses without imposing specific constraints on hidden concept mastery values does not guarantee the accuracy of these intermediate values as concept mastery values. To address this issue, we propose a principled approach called Counterfactual Monotonic Knowledge Tracing (CMKT), which builds on the implicit paradigm described above by using a counterfactual assumption to constrain the evolution of students' mastery of knowledge concepts. Specifically, CMKT first assesses students' knowledge concept mastery value based on their historical practice sequences. Then, CMKT sets the answer of the most recent practice as the opposite of the actual answer and, based on this counterfactual answer, assesses the student's corresponding counterfactual knowledge mastery value. During the model training process, CMKT constrains the update of the student's knowledge states by ensuring that the two types of knowledge mastery values of students satisfy

a fundamental educational theory, the monotonicity theory, to provide specific semantics for the assessed mastery values by the model. Finally, extensive experiments on five datasets demonstrate the superiority of CMKT over baseline models.

### **FATA-Trans: Field And Time-Aware Transformer for Sequential Tabular Data**

Dongyu Zhang, Liang Wang, Xin Dai, Shubham Jain, Junpeng Wang, Yujie Fan, Chin-Chia Michael Yeh, Yan Zheng, Zhongfang Zhuang, Wei Zhang

Sequential tabular data is one of the most commonly used data types in real-world applications. Different from conventional tabular data, where rows in a table are independent, sequential tabular data contains rich contextual and sequential information, where some fields are dynamically changing over time and others are static. Existing transformer-based approaches analyzing sequential tabular data overlook the differences between dynamic and static fields by replicating and filling static fields into each record, and ignore temporal information between rows, which leads to three major disadvantages: (1) computational overhead, (2) artificially simplified data for masked language modeling pre-training task that may yield less meaningful representations, and (3) disregarding the temporal behavioral patterns implied by time intervals. In this work, we propose FATA-Trans, a model with two field transformers for modeling sequential tabular data, where each processes static and dynamic field information separately. FATA-Trans is field- and time-aware for sequential tabular data. The field-type embedding in the method enables FATA-Trans to capture differences between static and dynamic fields. The time-aware position embedding exploits both order and time interval information between rows, which helps the model detect underlying temporal behavior in a sequence. Our experiments on three benchmark datasets demonstrate that the learned representations from FATA-Trans consistently outperform state-of-the-art solutions in the downstream tasks. We also present visualization studies to highlight the insights captured by the learned representations, enhancing our understanding of the underlying data. Our codes are available at <https://github.com/zdy93/FATA-Trans>.

### **Non-IID always Bad? Semi-Supervised Heterogeneous Federated Learning with Local Knowledge Enhancement**

Chao Zhang, Fangzhao Wu, Jingwei Yi, Derong Xu, Yang Yu, Jindong Wang, Yidong Wang, Tong Xu, Xing Xie, Enhong Chen

Federated learning (FL) is important for privacy-preserving services by training models without collecting raw user data. Most FL algorithms assume all data is annotated, which is impractical due to the high cost of labeling data in real applications. To alleviate the reliance on labeled data, semi-supervised federated learning (SSFL) has been proposed to utilize unlabeled data on clients to improve model performance. However, most existing methods either have privacy issues which share models trained on other clients, or generate pseudo-labels for unlabeled local datasets with the global model, which is usually biased towards the global data distribution. The latter may lead to sub-optimal accuracy of pseudo-labels, due to the gap between the local data distribution and the global model, especially in non-IID settings. In this paper, we propose a semi-supervised heterogeneous federated learning method with local knowledge enhancement, called FedLoKe, which aims to train an accurate global model from both labeled and unlabeled local data with non-IID distributions. Specifically, in FedLoKe, the server maintains a global model to capture global data distribution, and each client learns a local model to capture local data distribution. Since the distribution captured by the local model is aligned with the local data distribution, we utilize it to generate high-accuracy pseudo-labels of the unlabeled dataset for global model training. To prevent the local model from severely overfitting the small number of local labeled data, we further use the exponential moving average and apply the global model to generate pseudo-labels for local modeling training. Experiments on four datasets show the effectiveness of FedLoKe. Our code is available at: <https://github.com/zcfinal/FedLoKe>.

### **Mutual Information-Driven Multi-View Clustering**

Lei Zhang, Lele Fu, Tong Wang, Chuan Chen, Chuanfu Zhang

In deep multi-view clustering, three intractable problems are posed ahead of researchers, namely, the complementarity exploration problem, the information preservation problem, and the cluster structure discovery problem. In this paper, we consider the deep multi-view clustering from the perspective of mutual information (MI), and attempt to address the three important concerns with a Mutual Information-Driven Multi-View Clustering (MIMC) method, which extracts the common and view-specific information hidden in multi-view data and constructs a clustering-oriented comprehensive representation. Specifically, three constraints based on MI are devised in response to three issues. Correspondingly, we minimize the MI between the common representation and view-specific representations to exploit the inter-view complementary information. Further, we maximize the MI between the refined data representations and original data representations to preserve the principal information. Moreover, to learn a clustering-friendly comprehensive representation, the MI between the

comprehensive embedding space and cluster structure is maximized. Finally, we conduct extensive experiments on six benchmark datasets, and the experimental results indicate that the proposed MIMC outperforms other clustering methods.

### **Closed-form Machine Unlearning for Matrix Factorization**

Shuijing Zhang, Jian Lou, Li Xiong, Xiaoyu Zhang, Jing Liu

Matrix factorization (MF) is a fundamental model in data mining and machine learning, which finds wide applications in diverse application areas, including recommendation systems with user-item rating matrices, phenotype extraction from electronic health records, and spatial-temporal data analysis for check-in records. The "right to be forgotten" has become an indispensable privacy consideration due to the widely enforced data protection regulations, which allow personal users having contributed their data for model training to revoke their data through a data deletion request. Consequently, it gives rise to the emerging task of machine unlearning for the MF model, which removes the influence of the matrix rows/columns from the trained MF factors upon receiving the deletion requests from the data owners of these rows/columns. The central goal is to effectively remove the influence of the rows/columns to be forgotten, while avoiding the computationally prohibitive baseline approach of retraining from scratch. Existing machine unlearning methods are either designed for single-variable models and not compatible with MF that has two factors as coupled model variables, or require alternative updates that are not efficient enough. In this paper, we propose a closed-form machine unlearning method. In particular, we explicitly capture the implicit dependency between the two factors, which yields the total Hessian-based Newton step as the closed-form unlearning update. In addition, we further introduce a series of efficiency-enhancement strategies by exploiting the structural properties of the total Hessian. Extensive experiments on five real-world datasets from three application areas as well as synthetic datasets validate the efficiency, effectiveness, and utility of the proposed method.

### **Time-aware Graph Structure Learning via Sequence Prediction on Temporal Graphs**

Haozhen Zhang, Xueting Han, Xi Xiao, Jing Bai

Temporal Graph Learning, which aims to model the time-evolving nature of graphs, has gained increasing attention and achieved remarkable performance recently. However, in reality, graph structures are often incomplete and noisy, which hinders temporal graph networks (TGNs) from learning informative representations. Graph contrastive learning uses data augmentation to generate plausible variations of existing data and learn robust representations. However, rule-based augmentation approaches may be suboptimal as they lack learnability and fail to leverage rich information from downstream tasks. To address these issues, we propose a Time-aware Graph Structure Learning (TGSL) approach via sequence prediction on temporal graphs, which learns better graph structures for downstream tasks through adding potential temporal edges. In particular, it predicts time-aware context embedding based on previously observed interactions and uses the Gumble-Top-K to select the closest candidate edges to this context embedding. Additionally, several candidate sampling strategies are proposed to ensure both efficiency and diversity. Furthermore, we jointly learn the graph structure and TGNs in an end-to-end manner and perform inference on the refined graph. Extensive experiments on temporal link prediction benchmarks demonstrate that TGSL yields significant gains for the popular TGNs such as TGAT and GraphMixer, and it outperforms other contrastive learning methods on temporal graphs. We release the code at <https://github.com/ViktorAxelsen/TGSL>.

### **Mask- and Contrast-Enhanced Spatio-Temporal Learning for Urban Flow Prediction**

Xu Zhang, Yongshun Gong, Xinxin Zhang, Xiaoming Wu, Chengqi Zhang, Xiangjun Dong

As a critical mission of intelligent transportation systems, urban flow prediction (UFP) benefits in many city services including trip planning, congestion control, and public safety. Despite the achievements of previous studies, limited efforts have been observed on simultaneous investigation of the heterogeneity in both space and time aspects. That is, regional correlations would be variable at different timestamps. In this paper, we propose a spatio-temporal learning framework with mask and contrast enhancements to capture spatio-temporal variabilities among city regions. We devise a mask-enhanced pre-training task to learn latent correlations across the spatial and temporal dimensions, and then a graph-based method is developed to extract the significance of regions by using the inter-regional attention weights. To further acquire contrastive correlations of regions, we elaborate a pre-trained contrastive learning task with the global-local cross-attention mechanism. Thereafter, two well-trained encoders have strong capability to capture latent spatio-temporal representations for the flow forecasting with time-varying. Extensive experiments conducted on real-world urban flow datasets demonstrate that our method compares favorably with other state-of-the-art models.

### **A Co-training Approach for Noisy Time Series Learning**

Weiqi Zhang, Jianfeng Zhang, Jia Li, Fugee Tsung

In this work, we focus on robust time series representation learning. Our assumption is that real-world time series is noisy and complementary information from different views of the same time series plays an important role while analyzing noisy input. Based on this, we create two views for the input time series through two different encoders. We conduct co-training based contrastive learning iteratively to learn the encoders. Our experiments demonstrate that this co-training approach leads to a significant improvement in performance. Especially, by leveraging the complementary information from different views, our proposed TS-CoT method can mitigate the impact of data noise and corruption. Empirical evaluations on four time series benchmarks in unsupervised and semi-supervised settings reveal that TS-CoT outperforms existing methods. Furthermore, the representations learned by TS-CoT can transfer well to downstream tasks through fine-tuning<sup>1</sup>.

#### [Task Relation Distillation and Prototypical Pseudo Label for Incremental Named Entity Recognition](#)

Duzhen Zhang, Hongliu Li, Wei Cong, Rongtao Xu, Jiahua Dong, Xiuyi Chen

Incremental Named Entity Recognition (INER) involves the sequential learning of new entity types without accessing the training data of previously learned types. However, INER faces the challenge of catastrophic forgetting specific for incremental learning, further aggravated by background shift (i.e., old and future entity types are labeled as the non-entity type in the current task). To address these challenges, we propose a method called task Relation Distillation and Prototypical pseudo label (RDP) for INER. Specifically, to tackle catastrophic forgetting, we introduce a task relation distillation scheme that serves two purposes: 1) ensuring inter-task semantic consistency across different incremental learning tasks by minimizing inter-task relation distillation loss, and 2) enhancing the model's prediction confidence by minimizing intra-task self-entropy loss. Simultaneously, to mitigate background shift, we develop a prototypical pseudo label strategy that distinguishes old entity types from the current non-entity type using the old model. This strategy generates high-quality pseudo labels by measuring the distances between token embeddings and type-wise prototypes. We conducted extensive experiments on ten INER settings of three benchmark datasets (i.e., CoNLL2003, I2B2, and OntoNotes5). The results demonstrate that our method achieves significant improvements over the previous state-of-the-art methods, with an average increase of 6.08% in Micro F1 score and 7.71% in Macro F1 score.

#### [Communication-Efficient Decentralized Online Continuous DR-Submodular Maximization](#)

Qixin Zhang, Zengde Deng, Xiangru Jian, Zaiyi Chen, Haoyuan Hu, Yu Yang

Maximizing a monotone submodular function is a fundamental task in data mining, machine learning, economics, and statistics. In this paper, we present two communication-efficient decentralized online algorithms for the monotone continuous DR-submodular maximization problem, both of which reduce the number of per-function gradient evaluations and per-round communication complexity from  $T^{3/2}$  to 1. The first one, One-shot Decentralized Meta-Frank-Wolfe~(Mono-DMFW), achieves a  $(1-1/e)$ -regret bound of  $O(T^{4/5})$ . As far as we know, this is the first one-shot and projection-free decentralized online algorithm for monotone continuous DR-submodular maximization. Next, inspired by the non-oblivious boosting function[29], we propose the Decentralized Online Boosting Gradient Ascent (DOBGA) algorithm, which attains a  $(1-1/e)$ -regret of  $O(\sqrt{T})$ . To the best of our knowledge, this is the first result to obtain the optimal  $O(\sqrt{T})$  against a  $(1-1/e)$ -approximation with only one gradient inquiry for each local objective function per step. Finally, various experimental results confirm the effectiveness of the proposed methods.

#### [Unleashing the Power of Shared Label Structures for Human Activity Recognition](#)

Xiyuan Zhang, Ranak Roy Chowdhury, Jiayun Zhang, Dezhi Hong, Rajesh K. Gupta, Jingbo Shang

Current human activity recognition (HAR) techniques regard activity labels as integer class IDs without explicitly modeling the semantics of class labels. We observe that different activity names often have shared structures. For example, "open door" and "open fridge" both have "open" as the action; "kicking soccer ball" and "playing tennis ball" both have "ball" as the object. Such shared structures in label names can be translated to the similarity in sensory data and modeling common structures would help uncover knowledge across different activities, especially for activities with limited samples. In this paper, we propose SHARE, a HAR framework that takes into account shared structures of label names for different activities. To exploit the shared structures, SHARE comprises an encoder for extracting features from input sensory time series and a decoder for generating label names as a token sequence. We also propose three label augmentation techniques to help the model more effectively capture semantic structures across activities, including a basic token-level augmentation, and two enhanced embedding-level and sequence-level augmentations utilizing the capabilities of pre-trained models. SHARE outperforms

state-of-the-art HAR models in extensive experiments on seven HAR benchmark datasets. We also evaluate in few-shot learning and label imbalance settings and observe even more significant performance gap.

### [Temporal Convolutional Explorer Helps Understand 1D-CNN's Learning Behavior in Time Series Classification from Frequency Domain](#)

Junru Zhang, Lang Feng, Yang He, Yuhang Wu, Yabo Dong

While one-dimensional convolutional neural networks (1D-CNNs) have been empirically proven effective in time series classification tasks, we find that there remain undesirable outcomes that could arise in their application, motivating us to further investigate and understand their underlying mechanisms. In this work, we propose a Temporal Convolutional Explorer (TCE) to empirically explore the learning behavior of 1D-CNNs from the perspective of the frequency domain. Our TCE analysis highlights that deeper 1D-CNNs tend to distract the focus from the low-frequency components leading to the accuracy degradation phenomenon, and the disturbing convolution is the driving factor. Then, we leverage our findings to the practical application and propose a regulatory framework, which can easily be integrated into existing 1D-CNNs. It aims to rectify the suboptimal learning behavior by enabling the network to selectively bypass the specified disturbing convolutions. Finally, through comprehensive experiments on widely-used UCR, UEA, and UCI benchmarks, we demonstrate that 1) TCE's insight into 1D-CNN's learning behavior; 2) our regulatory framework enables state-of-the-art 1D-CNNs to get improved performances with less consumption of memory and computational overhead.

### [AspectMMKG: A Multi-modal Knowledge Graph with Aspect-aware Entities](#)

Jingdan Zhang, Jian Wang, Xiaodan Wang, Zhixu Li, Yanghua Xiao

Multi-modal knowledge graphs (MMKGs) combine different modal data (e.g., text and image) for a comprehensive understanding of entities. Despite the recent progress of large-scale MMKGs, existing MMKGs neglect the multi-aspect nature of entities, limiting the ability to comprehend entities from various perspectives. In this paper, we construct AspectMMKG, the first MMKG with aspect-related images by matching images to different entity aspects. Specifically, we collect aspect-related images from a knowledge base, and further extract aspect-related sentences from the knowledge base as queries to retrieve a large number of aspect-related images via an online image search engine. Finally, AspectMMKG contains 2,380 entities, 18,139 entity aspects, and 645,383 aspect-related images. We demonstrate the usability of AspectMMKG in entity aspect linking (EAL) downstream task and show that previous EAL models achieve a new state-of-the-art performance with the help of AspectMMKG. To facilitate the research on aspect-related MMKG, we further propose an aspect-related image retrieval (AIR) model, that aims to correct and expand aspect-related images in AspectMMKG. We train an AIR model to learn the relationship between entity image and entity aspect-related images by incorporating entity image, aspect, and aspect image information. Experimental results indicate that the AIR model could retrieve suitable images for a given entity w.r.t different aspects.

### [Towards Dynamic and Reliable Private Key Management for Hierarchical Access Structure in Decentralized Storage](#)

Yifang Zhang, Mingyue Wang, Yu Guo, Fangda Guo

With the widespread development of decentralized storage, it is increasingly popular for users to store their data to the decentralized database systems for the well-understood benefits of outsourced storage. To ensure the data privacy, systems commonly require users to securely keep their private keys. Thus, the secure storage of private keys is an important issue in these systems. However, existing key-management schemes commonly rely on a Trusted Third Party (TTP), which raises critical security concerns such as the single point of failure and Distributed Denial of Service (DDoS) attacks. In this paper, we propose HasDPSS, a secure and efficient blockchain-based key-management scheme for decentralized storage systems. It uses secret sharing, a lightweight cryptographic technique, to build the decentralized key-management scheme. Considering that the reliability of managing participants has inherent heterogeneity, we introduce the hierarchical access structure to achieve fine-grained key management. Meanwhile, to adapt the node churn of decentralized key management, HasDPSS enables a dynamic management committee to provide reliable services with a proactive refresh mechanism while protecting the integrity and security of private keys. In our design, we use the dimension switch method of polynomials in the evolving process to achieve the committee change of the hierarchical access structure. The reliability of participants is guaranteed by the customized commitment protocol and the immutable property of the blockchain. We thoroughly analyze security strengths and conduct extensive experiments to demonstrate the practicality of our design.

### [MLPST: MLP is All You Need for Spatio-Temporal Prediction](#)

Zijian Zhang, Ze Huang, Zhiwei Hu, Xiangyu Zhao, Wanyu Wang, Zitao Liu, Junbo Zhang, S. Joe Qin, Hongwei Zhao



Traffic prediction is a typical spatio-temporal data mining task and has great significance to the public transportation system. Considering the demand for its grand application, we recognize key factors for an ideal spatio-temporal prediction method: efficient, lightweight, and effective. However, the current deep model-based spatio-temporal prediction solutions generally own intricate architectures with cumbersome optimization, which can hardly meet these expectations. To accomplish the above goals, we propose an intuitive and novel framework, MLPST, a pure multi-layer perceptron architecture for traffic prediction. Specifically, we first capture spatial relationships from both local and global receptive fields. Then, temporal dependencies in different intervals are comprehensively considered. Through compact and swift MLP processing, MLPST can well capture the spatial and temporal dependencies while requiring only linear computational complexity, as well as model parameters that are more than an order of magnitude lower than baselines. Extensive experiments validated the superior effectiveness and efficiency of MLPST against advanced baselines, and among models with optimal accuracy, MLPST achieves the best time and space efficiency.

#### **Efficient Exact Minimum k-Core Search in Real-World Graphs**

Qifan Zhang, Shengxin Liu

The k-core, which refers to the induced subgraph with a minimum degree of at least k, is widely used in cohesive subgraph discovery and has various applications. However, the k-core in real-world graphs tends to be extremely large, which hinders its effectiveness in practical applications. This challenge has motivated researchers to explore a variant of the k-core problem known as the minimum k-core search problem. This problem has been proven to be NP-Hard, and most of the existing studies naturally either deal with approximate solutions or suffer from inefficiency in practice. In this paper, we focus on designing efficient exact algorithms for the minimum k-core search problem. In particular, we develop an iterative-based framework that decomposes an instance of the minimum k-core search problem into a list of problem instances on another well-structured graph pattern. Based on this framework, we propose an iterative-based branch-and-bound algorithm, namely IBB, with additional pruning and reduction techniques. We show that, with a n-vertex graph, IBB runs in  $cn \cdot nO(1)$  time for some  $c < 2$ , achieving better theoretical performance than the trivial bound of  $2n \cdot nO(1)$ . Finally, our experiments on real-world graphs demonstrate that IBB is up to three orders of magnitude faster than the state-of-the-art algorithms on real-world datasets.

#### **HST-GT: Heterogeneous Spatial-Temporal Graph Transformer for Delivery Time Estimation in Warehouse-Distribution Integration E-Commerce**

Xiaohui Zhao, Shuai Wang, Hai Wang, Tian He, Desheng Zhang, Guang Wang

Warehouse-distribution integration has been adopted by many e-commerce retailers (e.g., Amazon, TAOBAO, and JD) as an efficient business mode. In warehouse-distribution integration e-commerce, one of the most important problems is to estimate the full-link delivery time for better decision-making. Existing solutions for traditional warehouse-distribution separation mode are challenging to address this problem due to two unique features in the integration mode including (i) contextual influence caused by neighbor units in heterogeneous delivery networks, (ii) uncertain delivery time caused by the dynamic temporal data (e.g., online sales volume) and heterogeneity of delivery units. To incorporate these new factors, we propose Heterogeneous Spatial-Temporal Graph Transformer (HST-GT), a novel full-link delivery time estimation method under the warehouse-distribution integration mode, where we (i) develop heterogeneous graph transformers to capture hierarchical heterogeneous information; and (ii) design a set of spatial-temporal transformers based on heterogeneous features to fully exploit the correlation of spatial and temporal information. We extensively evaluate our method based on one-month real-world data consisting of hundreds of warehouses and sorting centers, and millions of historical orders collected from one of the largest e-commerce retailers in the world. Experimental results demonstrate that our method outperforms state-of-the-art baselines in various metrics.

#### **Geometric Graph Learning for Protein Mutation Effect Prediction**

Kangfei Zhao, Yu Rong, Biaobin Jiang, Jianheng Tang, Hengtong Zhang, Jeffrey Xu Yu, Peilin Zhao

Proteins govern a wide range of biological systems. Evaluating the changes in protein properties upon protein mutation is a fundamental application of protein design, where modeling the 3D protein structure is a principal task for AI-driven computational approaches. Existing deep learning (DL) approaches represent the protein structure as a 3D geometric graph and simplify the graph modeling to different degrees, thereby failing to capture the low-level atom patterns and high-level amino acid patterns simultaneously. In addition, limited training samples with ground truth labels and protein structures further restrict the effectiveness of DL approaches. In this paper, we propose a new graph learning framework, Hierarchical Graph Invariant Network (HGIN), a fine-grained and data-efficient graph neural encoder for encoding protein structures and

predicting the mutation effect on protein properties. For fine-grained modeling, HGIN hierarchically models the low-level interactions of atoms and the high-level interactions of amino acid residues by Graph Neural Networks. For data efficiency, HGIN preserves the invariant encoding for atom permutation and coordinate transformation, which is an intrinsic inductive bias of property prediction that bypasses data augmentations. We integrate HGIN into a Siamese network to predict the quantitative effect on protein properties upon mutations. Our approach outperforms 9 state-of-the-art approaches on 3 protein datasets. More inspiringly, when predicting the neutralizing ability of human antibodies against COVID-19 mutant viruses, HGIN achieves an absolute improvement of 0.23 regarding the Spearman coefficient.

### [Simulating Student Interactions with Two-stage Imitation Learning for Intelligent Educational Systems](#)

Guanhao Zhao, Zhenya Huang, Yan Zhuang, Jiayu Liu, Qi Liu, Zhiding Liu, Jinze Wu, Enhong Chen

The fundamental task of intelligent educational systems is to offer adaptive learning services to students, such as exercise recommendations and computerized adaptive testing. However, optimizing required models in these systems would always encounter the collection difficulty of high-quality interaction data in practice. Therefore, establishing a student simulator is of great value since it can generate valid interactions to help optimize models. Existing advances have achieved success but generally suffer from exposure bias and overlook long-term intentions. To tackle these problems, we propose a novel Direct-Adversarial Imitation Student Simulator (DAISim) by formulating it as a Markov Decision Process (MDP), which unifies the workflow of the simulator in training and generating to alleviate the exposure bias and single-step optimization problems. To construct the intentions underlying the complex student interactions, we first propose a direct imitation strategy to mimic the interactions with a simple reward function. Then, we propose an adversarial imitation strategy to learn a rational distribution with the reward given by a parameterized discriminator. Furthermore, we optimize the discriminator in adversarial imitation in a pairwise manner, and the theoretical analysis shows that the pairwise discriminator would improve the generation quality. We conduct extensive experiments on real-world datasets, where the results demonstrate that our DAISim can simulate high-quality student interactions whose distribution is close to real distribution and can promote several downstream services.

### [HTMapper: Bidirectional Head-Tail Mapping for Nested Named Entity Recognition](#)

Jin Zhao, Zhixu Li, Yanghua Xiao, Jiaqing Liang, Jingping Liu

Nested named entity recognition (Nested NER) aims to identify entities with nested structures from the given text, which is a fundamental task in Natural Language Processing. The region-based approach is the current mainstream approach, which first generates candidate spans and then classifies them into predefined categories. However, this method suffers from several drawbacks, including over-reliance on span representation, vulnerability to unbalanced category distribution, and inaccurate span boundary detection. To address these problems, we propose to model the nested NER problem into a head-tail mapping problem, namely, HTMapper, which detects head boundaries first and then models a conditional mapping from head to tail under a given category. Based on this mapping, we can find corresponding tails under different categories for each detected head by enumerating all entity categories. Our approach directly models the head boundary and tail boundary of entities, avoiding over-reliance on the span representation. Additionally, Our approach utilizes category information as an indicator signal to address the imbalance of category distribution during category prediction. Furthermore, our approach enhances the detection of span boundaries by capturing the correlation between head and tail boundaries. Extensive experiments on three nested NER datasets and two flat NER datasets demonstrate that our HTMapper achieves excellent performance with F1 scores of 89.09%, 88.30%, 81.57% on ACE2004, ACE2005, GENIA, and 94.26%, 91.40% on CoNLL03, OntoNotes, respectively.

### [Highly-Optimized Forgetting for Creating Signature-Based Views of Ontologies](#)

Yizheng Zhao

Uniform interpolation (UI) is a non-standard reasoning service that seeks to project an ontology down to its sub-signature --- given an ontology taking a certain signature, and a subset  $\Sigma$  of "relevant names" of that signature, compute a new ontology, called a uniform interpolant, that uses only the relevant names while preserving the semantics of the relevant names in the uniform interpolant. UI is of great potential importance since it may be used in a variety of applications where suitable views of ontologies need to be computed. However, this potential can only be fully realized if a highly optimized method for computing such views exists. Previous research has shown that computing uniform interpolants of ELH-ontologies is a computationally extremely hard problem --- a finite uniform interpolant does not always exist for ELH, and if it exists, then there exists one of at most triple exponential size in terms of the original ontology, and that, in the worst case, no shorter interpolant exists. Despite the inherent difficulty of the problem, in this paper, we present a highly optimized forgetting method for computing uniform interpolants of ELH-ontologies, and show however that, with good reduction and inference strategies,

such uniform interpolants can be efficiently computed. The method is an improvement of the one presented in our previous work. What sets it apart is its flexibility to treat concept names of different types differently, effectively cutting down on the inferences involved. This treatment is primarily driven by the polarities of the concept names within an ontology. A comprehensive evaluation with a prototypical implementation of the method shows >95% average success rates over two popular benchmark datasets and demonstrates a clear computational advantage over state-of-the-art systems.

### [Sequential Recommendation via an Adaptive Cross-domain Knowledge Decomposition](#)

Chuang Zhao, Xinyu Li, Ming He, Hongke Zhao, Jianping Fan

Cross-domain recommendation, as an intelligent machine to alleviate data sparsity and cold start problems, has attracted extensive attention from scholars. Existing cross-domain recommendation frameworks usually leverage overlapping entities for knowledge transfer, the most popular of which are information aggregation and consistency maintenance. Despite decent improvements, the neglect of dynamic perspectives, the presence of confounding factors, and the disparities in domain properties inevitably constrain model performance. In view of this, this paper proposes a sequential recommendation framework via adaptive cross-domain knowledge decomposition, namely ARISEN, which focuses on employing adaptive causal learning to improve recommendation performance. Specifically, in order to facilitate sequence transfer, we align the user's behaviour sequences in the source domain and target domain according to the timestamps, expecting to use the abundant semantics of the former to augment the information of the latter. Regarding confounding factor removal, we introduce the causal learning technique and promote it as an adaptive representation decomposition framework on the basis of instrumental variables. For the sake of alleviating the impact of domain disparities, this paper endeavors to employ two mutually orthogonal transformation matrices for information fusion. Extensive experiments and detailed analyzes on large industrial and public data sets demonstrate that our framework can achieve substantial improvements over state-of-the-art algorithms.

### [GCformer: An Efficient Solution for Accurate and Scalable Long-Term Multivariate Time Series Forecasting](#)

Yanjun Zhao, Ziqing` Ma, Tian Zhou, Mengni Ye, Liang Sun, Yi Qian

Transformer-based models have emerged as promising tools for time series forecasting. However, these models cannot make accurate prediction for long input time series. On the one hand, they failed to capture long-range dependency within time series data. On the other hand, the long input sequence usually leads to large model size and high time complexity. To address these limitations, we present GCformer, which combines a structured global convolutional branch for processing long input sequences with a local Transformer-based branch for capturing short, recent signals. A cohesive framework for a global convolution kernel has been introduced, utilizing three distinct parameterization methods. The selected structured convolutional kernel in the global branch has been specifically crafted with sublinear complexity, thereby allowing for the efficient and effective processing of lengthy and noisy input signals. Empirical studies on six benchmark datasets demonstrate that GCformer outperforms state-of-the-art methods, reducing MSE error in multivariate time series benchmarks by 4.38% and model parameters by 61.92%. In particular, the global convolutional branch can serve as a plug-in block to enhance the performance of other models, with an average improvement of 31.93%, including various recently published Transformer-based models. Our code is publicly available at <https://github.com/Yanjun-Zhao/GCformer>.

### [Unveiling the Role of Message Passing in Dual-Privacy Preservation on GNNs](#)

Tianyi Zhao, Hui Hu, Lu Cheng

Graph Neural Networks (GNNs) are powerful tools for learning representations on graphs, such as social networks. However, their vulnerability to privacy inference attacks restricts their practicality, especially in high-stake domains. To address this issue, privacy-preserving GNNs have been proposed, focusing on preserving node and/or link privacy. This work takes a step back and investigates how GNNs contribute to privacy leakage. Through theoretical analysis and simulations, we identify message passing under structural bias as the core component that allows GNNs to propagate and amplify privacy leakage. Building upon these findings, we propose a principled privacy-preserving GNN framework that effectively safeguards both node and link privacy, referred to as dual-privacy preservation. The framework comprises three major modules: a Sensitive Information Obfuscation Module that removes sensitive information from node embeddings, a Dynamic Structure Debiasing Module that dynamically corrects the structural bias, and an Adversarial Learning Module that optimizes the privacy-utility trade-off. Experimental results on four benchmark datasets validate the effectiveness of the proposed model in protecting both node and link privacy while preserving high utility for downstream tasks, such as node classification.

### [Task-Difficulty-Aware Meta-Learning with Adaptive Update Strategies for User Cold-Start Recommendation](#)

Xuhao Zhao, Yanmin Zhu, Chunyang Wang, Mengyuan Jing, Jiadi Yu, Feilong Tang

User cold-start recommendation is one of the most challenging problems that limit the effectiveness of recommender systems. Meta-learning-based methods are introduced to address this problem by learning initialization parameters for cold-start tasks. Recent studies attempt to enhance the initialization methods. They first represent each task by the cold-start user and interacted items. Then they distinguish tasks based on the task relevance to learn adaptive initialization. However, this manner is based on the assumption that user preferences can be reflected by the interacted items saliently, which is not always true in reality. In addition, we argue that previous approaches suffer from their adaptive framework (e.g., adaptive initialization), which reduces the adaptability in the process of transferring meta-knowledge to personalized RSs. In response to the issues, we propose a task-difficulty-aware meta-learning with adaptive update strategies (TDAS) for user cold-start recommendation. First, we design a task difficulty encoder, which can represent user preference salience, task relevance, and other task characteristics by modeling task difficulty information. Second, we adopt a novel framework with task-adaptive local update strategies by optimizing the initialization parameters with task-adaptive per-step and per-layer hyperparameters. Extensive experiments based on three real-world datasets demonstrate that our TDAS outperforms the state-of-the-art methods. The source code is available at <https://github.com/XuHao-bit/TDAS>.

#### **Decentralized Graph Neural Network for Privacy-Preserving Recommendation**

Xiaolin Zheng, Zhongyu Wang, Chaochao Chen, Jiashu Qian, Yao Yang

Building a graph neural network (GNN)-based recommender system without violating user privacy proves challenging. Existing methods can be divided into federated GNNs and decentralized GNNs. But both methods have undesirable effects, i.e., low communication efficiency and privacy leakage. This paper proposes DGREC, a novel decentralized GNN for privacy-preserving recommendations, where users can choose to publicize their interactions. It includes three stages, i.e., graph construction, local gradient calculation, and global gradient passing. The first stage builds a local inner-item hypergraph for each user and a global inter-user graph. The second stage models user preference and calculates gradients on each local device. The third stage designs a local differential privacy mechanism named secure gradient-sharing, which proves strong privacy-preserving of users' private data. We conduct extensive experiments on three public datasets to validate the consistent superiority of our framework.

#### **DiffUFlow: Robust Fine-grained Urban Flow Inference with Denoising Diffusion Model**

Yuhao Zheng, Lian Zhong, Senzhang Wang, Yu Yang, Weixi Gu, Junbo Zhang, Jianxin Wang

Inferring the fine-grained urban flows based on the coarse-grained flow observations is practically important to many smart city-related applications. However, the collected human/vehicle trajectory flows are usually rather unreliable, may contain various noise and sometimes are incomplete, thus posing great challenges to existing approaches. In this paper, we present a pioneering study on robust fine-grained urban flow inference with noisy and incomplete urban flow observations, and propose a denoising diffusion model named DiffUFlow to effectively address it. Specifically, we propose an improved reverse diffusion strategy. A spatial-temporal feature extraction network called STFormer and a semantic features extraction network called ELFletcher are also proposed. Then, we overlay the spatial-temporal feature map extracted by STFormer onto the coarse-grained flow map, serving as a conditional guidance for the reverse diffusion process. We further integrate the semantic features extracted by ELFletcher to cross-attention layers, enabling the comprehensive consideration of semantic information encompassing the entirety of urban data in fine-grained inference. Extensive experiments on two large real-world datasets validate the effectiveness of our method compared with the state-of-the-art baselines.

#### **FedPSE: Personalized Sparsification with Element-wise Aggregation for Federated Learning**

Longfei Zheng, Yingting Liu, Xiaolong Xu, Chaochao Chen, Yuzhou Tang, Lei Wang, Xiaolong Hu

Federated learning (FL) is a popular distributed machine learning framework in which clients aggregate models' parameters instead of sharing individual data. In FL, clients frequently communicate with the server under limited network bandwidth, raising the communication challenge. Multiple compression methods have been proposed to reduce the transmitted parameters. However, these techniques show that the federated performance degrades significantly with Non-IID (non-identically independently distributed) datasets. To address this issue, we propose an effective method called FedPSE, which solves the efficiency challenge of FL with heterogeneous data. FedPSE compresses the local updates on clients using Top-K sparsification and aggregates these updates on the server by element-wise aggregation. Then clients download the personalized sparse updates from the server to update their individual local models. We then theoretically analyze the convergence of FedPSE under the non-convex setting. Moreover, extensive experiments on four benchmark tasks

demonstrate that our FedPSE outperforms the state-of-the-art methods on Non-IID datasets in terms of efficiency and accuracy.

### **Assessing the Continuous Causal Responses of Typhoon-related Weather on Human Mobility: An Empirical Study in Japan**

Zhang Zhiwen, Hongjun Wang, Zipei Fan, Ryosuke Shibasaki, Xuan Song

To understand human mobility following the typhoon, analyzing the causal impact of extreme typhoon weather on human mobility is important for disaster emergency management. However, the unobserved confounders (e.g., the characteristic of each region) correlate with the strength of typhoon weather and also affect human mobility during typhoon, which may generate biased influences on the causal analysis process. Besides, these confounders may be time-varying following the dynamic movements of typhoon. In this work, we develop a neural network-based continuous causal effect estimation framework to mitigate the interference from (unobserved) confounders and assess the continuous causal responses of typhoon-related weather (treatment) on several types of human mobility (outcome) across different counties at any given period. To this end, we integrate the big data from two huge typhoons in Japan (i.e., Typhoon Faxai and Hagibis) and leverage multiple sources of covariates (i.e., residents' vigilance and basic mobility patterns) from different counties to learn the representations of time-varying confounders. The experimental results indicate the effectiveness of our proposed framework in capturing the confounders for quantifying the causal impact of extreme weather during the typhoon process, compared with several existing causal studies.

### **Personalized Location-Preference Learning for Federated Task Assignment in Spatial Crowdsourcing**

Xiaolong Zhong, Hao Miao, Dazhuo Qiu, Yan Zhao, Kai Zheng

With the proliferation of wireless and mobile devices, Spatial Crowdsourcing (SC) attracts increasing attention, where task assignment plays a critically important role. However, recent task assignment solutions in SC often assume that data is stored in a central station while ignoring the issue of privacy leakage. To enable decentralized training and privacy protection, we propose a federated task assignment framework with personalized location-preference learning, which performs efficient task assignment while keeping the data decentralized and private in each platform center (e.g., a delivery center of an SC company). The framework consists of two phases: personalized federated location-preference learning and task assignment. Specifically, in the first phase, we design a personalized location-preference learning model for each platform center by simultaneously considering the location information and data heterogeneity across platform centers. Based on workers' location preference, the task assignment phase aims to achieve effective and efficient task assignment by means of the Kuhn-Munkres (KM) algorithm and the newly proposed conditional degree-reduction algorithm. Extensive experiments on real-world data show the effectiveness of the proposed framework.

### **RLIFE: Remaining Lifespan Prediction for E-scooters**

Shuxin Zhong, William Yubeaton, Wenjun Lyu, Guang Wang, Desheng Zhang, Yu Yang

Shared electric scooters (e-scooters) have been increasingly popular because of their characteristics of convenience and eco-friendliness. Due to their shared nature and widespread usage, e-scooters usually have a short lifespan (e.g., two to five months[2]), which makes it important to predict the remaining lifespan accurately, ensuring timely replacements. While several studies have focused on the lifespan prediction of various systems, such as batteries and bridges, they present a two-fold drawback. Firstly, they require significant manual labor or additional sensor resources to ascertain the explicit status of the object, rendering them cost-ineffective. Secondly, these studies assume that future usage is similar as the historical usage. To solve these limitations, we aim at accurately predicting the remaining lifespan of e-scooters without extra cost, and its essence is to accurately represent its current status and anticipate its future usage. However, it is challenging because: i) lack of explicit rules for the e-scooters' status representation; and ii) e-scooters' future usage may significantly differ from their historical usage. In this paper, we design a framework called RLIFE, whose key insight is modeling user behaviors from trip transactions is of great importance in predicting the Remaining Lifespan of shared E-scooters. Specifically, we introduce an unsupervised contrastive learning component to learn the e-scooters' status representation over time considering degradation, where user preferences are served as a status reflector; We further design an LSTM-based recursive component to dynamically predict uncertain future usage, upon which we fuse the current status and predicted usage of the e-scooter for its remaining lifespan prediction. Extensive experiments are conducted on large-scale, real-world datasets collected from an e-scooter company. It shows that RLIFE improves the baselines by 35.67% and benefits from the learned user preferences and predicted future usage.



### **A Generalized Propensity Learning Framework for Unbiased Post-Click Conversion Rate Estimation**

Yuqing Zhou, Tianshu Feng, Mingrui Liu, Ziwei Zhu

This paper addresses the critical gap in the unbiased estimation of post-click conversion rate (CVR) in recommender systems. Existing CVR prediction methods, such as Inverse Propensity Score (IPS) and various Doubly Robust (DR) based estimators, overlook the impact of propensity estimation on the model bias and variance, thus leading to a debiasing performance gap. We propose a Generalized Propensity Learning (GPL) framework to directly minimize the bias and variance in CVR prediction models. The proposed method works as a complement to existing methods like IPS, DR, MRDR, and DRMSE to improve prediction performance by reducing their bias and variance. Extensive experiments on real-world datasets and semi-synthetic datasets demonstrate the significant performance promotion brought by our proposed method. Data and code can be found at: <https://github.com/yuqing-zhou/GPL>.

### **Contrastive Counterfactual Learning for Causality-aware Interpretable Recommender Systems**

Guanglin Zhou, Chengkai Huang, Xiaocong Chen, Xiwei Xu, Chen Wang, Liming Zhu, Lina Yao

The field of generating recommendations within the framework of causal inference has seen a recent surge. This approach enhances insights into the influence of recommendations on user behavior and helps in identifying the underlying factors. Existing research has often leveraged propensity scores to mitigate bias, albeit at the risk of introducing additional variance. Others have explored the use of unbiased data from randomized controlled trials, although this comes with assumptions that may prove challenging in practice. In this paper, we first present the causality-aware interpretation of recommendations and reveal how the underlying exposure mechanism can bias the maximum likelihood estimation (MLE) of observational feedback. Recognizing that confounders may be elusive, we propose a contrastive self-supervised learning to minimize exposure bias, employing inverse propensity scores and expanding the positive sample set. Building on this foundation, we present a novel contrastive counterfactual learning method (CCL) that incorporates three unique positive sampling strategies grounded in estimated exposure probability or random counterfactual samples. Through extensive experiments on two real-world datasets, we demonstrate that our CCL outperforms the state-of-the-art methods.

### **All about Sample-Size Calculations for A/B Testing: Novel Extensions & Practical Guide**

Jing Zhou, Jiannan Lu, Anas Shallah

While there exists a large amount of literature on the general challenges and best practices for trustworthy online A/B testing, there are limited studies on sample size estimation, which plays a crucial role in trustworthy and efficient A/B testing that ensures the resulting inference has a sufficient power and type I error control. For example, when sample size is underestimated, the statistical inference, even with the correct analysis methods, will not be able to detect the true significant improvement leading to misinformed and costly decisions. This paper addresses this fundamental gap by developing new sample size calculation methods for correlated data, as well as absolute vs. relative treatment effects, both ubiquitous in online experiments. Additionally, we address a practical question of the minimal observed difference that will be statistically significant and how it relates to average treatment effect and sample size calculation. All proposed methods are accompanied by mathematical proofs, illustrative examples, and simulations. We end by sharing some best practices on various practical topics on sample size calculation and experimental design.

### **Learning Node Abnormality with Weak Supervision**

Qinghai Zhou, Kaize Ding, Huan Liu, Hanghang Tong

Graph anomaly detection aims to identify the atypical substructures and has attracted an increasing amount of research attention due to its profound impacts in a variety of application domains, including social network analysis, security, finance, and many more. The lack of prior knowledge of the ground-truth anomaly has been a major obstacle in acquiring fine-grained annotations (e.g., anomalous nodes), therefore, a plethora of existing methods have been developed either with a limited number of node-level supervision or in an unsupervised manner. Nonetheless, annotations for coarse-grained graph elements (e.g., a suspicious group of nodes), which often require marginal human effort in terms of time and expertise, are comparatively easier to obtain. Therefore, it is appealing to investigate anomaly detection in a weakly-supervised setting and to establish the intrinsic relationship between annotations at different levels of granularity. In this paper, we tackle the challenging problem of weakly-supervised graph anomaly detection with coarse-grained supervision by (1) proposing a novel architecture of graph neural network with attention mechanism named WEDGE that can identify the critical node-level anomaly given a few labels of anomalous subgraphs, and (2) designing a novel objective with contrastive loss that facilitates node representation learning by enforcing distinctive representations between normal and abnormal graph elements.

Through extensive evaluations on real-world datasets, we corroborate the efficacy of our proposed method, improving AUC-ROC by up to 16.48% compared to the best competitor.

### **Attention Calibration for Transformer-based Sequential Recommendation**

Peilin Zhou, Qichen Ye, Yueqi Xie, Jingqi Gao, Shoujin Wang, Jae Boum Kim, Chenyu You, Sunghun Kim

Transformer-based sequential recommendation (SR) has been booming in recent years, with the self-attention mechanism as its key component. Self-attention has been widely believed to be able to effectively select those informative and relevant items from a sequence of interacted items for next-item prediction via learning larger attention weights for these items. However, this may not always be true in reality. Our empirical analysis of some representative Transformer-based SR models reveals that it is not uncommon for large attention weights to be assigned to less relevant items, which can result in inaccurate recommendations. Through further in-depth analysis, we find two factors that may contribute to such inaccurate assignment of attention weights: sub-optimal position encoding and noisy input. To this end, in this paper, we aim to address this significant yet challenging gap in existing works. To be specific, we propose a simple yet effective framework called Attention Calibration for Transformer-based Sequential Recommendation (AC-TSR). In AC-TSR, a novel spatial calibrator and adversarial calibrator are designed respectively to directly calibrate those incorrectly assigned attention weights. The former is devised to explicitly capture the spatial relationships (i.e., order and distance) among items for more precise calculation of attention weights. The latter aims to redistribute the attention weights based on each item's contribution to the next-item prediction. AC-TSR is readily adaptable and can be seamlessly integrated into various existing transformer-based SR models. Extensive experimental results on four benchmark real-world datasets demonstrate the superiority of our proposed AC-TSR via significant recommendation performance enhancements. The source code is available at <https://github.com/AIM-SE/AC-TSR>.

### **Privacy-Preserving Federated Learning via Disentanglement**

Wenjie Zhou, Piji Li, Zhaoyang Han, Xiaozhen Lu, Juan Li, Zhaochun Ren, Zhe Liu

The trade-off between privacy and accuracy presents a challenge for current federated learning (FL) frameworks, hindering their progress from theory to application. The main issues with existing FL frameworks stem from a lack of interpretability and targeted privacy protections. To cope with these, we proposed Disentangled Federated Learning for Privacy (DFLP) which employs disentanglement, one of interpretability techniques, in private FL frameworks. Since sensitive properties are client-specific in nature, our main idea is to turn this feature into a tool that strikes the balance between data privacy and FL model performance, enabling the sensitive attributes to be private. DFLP disentangles the client-specific and class-invariant attributes to mask the sensitive attributes precisely. To our knowledge, this is the first work that successfully integrates disentanglement and the nature of sensitive attributes to achieve privacy protection while ensuring high FL model performance. Extensive experiments validate that disentanglement is an effective method for accuracy-aware privacy protection in FL frameworks.

### **CANDY: A Causality-Driven Model for Hotel Dynamic Pricing**

Ruitao Zhu, Wendong Xiao, Yao Yu, Yizhi Yu, Zhenzhe Zheng, Ke Bu, Dong Li, Fan Wu

Broad adoption of online travel platforms (OTPs) has led to increasing focus on hotel dynamic pricing algorithms, which directly affect the revenue of platform and hotels. Existing approaches, which directly model the correlation between price and occupancy, have limitations in improving occupancy prediction accuracy while ensuring interpretability for dynamic pricing. Moreover, these methods struggle to address the significant data sparsity issue in hotel pricing scenarios. To overcome these limitations, we propose a novel Causality-driven Hotel Dynamic Pricing Model (CANDY) that captures the essential causal relationship between price and occupancy, enhancing occupancy prediction accuracy and interpretability for dynamic pricing. Specifically, we decompose confounders into three orthogonal groups of factors: characteristic factors, competitive factors, and temporal factors, and design submodules to capture the features of each dimension. To address the treatment bias and sample imbalance issues faced by existing causal inference methods in hotel pricing scenarios, we propose a novel data augmentation method based on the monotonic relationship between price and occupancy, and further design a multi-task learning framework tailored to multi-valued treatment scenarios, simultaneously alleviating the data sparsity issue. Both offline and online experiments demonstrate the effectiveness of CANDY in occupancy prediction and dynamic pricing. CANDY has been successfully deployed to provide price suggestion service at Fliggy, a leading OTP in China, serving thousands of hotel operators.

### **Improving Adversarial Transferability via Frequency-based Stationary Point Search**

Zhiyu Zhu, Huaming Chen, Jiayu Zhang, Xinyi Wang, Zhibo Jin, Qinghua Lu, Jun Shen, Kim-Kwang Raymond Choo

Deep neural networks (DNNs) have been shown vulnerable to interference from adversarial samples, leading to erroneous predictions. Investigating adversarial attacks can effectively improve the reliability as well as the performance of deep neural models in real-world applications. Since it is generally challenging to infer the parameters in black-box models, high transferability becomes an important factor for the success rate of an attack method. Recently, the Spectrum Simulation Attack method exhibits promising results based on the frequency domain. In light of SSA, we propose a novel attack approach in this paper, which achieves the best results among diverse state-of-the-art transferable adversarial attack methods. Our method aims to find a stationary point, which extends the ability to find multiple local optima with the optimal local attack effect. After finding the stationary point, a frequency-based search is employed to explore the best adversarial samples in the neighbouring space, ultimately determining the final adversarial direction. We compare our method against a variety of cutting-edge transferable adversarial methods. Extensive experiments validate that our method improves the attack success rate by 4.7% for conventionally trained models and 53.1% for adversarially trained models. Our code is available at <https://github.com/LMBTough/FSPS>

### **Scalable Neural Contextual Bandit for Recommender Systems**

Zheqing Zhu, Benjamin Van Roy

High-quality recommender systems ought to deliver both innovative and relevant content through effective and exploratory interactions with users. Yet, supervised learning-based neural networks, which form the backbone of many existing recommender systems, only leverage recognized user interests, falling short when it comes to efficiently uncovering unknown user preferences. While there has been some progress with neural contextual bandit algorithms towards enabling online exploration through neural networks, their onerous computational demands hinder widespread adoption in real-world recommender systems. In this work, we propose a scalable sample-efficient neural contextual bandit algorithm for recommender systems. To do this, we design an epistemic neural network architecture, Epistemic Neural Recommendation (ENR), that enables Thompson sampling at a large scale. In two distinct large-scale experiments with real-world tasks, ENR significantly boosts click-through rates and user ratings by at least 9% and 6% respectively compared to state-of-the-art neural contextual bandit algorithms. Furthermore, it achieves equivalent performance with at least 29% fewer user interactions compared to the best-performing baseline algorithm. Remarkably, while accomplishing these improvements, ENR demands orders of magnitude fewer computational resources than neural contextual bandit baseline algorithms.

### **HCL4QC: Incorporating Hierarchical Category Structures Into Contrastive Learning for E-commerce Query**

#### **Classification**

Lvxing Zhu, Kexin Zhang, Hao Chen, Chao Wei, Weiru Zhang, Haihong Tang, Xiu Li

Query classification plays a crucial role in e-commerce, where the goal is to assign user queries to appropriate categories within a hierarchical product category taxonomy. However, existing methods rely on a limited number of words from the category description and often neglect the hierarchical structure of the category tree, resulting in suboptimal category representations. To overcome these limitations, we propose a novel approach named hierarchical contrastive learning framework for query classification (HCL4QC), which leverages the hierarchical category tree structure to improve the performance of query classification. Specifically, HCL4QC is designed as a plugin module that consists of two innovative losses, namely local hierarchical contrastive loss (LHCL) and global hierarchical contrastive loss (GHCL). LHCL adjusts representations of categories according to their positional relationship in the hierarchical tree, while GHCL ensures the semantic consistency between the parent category and its child categories. Our proposed method can be adapted to any query classification tasks that involve a hierarchical category structure. We conduct experiments on two real-world datasets to demonstrate the superiority of our hierarchical contrastive learning. The results demonstrate significant improvements in the query classification task, particularly for long-tail categories with sparse supervised information.

### **FVW: Finding Valuable Weight on Deep Neural Network for Model Pruning**

Zhiyu Zhu, Huaming Chen, Zhibo Jin, Xinyi Wang, Jiayu Zhang, Minhui Xue, Qinghua Lu, Jun Shen, Kim-Kwang Raymond Choo

The rapid development of deep learning has demonstrated its potential for deployment in many intelligent service systems. However, some issues such as optimisation (e.g., how to reduce the deployment resources costs and further improve the detection speed), especially in scenarios where limited resources are available, remain challenging to address. In this paper, we aim to delve into the principles of deep neural networks, focusing on the importance of network neurons. The goal is to

identify the neurons that exert minimal impact on model performances, thereby aiding in the process of model pruning. In this work, we have thoroughly considered the deep learning model pruning process with and without fine-tuning step, ensuring the model performance consistency. To achieve our objectives, we propose a methodology that employs adversarial attack methods to explore deep neural network parameters. This approach is combined with an innovative attribution algorithm to analyse the level of network neurons involvement. In our experiments, our approach can effectively quantify the importance of network neuron. We extend the evaluation through comprehensive experiments conducted on a range of datasets, including CIFAR-10, CIFAR-100 and Caltech101. The results demonstrate that, our method have consistently achieved the state-of-the-art performance over many existing methods. We anticipate that this work will help to reduce the heavy training and inference cost of deep neural network models where a lightweight deep learning enhanced service and system is possible. The source code is open source at <https://github.com/LMBTough/FVW>.

#### **PCENet: Psychological Clues Exploration Network for Multimodal Personality Assessment**

Yangfu Zhu, Yuting Wei, Meiling Li, Tingting Zhang, Siqi Wei, Bin Wu

Multimodal personality assessment aims to identify and express human personality traits in videos. Existing methods primarily focus on multimodal fusion while ignoring the inherent psychological clues essential for this interdisciplinary task. Modality clues: personality traits are stable over time due to their genetic and environmental origins, resulting in stable personality traits in the multimodal data. Trait clues: multiple traits often co-occur with non-negligible correlations, which can collectively aid trait identification. To simultaneously capture the above psychological clues, we propose a novel Psychological Clues Exploration Network (PCENet) for multimodal personality assessment, which is a human-like judgment paradigm with more generalization capability. Specifically, we first devise a multimodal hierarchical disentanglement, which clearly aligns stable representations among different modalities and separates the mutability of each modality. Subsequently, a Transformer-backbone decoder equipped with modality-to-trait attention is exploited to adaptively generate a tailored representation for each trait with the guidance of trait semantics. The trait semantics are obtained by exploiting trait correlations through self-attention. Extensive experiments on the First Impression V2 dataset demonstrate that our PCENet outperforms the state-of-the-art methods for multimodal personality assessment.

#### **G-STO: Sequential Main Shopping Intention Detection via Graph-Regularized Stochastic Transformer**

Yuchen Zhuang, Xin Shen, Yan Zhao, Chaosheng Dong, Ming Wang, Jin Li, Chao Zhang

Sequential recommendation requires understanding the dynamic patterns of users' behaviors, contexts, and preferences from their historical interactions. While most research emphasizes item-level user-item interactions, they often overlook underlying shopping intentions, such as preferences for ballpoint pens or miniatures. Identifying these latent intentions is vital for enhancing shopping experiences on platforms like Amazon. Despite its significance, the area of main shopping intention detection remains under-investigated in the academic literature. To fill this gap, we introduce a graph-regularized stochastic Transformer approach, G-STO. It considers intentions as product sets and user preferences as intention composites, both modeled as stochastic Gaussian embeddings in latent space. We also employ a global intention relational graph as prior knowledge for regularization, ensuring related intentions are distributionally close. These regularized embeddings are then input into Transformer-based models to capture sequential intention transitions. On testing our model with three real-world datasets, it outperformed the baselines by 18.08% in Hit@1, 7.01% in Hit@10, and 6.11% in NDCG@10.

### **SESSION: Short Papers**

#### **DynED: Dynamic Ensemble Diversification in Data Stream Classification**

Soheil Abadifard, Sepehr Bakhshi, Sanaz Gheibuni, Fazli Can

Ensemble methods are commonly used in classification due to their remarkable performance. Achieving high accuracy in a data stream environment is a challenging task considering disruptive changes in the data distribution, also known as concept drift. A greater diversity of ensemble components is known to enhance prediction accuracy in such settings. Despite the diversity of components within an ensemble, not all contribute as expected to its overall performance. This necessitates a method for selecting components that exhibit high performance and diversity. We present a novel ensemble construction and maintenance approach based on MMR (Maximal Marginal Relevance) that dynamically combines the diversity and prediction accuracy of components during the process of structuring an ensemble. The experimental results on both four real and 11 synthetic datasets demonstrate that the proposed approach (DynED) provides a higher average mean accuracy compared to the five state-of-the-art baselines.

#### **Retrievability Bias Estimation Using Synthetically Generated Queries**

Amin Abolghasemi, Suzan Verberne, Arian Askari, Leif Azzopardi

Ranking with pre-trained language models (PLMs) has shown to be highly effective for various Information Retrieval tasks. Previous studies investigated the performance of these models in terms of effectiveness and efficiency. However, there is no prior work on evaluating PLM-based rankers in terms of their retrievability bias. In this paper, we evaluate the retrievability bias of PLM-based rankers with the use of synthetically generated queries. We compare the retrievability bias in two of the most common PLM-based rankers, a Bi-Encoder BERT ranker and a Cross-Encoder BERT re-ranker against BM25, which was found to be one of the least biased models in prior work. We conduct a series of experiments with which we explore the plausibility of using synthetic queries generated with a generative model, docT5query, in the evaluation of retrievability bias. Our experiments show promising results on the use of synthetically generated queries for the purpose of retrievability bias estimation. Moreover, we find that the estimated bias values resulting from synthetically generated queries are lower than the ones estimated with user-generated queries on the MS MARCO evaluation benchmark. This indicates that synthetically generated queries might cause less bias than user-generated queries and therefore, by using such queries in training PLM-based rankers, we might be able to reduce the retrievability bias in these models.

#### **Fine-Grained Socioeconomic Prediction from Satellite Images with Distributional Adjustment**

Donghyun Ahn, Minhyuk Song, Seungeon Lee, Yubin Choi, Jihee Kim, Sangyoon Park, Hyunjoon Yang, Meeyoung Cha

While measuring socioeconomic indicators is critical for local governments to make informed policy decisions, such measurements are often unavailable at fine-grained levels like municipality. This study employs deep learning-based predictions from satellite images to close the gap. We propose a method that assigns a socioeconomic score to each satellite image by capturing the distributional behavior observed in larger areas based on the ground truth. We train an ordinal regression scoring model and adjust the scores to follow the common power law within and across regions. Evaluation based on official statistics in South Korea shows that our method outperforms previous models in predicting population and employment size at both the municipality and grid levels. Our method also demonstrates robust performance in districts with uneven development, suggesting its potential use in developing countries where reliable, fine-grained data is scarce.

#### **Noisy Perturbations for Estimating Query Difficulty in Dense Retrievers**

Negar Arabzadeh, Radin Hamidi Rad, Maryam Khodabakhsh, Ebrahim Bagheri

Estimating query difficulty, also known as Query Performance Prediction (QPP), is concerned with assessing the retrieval quality of a ranking method for an input query. Most traditional unsupervised frequency-based models and many recent supervised neural methods have been designed specifically for predicting the performance of sparse retrievers such as BM25. In this paper we propose an unsupervised QPP method for dense neural retrievers which operates by redefining the well-known concept of query robustness i.e., a more robust query to perturbations is an easier query to handle. We propose to generate query perturbations for measuring query robustness by systematically injecting noise into the contextualized neural representation of each query. We then compare the retrieved list for the original query with that of the perturbed query as a way to measure query robustness. Our experiments on four different query sets including MS MARCO, TREC Deep Learning track 2019 and 2020 and TREC DL-Hard show consistently improved performance on linear and ranking correlation metrics over the state of the art.

#### **HOVER: Homophilic Oversampling via Edge Removal for Class-Imbalanced Bot Detection on Graphs**

Bradley Ashmore, Lingwei Chen

As malicious bots reside in a network to disrupt network stability, graph neural networks (GNNs) have emerged as one of the most popular bot detection methods. However, in most cases these graphs are significantly class-imbalanced. To address this issue, graph oversampling has recently been proposed to synthesize nodes and edges, which still suffers from graph heterophily, leading to suboptimal performance. In this paper, we propose HOVER, which implements Homophilic Oversampling Via Edge Removal for bot detection on graphs. Instead of oversampling nodes and edges within initial graph structure, HOVER designs a simple edge removal method with heuristic criteria to mitigate heterophily and learn distinguishable node embeddings, which are then used to oversample minority bots to generate a balanced class distribution without edge synthesis. Experiments on TON IoT networks demonstrate the state-of-the-art performance of HOVER on bot detection with high graph heterophily and extreme class imbalance.

#### **Accelerating Concept Learning via Sampling**

Alkid Baci, Stefan Heindorf



Node classification is an important task in many fields, e.g., predicting entity types in knowledge graphs, classifying papers in citation graphs, or classifying nodes in social networks. In many cases, it is crucial to explain why certain predictions are made. Towards this end, concept learning has been proposed as a means of interpretable node classification: given positive and negative examples in a knowledge base, concepts in description logics are learned that serve as classification models. However, state-of-the-art concept learners, including EvoLearner and CELOE exhibit long runtimes. In this paper, we propose to accelerate concept learning with graph sampling techniques. We experiment with seven techniques and tailor them to the setting of concept learning. In our experiments, we achieve a reduction in training size by over 90% while maintaining a high predictive performance.

#### [Logarithmic Dimension Reduction for Quantum Neural Networks](#)

Hankyul Baek, Soohyun Park, Joongheon Kim

In recent years, quantum neural network (QNN) based on quantum computing has attracted attention due to its potential for computation-acceleration and parallelism. However, the intrinsic limitations of QNN, where the output (i.e., observables) can only be obtained through a measurement process, pose scalability challenges. Motivated by this, this paper aims to address the scalability challenges by incorporating Pauli-Z measurement and Basis measurement. In conventional frameworks, QNN typically relies on classical fully connected networks (FCNs) or increases the number of qubits to achieve large output dimensions. However, by leveraging our proposed framework, this paper successfully expands the output dimensions to an exponential scale, surpassing the limitations imposed by the limited number of qubits without relying on FCNs. Through extensive experiments, this paper demonstrates that the proposed framework outperforms existing QNN frameworks in multi-class classification tasks that require numerous output dimensions.

#### [A Comparative Study of Reference Reliability in Multiple Language Editions of Wikipedia](#)

Aitolkyn Baigutanova, Diego Saez-Trumper, Miriam Redi, Meeyoung Cha, Pablo Aragón

Information presented in Wikipedia articles must be attributable to reliable published sources in the form of references. This study examines over 5 million Wikipedia articles to assess the reliability of references in multiple language editions. We quantify the cross-lingual patterns of the perennial sources list, a collection of reliability labels for web domains identified and collaboratively agreed upon by Wikipedia editors. We discover that some sources (or web domains) deemed untrustworthy in one language (i.e., English) continue to appear in articles in other languages. This trend is especially evident with sources tailored for smaller communities. Furthermore, non-authoritative sources found in the English version of a page tend to persist in other language versions of that page. We finally present a case study on the Chinese, Russian, and Swedish Wikipedias to demonstrate a discrepancy in reference reliability across cultures. Our finding highlights future challenges in coordinating global knowledge on source reliability.

#### [Non-Recursive Cluster-Scale Graph Interacted Model for Click-Through Rate Prediction](#)

Yuanchen Bei, Hao Chen, Shengyuan Chen, Xiao Huang, Sheng Zhou, Feiran Huang

Extracting users' interests from their behavior, particularly their 1-hop neighbors, has been shown to enhance Click-Through Rate (CTR) prediction performance. However, online recommender systems impose strict constraints on the inference time of CTR models, which necessitates pruning or filtering users' 1-hop neighbors to reduce computational complexity. Furthermore, while the graph information of users and items has been proven effective in collaborative filtering models, recursive graph convolution can be computationally costly and expensive to implement. To address these challenges, we propose the Non-Recursive Cluster-scale Graph Interacted (NRCGI) model, which reorganizes graph convolutional networks in a non-recursive and cluster-scale view to enable CTR models to consider deep graph information with low computational cost. NRCGI employs non-recursive cluster-scale graph aggregation, which allows the online recommendation computational complexity to shrink from tens of thousands of items to tens to hundreds of clusters. Additionally, since NRCGI aggregates neighbors in a non-recursive view, each hop of neighbors has a clear physical meaning. NRCGI explicitly constructs meaningful interactions between the hops of neighbors of users and items to fully model users' intent towards the given item. Experimental results demonstrate that NRCGI outperforms state-of-the-art baselines in three public datasets and one industrial dataset while maintaining efficient inference.

#### [Counterfactual Graph Augmentation for Consumer Unfairness Mitigation in Recommender Systems](#)

Ludovico Boratto, Francesco Fabbri, Gianni Fenu, Mirko Marras, Giacomo Medda

In recommendation literature, explainability and fairness are becoming two prominent perspectives to consider. However, prior works have mostly addressed them separately, for instance by explaining to consumers why a certain item was

recommended or mitigating disparate impacts in recommendation utility. None of them has leveraged explainability techniques to inform unfairness mitigation. In this paper, we propose an approach that relies on counterfactual explanations to augment the set of user-item interactions, such that using them while inferring recommendations leads to fairer outcomes. Modeling user-item interactions as a bipartite graph, our approach augments the latter by identifying new user-item edges that not only can explain the original unfairness by design, but can also mitigate it. Experiments on two public data sets show that our approach effectively leads to a better trade-off between fairness and recommendation utility compared with state-of-the-art mitigation procedures. We further analyze the characteristics of added edges to highlight key unfairness patterns. Source code available at <https://github.com/jackmedda/RS-BGExplainer/tree/cikm2023>.

### [Linkage Attack on Skeleton-based Motion Visualization](#)

Thomas Carr, Aidong Lu, Depeng Xu

Skeleton-based motion capture and visualization is an important computer vision task, especially in the virtual reality (VR) environment. It has grown increasingly popular due to the ease of gathering skeleton data and the high demand of virtual socialization. The captured skeleton data seems anonymous but can still be used to extract personal identifiable information (PII). This can lead to an unintended privacy leakage inside a VR meta-verse. We propose a novel linkage attack on skeleton-based motion visualization. It detects if a target and a reference skeleton are the same individual. The proposed model, called Linkage Attack Neural Network (LAN), is based on the principles of a Siamese Network. It incorporates deep neural networks to embed the relevant PII then uses a classifier to match the reference and target skeletons. We also employ classical and deep motion retargeting (MR) to cast the target skeleton onto a dummy skeleton such that the motion sequence is anonymized for privacy protection. Our evaluation shows that the effectiveness of LAN in the linkage attack and the effectiveness of MR in anonymization.

### [Region-Wise Attentive Multi-View Representation Learning For Urban Region Embedding](#)

Weiliang Chan, Qianqian Ren

Urban region embedding is an important and yet highly challenging issue due to the complexity and constantly changing nature of urban data. To address the challenges, we propose a Region-Wise Multi-View Representation Learning (ROMER) to capture multi-view dependencies and learn expressive representations of urban regions without the constraints of rigid neighbourhood region conditions. Our model focuses on learn urban region representation from multi-source urban data. First, we capture the multi-view correlations from mobility flow patterns, POI semantics and check-in dynamics. Then, we adopt global graph attention networks to learn similarity of any two vertices in graphs. To comprehensively consider and share features of multiple views, a two-stage fusion module is further proposed to learn weights with external attention to fuse multi-view embeddings. Extensive experiments for two downstream tasks on real-world datasets demonstrate that our model outperforms state-of-the-art methods by up to 17% improvement.

### ['Choose your Data Wisely': Active Learning based Selection with Multi-Objective Optimisation for Mitigating Stereotypes](#)

Manish Chandra, Debasis Ganguly, Tulika Saha, Iadh Ounis

Data-driven (deep) learning methods has led to parameterised abstractions of the data, often leading to stereotype societal biases in their predictions, e.g., predicting more frequently that women are weaker than men, or that African Americans are more likely to commit crimes than Caucasians. Standard approaches of mitigating such stereotypical biases from deep neural models include modifying the training dataset (pre-processing), or adjusting the model parameters with a bias-specific objective (in-processing). In our work, we approach this bias mitigation from a different perspective - that of an active learning-based selection of a subset of data instances towards training a model optimised for both effectiveness and fairness. Specifically speaking, the imbalances in the attribute value priors can be alleviated by constructing a balanced subset of the data instances with two selection objectives - first, of improving the model confidence of the primary task itself (a standard practice in active learning), and the second, of taking into account the parity of the model predictions with respect to the sensitive attributes, such as gender and race etc. We demonstrate that our proposed selection function achieves better results in terms of both the primary task effectiveness and fairness. The results are further shown to improve when this active learning-based data selection is combined with an in-process method of multi-objective training.

### [Semi-supervised Curriculum Ensemble Learning for Financial Precision Marketing](#)

Hsin-Yu Chen, Cheng-Te Li, Ting-Yu Chen

This paper tackles precision marketing in financial technology, focusing on the accurate prediction of potential customers' interest in specific financial products amidst extreme class imbalance and a significant volume of unlabeled data. We propose the innovative Semi-supervised Curriculum Ensemble (SSCE) framework, which integrates curriculum pseudo-labeling and balanced bagging with tree-based models. This novel approach enables the effective utilization of high-confidence predicted instances from unlabeled data and mitigates the impact of extreme class imbalance. Experiments conducted on a large-scale real-world banking dataset, featuring five financial products, demonstrate that the SSCE consistently outperforms existing methods, thereby promising significant advances in the domain of financial precision marketing.

#### **TCCM: Time and Content-Aware Causal Model for Unbiased News Recommendation**

Yewang Chen, Weiyao Ye, Guipeng Xv, Chen Lin, Xiaomin Zhu

Popularity bias significantly impacts news recommendation systems, as popular news articles receive more exposure and are often delivered to irrelevant users, resulting in unsatisfactory performance. Existing methods have not adequately addressed the issue of popularity bias in news recommendations, largely due to the neglect of the time factor and the impact of news content on popularity. In this paper, we propose a novel approach called Time and Content-aware Causal Model, namely TCCM. It models the effects of three factors on user interaction behavior, i.e., the time factor, the news popularity, and the matching between news content and user interest. TCCM also estimates news popularity more accurately by incorporating the news content, i.e., the popularity of entity and words. Causal intervention techniques are applied to obtain debiased recommendations. Extensive experiments on well-known benchmark datasets demonstrate that the proposed approach outperforms a range of state-of-the-art techniques.

#### **Unsupervised Anomaly Detection & Diagnosis: A Stein Variational Gradient Descent Approach**

Zhichao Chen, Leilei Ding, Jianmin Huang, Zhixuan Chu, Qingyang Dai, Hao Wang

Detecting and diagnosing anomalies in observational data plays a crucial role in various real-world applications, such as e-commerce applet maintenance. Unsupervised machine learning techniques are typically employed for anomaly detection and diagnosis due to their convenience and independence from labeled data. Density estimation (DE), as one of the most widely used unsupervised machine learning techniques for anomaly detection, can be categorized into kernel density estimation (KDE)-based methods and normalizing flow (NF)-based methods. While KDE-based methods offer fast computation speed, they often ignore the complex manifold structure present in observational data. On the other hand, NF-based methods address the manifold issue but suffer from longer computation times. In this study, we propose a novel DE-based anomaly detection & diagnosis method using Stein Variational Gradient Descent (SVGD), aiming to leverage the strengths of KDE and NF approaches. Firstly, we rigorously derive the DE capability of SVGD through mathematical analysis. Subsequently, we demonstrate the ability of the SVGD method to perform anomaly diagnosis based on input feature attribution. Finally, to validate the effectiveness of our approach, we conduct experiments using synthetic, benchmark, and industrial datasets. The results demonstrate the superior performance and practical applicability of our proposed method.

#### **Segment Augmentation and Prediction Consistency Neural Network for Multi-label Unknown Intent Detection**

Miaoxin Chen, Cao Liu, Boqi Dai, Hai-Tao Zheng, Ting Song, Jiansong Chen, Guanglu Wan, Rui Xie

Multi-label unknown intent detection is a challenging task where each utterance may contain not only multiple known but also unknown intents. To tackle this challenge, pioneers proposed to predict the intent number of the utterance first, then compare it with the results of known intent matching to decide whether the utterance contains unknown intent(s). Though they have made remarkable progress on this task, their method still suffers from two important issues: 1) It is inadequate to extract multiple intents using only utterance encoding; 2) Optimizing two sub-tasks (intent number prediction and known intent matching) independently leads to inconsistent predictions. In this paper, we propose to incorporate segment augmentation rather than only use utterance encoding to better detect multiple intents. We also design a prediction consistency module to bridge the gap between the two sub-tasks. Empirical results on MultiWOZ2.3 show that our method achieves state-of-the-art performance and improves the best baseline significantly.

#### **Attribute-enhanced Dual Channel Representation Learning for Session-based Recommendation**

Qian Chen, Jianjun Li, Zhiqiang Guo, Guohui Li, Zhiying Deng

Session-based recommendation (SBR) aims to predict the anonymous user's next-click items by modeling the short-term sequence pattern. As most existing SBR models generally generate item representations based only on information propagation over the short sequence while ignoring additional valuable knowledge, their expressive abilities are somewhat limited by data sparsity caused by short sequence. Though there have been some attempts on utilizing items' attributes, they

basically embed attributes into items directly, ignoring the fact that 1) there is no contextual relationship among attributes; and 2) users have varying levels of attention to different attributes, which still leads to unsatisfactory performance. To tackle the issues, we propose a novel Attribute-enhanced Dual Channel Representation Learning (ADRL) model for SBR, in which we independently model session representations in attribute-related pattern and sequence-related pattern. Specifically, we learn session representations with sequence patterns from the session graph, and we further design an frequency-driven attribute aggregator to generate the attribute-related session representations within a session. The proposed attribute aggregator is plug-and-play, as it can be coupled with most existing SBR models. Extensive experiments on three real-world public datasets demonstrate the superiority of the proposed ADRL over several state-of-the-art baselines, as well as the effectiveness and efficiency of our attribute aggregator module.

### **Assessing Student Performance with Multi-granularity Attention from Online Classroom Dialogue**

Jiahao Chen, Zitao Liu, Shuyan Huang, Yaying Huang, Xiangyu Zhao, Boyu Gao, Weiqi Luo

Accurately judging students' ongoing performance is very crucial for real-world educational scenarios. In this work, we focus on the task of automatically predicting students' levels of mastery of math questions from teacher-student classroom dialogue data in the online learning environment. We propose a novel neural network armed with a multi-granularity attention mechanism to capture the personalized pedagogical instructions from the very noisy teacher-student dialogue transcriptions. We conduct experiments on a real-world educational dataset and the results demonstrate the superiority and availability of our model in terms of various evaluation metrics.

### **MI-DPG: Decomposable Parameter Generation Network Based on Mutual Information for Multi-Scenario**

#### **Recommendation**

Wenzhuo Cheng, Ke Ding, Xin Dong, Yong He, Liang Zhang, Linjian Mo

Conversion rate (CVR) prediction models play a vital role in recommendation systems. Recent research shows that learning a unified model to serve multiple scenarios is effective for improving overall performance. However, it remains challenging to improve model prediction performance across scenarios at low model parameter cost, and current solutions are hard to robustly model multi-scenario diversity. In this paper, we propose MI-DPG for the multi-scenario CVR prediction, which learns scenario-conditioned dynamic model parameters for each scenario in a more efficient and effective manner. Specifically, we introduce an auxiliary network to generate scenario-conditioned dynamic weighting matrices, which are obtained by combining decomposed scenario-specific and scenario-shared low-rank matrices with parameter efficiency. For each scenario, weighting the backbone model parameters by the weighting matrix helps to specialize the model parameters for different scenarios. It can not only modulate the complete parameter space of the backbone model but also improve the model effectiveness. Furthermore, we design a mutual information regularization to enhance the diversity of model parameters across scenarios by maximizing the mutual information between the scenario-aware input and the scenario-conditioned dynamic weighting matrix. Experiments from three real-world datasets show that MI-DPG outperforms previous multi-scenario recommendation models.

### **Enhancing Information Diffusion Prediction with Self-Supervised Disentangled User and Cascade Representations**

Zhangtao Cheng, Wenxue Ye, Leyuan Liu, Wenxin Tai, Fan Zhou

Accurately predicting information diffusion is critical for a vast range of applications. Existing methods generally consider user re-sharing behaviors to be driven by a single intent, and/or assume cascade temporal influence to be unchanged, which might not be consistent with real-world scenarios. To address these issues, we propose a self-supervised disentanglement framework (DisenIDP) for information diffusion prediction. First, we construct intent-aware hypergraphs to capture users' potential intents from different perspectives, and then perform the light hypergraph convolution to adaptively activate disentangled intents. Second, we extract long-term and short-term cascade influence via independent attention-based encoders. Finally, we set a self-supervised disentanglement task to alleviate the information loss and learn better-disentanglement representations. Extensive experiments conducted on two real-world social datasets demonstrate that DisenIDP outperforms state-of-the-art models across several settings.

### **Identify Risky Rules to Reduce Side Effects in Association Rule Hiding**

Peng Cheng

Data sharing is necessary for many practical applications. People do, however, frequently worry about the problem of privacy leaking. This study focuses on preventing the disclosure of sensitive information using association rules and frequent itemsets, which are frequently utilized in numerous applications. How to minimize side effects while hiding, particularly side

effects on non-sensitive knowledge, is the difficult part of the problem. The majority of association rule hiding techniques currently in use solely consider reducing side effects on frequent itemsets (patterns), rather than rules, in order to conceal sensitive rules by reducing the statistical disclosure of the itemsets that generate such rules. In this study, we provide a concealment technique utilizing potentially risky rules to lessen adverse impacts on non-sensitive rules, not only itemsets. In addition, this method can be tailored to conceal sensitive itemsets, instead of rules. Extensive experiments show that in most cases the proposed solution can bring fewer side effects on rules, frequent patterns or data quality than existing methods.

#### [Patient Clustering via Integrated Profiling of Clinical and Digital Data](#)

Dongjin Choi, Andy Xiang, Ozgur Ozturk, Deep Shrestha, Barry Drake, Hamid Haidarian, Faizan Javed, Haesun Park

We introduce a novel profile-based patient clustering model designed for healthcare clinical data. By utilizing a method grounded on constrained low-rank approximation, our model takes advantage of patients' clinical data and digital interaction data, including browsing and search, to construct patient profiles. As a result of the method, nonnegative embedding vectors are generated, serving as a low-dimensional representation of the patients. Our model was assessed using real-world patient data from a healthcare web portal, with a comprehensive evaluation approach which considered clustering and recommendation capabilities. In comparison to other baselines, our approach demonstrated superior performance in terms of clustering coherence and recommendation accuracy.

#### [Incorporating Co-purchase Correlation for Next-basket Recommendation](#)

Yu Hao Chou, Pu Jen Cheng

Next-basket recommendation (NBR) aims to recommend a set of items that users would most likely purchase together. Existing approaches use deep learning to capture basket-level preference and traditional statistical methods to model user behavior sequences. However, these methods neglect the correlation of co-purchase items among users. We, therefore, propose a novel model that incorporates Co-purchase Correlation with Bidirectional Transformer (CCBT) to enhance item representation by exploiting the correlation among users' baskets. The results of experiments conducted on four real-world datasets demonstrate the proposed model outperforms state-of-the-art NBR methods. The relative improvement for Recall@20 ranges from 11% to 27%.

#### [Learning Invariant Representations for New Product Sales Forecasting via Multi-Granularity Adversarial Learning](#)

Zhenzhen Chu, Chengyu Wang, Cen Chen, Dawei Cheng, Yuqi Liang, Weining Qian

Sales forecasting during the launch of new products has always been a challenging task, due to the lack of historical sales data. The dynamic market environment and consumer preferences also increase the uncertainty of predictions. Large chains face even greater difficulties due to their extensive presence across various regions. Traditional time-series forecasting methods usually rely on statistical models and empirical judgments, which are difficult to handle large, variable data and often fail to achieve satisfactory performance for new products. In this paper, we propose a Multi-granularity Adversarial Learning framework (MARL) to leverage knowledge from old products and improve the quality of invariant representations for more accurate sales predictions. To evaluate our proposed method, we conducted extensive experiments on both a real-world dataset from a prominent international Café chain and a public dataset. The results demonstrated that our method is more effective than the existing state-of-the-art baselines for new product sales forecasting.

#### [Reconciling Training and Evaluation Objectives in Location Agnostic Surrogate Explainers](#)

Matt Clifford, Jonathan Erskine, Alexander Hepburn, Peter Flach, Raúl Santos-Rodríguez

Transparency in AI models is crucial to designing, auditing, and deploying AI systems. However, 'black box' models are still used in practice for their predictive power despite their lack of transparency. This has led to a demand for post-hoc, model-agnostic surrogate explainers which provide explanations for decisions of any model by approximating its behaviour close to a query point with a surrogate model. However, it is often overlooked how the location of the query point in the decision surface of the black box model affects the faithfulness of the surrogate explainer. Here, we show that when using standard techniques, there is a decrease in agreement between the black box and the surrogate model for query points towards the edge of the test dataset and when moving away from the decision boundary. This originates from a mismatch between the data distributions used to train and evaluate surrogate explainers. We address this by leveraging knowledge about the test data distribution captured in the class labels of the black box model. By addressing this and encouraging users to take care in understanding the alignment of training and evaluation objectives, we empower them to construct more faithful surrogate explainers.



**DPAN: Dynamic Preference-based and Attribute-aware Network for Relevant Recommendations**

Wei Dai, Yingmin Su, Xiaofeng Pan, Yufeng Wang, Zhenyu Zhu, Nan Xu, Chengjun Mao, Bo Cao

In e-commerce platforms, the relevant recommendation is a unique scenario providing related items for a trigger item that users are interested in. However, users' preferences for the similarity and diversity of recommendation results are dynamic and vary under different conditions. Moreover, individual item-level diversity is too coarse-grained since all recommended items are related to the trigger item. Thus, the two main challenges are to learn fine-grained representations of similarity and diversity and capture users' dynamic preferences for them under different conditions. To address these challenges, we propose a novel method called the Dynamic Preference-based and Attribute-aware Network (DPAN) for predicting Click-Through Rate (CTR) in relevant recommendations. Specifically, based on Attribute-aware Activation Values Generation (AAVG), Bi-dimensional Compression-based Re-expression (BCR) is designed to obtain similarity and diversity representations of user interests and item information. Then Shallow and Deep Union-based Fusion (SDUF) is proposed to capture users' dynamic preferences for the diverse degree of recommendation results according to various conditions. DPAN has demonstrated its effectiveness through extensive offline experiments and online A/B testing, resulting in a significant 7.62% improvement in CTR. Currently, DPAN has been successfully deployed on our e-commerce platform serving the primary traffic for relevant recommendations.

**Efficient Variant Calling on Human Genome Sequences Using a GPU-Enabled Commodity Cluster**

Manas Jyoti Das, Khawar Shehzad, Praveen Rao

Human genome sequences are very large in size and require significant compute and storage resources for processing and analysis. Variant calling is a key task performed on an individual's genome to identify different types of variants. Knowing these variants can lead to new advances in disease diagnosis and treatment. In this work, we propose a new approach for accelerating variant calling pipelines on a large workload of human genomes using a commodity cluster with graphics processing units (GPUs). Our approach has two salient features: First, it enables a pipeline stage to use GPUs and/or CPUs based on the availability of resources in the cluster. Second, it employs a mutual exclusion strategy for executing a pipeline stage on the GPUs of a cluster node so that the stages (for other sequences) can be executed using CPUs if needed. We evaluated our approach on a 8-node cluster with bare metal servers and virtual machines (VMs) containing different types of GPUs. On publicly available genome sequences, our approach was 3.6X-5X faster compared to an approach that used only the cluster CPUs.

**Self-supervised Learning and Graph Classification under Heterophily**

Yilin Ding, Zhen Liu, Hao Hao

Most existing pre-training strategies usually choose the popular Graph Neural Networks (GNNs), which can be seen as a special form of low-pass filter, but fail to effectively capture heterophily. In this paper, we first present an experimental investigation exploring the performance of low-pass and high-pass filters in heterophily graph classification, where the results clearly show that high-frequency signal is important for learning heterophily graph representation. In addition, it is still unclear how to effectively capture the structural pattern of graphs and how to measure the capability of the self-supervised pre-training strategy in capturing graph structure. To address the problem, we first design a quantitative Metric for Graph Structure (MGS), which analyzes the correlation between structural similarity and embedding similarity of graph pairs. Then, to enhance the graph structural information captured by self-supervised learning, we propose a novel self-supervised strategy for Pre-training GNNs based on the Metric (PGM). Extensive experiments validate our pre-training strategy achieves state-of-the-art performance for molecular property prediction and protein function prediction. In addition, we find choosing a suitable filter sometimes may be better than designing good pre-training strategies for heterophily graph classification.

**Bridging the KB-Text Gap: Leveraging Structured Knowledge-aware Pre-training for KBQA**

Guanting Dong, Rumei Li, Sirui Wang, Yupeng Zhang, Yunsen Xian, Weiran Xu

Knowledge Base Question Answering (KBQA) aims to answer natural language questions with factual information such as entities and relations in KBs. However, traditional Pre-trained Language Models (PLMs) are directly pre-trained on large-scale natural language corpus, which poses challenges for them in understanding and representing complex subgraphs in structured KBs. To bridge the gap between texts and structured KBs, we propose a Structured Knowledge-aware Pre-training method (SKP). In the pre-training stage, we introduce two novel structured knowledge-aware tasks, guiding the model to effectively learn the implicit relationship and better representations of complex subgraphs. In downstream KBQA task, we further design an efficient linearization strategy and an interval attention mechanism, which assist the model to better encode

complex subgraphs and shield the interference of irrelevant subgraphs during reasoning respectively. Detailed experiments and analyses on WebQSP verify the effectiveness of SKP, especially the significant improvement in subgraph retrieval (+4.08% H@10).

### [Geometric Matrix Completion via Sylvester Multi-Graph Neural Network](#)

Boxin Du, Changhe Yuan, Fei Wang, Hanghang Tong

Despite the success of the Sylvester equation empowered methods on various graph mining applications, such as semi-supervised label learning and network alignment, there also exists several limitations. The Sylvester equation's inability of modeling non-linear relations and the inflexibility of tuning towards different tasks restrict its performance. In this paper, we propose an end-to-end neural framework, SYMGNN, which consists of a multi-network neural aggregation module and a prior multi-network association incorporation learning module. The proposed framework inherits the key ideas of the Sylvester equation, and meanwhile generalizes it to overcome aforementioned limitations. Empirical evaluations on real-world datasets show that the instantiations of SYMGNN overall outperform the baselines in geometric matrix completion task, and its low-rank instantiation could further reduce the memory consumption by 16.98% on average.

### [Learning Sparse Lexical Representations Over Specified Vocabularies for Retrieval](#)

Jeffrey M. Dudek, Weize Kong, Cheng Li, Mingyang Zhang, Michael Bendersky

A recent line of work in first-stage Neural Information Retrieval has focused on learning sparse lexical representations instead of dense embeddings. One such work is SPLADE, which has been shown to lead to state-of-the-art results in both the in-domain and zero-shot settings, can leverage inverted indices for efficient retrieval, and offers enhanced interpretability. However, existing SPLADE models are fundamentally limited to learning a sparse representation based on the native BERT WordPiece vocabulary.

In this work, we extend SPLADE to support learning sparse representations over arbitrary sets of tokens to improve flexibility and aid integration with existing retrieval systems. As an illustrative example, we focus on learning a sparse representation over a large (300k) set of unigrams. We add an unsupervised pretraining task on C4 to learn internal representations for new tokens. Our experiments show that our Expanded-SPLADE model maintains the performance of WordPiece-SPLADE on both in-domain and zero-shot retrieval while allowing for custom output vocabularies.

### [OnlineAutoClust: A Framework for Online Automated Clustering](#)

Radwa El Shawi, Dmitri Rozgonjuk

Automated Machine Learning (AutoML) has been successful when the learning task is assumed to be static. However, it remains unclear whether AutoML methods can efficiently create online pipelines in dynamic environments. The current online AutoML frameworks primarily focus on supervised learning. However, unsupervised learning, particularly clustering, also requires AutoML solutions, especially with the ambiguity associated with evaluating clustering results. In this paper, we introduce OnlineAutoClust, a framework for online automated clustering for algorithm selection and hyperparameter tuning. OnlineAutoClust combines the inherent adaptation capabilities of online learners with automated pipeline optimization using Bayesian optimization. OnlineAutoClust develops a collaborative mechanism based on clustering ensemble to combine optimized pipelines based on different internal cluster validity indices. The proposed framework is based on River library and utilizes five clustering algorithms. Empirical evaluation on several real and synthetic data streams with varying types of concept drift demonstrates the effectiveness of the proposed approach compared to existing methods

### [Simulating Users in Interactive Web Table Retrieval](#)

Björn Engelmann, Timo Breuer, Philipp Schaer

Considering the multimodal signals of search items is beneficial for retrieval effectiveness. Especially in web table retrieval (WTR) experiments, accounting for multimodal properties of tables boosts effectiveness. However, it still remains an open question how the single modalities affect user experience in particular. Previous work analyzed WTR performance in ad-hoc retrieval benchmarks, which neglects interactive search behavior and limits the conclusion about the implications for real-world user environments.

To this end, this work presents an in-depth evaluation of simulated interactive WTR search sessions as a more cost-efficient and reproducible alternative to real user studies. As a first of its kind, we introduce interactive query reformulation strategies based on Doc2Query, incorporating cognitive states of simulated user knowledge. Our evaluations include two perspectives

on user effectiveness by considering different cost paradigms, namely query-wise and time-oriented measures of effort. Our multi-perspective evaluation scheme reveals new insights about query strategies, the impact of modalities, and different user types in simulated WTR search sessions.

### **KGPR: Knowledge Graph Enhanced Passage Ranking**

Jinyuan Fang, Zaiqiao Meng, Craig Macdonald

Passage ranking aims to rank a set of passages based on their relevance to a query. Current state-of-the-art models for this task typically employ a cross-encoder structure. However, these models lack access to background knowledge, i.e., information related to the query that can be helpful in retrieving relevant passages. Knowledge Graphs (KGs) provide a structured way of storing information about entities and their relationships, offering valuable background knowledge about entities. While KGs have been used to augment pretrained language models (LMs) to perform several reasoning tasks such as question answering, it remains an open question of how to utilise the information from KGs to enhance the performance of cross-encoders on the passage ranking task. Therefore, we propose KGPR, a KG-enhanced cross-encoder for the Passage Retrieval task. KGPR is built upon LUKE, an entity-aware pretrained LM, with an additional module that fuses information from KGs into LUKE. By leveraging the background knowledge from KGs, KGPR enhances the model's comprehension of queries and passages, resulting in improved ranking performance. Experimental results demonstrate that using KGs can enhance the performance of LUKE in the passage retrieval task, and KGPR can outperform state-of-the-art monoT5 cross-encoder by 3.32% and 10.77% on the MS MARCO development set and TREC DL-HARD query set respectively, using a model with a similar number of parameters.

### **Multi-step Prompting for Few-shot Emotion-Grounded Conversations**

Mauzama Firdaus, Gopendra Singh, Asif Ekbal, Pushpak Bhattacharyya

Conversational systems have shown immense growth in their ability to communicate like humans. With the emergence of large pre-trained language models (PLMs) the ability to provide informative responses have improved significantly. Despite the success of PLMs, the ability to identify and generate engaging and empathetic responses is largely dependent on labelled-data. In this work, we design a prompting approach that identifies the emotion of a given utterance and uses the emotion information for generating the appropriate responses for conversational systems. We propose a two-step prompting method that first recognises the emotion in the dialogue utterance and in the second-step uses the predicted emotion to prompt the PLM to generate the corresponding empathetic response in a few-shot setting. Experimental results on three publicly available datasets show that our proposed approach outperforms the state-of-the-art approaches for both automatic and manual evaluation.

### **Synergistic Disease Similarity Measurement via Unifying Hierarchical Relation Perception and Association Capturing**

Zihao Gao, Huifang Ma, Yike Wang, Zhixin Li, Liang Chang

Quantifying similarities among human diseases is crucial to enhance our understanding of disease biology. Deep learning efforts have been devoted to quantifying disease similarity by integrating multi-view data sources from disparate biological data. However, disease data are often sparse, leading to suboptimal representation of disease given biological entity relationships and labeled disease data are not adequately modeled. In this paper, we propose an effective Synergistic disease Similarity measurement model called SynerSim. SynerSim possesses two key components: a hierarchical biological entity relation perception module to capture disease features from various biological entities, and a disease association capturing module based on signed random walk to model precious disease data. Additionally, SynerSim leverages dual granularity contrastive learning to enhance the representation of diverse biological entities, owing to the ability to enable the synergistic supervision of diseases represented by both homogeneous and heterogeneous information. Experimental results demonstrate that SynerSim achieves outstanding performance in the disease similarity measurement.

### **Extracting Methodology Components from AI Research Papers: A Data-driven Factored Sequence Labeling Approach**

Madhusudan Ghosh, Debasis Ganguly, Partha Basuchowdhuri, Sudip Kumar Naskar

Extraction of methodology component names from scientific articles is a challenging task due to the diversified contexts around the occurrences of these entities, and the different levels of granularity and containment relationships exhibited by these entities. We hypothesize that standard sequence labeling approaches may not adequately model the dependence of methodology name mentions with their contexts, due to the problems of their large, fast evolving, and domain-specific vocabulary. As a solution, we propose a factored approach, where the mention-context dependencies are represented in a more fine-grained manner, thus allowing the model parameters to better adjust to the different characteristic patterns inherent

within the data. In particular, we experiment with two variants of this factored approach - one that uses the per-entity category information derived from an ontology, and the other that makes use of the topology of the sentence embedding space to infer a category for each entity constituting that sentence. We demonstrate that both these factored variants of SciBERT outperform their non-factored counterpart, a state-of-the-art model for scientific concept extraction.

#### **Knowledge-Enhanced Multi-Label Few-Shot Product Attribute-Value Extraction**

Jiaying Gong, Wei-Te Chen, Hoda Eldardiry

Existing attribute-value extraction (AVE) models require large quantities of labeled data for training. However, new products with new attribute-value pairs enter the market every day in real-world e-Commerce. Thus, we formulate AVE in multi-label few-shot learning (FSL), aiming to extract unseen attribute value pairs based on a small number of training examples. We propose a Knowledge-Enhanced Attentive Framework (KEAF) based on prototypical networks, leveraging the generated label description and category information to learn more discriminative prototypes. Besides, KEAF integrates with hybrid attention to reduce noise and capture more informative semantics for each class by calculating the label-relevant and query-related weights. To achieve multi-label inference, KEAF further learns a dynamic threshold by integrating the semantic information from both the support set and the query set. Extensive experiments with ablation studies conducted on two datasets demonstrate that our proposed model significantly outperforms other SOTA models for information extraction in few-shot learning.

#### **Neighborhood Homophily-based Graph Convolutional Network**

Shengbo Gong, Jiajun Zhou, Chenxuan Xie, Qi Xuan

Graph neural networks (GNNs) have been proved powerful in graph-oriented tasks. However, many real-world graphs are heterophilous, challenging the homophily assumption of classical GNNs. To solve the universality problem, many studies deepen networks or concatenate intermediate representations, which does not inherently change neighbor aggregation and introduces noise. Recent studies propose new metrics to characterize the homophily, but rarely consider the correlation of the proposed metrics and models. In this paper, we first design a new metric, Neighborhood Homophily (NH), to measure the label complexity or purity in node neighborhoods. Furthermore, we incorporate the metric into the classical graph convolutional network (GCN) architecture and propose Neighborhood Homophily-based Graph Convolutional Network (NHGCN). In this framework, neighbors are grouped by estimated NH values and aggregated from different channels, and the resulting node predictions are then used in turn to estimate and update NH values. The two processes of metric estimation and model inference are alternately optimized to achieve better node classification. NHGCN achieves top overall performance on both homophilous and heterophilous benchmarks, with an improvement of up to 7.4% compared to the current SOTA methods.

#### **Explainable and Accurate Natural Language Understanding for Voice Assistants and Beyond**

Kalpa Gunaratna, Vijay Srinivasan, Hongxia Jin

Joint intent detection and slot filling, which is also termed as joint NLU (Natural Language Understanding) is invaluable for smart voice assistants. Recent advancements in this area have been heavily focusing on improving accuracy using various techniques. Explainability is undoubtedly an important aspect for deep learning-based models including joint NLU models. Without explainability, their decisions are opaque to the outside world and hence, have tendency to lack user trust. Therefore to bridge this gap, we transform the full joint NLU model to be 'inherently' explainable at granular levels without compromising on accuracy. Further, as we enable the full joint NLU model explainable, we show that our extension can be successfully used in other general classification tasks. We demonstrate this using sentiment analysis and named entity recognition.

#### **Leveraging Post-Click User Behaviors for Calibrated Conversion Rate Prediction Under Delayed Feedback in Online Advertising**

Yuyao Guo, Xiang Ao, Qiming Liu, Qing He

Obtaining accurately calibrated conversion rate predictions is essential for the bidding and ranking process in online advertising systems. Nevertheless, the inherent latency between clicks and conversions leads to delayed feedback, which may introduce bias into the prediction models. Compared to indefinitely long conversion delays, post-click user behaviors manifest within a relatively brief time and have been empirically validated to exert a favorable influence on the precision of conversion rate estimates. In light of this, we propose a novel approach that leverages post-click user behaviors to calibrate conversion rate predictions. Specifically, we treat user behaviors as predictable targets to improve accuracy and enhance timeliness. An adaptive loss function based on task uncertainty is employed for multi-task learning. To further reduce

calibration error, we integrate the modified prediction model with a parameterized scaling technique. Experiments conducted on two real-world datasets demonstrate that our proposed method outperforms existing models in providing more calibrated predictions.

### **Hateful Comment Detection and Hate Target Type Prediction for Video Comments**

Shrey Gupta, Pratyush Priyadarshi, Manish Gupta

With the widespread increase in hateful content on the web, hate detection has become more crucial than ever. Although vast literature exists on hate detection from text, images and videos, interestingly, there has been no previous work on hateful comment detection (HCD) from video pages. HCD is critical for comment moderation and for flagging controversial videos. Comments are often short, contextual and convoluted making the problem challenging. Toward solving this problem, we contribute a dataset, HateComments, consisting of 2071 comments for 401 videos obtained from two popular video sharing platforms. We investigate two related tasks: binary HCD and 4-class multi-label hate target-type prediction (HTP). We systematically explore the importance of various forms of context for effective HCD. Our initial experiments show that our best method which leverages rich video context (like description, transcript and visual input) leads to an HCD accuracy of ~78.6% and an ROC AUC score of ~0.61 for HTP. Code and data is at <https://drive.google.com/file/d/1EubWDUokv1CYkWKlwByUC6yluBGUw2MN/>.

### **A Deep Conditional Generative Approach for Constrained Community Detection**

Chaobo He, Junwei Cheng, Quanlong Guan, Xiang Fei, Hanchao Li, Yong Tang

Constrained community detection is one of the popular topics in graph data mining, and it aims to improve the performance by exploiting prior pairwise constraints, such as must-link and cannot-link constraints. However, most of existing methods for constrained community detection are shallow approaches, and are also not robust to handle constraints information with noises. In view of this, we propose a deep conditional generative approach CGMVGAE. It firstly treats pairwise constraints as the priors with different degrees of certainty, and then integrates them into the conditional Gaussian mixture model. By further combining variational graph auto-encoders and the Wasserstein regularization, CGMVGAE can learn the latent node representations preserving community structures in a deep generative manner. Experimental results show that CGMVGAE outperforms state-of-the-art approaches, and is also more robust.

### **Perturbation-Based Two-Stage Multi-Domain Active Learning**

Rui He, Zeyu Dai, Shan He, Ke Tang

In multi-domain learning (MDL) scenarios, high labeling effort is required due to the complexity of collecting data from various domains. Active Learning (AL) presents an encouraging solution to this issue by annotating a smaller number of highly informative instances, thereby reducing the labeling effort. Previous research has relied on conventional AL strategies for MDL scenarios, which underutilize the domain-shared information of each instance during the selection procedure. To mitigate this issue, we propose a novel perturbation-based two-stage multi-domain active learning (P2S-MDAL) method incorporated into the well-regarded ASP-MTL model. Specifically, P2S-MDAL involves allocating budgets for domains and establishing regions for diversity selection, which are further used to select the most cross-domain influential samples in each region. A perturbation metric has been introduced to evaluate the robustness of the shared feature extractor of the model, facilitating the identification of potentially cross-domain influential samples. Experiments are conducted on three real-world datasets, encompassing both texts and images. The superior performance over conventional AL strategies shows the effectiveness of the proposed strategy. Additionally, an ablation study has been carried out to demonstrate the validity of each component. Finally, we outline several intriguing potential directions for future MDAL research, thus catalyzing the field's advancement.

### **KGrEaT: A Framework to Evaluate Knowledge Graphs via Downstream Tasks**

Nicolas Heist, Sven Hertling, Heiko Paulheim

In recent years, countless research papers have addressed the topics of knowledge graph creation, extension, or completion in order to create knowledge graphs that are larger, more correct, or more diverse. This research is typically motivated by the argumentation that using such enhanced knowledge graphs to solve downstream tasks will improve performance. Nonetheless, this is hardly ever evaluated. Instead, the predominant evaluation metrics - aiming at correctness and completeness - are undoubtedly valuable but fail to capture the complete picture, i.e., how useful the created or enhanced knowledge graph actually is. Further, the accessibility of such a knowledge graph is rarely considered (e.g., whether it contains expressive labels, descriptions, and sufficient context information to link textual mentions to the entities of the



knowledge graph). To better judge how well knowledge graphs perform on actual tasks, we present KGrEaT - a framework to estimate the quality of knowledge graphs via actual downstream tasks like classification, clustering, or recommendation. Instead of comparing different methods of processing knowledge graphs with respect to a single task, the purpose of KGrEaT is to compare various knowledge graphs as such by evaluating them on a fixed task setup. The framework takes a knowledge graph as input, automatically maps it to the datasets to be evaluated on, and computes performance metrics for the defined tasks. It is built in a modular way to be easily extendable with additional tasks and datasets.

#### [Latent Aspect Detection via Backtranslation Augmentation](#)

Farinam Hemmatizadeh, Christine Wong, Alice Yu, Hossein Fani

Within the context of review analytics, aspects are the features of products and services at which customers target their opinions and sentiments. Aspect detection helps product owners and service providers identify shortcomings and prioritize customers' needs. Existing methods focus on detecting the surface form of an aspect falling short when aspects are latent in reviews, especially in an informal context like in social posts. In this paper, we propose data augmentation via natural language backtranslation to extract latent occurrences of aspects. We presume that backtranslation (1) can reveal latent aspects because they may not be commonly known in the target language and can be generated through backtranslation; (2) augments context-aware synonymous aspects from a target language to the original language, hence addressing the out-of-vocabulary issue; and (3) helps with the semantic disambiguation of polysemous words and collocations. Through our experiments on well-known aspect detection methods across semeval datasets of restaurant and laptop reviews, we demonstrate that review augmentation via backtranslation yields a steady performance boost in baselines. We further contribute LADy at <https://github.com/fani-lab/LADy>, a benchmark library to support the reproducibility of our research.

#### [Deep Context Interest Network for Click-Through Rate Prediction](#)

Xuyang Hou, Zhe Wang, Qi Liu, Tan Qu, Jia Cheng, Jun Lei

Click-Through Rate (CTR) prediction, estimating the probability of a user clicking on an item, is essential in industrial applications, such as online advertising. Many works focus on user behavior modeling to improve CTR prediction performance. However, most of those methods only model users' positive interests from users' click items while ignoring the context information, which is the display items around the clicks, resulting in inferior performance. In this paper, we highlight the importance of context information on user behavior modeling and propose a novel model named Deep Context Interest Network (DCIN), which integrally models the click and its display context to learn users' context-aware interests. DCIN consists of three key modules: 1) Position-aware Context Aggregation Module (PCAM), which performs aggregation of display items with an attention mechanism; 2) Feedback-Context Fusion Module (FCFM), which fuses the representation of clicks and display contexts through non-linear feature interaction; 3) Interest Matching Module (IMM), which activates interests related with the target item. Moreover, we provide our hands-on solution to implement DCIN on large-scale industrial systems. The significant improvements in both offline and online evaluations demonstrate the superiority of our proposed DCIN method. Notably, DCIN has been deployed on our online advertising system serving the main traffic, which brings 1.5% CTR and 1.5% RPM lift.

#### [Unlocking the Potential of User Feedback: Leveraging Large Language Model as User Simulators to Enhance Dialogue System](#)

Zhiyuan Hu, Yue Feng, Anh Tuan Luu, Bryan Hooi, Aldo Lipani

Dialogue systems and large language models (LLMs) have gained considerable attention. However, the direct utilization of LLMs as task-oriented dialogue (TOD) models has been found to underperform compared to smaller task-specific models. Nonetheless, it is crucial to acknowledge the significant potential of LLMs and explore improved approaches for leveraging their impressive abilities. Motivated by the goal of leveraging LLMs, we propose an alternative approach called User-Guided Response Optimization (UGRO) to combine it with a smaller TOD model. This approach uses LLM as an annotation-free user simulator to assess dialogue responses, combining them with smaller fine-tuned end-to-end TOD models. By utilizing the satisfaction feedback generated by LLMs, UGRO further optimizes the supervised fine-tuned TOD model. Specifically, the TOD model takes the dialogue history as input and, with the assistance of the user simulator's feedback, generates high-satisfaction responses that meet the user's requirements. Through empirical experiments on two TOD benchmarks, we validate the effectiveness of our method. The results demonstrate that our approach outperforms previous state-of-the-art (SOTA) results.

#### [Forgetting-aware Linear Bias for Attentive Knowledge Tracing](#)

Yoonjin Im, Eunseong Choi, Heejin Kook, Jongwuk Lee

Knowledge Tracing (KT) aims to track proficiency based on a question-solving history, allowing us to offer a streamlined curriculum. Recent studies actively utilize attention-based mechanisms to capture the correlation between questions and combine it with the learner's characteristics for responses. However, our empirical study shows that existing attention-based KT models neglect the learner's forgetting behavior, especially as the interaction history becomes longer. This problem arises from the bias that overprioritizes the correlation of questions while inadvertently ignoring the impact of forgetting behavior. This paper proposes a simple-yet-effective solution, namely Forgetting-aware Linear Bias (FoLiBi), to reflect forgetting behavior as a linear bias. Despite its simplicity, FoLiBi is readily equipped with existing attentive KT models by effectively decomposing question correlations with forgetting behavior. FoLiBi plugged with several KT models yields a consistent improvement of up to 2.58% in AUC over state-of-the-art KT models on four benchmark datasets.

#### **Stochastic Subgraph Neighborhood Pooling for Subgraph Classification**

Shweta Ann Jacob, Paul Louis, Amirali Salehi-Abari

Subgraph classification is an emerging field in graph representation learning where the task is to classify a group of nodes (i.e., a subgraph) within a graph (e.g., identifying rare diseases given a collection of phenotypes). Graph neural network (GNN) solutions for node, link, and graph tasks fail to perform well on subgraph classification as they do not capture the external topology of the subgraph (i.e., how the subgraph is located within the larger graph). The current state-of-the-art models address this shortcoming through either labeling tricks or multiple message-passing channels, which are computationally expensive and not scalable to large graphs. To address the scalability issue while maintaining generalization, we propose Stochastic Subgraph Neighborhood Pooling (SSNP), which jointly aggregates the subgraph and its neighborhood (i.e., external topology) information while removing the need for any computationally expensive operations (e.g. labeling tricks). Our extensive experiments demonstrate that SSNP outperforms or is comparable to state-of-the-art methods while being up to 13x faster in runtime.

#### **Lightweight Adaptation of Neural Language Models via Subspace Embedding**

Amit Kumar Jaiswal, Haiming Liu

Traditional neural word embeddings are usually dependent on a richer diversity of vocabulary. However, the language models decline to cover major vocabularies via the word embedding parameters, in particular, for multilingual language models that generally cover a significant part of their overall learning parameters. In this work, we present a new compact embedding structure to reduce the memory footprint of the pre-trained language models with a sacrifice of up to 4% absolute accuracy. The embeddings vectors reconstruction follows a set of subspace embeddings and an assignment procedure via the contextual relationship among tokens from pre-trained language models. The subspace embedding structure<sup>1</sup> calibrates to masked language models, to evaluate our compact embedding structure on similarity and textual entailment tasks, sentence and paraphrase tasks. Our experimental evaluation shows that the subspace embeddings achieve compression rates beyond 99.8% in comparison with the original embeddings for the language models on XNLI and GLUE benchmark suites.

#### **Multi-Granularity Attention Model for Group Recommendation**

Jianye Ji, Jiayan Pei, Shaochuan Lin, Taotao Zhou, Hengxu He, Jia Jia, Ning Hu

Group recommendation provides personalized recommendations to a group of users based on their shared interests, preferences, and characteristics. Current studies have explored different methods for integrating individual preferences and making collective decisions that benefit the group as a whole. However, most of them heavily rely on users with rich behavior and ignore latent preferences of users with relatively sparse behavior, leading to insufficient learning of individual interests. To address this challenge, we present the Multi-Granularity Attention Model (MGAM), a novel approach that utilizes multiple levels of granularity (i.e., subsets, groups, and supersets) to uncover group members' latent preferences and mitigate recommendation noise. Specially, we propose a Subset Preference Extraction module that enhances the representation of users' latent subset-level preferences by incorporating their previous interactions with items and utilizing a hierarchical mechanism. Additionally, our method introduces a Group Preference Extraction module and a Superset Preference Extraction module, which explore users' latent preferences on two levels: the group-level, which maintains users' original preferences, and the superset-level, which includes group-group exterior information. By incorporating the subset-level embedding, group-level embedding, and superset-level embedding, our proposed method effectively reduces group recommendation noise across multiple granularities and comprehensively learns individual interests. Extensive offline and online experiments have demonstrated the superiority of our method in terms of performance.

**CSPM: A Contrastive Spatiotemporal Preference Model for CTR Prediction in On-Demand Food Delivery Services**

Guyu Jiang, Xiaoyun Li, Rongrong Jing, Ruoqi Zhao, Xingliang Ni, Guodong Cao, Ning Hu

Click-through rate (CTR) prediction is a crucial task in the context of an online on-demand food delivery (OFD) platform for precisely estimating the probability of a user clicking on food items. Unlike universal e-commerce platforms such as Taobao and Amazon, user behaviors and interests on the OFD platform are more location and time-sensitive due to limited delivery ranges and regional commodity supplies. However, existing CTR prediction algorithms in OFD scenarios concentrate on capturing interest from historical behavior sequences, which fails to effectively model the complex spatiotemporal information within features, leading to poor performance. To address this challenge, this paper introduces the Contrastive Spatiotemporal Preference Model (CSPM), which disentangles users' spatiotemporal preferences from multiple-field features under different search states using three modules: contrastive spatiotemporal representation learning (CSRL), spatiotemporal preference extractor (StPE), and spatiotemporal information filter (StIF). CSRL utilizes a contrastive learning framework to generate a spatiotemporal activation representation (SAR) for the search action. StPE employs SAR to activate users' diverse preferences related to location and time from the historical behavior sequence field, using a multi-head attention mechanism. StIF incorporates SAR into a gating network to automatically capture important features with latent spatiotemporal effects. Extensive experiments conducted on two large-scale industrial datasets demonstrate the state-of-the-art performance of CSPM. Notably, CSPM has been successfully deployed in Alibaba's online OFD platform Ele.me, resulting in a significant 0.88% lift in CTR, which has substantial business implications.

**Uncertainty Quantification via Spatial-Temporal Tweedie Model for Zero-inflated and Long-tail Travel Demand Prediction**

Xinke Jiang, Dingyi Zhuang, Xianghui Zhang, Hao Chen, Jiayuan Luo, Xiaowei Gao

Understanding Origin-Destination (O-D) travel demand is crucial for transportation management. However, traditional spatial-temporal deep learning models grapple with addressing the sparse and long-tail characteristics in high-resolution O-D matrices and quantifying prediction uncertainty. This dilemma arises from the numerous zeros and over-dispersed demand patterns within these matrices, which challenge the Gaussian assumption inherent to deterministic deep learning models. To address these challenges, we propose a novel approach: the Spatial-Temporal Tweedie Graph Neural Network (STTD). The STTD introduces the Tweedie distribution as a compelling alternative to the traditional 'zero-inflated' model and leverages spatial and temporal embeddings to parameterize travel demand distributions. Our evaluations using real-world datasets highlight STTD's superiority in providing accurate predictions and precise confidence intervals, particularly in high-resolution scenarios. GitHub code is available online(<https://github.com/STTDAnonymous/STTD>).

**MSRA: A Multi-Aspect Semantic Relevance Approach for E-Commerce via Multimodal Pre-Training**

Hanqi Jin, Jiwei Tan, Lixin Liu, Lisong Qiu, Shaowei Yao, Xi Chen, Xiaoyi Zeng

To enhance the effectiveness of matching user requests with millions of online products, practitioners invest significant efforts in developing semantic relevance models on large-scale e-commerce platforms. Generally, such semantic relevance models are formulated as text-matching approaches, which measure the relevance between users' search queries and the titles of candidate items (i.e., products). However, we argue that conventional relevance methods may lead to sub-optimal performance due to the limited information provided by the titles of candidate items. To alleviate this issue, we suggest incorporating additional information about candidate items from multiple aspects, including their attributes and images. This could supplement the information that may not be fully provided by titles alone. To this end, we propose a multi-aspect semantic relevance model that takes into account the match between search queries and the title, attribute and image information of items simultaneously. The model is further enhanced through pre-training using several well-designed self-supervised and weakly-supervised tasks. Furthermore, the proposed model is fine-tuned using annotated data and distilled into a representation-based architecture for efficient online deployment. Experimental results show the proposed approach significantly improves relevance and leads to considerable enhancements in business metrics.

**SAFE: Sequential Attentive Face Embedding with Contrastive Learning for Deepfake Video Detection**

Juho Jung, Chaewon Kang, Jeewoo Yoon, Simon S. Woo, Jinyoung Han

The emergence of hyper-realistic deepfake videos has raised significant concerns regarding their potential misuse. However, prior research on deepfake detection has primarily focused on image-based approaches, with little emphasis on video. With the advancement of generation techniques enabling intricate and dynamic manipulation of entire faces as well as specific facial components in a video sequence, capturing dynamic changes in both global and local facial features becomes crucial in detecting deepfake videos. This paper proposes a novel sequential attentive face embedding, SAFE, that can capture

facial dynamics in a deepfake video. The proposed SAFE can effectively integrate global and local dynamics of facial features revealed in a video sequence using contrastive learning. Through a comprehensive comparison with the state-of-the-art methods on the DFDC (Deepfake Detection Challenge) dataset and the FaceForensic++ benchmark, we show that our model achieves the highest accuracy in detecting deepfake videos on both datasets.

#### Effective Slogan Generation with Noise Perturbation

Jongeun Kim, MinChung Kim, Taehwan Kim

Slogans play a crucial role in building the brand's identity of the firm. A slogan is expected to reflect firm's vision and the brand's value propositions in memorable and likeable ways. Automating the generation of slogans with such characteristics is challenging. Previous studies developed and tested slogan generation with syntactic control and summarization models which are not capable of generating distinctive slogans. We introduce a novel approach that leverages pre-trained transformer T5 model with noise perturbation on newly proposed 1:N matching pair dataset. This approach serves as a contributing factor in generating distinctive and coherent slogans. Furthermore, the proposed approach incorporates descriptions about the firm and brand into the generation of slogans. We evaluate generated slogans based on ROUGE-1, ROUGE-L and Cosine Similarity metrics and also assess them with human subjects in terms of slogan's distinctiveness, coherence, and fluency. The results demonstrate that our approach yields better performance than baseline models and other transformer-based models.

#### S-Mixup: Structural Mixup for Graph Neural Networks

Junghurn Kim, Sukwon Yun, Chanyoung Park

Existing studies for applying the mixup technique on graphs mainly focus on graph classification tasks, while the research in node classification is still under-explored. In this paper, we propose a novel mixup augmentation for node classification called Structural Mixup (S-Mixup). The core idea is to take into account the structural information while mixing nodes. Specifically, S-Mixup obtains pseudo-labels for unlabeled nodes in a graph along with their prediction confidence via a Graph Neural Network (GNN) classifier. These serve as the criteria for the composition of the mixup pool for both inter and intra-class mixups. Furthermore, we utilize the edge gradient obtained from the GNN training and propose a gradient-based edge selection strategy for selecting edges to be attached to the nodes generated by the mixup. Through extensive experiments on real-world benchmark datasets, we demonstrate the effectiveness of S-Mixup evaluated on the node classification task. We observe that S-Mixup enhances the robustness and generalization performance of GNNs, especially in heterophilous situations. The source code of S-Mixup can be found at <https://github.com/SukwonYun/S-Mixup>

#### Class Label-aware Graph Anomaly Detection

Junghoon Kim, Yeonjun In, Kanghoon Yoon, Junmo Lee, Chanyoung Park

Unsupervised GAD methods assume the lack of anomaly labels, i.e., whether a node is anomalous or not. One common observation we made from previous unsupervised methods is that they not only assume the absence of such anomaly labels, but also the absence of class labels (the class a node belongs to used in a general node classification task). In this work, we study the utility of class labels for unsupervised GAD; in particular, how they enhance the detection of structural anomalies. To this end, we propose a Class Label-aware Graph Anomaly Detection framework (CLAD) that utilizes a limited amount of labeled nodes to enhance the performance of unsupervised GAD. Extensive experiments on ten datasets demonstrate the superior performance of CLAD in comparison to existing unsupervised GAD methods, even in the absence of ground-truth class label information. The source code for CLAD is available at <https://github.com/jhkim611/CLAD>.

#### Exploring Cohesive Subgraphs in Hypergraphs: The (k,g)-core Approach

Dahee Kim, Junghoon Kim, Sungsu Lim, Hyun Ji Jeong

Identifying cohesive subgraphs in hypergraphs is a fundamental problem that has received recent attention in data mining and engineering fields. Existing approaches mainly focus on a strongly induced subhypergraph or edge cardinality, overlooking the importance of the frequency of co-occurrence. In this paper, we propose a new cohesive subgraph named (k,g)-core, which considers both neighbour and co-occurrence simultaneously. The (k,g)-core has various applications including recommendation system, network analysis, and fraud detection. To the best of our knowledge, this is the first work to combine these factors. We extend an existing efficient algorithm to find solutions for (k,g)-core. Finally, we conduct extensive experimental studies that demonstrate the efficiency and effectiveness of our proposed algorithm.

### **Can a Chatbot be Useful in Childhood Cancer Survivorship? Development of a Chatbot for Survivors of Childhood Cancer**

Mirae Kim, Kyubum Hwang, Hayoung Oh, Heejin Kim, Min Ah Kim

This study introduces an informational and empathetic chatbot for childhood cancer survivors. As the survival rates for childhood cancer around the world have increased, survivors often face various psychosocial challenges during and after cancer treatment. However, they rarely seek support from psychosocial professionals due to the low availability of resources and stigma toward cancer survivors in countries like South Korea. This study aimed to develop a chatbot tailored to the unique characteristics of childhood cancer survivors in need of informational and emotional support. Given the limited availability of empirical data on childhood cancer survivors, quotes from survivors were gathered from academic articles and social media, then large language models were employed to generate appropriate responses. Furthermore, we incorporated domain learning techniques to ensure a more tailored and suitable model for addressing the needs of survivors.

### **Test-Time Embedding Normalization for Popularity Bias Mitigation**

Dain Kim, Jinhyeok Park, Dongwoo Kim

Popularity bias is a widespread problem in the field of recommender systems, where popular items tend to dominate recommendation results. In this work, we propose 'Test Time Embedding Normalization' as a simple yet effective strategy for mitigating popularity bias, which surpasses the performance of the previous mitigation approaches by a significant margin. Our approach utilizes the normalized item embedding during the inference stage to control the influence of embedding magnitude, which is highly correlated with item popularity. Through extensive experiments, we show that our method combined with the sampled softmax loss effectively reduces popularity bias compare to previous approaches for bias mitigation. We further investigate the relationship between user and item embeddings and find that the angular similarity between embeddings distinguishes preferable and non-preferable items regardless of their popularity. The analysis explains the mechanism behind the success of our approach in eliminating the impact of popularity bias. Our code is available at <https://github.com/ml-postech/TTEN>.

### **AmpliBias: Mitigating Dataset Bias through Bias Amplification in Few-shot Learning for Generative Models**

Donggeun Ko, Dongjun Lee, Namjun Park, Kyoungrae Noh, Hyeonjin Park, Jaekwang Kim

Deep learning models exhibit a dependency on peripheral attributes of input data, such as shapes and colors, leading the models to become biased towards these certain attributes that result in subsequent degradation of performance. In this paper, we alleviate this problem by presenting  $\sim$ sysname, a novel framework that tackles dataset bias by leveraging generative models to amplify bias and facilitate the learning of debiased representations of the classifier. Our method involves three major steps. We initially train a biased classifier, denoted as  $f_b$ , on a biased dataset and extract the top-K biased-conflict samples. Next, we train a generator solely on a bias-conflict dataset comprised of these top-K samples, aiming to learn the distribution of bias-conflict samples. Finally, we re-train the classifier on the newly constructed debiased dataset, which combines the original and amplified data. This allows the biased classifier to competently learn debiased representation. Extensive experiments validate that our proposed method effectively debiases the biased classifier.

### **You're Not Alone in Battle: Combat Threat Analysis Using Attention Networks and a New Open Benchmark**

Soo Yong Lee, Juwon Kim, Kiwoong Park, Dong Kuk Ryu, Sangheun Shim, Kijung Shin

For military commands, combat threat analysis is crucial in predicting future outcomes and informing consequent decisions. Its primary objectives include determining the intention and attack likelihood of the hostiles. The complex, dynamic, and noisy nature of combat, however, presents significant challenges in its analysis. The prior research has been limited in accounting for such characteristics, assuming independence of each entity, no unobserved tactics, and clean combat data. As such, we present spatio-temporal attention for threat analysis (SAFETY) to encode complex interactions that arise within combat. We test the model performance for unobserved tactics and with various perturbations. To do so, we also present the first open-source benchmark for combat threat analysis with two downstream tasks of predicting entity intention and attack probability. Our experiments show that SAFETY achieves a significant improvement in model performance, with enhancements of up to 13% in intention prediction and 7% in attack prediction compared to the strongest competitor, even when confronted with noisy or missing data. This result highlights the importance of encoding dynamic interactions among entities for combat threat analysis. Our codes and dataset are available at <https://github.com/syleeheal/SAFETY>.

### **Look At Me, No Replay! SurpriseNet: Anomaly Detection Inspired Class Incremental Learning**

Anton Lee, Yaqian Zhang, Heitor Murilo Gomes, Albert Bifet, Bernhard Pfahringer



Continual learning aims to create artificial neural networks capable of accumulating knowledge and skills through incremental training on a sequence of tasks. The main challenge of continual learning is catastrophic interference, wherein new knowledge overrides or interferes with past knowledge, leading to forgetting. An associated issue is the problem of learning "cross-task knowledge," where models fail to acquire and retain knowledge that helps differentiate classes across task boundaries. A common solution to both problems is "replay," where a limited buffer of past instances is utilized to learn cross-task knowledge and mitigate catastrophic interference. However, a notable drawback of these methods is their tendency to overfit the limited replay buffer. In contrast, our proposed solution, SurpriseNet, addresses catastrophic interference by employing a parameter isolation method and learning cross-task knowledge using an auto-encoder inspired by anomaly detection. SurpriseNet is applicable to both structured and unstructured data, as it does not rely on image-specific inductive biases. We have conducted empirical experiments demonstrating the strengths of SurpriseNet on various traditional vision continual-learning benchmarks, as well as on structured data datasets. Source code made available at <https://doi.org/10.5281/zenodo.8247906> and <https://github.com/tachyonicClock/SurpriseNet-CIKM-23>

#### **UNDO: Effective and Accurate Unlearning Method for Deep Neural Networks**

Sangyong Lee, Simon S. Woo

Machine learning has evolved through extensive data usage, including personal and private information. Regulations like GDPR highlight the "Right to be forgotten" for user and data privacy. Research in machine unlearning aims to remove specific data from pre-trained models. We introduce a novel two-step unlearning method, UNDO. First, we selectively disrupt the decision boundary of forgetting data at the coarse-grained level. However, this can also inadvertently affect the decision boundary of other remaining data, lowering the overall performance of classification task. Hence, we subsequently repair and refining the decision boundary for each class at the fine-grained level by introducing a loss for maintain the overall performance, while completely removing the class. Our approach is validated through experiments on two datasets, outperforming other methods in effectiveness and efficiency.

#### **MvFS: Multi-view Feature Selection for Recommender System**

Youngjune Lee, Yeongjong Jeong, Keunchan Park, SeongKu Kang

Feature selection, which is a technique to select key features in recommender systems, has received increasing research attention. Recently, Adaptive Feature Selection (AdaFS) has shown remarkable performance by adaptively selecting features for each data instance, considering that the importance of a given feature field can vary significantly across data. However, this method still has limitations in that its selection process could be easily biased to major features that frequently occur. To address these problems, we propose Multi-view Feature Selection (MvFS), which selects informative features for each instance more effectively. Most importantly, MvFS employs a multi-view network consisting of multiple sub-networks, each of which learns to measure the feature importance of a part of data with different feature patterns. By doing so, MvFS mitigates the bias problem towards dominant patterns and promotes a more balanced feature selection process. Moreover, MvFS adopts an effective importance score modeling strategy which is applied independently to each field without incurring dependency among features. Experimental results on real-world datasets demonstrate the effectiveness of MvFS compared to state-of-the-art baselines.

#### **ST-RAP: A Spatio-Temporal Framework for Real Estate Appraisal**

Hojoon Lee, Hawon Jeong, Byungkun Lee, Kyungyup Daniel Lee, Jaegul Choo

In this paper, we introduce ST-RAP, a novel Spatio-Temporal framework for Real estate Appraisal. ST-RAP employs a hierarchical architecture with a heterogeneous graph neural network to encapsulate temporal dynamics and spatial relationships simultaneously. Through comprehensive experiments on a large-scale real estate dataset, ST-RAP outperforms previous methods, demonstrating the significant benefits of integrating spatial and temporal aspects in real estate appraisal. Our code and dataset are available at <https://github.com/dojeon-ai/STRAP>.

#### **Temporal and Topological Augmentation-based Cross-view Contrastive Learning Model for Temporal Link Prediction**

Dongyuan Li, Shiyin Tan, Yusong Wang, Kotaro Funakoshi, Manabu Okumura

With the booming development of social media, temporal link prediction (TLP), as a core technology, has been receiving increasing attention. However, current methods are based on graph neural networks, which suffer from the over-smoothing issue and easily yield indistinguishable node representations, degrading the prediction accuracy. Besides, they lack the ability to eliminate noisy temporal information and ignore the importance of high-order neighbor information for measuring the link probability between nodes. To solve these issues, we design a cross-view graph contrastive learning (GCL) framework for

TLP, called Tacl. We first design two augmented views for GCL by enhancing the temporal and topological information to obtain distinguishable node representations. Then, we learn the evolution rule of temporal networks to help constrain consistency of node representations and eliminate noise. Finally, we incorporate the high-order neighbor information to measure the link probability between nodes. Extensive experiments demonstrate the effectiveness and robustness of Tacl.

#### **HEPT Attack: Heuristic Perpendicular Trial for Hard-label Attacks under Limited Query Budgets**

Qi Li, Xingyu Li, Xiaodong Cui, Keke Tang, Peican Zhu

Exploring adversarial attacks on deep neural networks (DNNs) is crucial for assessing and enhancing their adversarial robustness. Among various attack types, hard-label attacks that rely only on predicted labels offer a practical approach. This paper focuses on the challenging task of hard-label attacks within an extremely limited query budget, which is a significant achievement rarely accomplished by existing methods. To tackle this, we propose an attack framework that leverages geometric information from previous perturbation directions to form triangles and employs a heuristic perpendicular trial to effectively utilize the intermediate directions. Extensive experiments validate the effectiveness of our approach under strict query constraints and demonstrate its superiority to the state-of-the-art methods.

#### **CORD: A Three-Stage Coarse-to-Fine Framework for Relation Detection in Knowledge Base Question Answering**

Yanzeng Li, Sen Hu, Wenjuan Han, Lei Zou

As a fundamental subtask of Knowledge Base Question Answering (KBQA), Relation Detection (KBQA-RD) plays a crucial role to detect the KB relations between entities or variables in natural language questions. It remains, however, a challenging task, particularly for significant large-scale relations and in the presence of easily confused relations. Recent state-of-the-art methods not only struggle with such scenarios, but often take into account only one facet and fail to incorporate the subtle discrepancy among the relations. In this paper, we propose a simple and efficient three-stage framework to exploit the coarse-to-fine paradigm. Specifically, we employ a natural clustering over all KB relations and perform a coarse-to-fine relation recognition process based on the relation clustering. In this way, our framework (i.e., CORD) refines the detection of relations, so as to scale well with large-scale relations. Experiments on both single-relation (i.e., SimpleQuestions (SQ)) and multi-relation (i.e., WebQSP (WQ)) benchmarks show that CORD not only achieves the outstanding relation detection performance in KBQA-RD subtask; but more importantly, further improves the accuracy of KBQA systems.

#### **Causal Discovery in Temporal Domain from Interventional Data**

Peiwen Li, Yuan Meng, Xin Wang, Fang Shen, Yue Li, Jialong Wang, Wenwu Zhu

Causal learning from observational data has garnered attention as controlled experiments can be costly. To enhance identifiability, incorporating intervention data has become a mainstream approach. However, these methods have yet to be explored in the context of time series data, despite their success in static data. To address this research gap, this paper presents a novel contribution. Firstly, a temporal interventional dataset with causal labels is introduced, derived from a data center IT room of a cloud service company. Secondly, this paper introduces TECDI, a novel approach for temporal causal discovery. TECDI leverages the smooth, algebraic characterization of acyclicity in causal graphs to efficiently uncover causal relationships. Experimental results on simulated and proposed real-world datasets validate the effectiveness of TECDI in accurately uncovering temporal causal relationships. The introduction of the temporal interventional dataset and the superior performance of TECDI contribute to advancing research in temporal causal discovery. Our datasets and codes have released at <https://github.com/lpwwpower/TECDI> <https://github.com/lpwwpower/TECDI>.

#### **Pseudo Triplet Networks for Classification Tasks with Cross-Source Feature Incompleteness**

Cayon Liow, Cheng-Te Li, Chun-Pai Yang, Shou-De Lin

Cross-source feature incompleteness -- a scenario where certain features are only available in one data source but missing in another -- is a common and significant challenge in machine learning. It typically arises in situations where the training data and testing data are collected from different sources with distinct feature sets. Addressing this challenge has the potential to greatly improve the utility of valuable datasets that might otherwise be considered incomplete and enhance model performance. This paper introduces the novel Pseudo Triplet Network (PTN) to address cross-source feature incompleteness. PTN fuses two Siamese network architectures -- Triplet Networks and Pseudo Networks. By segregating data into instance, positive, and negative subsets, PTN facilitates effectively contrastive learning through a hybrid loss function design. The model was rigorously evaluated on six benchmark datasets from the UCI Repository, in comparison with five other methods for managing missing data, under a range of feature overlap and missing data scenarios. The PTN

consistently exhibited superior performance, displaying resilience in high missing ratio situations and maintaining robust stability across various data scenarios.

### Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction

Mutong Liu, Yang Liu, Jiming Liu

Infectious disease risk prediction plays a vital role in disease control and prevention. Recent studies in machine learning have attempted to incorporate epidemiological knowledge into the learning process to enhance the accuracy and informativeness of prediction results for decision-making. However, these methods commonly involve single-patch mechanistic models, overlooking the disease spread across multiple locations caused by human mobility. Additionally, these methods often require extra information beyond the infection data, which is typically unavailable in reality. To address these issues, this paper proposes a novel epidemiology-aware deep learning framework that integrates a fundamental epidemic component, the next-generation matrix (NGM), into the deep architecture and objective function. This integration enables the inclusion of both mechanistic models and human mobility in the learning process to characterize within- and cross-location disease transmission. From this framework, two novel methods, Epi-CNNRNN-Res and Epi-Cola-GNN, are further developed to predict epidemics, with experimental results validating their effectiveness.

### Towards Trustworthy Rumor Detection with Interpretable Graph Structural Learning

Leyuan Liu, Junyi Chen, Zhangtao Cheng, Wenxin Tai, Fan Zhou

The exponential growth of digital information has amplified the necessity for effective rumor detection on social media. However, existing approaches often neglect the inherent noise and uncertainty in rumor propagation, leading to obscure learning mechanisms. Moreover, current deep-learning methodologies, despite their top-tier performance, are heavily dependent on supervised learning, which is labor-intensive and inefficient. Their prediction credibility is also questionable. To tackle these issues, we present a new framework, TrustRD, for reliable rumor detection. Our framework incorporates a self-supervised learning module, designed to derive interpretable and informative representations with less reliance on large labeled data sets. A downstream model based on Bayesian networks, which is further refined with adversarial training, enhances performance while providing a quantifiable trustworthiness assessment of results. Our methods' effectiveness is confirmed through experiments on two benchmark datasets.

### Homogeneous Cohort-Aware Group Cognitive Diagnosis: A Multi-grained Modeling Perspective

Shuhuan Liu, Xiaoshan Yu, Haiping Ma, Ziwen Wang, Chuan Qin, Xingyi Zhang

Cognitive Diagnosis has been widely investigated as a fundamental task in the field of education, aiming at effectively assessing the students' knowledge proficiency level by mining their exercise records. Recently, group-level cognitive diagnosis is also attracting attention, which measures the group-level knowledge proficiency on specific concepts by modeling the response behaviors of all students within the classes. However, existing work tends to explore group characteristics with a coarse-grained perspective while ignoring the inter-individual variability within groups, which is prone to unstable diagnosis results. To this end, in this paper, we propose a novel Homogeneous cohort-aware Group Cognitive Diagnosis model, namely HomoGCD, to effectively model the group's knowledge proficiency level from a multi-grained modeling perspective. Specifically, we first design a homogeneous cohort mining module to explore subgroups of students with similar ability status within a class by modeling their routine exercising performance. Then, we construct the mined cohorts into fine-grained organizations for exploring stable and uniformly distributed features of groups. Subsequently, we develop a multi-grained modeling module to comprehensively learn the cohort and group ability status, which jointly trains both interactions with the exercises. In particular, an extensible diagnosis module is introduced to support the incorporation of different diagnosis functions. Finally, extensive experiments on two real-world datasets clearly demonstrate the generality and effectiveness of our HomoGCD in group as well as cohort~assessments.

### Retrieval-Based Unsupervised Noisy Label Detection on Text Data

Peiyang Liu, Jinyu Yang, Lin Wang, Sen Wang, Yunlai Hao, Huihui Bai

The success of deep neural networks hinges on both high-quality annotations and copious amounts of data; however, in practice, a compromise between dataset size and quality frequently arises. Data collection and cleansing are often resource-intensive and time-consuming, leading to real-world datasets containing label noise that can introduce incorrect correlation patterns, adversely affecting model generalization capabilities. The efficient identification of corrupted patterns is indispensable, with prevalent methods predominantly concentrating on devising robust training techniques to preclude models from internalizing these patterns. Nevertheless, these supervised approaches often necessitate tailored training

procedures, potentially resulting in overfitting corrupted patterns and a decline in detection performance. This paper presents a retrieval-based unsupervised solution for the detection of noisy labels, surpassing the performance of three current competitive methods in this domain.

### **Boosting Meta-Learning Cold-Start Recommendation with Graph Neural Network**

Han Liu, Hongxiang Lin, Xiaotong Zhang, Fenglong Ma, Hongyang Chen, Lei Wang, Hong Yu, Xianchao Zhang

Meta-learning methods have shown to be effective in dealing with cold-start recommendation. However, most previous methods rely on an ideal assumption that there exists a similar data distribution between source and target tasks, which are unsuitable for the scenario that only extremely limited number of new user or item interactions are available. In this paper, we propose to boost meta-learning cold-start recommendation with graph neural network (MeGNN). First, it utilizes the global neighborhood translation learning to obtain consistent potential interactions for all new user and item nodes, which can refine their representations. Second, it employs the local neighborhood translation learning to predict specific potential interactions for each node, thus guaranteeing the personalized requirement. In experiments, we combine MeGNN with two representative meta-learning models MeLU and TaNP. Extensive results on two widely-used datasets show the superiority of MeGNN in four different scenarios.

### **Understanding the Multi-vector Dense Retrieval Models**

Qi Liu, Jiaxin Mao

While dense retrieval has become a promising alternative to the traditional text retrieval models, such as BM25, some recent studies show that multi-vector dense retrieval models are more effective than the single-vector method in retrieval tasks. However, due to a lack of interpretability, why the multi-vector method outperforms its single-vector counterpart has not been fully studied. To fill this research gap, in this work, we investigate and compare the behaviors of single-vector and multi-vector models in retrieval. Specifically, we analyze the vocabulary distribution of dense representations by mapping them back to the sparse, vocabulary space. Our empirical findings show that the multi-vector representation has more lexical overlaps between queries and passages. Additionally, we show that this feature of multi-vector representation can enhance its ranking performance when a given passage can fulfill different information needs and thus can be retrieved by different queries. These results shed light on the internal mechanisms of multi-vector representation and may provide new perspectives for future research.

### **Counterfactual Adversarial Learning for Recommendation**

Jialin Liu, Zijian Zhang, Xiangyu Zhao, Jun Li

Long-term user responses, i.e., clicks or purchases on e-commerce platforms, are crucial for sequential recommender systems. Recent off-policy evaluation methods involve these responses by simultaneously maximizing expected cumulative rewards. However, two aspects of these methods require further consideration. Firstly, from the system's point of view, candidates with various values are interchangeable, which may result in contradictory future recommendations despite having the same interaction history. Secondly, rewards are manually designed, which necessitates a trial-and-error approach to strike a balance between training stabilization and reward distinction. To address these issues, we propose a new sequential recommender system called NCM4Rec. Specifically, for the distinction problem, NCM4Rec achieves counterfactual consistency via a neural causal model, which is learnable yet equally expressive as classic structural causal models. Such consistency is maintained by a Gumbel-Max design. For the representing problem, NCM4Rec encodes different types of responses as one-hot vectors and captures the long-term preference via adversarial learning. As a consequence, NCM4Rec is both adaptive and identifiable. Both theoretical analyses of the consistency and empirical studies over two real-world datasets demonstrate the effectiveness of our method.

### **STGIN: Spatial-Temporal Graph Interaction Network for Large-scale POI Recommendation**

Shaohua Liu, Yu Qi, Gen Li, Mingjian Chen, Teng Zhang, Jia Cheng, Jun Lei

In Location-Based Services, Point-Of-Interest(POI) recommendation plays a crucial role in both user experience and business opportunities. Graph neural networks have been proven effective in providing personalized POI recommendation services. However, there are still two critical challenges. First, existing graph models attempt to capture users' diversified interests through a unified graph, which limits their ability to express interests in various spatial-temporal contexts. Second, the efficiency limitations of graph construction and graph sampling in large-scale systems make it difficult to adapt quickly to new real-time interests. To tackle the above challenges, we propose a novel Spatial-Temporal Graph Interaction Network. Specifically, we construct subgraphs of spatial, temporal, spatial-temporal, and global views respectively to precisely

characterize the user's interests in various contexts. In addition, we design an industry-friendly framework to track the user's latest interests. Extensive experiments on the real-world dataset show that our method outperforms state-of-the-art models. This work has been successfully deployed in a large e-commerce platform, delivering a 1.1% CTR and 6.3% RPM improvement.

### **Spatio-Temporal Adaptive Embedding Makes Vanilla Transformer SOTA for Traffic Forecasting**

Hangchen Liu, Zheng Dong, Renhe Jiang, Jiewen Deng, Jinliang Deng, Quanjun Chen, Xuan Song

With the rapid development of the Intelligent Transportation System (ITS), accurate traffic forecasting has emerged as a critical challenge. The key bottleneck lies in capturing the intricate spatio-temporal traffic patterns. In recent years, numerous neural networks with complicated architectures have been proposed to address this issue. However, the advancements in network architectures have encountered diminishing performance gains. In this study, we present a novel component called spatio-temporal adaptive embedding that can yield outstanding results with vanilla transformers. Our proposed Spatio-Temporal Adaptive Embedding transformer (STAEformer) achieves state-of-the-art performance on five real-world traffic forecasting datasets. Further experiments demonstrate that spatio-temporal adaptive embedding plays a crucial role in traffic forecasting by effectively capturing intrinsic spatio-temporal relations and chronological information in traffic time series.

### **TemDep: Temporal Dependency Priority for Multivariate Time Series Prediction**

Shu Liu, Jiaheng Wang, Jiamin Chen, Jianliang Gao, Yuhui Zhong

The multivariate fusion transformation is ubiquitous in multivariate time series prediction (MTSP) problems. The previous multivariate fusion transformation fuses the feature of different variates at a time step, then projects them to a new feature space for effective feature representation. However, temporal dependency is the most fundamental property of time series. The previous manner fails to capture the temporal dependency of the feature, which is destroyed in the transformed feature matrix. Multivariate feature extraction based on the feature matrix with missing temporal dependency leads to the loss of predictive performance of MTSP. To address this problem, we propose the Temporal Dependency Priority for Multivariate Time Series Prediction (TemDep) method. Specifically, TemDep extracts feature temporal dependency of multivariate time series first and then considers multivariate feature fusion. Moreover, the low-dimensional and high-dimensional feature fusion manners are designed with the temporal dependency priority to fit different dimensional multivariate time series. The extensive experimental results of different datasets show that our proposed method can outperform all state-of-the-art baseline methods. It proves the significance of temporal dependency priority for MTSP.

### **FairGraph: Automated Graph Debiasing with Gradient Matching**

Yezi Liu

As a prevalence data structure in the real world, graphs have found extensive applications ranging from modeling social networks to molecules. However, the existence of diverse biases within graphs gives rise to unfair representations learned by graph neural networks (GNNs). Addressing this issue has typically been approached from a modeling perspective, which not only compromises the integrity of the model structure but also entails additional effort and cost for retraining model parameters when the architecture changes. In this study, we adopt a data-centric standpoint to tackle the problem of fairness, focusing on graph debiasing for Graph Neural Networks. Our specific objective is to eliminate various biases from the input graph by generating a fair synthetic graph. By training GNNs on this fair graph, we aim to achieve an optimal accuracy-fairness trade-off. To this end, we propose FairGraph, which approaches the graph debiasing problem by mimicking the GNN training trajectory of the input graph through an optimization process involving a gradient-matching loss and fairness constraints. Through extensive experiments conducted on three benchmark datasets, we demonstrate the effectiveness of FairGraph and its ability to automatically generate fair graphs that are transferable across different GNN architectures.

### **Personalized Differentially Private Federated Learning without Exposing Privacy Budgets**

Junxu Liu, Jian Lou, Li Xiong, Xiaofeng Meng

The meteoric rise of cross-silo Federated Learning (FL) is due to its ability to mitigate data breaches during collaborative training. To further provide rigorous privacy protection with consideration of the varying privacy requirements across different clients, a privacy-enhanced line of work on personalized differentially private federated learning (PDP-FL) has been proposed. However, the existing solution for PDP-FL [20] assumes the raw privacy budgets of all clients should be collected by the server. These values are then directly utilized to improve the model utility via facilitating the privacy preferences partitioning (i.e., partitioning all clients into multiple privacy groups). It is however non-realistic because the raw privacy budgets can be quite informative and sensitive.



In this work, our goal is to achieve PDP-FL without exposing clients' raw privacy budgets by indirectly partitioning the privacy preferences solely based on clients' noisy model updates. The crux lies in the fact that the noisy updates could be influenced by two entangled factors of DP noises and non-IID clients' data, leaving it unknown whether it is possible to uncover privacy preferences by disentangling the two affecting factors. To overcome the hurdle, we systematically investigate the unexplored question of under what conditions can the model updates of clients be primarily influenced by noise levels rather than data distribution. Then, we propose a simple yet effective strategy based on clustering the L2 norm of the noisy updates, which can be integrated into the vanilla PDP-FL to maintain the same performance. Experimental results demonstrate the effectiveness and feasibility of our privacy-budget-agnostic PDP-FL method.

#### [Personalized Interest Sustainability Modeling for Sequential POI Recommendation](#)

Zewen Long, Liang Wang, Qiang Liu, Shu Wu

Sequential point-of-interest (POI) recommendation endeavors to capture users' dynamic interests based on their historical check-ins, subsequently predicting the next POIs that they are most likely to visit. Existing methods conventionally capture users' personalized dynamic interests from their chronological sequences of visited POIs. However, these methods fail to explicitly consider personalized interest sustainability, which means whether each user's interest in specific POIs will sustain beyond the training time. In this work, we propose a personalized Interest Sustainability modeling framework for sequential POI REcommendation, INSPIRE for brevity. Different from existing methods that directly recommend next POIs through users' historical trajectories, our proposed INSPIRE focuses on users' personalized interest sustainability. Specifically, we first develop a new task to predict whether each user will visit the POIs in the recent period of the training time. Afterwards, to remedy the sparsity issue of users' check-in history, we propose to augment users' check-in history in three ways: geographical, intrinsic, and extrinsic schemes. Extensive experiments are conducted on two real-world datasets and results show that INSPIRE outperforms existing next POI solutions.

#### [Can Embeddings Analysis Explain Large Language Model Ranking?](#)

Claudio Lucchese, Giorgia Minello, Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, Alberto Veneri

Understanding the behavior of deep neural networks for Information Retrieval (IR) is crucial to improve trust in these effective models. Current popular approaches to diagnose the predictions made by deep neural networks are mainly based on: i) the adherence of the retrieval model to some axiomatic property of the IR system, ii) the generation of free-text explanations, or iii) feature importance attributions. In this work, we propose a novel approach that analyzes the changes of document and query embeddings in the latent space and that might explain the inner workings of IR large pre-trained language models. In particular, we focus on predicting query/document relevance, and we characterize the predictions by analyzing the topological arrangement of the embeddings in their latent space and their evolution while passing through the layers of the network. We show that there exists a link between the embedding adjustment and the predicted score, based on how tokens cluster in the embedding space. This novel approach, grounded in the query and document tokens interplay over the latent space, provides a new perspective on neural ranker explanation and a promising strategy for improving the efficiency of the models and Query Performance Prediction (QPP).

#### [DCGNN: Dual-Channel Graph Neural Network for Social Bot Detection](#)

Nuoyan Lyu, Bingbing Xu, Fangda Guo, Huawei Shen

The importance of social bot detection has been increasingly recognized due to its profound impact on information dissemination. Existing methodologies can be categorized into feature engineering and deep learning-based methods, which mainly focus on static features, e.g., post characteristics and user profiles. However, existing methods often overlook the burst phenomena when distinguishing social bots and genuine users, i.e., the sudden and intense activity or behavior of bots after prolonged inter. Through comprehensive analysis, we find that both burst behavior and static features play pivotal roles in social bot detection. To capture such properties, the dual-channel GNN (DCGNN) is proposed which consists of a burst-aware channel with an adaptive-pass filter and a static-aware channel with a low-pass filter to model user characteristics effectively. Experimental results demonstrate the superiority of this method over competitive baselines.

#### [Contrastive Learning for Rumor Detection via Fitting Beta Mixture Model](#)

Jiachen Ma, Jing Dai, Yong Liu, Meng Han, Chunyu Ai

The rise of social media has posed a challenging problem of effectively identifying rumors. With the great success of contrastive learning in many fields, many contrastive learning models for rumor detection have been proposed. However, existing models usually use the propagation structure of other events as negative samples and regard more similar samples

to anchor events as hard ones across all the training processes, resulting in undesirably pushing away the samples of the same class. Thus, we propose a novel contrastive learning model (CRFB) to solve the above problem. Specifically, we employ contrastive learning between two augmented propagation structure and fit a two-component (true-false) beta mixture model (BMM) to measure the probability of negative samples being true. In addition, we propose a CNN-based model to capture the consistent and complementary information between two augmented propagation structure. The experimental results on public datasets demonstrate that our CRFB outperforms the existing state-of-the-art models for rumor detection.

### [Age-Aware Guidance via Masking-Based Attention in Face Aging](#)

Junyeong Maeng, Kwanseok Oh, Heung-II Suk

Face age transformation aims to convert reference images into synthesized images so that they portray the specified target ages. The crux of this task is to change only age-related areas of the given image while maintaining the age-irrelevant areas unchanged. Nevertheless, a common limitation among most existing models is the struggle to generate high-quality aging images that effectively consider both crucial properties. To address this problem, we propose a novel GAN-based face-aging framework that utilizes age-aware Guidance via Masking-Based Attention (GMBA). Specifically, we devise an age-aware guidance module to adjust age-relevant and age-irrelevant attributes within the image seamlessly. By virtue of its capability, it enables the model to produce realistic age-transformed images that certainly preserve the input's identities while delicately imposing age-related properties. Experimental results show that our proposed GMBA outperformed other state-of-the-art methods in terms of identity preservation and accurate age conversion, as well as providing superior visual quality for age-transformed images.

### [Generating News-Centric Crossword Puzzles As A Constraint Satisfaction and Optimization Problem](#)

Kaito Majima, Shotaro Ishihara

Crossword puzzles have traditionally served not only as entertainment but also as an educational tool that can be used to acquire vocabulary and language proficiency. One strategy to enhance the educational purpose is personalization, such as including more words on a particular topic. This paper focuses on the case of encouraging people's interest in news and proposes a framework for automatically generating news-centric crossword puzzles. We designed possible scenarios and built a prototype as a constraint satisfaction and optimization problem, that is, containing as many news-derived words as possible. Our experiments reported the generation probabilities and time required under several conditions. The results showed that news-centric crossword puzzles can be generated even with few news-derived words. We summarize the current issues and future research directions through a qualitative evaluation of the prototype. This is the first proposal that a formulation of a constraint satisfaction and optimization problem can be beneficial as an educational application.

### [Metapath-Guided Data-Augmentation For Knowledge Graphs](#)

Saurav Manchanda

Knowledge graph (KG) embedding techniques use relationships between entities to learn low-dimensional representations of entities and relations. The traditional KG embedding techniques (such as TransE and DistMult) estimate these embeddings using the observed KG triplets and differ in their triplet scoring loss functions. As these models only use the observed triplets to estimate the embeddings, they are prone to suffer through data sparsity that usually occurs in the real-world knowledge graphs, i.e., the lack of enough triplets per entity. In this paper, we propose an efficient method to augment the triplets to address the problem of data sparsity. We use random walks to create additional triplets, such that the relations carried by these introduced triplets correspond to the metapath (sequence of underlying relations) induced by the random walks. We also provide approaches to accurately and efficiently choose the informative metapaths from the possible set of metapaths. The proposed augmentation approaches can be used with any KG embedding approach out of the box. Experimental results on benchmarks show the advantages of the proposed approach.

### [Learning Visibility Attention Graph Representation for Time Series Forecasting](#)

Shengzhong Mao, Xiao-Jun Zeng

Visibility algorithm acts as a mapping that bridges graph representation learning with time series analysis, which has been broadly investigated for forecasting tasks. However, the intrinsic nature of visibility encoding yields graphs structured exclusively by binary adjacency matrix, leading to inevitable information loss of temporal sequence during the mapping. To this end, we introduce Angular Visibility Graph Networks (AVGNets), designed with two core features: (i) The framework reconstructs weighted graphs to encode time series by leveraging topological insights derived from visual angles of visibility networks, which capture sequential and structural information within weighted angular matrix. (ii) ProbAttention module is

proposed for evaluating probabilistic attention of weighted networks, with remarkable capabilities to extract intrinsic and extrinsic temporal dependencies across multi-layer graphs. Extensive experiments and ablation studies on real-world datasets covering diverse ranges demonstrate that AVGNets achieve state-of-the-art performance, offering an innovative perspective on graph representation for sequence modeling.

#### [On the Reliability of User Feedback for Evaluating the Quality of Conversational Agents](#)

Jordan Massiah, Emine Yilmaz, Yunlong Jiao, Gabriella Kazai

We analyse the reliability of users' explicit feedback for evaluating the quality of conversational agents. Using data from a commercial conversational system, we analyse how user feedback compares with human annotations; how well it aligns with implicit user satisfaction signals, such as retention; and how much user feedback is needed to reliably evaluate the quality of a conversational system.

#### [A Robust Backward Compatibility Metric for Model Retraining](#)

Ryuta Matsuno, Keita Sakuma

Model retraining and updating are essential processes in AI applications. However, during updates, there is a potential for performance degradation, in which the overall performance improves, but local performance deteriorates. This study proposes a backward compatibility metric that focuses on the compatibility of local predictive performance. The score of the proposed metric increases if the accuracy over the conditional distribution for each input is higher than before. Furthermore, we propose a model retraining method based on the proposed metric. Due to the use of the conditional distribution, our metric and retraining method are robust against label noises, while existing sample-based backward compatibility metrics are often affected by noise. We perform a theoretical analysis of our method and derive an upper bound for the generalization error. Numerical experiments demonstrate that our retraining method enhances compatibility while achieving equal or better trade-offs in overall performance compared to existing methods.

#### [Graph Contrastive Learning with Graph Info-Min](#)

En Meng, Yong Liu

The complexity of the graph structure poses a challenge for graph representation learning. Contrastive learning offers a straightforward and efficient unsupervised framework for graph representation learning. It achieves unsupervised learning by augmenting the original views and comparing them with the augmented views. Several methods based on this framework have achieved significant progress in the field of graph representation learning. Despite its success, the factors contributing to good augmented views in graph contrast learning have received less attention. In order to address this issue, we introduce the graph info-min principle. We investigate the relationship between mutual information (MI) and good augmented views through experimental and theoretical analysis. Additionally, we present a new contrastive learning method called Info-min Contrastive Learning (IMCL). Specifically, The method comprises an adaptive graph augmentation generator and a pseudo-label generator. The graph augmentation generator ensures sufficient differentiation between the augmented and original views. The pseudo-label generator generates pseudo-labels as supervision signals, ensuring consistency between the classification results of augmented views and original views. Our method demonstrates excellent performance through extensive experimental results on various datasets.

#### [Generative Graph Augmentation for Minority Class in Fraud Detection](#)

Lin Meng, Hesham Mostafa, Marcel Nassar, Xiaonan Zhang, Jiawei Zhang

Class imbalance is a well-recognized challenge in GNN-based fraud detection. Traditional methods like re-sampling and re-weighting address this issue by balancing class distribution. However, node class balancing with simple re-sampling or re-weighting may greatly distort the data distributions and eventually lead to the ineffective performance of GNNs. In this paper, we propose a novel approach named Graph Generative Node Augmentation (GGA), which improves GNN-based fraud detection models by augmenting synthetic nodes of the minority class. GGA utilizes the GAN framework to synthesize node features and related edges of fake fraudulent nodes. To introduce greater variety in the generated nodes, we employ an MLP for feature generation. We also introduce an attention module to encode feature-level information before graph convolutional layers for edge generation. Our empirical results on two real-world fraud datasets demonstrate that GGA improves the performance of GNN-based fraud detection models by a large margin with much fewer nodes than traditional class balance methods, and outperforms recent graph augmentation methods with the same number of synthetic nodes.

#### [SeqGen: A Sequence Generator via User Side Information for Behavior Sparsity in Recommendation](#)

Xu Min, Xiaolu Zhang, Bin Shen, Shuhan Wang, Yong He, Changsheng Li, Jun Zhou

In real-world industrial advertising systems, user behavior sparsity is a key issue that affects online recommendation performance. We observe that users with rich behaviors can obtain better recommendation results than those with sparse behaviors in a conversion-rate (CVR) prediction model. Inspired by this phenomenon, we propose a new method SeqGen, in an effort to exploit user side information to bridge the gap between rich and sparse behaviors. SeqGen is a learnable and pluggable module, which can be easily integrated into any CVR model and no longer requires two-stage training as in previous works. In particular, SeqGen learns a mapping relationship between the user side information and behavior sequences, only on the basis of the users with long behavior sequences. After that, SeqGen can generate rich sequence features for users with sparse behaviors based on their side information, so as to alleviate the issue of user behavior sparsity. The generated sequence features will then be fed into the classifier tower of an arbitrary CVR model together with the original sequence features. To the best of our knowledge, our approach constitutes the first attempt to exploit user side information for addressing the user behavior sparsity issue. We validate the effectiveness of SeqGen on the publicly available dataset MovieLens-1M, and our method receives an improvement of up to 0.5% in terms of the AUC score. More importantly, we successfully deploy SeqGen in the commercial advertising system Xlight of Alipay, which improves the grouped AUC of the CVR model by 0.6% and brings a boost of 0.49% in terms of the conversion rate on A/B testing.

#### [A Flash Attention Transformer for Multi-Behaviour Recommendation](#)

Tendai Mukande, Esraa Ali, Annalina Caputo, Ruihai Dong, Noel E. O'Connor

Recently, modelling heterogeneous interactions in recommender systems has attracted research interest. Real-world scenarios involve sequential multi-type user-item interactions such as "shape view", "shape add-to-favourites", "shape add-to-cart" and "shape purchase". Graph Neural Network (GNN) methods have been widely adopted in Representation Learning of similar sequential user-item interactions. Promising results have been achieved by the integration of GNNs and transformers for self-attention. However, GNN based methods suffer from limited capability in handling global user-item interaction dependencies, particularly for long sequences. Moreover, these models require high computational cost of transformers, due to the quadratic memory and time complexity with respect to sequence length. This results in memory bottlenecks and slow training especially in computational resource-constrained environments. To address these challenges, we propose the FATH model which employs Flash Attention mechanism to reduce the high-bandwidth memory usage over higher-order user-item interaction sequences. Experimental results show that our model improves the training speed and reduces the memory usage with better recommendation performance in comparison with the state-of-the-art baselines.

#### [Product Entity Matching via Tabular Data](#)

Ali Naeim abadi, Mir Tafseer Nayeem, Davood Rafiei

Product Entity Matching (PEM)--a subfield of record linkage that focuses on linking records that refer to the same product--is a challenging task for many entity matching models. For example, recent transformer models report a near-perfect performance score on many datasets while their performance is the lowest on PEM datasets. In this paper, we study PEM under the common setting where the information is spread over text and tables. We show that adding tables can enrich the existing PEM datasets and those tables can act as a bridge between the entities being matched. We also propose TATEM, an effective solution that leverages Pre-trained Language Models (PLMs) with a novel serialization technique to encode tabular product data and an attribute ranking module to make our model more data-efficient. Our experiments on both current benchmark datasets and our proposed datasets show significant improvements compared to state-of-the-art methods, including Large Language Models (LLMs) in zero-shot and few-shot settings.

#### [Efficient Differencing of System-level Provenance Graphs](#)

Yuta Nakamura, Iyad Kanj, Tanu Malik

Data provenance, when audited at the operating system level, generates a large volume of low-level events. Current provenance systems infer causal flow from these event traces, but do not infer application structure, such as loops and branches. The absence of these inferred structures decreases accuracy when comparing two event traces, leading to low-quality answers from a provenance system. In this paper, we infer nested natural and unnatural loop structures over a collection of provenance event traces. We describe an 'unrolling method' that uses the inferred nested loop structure to systematically mark loop iterations. Our loop-based unrolling improves the accuracy of trace comparison by 20-70% over trace comparisons that do not rely on inferred structures.

#### [Differential Privacy in HyperNetworks for Personalized Federated Learning](#)

Vaisnavi Nemala, Phung Lai, Nhathai Phan

Federated learning (FL) is a framework for collaborative learning among users through a coordinating server. A recent HyperNetwork-based personalized FL framework, called HyperNetFL, is used to generate local models using personalized descriptors optimized for each user independently. However, HyperNetFL introduces unknown privacy risks. This paper introduces a novel approach to preserve user-level differential privacy, dubbed User-level DP, by providing formal privacy protection for data owners in training a HyperNetFL model. To achieve that, our proposed algorithm, called UDP-Alg, optimizes the trade-off between privacy loss and model utility by tightening sensitivity bounds. An intensive evaluation using benchmark datasets shows that our proposed UDP-Alg significantly improves privacy protection at a modest cost in utility.

#### **Camaraderie: Content-based Knowledge Transfer for Medical Image Labelling using Supervised Autoencoders in a Decentralized Setting**

Advait Padhye, Shreeja Bhakat, Humaira Firdowse, Atharv Savarkar, Ganesh Ramakrishnan, Kshitij S. Jadhav

Deep neural networks for medical imaging require large high-quality labelled data, a huge bottleneck for resource poor settings. Given the privacy requirements of medical data, institutes are un-willing to share data, causing an hindrance in resource poor settings. In the present paper, (Camaraderie: Content-based Knowledge Transfer for Medical Image Labelling using Supervised Autoencoders in a Decentralized Setting) we propose to use Discrete Classifier Supervised Autoencoder (DC-SAE) to generate low-dimensional representations of a few annotated images at the Donor client and transfer both the DC-SAE's encoder part and the latent space representations to the Recipient client without sharing raw data. We then pass the unlabelled images of the Recipient Client through this encoder to obtain their latent space representation. In a supervised setting, using latent space representation of Donor client's labelled images, we accurately annotate images of Recipient client. Camaraderie demonstrates that DC-SAE outperforms Recipient end label accuracy beyond classical VAE based classification and anomaly detection based VAE. Thus, given a limited amount of labelled data in a decentralized privacy preserving scenario, one can transfer latent space representation across clients to annotate large number of unlabelled images with high accuracy.

#### **Satisfaction-Aware User Interest Network for Click-Through Rate Prediction**

Mao Pan, Wen Shi, Kun Zhou, Zheng Li, Dongyue Wang, Zhuoye Ding, Xiwei Zhao, Sulong Xu

Click-Through Rate (CTR) prediction plays a pivotal role in numerous industrial applications, including online advertising and recommender systems. Existing approaches primarily focus on modeling the correlation between user interests and candidate items. However, we argue that personalized user preferences for candidate items depend not only on correlation but also on the satisfaction of associated interests. To address this limitation, we propose SUIN, a novel CTR model that integrates satisfaction factors into user interest modeling for enhanced click-through rate prediction. Specifically, we employ a user interest satisfaction-aware network to capture the degree of satisfaction for each interest, thereby enabling adaptation of the user's personalized preference based on satisfaction levels. Additionally, we leverage the exposure-unclicked signal (recommended to the user but not clicked) as supervision during training, facilitating the interest satisfaction module to better model the satisfaction degree of user interests. Besides, this module serves as a foundational building block suitable for integration into mainstream sequential-based CTR models. Extensive experiments conducted on two real-world datasets demonstrate the superiority of our proposed model, outperforming state-of-the-art methods across various evaluation metrics. Furthermore, an online A/B test deployed on large-scale recommender systems shows significant improvements achieved by our model in diverse evaluation metrics.

#### **Quantum Split Learning for Privacy-Preserving Information Management**

Soohyun Park, Hankyul Baek, Joongheon Kim

Recently, research on quantum neural network (QNN) architectures has been attracted in various fields. Among them, the distributed computation of QNN has been actively discussed for privacy-preserving information management due to data and model distribution over multiple computing devices. Based on this concept, this paper proposes quantum split learning (QSL) which splits a single QNN architecture across multiple distributed computing devices to avoid entire QNN architecture exposure. In order to realize QSL design, this paper also proposes cross-channel pooling, which utilizes quantum state tomography. Our evaluation results verifies that QSL preserves privacy in classification tasks and also improves accuracy at most by 6.83% compared to existing methods.

#### **T-SaS: Toward Shift-aware Dynamic Adaptation for Streaming Data**

Weijieying Ren, Tianxiang Zhao, Wei Qin, Kunpeng Liu



In many real-world scenarios, distribution shifts exist in the streaming data across time steps. Many complex sequential data can be effectively divided into distinct regimes that exhibit persistent dynamics. Discovering the shifted behaviors and the evolving patterns underlying the streaming data are important to understand the dynamic system. Existing methods typically train one robust model to work for the evolving data of distinct distributions or sequentially adapt the model utilizing explicitly given regime boundaries. However, there are two challenges: (1) shifts in data streams could happen drastically and abruptly without precursors. Boundaries of distribution shifts are usually unavailable, and (2) training a shared model for all domains could fail to capture varying patterns. This paper aims to solve the problem of sequential data modeling in the presence of sudden distribution shifts that occur without any precursors. Specifically, we design a Bayesian framework, dubbed as T-SaS, with a discrete distribution-modeling variable to capture abrupt shifts of data. Then, we design a model that enable adaptation with dynamic network selection conditioned on that discrete variable. The proposed method learns specific model parameters for each distribution by learning which neurons should be activated in the full network. A dynamic masking strategy is adopted here to support inter-distribution transfer through the overlapping of a set of sparse networks. Extensive experiments show that our proposed method is superior in both accurately detecting shift boundaries to get segments of varying distributions and effectively adapting to downstream forecast or classification tasks.

#### [A Self-Learning Resource-Efficient Re-Ranking Method for Clinical Trials Search](#)

Maciej Rybinski, Vincent Nguyen, Sarvnaz Karimi

Complex search scenarios, such as those in biomedical settings, can be challenging. One such scenario is matching a patient's profile to relevant clinical trials. There are multiple criteria that should match for a document (clinical trial) to be considered relevant to a query (patient's profile represented with an admission note). While different neural ranking methods have been proposed for searching clinical trials, resource-efficient approaches to ranker training are less studied. A resource-efficient method uses training data in moderation. We propose a self-learning reranking method that achieves results comparable to those of more complicated, fully supervised, systems. Our experiments demonstrate our method's robustness and competitiveness compared to the state-of-the-art approaches in clinical trial search.

#### [Adversarial Density Ratio Estimation for Change Point Detection](#)

Shreyas S, Prakash Mandayam Comar, Sivaramakrishnan Kaveri

Change Point Detection (CPD) models are used to identify abrupt changes in the distribution of a data stream and have a widespread practical use. CPD methods generally compare the distribution of data sequences before and after a given time step to infer if there is a shift in distribution at the said time step. Numerous divergence measures, which measure distance between data distributions of sequence pairs, have been proposed for CPD \cite{MStatisticNIPS, BergCPD} and often the choice of divergence measure depends on the data used. Density Ratio Estimation (DRE) \cite{RelDivCPD, BergCPD} can be used to estimate a broad family of f-divergences, which includes widely used CPD divergences like Kullback-Leibler (KL) and Pearson, and thus DRE is a popular approach for CPD. In this work, we improve upon the existing DRE techniques for CPD, by proposing a novel objective that combines DRE seamlessly with adversarial sample generation. The adversarial samples allows for a robust CPD with DRE to track subtle changes in distribution, leading to a reduction in false negatives. We experiment on a wide variety of real-world, public benchmark datasets to show that our approach improves upon existing state-of-the-art (SoTA) methods, including DRE based CPD methods, by demonstrating an  $\sim 5\%$  increase in F-score.

#### [Quantitative Decomposition of Prediction Errors Revealing Multi-Cause Impacts: An Insightful Framework for MLOps](#)

Keita Sakuma, Ryuta Matsuno, Yoshio Kameda

As machine learning applications expand in various industries, MLOps, which enables continuous model operation and improvement, becomes increasingly significant. Identifying causes of prediction errors, such as low model performance or anomalous samples, and implementing appropriate countermeasures are essential for effective MLOps. Furthermore, quantitatively evaluating each cause's impact is necessary to determine the effectiveness of countermeasures. In this study, we propose a method to quantitatively decompose a single sample's prediction error into contributions from multiple causes. Our method involves four steps: calculating the prediction error, computing metrics related to error causes, using a regression model to learn the relationship between the error and metrics, and applying SHAP to interpret the model's predictions and calculate the contribution of each cause to the prediction error. Numerical experiments with open data show that our method offers valuable insights for model improvement, confirming the effectiveness of our approach.

#### [Neural Disentanglement of Query Difficulty and Semantics](#)

Sara Salamat, Negar Arabzadeh, Shirin Seyedsalehi, Amin Bigdeli, Morteza Zihayat, Ebrahim Bagheri

Researchers have shown that the retrieval effectiveness of queries may depend on other factors in addition to the semantics of the query. In other words, several queries expressed with the same intent, and even using overlapping keywords, may exhibit completely different degrees of retrieval effectiveness. As such, the objective of our work in this paper is to propose a neural disentanglement method that is able to disentangle query semantics from query difficulty. The disentangled query semantics representation provides the means to determine semantic association between queries whereas the disentangled query difficulty representation would allow for the estimation of query effectiveness. We show through our experiments on the query performance prediction; and, query similarity calculation tasks that our proposed disentanglement method is able to show better performance compared to the state of the art.

#### **[VN-Solver: Vision-based Neural Solver for Combinatorial Optimization over Graphs](#)**

Mina Samizadeh, Guangmo Tong

Data-driven approaches have been proven effective in solving combinatorial optimization problems over graphs such as the traveling salesman problems and the vehicle routing problem. The rationale behind such methods is that the input instances may follow distributions with salient patterns that can be leveraged to overcome the worst-case computational hardness. For optimization problems over graphs, the common practice of neural combinatorial solvers consumes the inputs in the form of adjacency matrices. In this paper, we explore a vision-based method that is conceptually novel: can neural models solve graph optimization problems by taking a look at the graph pattern - Our results suggest that the performance of such vision-based methods is not only non-trivial but also comparable to the state-of-the-art matrix-based methods, which opens a new avenue for developing data-driven optimization solvers.

#### **[EdgeNet : Encoder-decoder generative Network for Auction Design in E-commerce Online Advertising](#)**

Guangyuan Shen, Shengjie Sun, Dehong Gao, Duanxiao Song, Libin Yang, Zhen Wang, Yongping Shi, Wei Ning

We present a new encoder-decoder generative network dubbed EdgeNet, which introduces a novel encoder-decoder framework for data-driven auction design in online e-commerce advertising. We break the neural auction paradigm of Generalized-Second-Price(GSP), and improve the utilization efficiency of data while ensuring the economic characteristics of the auction mechanism. Specifically, EdgeNet introduces a transformer-based encoder to better capture the mutual influence among different candidate advertisements. In contrast to GSP based neural auction model, we design an autoregressive decoder to better utilize the rich context information in online advertising auctions. EdgeNet is conceptually simple and easy to extend to the existing end-to-end neural auction framework. We validate the efficiency of EdgeNet on a wide range of e-commercial advertising auction, demonstrating its potential in improving user experience and platform revenue.

#### **[DAE: Distribution-Aware Embedding for Numerical Features in Click-Through Rate Prediction](#)**

Bin Shen, Jingran Xu, Xu Min, Zeyu Ke, Yong He, Liang Zhang, Xin Dong, Linjian Mo

Numerical features are an important type of input for CTR prediction models. Recently, several discretization and numerical transformation methods have been proposed to deal with numerical features. However, existing approaches do not fully consider compatibility with different distributions. Here, we propose a novel numerical feature embedding framework, called Distribution-Aware Embedding (DAE), which is applicable to various numerical feature distributions. First, DAE efficiently approximates the cumulative distribution function by estimating the expectation of the order statistics. Then, the distribution information is applied to the embedding layer by nonlinear interpolation. Finally, to capture both local and global information, we aggregate the embeddings at multiple scales to obtain the final representation. Empirical results validate the effectiveness of DAE compared to the baselines, while demonstrating the adaptability to different CTR models and distributions.

#### **[Learning to Simulate Complex Physical Systems: A Case Study](#)**

Jiasheng Shi, Fu Lin, Weixiong Rao

Complex physical system simulation is important in many real world applications. We study the general simulation scenario to generate the response result when a physical object is applied by external factors. Traditional solvers on Partial Differential Equations (PDEs) suffer from significantly high computational cost. Many recent learning-based approaches focus on multivariate time series alike simulation prediction problem and do not work for our case. In this paper, we propose a novel two-level graph neural networks (GNNs) to learn the simulation result of a physical object applied by external factors. The key is a two-level graph structure where one fine mesh graph is mapped to multiple coarse one. Our preliminary evaluation on both synthetic and real datasets demonstrates that our work outperforms three state-of-the-arts by much lower errors.

#### **[Findability: A Novel Measure of Information Accessibility](#)**

Aman Sinha, Priyanshu Raj Mall, Dwaipayan Roy

The overwhelming volume of data generated and indexed by search engines poses a significant challenge in retrieving documents from the index efficiently and effectively. Even with a well-crafted query, several relevant documents often get buried among a multitude of competing documents, resulting in reduced accessibility or "findability" of the desired document. Consequently, it is crucial to develop a robust methodology for assessing this dimension of Information Retrieval (IR) system performance. While previous studies have focused on measuring document accessibility disregarding user queries and document relevance, there exists no metric to quantify the findability of a document within a given IR system without resorting to manual labor. This paper aims to address this gap by defining and deriving a metric to evaluate the findability of documents as perceived by end-users. Through experiments, we demonstrate the varying impact of different retrieval models and collections on the findability of documents. Furthermore, we establish the findability measure as an independent metric distinct from retrievability, an accessibility measure introduced in prior literature.

### [Improving Diversity in Unsupervised Keyphrase Extraction with Determinantal Point Process](#)

Mingyang Song, Huafeng Liu, Liping Jing

Keyphrase extraction aims to provide readers with high-level information about the central ideas or important topics described in a given source text. Recent advances in embedding-based models have made remarkable progress on unsupervised keyphrase extraction, demonstrated through improved quality metrics such as F1-score. However, the diversity in the keyphrase extraction task needs to be addressed. In this paper, we focus on diverse keyphrase extraction, which entails extracting keyphrases that cover different central information or essential topics in the document. To achieve this goal, we propose a re-ranking-based approach that employs determinantal point processes utilizing BERT as kernels, which we call DiversityRank. Specifically, DiversityRank jointly considers phrase-document relevance and cross-phrase similarities to select candidate keyphrases that are document-relevant and diverse. Results demonstrate that our re-ranking strategy outperforms the state-of-the-art unsupervised keyphrase extraction baselines on three benchmark datasets.

### [Pre-training with Aspect-Content Text Mutual Prediction for Multi-Aspect Dense Retrieval](#)

Xiaojie Sun, Keping Bi, Jiafeng Guo, Xinyu Ma, Yixing Fan, Hongyu Shan, Qishen Zhang, Zhongyi Liu

Grounded on pre-trained language models (PLMs), dense retrieval has been studied extensively on plain text. In contrast, there has been little research on retrieving data with multiple aspects using dense models. In the scenarios such as product search, the aspect information plays an essential role in relevance matching, e.g., category: Electronics, Computers, and Pet Supplies. A common way of leveraging aspect information for multi-aspect retrieval is to introduce an auxiliary classification objective, i.e., using item contents to predict the annotated value IDs of item aspects. However, by learning the value embeddings from scratch, this approach may not capture the various semantic similarities between the values sufficiently. To address this limitation, we leverage the aspect information as text strings rather than class IDs during pre-training so that their semantic similarities can be naturally captured in the PLMs. To facilitate effective retrieval with the aspect strings, we propose mutual prediction objectives between the text of the item aspect and content. In this way, our model makes more sufficient use of aspect information than conducting undifferentiated masked language modeling (MLM) on the concatenated text of aspects and content. Extensive experiments on two real-world datasets (product and mini-program search) show that our approach can outperform competitive baselines both treating aspect values as classes and conducting the same MLM for aspect and content strings. Code and related dataset will be available at the URL

[/footnotehttps://github.com/sunxiaojie99/ATTEMPT](https://github.com/sunxiaojie99/ATTEMPT).

### [Sequential Text-based Knowledge Update with Self-Supervised Learning for Generative Language Models](#)

Hao-Ru Sung, Ying-Jhe Tang, Yu-Chung Cheng, Pai-Lin Chen, Tsai-Yen Li, Hen-Hsen Huang

This work proposes a new natural language processing (NLP) task to tackle the issue of multi-round, sequential text-based knowledge update. The study introduces a hybrid learning architecture and a novel self-supervised training strategy to enable generative language models to consolidate knowledge in the same way as humans. A dataset was also created for evaluation and results showed the effectiveness of our methodology. Experimental results confirm the superiority of the proposed approach over existing models and large language models (LLMs). The proposed task and model framework have the potential to significantly improve the automation of knowledge organization, making text-based knowledge an increasingly crucial resource for powerful LLMs to perform various tasks for humans.

### [Higher-Order Peak Decomposition](#)

Xingyu Tan, Jingya Qian, Chen Chen, Sima Qing, Yanping Wu, Xiaoyang Wang, Wenjie Zhang

k-peak is a well-regarded cohesive subgraph model in graph analysis. However, the k-peak model only considers the direct neighbors of a vertex, consequently limiting its capacity to uncover higher-order structural information of the graph. To address this limitation, we propose a new model in this paper, named (k,h)-peak, which incorporates higher-order (h-hops) neighborhood information of vertices. Employing the (k,h)-peak model, we explore the higher-order peak decomposition problem that calculates the vertex peakness for all conceivable k values given a particular h. To tackle this problem efficiently, we propose an advanced local computation based algorithm, which is parallelizable, and additionally, devise novel pruning strategies to mitigate unnecessary computation. Experiments as well as case studies are conducted on real-world datasets to evaluate the efficiency and effectiveness of our proposed solutions.

#### [Exposing Model Theft: A Robust and Transferable Watermark for Thwarting Model Extraction Attacks](#)

Ruixiang Tang, Hongye Jin, Mengnan Du, Curtis Wigington, Rajiv Jain, Xia Hu

The increasing prevalence of Deep Neural Networks (DNNs) in cloud-based services has led to their widespread use through various APIs. However, recent studies reveal the susceptibility of these public APIs to model extraction attacks, where adversaries attempt to create a local duplicate of the private model using data and API-generated predictions. Existing defense methods often involve perturbing prediction distributions to hinder an attacker's training goals, inadvertently affecting API utility. In this study, we extend the concept of digital watermarking to protect DNNs' APIs. We suggest embedding a watermark into the safeguarded APIs; thus, any model attempting to copy will inherently carry the watermark, allowing the defender to verify any suspicious models. We propose a simple yet effective framework to increase watermark transferability. By requiring the model to memorize the preset watermarks in the final decision layers, we significantly enhance the transferability of watermarks. Comprehensive experiments show that our proposed framework not only successfully watermarks APIs but also maintains their utility.

#### [Leveraging Knowledge and Reinforcement Learning for Enhanced Reliability of Language Models](#)

Nancy Tyagi, Surjodeep Sarkar, Manas Gaur

The Natural Language Processing (NLP) community has been using crowd-sourcing techniques to create benchmark datasets such as General Language Understanding and Evaluation (GLUE) for training modern Language Models (LMs) such as BERT. GLUE tasks measure the reliability scores using inter-annotator metrics - Cohen's Kappa (K). However, the reliability aspect of LMs has often been overlooked. To counter this problem, we explore a knowledge-guided LM ensembling approach that leverages reinforcement learning to integrate knowledge from ConceptNet and Wikipedia as knowledge graph embeddings. This approach mimics human annotators resorting to external knowledge to compensate for information deficits in the datasets. Across nine GLUE datasets, our research shows that ensembling strengthens reliability and accuracy scores, outperforming state-of-the-art.

#### [RecRec: Algorithmic Recourse for Recommender Systems](#)

Sahil Verma, Ashudeep Singh, Varich Boonsanong, John P. Dickerson, Chirag Shah

Recommender systems play an essential role in the choices people make in domains such as entertainment, shopping, food, news, employment, and education. The machine learning models underlying these recommender systems are often enormously large and black-box in nature for users, content providers, and system developers alike. It is often crucial for all stakeholders to understand the model's rationale behind making certain predictions and recommendations. This is especially true for the content providers whose livelihoods depend on the recommender system. Drawing motivation from the practitioners' need, in this work, we propose a recourse framework for recommender systems, targeted towards the content providers. Algorithmic recourse in the recommendation setting is a set of actions that, if executed, would modify the recommendations (or ranking) of an item in the desired manner. A recourse suggests actions of the form: "if a feature changes X to Y, then the ranking of that item for a set of users will change to X." Furthermore, we demonstrate that RecRec is highly effective in generating valid, sparse, and actionable recourses through an empirical evaluation of recommender systems trained on three real-world datasets. To the best of our knowledge, this work is the first to conceptualize and empirically test a generalized framework for generating recourses for recommender systems.

#### [Network Embedding with Adaptive Multi-hop Contrast](#)

Chenhao Wang, Yong Liu, Yan Yang

Graph neural networks (GNNs) have shown strong performance in graph-based analysis tasks. Despite their remarkable success, the inherent homophilic message-passing mechanism (MP) makes GNNs challenging to generalize to heterophilic graphs. In addition, the MP explicitly exploits the connection relationships between local neighbor nodes making

GNNs unable to maintain stable performance in the face of adversarial perturbation attacks. In this paper, we propose a new method to explore graph structure by removing explicit message-passing mechanisms and present a network embedding framework AMCNE with Adaptive Multi-hop Contrast loss (AMCloss) to address these challenges. AMCNE only relies on a simple autoencoder to obtain node representations for classification and uses elaborate contrastive loss to drive nodes capturing complex structural information on heterophilic graphs. The comprehensive experiments show that AMCNE outperforms state-of-the-art baseline models on homophilic and heterophilic graphs and is more robust in the node classification task. \endabstract

### [Efficient Multi-Task Learning via Generalist Recommender](#)

Luyang Wang, Cangcheng Tang, Chongyang Zhang, Jun Ruan, Kai Huang, Jason Dai

Multi-task learning (MTL) is a common machine learning technique that allows the model to share information across different tasks and improve the accuracy of recommendations for all of them. Many existing MTL implementations suffer from scalability issues as the training and inference performance can degrade with the increasing number of tasks, which can limit production use case scenarios for MTL-based recommender systems. Inspired by the recent advances of large language models, we developed an end-to-end efficient and scalable Generalist Recommender (GRec). GRec takes comprehensive data signals by utilizing NLP heads, parallel Transformers, as well as a wide and deep structure to process multi-modal inputs. These inputs are then combined and fed through a newly proposed task-sentence level routing mechanism to scale the model capabilities on multiple tasks without compromising performance. Offline evaluations and online experiments show that GRec significantly outperforms our previous recommender solutions. GRec has been successfully deployed on one of the largest telecom websites and apps, effectively managing high volumes of online traffic every day.

### [Clustering-property Matters: A Cluster-aware Network for Large Scale Multivariate Time Series Forecasting](#)

Yuan Wang, Zezhi Shao, Tao Sun, Chengqing Yu, Yongjun Xu, Fei Wang

Large-scale Multivariate Time Series(MTS) widely exist in various real-world systems, imposing significant demands on model efficiency. A recent work, STID, addressed the high complexity issue of popular Spatial-Temporal Graph Neural Networks(STGNNs). Despite its success, when applied to large-scale MTS data, the number of parameters of STID for modeling spatial dependencies increases substantially, leading to over-parameterization issues and suboptimal performance. These observations motivate us to explore new approaches for modeling spatial dependencies in a parameter-friendly manner. In this paper, we argue that the spatial properties of variables are essentially the superposition of multiple cluster centers. Accordingly, we propose a Cluster-Aware Network(CANet), which effectively captures spatial dependencies by mining the implicit cluster centers of variables. CANet solely optimizes the cluster centers instead of the spatial information of all nodes, thereby significantly reducing the parameter amount. Extensive experiments on two large-scale datasets validate our motivation and demonstrate the superiority of CANet.

### [Training Heterogeneous Graph Neural Networks using Bandit Sampling](#)

Ta-Yang Wang, Rajgopal Kannan, Viktor Prasanna

Graph neural networks (GNNs) have gained significant attention across diverse areas due to their superior performance in learning graph representations. While GNNs exhibit superior performance compared to other methods, they are primarily designed for homogeneous graphs, where all nodes and edges are of the same type. Training a GNN model for large-scale graphs incurs high computation and storage costs, especially when considering the heterogeneous structural information of each node. To address the demand for efficient GNN training, various sampling methods have been proposed. In this paper, we propose a sampling method based on bandit sampling, an online learning algorithm with provable convergence under weak assumptions on the learning objective. To the best of our knowledge, this is the first bandit-based sampling method applied to heterogeneous GNNs with a theoretical guarantee. The main idea is to prioritize node types with more informative connections with respect to the learning objective. Compared with existing techniques for GNN training on heterogeneous graphs, extensive experiments using the Open Academic Graph (OAG) dataset demonstrate that our proposed method outperforms the state-of-the-art in terms of the runtime across various tasks with a speed-up of 1.5-2x, while achieving similar accuracy.

### [Adaptive Graph Neural Diffusion for Traffic Demand Forecasting](#)

Yiling Wu, Xinfeng Zhang, Yaowei Wang

This paper studies the problem of spatial-temporal modeling for traffic demand forecasting. In practice, the temporal-spatial dependencies are complex. Conventional methods using graph convolutional networks and gated recurrent units cannot fully



explore the patterns of demand evolution. Therefore, we propose Adaptive Graph Neural Diffusion (AGND) for spatial-temporal graph modeling. Specifically, complex spatial relations are modeled with a diffusion process by the graph neural diffusion. The spatial attention mechanism and a data-driven semantic adjacency matrix are used to describe the diffusivity function in the graph neural diffusion, which provides both local and global spatial information. Long-term temporal dependencies are modeled by the temporal attention mechanism. The proposed method is applied to two real-world datasets, and the results show that the proposed method outperforms state-of-the-art methods.

### **Promoting Diversity in Mixed Complex Cooperative and Competitive Multi-Agent Environment**

Jia Wu, ZiXiao Huang

This paper introduces a new approach for promoting diversity of behavior in complex multi-agent environments that pose three challenges: 1) competition or collaboration among agents of diverse types, 2) the need for complex multi-agent coordination, which makes it challenging to achieve risky cooperation strategies, and 3) a large number of agents in the environment, leading to increased complexity when considering agent-to-agent relationships. To address the first two challenges, we leverage Reward Randomization in combination with Bayesian Optimization to train agents to exhibit diverse strategic behaviors, thereby mitigating the issue of risky cooperation. To address the challenge of learning in a large number of agents, we utilize MAPPO with parameter sharing to enhance learning efficiency. Experimental results demonstrate that within this multi-agent environment, agents can effectively learn multiple visually distinct behaviors, and the incorporation of these two techniques significantly improves agents' performance.

### **MTKDN: Multi-Task Knowledge Disentanglement Network for Recommendation**

Haotian Wu, Bowen Xing, Ivor Tsang

Multi-task learning (MTL) is a widely adopted machine learning paradigm in recommender systems. However, existing MTL models often suffer from performance degeneration with negative transfer and seesaw phenomena. Some works attempt to alleviate the negative transfer and seesaw issues by separating task-specific and shared experts to mitigate the harmful interference between task-specific and shared knowledge. Despite the success of these efforts, task-specific and shared knowledge have still not been thoroughly decoupled. There may still exist unnecessary mixture between the shared and task-specific knowledge, which may harm MTL models' performances. To tackle this problem, in this paper, we propose multi-task knowledge disentanglement network (MTKDN) to further reduce harmful interference between the shared and task-specific knowledge. Specifically, we propose a novel contrastive disentanglement mechanism to explicitly decouple the shared and task-specific knowledge in corresponding hidden spaces. In this way, the unnecessary mixture between shared and task-specific knowledge can be reduced. As for optimization objectives, we propose individual optimization objectives for shared and task-specific experts, by which we can encourage these two kinds of experts to focus more on extracting the shared and task-specific knowledge, respectively. Additionally, we propose a margin regularization to ensure that the fusion of shared and task-specific knowledge can outperform exploiting either of them alone. We conduct extensive experiments on open-source large-scale recommendation datasets. The experimental results demonstrate that MTKDN significantly outperforms state-of-the-art MTL models. In addition, the ablation experiments further verify the necessity of our proposed contrastive disentanglement mechanism and the novel loss settings.

### **G-Meta: Distributed Meta Learning in GPU Clusters for Large-Scale Recommender Systems**

Youshao Xiao, Shangchun Zhao, Zhenglei Zhou, Zhaoxin Huan, Lin Ju, Xiaolu Zhang, Lin Wang, Jun Zhou

Recently, a new paradigm, meta learning, has been widely applied to Deep Learning Recommendation Models (DLRM) and significantly improves statistical performance, especially in cold-start scenarios. However, the existing systems are not tailored for meta learning based DLRM models and have critical problems regarding efficiency in distributed training in the GPU cluster. It is because the conventional deep learning pipeline is not optimized for two task-specific datasets and two update loops in meta learning. This paper provides a high-performance framework for large-scale training for Optimization-based Meta DLRM models over the GPU cluster, namely G-Meta. Firstly, G-Meta utilizes both data parallelism and model parallelism with careful orchestration regarding computation and communication efficiency, to enable high-speed distributed training. Secondly, it proposes a Meta-IO pipeline for efficient data ingestion to alleviate the I/O bottleneck. Various experimental results show that G-Meta achieves notable training speed without loss of statistical performance. Since early 2022, G-Meta has been deployed in Alipay's core advertising and recommender system, shrinking the continuous delivery of models by four times. It also obtains 6.48% improvement in Conversion Rate (CVR) and 1.06% increase in CPM (Cost Per Mille) in Alipay's homepage display advertising, with the benefit of larger training samples and tasks.

### [\*\*A Joint Training-Calibration Framework for Test-Time Personalization with Label Shift in Federated Learning\*\*](#)

Jian Xu, Shao-Lun Huang

The data heterogeneity has been a challenging issue in federated learning in both training and inference stages, which motivates a variety of approaches to learn either personalized models for participating clients or test-time adaptations for unseen clients. One such approach is employing a shared feature representation and a customized classifier head for each client. However, previous works either neglect the global head with rich knowledge or assume the new clients have enough labeled data, which significantly limit their broader practicality. In this work, we propose a lightweight framework to tackle with the label shift issue during the model deployment by test priors estimation and model prediction calibration. We also demonstrate the importance of training a balanced global model in FL so as to guarantee the general effectiveness of prior estimation approaches. Evaluation results on benchmark datasets demonstrate the superiority of our framework for model adaptation in unseen clients with unknown label shifts.

### [\*\*Geometry Interaction Augmented Graph Collaborative Filtering\*\*](#)

Jie Xu, Chaozhuo Li

Graph collaborative filtering, which could capture the abundant collaborative signal from the high-order connectivity of the tree-likeness user-item interaction graph, has received considerable research attention recently. Most graph collaborative filtering methods embed graphs in the Euclidean spaces, but that could have high distortion when embedding graphs with tree-likeness structure. Recently, some researchers address this problem by learning the feature representations in the hyperbolic spaces. However, because the user-item interaction graphs also have cyclic structure, the high-order collaborative signal cannot be well captured by hyperbolic spaces. From this point of view, neither Euclidean spaces nor hyperbolic spaces can capture the full information from the complexity of user-item interactions. Therefore, how to construct a suitable embedding space for graph collaboration filtering is an important problem. In this paper, we analyze the properties of hyperbolic geometry in graph collaborative filtering tasks and proposed a novel geometry interaction augmented graph collaborative filtering (GeoGCF) method, which leverages both Euclidean and hyperbolic geometry to model the user-item interactions. Experimental results show the effectiveness of the proposed method.

### [\*\*Mitigating Semantic Confusion from Hostile Neighborhood for Graph Active Learning\*\*](#)

Tianmeng Yang, Min Zhou, Yujing Wang, Zhengjie Lin, Lujia Pan, Bin Cui, Yunhai Tong

Graph Active Learning (GAL), which aims to find the most informative nodes in graphs for annotation to maximize the Graph Neural Networks (GNNs) performance, has attracted many research efforts but remains non-trivial challenges. One major challenge is that existing GAL strategies may introduce semantic confusion to the selected training set, particularly when graphs are noisy. Specifically, most existing methods assume all aggregating features to be helpful, ignoring the semantically negative effect between inter-class edges under the message-passing mechanism. In this work, we present Semantic-aware Active learning framework for Graphs (SAG) to mitigate the semantic confusion problem. Pairwise similarities and dissimilarities of nodes with semantic features are introduced to jointly evaluate the node influence. A new prototype-based criterion and query policy are also designed to maintain diversity and class balance of the selected nodes, respectively. Extensive experiments on the public benchmark graphs and a real-world financial dataset demonstrate that SAG significantly improves node classification performances and consistently outperforms previous methods. Moreover, comprehensive analysis and ablation study also verify the effectiveness of the proposed framework.

### [\*\*MC-DRE: Multi-Aspect Cross Integration for Drug Event/Entity Extraction\*\*](#)

Jie Yang, Soyeon Caren Han, Siqu Long, Josiah Poon, Goran Nenadic

Extracting meaningful drug-related information chunks, such as adverse drug events (ADE), is crucial for preventing morbidity and saving many lives. Most ADEs are reported via an unstructured conversation with the medical context, so applying a general entity recognition approach is not sufficient enough. In this paper, we propose a new multi-aspect cross-integration framework for drug entity/event detection by capturing and aligning different context/language/knowledge properties from drug-related documents. We first construct multi-aspect encoders to describe semantic, syntactic, and medical document contextual information by conducting those slot tagging tasks, main drug entity/event detection, part-of-speech tagging, and general medical named entity recognition. Then, each encoder conducts cross-integration with other contextual information in three ways: the key-value cross, attention cross, and feedforward cross, so the multi-encoders are integrated in depth. Our model outperforms all SOTA on two widely used tasks, flat entity detection and discontinuous event extraction.

### **Positive-Unlabeled Node Classification with Structure-aware Graph Learning**

Hansi Yang, Yongqi Zhang, Quanming Yao, James Kwok

Node classification on graphs is an important research problem with many applications. Real-world graph data sets may not be balanced and accurate as assumed by most existing works. A challenging setting is positive-unlabeled (PU) node classification, where labeled nodes are restricted to positive nodes. It has diverse applications, e.g., pandemic prediction or network anomaly detection. Existing works on PU node classification overlook information in the graph structure, which can be critical. In this paper, we propose to better utilize graph structure for PU node classification. We first propose a distance-aware PU loss that uses homophily in graphs to introduce more accurate supervision. We also propose a regularizer to align the model with graph structure. Theoretical analysis shows that minimizing the proposed loss also leads to minimizing the expected loss with both positive and negative labels. Extensive empirical evaluation on diverse graph data sets demonstrates its superior performance over existing state-of-the-art methods.

### **Graph-based Alignment and Uniformity for Recommendation**

Liangwei Yang, Zhiwei Liu, Chen Wang, Mingdai Yang, Xiaolong Liu, Jing Ma, Philip S. Yu

Collaborative filtering-based recommender systems (RecSys) rely on learning representations for users and items to predict preferences accurately. Representation learning on the hypersphere is a promising approach due to its desirable properties, such as alignment and uniformity. However, the sparsity issue arises when it encounters RecSys. To address this issue, we propose a novel approach, graph-based alignment and uniformity (GraphAU), that explicitly considers high-order connectivities in the user-item bipartite graph. GraphAU aligns the user/item embedding to the dense vector representations of high-order neighbors using a neighborhood aggregator, eliminating the need to compute the burdensome alignment to high-order neighborhoods individually. To address the discrepancy in alignment losses, GraphAU includes a layer-wise alignment pooling module to integrate alignment losses layer-wise. Experiments on four datasets show that GraphAU significantly alleviates the sparsity issue and achieves state-of-the-art performance. We open-source GraphAU at <https://github.com/YangLiangwei/GraphAU>.

### **Toward a Foundation Model for Time Series Data**

Chin-Chia Michael Yeh, Xin Dai, Huiyuan Chen, Yan Zheng, Yujie Fan, Audrey Der, Vivian Lai, Zhongfang Zhuang, Junpeng Wang, Liang Wang, Wei Zhang

A foundation model is a machine learning model trained on a large and diverse set of data, typically using self-supervised learning-based pre-training techniques, that can be adapted to various downstream tasks. However, current research on time series pre-training has predominantly focused on models trained exclusively on data from a single domain. As a result, these models possess domain-specific knowledge that may not be easily transferable to time series from other domains. In this paper, we aim to develop an effective time series foundation model by leveraging unlabeled samples from multiple domains. To achieve this, we repurposed the publicly available UCR Archive and evaluated four existing self-supervised learning-based pre-training methods, along with a novel method, on the datasets. We tested these methods using four popular neural network architectures for time series to understand how the pre-training methods interact with different network designs. Our experimental results show that pre-training improves downstream classification tasks by enhancing the convergence of the fine-tuning process. Furthermore, we found that the proposed pre-training method, when combined with the Transformer, outperforms the alternatives. The proposed method outperforms or achieves equal performance compared to the second best method in ~93% of downstream tasks.

### **Simplex2vec Backward: From Vectors Back to Simplicial Complex**

Huixin Zhan, Kun Zhang, Zhong Chen, Victor S. Sheng

Simplicial neural networks (SNNs) were proposed to generate higher-order simplicial complex representations as vectors that encode not only pairwise relationships but also higher-order interactions between nodes. Although these vectors allow us to consider richer data representations compared to typical graph convolution, most real-world graphs associated with molecule or human-related activities are often sensitive and might contain confidential information, e.g., molecular geometry or friend lists. However, little works investigate the potential threats for these simplicial complexes (higher-order interactions between nodes). We name this threat by Simplicial Complexes Reconstruction Attack (SCRA) and conduct this attack by studying whether the vectors can be inverted to (approximately) recover the simplicial complexes who used to generate them. Specifically, we first generate the vectors via a k-simplex2vec approach that extends the node2vec algorithm to simplices of higher dimensions to associate Euclidean vectors to simplicial complexes. We then present a Simplex2vec

Backward algorithm to perform the SCRA on  $k$ -simplex2vec vectors by pointwise mutual information (PMI) matrix reconstruction.

### **BI-GCN: Bilateral Interactive Graph Convolutional Network for Recommendation**

Yinan Zhang, Pei Wang, Congcong Liu, Xiwei Zhao, Hao Qi, Jie He, Junsheng Jin, Changping Peng, Zhangang Lin, Jingping Shao

Recently, Graph Convolutional Network (GCN) based methods have become novel state-of-the-arts for Collaborative Filtering (CF) based Recommender Systems. To obtain users' preferences over different items, it is a common practice to learn representations of users and items by performing embedding propagation on a user-item bipartite graph, and then calculate the preference scores based on the representations. However, in most existing algorithms, user/item representations are generated independently of target items/users. To address this problem, we propose a novel graph attention model named Bilateral Interactive GCN (BI-GCN), which introduces bilateral interactive guidance into each user-item pair and thus leads to target-aware representations for preference prediction. Specifically, to learn the user/item representation from its neighborhood, we assign higher attention weights to those neighbors similar to the target item/user. By this manner, we can obtain target-aware representations, i.e., the information of the target item/user is explicitly encoded in the corresponding user/item representation, for more precise matching. Extensive experiments on three benchmark datasets demonstrate the effectiveness and robustness of BI-GCN.

### **Unlocking the Potential of Deep Learning in Peak-Hour Series Forecasting**

Zhenwei Zhang, Xin Wang, Jingyuan Xie, Heling Zhang, Yuntao Gu

Unlocking the potential of deep learning in Peak-Hour Series Forecasting (PHSF) remains a critical yet underexplored task in various domains. While state-of-the-art deep learning models excel in regular Time Series Forecasting (TSF), they struggle to achieve comparable results in PHSF. This can be attributed to the challenges posed by the high degree of non-stationarity in peak-hour series, which makes direct forecasting more difficult than standard TSF. Additionally, manually extracting the maximum value from regular forecasting results leads to suboptimal performance due to models minimizing the mean deficit. To address these issues, this paper presents Seq2Peak, a novel framework designed specifically for PHSF tasks, bridging the performance gap observed in TSF models. Seq2Peak offers two key components: the CyclicNorm pipeline to mitigate the non-stationarity issue and a simple yet effective trainable-parameter-free peak-hour decoder with a hybrid loss function that utilizes both the original series and peak-hour series as supervised signals. Extensive experimentation on publicly available time series datasets demonstrates the effectiveness of the proposed framework, yielding a remarkable average relative improvement of 37.7% across four real-world datasets for both transformer- and non-transformer-based TSF models.

### **POSPAN: Position-Constrained Span Masking for Language Model Pre-training**

Zhenyu Zhang, Lei Shen, Yuming Zhao, Meng Chen, Xiaodong He

Span-level masked language modeling (MLM) has shown to be advantageous to pre-trained language models over the original single-token MLM, as entities/phrases and their dependencies are critical to language understanding. Previous works only consider span length with some discrete distributions, while the dependencies among spans are ignored, i.e., assuming that the positions of masked spans are uniformly distributed. In this paper, we present POSPAN, a general framework to allow diverse position-constrained span masking strategies via the combination of span length distribution and position constraint distribution, which unifies all existing span-level masking methods. To verify the effectiveness of POSPAN in pre-training, we evaluate it on the datasets from several NLU benchmarks. Experimental results indicate that the position constraint is capable of enhancing span-level masking broadly, and our best POSPAN setting consistently outperforms its span-length-only counterparts and vanilla MLM. We also conduct theoretical analysis for the position constraint in masked language models to shed light on the reason why POSPAN works well, demonstrating the rationality and necessity of POSPAN.

### **FiBiNet++: Reducing Model Size by Low Rank Feature Interaction Layer for CTR Prediction**

Pengtao Zhang, Zheng Zheng, Junlin Zhang

Click-Through Rate (CTR) estimation has become one of the most fundamental tasks in many real-world applications and various deep models have been proposed. Some research has proved that FiBiNet is one of the best performance models and outperforms all other models on Avazu dataset. However, the large model size of FiBiNet hinders its wider application. In this paper, we propose a novel FiBiNet++ model to redesign FiBiNet's model structure, which greatly reduces model size while further improves its performance. One of the primary techniques involves our proposed "Low Rank Layer" focused on

feature interaction, which serves as a crucial driver of achieving a superior compression ratio for models. Extensive experiments on three public datasets show that FiBiNet++ effectively reduces non-embedding model parameters of FiBiNet by 12x to 16x on three datasets. On the other hand, FiBiNet++ leads to significant performance improvements compared to state-of-the-art CTR methods, including FiBiNet. The source code is in <https://github.com/recommendation-algorithm/FiBiNet>.

### **Knowledge Graph Error Detection with Hierarchical Path Structure**

Zhao Zhang, Fuwei Zhang, Fuzhen Zhuang, Yongjun Xu

Knowledge graphs (KGs) play a pivotal role in AI-related applications. In order to construct or continuously enrich KGs, automatic knowledge construction and update mechanisms are usually utilized, which inevitably bring in plenty of noise, and noise would degrade the performance of downstream applications. Existing KG error detection methods utilize the embeddings of entities and relations, or directly leverage the paths between entities to test the plausibility of triples, while ignore the valuable hierarchical information contained in the paths between entities. Indeed, the paths between a pair of entities conform to a hierarchical structure. Specifically, there may be a number of paths between two entities, and each path is comprised of several relations. The hierarchical structure is able to provide precious information, and is beneficial to leverage the path information in a fine-grained manner. To this end, in this paper, we propose a novel model named KG error detection with HiErarchical pAth stRucture (HEAR for short). Particularly, for a given triple, HEAR first learns path representations with the relations contained in the path, then integrates all path representations, and at last predicts the plausibility of the triple. Finally, we extensively validate the superiority of HEAR against various state-of-the-art baselines.

### **XuanYuan 2.0: A Large Chinese Financial Chat Model with Hundreds of Billions Parameters**

Xuanyu Zhang, Qing Yang

Recently, with the popularity of ChatGPT, large-scale language models have experienced rapid development. However, there is a scarcity of open-sourced chat models specifically designed for the Chinese language, especially in the field of Chinese finance, at the scale of hundreds of billions. To address this gap, we introduce XuanYuan 2.0, the largest Chinese chat model to date, built upon the BLOOM-176B architecture. Additionally, we propose a novel training method called hybrid-tuning to mitigate catastrophic forgetting. By integrating general and domain-specific knowledge, as well as combining the stages of pre-training and fine-tuning, XuanYuan 2.0 is capable of providing accurate and contextually appropriate responses in the Chinese financial domain.

### **Weight Matters: An Empirical Investigation of Distance Oracles on Knowledge Graphs**

Ke Zhang, Jiageng Chen, Zixian Huang, Gong Cheng

Distance computation is a bottleneck that limits the performance of many applications based on knowledge graphs (KGs). One common approach to improving online distance computation is to offline precompute certain information to be stored in an index called distance oracle. However, its effectiveness remains under-studied in the setting where edges are methodologically weighted to capture the structure and semantics of edge types in a KG. To fill the gap, in this paper, we present the first evaluation of representative distance oracles on KGs with commonly used edge weighting schemes. Our negative results and empirical justifications provide insights and a motivation for future studies of this unique setting.

### **Differentiable Retrieval Augmentation via Generative Language Modeling for E-commerce Query Intent Classification**

Chenyu Zhao, Yunjiang Jiang, Yiming Qiu, Han Zhang, Wen-Yun Yang

Retrieval augmentation, which enhances downstream models by a knowledge retriever and an external corpus instead of by merely increasing the number of model parameters, has been successfully applied to many natural language processing(NLP) tasks such as text classification, question answering and so on. However, existing methods that separately or asynchronously train the retriever and downstream model mainly due to the non-differentiability between the two parts, usually lead to degraded performance compared to end-to-end joint training. In this paper, we propose Differentiable Retrieval Augmentation via Generative LAnguage modeling(Dragan), to address this problem by a novel differentiable reformulation. We demonstrate the effectiveness of our proposed method on a challenging NLP task in e-commerce search, namely query intent classification. Both the experimental results and ablation study show that the proposed method significantly and reasonably improves the state-of-the-art baselines on both offline evaluation and online A/B test.

### **FCT-GAN: Enhancing Global Correlation of Table Synthesis via Fourier Transform**

Zilong Zhao, Robert Birke, Lydia Y. Chen



An alternative method for sharing knowledge while complying with strict data access regulations, such as the European General Data Protection Regulation (GDPR), is the emergence of synthetic tabular data. Mainstream table synthesizers utilize methodologies derived from Generative Adversarial Networks (GAN). Although several state-of-the-art (SOTA) tabular GAN algorithms inherit Convolutional Neural Network (CNN)-based architectures, which have proven effective for images, they tend to overlook two critical properties of tabular data: (i) the global correlation across columns, and (ii) the semantic invariance to the column order. Permuting columns in a table does not alter the semantic meaning of the data, but features extracted by CNNs can change significantly due to their limited convolution filter kernel size. To address the above problems, we propose FCT-GAN the first conditional tabular GAN to adopt Fourier networks into table synthesis. FCT-GAN enhances permutation invariant GAN training by strengthening the learning of global correlations via Fourier layers. Extensive evaluation on benchmarks and real-world datasets show that FCT-GAN can synthesize tabular data with better (up to 27.8%) machine learning utility (i.e. a proxy of global correlations) and higher (up to 26.5%) statistical similarity to real data. FCT-GAN also has the least variation on synthetic data quality among 7 SOTA baselines on 3 different training-data column orders.

#### [A Semi-Supervised Anomaly Network Traffic Detection Framework via Multimodal Traffic Information Fusion](#)

Yu Zheng, Xinglin Lian, Zhangxuan Dang, Chunlei Peng, Chao Yang, Jianfeng Ma

Anomaly traffic detection is a crucial issue in the cyber-security field. Previously, many researchers regarded anomaly traffic detection as a supervised classification problem. However, in real scenarios, anomaly network traffic is unpredictable, dynamically changing and difficult to collect. To address these limitations, we employ anomaly detection setting to propose a novel semi-supervised anomaly network traffic detection framework. It only learns features of normal samples during the training phase. Our framework utilizes low-pass filtering to extract multi-scale low-frequency information from 2-D traffic image. Furthermore, we design a two-stage fusion scheme to incorporate information from original and multi-scale low-frequency traffic image modalities. We conduct experiments on two public datasets: ISCX Tor-nonTor and USTC-TFC2016. The experimental results show that our method outperforms current state-of-the-art anomaly detection methods.

#### [LEAD-ID: Language-Enhanced Denoising and Intent Distinguishing Graph Neural Network for Sponsored Search Broad Retrievals](#)

Xiao Zhou, Ran Wang, Haorui Li, Qiang Liu, Xingxing Wang, Dong Wang

As a local-based service (LBS), search ad retrieval in online meal delivery platforms should be broader to bridge the gap between vague consumption intentions of users and shortage of ad candidates limited by users' queries and positions. Recently, graph neural networks (GNNs) have been successfully applied to search ad retrieval task. However, directly applying GNNs suffer from noisy interactions and intents indistinguishability, which seriously degrades systems' effectiveness in the broad retrieval. In this paper, we propose a Language-Enhanced Denoising and Intent Distinguishing graph neural network, LEAD-ID, which is developed and deployed at Meituan for sponsored search broad retrieval. To denoise interaction data, LEAD-ID designs hard- and soft- denoising strategies for GNNs based on a pretrained language model. A variational EM method is also employed to reduce high computational complexity of combining LMs and GNNs jointly. To distinguish various intents, LEAD-ID generates intent-aware node representations based on meticulously crafted LMs (language model) and GNNs; and then, it is guided by a contrastive learning object in an explicit and effective manner. According to offline experiments and online A/B tests, our framework significantly outperforms baselines in terms of recall and revenue.

#### [Target-oriented Few-shot Transferring via Measuring Task Similarity](#)

Zhipeng Zhou, Wei Gong, Haoquan Zhou

Despite significant progress in recent years, few-shot learning (FSL) still faces two critical challenges. Firstly, most FSL solutions in the training phase rely on exploiting auxiliary tasks, while target tasks are underutilized. Secondly, current benchmarks sample numerous target tasks, each with only an N-way C-shot query set in the evaluation phase, which is not representative of real-world scenarios. To address these issues, we propose Guidepost, a target-oriented FSL method that can implicitly learn task similarities using a task-level learn-to-learn mechanism and then re-weight auxiliary tasks. Additionally, we introduce a new FSL benchmark that satisfies realistic needs and aligns with our target-oriented approach. Mainstream FSL methods struggle under this new experimental setting. Extensive experiments demonstrate that Guidepost outperforms two classical few-shot learners, i.e., MAML and ProtoNet, and one state-of-the-art few-shot learner, i.e., RENet, on several FSL image datasets. Furthermore, we implement Guidepost as a domain adaptor to achieve high accuracy wireless sensing on our collected WiFi-based human activity recognition dataset.

**SESSION: Applied Research Papers****Enhancing E-commerce Product Search through Reinforcement Learning-Powered Query Reformulation**

Sanjay Agrawal, Srujana Merugu, Vivek Sembium

Query reformulation (QR) is a widely used technique in web and product search. In QR, we map a poorly formed or low coverage user query to a few semantically similar queries that are rich in product coverage, thereby enabling effective targeted searches with less cognitive load on the user. Recent QR approaches based on generative language models are superior to informational retrieval-based methods but exhibit key limitations: (i) generated reformulations often have low lexical diversity and fail to retrieve a large set of relevant products of a wider variety, (ii) the training objective of generative models does not incorporate a our goal of improving product coverage. In this paper, we propose RLQR (Reinforcement Learning for Query Reformulations), for generating high quality diverse reformulations which aim to maximize the product coverage (number of distinct relevant products returned). We evaluate our approach against supervised generative models and strong RL-based methods. Our experiments demonstrate a 28.6% increase in product coverage compared to a standard generative model, outperforming SOTA benchmarks by a significant margin. We also conduct our experiments on an external Amazon shopping dataset and demonstrate increased product coverage over SOTA algorithms.

**Nowcast-to-Forecast: Token-Based Multiple Remote Sensing Data Fusion for Precipitation Forecast**

Sojung An

Accurate short-term precipitation forecast is of social and economic significance for preventing severe weather damage. Deep learning has been rapidly adopted in nowcasting based on weather radar, which plays a key role in preventing dangerous weather conditions such as torrential rainfall. However, the limited observation range of the radar imposes constraints on shorter forecast lead times. Securing a sufficient lead time for timely flood warnings and emergency responses is crucial. Here, we propose a novel GAN-based framework that combines radar and satellite data to extend forecast lead time. First, we tokenize the satellite image to align with radar dimensions and combine the satellite and radar data. We then apply positional encoding to add positional information. Second, we design the self-conditioned generator to estimate distributions of various rainfall intensities. Finally, we employ Gaussian Fourier features to map the input noise into a continuous representation. The proposed framework realistically and accurately produces time series images of various precipitation types. Furthermore, our multisource data-driven system outperforms numerical weather prediction at forecasts of up to 6 hours in South Korea.

**Regression Compatible Listwise Objectives for Calibrated Ranking with Binary Relevance**

Aijun Bai, Rolf Jagerman, Zhen Qin, Le Yan, Pratyush Kar, Bing-Rong Lin, Xuanhui Wang, Michael Bendersky, Marc Najork

As Learning-to-Rank (LTR) approaches primarily seek to improve ranking quality, their output scores are not scale-calibrated by design. This fundamentally limits LTR usage in score-sensitive applications. Though a simple multi-objective approach that combines a regression and a ranking objective can effectively learn scale-calibrated scores, we argue that the two objectives are not necessarily compatible, which makes the trade-off less ideal for either of them. In this paper, we propose a practical regression compatible ranking (RCR) approach that achieves a better trade-off, where the two ranking and regression components are proved to be mutually aligned. Although the same idea applies to ranking with both binary and graded relevance, we mainly focus on binary labels in this paper. We evaluate the proposed approach on several public LTR benchmarks and show that it consistently achieves either best or competitive result in terms of both regression and ranking metrics, and significantly improves the Pareto frontiers in the context of multi-objective optimization. Furthermore, we evaluated the proposed approach on YouTube Search and found that it not only improved the ranking quality of the production pCTR model, but also brought gains to the click prediction accuracy. The proposed approach has been successfully deployed in the YouTube production system.

**CallMine: Fraud Detection and Visualization of Million-Scale Call Graphs**

Mirela Cazzolato, Saranya Vijayakumar, Meng-Chieh Lee, Catalina Vajiac, Namyong Park, Pedro Fidalgo, Agma J.M. Traina, Christos Faloutsos

Given a million-scale dataset of who-calls-whom data containing imperfect labels, how can we detect existing and new fraud patterns? We propose CallMine, with carefully designed features and visualizations. Our CallMine method has the following properties: (a) Scalable, being linear on the input size, handling about 35 million records in around one hour on a stock laptop; (b) Effective, allowing natural interaction with human analysts; (c) Flexible, being applicable in both supervised and unsupervised settings; (d) Automatic, requiring no user-defined parameters.

In the real world, in a multi-million-scale dataset, CallMine was able to detect fraudsters 7,000x faster, namely in a matter of hours, while expert humans took over 10 months to detect them.

CIKM-ARP Categories: Application; Analytics and machine learning; Data presentation.

### **Beyond Semantics: Learning a Behavior Augmented Relevance Model with Self-supervised Learning**

Zeyuan Chen, Wei Chen, Jia Xu, Zhongyi Liu, Wei Zhang

Relevance modeling aims to locate desirable items for corresponding queries, which is crucial for search engines to ensure user experience. Although most conventional approaches address this problem by assessing the semantic similarity between the query and item, pure semantic matching is not everything. In reality, auxiliary query-item interactions extracted from user historical behavior data of the search log could provide hints to reveal users' search intents further. Drawing inspiration from this, we devise a novel Behavior Augmented Relevance Learning model for Alipay Search (BARL-ASe) that leverages neighbor queries of target item and neighbor items of target query to complement target query-item semantic matching. Specifically, our model builds multi-level co-attention for distilling coarse-grained and fine-grained semantic representations from both neighbor and target views. The model subsequently employs neighbor-target self-supervised learning to improve the accuracy and robustness of BARL-ASe by strengthening representation and logit learning. Furthermore, we discuss how to deal with the long-tail query-item matching of the mini apps search scenario of Alipay practically. Experiments on real-world industry data and online A/B testing demonstrate our proposal achieves promising performance with low latency.

### **Monotonic Neural Ordinary Differential Equation: Time-series Forecasting for Cumulative Data**

Zhichao Chen, Leilei Ding, Zhixuan Chu, Yucheng Qi, Jianmin Huang, Hao Wang

Time-Series Forecasting based on Cumulative Data (TSFCD) is a crucial problem in decision-making across various industrial scenarios. However, existing time-series forecasting methods often overlook two important characteristics of cumulative data, namely monotonicity and irregularity, which limit their practical applicability. To address this limitation, we propose a principled approach called Monotonic neural Ordinary Differential Equation (MODE) within the framework of neural ordinary differential equations. By leveraging MODE, we are able to effectively capture and represent the monotonicity and irregularity in practical cumulative data. Through extensive experiments conducted in a bonus allocation scenario, we demonstrate that MODE outperforms state-of-the-art methods, showcasing its ability to handle both monotonicity and irregularity in cumulative data and delivering superior forecasting performance.

### **Towards Understanding of Deepfake Videos in the Wild**

Beomsang Cho, Binh M. Le, Jiwon Kim, Simon Woo, Shahroz Tariq, Alsharif Abuadbba, Kristen Moore

**Abstract:** Deepfakes have become a growing concern in recent years, prompting researchers to develop benchmark datasets and detection algorithms to tackle the issue. However, existing datasets suffer from significant drawbacks that hamper their effectiveness. Notably, these datasets fail to encompass the latest deepfake videos produced by state-of-the-art methods that are being shared across various platforms. This limitation impedes the ability to keep pace with the rapid evolution of generative AI techniques employed in real-world deepfake production. Our contributions in this IRB-approved study are to bridge this knowledge gap from current real-world deepfakes by providing in-depth analysis. We first present the largest and most diverse and recent deepfake dataset, RWDF-23, collected from the wild to date, consisting of 2,000 deepfake videos collected from 4 platforms targeting 4 different languages span created from 21 countries: Reddit, YouTube, TikTok, and Bilibili. By expanding the dataset's scope beyond the previous research, we capture a broader range of real-world deepfake content, reflecting the ever-evolving landscape of online platforms. Also, we conduct a comprehensive analysis encompassing various aspects of deepfakes, including creators, manipulation strategies, purposes, and real-world content production methods. This allows us to gain valuable insights into the nuances and characteristics of deepfakes in different contexts. Lastly, in addition to the video content, we also collect viewer comments and interactions, enabling us to explore the engagements of internet users with deepfake content. By considering this rich contextual information, we aim to provide a holistic understanding of the evolving deepfake phenomenon and its impact on online platforms.

### **Continually-Adaptive Representation Learning Framework for Time-Sensitive Healthcare Applications**

Akash Choudhuri, Hankyu Jang, Alberto M. Segre, Philip M. Polgreen, Kishlay Jha, Bijaya Adhikari

Continual learning has emerged as a powerful approach to address the challenges of non-stationary environments, allowing machine learning models to adapt to new data while retaining the previously acquired knowledge. In time-sensitive healthcare applications, where entities such as physicians, hospital rooms, and medications exhibit continuous changes over

time, continual learning holds great promise, yet its application remains relatively unexplored. This paper aims to bridge this gap by proposing a novel framework, i.e., Continually-Adaptive Representation Learning, designed to adapt representations in response to changing data distributions in evolving healthcare applications. Specifically, the proposed approach develops a continual learning strategy wherein the context information (e.g., interactions) of healthcare entities is exploited to continually identify and retrain the representations of those entities whose context evolved over time. Moreover, different from existing approaches, the proposed approach leverages the valuable patient information present in clinical notes to generate accurate and robust healthcare embeddings. Notably, the proposed continually-adaptive representations have practical benefits in low-resource clinical settings where it is difficult to training machine learning models from scratch to accommodate the newly available data streams. Experimental evaluations on real-world healthcare datasets demonstrate the effectiveness of our approach in time-sensitive healthcare applications such as *Clostridioides difficile* (C.diff) Infection (CDI) incidence prediction task and medical intensive care unit transfer prediction task.

#### **Enhancing Catalog Relationship Problems with Heterogeneous Graphs and Graph Neural Networks Distillation**

Boxin Du, Rob Barton, Grant Galloway, Junzhou Huang, Shioulin Sam, Ismail Tutar, Changhe Yuan

Traditionally, catalog relationship problems in e-commerce stores have been handled as pairwise classification tasks, which limit the ability of machine learning models to learn from the diverse relationships among different entities in the catalog. In this paper, we leverage heterogeneous graphs and Graph Neural Networks (GNNs) for improving catalog relationship inference. We start from investigating how to create multi-entity, multi-relationship graphs from diverse relationship data sources, and then explore how to utilizing GNNs to leverage the knowledge of the constructed graph in a self-supervised fashion. We finally propose a distillation approach to transfer the knowledge learned by GNNs into a pairwise neural network for seamless deployment in the catalog pipeline that relies on pairwise input for inductive relationship inference. Our experiments exhibit that in two of the representative catalog relationship problems, Title Authority/Contributor Authority and Broken Variation, the proposed framework is able to improve the recall at 95% precision of a pairwise baseline by up to 33.6% and 14.0%, respectively. Our findings highlight the effectiveness of this approach in advancing catalog quality maintenance and accurate relationship modeling, with potential for broader industry adoption.

#### **DoRA: Domain-Based Self-Supervised Learning Framework for Low-Resource Real Estate Appraisal**

Wei-Wei Du, Wei-Yao Wang, Wen-Chih Peng

The marketplace system connecting demands and supplies has been explored to develop unbiased decision-making in valuing properties. Real estate appraisal serves as one of the high-cost property valuation tasks for financial institutions since it requires domain experts to appraise the estimation based on the corresponding knowledge and the judgment of the market. Existing automated valuation models reducing the subjectivity of domain experts require a large number of transactions for effective evaluation, which is predominantly limited to not only the labeling efforts of transactions but also the generalizability of new developing and rural areas. To learn representations from unlabeled real estate sets, existing self-supervised learning (SSL) for tabular data neglects various important features, and fails to incorporate domain knowledge. In this paper, we propose DoRA, a Domain-based self-supervised learning framework for low-resource Real estate Appraisal. DoRA is pre-trained with an intra-sample geographic prediction as the pretext task based on the metadata of the real estate for equipping the real estate representations with prior domain knowledge. Furthermore, inter-sample contrastive learning is employed to generalize the representations to be robust for limited transactions of downstream tasks. Our benchmark results on three property types of real-world transactions show that DoRA significantly outperforms the SSL baselines for tabular data, the graph-based methods, and the supervised approaches in the few-shot scenarios by at least 7.6% for MAPE, 11.59% for MAE, and 3.34% for HR10%. We expect DoRA to be useful to other financial practitioners with similar marketplace applications who need general models for properties that are newly built and have limited records. The source code is available at <https://github.com/wwwiwei/DoRA>.

#### **Content-Based Email Classification at Scale**

Kirstin Early, Neil O'Hare, Christopher Luvogt

Understanding the content of email messages can enable new features that highlight what matters to users, making email a more useful tool for people to manage their lives. We present work from a consumer email platform to build multilabel models to classify messages according to a mail-specific, content-based taxonomy that represents the topic, type, and objective of an email. While state-of-the-art Transformer-based language models can achieve impressive results for text classification, these models are too costly to deploy at the scale of email. Using a knowledge distillation framework, we first build a complex, accurate teacher model from limited human-labeled training data and then use a large amount of teacher-labeled

data to train lightweight student models that are suitable for deployment. The student models retain up to 91% of the predictive performance of the teacher model while reducing inference cost by three orders of magnitude. Deployed to production in Yahoo Mail, these models classify billions of emails every day and power features that help people tackle their inboxes.

### **Robust User Behavioral Sequence Representation via Multi-scale Stochastic Distribution Prediction**

Chilin Fu, Weichang Wu, Xiaolu Zhang, Jun Hu, Jing Wang, Jun Zhou

User behavior representation learned by self-supervised pre-training tasks is widely used in various domains and applications. Conventional methods usually follow the methodology in Natural Language Processing (NLP) to set the pre-training tasks. They either randomly mask some of the behaviors in the sequence and predict the masked ones or predict the next  $k$  behaviors. These methods fit for text sequence, in which the tokens are sequentially arranged subject to linguistic criterion. However, the user behavior sequences can be stochastic with noise and randomness. The same paradigm is intractable for learning a robust user behavioral representation.

Though the next user behavior can be stochastic, the behavior distribution over a period of time is much more stable and less noisy. Based on this, we propose a Multi-scale Stochastic Distribution Prediction (MSDP) algorithm for learning robust user behavioral sequence representation. Instead of using predictions on concrete behavior as pre-training tasks, we take the prediction on user's behaviors distribution over a period of time as the self-supervision signal. Moreover, inspired by the recent success of the multi-task prompt training method on Large Language Models (LLM), we propose using the window size of the predicted time period as a prompt, enabling the model to learn user behavior representations that can be applied to prediction tasks across various future time periods. We generate different window size prompts through stochastic sampling. It effectively improves the generalization capability of the learned sequence representation. Extensive experiments demonstrate that our approach can learn robust user behavior representation successfully, which significantly outperforms state-of-the-art (SOTA) baselines.

### **Rec4Ad: A Free Lunch to Mitigate Sample Selection Bias for Ads CTR Prediction in Taobao**

Jingyue Gao, Shuguang Han, Han Zhu, Siran Yang, Yuning Jiang, Jian Xu, Bo Zheng

Click-Through Rate (CTR) prediction serves as a fundamental component in online advertising. A common practice is to train a CTR model on advertisement (ad) impressions with user feedback. Since ad impressions are purposely selected by the model itself, their distribution differs from the inference distribution and thus exhibits sample selection bias (SSB) that affects model performance. Existing studies on SSB mainly employ sample re-weighting techniques which suffer from high variance and poor model calibration. Another line of work relies on costly uniform data that is inadequate to train industrial models. Thus mitigating SSB in industrial models with a uniform-data-free framework is worth exploring. Fortunately, many platforms display mixed results of organic items (i.e., recommendations) and sponsored items (i.e., ads) to users, where impressions of ads and recommendations are selected by different systems but share the same user decision rationales. Based on the above characteristics, we propose to leverage recommendations samples as a free lunch to mitigate SSB for ads CTR model (Rec4Ad). After elaborating data augmentation, Rec4Ad learns disentangled representations with alignment and decorrelation modules for enhancement. When deployed in Taobao display advertising system, Rec4Ad achieves substantial gains in key business metrics, with a lift of up to +6.6% CTR and +2.9% RPM.

### **Predicting Interaction Quality of Conversational Assistants With Spoken Language Understanding Model Confidences**

Yue Gao, Enrico Piovano, Tamer Soliman, Monir Moniruzzaman, Anoop Kumar, Melanie Bradford, Subhrangshu Nandi

In conversational AI assistants, SLU models are part of a complex pipeline composed of several modules working in harmony. Hence, an update to the SLU model needs to ensure improvements not only in the model specific metrics but also in the overall conversational assistant performance. Specifically, the impact on user interaction quality metrics must be factored in, while integrating interactions with distal modules upstream and downstream of the SLU component. We develop a ML model that makes it possible to gauge the interaction quality metrics due to SLU model changes before a production launch. The proposed model is a multi-modal transformer with a gated mechanism that conditions on text embeddings, output of a BERT model pre-trained on conversational data, and the hypotheses of the SLU classifiers with the corresponding confidence scores. We show that the proposed model predicts defect with more than 76% correlation with live interaction quality defects, compared to 46% baseline.

### **MEBS: Multi-task End-to-end Bid Shading for Multi-slot Display Advertising**



Zhen Gong, Lvyin Niu, Yang Zhao, Miao Xu, Haoqi Zhang, Zhenzhe Zheng, Zhilin Zhang, Rongquan Bai, Chuan Yu, Jian Xu, Bo Zheng, Fan Wu

Online bidding and auction are crucial aspects of the online advertising industry. Conventionally, there is only one slot for ad display and most current studies focus on it. Nowadays, multi-slot display advertising is gradually becoming popular where many ads could be displayed in a list and shown as a whole to users. However, multi-slot display advertising leads to different cost-effectiveness. Advertisers have the incentive to adjust bid prices so as to win the most economical ad positions. In this study, we introduce bid shading into multi-slot display advertising for bid price adjustment with a Multi-task End-to-end Bid Shading~(MEBS) method. We prove the optimality of our method theoretically and examine its performance experimentally. Through extensive offline and online experiments, we demonstrate the effectiveness and efficiency of our method, and we obtain a 7.01% lift in Gross Merchandise Volume, a 7.42% lift in Return on Investment, and a 3.26% lift in ad buy count.

#### **An Unified Search and Recommendation Foundation Model for Cold-Start Scenario**

Yuqi Gong, Xichen Ding, Yehui Su, Kaiming Shen, Zhongyi Liu, Guannan Zhang

In modern commercial search engines and recommendation systems, data from multiple domains is available to jointly train the multi-domain model. Traditional methods train multi-domain models in the multi-task setting, with shared parameters to learn the similarity of multiple tasks, and task-specific parameters to learn the divergence of features, labels, and sample distributions of individual tasks. With the development of large language models, LLM can extract global domain-invariant text features that serve both search and recommendation tasks. We propose a novel framework called S&R Multi-Domain Foundation, which uses LLM to extract domain invariant features, and Aspect Gating Fusion to merge the ID feature, domain invariant text features and task-specific heterogeneous sparse features to obtain the representations of query and item. Additionally, samples from multiple search and recommendation scenarios are trained jointly with Domain Adaptive Multi-Task module to obtain the multi-domain foundation model. We apply the S&R Multi-Domain foundation model to cold start scenarios in the pretrain-finetune manner, which achieves better performance than other SOTA transfer learning methods. The S&R Multi-Domain Foundation model has been successfully deployed in Alipay Mobile Application's online services, such as content query recommendation and service card recommendation, etc.

#### **DFFM: Domain Facilitated Feature Modeling for CTR Prediction**

Wei Guo, Chenxu Zhu, Fan Yan, Bo Chen, Weiwen Liu, Huifeng Guo, Hongkun Zheng, Yong Liu, Ruiming Tang

CTR prediction is critical to industrial recommender systems. Recently, with the growth of business domains in enterprises, much attention has been focused on the multi-domain CTR recommendation. Numerous models have been proposed that attempt to use a unified model to serve multiple domains. Although much progress has been made, we argue that they ignore the importance of feature interactions and user behaviors when modeling cross-domain relations, which is a coarse-grained utilizing of domain information. To solve this problem, we propose Domain Facilitated Feature Modeling (DFFM) for CTR prediction. It incorporates domain-related information into the parameters of the feature interaction and user behavior modules, allowing for domain-specific learning of these two aspects. Extensive experiments are conducted on two public datasets and one industrial dataset to demonstrate the effectiveness of DFFM. We deploy the DFFM model in Huawei advertising platform and gain a 4.13% improvement of revenue on a two week online A/B test. Currently DFFM model has been used as the main traffic model, serving for hundreds of millions of people.

#### **Learning To Rank Diversely At Airbnb**

Malay Haldar, Mustafa Abdool, Liwei He, Dillon Davis, Huiji Gao, Sanjeev Katariya

Airbnb is a two-sided marketplace, bringing together hosts who own listings for rent, with prospective guests from around the globe. Applying neural network-based learning to rank techniques has led to significant improvements in matching guests with hosts. These improvements in ranking were driven by a core strategy: order the listings by their estimated booking probabilities, then iterate on techniques to make these booking probability estimates more and more accurate. Embedded implicitly in this strategy was an assumption that the booking probability of a listing could be determined independently of other listings in search results. In this paper we discuss how this assumption, pervasive throughout the commonly-used learning to rank frameworks, is false. We provide a theoretical foundation correcting this assumption, followed by efficient neural network architectures based on the theory. Explicitly accounting for possible similarities between listings, and reducing them to diversify the search results generated strong positive impact. We discuss these metric wins as part of the online A/B

tests of the theory. Our method provides a practical way to diversify search results for large-scale production ranking systems.

### [\*\*Continual Learning in Predictive Autoscaling\*\*](#)

Hongyan Hao, Zhixuan Chu, Shiyi Zhu, Gangwei Jiang, Yan Wang, Caigao Jiang, James Y. Zhang, Wei Jiang, Siqiao Xue, Jun Zhou

Predictive Autoscaling is used to forecast the workloads of servers and prepare the resources in advance to ensure service level objectives (SLOs) in dynamic cloud environments. However, in practice, its prediction task often suffers from performance degradation under abnormal traffics caused by external events (such as sales promotional activities and applications' re-configurations), for which a common solution is to re-train the model with data of a long historical period, but at the expense of high computational and storage costs. To better address this problem, we propose a replay-based continual learning method, i.e., Density-based Memory Selection and Hint-based Network Learning Model (DMSHM), using only a small part of the historical log to achieve accurate predictions. First, we discover the phenomenon of sample overlap when applying replay-based continual learning in prediction tasks. In order to surmount this challenge and effectively integrate new sample distribution, we propose a density-based sample selection strategy that utilizes kernel density estimation to calculate sample density as a reference to compute sample weight, and employs weight sampling to construct a new memory set. Then we implement hint-based network learning based on hint representation to optimize the parameters. Finally, we conduct experiments on public and industrial datasets to demonstrate that our proposed method outperforms state-of-the-art continual learning methods in terms of memory capacity and prediction accuracy. Furthermore, we demonstrate remarkable practicability of DMSHM in real industrial applications.

### [\*\*AutoBuild: Automatic Community Building Labeling for Last-mile Delivery\*\*](#)

Zhiqing Hong, Dongjiang Cao, Haotian Wang, Guang Wang, Tian He, Desheng Zhang

Fine-grained community-building information, such as building names and accurate geographical coordinates, is critical for a range of practical applications like navigation and door-to-door services (e.g., on-demand delivery and last-mile delivery). A common practice of traditional methods to gather community-building information usually relies on manual collection, which is typically labor-intensive and time-consuming. To address these issues, we utilize the massive data generated from e-commerce delivery services and design a framework, AutoBuild, for fine-grained large-scale community-building labeling. AutoBuild consists of two main components: (i) a Location Candidate Detection Module that identifies potential building names and coordinates from multi-source delivery data, and (ii) a Progressive Building Matching Model that employs trajectory modeling, human behavior analysis, and heterogeneous graph alignment to match building names and coordinates. To evaluate the performance of AutoBuild, we applied it to two real-world multi-modal datasets from Beijing City and Chengdu City. The results reveal that AutoBuild significantly outperforms multiple baseline models by 50-meter accuracy of 81.8% and 100-meter accuracy of 95.9% in Beijing City. More importantly, we conduct a real-world case study to demonstrate the practical impact of AutoBuild in last-mile delivery.

### [\*\*Urban-scale POI Updating with Crowd Intelligence\*\*](#)

Zhiqing Hong, Haotian Wang, Wenjun Lyu, Hai Wang, Yunhuai Liu, Guang Wang, Tian He, Desheng Zhang

Points of Interest (POIs), such as entertainment, dining, and living, are crucial for urban planning and location-based services. However, the high dynamics and expensive updating costs of POIs pose a key roadblock for their urban applications. This is especially true for developing countries, where active economic activities lead to frequent POI updates (e.g., merchants closing down and new ones opening). Therefore, POI updating, i.e., detecting new POIs and different names of the same POIs (alias) to update the POI database, has become an urgent but challenging problem to address. In this paper, we attempt to answer the research question of how to detect and update large-scale POIs via a low-cost approach. To do so, we propose a novel framework called UrbanPOI, which formulates the POI updating problem as a tagging and detection problem based on multi-modal logistics delivery data. UrbanPOI consists of two key modules: (i) a hierarchical POI candidate generation module based on the POINet model that detects POIs from shipping addresses; and (ii) a new POI detection module based on the Siamese Attention Network that models multi-modal data and crowd intelligence. We evaluate our framework on real-world logistics delivery datasets from two Chinese cities. Extensive results show that our model outperforms state-of-the-art models in Beijing City by 26.2% in precision and 10.7% in F1-score, respectively.

### [\*\*PS-SA: An Efficient Self-Attention via Progressive Sampling for User Behavior Sequence Modeling\*\*](#)

Jiacen Hu, Zhangming Chan, Yu Zhang, Shuguang Han, Siyuan Lou, Baolin Liu, Han Zhu, Yuning Jiang, Jian Xu, Bo Zheng

As the self-attention mechanism offers powerful capabilities for capturing sequential relationships, it has become increasingly popular to use it for modeling user behavior sequences in recommender systems. However, the self-attention mechanism has a quadratic computational complexity of  $O(n^2)$ , as it conducts interactions among all item pairs in the sequence. This can lead to expensive model training and slow inference speeds, which may hinder practical deployment. To this end, we pursue to develop alternative approaches to improve the efficiency of the self-attention mechanism. We observe that the attention scores calculated from each item interacting with other items (including itself) are sparse, indicating that there are limited valuable item pairs (with non-zero attention weight) that contribute to the final output. This motivates us to develop effective strategies for discerning valuable items and computing attention scores solely for these items, thereby minimizing the consumption of unnecessary computations. Herein, we present a novel Progressive Sampling-based Self-Attention (PS-SA) mechanism, which utilizes a learnable progressive sampling strategy to identify the most valuable items. Subsequently, we solely utilize these selected items to produce the final output. Experiments on academic and production datasets demonstrate PS-SA could still achieve promising results while reducing computational costs. It is notable that we have successfully deployed it on Alibaba display advertising system, resulting in a 2.6% CTR and 1.3% RPM increase.

#### [A Stochastic Online Forecast-and-Optimize Framework for Real-Time Energy Dispatch in Virtual Power Plants under Uncertainty](#)

Wei Jiang, Zhongkai Yi, Li Wang, Hanwei Zhang, Jihai Zhang, Fangquan Lin, Cheng Yang

Aggregating distributed energy resources in power systems significantly increases uncertainties, in particular caused by the fluctuation of renewable energy generation. This issue has driven the necessity of widely exploiting advanced predictive control techniques under uncertainty to ensure long-term economics and decarbonization. In this paper, we propose a real-time uncertainty-aware energy dispatch framework, which is composed of two key elements: (i) A hybrid forecast-and-optimize sequential task, integrating deep learning-based forecasting and stochastic optimization, where these two stages are connected by the uncertainty estimation at multiple temporal resolutions; (ii) An efficient online data augmentation scheme, jointly involving model pre-training and online fine-tuning stages. In this way, the proposed framework is capable to rapidly adapt to the real-time data distribution, as well as to target on uncertainties caused by data drift, model discrepancy and environment perturbations in the control process, and finally to realize an optimal and robust dispatch solution. The proposed framework won the championship in CityLearn Challenge 2022, which provided an influential opportunity to investigate the potential of AI application in the energy domain. In addition, comprehensive experiments are conducted to interpret its effectiveness in the real-life scenario of smart building energy management.

#### [Climate Intervention Analysis using AI Model Guided by Statistical Physics Principles](#)

Sookyung Kim, Kalai Ramea, Salva Rühling Cachay, Haruki Hirasawa, Subhashis Hazarika, Dipti Hingmire, Peetak Mitra, Philip J. Rasch, Hansi Singh

In this study, we propose a solution to estimating system responses to external forcings or perturbations. We utilize the Fluctuation-Dissipation Theorem (FDT) from statistical physics to extract knowledge using an AI model that can rapidly produce scenarios for different external forcings by leveraging FDT and analyzing a large dataset from Earth System Models. Our model, AiBEDO, accurately captures the complex effects of radiation perturbations on global and regional surface climate, enabling faster exploration of the impacts of spatially-heterogeneous climate forcings. We demonstrate its effectiveness by applying AiBEDO to Marine Cloud Brightening, a climate intervention technique, aiming to optimize cloud brightening patterns for regional climate targets and prevent climate tipping points. Our approach has broader applicability to other scientific disciplines with computationally demanding simulation models. Source code of AiBEDO framework is made available at [https://github.com/kramea/cikm\\_aibedo](https://github.com/kramea/cikm_aibedo). A sample dataset is made available at <https://doi.org/10.5281/zenodo.7597027>. Additional data available upon request.

#### [Generating Product Insights from Community Q&A](#)

Lital Kuchy, Ran Levy, Avihai Mejer, Noam Segev, Shunit Agmon, Miriam Farber

In e-commerce sites, customer questions on the product details-page express the customers' information needs about the product. The answers to these questions often provide the necessary information. In this work, we present and address the novel task of generating product insights from community questions and answers (Q&A). These insights can be presented to customers to assist them in their shopping journey. Our method first generates concise, self-contained sentences based on the information in the Q&A. Then insights are selected based on the prominence of their associated questions. Empirical

evaluation attests to the effectiveness of our approach in generating well-formed, objective, and helpful insights that are often not available in the product description or in summaries of customer reviews.

### **Unsupervised Multi-Modal Representation Learning for High Quality Retrieval of Similar Products at E-commerce Scale**

Kushal Kumar, Tarik Arici, Tal Neiman, Jinyu Yang, Shioulin Sam, Yi Xu, Hakan Ferhatosmanoglu, Ismail Tutar

Identifying similar products in e-commerce is useful in discovering relationships between products, making recommendations, and increasing diversity in search results. Product representation learning is the first step to define a generalized product similarity metric for search. The second step is to extend similarity search to a large scale (e.g., e-commerce catalog scale) without sacrificing quality. In this work, we present a solution that interweaves both steps, i.e., learn representations suited to high quality retrieval using contrastive learning (CL) and retrieve similar items from a large search space using approximate nearest neighbor search (ANNS) to trade-off quality for speed. We propose a CL training strategy for learning uni-modal encoders suited to multi-modal similarity search for e-commerce. We study ANNS retrieval by generating Pareto Frontiers (PFs) without requiring labels. Our CL training strategy doubles retrieval@1 metric across categories (e.g., from 36% to 88% in category C). We also demonstrate that ANNS engine optimization using PFs help select configurations appropriately (e.g., we achieve 6.8× search speed with just 2% drop from the maximum retrieval accuracy in medium size datasets).

### **Addressing Selection Bias in Computerized Adaptive Testing: A User-Wise Aggregate Influence Function Approach**

Soonwoo Kwon, Sojung Kim, Seunghyun Lee, Jin-Young Kim, Suyeong An, Kyuseok Kim

Computerized Adaptive Testing (CAT) is a widely used, efficient test mode that adapts to the examinee's proficiency level in the test domain. CAT requires pre-trained item profiles, for CAT iteratively assesses the student real-time based on the registered items' profiles, and selects the next item to administer using candidate items' profiles. However, obtaining such item profiles is a costly process that involves gathering a large, dense item-response data, then training a diagnostic model on the collected data. In this paper, we explore the possibility of leveraging response data collected in the CAT service. We first show that this poses a unique challenge due to the inherent selection bias introduced by CAT, i.e., more proficient students will receive harder questions. Indeed, when naïvely training the diagnostic model using CAT response data, we observe that item profiles deviate significantly from the ground-truth. To tackle the selection bias issue, we propose the user-wise aggregate influence function method. Our intuition is to filter out users whose response data is heavily biased in an aggregate manner, as judged by how much perturbation the added data will introduce during parameter estimation. This way, we may enhance the performance of CAT while introducing minimal bias to the item profiles. We provide extensive experiments to demonstrate the superiority of our proposed method based on the three public datasets and one dataset that contains real-world CAT response data.

### **The Price is Right: Removing A/B Test Bias in a Marketplace of Expirable Goods**

Thu Le, Alex Deng

Pricing Guidance tools at Airbnb aim to help hosts maximize the earning for each night of stay. For a given listing, the earning-maximization price point of a night can vary greatly with lead-day - the number of days from now until the night of stay. This introduces systematic bias in running marketplace A/B tests to compare the performances of two pricing strategies. Lead-day bias can cause the short-term experiment result to move in the opposite direction to the long-term impact, possibly leading to the suboptimal business decision and customer dissatisfaction. We propose an efficient experimentation approach that corrects for the bias, minimizes the possible negative impact of experimenting, and greatly accelerates the R&D cycle. This paper is the first of its kind to lays out the theoretical framework along with the real-world example that demonstrates the magnitude of the bias. It serves as a conversation starter for such insidious type of experimentation bias that is likely present in other marketplaces of expirable goods such as vacation nights, car rentals, and airline tickets, concert passes, or ride-hailings.

### **Fragment and Integrate Network (FIN): A Novel Spatial-Temporal Modeling Based on Long Sequential Behavior for Online Food Ordering Click-Through Rate Prediction**

Jun Li, Ge Zhang

Spatial-temporal information has been proven to be of great significance for click-through rate prediction tasks in online Location-Based Services (LBS), especially in mainstream food ordering platforms such as DoorDash, Uber Eats, Meituan, and Ele.me. Modeling user spatial-temporal preferences with sequential behavior data has become a hot topic in recommendation systems and online advertising. However, most of existing methods either lack the representation of rich spatial-temporal information or only handle user behaviors with limited length, e.g. 100. In this paper, we tackle these

problems by designing a new spatial-temporal modeling paradigm named Fragment and Integrate Network (FIN). FIN consists of two networks: (i) Fragment Network (FN) extracts Multiple Sub-Sequences (MSS) from lifelong sequential behavior data, and captures the specific spatial-temporal representation by modeling each MSS respectively. Here both a simplified attention and a complicated attention are adopted to balance the performance gain and resource consumption. (ii) Integrate Network (IN) builds a new integrated sequence by utilizing spatial-temporal interaction on MSS and captures the comprehensive spatial-temporal representation by modeling the integrated sequence with a complicated attention. Both public datasets and production datasets have demonstrated the accuracy and scalability of FIN. Since 2022, FIN has been fully deployed in the recommendation advertising system of Ele.me, one of the most popular online food ordering platforms in China, obtaining 5.7% improvement on Click-Through Rate (CTR) and 7.3% increase on Revenue Per Mille (RPM).

#### [\*\*A Hierarchical Imitation Learning-based Decision Framework for Autonomous Driving\*\*](#)

Hebin Liang, Zibin Dong, Yi Ma, Xiaotian Hao, Yan Zheng, Jianye Hao

In this paper, we focus on the decision-making challenge in autonomous driving, a central and intricate problem influencing the safety and practicality of autonomous vehicles. We propose an innovative hierarchical imitation learning framework that effectively alleviates the complexity of learning in autonomous driving decision-making problems by decoupling decision-making tasks into sub-problems. Specifically, the decision-making process is divided into two levels of sub-problems: the upper level directs the vehicle's lane selection and qualitative speed management, while the lower level implements precise control of the driving speed and direction. We harness Transformer-based models for solving each sub-problem, enabling overall hierarchical framework to comprehend and navigate diverse and various road conditions, ultimately resulting in improved decision-making. Through an evaluation in several typical driving scenarios within the SMARTS autonomous driving simulation environment, our proposed hierarchical decision-making framework significantly outperforms end-to-end reinforcement learning algorithms and behavior cloning algorithm, achieving an average pass rate of over 90%. Our framework's effectiveness is substantiated by its commendable achievements at the NeurIPS 2022 Driving SMARTS competition, where it secures dual track championships.

#### [\*\*Enhancing Dynamic On-demand Food Order Dispatching via Future-informed and Spatial-temporal Extended Decisions\*\*](#)

Yile Liang, Donghui Li, Jiuxia Zhao, Xuetao Ding, Huanjia Lian, Jinghua Hao, Renqing He

On-demand food delivery (OFD) service has gained fast-growing popularity all around the world. Order dispatching is instrumental to large-scale OFD platforms, such as Meituan, which continuously match food order requests to couriers at a scale of tens of millions each day to satisfy the needs of consumers, couriers, and merchants. However, due to high dynamism and inevitable uncertainties in the real-world environment, it is not an easy task to achieve long-term global objective optimization through continuous isolated optimization decisions at each dispatch moment. Our work proposes the concept of "courier occupancy" (CO) to precisely quantify the impact of order assignment on the courier's delivery efficiency, realizing a decomposition of long-term and macro goals into various dispatch moments and micro decision-making dimensions. Then in the prediction phase, an improved and universally applicable distribution estimation method is designed to quantify CO which is a stochastic variable and contains future information, combining Monte Carlo dropout and knowledge distillation. In the optimization phase, we use CO to model the objective function at each dispatch moment to introduce future information and extend dispatch decisions from merely who to assign the order to both when and who to assign it, significantly enhancing the long-term optimization capability of dispatching decisions and avoiding local greed. We conduct extensive offline simulations based on real dispatching data as well as online AB tests through Meituan's platform. Results show that our method consistently improves the couriers' delivery efficiency and consumers' satisfaction.

#### [\*\*Multitask Ranking System for Immersive Feed and No More Clicks: A Case Study of Short-Form Video Recommendation\*\*](#)

Qingyun Liu, Zhe Zhao, Liang Liu, Zhen Zhang, Junjie Shan, Yuening Li, Shuchao Bi, Lichan Hong, Ed H. Chi

In recent years, social media users spend significant amount of time on Short-Form Video (SFV) platforms. Its success in creating an immersive viewership experience is not only from the content, but also due to its unique UI innovation: instead of providing choices for users to click, SFV platforms actively recommend content to users to watch one at a time. In this paper, we highlight unique challenges rooted from such UI changes for SFV recommendation system design. Firstly, there is yet much unexplored for sources of system biases under the new UI, as there are no clicks nor the common click-based position biases. Additionally, when training multiple types of user activities, positive labels for activities like sharing and commenting can be much sparser and more skewed than traditional click-based recommendation systems, as the latter can filter non-click impressions when generating "post-click" activities.



To tackle these challenges, we introduce a unified multi-task ranking framework which puts two novel components all together into an overall system for SFV recommendation. First, we identify that there are position biases of SFVs in the recommendation sequence, namely "watch trail biases", and introduce biases correction using trail-related information. Second, to get the most benefits from multi-task learning, especially co-training tasks with extremely skewed and sparse labels, we adapt a disentangle regularization to mitigate task conflicts, introduce loss upweighting for sparse task co-training and adopt a meta-learning algorithm for efficient weight selection. We demonstrate the effectiveness and efficiency of the framework on one of today's largest SFV platforms. Our framework has been deployed to the production system for more than 6 months.

#### **FAF: A Risk Detection Framework on Industry-Scale Graphs**

Yice Luo, Guannan Wang, Yongchao Liu, Jiaxin Yue, Weihong Cheng, Binjie Fei

It is neither effective nor profitable for individuals to attempt to bypass Ant Group's comprehensive risk control system. However, there are still criminals (such as underground loan sharks and professional hackers,), who may teach borrowers how to use hacking techniques to circumvent Ant Group and its partners' risk control system. Despite the fact that our risk control system has greatly reduced fraud losses for merchants in practice, a significant number of intermediary-related frauds still occur. During our investigation into fraud events at Zhima Credit Renting (ZCR), we discovered that more than 30 percent of fraud cases were directly linked to malicious intermediaries. To address this issue, we propose an anti-abettor fraud detection framework, called FAF (Fraud-Abettor-Fraud), specifically designed to combat intermediary-related frauds on an industry-wide scale. We have developed a series of algorithms under the FAF framework, which outperform commonly used risk detection methods and meet real-world business requirements. In this paper, we use ZCR's risk management application as a real-world example-which has deployed FAF for over 1 year-to demonstrate the superiority of the FAF framework compared to existing methods.

#### **pADR: Towards Personalized Adverse Drug Reaction Prediction by Modeling Multi-sourced Data**

Junyu Luo, Cheng Qian, Xiaochen Wang, Lucas Glass, Fenglong Ma

Predicting adverse drug reactions (ADRs) of drugs is one of the most critical steps in drug development. By pre-estimating the adverse reactions, researchers and drug development companies can greatly prevent the potential ADR risks and tragedies. However, the current ADR prediction methods suffer from several limitations. First, the prediction results are based on pure drug-related information, which makes them impossible to be directly applied for the personalized ADR prediction task. The lack of personalization of models also makes rare adverse events hard to be predicted. Therefore, it is of great interest to develop a new personalized ADR prediction method by introducing additional sources, e.g., patient health records. However, few methods have tried to use additional sources. In the meantime, the variety of different source formats and structures makes this task more challenging. To address the above challenges, we propose a novel personalized multi-sourced-based drug adverse reaction prediction model named pADR. pADR first works on every single source to transform them into proper representations. Next, a hierarchical multi-sourced Transformer is designed to automatically model the interactions between different sources and fuse them together for the final adverse event prediction. Experimental results on a new multi-sourced ADR prediction dataset show that pADR1 outperforms state-of-the-art drug-based baselines. Moreover, the case and ablation studies also illustrate the effectiveness of our proposed fusion strategies and the reasonableness of each module design.

#### **Dynamic Group Parameter Modeling for Click-Through-Rate Prediction**

Xuan Ma, Jian Wang, Zhiyuan Chen, Zehua Zhang, Jie He, Changping Peng, Zhangang Lin, Jingping Shao

It is noted that Click-Through-Rate(CTR) prediction plays an important part in recommendation systems and online advertising. Over the past few years, numerous studies have been conducted to improve the accuracy of CTR prediction by exploring data inherent patterns. These studies indicate that training CTR models with group-specific parameters on divided data groups can lead to significant improvements. However, most works generally divide groups manually with some prior knowledge, and such a fixed group division method may hinder the expression of user common interests. To address this limitation, we propose a novel group parameter modeling method, where the user group division and group parameter learning processes are completed in an automatic and dynamic way. Our method employs a three-stage approach, consisting of group information selection, group representation learning, and group parameter generation, which allows the efficient expression of user common interests. We conduct experiments on both public datasets and industrial datasets, and the experimental results demonstrate the effectiveness of our method. We have also deployed the model in an online advertising system and observed significant improvements in both CTR and Revenue Per Mille (RPM).

### **BOLT: An Automated Deep Learning Framework for Training and Deploying Large-Scale Search and Recommendation Models on Commodity CPU Hardware**

Nicholas Meisburger, Vihan Lakshman, Benito Geordie, Joshua Engels, David Torres Ramos, Pratik Pranav, Benjamin Coleman, Benjamin Meisburger, Shubh Gupta, Yashwanth Adunukota, Siddharth Jain, Tharun Medini, Anshumali Shrivastava

Efficient large-scale neural network training and inference on commodity CPU hardware is of immense practical significance in democratizing deep learning (DL) capabilities. Presently, the process of training massive models consisting of hundreds of millions to billions of parameters requires the extensive use of specialized hardware accelerators, such as GPUs, which are only accessible to a limited number of institutions with considerable financial resources. Moreover, there is often an alarming carbon footprint associated with training and deploying these models. In this paper, we take a step towards addressing these challenges by introducing BOLT, a sparse deep learning library for training large-scale search and recommendation models on standard CPU hardware. BOLT provides a flexible, high-level API for constructing models that will be familiar to users of existing popular DL frameworks. By automatically tuning specialized hyperparameters, BOLT also abstracts away the algorithmic details of sparse network training. We evaluate BOLT on a number of information retrieval tasks including product recommendations, text classification, graph neural networks, and personalization. We find that our proposed system achieves competitive performance with state-of-the-art techniques at a fraction of the cost and energy consumption and an order-of-magnitude faster inference time. BOLT has also been successfully deployed by multiple businesses to address critical problems, and we highlight one customer case study in the field of e-commerce.

### **Unified Generative & Dense Retrieval for Query Rewriting in Sponsored Search**

Akash Kumar Mohankumar, Bhargav Dodla, Gururaj K, Amit Singh

Sponsored search is a key revenue source for search engines, where advertisers bid on keywords to target users or search queries of interest. However, finding relevant keywords for a given query is challenging due to the large and dynamic keyword space, ambiguous user/advertiser intents, and diverse possible topics and languages. In this work, we present a comprehensive comparison between two paradigms for online query rewriting: Generative (NLG) and Dense Retrieval (DR) methods. We observe that both methods offer complementary benefits that are additive. As a result, we show that around 40% of the high-quality keywords retrieved by the two approaches are unique and not retrieved by the other. To leverage the strengths of both methods, we propose CLOVER-Unity, a novel approach that unifies generative and dense retrieval methods in one single model. Through offline experiments, we show that the NLG and DR components of CLOVER-Unity consistently outperform individually trained NLG and DR models on public and internal benchmarks. Furthermore, we show that CLOVER-Unity achieves 9.8% higher good keyword density than the ensemble of two separate DR and NLG models while reducing computational costs by almost half. We conduct extensive online A/B experiments on Microsoft Bing in 140+ countries and achieve improved user engagement, with an average increase in total clicks by 0.89% and increased revenue by 1.27%. We also share our practical lessons and optimization tricks for deploying such unified models in production.

### **Retention is All You Need**

Karishma Mohiuddin, Mirza Ariful Alam, Mirza Mohtashim Alam, Pascal Welke, Michael Martin, Jens Lehmann, Sahar Vahdati

Skilled employees are the most important pillars of an organization. Despite this, most organizations face high attrition and turnover rates. While several machine learning models have been developed to analyze attrition and its causal factors, the interpretations of those models remain opaque. In this paper, we propose the HR-DSS approach, which stands for Human Resource (HR) Decision Support System, and uses explainable AI for employee attrition problems. The system is designed to assist HR departments in interpreting the predictions provided by machine learning models. In our experiments, we employ eight machine learning models to provide predictions. We further process the results achieved by the best-performing model by the SHAP explainability process and use the SHAP values to generate natural language explanations which can be valuable for HR. Furthermore, using "What-if-analysis", we aim to observe plausible causes for attrition of an individual employee. The results show that by adjusting the specific dominant features of each individual, employee attrition can turn into employee retention through informative business decisions.

### **Exploiting Sequential Music Preferences via Optimisation-Based Sequencing**

Dmitrii Moor, Yi Yuan, Rishabh Mehrotra, Zhenwen Dai, Mounia Lalmas

Users in music streaming platforms typically consume tracks sequentially in sessions by interacting with personalised playlists. To satisfy users, music platforms usually rely on recommender systems that learn users' preferences over individual tracks and rank the tracks within each playlist according to the learned preferences. However, such rankings often do not

fully exploit the sequential nature of the users' consumption, which may result in a lower within-a-session consumption. In this paper, we model the sequential within-a-session preferences of users and propose an optimisation-based sequencing approach that allows for optimally incorporating such preferences into the rankings. To this end, we rely on interaction data of a major music streaming service to identify two most common aspects of the users' sequential preferences: (1) Position-Aware preferences, and (2) Local-Sequential preferences. We propose a sequencing model that can leverage each of these aspects optimally to maximise the expected total consumption from the session. We further perform an extensive offline and off-policy evaluation of our model, and carry out a large scale online randomised control trial with 7M users across 80 countries. Our findings confirm that we can effectively incorporate sequential preferences of users into our sequencer to make users complete more and skip less tracks within their listening sessions.

#### **Optimal Real-Time Bidding Strategy for Position Auctions in Online Advertising**

Weitong Ou, Bo Chen, Weiwen Liu, Xinyi Dai, Weinan Zhang, Wei Xia, Xuan Li, Ruiming Tang, Yong Yu

Position auctions are widely studied in the context of sponsored search advertising, where multiple ad slots are sold in a single auction. In traditional sponsored search, bids are submitted at the keyword level, while recent works have explored transitioning to impression-level bidding using Real-Time Bidding (RTB) techniques to achieve finer bidding. However, position auctions introduce varying user appeal across different positions and more dynamic auction landscape, which RTB, originally devised for single-slot display advertising, fails to address adequately.

In this work, we are the first to study the optimal bidding strategy for position auctions in RTB. The position auctions we study belong to a broader class, including both sponsored search and display advertising with multiple ad slots. In particular, we aim at maximizing the advertising value within the budget constraint for an advertiser. We mathematically formulate the problem with explicit modeling of position effects. By leveraging the Lagrange multiplier, we derive the optimal bid price and prove its uniqueness under mild assumptions. Efficient numeric methods are applied to obtain the solution practically. Extensive experiments are conducted on a public semi-synthetic dataset and a private industry dataset to validate the effectiveness and feasibility in practice.

#### **Masked Multi-Domain Network: Multi-Type and Multi-Scenario Conversion Rate Prediction with a Single Model**

Wentao Ouyang, Xiuwu Zhang, Chaofeng Guo, Shukui Ren, Yupei Sui, Kun Zhang, Jinmei Luo, Yunfeng Chen, Dongbo Xu, Xiangzheng Liu, Yanlong Du

In real-world advertising systems, conversions have different types in nature and ads can be shown in different display scenarios, both of which highly impact the actual conversion rate (CVR). This results in the multi-type and multi-scenario CVR prediction problem. A desired model for this problem should satisfy the following requirements: 1) Accuracy: the model should achieve fine-grained accuracy with respect to any conversion type in any display scenario. 2) Scalability: the model parameter size should be affordable. 3) Convenience: the model should not require a large amount of effort in data partitioning, subset processing and separate storage.

Existing approaches cannot simultaneously satisfy these requirements. For example, building a separate model for each (conversion type, display scenario) pair is neither scalable nor convenient. Building a unified model trained on all the data with conversion type and display scenario included as two features is not accurate enough.

In this paper, we propose the Masked Multi-domain Network (MMN) to solve this problem. To achieve the accuracy requirement, we model domain-specific parameters and propose a dynamically weighted loss to account for the loss scale imbalance issue within each mini-batch. To achieve the scalability requirement, we propose a parameter sharing and composition strategy to reduce model parameters from a product space to a sum space. To achieve the convenience requirement, we propose an auto-masking strategy which can take mixed data from all the domains as input. It avoids the overhead caused by data partitioning, individual processing and separate storage. Both offline and online experimental results validate the superiority of MMN for multi-type and multi-scenario CVR prediction. MMN is now the serving model for real-time CVR prediction in UC Toutiao.

#### **Graph Learning for Exploratory Query Suggestions in an Instant Search System**

Enrico Palumbo, Andreas Damianou, Alice Wang, Alva Liu, Ghazal Fazelnia, Francesco Fabbri, Rui Ferreira, Fabrizio Silvestri, Hugues Bouchard, Claudia Hauff, Mounia Lalmas, Ben Carterette, Praveen Chandar, David Nyhan

Search systems in online content platforms are typically biased toward a minority of highly consumed items, reflecting the most common user behavior of navigating toward content that is already familiar and popular. Query suggestions are a

powerful tool to support query formulation and to encourage exploratory search and content discovery. However, classic approaches for query suggestions typically rely either on semantic similarity, which lacks diversity and does not reflect user searching behavior, or on a collaborative similarity measure mined from search logs, which suffers from data sparsity and is biased by highly popular queries. In this work, we argue that the task of query suggestion can be modelled as a link prediction task on a heterogeneous graph including queries and documents, enabling Graph Learning methods to effectively generate query suggestions encompassing both semantic and collaborative information. We perform an offline evaluation on an internal Spotify dataset of search logs and on two public datasets, showing that node2vec leads to an accurate and diversified set of results, especially on the large scale real-world data. We then describe the implementation in an instant search scenario and discuss a set of additional challenges tied to the specific production environment. Finally, we report the results of a large scale A/B test involving millions of users and prove that node2vec query suggestions lead to an increase in online metrics such as coverage (+1.42% shown search results pages with suggestions) and engagement (+1.21% clicks), with a specifically notable boost in the number of clicks on exploratory search queries (+9.37%).

#### **Learning and Optimization of Implicit Negative Feedback for Industrial Short-video Recommender System**

Yunzhu Pan, Nian Li, Chen Gao, Jianxin Chang, Yanan Niu, Yang Song, Depeng Jin, Yong Li

Short-video recommendation is one of the most important recommendation applications in today's industrial information systems. Compared with other recommendation tasks, the enormous amount of feedback is the most typical characteristic. Specifically, in short-video recommendation, the easiest-to-collect user feedback is the skipping behavior, which leads to two critical challenges for the recommendation model. First, the skipping behavior reflects implicit user preferences, and thus, it is challenging for interest extraction. Second, this kind of special feedback involves multiple objectives, such as total watching time and skipping rate, which is also very challenging. In this paper, we present our industrial solution in Kuaishou1, which serves billion-level users every day. Specifically, we deploy a feedback-aware encoding module that extracts user preferences, taking the impact of context into consideration. We further design a multi-objective prediction module which well distinguishes the relation and differences among different model objectives in the short-video recommendation. We conduct extensive online A/B tests, along with detailed and careful analysis, which verify the effectiveness of our solution.

#### **GBTTE: Graph Attention Network Based Bus Travel Time Estimation**

Yuecheng Rong, Juntao Yao, Jun Liu, Yifan Fang, Wei Luo, Hao Liu, Jie Ma, Zepeng Dan, Jinzhu Lin, Zhi Wu, Yan Zhang, Chuanming Zhang

Real-time bus travel time is crucial for the smart public transportation system and is beneficial for improving user satisfaction for online map services. However, it faces great challenges due to fine-grained spatial dependencies and dynamic temporal dependencies. To address the above problem, we propose GBTTE, a novel end-to-end graph attention network framework to estimate bus travel time. Specifically, we construct a novel graph structure of bus routes and use a graph attention network to capture the fine-grained spatial features of bus routes. Then, we fully exploit the joint spatial-temporal relations of bus stops through a spatial-temporal graph attention network and also capture the dynamic correlation between the route and the bus transportation network with a cross graph attention network. Finally, we integrate the route representation, the spatial-temporal representation and contextual information to estimate bus travel time. Extensive experiments carried out on two large-scale real-world datasets demonstrate the effectiveness of GBTTE. In addition, GBTTE has been deployed in production at Baidu Maps, handling tens of millions of requests every day.

#### **Deep Query Rewriting For Geocoding**

Pravakar Roy, Chirag Sharma, Chao Gao, Kumarswamy Valegarepura

Query rewriting aims to bridge the gap between user queries and indexed data. In the context of geocoding, it augments the user query to match all relevant entities in indexed data. This work presents an end-to-end trainable attention-based query rewriting model for geocoding. Most recent works attempt to solve query rewriting with sequence-to-sequence models, which often fail to satisfy the tight latency constraints of production environments. Toward developing a low latency model, we formulate it as a combination of multiple token classification tasks. We introduce numerous novel techniques, such as constructing a separate output space for alterations containing whole words to tackle many-to-one/one-to-many associations, using regular expression rules as prediction classes to handle the limitations of having a non-word piece prediction space, and curbing the output space for each wordpiece in our vocabulary to reduce inference time by 20x in average. We also present a data generation pipeline that generates train/test data without human judgment from the query logs. Through offline and online experiments, we demonstrate that the proposed model is orders of magnitude faster than their seq2seq and SMT-based counterparts in inference time, achieves better/on-par performance, and improves the relevance of overall geocoding.

**TrendSpotter: Forecasting E-commerce Product Trends**

Gayatri Ryali, Shreyas S, Sivaramakrishnan Kaveri, Prakash Mandayam Comar

Internet users actively search for trending products on various social media services like Instagram and YouTube which serve as popular hubs for discovering and exploring fashionable and popular items. It is imperative for e-commerce giants to have the capability to accurately identify, predict and subsequently showcase these trending products to the customers. E-commerce stores can effectively cater to the evolving demands of the customer base and enhance the overall shopping experience by offering recent and most sought-after products in a timely manner. In this work we propose a framework for predicting and surfacing trending products in e-commerce stores, the first of its kind to the best of our knowledge. We begin by defining what constitutes a trending product using sound statistical tests. We then introduce a machine learning-based early trend prediction system called TrendSpotter to help users identify upcoming product trends. TrendSpotter is a unique adaptation of the state-of-the-art InceptionTime model\cite{InceptionTime} that predicts the future popularity of a product based on its current customer engagement, such as clicks, purchases, and other relevant product attributes. The effectiveness of our approach is demonstrated through A/B tests, where we first showcase the effectiveness of our statistical test based labeling strategy, resulting in an incremental sales lift of 59 bps\footnote{bps or basis points are a measure of percentages. 1 bps = 0.01%} across two experiments on home page and search page. Subsequently, we conduct a comparison between our machine learning model and the statistical labeling baseline and observe an additional sales gain of 14 bps, reflecting the importance of early identification of trending products.

**STREAMS: Towards Spatio-Temporal Causal Discovery with Reinforcement Learning for Streamflow Rate Prediction**

Paras Sheth, Ahmadreza Mosallanezhad, Kaize Ding, Reepal Shah, John Sabo, Huan Liu, K. Selçuk Candan

The capacity to anticipate streamflow is critical to the efficient functioning of reservoir systems as it gives vital information to reservoir operators about water release quantities as well as help quantify the impact of environmental factors on downstream water quality. Yet, streamflow modelling is difficult owing to the intricate interactions between different watershed outlets. In this paper, we argue that one possible solution to this problem is to identify the causal structure of these outlets, which would allow for the identification of crucial watershed outlets while capturing the spatiotemporally informed complex relationships leading to improved hydrological resource management. However, due to the inherent complexity of spatiotemporal causal learning problems, extending existing causal discovery methods to a whole basin is a major hurdle. To address these issues, we offer STREAMS, a new framework that uses Reinforcement Learning (RL) to optimize the search space for causal discovery and an LSTM-GCN based autoencoder to infer spatiotemporal causal features for streamflow rate prediction. We conduct extensive experiments on the Brazos river basin carried out within the scope of a US Army Corps of Engineers, Engineering With Nature Initiative project, including empirical studies of generalization performance to verify the nature of the inferred relationships.

**Combating Ad Fatigue via Frequency-Recency Features in Online Advertising Systems**

Natalia Silberstein, Or Shoham, Assaf Klein

Online advertising is a driving force of Internet services today. One of the main challenges in advertising systems is finding the right balance between user experience and overall revenue.

In this paper we address one of the problems that negatively impacts the user experience, specifically, the repeated display of identical ads to the same user. This problem leads to the phenomenon called "ad fatigue", characterized by diminished interest in the ad, resulting in a decrease in the click-through rate (CTR) as users encounter the same ad repeatedly. Naive solutions such as placing a hard limit on the number of times a specific ad is displayed to a specific user, usually come at the cost of reduced revenue.

To address the ad fatigue problem, we introduce a new family of features, called FoRI (Frequency over Recent Intervals). FoRI features integrate information about frequency and recency of previous user-ad interactions within the CTR prediction model. This approach involves allocating these interactions to unevenly distributed time intervals, enabling a higher emphasis on more recent interactions. Furthermore, we introduce new metrics to assess ad fatigue in terms of the repetitiveness and novelty of the displayed ads.

We conducted a comprehensive large-scale online evaluation which shows that integrating FoRI features into our CTR prediction model offers two-fold benefits. Firstly, it improves user experience by reducing the occurrence of repeated ads by



15%, and increasing the exposure to unseen ads by 5% (ads not previously displayed to the user), leading to a substantial boost in CTR. Secondly, it significantly increases revenue.

### **GraphFC: Customs Fraud Detection with Label Scarcity**

Karandeep Singh, Yu-Che Tsai, Cheng-Te Li, Meeyoung Cha, Shou-De Lin

Customs officials across the world encounter huge volumes of transactions. Associated with customs transactions is customs fraud-the intentional manipulation of goods declarations to avoid taxes and duties. Due to limited manpower, the customs offices can only manually inspect a small number of declarations, necessitating the automation of customs fraud detection by machine learning techniques. The limited availability of manually inspected ground truth data makes it essential for the ML approach to generalize well on unseen data. However, current customs fraud detection models are not well suited or designed for this setting. In this work, we propose GraphFC (Graph Neural networks for Customs Fraud), a model-agnostic, domain-specific, graph neural network based customs fraud detection model that is designed to work in a real-world setting with limited ground truth data. Extensive experimentation using real customs data from two countries demonstrates that GraphFC generalizes well over unseen data and outperforms various baselines and other models by a large margin.

### **PRODIGY: Product Design Guidance at Scale**

Sambeet Tiady, Anirban Majumder, Deepak Gupta

Growth of e-commerce has enabled the creation of thousands of small-scale brands. However, these brands lack information on a) what new products to develop and b) how to refine existing products to improve on business metrics. We present a comprehensive Product Design Insights and Guidance service (named PRODIGY) that mines product attributes data available on e-commerce platforms and surface insights on a) new product development and b) product refinement. Our core contribution is a novel demand forecasting model for product designs based on a notable extension of the recently proposed FTTransformer architecture combined with a self-supervised pre-training task, akin to Masked Language Modeling (MLM) objective. For the product refinement use-case, we present a novel algorithm by embedding the design search in a data-density approximator, namely Conditional Variational Autoencoder. We run a thorough and comprehensive set of experiments and establish that PRODIGY achieves significant improvement in demand prediction as compared to state-of-the-art alternatives. Finally, we present our findings from an online experiment where PRODIGY helps to launch new products with +20% lift in sales and +1.3% lift in product ratings.

### **LittleMu: Deploying an Online Virtual Teaching Assistant via Heterogeneous Sources Integration and Chain of Teach Prompts**

Shangqing Tu, Zheyuan Zhang, Jifan Yu, Chunyang Li, Siyu Zhang, Zijun Yao, Lei Hou, Juanzi Li

Teaching assistants have played essential roles in the long history of education. However, few MOOC platforms are providing human or virtual teaching assistants to support learning for massive online students due to the complexity of real-world online education scenarios and the lack of training data. In this paper, we present a virtual MOOC teaching assistant, LittleMu with minimum labeled training data, to provide question answering and chat services. Consisting of two interactive modules of heterogeneous retrieval and language model prompting, LittleMu first integrates structural, semi- and unstructured knowledge sources to support accurate answers for a wide range of questions. Then, we design delicate demonstrations named "Chain of Teach" prompts to exploit the large-scale pre-trained model to handle complex uncollected questions. Except for question answering, we develop other educational services such as knowledge-grounded chat. We test the system's performance via both offline evaluation and online deployment. Since May 2020, our LittleMu system has served over 80,000 users with over 300,000 queries from over 500 courses on XuetangX MOOC platform, which continuously contributes to a more convenient and fair education. Our code, services, and dataset will be available at <https://github.com/THU-KEG/VTa>.

### **Generating Optimized Molecules without Patent Infringement**

Sally Turutov, Kira Radinsky

Molecular optimization seeks to improve a given molecule's therapeutic profile. It is a key challenge in drug development, but it is difficult due to the constraints of molecular similarity to the original molecule and the size of the chemical space to explore. Numerous works tackled this problem with initial success. Unlike previous works that focus on generating molecules that optimize chemical properties, we focus on the optimization while attempting to "move away" from patented molecules. We present a novel loss function and its utilization in numerous types of molecular optimization algorithms. The loss allows to improve molecular properties while decreasing patent-infringement. We perform empirical evaluation showing superior

performance of state-of-the-art models when using the novel loss function. The deployment of the system is underway at the Targeted Drug Delivery and Personalized Medicine labs. It will be utilized to generate targeted carriers of mRNA, providing a new method of drug delivery. The system is also producing non-patented candidates for industrial use, making it a valuable tool in the field of personalized medicine.

#### **MARBLE: Hierarchical Multi-Armed Bandits for Human-in-the-Loop Set Expansion**

Muntasir Wahed, Daniel Gruhl, Ismini Lourentzou

The modern-day research community has an embarrassment of riches regarding pre-trained AI models. Even for a simple task such as lexicon set expansion, where an AI model suggests new entities to add to a predefined seed set of entities, thousands of models are available. However, deciding which model to use for a given set expansion task is non-trivial. In hindsight, some models can be 'off topic' for specific set expansion tasks, while others might work well initially but quickly exhaust what they have to offer. Additionally, certain models may require more careful priming in the form of samples or feedback before being finetuned to the task at hand. In this work, we frame this model selection as a sequential non-stationary problem, where there exist a large number of diverse pre-trained models that may or may not fit a task at hand, and an expert is shown one suggestion at a time to include in the set or not, i.e., accept or reject the suggestion. The goal is to expand the list with the most entities as quickly as possible. We introduce MARBLE, a hierarchical multi-armed bandit method for this task, and two strategies designed to address cold-start problems. Experimental results on three set expansion tasks demonstrate MARBLE's effectiveness compared to baselines.

#### **Voucher Abuse Detection with Prompt-based Fine-tuning on Graph Neural Networks**

Zhihao Wen, Yuan Fang, Yihan Liu, Yang Guo, Shuji Hao

Voucher abuse detection is an important anomaly detection problem in E-commerce. While many GNN-based solutions have emerged, the supervised paradigm depends on a large quantity of labeled data. A popular alternative is to adopt self-supervised pre-training using label-free data, and further fine-tune on a downstream task with limited labels. Nevertheless, the "pre-train, fine-tune" paradigm is often plagued by the objective gap between pre-training and downstream tasks. Hence, we propose VPGNN, a prompt-based fine-tuning framework on GNNs for voucher abuse detection. We design a novel graph prompting function to reformulate the downstream task into a similar template as the pretext task in pre-training, thereby narrowing the objective gap. Extensive experiments on both proprietary and public datasets demonstrate the strength of VPGNN in both few-shot and semi-supervised scenarios. Moreover, an online deployment of VPGNN in a production environment shows a 23.4% improvement over two existing deployed models.

#### **RUEL: Retrieval-Augmented User Representation with Edge Browser Logs for Sequential Recommendation**

Ning Wu, Ming Gong, Linjun Shou, Jian Pei, Daxin Jiang

Online recommender systems (RS) aim to match user needs with the vast amount of resources available on various platforms. A key challenge is to model user preferences accurately under the condition of data sparsity. To address this challenge, some methods have leveraged external user behavior data from multiple platforms to enrich user representation. However, all of these methods require a consistent user ID across platforms and ignore the information from similar users. In this study, we propose RUEL, a novel retrieval-based sequential recommender that can effectively incorporate external anonymous user behavior data from Edge browser logs to enhance recommendation. We first collect and preprocess a large volume of Edge browser logs over a one-year period and link them to target entities that correspond to candidate items in recommendation datasets. We then design a contrastive learning framework with a momentum encoder and a memory bank to retrieve the most relevant and diverse browsing sequences from the full browsing log based on the semantic similarity between user representations. After retrieval, we apply an item-level attentive selector to filter out noisy items and generate refined sequence embeddings for the final predictor. RUEL is the first method that connects user browsing data with typical recommendation datasets and can be generalized to various recommendation scenarios and datasets. We conduct extensive experiments on four real datasets for sequential recommendation tasks and demonstrate that RUEL significantly outperforms state-of-the-art baselines. We also conduct ablation studies and qualitative analysis to validate the effectiveness of each component of RUEL and provide additional insights into our method.

#### **Logistics Audience Expansion via Temporal Knowledge Graph**

Hua Yan, Yingqiang Ge, Haotian Wang, Desheng Zhang, Yu Yang

Logistics audience expansion, the process for logistics companies to find potential long-term customers, is one of the most important tasks for business growth. However, existing methods for conventional audience expansion fall short due to two

significant challenges, the intricate interplay of multiple complex factors in the logistics scenario and the emphasis on long-term logistics service usage instead of one-time promotions. To address the above limitations, we design LOGAE-TKG, a logistics audience expansion method based on a temporal knowledge graph, which consists of three components: (i) a temporal logistics knowledge graph pre-trained model to model the effect of multiple complex factors and build a solid logistics knowledge base for contracting and usage prediction; (ii) an intention learning model with data augmentation-based comparison to capture the contracting intention; (iii) a future pattern discovery model to uncover post-contract patterns. We evaluate and deploy our method on the JingDong e-commerce platform. Extensive offline experiment results and real-world deployment results demonstrate the effectiveness of our method.

#### [Practice on Effectively Extracting NLP Features for Click-Through Rate Prediction](#)

Hao Yang, Ziliang Wang, Weijie Bian, Yifan Zeng

Click-through rate (CTR) prediction is critical for industrial applications such as recommendation system and online advertising. Practically, there are a series of research proved that Natural Language Processing (NLP) features are helpful to improve CTR task performance. As these works show, there are different ways to extract NLP features. For example, keywords of item title are extracted as open-box feature by term frequency-inverse document frequency (tf-idf) method while item semantic embedding is extracted as black-box feature by shallow models (e.g., word2vec) or deep learning models (e.g., BERT). However, these NLP models are pre-trained on NLP task, which is very different from the CTR task. Then it leads to the limited improvement of Area Under the ROC Curve (AUC) in CTR task. On the other hand, traditional NLP models for CTR task only consider open-box feature or black-box feature separately, which also leads to the discounted effect. Lastly, many NLP models are mainly used to extract semantic features only on item side. These methods take little account of user side information, or only IDs related features (e.g., item's IDs) in user behavior sequence are introduced.

In our work, we proposed a new network named BERT Attention method based on both Item and User information (BAIU). The target of BAIU is consistent with the CTR task, which is helpful to extract more effective NLP features. Also, both open-box and black-box features are simultaneously extracted by this network, which makes the model to learn more useful NLP features for CTR task and also makes the model more interpretable. Extensive experiments on both public data and our commercial data validate the effectiveness of our approach. Finally, the online experiment brings 2.2% gain of CTR on recommendation.

#### [An Incremental Update Framework for Online Recommenders with Data-Driven Prior](#)

Chen Yang, Jin Chen, Qian Yu, Xiangdong Wu, Kui Ma, Zihao Zhao, Zhiwei Fang, Wenlong Chen, Chaosheng Fan, Jie He, Changping Peng, Zhangang Lin, Jingping Shao

Online recommenders have attained growing interest and created great revenue for businesses. Given numerous users and items, incremental update becomes a mainstream paradigm for learning large-scale models in industrial scenarios, where only newly arrived data within a sliding window is fed into the model, meeting the strict requirements of quick response. However, this strategy would be prone to overfitting to newly arrived data. When there exists a significant drift of data distribution, the long-term information would be discarded, which harms the recommendation performance. Conventional methods address this issue through native model-based continual learning methods, without analyzing the data characteristics for online recommenders. To address the aforementioned issue, we propose an incremental update framework for online recommenders with Data-Driven Prior (DDP), which is composed of Feature Prior (FP) and Model Prior (MP). The FP performs the click estimation for each specific value to enhance the stability of the training process. The MP incorporates previous model output into the current update while strictly following the Bayes rules, resulting in a theoretically provable prior for the robust update. In this way, both the FP and MP are well integrated into the unified framework, which is model-agnostic and can accommodate various advanced interaction models. Extensive experiments on two publicly available datasets as well as an industrial dataset demonstrate the superior performance of the proposed framework.

#### [Graph Exploration Matters: Improving both Individual-Level and System-Level Diversity in WeChat Feed Recommendation](#)

Shuai Yang, Lixin Zhang, Feng Xia, Leyu Lin

There are roughly three stages in real industrial recommendation systems, candidates generation (retrieval), ranking and reranking. Both individual-level diversity and system-level diversity are important in this framework. The former focus on each single user's experience, while the latter focus on the difference among users. Graph-based retrieval strategies are inevitably

hijacked by heavy users and popular items, leading to similar results for each user and the convergence among users. In the reranking phase, Determinantal Point Process (DPP) is widely deployed to increase individual-level diversity. Heavily relying on the semantic information of items, DPP suffers from clickbait and inaccurate categories. Besides, most studies only focus on one of the two levels of diversity, and ignore the mutual influence among different stages in real recommender systems. We argue that individual-level diversity and system-level diversity should be viewed as an integrated problem, and we provide an efficient and deployable solution for web-scale recommenders. Generally, we propose to employ the retrieval graph information in diversity-based reranking, by which to weaken the hidden similarity of items exposed to users, and consequently gain more graph explorations to improve the system-level diversity. Besides, we argue that users' propensity for diversity changes over time in content feed recommendation. Therefore, with the explored graph, we also propose to capture the user's real-time personalized propensity to the diversity. We implement and deploy the combined system in WeChat App's Top Stories used by hundreds of millions of users. Comparative and long-term online A/B tests show our solution can effectively improve both user engagement and system revenue.

#### **An Efficient Content-based Time Series Retrieval System**

Chin-Chia Michael Yeh, Huiyuan Chen, Xin Dai, Yan Zheng, Junpeng Wang, Vivian Lai, Yujie Fan, Audrey Der, Zhongfang Zhuang, Liang Wang, Wei Zhang, Jeff M. Phillips

A Content-based Time Series Retrieval (CTSR) system is an information retrieval system for users to interact with time series emerged from multiple domains, such as finance, healthcare, and manufacturing. For example, users seeking to learn more about the source of a time series can submit the time series as a query to the CTSR system and retrieve a list of relevant time series with associated metadata. By analyzing the retrieved metadata, users can gather more information about the source of the time series. Because the CTSR system is required to work with time series data from diverse domains, it needs a high-capacity model to effectively measure the similarity between different time series. On top of that, the model within the CTSR system has to compute the similarity scores in an efficient manner as the users interact with the system in real-time. In this paper, we propose an effective and efficient CTSR model that outperforms alternative models, while still providing reasonable inference runtimes. To demonstrate the capability of the proposed method in solving business problems, we compare it against alternative models using our in-house transaction data. Our findings reveal that the proposed model is the most suitable solution compared to others for our transaction data problem.

#### **DeepSTA: A Spatial-Temporal Attention Network for Logistics Delivery Timely Rate Prediction in Anomaly Conditions**

Jinhui Yi, Huan Yan, Haotian Wang, Jian Yuan, Yong Li

Prediction of couriers' delivery timely rates in advance is essential to the logistics industry, enabling companies to take preemptive measures to ensure the normal operation of delivery services. This becomes even more critical during anomaly conditions like the epidemic outbreak, during which couriers' delivery timely rate will decline markedly and fluctuates significantly. Existing studies pay less attention to the logistics scenario. Moreover, many works focusing on prediction tasks in anomaly scenarios fail to explicitly model abnormal events, e.g., treating external factors equally with other features, resulting in great information loss. Further, since some anomalous events occur infrequently, traditional data-driven methods perform poorly in these scenarios. To deal with them, we propose a deep spatial-temporal attention model, named DeepSTA. To be specific, to avoid information loss, we design an anomaly spatio-temporal learning module that employs a recurrent neural network to model incident information. Additionally, we utilize Node2vec to model correlations between road districts, and adopt graph neural networks and long short-term memory to capture the spatial-temporal dependencies of couriers. To tackle the issue of insufficient training data in abnormal circumstances, we propose an anomaly pattern attention module that adopts a memory network for couriers' anomaly feature patterns storage via attention mechanisms. The experiments on real-world logistics datasets during the COVID-19 outbreak in 2022 show the model outperforms the best baselines by 12.11% in MAE and 13.71% in MSE, demonstrating its superior performance over multiple competitive baselines.

#### **SPM: Structured Pretraining and Matching Architectures for Relevance Modeling in Meituan Search**

Wen Zan, Yaopeng Han, Xiaotian Jiang, Yao Xiao, Yang Yang, Dayao Chen, Sheng Chen

In e-commerce search, relevance between query and documents is an essential requirement for satisfying user experience. Different from traditional e-commerce platforms that offer products, users search on life service platforms such as Meituan mainly for product providers, which usually have abundant structured information, e.g. name, address, category, thousands of products. Modeling search relevance with these rich structured contents is challenging due to the following issues: (1) there is language distribution discrepancy among different fields of structured document, making it difficult to directly adopt off-the-

shelf pretrained language model based methods like BERT. (2) different fields usually have different importance and their length vary greatly, making it difficult to extract document information helpful for relevance matching.

To tackle these issues, in this paper we propose a novel two-stage pretraining and matching architecture for relevance matching with rich structured documents. At pretraining stage, we propose an effective pretraining method that employs both query and multiple fields of document as inputs, including an effective information compression method for lengthy fields. At relevance matching stage, a novel matching method is proposed by leveraging domain knowledge in search query to generate more effective document representations for relevance scoring. Extensive offline experiments and online A/B tests on millions of users verify that the proposed architectures effectively improve the performance of relevance modeling. The model has already been deployed online, serving the search traffic of Meituan for over a year.

### **SHARK: A Lightweight Model Compression Approach for Large-scale Recommender Systems**

Beichuan Zhang, Chenggen Sun, Jianchao Tan, Xinjun Cai, Jun Zhao, Mengqi Miao, Kang Yin, Chengru Song, Na Mou, Yang Song

Increasing the size of embedding layers has shown to be effective in improving the performance of recommendation models, yet gradually causing their sizes to exceed terabytes in industrial recommender systems, and hence the increase of computing and storage costs. To save resources while maintaining model performances, we propose SHARK, the model compression practice we have summarized in the recommender system of industrial scenarios. SHARK consists of two main components. First, we use the novel first-order component of Taylor expansion as importance scores to prune the number of embedding tables (feature fields). Second, we introduce a new row-wise quantization method to apply different quantization strategies to each embedding. We conduct extensive experiments on both public and industrial datasets, demonstrating that each component of our proposed SHARK framework outperforms previous approaches. We conduct A/B tests in multiple models on Kuaishou, such as short video, e-commerce, and advertising recommendation models. The results of the online A/B test showed SHARK can effectively reduce the memory footprint of the embedded layer. For the short-video scenarios, the compressed model without any performance drop significantly saves 70% storage and thousands of machines, improves 30% queries per second (QPS), and has been deployed to serve hundreds of millions of users and process tens of billions of requests every day.

### **Multi-gate Mixture-of-Contrastive-Experts with Graph-based Gating Mechanism for TV Recommendation**

Cong Zhang, Dongyang Liu, Lin Zuo, Junlan Feng, Chao Deng, Jian Sun, Haitao Zeng, Yaohong Zhao

With the rapid development of smart TV, TV recommendation is attracting more and more users. TV users usually distribute in multiple regions with different cultures and hence have diverse TV program preferences. From the perspective of engineering practice and performance improvement, it's very essential to model users from multiple regions with one single model. In previous work, Multi-gate Mixture-of-Expert (MMoE) has been widely adopted in multi-task and multi-domain recommendation scenarios. In practice, however, we first observe the embeddings generated by experts tend to be homogeneous which may result in high semantic similarities among embeddings that reduce the capability of Multi-gate Mixture-of-Expert (MMoE) model. Secondly, we also find there are lots of commonalities and differences between multiple regions regarding user preferences. Therefore, it's meaningful to model the complicated relationships between regions. In this paper, we first introduce contrastive learning to overcome the expert representation degeneration problem. The embeddings of two augmented samples generated by the same experts are pushed closer to enhance the alignment, and the embeddings of the same samples generated by different experts are pushed away in vector space to improve uniformity. Then we propose a Graph-based Gating Mechanism to empower typical Multi-gate Mixture-of-Experts. Graph-based MMoE is able to recognize the commonalities and differences among multiple regions by introducing a Graph Neural Network (GNN) with region similarity prior. We name our model Multi-gate Mixture-of-Contrastive-Experts model with Graph-based Gating Mechanism (MMoCEG). Extensive offline experiments and online A/B tests on a commercial TV service provider over 100 million users and 2.3 million items demonstrate the efficacy of MMoCEG compared to the existing models.

### **3MN: Three Meta Networks for Multi-Scenario and Multi-Task Learning in Online Advertising Recommender Systems**

Yifei Zhang, Hua Hua, Hui Guo, Shuangyang Wang, Chongyu Zhong, Shijie Zhang

Recommender systems are widely applied on web. For example, online advertising systems rely on recommender systems to accurately estimate the value of display opportunities, which is critical to maximize the profits of advertisers. To reduce computational resource consumption, the core tactic of Multi-Scenario Multi-Task Learning (MSMTL) is to devise a single recommender system that is adapted to all contexts instead of implementing multiple scenario-oriented or task-oriented



recommender systems. However, MSMTL is challenging because there are complicated task-task, scenario-scenario, and task-scenario interrelations; the characteristic of different tasks in different scenarios also largely varies; and samples of each context are often unevenly distributed. Previous MSMTL solutions focus on applying scenario knowledge to improve the performance of multi-task learning, while neglecting the complicated interrelations among tasks and scenarios. Moreover, samples derived from different scenarios are transferred into the latent embedding with the same dimension. This static embedding strategy impedes the practicality of model expressiveness, since the scenarios with sufficient samples are underrepresented and those with insufficient samples are over-represented.

In this paper, we propose a novel three meta networks-based solution (3MN) to MSMTL that addresses all the limitations discussed above. Specifically, we innovatively bind the meta network with scenario-related input in bottom embedding layer, so that the embedding layer is capable of learning the scenario-related knowledge explicitly. To counteract the imbalanced scenario-related data distributions, our flexible embedding layer adaptively learns the representation of samples. This innovative embedding layer is also able to boost other solutions as a plug-in. Moreover, to fully capture the interrelations among scenarios and tasks, we enforce the task and scenario information into the other two meta networks, and transfer the resulted meta-knowledge into the top components (i.e., backbone network and classifier) of the recommender system, respectively. These three meta networks contribute to the superiority of our 3MN solution over state-of-the-art MSMTL solutions, which is demonstrated by extensive offline experiments. 3MN has been successfully deployed in our industrial online advertising system.

#### [Leveraging Watch-time Feedback for Short-Video Recommendations: A Causal Labeling Framework](#)

Yang Zhang, Yimeng Bai, Jianxin Chang, Xiaoxue Zang, Song Lu, Jing Lu, Fuli Feng, Yanan Niu, Yang Song

With the proliferation of short video applications, the significance of short video recommendations has vastly increased. Unlike other recommendation scenarios, short video recommendation systems heavily rely on feedback from watch time. Existing approaches simply treat watch time as a direct label, failing to effectively harness its extensive semantics and introduce bias, thereby limiting the potential for modeling user interests based on watch time. To overcome this challenge, we propose a framework named Debiased Multiple-semantics-extracting Labeling (DML). DML constructs labels that encompass various semantics by utilizing quantiles derived from the distribution of watch time, prioritizing relative order rather than absolute label values. This approach facilitates easier model learning while aligning with the ranking objective of recommendations. Furthermore, we introduce a method inspired by causal adjustment to refine label definitions, thereby directly mitigating bias at the label level. We substantiate the effectiveness of our DML framework through both online and offline experiments. Extensive results demonstrate that our DML could effectively leverage watch time to discover users' real interests, enhancing their engagement in our application.

#### [Build Faster with Less: A Journey to Accelerate Sparse Model Building for Semantic Matching in Product Search](#)

Jiong Zhang, Yau-Shian Wang, Wei-Cheng Chang, Wei Li, Jyun-Yu Jiang, Cho-Jui Hsieh, Hsiang-Fu Yu

The semantic matching problem in product search seeks to retrieve all semantically relevant products given a user query. Recent studies have shown that extreme multi-label classification~(XMC) model enjoys both low inference latency and high recall in real-world scenarios. These XMC semantic matching models adopt TF-IDF vectorizers to extract query text features and use mainly sparse matrices for the model weights. However, limited availability of libraries for efficient parallel sparse modules may lead to tediously long model building time when the problem scales to hundreds of millions of labels. This incurs significant hardware cost and renders the semantic model stale even before it is deployed. In this paper, we investigate and accelerate the model building procedures in a tree-based XMC model. On a real-world semantic matching task with 100M labels, our enhancements achieve over 10 times acceleration (from 3.1 days to 6.7 hours) while reducing hardware cost by 25%.

#### [Popularity-aware Distributionally Robust Optimization for Recommendation System](#)

Jujia Zhao, Wenjie Wang, Xinyu Lin, Leigang Qu, Jizhi Zhang, Tat-Seng Chua

Collaborative Filtering (CF) has been widely applied for personalized recommendations in various industrial applications. However, due to the training strategy of Empirical Risk Minimization, CF models tend to favor popular items, resulting in inferior performance on sparse users and items. To enhance the CF representation learning of sparse users and items without sacrificing the performance of popular items, we propose a novel Popularity-aware Distributionally Robust Optimization (PDRO) framework. In particular, PDRO emphasizes the optimization of sparse users/items, while incorporating item popularity to preserve the performance of popular items through two modules. First, an implicit module develops a new

popularity-aware DRO objective, paying more attention to items that will potentially become popular over time. Second, an explicit module that directly predicts the popularity of items to help the estimation of user-item matching scores. We apply PDRO to a micro-video recommendation scenario and implement it on two representative backend models. Extensive experiments on a real-world industrial dataset, as well as two public benchmark datasets, validate the efficacy of our proposed PDRO. Additionally, we perform an offline A/B test on the industrial dataset, further demonstrating the superiority of PDRO in real-world application scenarios.

#### **COPR: Consistency-Oriented Pre-Ranking for Online Advertising**

Zhishan Zhao, Jingyue Gao, Yu Zhang, Shuguang Han, Siyuan Lou, Xiang-Rong Sheng, Zhe Wang, Han Zhu, Yuning Jiang, Jian Xu, Bo Zheng

Cascading architecture has been widely adopted in large-scale advertising systems to balance efficiency and effectiveness. In this architecture, the pre-ranking model is expected to be a lightweight approximation of the ranking model, which handles more candidates with strict latency requirements. Due to the gap in model capacity, the pre-ranking and ranking models usually generate inconsistent ranked results, thus hurting the overall system effectiveness. The paradigm of score alignment is proposed to regularize their raw scores to be consistent. However, it suffers from inevitable alignment errors and error amplification by bids when applied in online advertising. To this end, we introduce a consistency-oriented pre-ranking framework for online advertising, which employs a chunk-based sampling module and a plug-and-play rank alignment module to explicitly optimize consistency of ECPM-ranked results. A  $\Delta$ ND CG-based weighting mechanism is adopted to better distinguish the importance of inter-chunk samples in optimization. Both online and offline experiments have validated the superiority of our framework. When deployed in Taobao display advertising system, it achieves an improvement of up to +12.3% CTR and +5.6% RPM.

#### **Entire Space Cascade Delayed Feedback Modeling for Effective Conversion Rate Prediction**

Yunfeng Zhao, Xu Yan, Xiaoqiang Gui, Shuguang Han, Xiang-Rong Sheng, Guoxian Yu, Jufeng Chen, Zhao Xu, Bo Zheng

Conversion rate (CVR) prediction is an essential task for e-commerce platforms. However, refunds frequently occur after conversion in online shopping systems, which drives us to pay attention to effective conversion for building healthier services. This paper defines the probability of item purchasing without any subsequent refund as an effective conversion rate (ECVR). A simple paradigm for ECVR prediction is to decompose it into two sub-tasks: CVR prediction and post-conversion refund rate (RFR) prediction. However, RFR prediction suffers from data sparsity (DS) and sample selection bias (SSB) issues, as refund behaviors are only available after user purchase. Furthermore, there is delayed feedback in both sequentially dependent conversion and refund events, named cascade delayed feedback (CDF). Previous studies mainly focus on tackling DS and SSB or delayed feedback for a single event. To jointly tackle these issues in ECVR prediction, we propose an Entire space CAscade Delayed feedback modeling (ECAD) method. Specifically, ECAD deals with DS and SSB by constructing two tasks including CVR and conversion&refund rate (CVRFR) predictions using the entire space modeling framework. In addition, it carefully schedules auxiliary tasks to leverage both conversion and refund time within data to alleviate CDF. Experiments on the offline industrial dataset and online A/B testing demonstrate the effectiveness of ECAD. ECAD has been deployed in the Xianyu recommender system of Alibaba, contributing to a significant improvement of ECVR.

#### **Dual Interests-Aligned Graph Auto-Encoders for Cross-domain Recommendation in WeChat**

Jiawei Zheng, Hao Gu, Chonggang Song, Dandan Lin, Lingling Yi, Chuan Chen

Recently, cross-domain recommendation (CDR) has been widely studied in both research and industry since it can alleviate a long-standing challenge of traditional recommendation methods, i.e., data sparsity issue, by transferring the information from a relatively richer domain (termed source domain) to a sparser domain (termed target domain). To our best knowledge, most (if not all) existing CDR methods focus on transferring either the similar content information or the user preferences embedding from the source domain to the target domain. However, they fail to improve the recommendation performance in real-world recommendation scenarios where the items in the source domain are totally different from those in the target domain in terms of attributes. To solve the above issues, we analyzed the historical interactions of users from different domains in the WeChat platform, and found that if two users have similar interests (interactions) in one domain, they are very likely to have similar interests in another domain even though the items of these two domains are totally different in terms of attributes. Based on this observation, in this paper, we propose a novel model named Dual Interests-Aligned Graph Auto-Encoders (DIAGAE) by utilizing the inter-domain interest alignment of users. Besides, our proposed model DIAGAE also leverages graph decoding objectives to align intra-domain user interests, which makes the representation of two users who have similar interests in a single domain closer. Comprehensive experimental results demonstrate that our model DIAGAE

outperforms state-of-the-art methods on both public benchmark datasets and online A/B tests in WeChat live-stream recommendation scenario. Our model DIAGAE now serves the major online traffic in WeChat live-streaming recommendation scenario.

#### [Detecting Social Bot on the Fly using Contrastive Learning](#)

Ming Zhou, Dan Zhang, Yuandong Wang, Yangli-Ao Geng, Jie Tang

Social bot detection is becoming a task of wide concern in social security. All along, the development of social bot detection technology is hindered by the lack of high-quality annotated data. Besides, the rapid development of AI Generated Content (AIGC) technology is dramatically improving the creative ability of social bots. For example, the recently released ChatGPT [2] can fool the state-of-the-art AI-text-detection method with a probability of 74%, bringing a large challenge to content-based bot detection methods. To address the above drawbacks, we propose a Contrastive Learning-driven Social Bot Detection framework (CBD). The core of CBD is characterized by a two-stage model learning strategy: a contrastive pre-training stage to mine generalization patterns from massive unlabeled social graphs, followed by a semi-supervised fine-tuning stage to model task-specific knowledge latent in social graphs with a few annotations. The above strategy endows our model with promising detection performance under an extreme scarcity of labeled data. In terms of system architecture, we propose a smart feedback mechanism to further improve detection performance. Comprehensive experiments on a real bot detection dataset show that CBD consistently outperforms 10 state-of-the-art baselines by a large margin for few-shot bot detection using very little (5-shot) labeled data. CBD has been deployed online.

#### [DeepTagger: Knowledge Enhanced Named Entity Recognition for Web-Based Ads Queries](#)

Simiao Zuo, Pengfei Tang, Xinyu Hu, Qiang Lou, Jian Jiao, Denis Charles

Named entity recognition (NER) is a crucial task for online advertisement. State-of-the-art solutions leverage pre-trained language models for this task. However, three major challenges remain unresolved: web queries differ from natural language, on which pre-trained models are trained; web queries are short and lack contextual information; and labeled data for NER is scarce. We propose DeepTagger, a knowledge-enhanced NER model for web-based ads queries. The proposed knowledge enhancement framework leverages both model-free and model-based approaches. For model-free enhancement, we collect unlabeled web queries to augment domain knowledge; and we collect web search results to enrich the information of ads queries. We further leverage effective prompting methods to automatically generate labels using large language models such as ChatGPT. Additionally, we adopt a model-based knowledge enhancement method based on adversarial data augmentation. We employ a three-stage training framework to train DeepTagger models.

### **SESSION: Demo Papers**

#### [SNAKE Challenge: Sanitization Algorithms under Attack](#)

Tristan Allard, Louis Béziau, Sébastien Gambs

While there were already some privacy challenges organized in the domain of data sanitization, they have mainly focused on the defense side of the problem. To favor the organization of successful challenges focusing on attacks, we introduce the SNAKE framework that is designed to facilitate the organization of challenges dedicated to attacking existing data sanitization mechanisms. In particular, it enables to easily automate the redundant tasks that are inherent to any such challenge and exhibits the following salient features: genericity with respect to attacks, ease of use and extensibility. We propose to demonstrate the main features of the SNAKE framework through a specific instantiation focusing on membership inference attacks over differentially-private synthetic data generation schemes. This instance of the SNAKE framework is currently being used for supporting a challenge co-located with APVP 2023 (the French workshop on the protection of privacy).

#### [AQUAPLANE: The Argument Quality Explainer App](#)

Sebastian Britner, Lorik Dumani, Ralf Schenkel

In computational argumentation, so-called quality dimensions such as coherence or rhetoric are often used for ranking arguments. However, the literature often only predicts which argument is more persuasive, but not why this is the case. In this paper, we introduce AQUAPLANE, a transparent and easy-to-extend application that not only decides for a pair of arguments which one is more convincing with respect to a statement, but also provides an explanation.

#### [DiscoverPath: A Knowledge Refinement and Retrieval System for Interdisciplinarity on Biomedical Research](#)

Yu-Neng Chuang, Guanchu Wang, Chia-Yuan Chang, Kwei-Herng Lai, Daochen Zha, Ruixiang Tang, Fan Yang, Alfredo Costilla Reyes, Kaixiong Zhou, Xiaoqian Jiang, Xia Hu

The exponential growth in scholarly publications necessitates advanced tools for efficient article retrieval, especially in interdisciplinary fields where diverse terminologies are used to describe similar research. Traditional keyword-based search engines often fall short in assisting users who may not be familiar with specific terminologies. To address this, we present a knowledge graph based paper search engine for biomedical research to enhance the user experience in discovering relevant queries and articles. The system, dubbed DiscoverPath, employs Named Entity Recognition (NER) and part-of-speech (POS) tagging to extract terminologies and relationships from article abstracts to create a KG. To reduce information overload, DiscoverPath presents users with a focused subgraph containing the queried entity and its neighboring nodes and incorporates a query recommendation system enabling users to iteratively refine their queries. The system is equipped with an accessible Graphical User Interface that provides an intuitive visualization of the KG, query recommendations, and detailed article information, enabling efficient article retrieval, thus fostering interdisciplinary knowledge exploration.

### **Contrastive Keyword Extraction from Versioned Documents**

Lukas Eder, Ricardo Campos, Adam Jatowt

Versioned documents are common in many situations and play a vital part in numerous applications enabling an overview of the revisions made to a document or document collection. However, as documents increase in size, it gets difficult to summarize and comprehend all the changes made to versioned documents. In this paper, we propose a novel research problem of contrastive keyword extraction from versioned documents, and introduce an unsupervised approach that extracts keywords to reflect the key changes made to an earlier document version. In order to provide an easy-to-use comparison and summarization tool, an open-source demonstration is made available which can be found at <https://contrastive-keyword-extraction.streamlit.app/>

### **Learning What to Ask: Mining Product Attributes for E-commerce Sales from Massive Dialogue Corpora**

Yan Fan, Chengyu Wang, Fan Feng, Hengbin Cui, Yuchuan Wu, Yongbin Li

Conversational Recommender Systems (CRSs) are extensively applied in e-commercial platforms that recommend items to users. To ensure accurate recommendation, agents usually ask for users' preferences towards specific product attributes which are pre-defined by humans. In e-commercial platforms, however, the number of products easily reaches to billions, making it prohibitive to pre-define decisive attributes for efficient recommendation due to the lack of substantial human resources and the scarce domain expertise. In this work, we present AliMeMOSAIC, a novel knowledge mining and conversational assistance framework that extracts core product attributes from massive dialogue corpora for better conversational recommendation experience. It first extracts user-agent interaction utterances from massive corpora that contain product attributes. A Joint Attribute and Value Extraction (JAVE) network is designed to extract product attributes from user-agent interaction utterances. Finally, AliMeMOSAIC generates attribute sets that frequently appear in dialogues as the target attributes for agents to request, and serve as an assistant to guide the dialogue flow. To prove the effectiveness of AliMeMOSAIC, we show that it consistently improves the overall recommendation performance of our CRS system. An industrial demonstration scenario is further presented to show how it benefits online shopping experiences.

### **ParkFlow: Intelligent Dispersal for Mitigating Parking Shortages Using Multi-Granular Spatial-Temporal Analysis**

Yang Fan Chiang, Chun-Wei Shen, Jhe-Wei Tsai, Pei-Xuan Li, Tzu-Chang Lee, Hsun-Ping Hsieh

Parking behaviors near popular destinations often exhibit a preference for proximity, resulting in poor habits, limited parking availability, and a range of consequential issues such as traffic chaos, economic challenges due to congestion, and imbalanced parking utilization. Taiwan has also faced escalating challenges in this regard. To effectively address these issues, the Government of Taiwan has initiated the Smart City program, encompassing various initiatives to enhance urban functionality. One notable solution implemented under this program is the Smart Parking Meter System (SPMS), designed to enhance the overall parking experience. The SPMS incorporates intelligent billing and secure parking data transmission, ensuring a safer and improved parking environment. In this paper, we propose ParkFlow, a comprehensive software-based solution that seamlessly integrates with smart parking hardware, presenting a holistic approach to tackling these challenges. ParkFlow intelligently disperses parking shortages in highly frequented areas and addresses the problem from multiple perspectives, including user, engineering, and government scenarios. By exploring and addressing these scenarios, we aim to provide valuable insights and inspiration to regions worldwide grappling with similar parking-related difficulties. Based on historical data analysis, the implementation of ParkFlow in resolving the parking imbalance problem is anticipated to lead to a

significant increase of up to 10% to 20% in available parking hours in popular areas of Tainan, Taiwan. ParkFlow is in the process of being integrated into the Tainan City Government's Parking application, indicating its potential to address real-world parking challenges.

### **The $\mu$ -RA System for Recursive Path Queries over Graphs**

Amela Fejza, Pierre Genevès, Nabil Layaïda, Sarah Chlyah

We demonstrate a system for recursive query answering over graphs. The system is based on a complete implementation of the recursive relational algebra  $\mu$ -RA, extended with parsers and compilers adapted for queries over knowledge and property graphs. Each component of the system comes with novelty for processing recursion. As a result, one can formulate, optimize and efficiently answer expressive queries that navigate recursively along paths in different types of graphs. We demonstrate the system on real datasets and show how it performs considering other state-of-the-art systems.

### **DataDoc Analyzer: A Tool for Analyzing the Documentation of Scientific Datasets**

Joan Giner-Miguel, Abel Gómez, Jordi Cabot

Recent public regulatory initiatives and relevant voices in the ML community have identified the need to document datasets according to several dimensions to ensure the fairness and trustworthiness of machine learning systems. In this sense, the data-sharing practices in the scientific field have been quickly evolving in the last years, with more and more research works publishing technical documentation together with the data for replicability purposes. However, this documentation is written in natural language, and its structure, content focus, and composition vary, making them challenging to analyze.

We present DataDoc Analyzer, a tool for analyzing the documentation of scientific datasets by extracting the details of the main dimensions required to analyze the fairness and potential biases. We believe that our tool could help improve the quality of scientific datasets, aid dataset curators during its documentation process, and be a helpful tool for empirical studies on the overall quality of the datasets used in the ML field. The tool implements an ML pipeline that uses Large Language Models at its core for information retrieval. DataDoc is open-source, and a public demo is published online.

### **MORPHER: Structural Transformation of Ill-formed Rows**

Mazhar Hameed, Gerardo Vitagliano, Felix Naumann

Open data portals contain a plethora of data files, with comma-separated value (CSV) files being particularly popular with users and businesses due to their flexible standard. However, this flexibility comes with much responsibility for data consumers, as many files contain various structural problems, e.g., a different number of cells across data rows, multiple value formats within the same column, different variants of quoted fields due to user specifications, etc. We refer to rows that contain such structural inconsistencies as ill-formed. Consequently, ingesting them into a host system, such as a database or an analytics platform, often requires prior data preparation steps.

We propose to demonstrate MORPHER, a desktop-based system that incorporates our state-of-the-art error detection system, SURAGH and extends it to also clean the files at hand. MORPHER facilitates ingesting CSV files by automatically identifying and cleaning ill-formed rows while preserving all data. It comprises three key components: 1) The pattern modeler, which generates syntax-based patterns for each row of the input file. The system uses these patterns to classify rows into ill-formed and well-formed. 2) The pattern classifier obtains row patterns for ill-formed rows and uses them to distinguish ill-formed but wanted rows from ill-formed unwanted rows. 3) The pattern wrangler transforms the identified wanted rows into well-formed rows, effectively repairing a wide range of formatting problems.

### **SEDAR: A Semantic Data Reservoir for Heterogeneous Datasets**

Sayed Hoseini, Ahmed Ali, Haron Shaker, Christoph Quix

Data lakes have emerged as a solution for managing vast and diverse datasets for modern data analytics. To prevent them from becoming ungoverned, semantic data management techniques are crucial, which involve connecting metadata with knowledge graphs, following the principles of Linked Data. This semantic layer enables more expressive data management, integration from various sources and enhances data access utilizing the concepts and relations to semantically enrich the data. Some frameworks have been proposed, but requirements like data versioning, linking of datasets, managing machine learning projects, automated semantic modeling and ontology-based data access are not supported in one uniform system. We demonstrate SEDAR, a comprehensive semantic data lake that includes support for data ingestion, storage, processing, and governance with a special focus on semantic data management. The demo will showcase how the system allows for



various ingestion scenarios, metadata enrichment, data source linking, profiling, semantic modeling, data integration and processing inside a machine learning life cycle.

### **EFFECTS: Explorable and Explainable Feature Extraction Framework for Multivariate Time-Series Classification**

Ido Ikar, Amit Somech

We demonstrate EFFECTS, an automated system for explorable and explainable feature extraction for multivariate time series classification. EFFECTS has a twofold contribution: (1) It significantly facilitates the exploration of MTSC data, and (2) it generates informative yet intuitive and explainable features to be used by the classification model. EFFECTS first mines the MTS data and extracts a set of interpretable features using an optimized transform-slice-aggregate process. To evaluate the quality of EFFECTS features, we gauge how well each feature distinguishes between every two classes, and how well they characterize each single class. Users can then explore the MTS data via the EFFECTS Explorer, which facilitates the visual inspection of important features, dimensions, and time slices. Last, the user can use the top features for each class when building a classification pipeline.

We demonstrate EFFECTS on several real-world MTSC datasets, inviting the audience to investigate the data via EFFECTS Explorer and obtain initial insights on the time series data. Then, we will show how EFFECTS features are used in an ML model, and obtain accuracy that is on par with state-of-the-art MTSC models that do not optimize on explainability.

### **LARCH: Large Language Model-based Automatic Readme Creation with Heuristics**

Yuta Koreeda, Terufumi Morishita, Osamu Imaichi, Yasuhiro Sogawa

Writing a readme is a crucial aspect of software development as it plays a vital role in managing and reusing program code. Though it is a pain point for many developers, automatically creating one remains a challenge even with the recent advancements in large language models (LLMs), because it requires generating an abstract description from thousands of lines of code. In this demo paper, we show that LLMs are capable of generating a coherent and factually correct readmes if we can identify a code fragment that is representative of the repository. Building upon this finding, we developed LARCH (LLM-based Automatic Readme Creation with Heuristics) which leverages representative code identification with heuristics and weak supervision. Through human and automated evaluations, we illustrate that LARCH can generate coherent and factually correct readmes in the majority of cases, outperforming a baseline that does not rely on representative code identification. We have made LARCH open-source and provided a cross-platform Visual Studio Code interface and command-line interface, accessible at <https://github.com/hitachi-nlp/larch>. A demo video showcasing LARCH's capabilities is available at <https://youtu.be/ZUKkh5ED-O4>.

### **CRUISE-Screening: Living Literature Reviews Toolbox**

Wojciech Kusa, Petr Knoth, Allan Hanbury

Keeping up with research and finding related work is still a time-consuming task for academics. Researchers sift through thousands of studies to identify a few relevant ones. Automation techniques can help by increasing the efficiency and effectiveness of this task. To this end, we developed CRUISE-Screening, a web-based application for conducting living literature reviews -- a type of literature review that is continuously updated to reflect the latest research in a particular field. CRUISE-Screening is connected to several search engines via an API, which allows for updating the search results periodically. Moreover, it can facilitate the process of screening for relevant publications by using text classification and question answering models. CRUISE-Screening can be used both by researchers conducting literature reviews and by those working on automating the citation screening process to validate their algorithms. The application is open-source, and a demo is available under this URL: <https://citation-screening.ec.tuwien.ac.at>.

### **STREAMER 3.0: Towards Online Monitoring and Distributed Learning**

Baudouin Naline, Sandra Garcia-Rodriguez, Karine Zeitouni

Applications that generate continuous data have proliferated in recent years, and thus the challenge of processing those data streams has emerged. This requires Data Stream Processing frameworks with monitoring capabilities able to detect and react to any non-desired situation. Many streaming use cases deal with distributed sources of data which, for privacy and communication saving purposes, need to be tackled in a distributed manner. Based on the mentioned challenges, this paper presents STREAMER 3.0, an improvement on the former data stream framework with two new modules: (i) a monitoring manager with detection algorithms, alert raising and automatic model updater; and (ii) a distributed learning module relying

on federated learning. We showcase these new functionalities with an example of remaining useful life estimation of turbofan engines using an LSTM.

### **gFOV: A Full-Stack SPARQL Query Optimizer & Plan Visualizer**

Yue Pang, Linglin Yang, Lei Zou, M. Tamer Özsu

SPARQL is the standard query language for RDF data. A SPARQL query consists of basic graph patterns (BGPs), which are matched onto the data graph, and graph pattern operators, which specify how to merge the matched results. Despite the prevalence of graph pattern operators in real-world SPARQL workloads, the optimization of SPARQL queries with graph pattern operators has scarcely been studied. We hence propose gFOV, a full-stack SPARQL query optimizer and plan visualizer targeting both BGPs and graph pattern operators. As its basis, we propose a novel BGP-based evaluation tree (BE-tree) plan representation for SPARQL queries that integrates the physical plan for BGPs, which directly accesses the RDF store, and the logical plan for graph pattern operators, which operates on existent results in memory. On top of it, we devise a full-stack cost-based optimization scheme, combining logical and physical plan optimization, that outperforms the state-of-the-art. In the demonstration, we present an interactive interface that explains our optimization scheme and shows its efficiency by visualizing changes in the query plan and allowing the audience to inspect and execute alternative plans.

### **A Data-Driven Index Recommendation System for Slow Queries**

Gan Peng, Peng Cai, Kaikai Ye, Kai Li, Jinlong Cai, Yufeng Shen, Han Su

The Database Autonomy Service (DAS) is a platform designed to assist database administrators in managing a large number of database instances within major internet companies. One of the key tasks in DAS is to find missing indexes to improve the slow query execution. In Meituan, a vast array of business lines deploy tens of thousands of MySQL database instances. Consequently, a great number of human-generated index cases are accumulated in the DAS platform. This motivates us to build a data-driven index recommendation system, referred to as idxLearner, which can learn index creation knowledge from human-generated index cases. In this demonstration, users can interact with idxLearner by choosing source databases to construct the training data, training the recommendation model, inputting slow queries for various target databases, and observing the recommended indexes and their evaluation results.

### **Misinformation Concierge: A Proof-of-Concept with Curated Twitter Dataset on COVID-19 Vaccination**

Shakshi Sharma, Anwitaman Datta, Vigneshwaran Shankaran, Rajesh Sharma

We demonstrate the Misinformation Concierge, a proof-of-concept that provides actionable intelligence on misinformation prevalent in social media. Specifically, it uses language processing and machine learning tools to identify subtopics of discourse and discerns non/misleading posts; presents statistical reports for policy-makers to understand the big picture of prevalent misinformation in a timely manner; and recommends rebuttal messages for specific pieces of misinformation, identified from within the corpus of data - providing means to intervene and counter misinformation promptly. The Misinformation Concierge proof-of-concept using a curated dataset is accessible at: <https://demo-frontend-uy34.onrender.com/>

### **NumJoin: Discovering Numeric Joinable Tables with Semantically Related Columns**

Pranav Subramaniam, Udayan Khurana, Kavitha Srinivas, Horst Samulowitz

Join discovery is a crucial part of exploration on data lakes. It often involves finding joinable tables that are semantically relevant. However, data lakes often contain numeric tables with unreliable column headers, and ID columns whose text names have been lost. Finding semantically relevant joins over numeric tables is a challenge. State-of-the-art describes join discovery using semantic similarity, but do not consider purely numeric tables. In this paper, we describe a system, NumJoin that includes two novel approaches for discovering joinable tables in a data lake: one that maps tables to knowledge graphs, and another that leverages numeric types and distributions. We demonstrate the effectiveness of NumJoin on a large data lake, including transportation data and finance data.

### **IMinimize: A System for Negative Influence Minimization via Vertex Blocking**

Siyi Teng, Jiadong Xie, Mingkai Zhang, Kai Wang, Fan Zhang

The rapid rise and prevalence of social platforms have created great demands on effective schemes to limit the influence of negative information, e.g., blocking key vertices for influence minimization. However, there is currently no system providing practical schemes to solve the negative influence minimization problem with a blocking budget effectively and efficiently in

the literature. In this demo, we present IMinimize, the first interactive system that provides audiences with vertex-blocking schemes over different budgets and demonstrates via visualization for comparison vividly and directly, aiming to help minimize the negative influence spreading in networks. Our IMinimize system applies an advanced greedy algorithm to select blocked vertices with both high efficiency and effectiveness. Furthermore, we extend IMinimize to the application of epidemic controlling and prevention and show the usability of IMinimize through two case studies of real-life applications.

### **Cluster-Explorer: An interactive Framework for Explaining Black-Box Clustering Results**

Sariel Tutay, Amit Somech

Interpreting clustering results is a challenging, manual task, that often requires the user to perform additional analytical queries and visualizations. To this end, we demonstrate Cluster-Explorer, an interactive, easy-to-use framework that provides explanations for black-box clustering results. Cluster-Explorer takes as input the raw dataset alongside cluster labels, and automatically generates multiple coherent explanations that characterize each cluster.

We first propose a threefold quality measure that considers the conciseness, cluster coverage, and separation error of an explanation. We tackle the challenge of efficiently computing high-quality explanations using a modified version of a generalized frequent-itemsets mining (gFIM) algorithm. The gFIM algorithm is employed over multiple filter predicates which are extracted by applying various binning methods of different granularities. We implemented Cluster-Explorer as a Python library that can be easily used by data scientists in their ongoing workflows. After employing the clustering pipeline of their choice, Cluster-Explorer opens an integrated, interactive interface for the user to explore the various different explanations for each cluster.

In our demonstration, the audience is invited to use Cluster-Explorer on numerous real-life datasets and different clustering pipelines and examine the usefulness of the cluster explanations provided by the system, as well as its efficiency of computation.

### **HugNLP: A Unified and Comprehensive Library for Natural Language Processing**

Jianing Wang, Nuo Chen, Qiushi Sun, Wenkang Huang, Chengyu Wang, Ming Gao

In this paper, we introduce HugNLP, a unified and comprehensive library for natural language processing (NLP) with the prevalent backend of Hugging Face Transformers, which is designed for NLP researchers to easily utilize off-the-shelf algorithms and develop novel methods with user-defined models and tasks in real-world scenarios. HugNLP consists of a hierarchical structure including models, processors and applications that unifies the learning process of pre-trained language models (PLMs) on different NLP tasks. Additionally, we present some featured NLP applications to show the effectiveness of HugNLP, such as knowledge-enhanced PLMs, universal information extraction, low-resource mining, and code understanding and generation, etc. The source code will be released on GitHub (<https://github.com/HugAILab/HugNLP>).

### **PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis**

Heng Yang, Chen Zhang, Ke Li

The advancement of aspect-based sentiment analysis (ABSA) has highlighted the lack of a user-friendly framework that can significantly reduce the difficulty of reproducing state-of-the-art ABSA performance, especially for beginners. To meet this demand, we present PyABSA, a modularized framework built on PyTorch for reproducible ABSA. To facilitate ABSA research, PyABSA supports several ABSA subtasks, including aspect term extraction, aspect sentiment classification, and end-to-end aspect-based sentiment analysis. With just a few lines of code, the result of a model on a specific dataset can be reproduced. With a modularized design, PyABSA can also be flexibly extended to incorporate new models, datasets, and other related tasks. Additionally, PyABSA highlights its data augmentation and annotation features, which significantly address data scarcity. The project is available at: <https://github.com/yangheng95/PyABSA>.

### **NP-SSL: A Modular and Extensible Self-supervised Learning Library with Neural Processes**

Zesheng Ye, Jing Du, Yao Liu, Yihong Zhang, Lina Yao

Neural Processes (NPs) are a family of supervised density estimators devoted to probabilistic function approximation with meta-learning. Despite extensive research on the subject, the absence of a unified framework for NPs leads to varied architectural solutions across diverse studies. This non-consensus poses challenges to reproducing and benchmarking different NPs. Moreover, existing codebases mainly prioritize generative density estimation, yet rarely consider expanding the capability of NPs to self-supervised representation learning, which however has gained growing importance in data mining

applications. To this end, we present NP-SSL, a modular and configurable framework with built-in support, requiring minimal effort to 1) implement classical NPs architectures; 2) customize specific components; 3) integrate hybrid training scheme (e.g., contrastive); and 4) extend NPs to act as a self-supervised learning toolkit, producing latent representations of data, and facilitating diverse downstream predictive tasks. To illustrate, we discuss a case study that applies NP-SSL to model time-series data. We interpret that NP-SSL can handle different predictive tasks such as imputation and forecasting, by a simple switch in data samplings, without significant change to the underlying structure. We hope this study can reduce the workload of future research on leveraging NPs to tackle more a broader range of real-world data mining applications. Code and documentation are at <https://github.com/zyecs/NP-SSL>.

#### **MOSS: AI Platform for Discovery of Corrosion-Resistant Materials**

Biao Yin, Nicholas Josselyn, Ziming Zhang, Elke A. Rundensteiner, Thomas A. Considine, John V. Kelley, Berend C. Rinderspacher, Robert E. Jensen, James F. Snyder

Amid corrosion degradation of metallic structures causing expenses nearing 3 trillion or 4% of the GDP annually along with major safety risks, the adoption of AI technologies for accelerating the materials science life-cycle for developing materials with better corrosive properties is paramount. While initial machine learning models for corrosion assessment are being proposed in the literature, their incorporation into end-to-end tools for field experimentation by corrosion scientists remains largely unexplored. To fill this void, our university data science team in collaboration with the materials science unit at the Army Research Lab have jointly developed MOSS, an innovative AI-based digital platform to support material science corrosion research. MOSS features user-friendly iPadOS app for in-field corrosion progression data collection, deep-learning corrosion assessor, robust data repository system for long-term experimental data modeling, and visual analytics web portal for material science research. In this demonstration, we showcase the key innovations of the MOSS platform via use cases supporting the corrosion exploration processes, with the promise of accelerating the discovery of new materials. We open a MOSS video demo at: <https://www.youtube.com/watch?v=CzcMMRsxE>

#### **Demonstration of ViTA: Visualizing, Testing and Analyzing Index Advisors**

Wei Zhou, Chen Lin, Xuanhe Zhou, Guoliang Li, Tianqing Wang

Index advisors have become an essential tool to optimize index selection and accelerate query processing. Various index advisors have been developed in recent years, and comprehensively assessing their performance from multiple aspects is necessary. In this demonstration, we introduce VITA, a user-friendly and informative tool for interactively Visualizing, Testing, and Analyzing index advisors. For a user-given workload, VITA can visualize the main steps of the index selection procedure in ten existing index advisors to facilitate the management of index advisors. Moreover, VITA can assess the index advisor's robustness w.r.t. workload drift by generating testing workloads, i.e., potentially future workloads that may damage the index advisor's performance. Finally, VITA provides a comparative analysis across index advisors on four aspects, including the index advisor's utility (i.e., the ratio of the reduced workload cost), robustness (i.e., the performance under dynamic workload), overhead (i.e., the time to acquire the final configuration), and scalability (i.e., the volume of the enumerated index candidates). Therefore, VITA can thoroughly compare existing index advisors to help users determine the most suitable index advisor that meets their requirements. VITA is now being integrated into the openGauss platform as a plug-in.

#### **An AI-based Simulation and Optimization Framework for Logistic Systems**

Zefang Zong, Huan Yan, Hongjie Sui, Haoxiang Li, Peiqi Jiang, Yong Li

Improving logistics efficiency is a challenging task in logistic systems, since planning the vehicle routes highly relies on the changing traffic conditions and diverse demand scenarios. However, most existing approaches either neglect the dynamic traffic environment or adopt manually designed rules, which fails to efficiently find a high-quality routing strategy. In this paper, we present a novel artificial intelligence (AI) based framework for logistic systems. This framework can simulate the spatio-temporal traffic conditions to form a dynamic environment in a data-driven manner. Under such a simulated environment, it adopts deep reinforcement learning techniques to intelligently generate the optimized routing strategy. Meanwhile, we also design an interactive frontend to visualize the simulated environment and routing strategies, which help operators evaluate the task performance. We will showcase the results of AI-based simulation and optimization in our demonstration.

### **SESSION: Doctoral Consortium**

#### **Towards Effective Modeling and Exploitation of Search and User Context in Conversational Information Retrieval**

Praveen Acharya

Conversational information retrieval has garnered considerable attention in recent years. A major challenge in conversational search is formulating the most effective query during the dialogue between the searcher and the conversational agent. Unlike traditional information retrieval systems that assume users can independently create queries, conversational settings allow agents to assist users in query formulation. This alleviates the burden on users by leveraging the multi-turn nature of the conversation to aid them in reaching their information goals. Conversational context plays a vital role in the query process. In this work, we focus on understanding and leveraging conversational context from two dimensions: conversational history and knowledge history. The goal is to identify and model the relevant portions from the search dialogue and the knowledge history and to use this within the search process to improve the overall performance of conversational information retrieval systems.

#### [Towards Improving Accuracy and Computation Cost Optimization of Recommendation Systems](#)

Abdelghani Azri

It is hard to avoid recommender systems (RS) these days which play a vital role in various domains, such as e-commerce, online streaming platforms, and personalized content delivery. These systems assist users in discovering relevant items based on their preferences and past interactions. A variety of methods are developed by RS communities, to address the accuracy issue. But, most of these methods are sequential and omit the item features and user side information which contains relevant and rich information that can increase the accuracy of these systems. However, enhancing the accuracy of recommendations often comes at the expense of increased computational costs. This PhD thesis aims to address the challenge of improving the accuracy of RS following a hybrid approach that allows leveraging user-item features based on some relevant state-of-the-art models and using deep learning techniques such as the contractive autoencoder (CAE) while optimizing the cost of computation using parallel/distributed paradigms.

#### [Data and Decision Fusion with Uncertainty Quantification for ML-based Healthcare Decision Systems](#)

Grigor Bezirganyan

This paper outlines the PhD research plan to develop a comprehensive, uncertainty-aware, multimodal deep learning approach to be used in the healthcare domain. The goal is to design a multimodal deep learning framework that can leverage the complex interconnections between various modalities in order to generate highly precise predictions for intricate healthcare datasets. In addition, the framework should also incorporate methods for quantifying and dealing with uncertainty, which is an important consideration in many real-world healthcare applications. The approach will be tested on real-world multimodal datasets from Marseilles hospitals in France. We further represent some preliminary results of our early stage experiments with uncertainty quantification on multimodal datasets.

#### [Investigating Natural and Artificial Dynamics in Graph Data Mining and Machine Learning](#)

Dongqi Fu

The complexity of relationships between entities is increasing in the era of big data, leading to a growing interest in graph (network) data, owing to its ability to encode intricate relational information. Graph data mining and machine learning methods extract informative node and graph representations to support broad applications, which have been proven effective for various high-impact tasks in the fields such as computer vision, natural language processing, and recommendation systems. Despite their effectiveness, graph data mining and machine learning methods face practical challenges in real-world scenarios. First, the input graphs may change over time, making it necessary to integrate time-evolving information to enhance representation capabilities. Second, the input graphs may contain unreliable, noisy, or sub-optimal information, requiring researchers and practitioners to intentionally modify the graph topology and node features to improve downstream performance. Facing these two phenomena, our research works focus on natural and artificial dynamics for benefiting graph data mining and machine learning. In this paper, we will briefly introduce our recent works in investigative natural and artificial dynamics and point out some under-explored research problems in the interaction of these dynamics.

#### [A Neuro-symbolic Approach to Enhance Interpretability of Graph Neural Network through the Integration of External Knowledge](#)

Kislay Raj

Graph Neural Networks (GNNs) have shown remarkable performance in tackling complex tasks. However, interpreting the decision-making process of GNNs remains a challenge. To address the challenge, we explore representing the behaviour of a GNN in a representation space that is more transparent such as a knowledge graph, in a way that captures the behaviour of a GNN as a graph. Our initial experiments on the node classification task can represent the trained graph convolutional neural network (GCN) behaviour with some semantics uncovered by state-of-the-art approaches. This research offers a



promising direction for enhancing GNN interpretability and understanding by providing structured, human-understandable representations and incorporating external knowledge for more accurate predictions.

### [Intersectional Bias Mitigation in Pre-trained Language Models: A Quantum-Inspired Approach](#)

Omid Shokrollahi

The growing criticality of contextualized language models has raised concerns about the perpetuation of biases. Current fairness research often concentrates on single aspects of social groups. However, our study brings a multifaceted approach to the table, reflecting the intricate realities of intersectionality. We propose a bias mitigation algorithm inspired by quantum theory to fine-tune pre-trained language models. Our method applies Jensen-Shannon divergence alongside an entanglement entropy measure to quantify the complexity of entwined identities. Moreover, we utilize an innovative entanglement embedding neural network to address emergent features from different intersectional groups. An Actor-Critic setup facilitates effective fine-tuning. Our approach broadens the scope of intersectional fairness beyond just statistical parity, providing a strategic tool to tackle complex, interrelated biases. We anticipate that this fresh approach to bias mitigation will substantially enhance fairness in a wide range of language model applications.

### [Explaining Learning to Rank Methods to Improve Them](#)

Alberto Veneri

State-of-the-art methods for Learning to Rank (LtR), either designed for tabular or textual data, are incredibly complex. Increasing the complexity of the models has many drawbacks, including difficulties in understanding the logic behind each prediction and a lack of trust in the system during its deployment. In this paper, which describes the author's goals during his Ph.D., there is an analysis and discussion of how we can use the ideas and tools coming from the eXplainable Artificial Intelligence (XAI) field to make the most effective methods for LtR understandable to the practitioners with the final goal of making them more efficient and/or understand better when they can be improved. The strategies adopted to achieve the aforementioned goals are different and based on the type of models analyzed, which go from more traditional LtR models based on ensembles of decision trees and using handcrafted features to fairly new neural LtR models using text data.

### [Enhancing Badminton Player Performance via a Closed-Loop AI Approach: Imitation, Simulation, Optimization, and Execution](#)

Kuang-Da Wang

In recent years, the sports industry has witnessed a significant rise in interest in leveraging artificial intelligence to enhance players' performance. However, the application of deep learning to improve badminton athletes' performance faces challenges related to identifying weaknesses, generating winning suggestions, and validating strategy effectiveness. These challenges arise due to the limited availability of realistic environments and agents. This paper aims to address these research gaps and make contributions to the badminton community. To achieve this goal, we propose a closed-loop approach consisting of six key components: Badminton Data Acquisition, Imitating Players' Styles, Simulating Matches, Optimizing Strategies, Training Execution, and Real-World Competitions. Specifically, we developed a novel model called RallyNet, which excels at imitating players' styles, allowing agents to accurately replicate real players' behavior. Secondly, we created a sophisticated badminton simulation environment that incorporates real-world physics, faithfully recreating game situations. Thirdly, we employed reinforcement learning techniques to improve players' strategies, enhancing their chances of winning while preserving their unique playing styles. By comparing strategy differences before and after improvement, we provide winning suggestions to players, which can be validated against diverse opponents within our carefully designed environment. Lastly, through collaborations with badminton venues and players, we apply the generated suggestions to the players' training and competitions, ensuring the effectiveness of our approach. Moreover, we continuously gather data from training and competitions, incorporating it into the closed-loop cycle to refine strategies and suggestions. This research presents an innovative approach for continuously improving players' performance, contributing to the field of AI-driven sports performance enhancement. This dissertation is supervised by Wen-Chih Peng (wcpeng@cs.nycu.edu.tw) and Ping-Chun Hsieh (pinghsieh@nycu.edu.tw).

### [Exploiting Homeostatic Synaptic Modulation in Spiking Neural Networks for Semi-Supervised Graph Learning](#)

Mingkun Xu

Semi-supervised graph learning (SSL) is an important task in machine learning that aims to make predictions based on a limited amount of labeled data and a larger set of unlabeled structured data, which can be effectively processed by biological neural networks. In this paper, we investigate the effects of the underlying homeostatic synaptic modulation (HSM) in spiking

neural networks (SNNs) on such scenario. We propose a novel framework that integrates HSM into the spiking graph convolutional network to maintain stability by regulating the strength of synapses based on the activity of neurons, allowing for stable graph learning in a semi-supervised setting. Experimental results on citation benchmark datasets demonstrate that the proposed HSM mechanism can enable SNNs with superior capabilities of convergence and generalization, meanwhile possessing expected characteristics of sparsity and call-back phenomenon. The proposed framework provides a promising approach for exploiting HSM in neural network architectures for efficient graph learning.

## SESSION: Tutorial Proposals

### [Some Useful Things to Know When Combining IR and NLP: the Easy, the Hard and the Ugly](#)

Omar Alonso, Kenneth Church

Deep nets such as GPT are at the core of the current advances in many systems and applications. Things are moving very fast, and it appears that techniques are out of date within weeks. How can we take advantage of new discoveries and incorporate them into our existing work? Are these radical new developments, repetitions of older concepts, or both?

In this tutorial, we aim to bring interested researchers and practitioners up to speed on the recent and ongoing techniques around ML and Deep learning in the context of IR and NLP. Additionally, our goal is to clarify terminology, emphasize fundamentals, and outline new research opportunities.

### [Tutorial on User Simulation for Evaluating Information Access Systems](#)

Krisztian Balog, ChengXiang Zhai

With the emergence of various information access systems exhibiting increasing complexity, there is a critical need for sound and scalable means of automatic evaluation. To address this challenge, user simulation emerges as a promising solution. This half-day tutorial focuses on providing a thorough understanding of user simulation techniques designed specifically for evaluation purposes. We systematically review major research progress, covering both general frameworks for designing user simulators, and specific models and algorithms for simulating user interactions with search engines, recommender systems, and conversational assistants. We also highlight some important future research directions.

### [Proactive Streaming Analytics at Scale: A Journey from the State-of-the-art to a Production Platform](#)

Nikos Giatrakos, Elias Alevizos, Antonios Deligiannakis, Ralf Klinkenberg, Alexander Artikis

Proactive streaming analytics continuously extract real-time business value from massive data that stream in data centers or clouds. This requires (a) to process the data while they are still in motion; (b) to scale the processing to multiple machines, often over various, dispersed computer clusters, with diverse Big Data technologies; and (c) to forecast complex business events for proactive decision-making. Combining the necessary facilities for proactive streaming analytics at scale entails: (I) deep knowledge of the relevant state-of-the-art, (II) cherry-picking cutting edge research outcomes based on desired features and with the prospect of building interoperable components, and (III) building components and deploying them into a holistic architecture within a real-world platform. In this tutorial, we drive the audience through the whole journey from (I) to (III), delivering cutting edge research into a commercial analytics platform, for which we provide a hands-on experience.

### [Application of Deep Clustering Algorithms](#)

Collin Leiber, Lukas Miklautz, Claudia Plant, Christian Böhm

Deep clustering algorithms have gained popularity for clustering complex, large-scale data sets, but getting started is difficult because of numerous decisions regarding architecture, optimizer, and other hyperparameters. Theoretical foundations must be known to obtain meaningful results. At the same time, ease of use is necessary to get used by a broader audience. Therefore, we require a unified framework that allows for easy execution in diverse settings. While this applies to established clustering methods like k-Means and DBSCAN, deep clustering algorithms lack a standard structure, resulting in significant programming overhead. This complicates empirical evaluations, which are essential in both scientific and practical applications. We present a solution to this problem by providing a theoretical background on deep clustering as well as practical implementation techniques and a unified structure with predefined neural networks. For the latter, we use the Python package ClustPy. The aim is to share best practices and facilitate community participation in deep clustering research.

### [Uplift Modeling: From Causal Inference to Personalization](#)

Felipe Moraes, Hugo Manuel Proença, Anastasiia Kornilova, Javier Albert, Dmitri Goldenberg

Uplift modeling is a collection of machine learning techniques for estimating causal effects of a treatment at the individual or subgroup levels. Over the last years, causality and uplift modeling have become key trends in personalization at online e-commerce platforms, enabling the selection of the best treatment for each user in order to maximize the target business metric. Uplift modeling can be particularly useful for personalized promotional campaigns, where the potential benefit caused by a promotion needs to be weighed against the potential costs. In this tutorial we will cover basic concepts of causality and introduce the audience to state-of-the-art techniques in uplift modeling. We will discuss the advantages and the limitations of different approaches and dive into the unique setup of constrained uplift modeling. Finally, we will present real-life applications and discuss challenges in implementing these models in production.

### [Leveraging Graph Neural Networks for User Profiling: Recent Advances and Open Challenges](#)

Erasmo Purificato, Ludovico Boratto, Ernesto William De Luca

The proposed tutorial aims to familiarise the CIKM community with modern user profiling techniques that utilise Graph Neural Networks (GNNs). Initially, we will delve into the foundational principles of user profiling and GNNs, accompanied by an overview of relevant literature. We will subsequently systematically examine cutting-edge GNN architectures specifically developed for user profiling, highlighting the typical data utilised in this context. Furthermore, ethical considerations and beyond-accuracy perspectives, e.g. fairness and explainability, will be discussed regarding the potential applications of GNNs in user profiling. During the hands-on session, participants will gain practical insights into constructing and training recent GNN models for user profiling using open-source tools and publicly available datasets. The audience will actively explore the impact of these models through case studies focused on bias analysis and explanations of user profiles. To conclude the tutorial, we will analyse existing and emerging challenges in the field and discuss future research directions.

### [Data Augmentation for Conversational AI](#)

Heydar Soudani, Evangelos Kanoulas, Faegheh Hasibi

Advancements in conversational systems have revolutionized information access, surpassing the limitations of single queries. However, developing dialogue systems requires a large amount of training data, which is a challenge in low-resource domains and languages. Traditional data collection methods like crowd-sourcing are labor-intensive and time-consuming, making them ineffective in this context. Data augmentation (DA) is an effective approach to alleviate the data scarcity problem in conversational systems. This tutorial provides a comprehensive and up-to-date overview of DA approaches in the context of conversational systems. It highlights recent advances in conversation augmentation, open domain and task-oriented conversation generation, and different paradigms of evaluating these models. We also discuss current challenges and future directions in order to help researchers and practitioners to further advance the field in this area.

### [Tutorial: Data Denoising Metrics in Recommender Systems](#)

Pengfei Wang, Chenliang Li, Lixin Zou, Zhichao Feng, Kaiyuan Li, Xiaochen Li, Xialong Liu, Shangguang Wang

Recommender systems play a pivotal role in navigating users through vast reservoirs of information. However, data sparseness can compromise recommendation accuracy, making it challenging to improve recommendation performance. To address this issue, researchers have explored incorporating multiple data types. Yet, this approach can introduce noise that impairs the recommendations' accuracy. Therefore, it is crucial to denoise the data to enhance recommendation quality. This tutorial highlights the importance of data denoising metrics for improving the accuracy and quality of recommendations. Four groups of data denoising metrics are introduced: feature, item, pattern, and modality level. For each group, various denoising methods are presented. The tutorial emphasizes the significance of selecting the right data denoising methods to enhance recommendation quality. It provides valuable guidance for practitioners and researchers implementing reliable data denoising metrics in recommender systems. Finally, the tutorial proposes open research questions for future studies, making it a valuable resource for the research community.

### [Reasoning beyond Triples: Recent Advances in Knowledge Graph Embeddings](#)

Bo Xiong, Mojtaba Nayeri, Daniel Daza, Michael Cochez

Knowledge Graphs (KGs) are a collection of facts describing entities connected by relationships. KG embeddings map entities and relations into a vector space while preserving their relational semantics. This enables effective inference of missing knowledge from the embedding space. Most KG embedding approaches focused on triple-shaped KGs. A great amount of real-world knowledge, however, cannot simply be represented by triples. In this tutorial, we give a systematic introduction to KG embeddings that go beyond the triple representation. In particular, the tutorial will focus on temporal facts where the triples are enriched with temporal information, hyper-relational facts where the triples are enriched with qualifiers,

n-ary facts describing relationships between multiple entities, and also facts that are augmented with literal and text descriptions. During the tutorial, we will introduce both fundamental knowledge and advanced topics for understanding recent embedding approaches for beyond-triple representations.

## SESSION: Industry Day

### [Comparative Analysis of Open Source and Commercial Embedding Models for Question Answering](#)

Georgios Balikas

In this industry track presentation, we will provide a comprehensive tour of the best performing embedding models for question answering, as determined by the Massive Text Embedding Benchmark<sup>1</sup>. We showcase these models while also considering solutions offered by OpenAI and Cohere, renowned for their state-of-the-art performance. Through rigorous evaluations on internal Salesforce datasets tailored for Question Answering on Knowledge articles, we compare the performance of these models using standardized metrics.

Our analysis sheds light on the current state-of-the-art in question answering using embedding models across three diverse domains. We hope that this talk's outcomes will empower practitioners and researchers to make informed decisions when selecting the most suitable solution for their specific requirements.

### [RT2S: A Framework for Learning with Noisy Labels](#)

Indranil Bhattacharya, Ze Ye, Kaushik Pavani, Sunny Dasgupta

We introduce Robust Training with Trust Scores (RT2S), a framework to train machine learning classifiers with potentially noisy labels. RT2S calculates a trust score for each training sample, which indicates the quality of its corresponding label. These trust scores are employed as sample weights during training and optionally during threshold optimization. The trust scores are generated from two sources: (i) the model's confidence in the observed label, leveraging out-of-fold prediction scores to detect anomalous labels in the training data, and (ii) the probability of the correct label, ascertained by a Large Language Model with the ability to identify biased label noise. We evaluate RT2S by training machine learning models on 6 product classification datasets that utilize low-quality labels generated by a rule-based classification engine acting as a surrogate labeler. Our experimental findings indicate that RT2S outperforms all baselines, and achieves an average accuracy improvement of 4.38% (max 7.18%) over rule-based classifiers in particular.

### [From User Activity Traces to Navigation Graph for Software Enhancement: An Application of Graph Neural Network \(GNN\) on a Real-World Non-Attributed Graph](#)

Ikram Boukharouba, Florence Sèdes, Christophe Bortolaso, Florent Mouysset

Understanding software's user behavior is key to personalizing and enriching the user experience and improving the quality of the software. In this paper, we consider the use of user navigation graphs issued from user activity traces. The aim of our study is to do node classification over the user graph navigation in order to understand better the composition of the software and to offer a better experience to the users. Traditional baseline methods has shown good performance in the node classification task, but can't be applied for tasks as link prediction. Graph Neural Network on the contrary can satisfy both node classification and link prediction. However, GNN produce significant results when the features on the nodes are numerous enough. This is not always the case in real-world problems, because too many features implies too much data, storage issues, affect the performances of apps, etc. Indeed, due to the origin of the data and their uncontrolled generation, the resulting graphs contain few or no features (AKA non-attributed graphs). In addition, in industrial fields, some external requirements particularly legal may limit the collection and the use of data. In this article, we show that graphs issued from real-world data also have such limitations, and we propose the generation of artificial features on the nodes as a solution to this problem.

The obtained results showed that the usage of artificial node features is a promising solution to overcome the greediness of GNN in terms of node features and applying GNN on the non-attributed graphs.

### [Type Theory as a Unifying Paradigm for Modern Databases](#)

Christoph Dorn, Haikal Pribadi

Over the past decades, data modeling has become a highly diversified discipline with many competing paradigms emerging across various application domains. We argue that higher levels of abstraction including, in particular, the integration of high-level programming and reasoning techniques, will pave the way forward for future knowledge management systems. As a

programmatic foundation for this endeavor, we will discuss a novel type theoretical modeling and reasoning paradigm, which aims to strike a powerful balance between what can be naturally semantically modeled and what can be practically implemented. TypeQL is a multi-purpose database language rooted in these foundations: it is designed around the expressivity of natural language and backed by type theoretical principles. This rigorous high-level approach to database language reduces development and maintenance loads, preventing hard to spot data integrity and logic errors through its underlying type system, while also providing a unifying toolset for a large class of domain-specific applications ranging from querying connected data in knowledge graphs for drug discovery to reasoning and adaptive decision making in cognitive robotics.

### **Comparing Fine-Tuned Transformers and Large Language Models for Sales Call Classification: A Case Study**

Roy Eisenstadt, Abedelkader Asi, Royi Ronen

We present a research project carried out to enable a Call Categorization Service (CCS) for Dynamics 365 Sales Conversation Intelligence. CCS identifies prevalent types of sales calls based on their transcription, for the purpose of automating manual sales processes and making informed business decisions. We sift through R&D process, and provide clear evidence that purpose-focused fine-tuned transformers out-perform GPT-3 in this text classification task. Additionally we share: an efficient, non-trivial data annotation approach suited to the problem of finding data related to rare categories in a highly unbalanced data source; Considerations regarding zero-shot and in-context learning (i.e. few-shot learning) when using LLMs for classification and cost and performance analysis that opt in favor of fine-tuned transformers as well.

### **Proactive and Automatic Detection of Product Misclassifications at Massive Scale**

Ling Jiang, Xiaoyu Chu, Saaranish Gulati, Pulkit Garg, Andrew Borthwick, Gang Luo

In e-commerce, product classification is widely used for various purposes. Misclassifying products can cause compliance issues and hurt the company's reputation. To address this problem, we propose an automated system to proactively detect product misclassifications by overcoming several challenges. A large e-commerce retailer can sell billions of distinct products, on which many thousands of classification tasks are performed. At this massive scale, we need to quickly detect misclassifications under a limited budget. In this talk, we point out these challenges and show how we design our system to handle them. When evaluated on a set of Amazon's product classification data, at an overhead of <10% of the classification cost, our system automatically identified and corrected many misclassifications, which would take a human many thousand years to manually find and 14.6 years to manually review and correct if our system were not used.

### **Application and Evaluation of Large Language Models for the Generation of Survey Questions**

Antonio Maiorino, Zoe Padgett, Chun Wang, Misha Yakubovskiy, Peng Jiang

Generative Language Models have shown promising results in various domains, and some of the most successful applications are related to "concept expansion", which is the task of generating extensive text based on concise instructions provided through a "seed" prompt. In this presentation we will discuss the recent work conducted by the Data Science team at SurveyMonkey, where we have recently introduced a new feature that harnesses Generative AI models to streamline the survey design process. With this feature users can effortlessly initiate this process by specifying their desired objectives through a prompt, allowing them to automate the creation of surveys that include the critical aspects they wish to investigate.

We will share our findings regarding some of the challenges encountered during the development of this feature. These include techniques for conditioning the model outputs, integrating generated text with industry-standard questions, fine-tuning Language Models using semi-synthetic Data Generation techniques, and more. Moreover, we will showcase the Evaluation Methodology that we have developed to measure the quality of the generated surveys across several dimensions. This evaluation process is crucial in ensuring that the generated surveys align well with user expectations and serve their intended purpose effectively. Our goal is to demonstrate the promising potential of Generative Language Models in the context of Survey Research, and we believe that sharing our learnings on these challenges and how we addressed them will be useful for practitioners working with Language Models on similar problems.

### **Harnessing GPT for Topic-Based Call Segmentation in Microsoft Dynamics 365 Sales**

Itzik Malkiel, Uri Alon, Yakir Yehdua, Shahar Keren, Oren Barkan, Royi Ronen, Noam Koenigstein

Transcriptions of phone calls hold significant value in sales, customer service, healthcare, law enforcement, and more. However, analyzing recorded conversations can be a time-consuming process, especially for complex dialogues. In Microsoft Dynamics 365 Sales, a novel system, named GPT-Calls, is applied for efficient and accurate topic-based call segmentation.



GPT-Calls comprises offline and online phases. In the offline phase, the system leverages a GPT model to generate synthetic sentences and extract anchor vectors for predefined topics. This phase, performed once on a given topic list, significantly reduces the computational burden. The online phase scores the similarity between the transcribed conversation and the topic anchors from the offline phase, followed by time domain analysis to group utterances into segments and tag them with topics. The GPT-Calls scheme offers an accurate and efficient approach to call segmentation and topic extraction, eliminating the need for labeled data. It is a versatile solution applicable to various industry domains. GPT-Calls operates in production under Dynamics 365 Sales Conversation Intelligence, applied to real sales conversations from diverse Dynamics 365 Sales tenants, streamlining call analysis, and saving time and resources while ensuring accuracy and effectiveness.

#### **Astrolabe: Visual Graph Database Queries with Tabular Output**

Michael Miller

Graph databases are an established solution for large, highly connected datasets. One challenge associated with deploying graph databases in industrial settings is usability. Typically, developers interact with graph databases through queries in languages such as Cypher or GraphQL. Many end-users, analysts, and administrators are not familiar with these specialized languages. Additionally, these queries return hierarchical data in formats such as JSON (JavaScript Object Notation) or XML (Extensible Markup Language). Additional scripts and interfaces are needed to convert hierarchical data into more easily digested tables. To overcome these challenges, each graph database use-case typically involves significant custom software to explore, view, and export data.

We introduce Astrolabe, a generalized interface that addresses the challenges of querying graph databases. In Astrolabe, queries are constructed visually, so users do not need to learn new graph query languages. Results are returned as tables, which can be easily digested by end users or down-stream applications. Astrolabe was designed to function with arbitrary graph databases, so schema definition is not required. Astrolabe revolutionizes graph exploration and querying by allowing graph databases to be viewed as tables, without the need for custom software adapters.

#### **Vigil: Effective End-to-end Monitoring for Large-scale Recommender Systems at Glance**

Priyansh Saxena, Manisha R

The success of large-scale recommender systems hinges upon their ability to deliver accurate and timely recommendations to a diverse user base. At Glance, we deliver snackable personalized content to the lock screens of 200M smartphones. In this context, continuous monitoring is paramount as it safeguards data integrity, detects drifts, addresses evolving user preferences, optimizes system downtime, and ultimately augments the system's effectiveness and user satisfaction. In this talk, we delve into Vigil, a set of monitoring practices developed to provide comprehensive end-to-end monitoring of recommender systems at Glance. These practices revolve around three key pillars: mitigating developer fatigue, ensuring precise predictions, and establishing a centralized monitoring framework. By adopting these practices, we have observed a 30% reduction in compute cost, a 26% drop in downtime, and a surge in developer productivity demonstrated by a 45% decrease in turnaround time.

#### **Practical Lessons Learned From Detecting, Preventing and Mitigating Harmful Experiences on Facebook**

Prathyusha Senthil Kumar

Social media's explosive growth brings with it a variety of societal risks ranging from severely harmful issues such as dangerous organizations and child sexual exploitation to moderately harmful content like displays of aggression, borderline nudity to benign or distasteful contents like gross videos and baity content. In recent times, the multitude and magnitude of these harms is being further exacerbated with the advent of generative AI [5]. Meta is committed to ensuring that Facebook is a place where people feel empowered to communicate and we take our role seriously in keeping abuse off the platform [7]. In this talk, I will describe practical challenges and lessons learned from tackling bad experiences for users on Facebook, particularly in the subjective, borderline and low quality spectrum of harms using state of the art, scalable machine learning approaches to content understanding, user behavior understanding and personalized ranking.

#### **Prod2Vec-Var: A Session Based Recommendation System with Enhanced Diversity**

Hacer Turgut, Tan Doruk Yetki, Ömür Bali, Tayfun Arda Yücel

Understanding user behavior and leveraging this information in recommendation systems pose challenges for websites lacking a login system or with limited logged-in users. This study introduces the Prod2Vec-Var recommendation system, a modified version of a session-based recommendation algorithm aimed at enhancing the performance of product

recommendation systems with a cold-start extension. The proposed model builds upon the original prod2vec algorithm, incorporating an additional step to improve the diversity of product recommendations. The project entails a well-designed data pipeline, effectively processing user actions to align them with the model, and implementing unique functions that expand the range of products capturing users' attention. Moreover, a straightforward yet effective cold-start model is developed to address newly added products that have not been viewed by users. The outcome of our project, namely product suggestions, is presented to users of cimri.com, one of iLab's affiliated companies, which attracts millions of daily visits, thereby enabling seamless access to desired products. Experimental results demonstrated the superior performance of our model compared to the other two different strategies in running recommendation on popular products, as evidenced by favorable R@1, R@5, R@10, and R@15 metrics. Concerning less popular products, we observed an improvement in our model's performance as the value of K increased, ultimately achieving optimal results in terms of R@15. Additionally, our cold-start model for new products substantiated the efficacy of our methodology, yielding the highest scores across R@5, R@10, and R@15 metrics.

#### **LAMM: Language Aware Active Learning for Multilingual Models**

Ze Ye, Dantong Liu, Kaushik Pavani, Sunny Dasgupta

In industrial settings, it is often necessary to achieve language-level accuracy targets. For example, Amazon business teams need to build multilingual product classifiers that operate accurately in all European languages. It is unacceptable for the accuracy of product classification to meet the target in one language (e.g., English), while falling below the target in other languages (e.g., Portuguese). To fix such issues, we propose Language Aware Active Learning for Multilingual Models (LAMM), an active learning strategy that enables a classifier to learn from a small amount of labeled data in a targeted manner to improve the accuracy of Low-resource languages (LRLs) with limited amounts of data for model training. Our empirical results on two open-source datasets and two proprietary product classification datasets demonstrate that LAMM is able to improve the LRL performance by 4%--11% when compared to strong baselines.

#### **Unleashing the Power of Large Language Models for Legal Applications**

Dell Zhang, Alina Petrova, Dietrich Trautmann, Frank Schilder

The use of Large Language Models (LLMs) is revolutionizing the legal industry. In this technical talk, we would like to explore the various use cases of LLMs in legal tasks, discuss the best practices, investigate the available resources, examine the ethical concerns, and suggest promising research directions.

### **SESSION: Workshop Proposals**

#### **MODIMO: Workshop on Multi-Omics Data Integration for Modelling Biological Systems**

Simone Avesani, Vincenzo Bonnici, Simone Pernice, Marco Beccuti, Rosalba Giugno

Multi-omics analysis aims at extracting previously uncovered biological knowledge by integrating information across multiple single-omic sources. Past approaches have focused on the simultaneous analysis of a small number of omic data sets. Current challenges face the problem of integrating multiple omic sources into a unified complex model, or of combining already available tools for two-by-two omics analyses and merging their outcomes. By doing so and leveraging integrated system-level knowledge, multi-omic approaches ought to enable the development of better qualitative and quantitative models for descriptive and predictive analyses. To move this area forward, new statistical and algorithmic frameworks are needed, for example for generalizing classical graph theory results to heterogeneous networks, and applying them to diverse problems such as drug repurposing or understanding the immune response to infections. Thus, in short, this workshop aims at investigating novel methodologies for providing crucial insights into multi-omics data management, integration, and analysis to enable biological discoveries. The workshop will be sponsored by the InfoLife CINI National Laboratory (<https://www.consortio-cini.it/index.php/en/>).

#### **MUWS'2023: The 2nd International Workshop on Multimodal Human Understanding for the Web and Social Media**

Gullal S. Cheema, Sherzod Hakimov, Marc A. Kastner, Noa Garcia

Multimodal human understanding and analysis is an emerging research area that cuts through several disciplines like Computer Vision, Natural Language Processing (NLP), Speech Processing, Human-Computer Interaction, and Multimedia. Several multimodal learning techniques have recently shown the benefit of combining multiple modalities in image-text, audio-visual and video representation learning and various downstream multimodal tasks. At the core, these methods focus on modelling the modalities and their complex interactions by using large amounts of data, different loss functions and deep

neural network architectures. However, for many Web and Social media applications, there is the need to model the human, including the understanding of human behaviour and perception. For this, it becomes important to consider interdisciplinary approaches, including social sciences, semiotics and psychology. The core is understanding various cross-modal relations, quantifying bias such as social biases, and the applicability of models to real-world problems. Interdisciplinary theories such as semiotics or gestalt psychology can provide additional insights and analysis on perceptual understanding through signs and symbols via multiple modalities. In general, these theories provide a compelling view of multimodality and perception that can further expand computational research and multimedia applications on the Web and Social media.

The theme of the MUWS workshop, multimodal human understanding, includes various interdisciplinary challenges related to social bias analyses, multimodal representation learning, detection of human impressions or sentiment, hate speech, sarcasm in multimodal data, multimodal rhetoric and semantics, and related topics. The MUWS workshop will be an interactive event and include keynotes by relevant experts, poster and demo sessions, research presentations and discussion.

### **[The First Workshop on Personalized Generative AI @ CIKM 2023: Personalization Meets Large Language Models](#)**

Zheng Chen, Ziyang Jiang, Fan Yang, Zhankui He, Yupeng Hou, Eunah Cho, Julian McAuley, Aram Galstyan, Xiaohua Hu, Jie Yang

The First Workshop on Personalized Generative AI1 aims to be a cornerstone event fostering innovation and collaboration in the dynamic field of personalized AI. Leveraging the potent capabilities of Large Language Models (LLMs) to enhance user experiences with tailored responses and recommendations, the workshop is designed to address a range of pressing challenges including knowledge gap bridging, hallucination mitigation, and efficiency optimization in handling extensive user profiles. As a nexus for academics and industry professionals, the event promises rich discussions on a plethora of topics such as the development and fine-tuning of foundational models, strategies for multi-modal personalization, and the imperative ethical and privacy considerations in LLM deployment. Through a curated series of keynote speeches, insightful panel discussions, and hands-on sessions, the workshop aspires to be a catalyst in the development of more precise, contextually relevant, and user-centric AI systems. It aims to foster a landscape where generative AI systems are not only responsive but also anticipatory of individual user needs, marking a significant stride in personalized experiences.

### **[Workshop on Enterprise Knowledge Graphs using Large Language Models](#)**

Rajeev Gupta, Srinath Srinivasa

Knowledge graphs are used for organizing and connecting individual entities to integrate the information extracted from different data sources. Typically, knowledge graphs are used to connect various real-world entities like persons, places, things, actions, etc. For the knowledge graphs created using the enterprise data, the knowledge graph entities can be of different types-static entities (e.g., people, projects), communication entities (e.g., emails, meetings, documents), derived entities (e.g., rules, definitions, entities from emails), etc. The graphs are used to connect these entities with enriched context (as edges and node attributes) and used for powering various search and recommendations applications.

With the advent of large language models, the whole lifecycle of knowledge graphs involving -information extraction, graph construction, application of graphs, querying knowledge graphs, using the graph for recommendations, etc., - is impacted. With large language models such as GPT, LLaMA, PALM, etc., entity and relationship extraction can be improved. Similarly, one can answer different types of queries with the help of LLMs which were very difficult without them. This workshop is about improving the enterprise knowledge graphs and its applications using large language models.

Enterprise graphs can be of different scopes-whether they contain data from individual users/customers, a sub-organization, or the whole enterprise. This workshop will also cover various privacy and access control related issues which are typical for any enterprise graph. These include privacy preserving federated learning, using LLMs to extract information from private data, querying the knowledge graph in a privacy preserving manner, etc.

### **[Workshop on Document Intelligence Understanding](#)**

Soyeon Caren Han, Yihao Ding, Siwen Luo, Josiah Poon, Hee-Guen Yoon, Zhe Huang, Paul Duuring, Eun-Jung Holden

Document understanding and information extraction include different tasks to understand a document and extract valuable information automatically. Recently, there has been a rising demand for developing document understanding among different domains, including business, law, and medicine, to boost the efficiency of work that is associated with a large number of documents. This workshop aims to bring together researchers and industry developers in the field of document intelligence

and understanding diverse document types to boost automatic document processing and understanding techniques. We also release a data challenge on the recently introduced document-level VQA dataset, PDFVQA. The PDFVQA challenge examines the model's structural and contextual understandings on the natural full document level of multiple consecutive document pages by including questions with a sequence of answers extracted from multi-pages of the full document. This task helps to boost the document understanding step from the single-page level to the full document level understanding.

### [The 3rd International Workshop on Mining and Learning in the Legal Domain](#)

Masoud Makrehchi, Dell Zhang, Alina Petrova, John Armour

The increasing accessibility of legal corpora and databases create opportunities to develop data-driven techniques and advanced tools that can facilitate a variety of tasks in the legal domain, such as legal search and research, legal document review and summary, legal contract drafting, and legal outcome prediction. Compared with other application domains, the legal domain is characterized by the huge scale of natural language text data, the high complexity of specialist knowledge, and the critical importance of ethical considerations. The MLLD workshop aims to bring together researchers and practitioners to share the latest research findings and innovative approaches in employing data mining, machine learning, information retrieval, and knowledge management techniques to transform the legal sector. Building upon the previous successes, the third edition of the MLLD workshop will emphasize the exploration of new research opportunities brought about by recent rapid advances in Large Language Models and Generative AI. We encourage submissions that intersect computer science and law, from both academia and industry, embodying the interdisciplinary spirit of CIKM.

### [Info-Wild: Knowledge Extraction and Management for Wildlife Conservation](#)

Prasenjit Mitra, Shreya Ghosh, Bistra Dilkina, Thomas Müller

Our primary objective is to explore and enhance AI's role for wildlife conservation, in brief, Nature Through the Lens of AI. It seeks to address crucial challenges related to data heterogeneity, scale integration, data privacy, mitigating biases, and decision-making under uncertainty. This workshop is centred around leveraging AI's prowess in deciphering complex spatio-temporal data patterns for wildlife conservation, thereby contributing significantly to the broader canvas of AI for social good. The workshop intends to create an interdisciplinary platform bringing together computer scientists, data scientists, geospatial experts, ecologists, and conservation practitioners, fostering collaboration and driving real-world impact. The program will include keynote speeches, panel discussions, and interactive sessions focusing on efficient knowledge extraction and management, remote sensing technologies, predictive modeling, species distribution modeling, habitat quality assessment, and human-wildlife conflict mitigation. With an emphasis on CIKM's primary interests, our aim is not only to enrich understanding of AI's symbiotic potential with ecology but also to utilize it to address pressing societal and environmental challenges.

### [IoTBDH-2023: The 5th International Workshop on Internet of Things of Big Data for Healthcare](#)

Jun Qi, Hongqing Yu, Po Yang, Yun Yang, Zhibo Pang

Internet of Things (IoT) enabled technology has rapidly and efficiently facilitate healthcare diagnose and treatment with low-cost and lightweight devices. Big data generated from IoT offers valuable and crucial information to guide decision-making, improve patient outcomes, and decrease healthcare costs, etc. The workshop is aiming to provide an opportunity for researchers and practitioners from both academia and industry to present the state-of-the-art research and applications in utilizing IoT and big data technology for healthcare by presenting efficient scientific and engineering solutions, addressing the needs and challenges for integration with new technologies, and providing visions for future research and development.

### [International Workshop on Edge-Cloud Intelligence \(ECI@CIKM2023\)](#)

Lianyong Qi, Xuyun Zhang, Xiaolong Xu, Wenwen Gong

With the exponential growth of mobile devices and the increasing demand for personalized and context-aware services, service-oriented applications play a crucial role in alleviating information overload and improving user experience. At the same time, Edge-Cloud Intelligence (ECI) provides a distributed computing paradigm that leverages the computational capabilities of edge servers to bring services closer to end-users, enabling low-latency and high-bandwidth data processing. The convergence of service applications and ECI has emerged as a promising way to enhance the efficiency and quality of service provisioning in today's dynamic and resource-constrained mobile environment. This call for papers invites researchers and practitioners to contribute to the exploration of the synergy between service applications and ECI infrastructure.

**Workshop On Large Language Models' Interpretability and Trustworthiness (LLMIT)**

Tulika Saha, Debasis Ganguly, Sriparna Saha, Prasenjit Mitra

Large language models (LLMs), when scaled from millions to billions of parameters, have been demonstrated to exhibit the so-called 'emergence' effect, in that they are not only able to produce semantically correct and coherent text, but are also able to adapt themselves surprisingly well with small changes in contexts supplied as inputs (commonly called prompts). Despite producing semantically coherent and potentially relevant text for a given context, LLMs are vulnerable to yield incorrect information. This misinformation generation, or the so-called hallucination problem of an LLM, gets worse when an adversary manipulates the prompts to their own advantage, e.g., generating false propaganda to disrupt communal harmony, generating false information to trap consumers with target consumables etc. Not only does the consumption of an LLM-generated hallucinated content by humans pose societal threats, such misinformation, when used as prompts, may lead to detrimental effects for in-context learning (also known as few-shot prompt learning). With reference to the above-mentioned problems of LLM usage, we argue that it is necessary to foster research on topics related to not only identifying misinformation from LLM-generated content, but also to mitigate the propagation effects of this generated misinformation on downstream predictive tasks thus leading to more robust and effective leveraging in-context learning.

**Anomaly and Novelty detection for Satellite and Drone systems (ANSND '23)**

Shahroz Tariq, Daewon Chung, Simon Woo, Youjin Shin

In recent times, there has been a notable surge in the amount of vision and sensing/time-series data obtained from drones and satellites. This data can be utilized in various fields, such as precision agriculture, disaster management, environmental monitoring, and others. However, the analysis of such data poses significant challenges due to its complexity, heterogeneity, and scale. Furthermore, it is critical to identify anomalies and maintain/monitor the health of drones and satellite systems to enable the aforementioned applications and sciences. This workshop presents an excellent opportunity to explore solutions that specifically target the detection of anomalies and novel occurrences in drones and satellite systems and their data. For more information, visit our website at <https://sites.google.com/view/ansd23>

**Knowledge-driven Analytics and Systems Impacting Human Quality of Life- Neurosymbolic AI, Explainable AI and Beyond**

Arijit Ukil, Joao Gama, Antonio J. Jara, Leandro Marin

The management of knowledge-driven artificial intelligence technologies is essential in order to evaluate their impact on human life and society. Social networks and tech use can have a negative impact on us physically, emotionally, socially and mentally. On the other hand, intelligent systems can have a positive effect on people's lives. Currently, we are witnessing the power of large language models (LLMs) like chatGPT and its influence towards the society. The objective of the workshop is to contribute to the advancement of intelligent technologies designed to address the human condition. This could include precise and personalized medicine, better care for elderly people, reducing private data leaks, using AI to manage resources better, using AI to predict risks, augmenting human capabilities, and more. The workshop's objective is to present research findings and perspectives that demonstrate how knowledge-enabled technologies and applications improve human well-being. This workshop indeed focuses on the impacts at different granularity levels made by Artificial Intelligence (AI) research on the micro granular level, where the daily or regular functioning of human life is affected, and also the macro granulate level, where the long-term or far-future effects of artificial intelligence on people's lives and the human society could be pretty high. In conclusion, this workshop explores how AI research can potentially address the most pressing challenges facing modern societies, and how knowledge management can potentially contribute to these solutions.

**The 1st Workshop on Recommendation with Generative Models**

Wenjie Wang, Yong Liu, Yang Zhang, Weiwen Liu, Fuli Feng, Xiangnan He, Aixin Sun

The boom of generative models has paved the way for significant advances in recommender systems. For instance, pre-trained generative models offer unprecedented opportunities to improve recommender algorithms for user modeling. This workshop aims to provide a platform for researchers to actively explore and share innovative ideas on integrating generative models into recommender systems, mainly focusing on five key aspects: (i) enhancing recommender algorithms, (ii) generating personalized content in some scenarios such as micro-videos, (iii) changes in the user-system interaction paradigm, (iv) boosting trustworthiness checks, and (v) evaluation methodologies of generative recommendation. With the rapid development of generative models, a growing number of studies along the above directions are emerging, revealing the timeliness and necessity of this workshop. The related research will bring novel features to recommender systems and



contribute to new tasks and technologies in both academia and industry. In the long run, this research direction might revolutionize the traditional recommender paradigm and lead to the maturation of next-generation recommender systems.

#### **DRL4IR: 4th Workshop on Deep Reinforcement Learning for Information Retrieval**

Xin Xin, Xiangyu Zhao, Jin Huang, Weinan Zhang, Li Zhao, Dawei Yin, Grace Hui Yang

IR is one of the most important fields to help users find relevant information. The interaction between IR systems and users can be naturally formulated as a decision-making problem. In the last decade, deep reinforcement learning (DRL) has become a promising direction to utilize the high model capacity of deep learning to improve long-term gains. On the one hand, there have been emerging research works focusing on leveraging DRL for IR tasks while the fundamental information theory under DRL settings, the principle of RL methods for IR tasks, or the experimental evaluation protocols of DRL-based IR systems, has not been deeply investigated. On the other hand, the emerging ChatGPT also provides new insights and challenges for DRL-based IR.

Therefore, we propose the fourth DRL4IR workshop<sup>1</sup> at CIKM 2023, which provides a venue for both academia researchers and industry practitioners to present the recent advances of DRL-based IR system, to foster novel research, interesting findings, and new applications of DRL for IR. We will pay more attention to fundamental research topics and recent application advances such as ChatGPT, with an expectation of over 300 workshop participants.

#### **Knowledge-enhanced Artificial Intelligence in Drug Discovery (KAIDD)**

Qingpeng Zhang, Jiannan Yang

Artificial Intelligence (AI) in drug discovery is a rapidly evolving field that combines computational methods with biological knowledge and applications. Traditionally, the process of developing a new drug has been time-consuming and expensive, spanning several years and costing billions of dollars. The emergence of AI technologies offers the potential to significantly reduce both the timeline and cost involved in this critical endeavour. However, it is crucial to acknowledge that AI applications in pharmacy and drug discovery require a high degree of interpretability and transparency. The integration of domain knowledge into AI models becomes paramount to ensure the reliability and trustworthiness of the generated results. In light of these considerations, we propose a workshop on "Knowledge-enhanced Artificial Intelligence in Drug Discovery (KAIDD)." This workshop aims to explore the profound impact of incorporating various knowledge databases into the development of explainable AI models for drug discovery. Participants will have the opportunity to delve into cutting-edge research, methodologies, and practical applications that leverage the fusion of AI techniques with domain-specific knowledge. Authors of accepted papers will have the opportunity to submit extended versions of their work for a full-paper review process and potential publication in Philosophical Transactions of the Royal Society B.

### **SESSION: Resource Papers**

#### **A Test Collection of Synthetic Documents for Training Rankers: ChatGPT vs. Human Experts**

Arian Askari, Mohammad Aliannejadi, Evangelos Kanoulas, Suzan Verberne

In this resource paper, we investigate the usefulness of generative Large Language Models (LLMs) in generating training data for cross-encoder re-rankers in a novel direction: generating synthetic documents instead of synthetic queries. We introduce a new dataset, ChatGPT-RetrievalQA, and compare the effectiveness of strong models fine-tuned on both LLM-generated and human-generated data. We build ChatGPT-RetrievalQA based on an existing dataset, human ChatGPT Comparison Corpus (HC3), consisting of public question collections with human responses and answers from ChatGPT. We fine-tune a range of cross-encoder re-rankers on either human-generated or ChatGPT-generated data. Our evaluation on MS MARCO DEV, TREC DL'19, and TREC DL'20 demonstrates that cross-encoder re-ranking models trained on LLM-generated responses are significantly more effective for out-of-domain re-ranking than those trained on human responses. For in-domain re-ranking, the human-trained re-rankers outperform the LLM-trained re-rankers. Our novel findings suggest that generative LLMs have high potential in generating training data for neural retrieval models and can be used to augment training data, especially in domains with smaller amounts of labeled data. We believe that our dataset, ChatGPT-RetrievalQA, presents various opportunities for analyzing and improving rankers with human and synthetic data. We release our data, code, and model checkpoints for future work.

#### **Ontology Enrichment from Texts: A Biomedical Dataset for Concept Discovery and Placement**

Hang Dong, Jiaoyan Chen, Yuan He, Ian Horrocks

Mentions of new concepts appear regularly in texts and require automated approaches to harvest and place them into Knowledge Bases (KB), e.g., ontologies and taxonomies. Existing datasets suffer from three issues, (i) mostly assuming that a new concept is pre-discovered and cannot support out-of-KB mention discovery; (ii) only using the concept label as the input along with the KB and thus lacking the contexts of a concept label; and (iii) mostly focusing on concept placement w.r.t a taxonomy of atomic concepts, instead of complex concepts, i.e., with logical operators. To address these issues, we propose a new benchmark, adapting MedMentions dataset (PubMed abstracts) with SNOMED CT versions in 2014 and 2017 under the Diseases sub-category and the broader categories of Clinical finding, Procedure, and Pharmaceutical / biologic product. We provide usage on the evaluation with the dataset for out-of-KB mention discovery and concept placement, adapting recent Large Language Model based methods.

#### **FlaCGEC: A Chinese Grammatical Error Correction Dataset with Fine-grained Linguistic Annotation**

Hanyue Du, Yike Zhao, Qingyuan Tian, Jiani Wang, Lei Wang, Yunshi Lan, Xuesong Lu

Chinese Grammatical Error Correction (CGEC) has been attracting growing attention from researchers recently. In spite of the fact that multiple CGEC datasets have been developed to support the research, these datasets lack the ability to provide a deep linguistic topology of grammar errors, which is critical for interpreting and diagnosing CGEC approaches. To address this limitation, we introduce FlaCGEC, which is a new CGEC dataset featured with fine-grained linguistic annotation. Specifically, we collect raw corpus from the linguistic schema defined by Chinese language experts, conduct edits on sentences via rules, and refine generated samples manually, which results in 10k sentences with 78 instantiated grammar points and 3 types of edits. We evaluate various cutting-edge CGEC methods on the proposed FlaCGEC dataset and their unremarkable results indicate that this dataset is challenging in covering a large range of grammatical errors. In addition, we also treat FlaCGEC as a diagnostic dataset for testing generalization skills and conduct a thorough evaluation of existing CGEC models.

#### **Towards Filling the Gap in Conversational Search: From Passage Retrieval to Conversational Response Generation**

Weronika Lajewska, Krisztian Balog

Research on conversational search has so far mostly focused on query rewriting and multi-stage passage retrieval. However, synthesizing the top retrieved passages into a complete, relevant, and concise response is still an open challenge. Having snippet-level annotations of relevant passages would enable both (1) the training of response generation models that are able to ground answers in actual statements and (2) automatic evaluation of the generated responses in terms of completeness. In this paper, we address the problem of collecting high-quality snippet-level answer annotations for two of the TREC Conversational Assistance track datasets. To ensure quality, we first perform a preliminary annotation study, employing different task designs, crowdsourcing platforms, and workers with different qualifications. Based on the outcomes of this study, we refine our annotation protocol before proceeding with the full-scale data collection to gather annotations for 1.8k question-paragraph pairs. The process of collecting data at this magnitude also led to multiple insights about the problem that can inform the design of future response-generation methods.

#### **PGB: A PubMed Graph Benchmark for Heterogeneous Network Representation Learning**

Eric W. Lee, Joyce C. Ho

There has been rapid growth in biomedical literature, yet capturing the heterogeneity of the bibliographic information of these articles remains relatively understudied. Graph neural networks have gained popularity, however, they may not fully capture the information available in the PubMed database, a biomedical literature repository containing over 33 million articles. We introduce PubMed Graph Benchmark (PGB), a new benchmark dataset for evaluating heterogeneous graph representations. PGB is one of the largest heterogeneous networks to date and aggregates the rich metadata into a unified source including abstract, authors, citations, keywords, and the associated keyword hierarchy. The benchmark contains an evaluation task of 21 systematic review topics, an essential knowledge translation tool.

#### **MUSER: A Multi-View Similar Case Retrieval Dataset**

Qingquan Li, Yiran Hu, Feng Yao, Chaojun Xiao, Zhiyuan Liu, Maosong Sun, Weixing Shen

Similar case retrieval (SCR) is a representative legal AI application that plays a pivotal role in promoting judicial fairness. However, existing SCR datasets only focus on the fact description section when judging the similarity between cases, ignoring other valuable sections (e.g., the court's opinion) that can provide insightful reasoning process behind. Furthermore, the case similarities are typically measured solely by the textual semantics of the fact descriptions, which may fail to capture the full complexity of legal cases from the perspective of legal knowledge. In this work, we present MUSER, a similar case

retrieval dataset based on multi-view similarity measurement and comprehensive legal element with sentence-level legal element annotations. Specifically, we select three perspectives (legal fact, dispute focus, and law statutory) and build a comprehensive and structured label schema of legal elements for each of them, to enable accurate and knowledgeable evaluation of case similarities. The constructed dataset originates from Chinese civil cases and contains 100 query cases and 4,024 candidate cases. We implement several text classification algorithms for legal element prediction and various retrieval methods for retrieving similar cases on MUSER. The experimental results indicate that incorporating legal elements can benefit the performance of SCR models, but further efforts are still required to address the remaining challenges posed by MUSER. The source code and dataset are released at <https://github.com/THULawtech/MUSER>.

#### **CTCam: Enhancing Transportation Evaluation through Fusion of Cellular Traffic and Camera-Based Vehicle Flows**

ChungYi Lin, Shen-Lung Tung, Hung-Ting Su, Winston H. Hsu

Traffic prediction utility often faces infrastructural limitations, which restrict its coverage. To overcome this challenge, we present Geographical Cellular Traffic (GCT) flow that leverages cellular network data as a new source for transportation evaluation. The broad coverage of cellular networks allows GCT flow to capture various mobile user activities across regions, aiding city authorities in resource management through precise predictions. Acknowledging the complexity arising from the diversity of mobile users in GCT flow, we supplement it with camera-based vehicle flow data from limited deployments and verify their spatio-temporal attributes and correlations through extensive data analysis. Our two-stage fusion approach integrates these multi-source data, addressing their coverage and magnitude discrepancies, thereby enhancing the prediction of GCT flow for accurate transportation evaluation. Overall, we propose novel uses of telecom data in transportation and verify its effectiveness in multi-source fusion with vision-based data.

#### **Datasets and Interfaces for Benchmarking Heterogeneous Graph Neural Networks**

Yijian Liu, Hongyi Zhang, Cheng Yang, Ao Li, Yugang Ji, Luhao Zhang, Tao Li, Jinyu Yang, Tianyu Zhao, Juan Yang, Hai Huang, Chuan Shi

In recent years, Heterogeneous Graph Neural Networks (HGNNs) have gained increasing attention due to their excellent performance in applications. However, the lack of high-quality benchmarks in new fields has become a critical limitation for developing and applying HGNNs. To accommodate the urgent need for emerging fields and the advancement of HGNNs, we present two large-scale, real-world, and challenging heterogeneous graph datasets from real scenarios: risk commodity detection and takeout recommendation. Meanwhile, we establish standard benchmark interfaces that provide over 40 heterogeneous graph datasets. We provide initial data split, unified evaluation metrics, and baseline results for future work, making it fair and handy to explore state-of-the-art HGNNs. Our interfaces also offer a comprehensive toolkit to research the characteristics of graph datasets. The above new datasets are publicly available on <https://zenodo.org/communities/hgd>, and the interface codes are available at <https://github.com/BUPT-GAMMA/hgbi>.

#### **ContributionSum: Generating Disentangled Contributions for Scientific Papers**

Meng-Huan Liu, An-Zi Yen, Hen-Hsen Huang, Hsin-Hsi Chen

Contributions are essentially the core of every scientific research, highlighting their key values to the academic community. Systems that are capable of identifying the contributions from scientific papers precisely and organizing them into well-structured summaries can facilitate both text processing and human comprehension. In this paper, we present ContributionSum, a dataset consisting of 24K computer science papers with contributions explicitly listed by the authors, which are further classified into different contribution types based on a newly-proposed annotation scheme. In addition, we study the task of generating disentangled contributions that summarize the values of scientific papers into key points. We propose a fine-grained post-training strategy tailored to our task and leverage salient information of different contribution types in the papers. To assess the coherency and coverage of each contribution aspect, we perform summary-level and contribution-level evaluations for our task. Experimental results show that our method improves upon mainstream baselines.

#### **ClinicalRisk: A New Therapy-related Clinical Trial Dataset for Predicting Trial Status and Failure Reasons**

Junyu Luo, Zhi Qiao, Lucas Glass, Cao Xiao, Fenglong Ma

Clinical trials aim to study new tests and evaluate their effects on human health outcomes, which has a huge market size. However, carrying out clinical trials is expensive and time-consuming and often ends in no results. It will revolutionize clinical practice if we can develop an effective model to automatically estimate the status of a clinical trial and find out possible failure reasons. However, it is challenging to develop such a model because of the lack of a benchmark dataset. To address these challenges, in this paper, we first build a new dataset by extracting the publicly available clinical trial reports from

ClinicalTrials.gov. The associated status of each report is treated as the status label. To analyze the failure reasons, domain experts help us manually annotate each failed report based on the description associated with it. More importantly, we examine several state-of-the-art text classification baselines on this task and find out that the unique format of the clinical trial protocols plays an essential role in affecting prediction accuracy, demonstrating the need for specially designed clinical trial classification models.

#### [Commonsense Temporal Action Knowledge \(CoTAK\) Dataset](#)

Steven Lynden, Hailemariam Mehari Yohannes, Kyoung-Sook Kim, Adam Jatowt, Akiyoshi Matono, Hai-Tao Yu, Xin Liu, Yijun Duan

This paper presents a publicly available, large-scale dataset resource, CoTAK (COMmonsense Temporal Action Knowledge) consisting of short descriptions of action-describing sentences manually annotated with temporal commonsense knowledge. The dataset consists of over 300K instructional sentences extracted from WikiHow, which are annotated with commonsense knowledge-based temporal labels indicating implicitly understood information about the actions described by the sentences, including approximately how long an action takes to perform and approximately how long its effects last for. For short duration actions labeled as taking seconds or minutes, which would be of relevance to automated task planning, e.g. in robotics applications, the dataset also provides scalar values to accurately label the temporal durations of how long actions take to perform. Experimental results are presented demonstrating that state-of-the-art machine learning techniques such as fine-tuning of large language models are effective in making predictions of commonsense temporal knowledge using the dataset, with up to 80% accuracy, showing the high utility and promising impact of the constructed resource and its applicability towards generating commonsense temporal knowledge relevant to various

#### [Anserini Gets Dense Retrieval: Integration of Lucene's HNSW Indexes](#)

Xueguang Ma, Tommaso Teofili, Jimmy Lin

Anserini is a Lucene-based toolkit for reproducible information retrieval research in Java that has been gaining traction in the community. It provides retrieval capabilities for both "traditional" bag-of-words retrieval models such as BM25 as well as retrieval using learned sparse representations such as SPLADE. With Pyserini, which provides a Python interface to Anserini, users gain access to both sparse and dense retrieval models, as Pyserini implements bindings to the Faiss vector search library alongside Lucene inverted indexes in a uniform, consistent interface. Nevertheless, hybrid fusion techniques that integrate sparse and dense retrieval models need to stitch together results from two completely different "software stacks", which creates unnecessary complexities and inefficiencies. However, the introduction of HNSW indexes for dense vector search in Lucene promises the integration of both dense and sparse retrieval within a single software framework. We explore exactly this integration in the context of Anserini. Experiments on the MS MARCO passage and BEIR datasets show that our Anserini HNSW integration supports (reasonably) effective and (reasonably) efficient approximate nearest neighbor search for dense retrieval models, using only Lucene.

#### [ThyExp: An explainable AI-assisted Decision Making Toolkit for Thyroid Nodule Diagnosis based on Ultra-sound Images](#)

Jamie Morris, Zehao Liu, Huizhi Liang, Sidhartha Nagala, Xia Hong

Radiologists have an important task of diagnosing thyroid nodules present in ultra sound images. Although reporting systems exist to aid in the diagnosis process, these systems do not provide explanations about the diagnosis results. We present ThyExp -- a web based toolkit for it use by medical professionals, allowing for accurate diagnosis with explanations of thyroid nodules present in ultrasound images utilising artificial intelligence models. The proposed web-based toolkit can be easily incorporated into current medical workflows, and allows medical professionals to have the confidence of a highly accurate machine learning model with explanations to provide supplementary diagnosis data. The solution provides classification results with their probability accuracy, as well as the explanations in the form of presenting the key features or characteristics that contribute to the classification results. The experiments conducted on a real-world UK NHS hospital patient dataset demonstrate the effectiveness of the proposed approach. This toolkit can improve the trust of medical professional to understand the confidence of the model in its predictions. This toolkit can improve the trust of medical professionals in understanding the models reasoning behind its predictions.

#### [RePair: An Extensible Toolkit to Generate Large-Scale Datasets for Query Refinement via Transformers](#)

Yogeswar Lakshmi Narayanan, Hossein Fani

Query refinement is the process of transforming users' queries into newrefined versions without semantic drift to enhance the relevance of search results. Prior query refiners were benchmarked on web query logs followingweak assumptions that

users' input queries within a search session are about a single topic and improve gradually, which is not necessarily accurate in practice. In this paper, we contribute RePair, an open-source configurable toolkit to generate large-scale gold-standard benchmark datasets whose pairs of (original query, refined versions) are almost surely guaranteed to be in the same semantic context. RePair takes a dataset of queries and their relevance judgements (e.g., msmarco or aol), a sparse or dense retrieval method (e.g., bm25 or colbert), and an evaluation metric (e.g., map or mrr), and outputs refined versions of queries, each of which with the relevance improvement guarantees under the retrieval method in terms of the evaluation metric. RePair benefits from text-to-text-transfer-transformer (t5) to generate gold-standard datasets for any input query sets and is designed with extensibility in mind. Out of the box, RePair includes gold-standard datasets for aol and msmarco passage as well as benchmark results of state-of-the-art supervised query suggestion methods on the generated datasets at <https://github.com/fani-lab/RePair>.

#### **KID34K: A Dataset for Online Identity Card Fraud Detection**

Eun-Ju Park, Seung-Yeon Back, Jeongho Kim, Simon S. Woo

Though digital financial systems have provided users with convenient and accessible services, such as supporting banking or payment services anywhere, it is necessary to have robust security to protect against identity misuse. Thus, online digital identity (ID) verification plays a crucial role in securing financial services on mobile platforms. One of the most widely employed techniques for digital ID verification is that mobile applications request users to take and upload a picture of their own ID cards. However, this approach has vulnerabilities where someone takes pictures of the ID cards belonging to another person displayed on a screen, or printed on paper to be verified as the ID card owner. To mitigate the risks associated with fraudulent ID card verification, we present a novel dataset for classifying cases where the ID card images that users upload to the verification system are genuine or digitally represented. Our dataset is replicas designed to resemble real ID cards, making it available while avoiding privacy issues. Through extensive experiments, we demonstrate that our dataset is effective for detecting digitally represented ID card images, not only in our replica dataset but also in the dataset consisting of real ID cards. Our dataset is available at [https://github.com/DASH-Lab/idcard\\_fraud\\_detection](https://github.com/DASH-Lab/idcard_fraud_detection).

#### **ITA-ELECTION-2022: A Multi-Platform Dataset of Social Media Conversations Around the 2022 Italian General Election**

Francesco Pierri, Geng Liu, Stefano Ceri

Online social media play a major role in shaping public discourse and opinion, especially during political events. We present the first public multi-platform dataset of Italian-language political conversations, focused on the 2022 Italian general election taking place on September 25th. Leveraging public APIs and a keyword-based search, we collected millions of posts published by users, pages and groups on Facebook, Instagram and Twitter, along with metadata of TikTok and YouTube videos shared on these platforms, over a period of four months. We augmented the dataset with a collection of political ads sponsored on Meta platforms, and a list of social media handles associated with political representatives. Our data resource will allow researchers and academics to further our understanding of the role of social media in the democratic process.

#### **GOAL: A Challenging Knowledge-grounded Video Captioning Benchmark for Real-time Soccer Commentary Generation**

Ji Qi, Jifan Yu, Teng Tu, Kunyu Gao, Yifan Xu, Xinyu Guan, Xiaozhi Wang, Bin Xu, Lei Hou, Juanzi Li, Jie Tang

Despite the recent emergence of video captioning models, how to generate vivid, fine-grained video descriptions based on the background knowledge (i.e., long and informative commentary about the domain-specific scenes with appropriate reasoning) is still far from being solved, which however has great applications such as automatic sports narrative. Based on soccer game videos and synchronized commentary data, we present GOAL, a benchmark of over 8.9k soccer video clips, 22k sentences, and 42k knowledge triples for proposing a challenging new task setting as Knowledge-grounded Video Captioning (KGVC). We experimentally test existing state-of-the-art (SOTA) methods on this resource to demonstrate the future directions for improvement in this challenging task. We hope that our data resource (now available at <https://github.com/THU-KEG/goal>) can serve researchers and developers interested in knowledge-grounded cross-modal applications.

#### **OpenGDA: Graph Domain Adaptation Benchmark for Cross-network Learning**

Boshen Shi, Yongqing Wang, Fangda Guo, Jiangli Shao, Huawei Shen, Xueqi Cheng

Graph domain adaptation models are widely adopted in cross-network learning tasks to transfer labeling or structural knowledge. Currently, there mainly exist two limitations in evaluating graph domain adaptation models. On one side, they are primarily tested for the specific cross-network node classification task, leaving tasks at edge-level and graph-level largely under-explored. Moreover, they are primarily examined in limited scenarios, such as social networks or citation networks,



needing more validation in richer scenarios. As comprehensively assessing models could enhance model practicality in real-world applications, we propose a benchmark known as OpenGDA. It provides abundant pre-processed and unified datasets for different types of tasks (node, edge, graph). They originate from diverse scenarios, covering web information systems, urban systems and natural systems. Furthermore, it integrates state-of-the-art models with standardized and end-to-end pipelines. Overall, OpenGDA provides a user-friendly, scalable and reproducible benchmark for evaluating graph domain adaptation models. The benchmark experiments highlight the challenges of applying GDA models to real-world applications with consistent good performance, and they potentially provide insights to future research. As an emerging project, OpenGDA will be regularly updated with new datasets and models. It could be accessed from <https://github.com/Skyorca/OpenGDA>.

### **TEI2GO: A Multilingual Approach for Fast Temporal Expression Identification**

Hugo Sousa, Ricardo Campos, Alípio Jorge

Temporal expression identification is crucial for understanding texts written in natural language. Although highly effective systems such as HeidelTime exist, their limited runtime performance hampers adoption in large-scale applications and production environments. In this paper, we introduce the TEI2GO models, matching HeidelTime's effectiveness but with significantly improved runtime, supporting six languages, and achieving state-of-the-art results in four of them. To train the TEI2GO models, we used a combination of manually annotated reference corpus and developed "Professor HeidelTime", a comprehensive weakly labeled corpus of news texts annotated with HeidelTime. This corpus comprises a total of 138,069 documents (over six languages) with 1,050,921 temporal expressions, the largest open-source annotated dataset for temporal expression identification to date. By describing how the models were produced, we aim to encourage the research community to further explore, refine, and extend the set of models to additional languages and domains. Code, annotations, and models are openly available for community exploration and use. The models are conveniently on HuggingFace for seamless integration and application.

### **KuaiSAR: A Unified Search And Recommendation Dataset**

Zhongxiang Sun, Zihua Si, Xiaoxue Zang, Dewei Leng, Yanan Niu, Yang Song, Xiao Zhang, Jun Xu

The confluence of Search and Recommendation (S&R) services is vital to online services, including e-commerce and video platforms. The integration of S&R modeling is a highly intuitive approach adopted by industry practitioners. However, there is a noticeable lack of research conducted in this area within academia, primarily due to the absence of publicly available datasets. Consequently, a substantial gap has emerged between academia and industry regarding research endeavors in joint optimization using user behavior data from both S&R services. To bridge this gap, we introduce the first large-scale, real-world dataset KuaiSAR of integrated Search And Recommendation behaviors collected from Kuaishou, a leading short-video app in China with over 350 million daily active users. Previous research in this field has predominantly employed publicly available semi-synthetic datasets, with artificially fabricated search behaviors. Distinct from previous datasets, KuaiSAR contains genuine user behaviors, including the occurrence of each interaction within either search or recommendation service, and the users' transitions between the two services. This work aids in joint modeling of S&R, and utilizing search data for recommender systems (and recommendation data for search engines). Furthermore, due to the various feedback labels associated with user-video interactions, KuaiSAR also supports a broad range of tasks, including intent recommendation, multi-task learning, and modeling of long sequential multi-behavioral patterns. We believe this dataset will serve as a catalyst for innovative research and bridge the gap between academia and industry in understanding the S&R services in practical, real-world applications. The dataset is available at <https://ethan00si.github.io/KuaiSAR/>. The dataset is also shared at <https://zenodo.org/record/8181109>.

### **DynamicESG: A Dataset for Dynamically Unearthing ESG Ratings from News Articles**

Yu-Min Tseng, Chung-Chi Chen, Hen-Hsen Huang, Hsin-Hsi Chen

This paper introduces the DynamicESG dataset, a unique resource for dynamically extracting ESG ratings from news articles. The ESG rating, a novel metric employed annually to gauge a company's sustainability, relies heavily on corporate disclosure and other external information, especially news narratives. Our dataset, comprising a wide spectrum of news over a twelve-year span, annotates articles in accordance with MSCI ESG ratings methodology and SASB standards, with relevance to ESG issues. DynamicESG provides a comprehensive means of investigating the relationship between public discourse, ESG-related events, and subsequent ESG rating adjustments. We detail our data collection, curation, annotation procedure, and inter-rater agreement, ensuring high data quality and usability. Importantly, our dataset includes a temporal dimension, enabling the analysis of longitudinal trends in ESG ratings and their correlation with news coverage. Moreover, the dataset incorporates an opportunity/risk tendency, thus permitting analysis from diverse perspectives to discern if the

news is beneficial or detrimental to the company. We believe this dataset will serve as a valuable resource for researchers in fields such as corporate social responsibility, sustainable investing, machine learning, and natural language processing. Initial analysis using the dataset underscores its potential to facilitate new insights into the dynamics of ESG ratings and the influence of news media on these ratings.

#### **MDCC: A Multimodal Dynamic Dataset for Donation-based Crowdfunding Campaigns**

Xovee Xu, Jiayang Li, Fan Zhou

Crowdfunding platforms have become pivotal financial support avenues for diverse causes, yet the success rates are surprisingly low. Previous research has largely focused on reward-based crowdfunding, leaving donation-based platforms under-studied. In addition, the roles of multimodal data (e.g., textual descriptions and visual photos) and dynamic elements (e.g., sequences of donations, project updates, and comments) in influencing campaign success have been largely overlooked. This paper introduces MDCC, a Multimodal Dynamic dataset for donation-based Crowdfunding Campaigns, collected from 14,961 projects on GoFundMe, incorporates multimodal project information and captures project dynamics, thus providing a comprehensive tool for analyzing donation-based crowdfunding. The dataset is expected to inspire innovative methodologies and facilitate understanding of project success determinants. Our preliminary experiments demonstrate the significance of multimodal and dynamic crowdfunding data on predicting the success of donation-based projects.