Lecture Notes on Neural Information Retrieval

1 - Text Representations for Ranking

The problem of ranking, which involves learning a real-valued ranking function to induce an ordering over an instance space, is an important task in machine learning [2]. Recent advancements in neural networks, especially deep convolutional neural networks (CNNs), have offered the potential for multidimensional signal processing and improved the ability to automatically learn features from large data sets for classification tasks [3]. In order to tackle the challenges posed by the increasing amount of digital data for retrieval systems, a novel retrieval framework has been proposed, which consists of weighted subspace-based filtering and ranking components [4]. This framework addresses the difficulty of comparing semantics between low-level features and high-level concepts by applying multiple correspondence analysis (MCA) to explore relationships between feature categories and concept classes [4]. Consequently, these techniques help in efficiently filtering and ranking results in multimedia retrieval applications. Using advanced algorithms, such as kernel-based ranking algorithms that perform regularization in reproducing kernel Hilbert space, generalization bounds can be derived for ranking tasks based on the notion of algorithmic stability [2]. By considering algorithmic stability and the advancements in neural networks like CNNs, one can improve the ranking effectiveness and learn better text representations for various applications.

1.1 - BOW Encodings

Bag-of-Words (BoW) encodings represent documents as fixed-sized vectors, where each dimension corresponds to the frequency of a word in the vocabulary, offering an efficient way to represent and compare documents for information retrieval tasks [7]. However, traditional BoW representations do not capture the semantic relationships in the text and usually have limited performance in ranking tasks compared to more advanced methods using dense representations [1,4].

Recent research has sought ways to improve the BoW model performance with additional information and optimizations. In [2], the authors propose SparTerm, a method that combines an importance predictor and a gating controller to generate term-based sparse representations based on semantic relationships with the input text. This approach maintains the interpretability and efficiency of BoW while enhancing ranking performance. Moreover, the Entropy Optimized Feature-Based Bag-of-Words (EO-BoW) proposed in [7] improves retrieval performance by optimizing the dictionary construction using an entropy-based criterion, providing better average precision and lowering encoding time and storage requirements.

On the other hand, the success of transformer-based models in dense representation learning has cast a shadow on BoW-based approaches [1]. For instance, pretrained transformers like BERT and its variants have demonstrated substantial effectiveness gains over BoW methods in text ranking tasks [4]. Despite the benefits of the dense representation models, they come with higher computational costs [1,4], motivating researchers to seek ways to bridge the performance gap between the two representations while maintaining BoW's efficiency advantages [2,7]. The ongoing research in this area suggests that there is still room for improvements in BoW and term-based representations while keeping their inherent advantages in computational cost and interpretability.

1.2 - LTR Features

Learning to Rank (LTR) is a crucial technique employed in information retrieval and search engine ranking tasks [7]. The effectiveness of LTR models is highly influenced by the selection of relevant features and the proper usage of different sources of information [3, 5, 7]. Generally, LTR features can be categorized into query-dependent or dynamic features (e.g., BM25 score) [4], query-level features (e.g., number of words in query) [4], and document-dependent features (e.g., term frequency, spam score) [4]. It is important to pay attention to feature selection for LTR, as features in information retrieval are not always independent from each other, and reducing the feature subset can lead to improved performance [3].

One approach for feature selection in the context of LTR is through hierarchical feature selection, which includes a two-phase process for determining ranking functions [3]. Another approach leverages Learning-to-Rank with BERT, which combines ranking losses with BERT representations for passage ranking [2]. This method demonstrated the effectiveness of fine-tuning BERT representations for query-document pairs within a ranking framework [2]. Other feature selection algorithms specifically designed for LTR also exist, such as the ones proposed by [5], which have shown to outperform well-known state-of-the-art competitors. In summary, selecting the right set of features and exploiting all available sources of information are essential elements for optimizing LTR models and improving search engine rankings.

1.3 - Word Embeddings

Word embeddings have emerged as a powerful tool in natural language processing (NLP), offering a compact and effective representation of words that capture their semantics and syntactic properties. Several methods have been proposed to generate word embeddings, such as the Neural Network Language Model [1], Sparse Coding Approach [1], and the popular Word2Vec [3]. Incorporating semantic knowledge can significantly improve the quality of the learned embeddings, as demonstrated by the work on learning semantic word embeddings (SWE) based on ordinal knowledge constraints [2]. These improvements can be seen in various NLP tasks, including word similarity measure, sentence completion, name entity recognition, and TOEFL synonym selection [2].

Moreover, word embeddings have been successfully applied in a variety of NLP applications such as document and query representation for information retrieval [4], sentiment-aware word embedding for emotion classification [6], and integrating extra knowledge into word embedding for biomedical NLP tasks [7]. The development of Dual Word Embeddings [8] has enhanced the document ranking process by utilizing both input and output embeddings in Word2Vec, offering a more suitable similarity measure for document ranking. Additionally, simple word embedding-based models with associated pooling mechanisms have demonstrated their potential as strong baselines for text representation learning and highlighted the computational expressiveness tradeoff while selecting compositional functions for distinct NLP problems [5].

2 - Interaction-focused Systems

Interaction-focused systems have gained significant popularity and interest in recent years due to the increasing need for personalized and efficient information processing. Recommender systems are one example of such technology, where the aim is to provide personalized suggestions based on users' preferences and behavior [1,3]. These systems rely on algorithms that gather and analyze data to optimize the user experience, making them highly relevant in the age of information overload. An important aspect of interaction-focused systems is managing the continuous and rapidly changing data streams that characterize modern applications [8]. These data streams pose unique challenges in terms of query processing, algorithmic design, and understanding the underlying structure and relationships within the data.

Another key aspect of interaction-focused systems is the development and integration of real-time systems, which are essential for distributed embedded applications that require deterministic, composable, and fault-tolerant designs [7]. Real-time systems must meet stringent requirements in terms of timing, stability, and accuracy to ensure that the desired outcomes are achieved in a timely and consistent manner.

Switching in systems and control is another important topic within the realm of interaction-focused systems [6]. Switching mechanisms help stabilize systems under arbitrary conditions and deal with the inherent complexities associated with sensor or actuator constraints and large modeling uncertainties. Additionally, the incorporation of deep learning systems, particularly in the area of automated whitebox testing, has opened new doors for improving the design, accuracy, and efficiency of interaction-focused applications [5].

In conclusion, the development of interaction-focused systems involves a multifaceted approach that combines insights from diverse fields such as recommender systems, data stream processing, real-time systems, switching mechanisms, and deep learning algorithms. Such systems hold the potential to revolutionize the way users interact with technology and process information in an increasingly interconnected world.

2.1 - Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a type of neural network that leverage convolution operations as a core component in their design, allowing them to efficiently process spatially structured data and extract local patterns [2]. These networks have found widespread success across numerous fields, including computer vision [5], natural language processing [3], and even molecular neural networks [7]. Advancements in Graph Convolutional Networks (GCNs) have further extended the power of CNNs by incorporating graph-structured data, enabling the propagation of information across local graph neighborhoods through iterative aggregation and feature transformation [4]. Such versatility makes CNNs ideal candidates for applications involving large datasets and complex feature extraction processes. Moreover, developments in neuromorphic photonic systems have led to Digital Electronic and Analog Photonic (DEAP) CNN hardware architectures, which hold the potential to be faster and more energy-efficient than current state-of-the-art GPUs [8]. Through such innovations, CNNs continue to augment our ability to process and understand complex datasets, setting new standards in various fields of research and application.

2.2 - Pre-trained Language Models

Pre-trained language models have demonstrated significant potential in various natural language processing tasks, transforming the way we approach problems in the field. Models such as GPT-2, GPT-3, BERT, and RoBERTa have led groundbreaking advancements [1,4,5]. In task-oriented dialogue systems, authors in [4] propose building on top of the TransferTransfo framework and the generative model pre-training to adapt these language models to generate task-specific responses. TOD-BERT [1] is an example of a pre-trained language model specifically aimed at task-oriented dialogues, demonstrating the adaptability of these models.

Additionally, pre-trained language models have proven effective in other domains, such as mental healthcare with MentalBERT [5] and recommendation systems with M6-Rec [2]. Pre-trained language models have also been employed for multilingual applications, as demonstrated in mLUKE [3], which highlights the power of multilingual models for cross-lingual transfer learning. Moreover, the development of Meta Language Models, such as METALM [8], showcases their potential as general-purpose interfaces, allowing users to interact with models using natural language and supporting multi-turn conversational interactions. Ultimately, pre-trained language models have transformed the landscape of NLP, demonstrating remarkable performance across a breadth of tasks and domains.

2.3 - Dealing with long texts

Approaches to handling long texts vary across different disciplines and applications. In the context of speaker verification, researchers have proposed algorithms to improve the robustness of embedding-based deep convolution neural networks (CNNs) for longer duration utterances [1]. Specifically, the discriminability of embeddings in speaker recognition is enhanced by reducing intra-speaker variation using center loss and increasing inter-speaker discrepancy with softmax loss [1]. When examining long texts in the context of ancient Near East archives, archaeologists have found that these archives were accumulated as long as they were considered useful, and they were organized using a classification system, giving insights into their application and utility [4]. Evaporation studies across different fields, such as oceanography and hydrology, involve varied approaches depending on the constraints of the system of interest [5]. For example, oceanographers focus on the aerodynamic approach when dealing with evaporation from large bodies of water, while hydrologists may use different methods such as water budgets or empirical approaches when evaluating long-term losses from heterogeneous basins [5]. These examples demonstrate the diverse methodologies that various disciplines employ to tackle the challenges associated with processing and analyzing long texts.

References

[1]-@Article{Duchi2012TheAO, author = {John C. Duchi and Lester W. Mackey and Michael I. Jordan}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {The Asymptotics of Ranking Algorithms}, volume = {abs/1204.1688}, year = {2012} }

[2]-@Article{Agarwal2009GeneralizationBF, author = {S. Agarwal and P. Niyogi}, booktitle = {Journal of machine learning research}, journal = {J. Mach. Learn. Res.}, pages = {441-474}, title = {Generalization Bounds for Ranking Algorithms via Algorithmic Stability}, volume = {10}, year = {2009} }

[3]-@Article{Gonzalez2018DeepCN, author = {Rafael C. Gonzalez}, booktitle = {IEEE Signal Processing Magazine}, journal = {IEEE Signal Processing Magazine}, pages = {79-87}, title = {Deep Convolutional Neural Networks [Lecture Notes]}, volume = {35}, year = {2018} }

[4]-@Article{Lin2011WeightedSF, author = {Lin Lin and Chao Chen and M. Shyu and Shu-Ching Chen}, booktitle = {IEEE Multimedia}, journal = {IEEE MultiMedia}, pages = {32-43}, title = {Weighted Subspace Filtering and Ranking Algorithms for Video Concept Retrieval}, volume = {18}, year = {2011} }

[5]-@Article{Baraniuk2007CompressiveS, author = {Richard Baraniuk}, booktitle = {IEEE Signal

Processing Magazine}, journal = {IEEE Signal Processing Magazine}, pages = {118-121}, title = {Compressive Sensing [Lecture Notes]}, volume = {24}, year = {2007} }

[6]-@Article{Kuo2017TheCA, author = {C.-C. Jay Kuo}, booktitle = {IEEE Signal Processing Magazine}, journal = {IEEE Signal Processing Magazine}, pages = {81-89}, title = {The CNN as a Guided Multilayer RECOS Transform [Lecture Notes]}, volume = {34}, year = {2017} }

[7]-@Article{Chatzivasileiadis2018LectureNO, author = {Spyros Chatzivasileiadis}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Lecture Notes on Optimal Power Flow (OPF)}, volume = {abs/1811.00943}, year = {2018} }

[8]-@Article{Lidar2019LectureNO, author = {Daniel A. Lidar}, journal = {arXiv: Quantum Physics}, title = {Lecture Notes on the Theory of Open Quantum Systems}, year = {2019} }

[1]-@Article{Lin2020PretrainedTF, author = {Jimmy J. Lin and Rodrigo Nogueira and Andrew Yates}, booktitle = {North American Chapter of the Association for Computational Linguistics}, journal = {Proceedings of the 14th ACM International Conference on Web Search and Data Mining}, title = {Pretrained Transformers for Text Ranking: BERT and Beyond}, year = {2020} }

[2]-@Article{Bai2020SparTermLT, author = {Yang Bai and Xiaoguang Li and Gang Wang and Chaoliang Zhang and Lifeng Shang and Jun Xu and Zhaowei Wang and Fangshan Wang and Qun Liu}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {SparTerm: Learning Term-based Sparse Representation for Fast Text Retrieval}, volume = {abs/2010.00768}, year = {2020} }

[3]-@Article{Wang2015IncorporatingLK, author = {Yan Wang and Zhiyuan Liu and Maosong Sun}, booktitle = {PLoS ONE}, journal = {PLoS ONE}, title = {Incorporating Linguistic Knowledge for Learning Distributed Word Representations}, volume = {10}, year = {2015} }

[4]-@Article{Zhuang2021FastPR, author = {Shengyao Zhuang and G. Zuccon}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Fast Passage Re-ranking with Contextualized Exact Term Matching and Efficient Passage Expansion}, volume = {abs/2108.08513}, year = {2021} }

[5]-@Article{Lin2020DistillingDR, author = {Sheng-Chieh Lin and Jheng-Hong Yang and Jimmy J. Lin}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Distilling Dense Representations for Ranking using Tightly-Coupled Teachers}, volume = {abs/2010.11386}, year = {2020} }

[6]-@Article{Paul2019RankingAS, author = {Debjit Paul and A. Frank}, booktitle = {North American Chapter of the Association for Computational Linguistics}, pages = {3671-3681}, title = {Ranking and Selecting Multi-Hop Knowledge Paths to Better Predict Human Needs}, year = {2019} }

[7]-@Article{Passalis2016EntropyOF, author = {N. Passalis and A. Tefas}, booktitle = {IEEE Transactions on Knowledge and Data Engineering}, journal = {IEEE Transactions on Knowledge and Data Engineering}, pages = {1664-1677}, title = {Entropy Optimized Feature-Based Bag-of-Words Representation for Information Retrieval}, volume = {28}, year = {2016} }

[8]-@Article{Lin2021InBatchNF, author = {Sheng-Chieh Lin and Jheng-Hong Yang and Jimmy J. Lin}, booktitle = {Workshop on Representation Learning for NLP}, pages = {163-173}, title = {In-Batch Negatives for Knowledge Distillation with Tightly-Coupled Teachers for Dense Retrieval}, year = {2021} }

[1]-@Book{Zamani2017NeuralRM, author = {Hamed Zamani and Bhaskar Mitra and Xia Song and Nick Craswell and Saurabh Tiwary}, booktitle = {Web Search and Data Mining}, journal = {Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining}, title = {Neural Ranking Models with Multiple Document Fields}, year = {2017} }

[2]-@Article{Han2020LearningtoRankWB, author = {Shuguang Han and Xuanhui Wang and Michael Bendersky and Marc Najork}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Learning-to-Rank with BERT in TF-Ranking}, volume = {abs/2004.08476}, year = {2020} }

[3]-@Article{Hua2010HierarchicalFS, author = {G. Hua and Min Zhang and Yiqun Liu and Shaoping Ma and Liyun Ru}, booktitle = {The Web Conference}, pages = {1113-1114}, title = {Hierarchical feature selection for ranking}, year = {2010} }

[4]-@Article{Mitra2017NeuralMF, author = {Bhaskar Mitra and Nick Craswell}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Neural Models for Information Retrieval}, volume = {abs/1705.01509}, year = {2017} }

[5]-@Article{Gigli2016FastFS, author = {Andrea Gigli and C. Lucchese and F. M. Nardini and R. Perego}, booktitle = {International Conference on the Theory of Information Retrieval}, journal =

{Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval}, title = {Fast Feature Selection for Learning to Rank}, year = {2016} }

[6]-@Article{Joachims2016UnbiasedLW, author = {T. Joachims and Adith Swaminathan and Tobias Schnabel}, booktitle = {Web Search and Data Mining}, journal = {Proceedings of the Tenth ACM International Conference on Web Search and Data Mining}, title = {Unbiased Learning-to-Rank with Biased Feedback}, year = {2016} }

[7]-@Article{Azarbonyad2020LearningTR, author = {H. Azarbonyad and M. Dehghani and M. Marx and J. Kamps}, booktitle = {Natural Language Engineering}, journal = {Natural Language Engineering}, pages = {89 - 111}, title = {Learning to rank for multi-label text classification: Combining different sources of information}, volume = {27}, year = {2020} }

[8]-@Article{Oosterhuis2018DifferentiableUO, author = {Harrie Oosterhuis and M. de Rijke}, booktitle = {International Conference on Information and Knowledge Management}, journal = {Proceedings of the 27th ACM International Conference on Information and Knowledge Management}, title = {Differentiable Unbiased Online Learning to Rank}, year = {2018} }

[1]-@Inproceedings{Li2018WordEF, author = {Yang Li and Tao Yang}, pages = {83-104}, title = {Word Embedding for Understanding Natural Language: A Survey}, year = {2018} }

[2]-@Article{LIU2015LearningSW, author = {QUAN LIU and Hui Jiang and Si Wei and Zhenhua Ling and Yu Hu}, booktitle = {Annual Meeting of the Association for Computational Linguistics}, pages = {1501-1511}, title = {Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints}, year = {2015} }

[3]-@Article{Ji2015WordRankLW, author = {Shihao Ji and Hyokun Yun and Pinar Yanardag and Shin Matsushima and S. Vishwanathan}, booktitle = {Conference on Empirical Methods in Natural Language Processing}, pages = {658-668}, title = {WordRank: Learning Word Embeddings via Robust Ranking}, year = {2015} }

[4]-@Article{Roy2016RepresentingDA, author = {Dwaipayan Roy and Debasis Ganguly and Mandar Mitra and G. Jones}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Representing Documents and Queries as Sets of Word Embedded Vectors for Information Retrieval}, volume = {abs/1606.07869}, year = {2016} }

[5]-@Article{Shen2018BaselineNM, author = {Dinghan Shen and Guoyin Wang and Wenlin Wang and Martin Renqiang Min and Qinliang Su and Yizhe Zhang and Chunyuan Li and Ricardo Henao and L. Carin}, booktitle = {Annual Meeting of the Association for Computational Linguistics}, pages = {440-450}, title = {Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms}, year = {2018} }

[6]-@Article{Mao2019SentimentAwareWE, author = {Xingliang Mao and Shuai Chang and Jinjing Shi and Fangfang Li and Ronghua Shi}, booktitle = {Applied Sciences}, journal = {Applied Sciences}, title = {Sentiment-Aware Word Embedding for Emotion Classification}, year = {2019} }

[7]-@Article{Ling2017IntegratingEK, author = {Yuan Ling and Yuan An and Mengwen Liu and Sadid A. Hasan and Ye-tian Fan and Xiaohua Hu}, booktitle = {IEEE International Joint Conference on Neural Network}, journal = {2017 International Joint Conference on Neural Networks (IJCNN)}, pages = {968-975}, title = {Integrating extra knowledge into word embedding models for biomedical NLP tasks}, year = {2017} }

[8]-@Article{Nalisnick2016ImprovingDR, author = {Eric T. Nalisnick and Bhaskar Mitra and Nick Craswell and R. Caruana}, booktitle = {The Web Conference}, journal = {Proceedings of the 25th International Conference Companion on World Wide Web}, title = {Improving Document Ranking with Dual Word Embeddings}, year = {2016} }

[1]-@Inproceedings{Ricci2011IntroductionTR, author = {F. Ricci and L. Rokach and Bracha Shapira}, booktitle = {Recommender Systems Handbook}, pages = {1-35}, title = {Introduction to Recommender Systems Handbook}, year = {2011} }

[2]-@Inproceedings{Cassandras1999IntroductionTD, author = {C. Cassandras and S. Lafortune}, booktitle = {The Kluwer International Series on Discrete Event Dynamic Systems}, pages = {1-823}, title = {Introduction to Discrete Event Systems}, volume = {11}, year = {1999} }

[3]-@Inproceedings{Ricci2015RecommenderSI, author = {F. Ricci and L. Rokach and Bracha Shapira}, booktitle = {Recommender Systems Handbook}, pages = {1-34}, title = {Recommender Systems: Introduction and Challenges}, year = {2015} }

[4]-@Article{Kriegeskorte2008RepresentationalSA, author = {N. Kriegeskorte and Marieke Mur and P. Bandettini}, booktitle = {Frontiers in Systems Neuroscience}, journal = {Frontiers in Systems Neuroscience}, title = {Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience}, volume = {2}, year = {2008} }

[5]-@Article{Pei2017DeepXploreAW, author = {Kexin Pei and Yinzhi Cao and Junfeng Yang and S. Jana}, booktitle = {Symposium on Operating Systems Principles}, journal = {Proceedings of the 26th Symposium on Operating Systems Principles}, title = {DeepXplore: Automated Whitebox Testing of Deep Learning Systems}, year = {2017} }

[6]-@Inproceedings{Liberzon2003SwitchingIS, author = {D. Liberzon}, booktitle = {Systems & Control: Foundations & Applications}, pages = {1-233}, title = {Switching in Systems and Control}, year = {2003} }

[7]-@Book{Kopetz2011RealTimeSD, author = {H. Kopetz}, booktitle = {Real-Time Systems Series}, journal = {Real-Time Systems}, title = {Real-Time Systems: Design Principles for Distributed Embedded Applications}, year = {2011} }

[8]-@Article{Babcock2002ModelsAI, author = {Brian Babcock and Shivnath Babu and Mayur Datar and R. Motwani and J. Widom}, booktitle = {ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems}, pages = {1-16}, title = {Models and issues in data stream systems}, year = {2002} }

[1]-@Article{Xu2018HowPA, author = {Keyulu Xu and Weihua Hu and J. Leskovec and S. Jegelka}, booktitle = {International Conference on Learning Representations}, journal = {ArXiv}, title = {How Powerful are Graph Neural Networks?}, volume = {abs/1810.00826}, year = {2018} }

[2]-@Article{Ketkar2021ConvolutionalNN, author = {Nikhil S. Ketkar}, booktitle = {Deep Learning with Python}, journal = {Deep Learning with Python}, title = {Convolutional Neural Networks}, year = {2021} }

[3]-@Article{Li2016MiningOS, author = {Qiudan Li and Zhipeng Jin and Can Wang and D. Zeng}, booktitle = {Knowledge-Based Systems}, journal = {Knowl. Based Syst.}, pages = {289-300}, title = {Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems}, volume = {107}, year = {2016} }

[4]-@Article{Ying2018GraphCN, author = {Rex Ying and Ruining He and Kaifeng Chen and Pong Eksombatchai and William L. Hamilton and J. Leskovec}, booktitle = {Knowledge Discovery and Data Mining}, journal = {Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining}, title = {Graph Convolutional Neural Networks for Web-Scale Recommender Systems}, year = {2018} }

[5]-@Article{Schirrmeister2017DeepLW, author = {R. Schirrmeister and J. T. Springenberg and L. Fiederer and M. Glasstetter and Katharina Eggensperger and M. Tangermann and F. Hutter and W. Burgard and T. Ball}, booktitle = {Human Brain Mapping}, journal = {Human Brain Mapping}, pages = {5391 - 5420}, title = {Deep learning with convolutional neural networks for EEG decoding and visualization}, volume = {38}, year = {2017} }

[6]-@Article{Choy20194DSC, author = {C. Choy and JunYoung Gwak and S. Savarese}, booktitle = {Computer Vision and Pattern Recognition}, journal = {2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)}, pages = {3070-3079}, title = {4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks}, year = {2019} }

[7]-@Article{Finzi2020GeneralizingCN, author = {Marc Finzi and S. Stanton and Pavel Izmailov and A. Wilson}, booktitle = {International Conference on Machine Learning}, pages = {3165-3176}, title = {Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data}, year = {2020} }

[8]-@Article{Bangari2019DigitalEA, author = {V. Bangari and B. Marquez and Heidi B. Miller and A. Tait and M. Nahmias and T. F. de Lima and Hsuan-Tung Peng and P. Prucnal and B. Shastri}, booktitle = {IEEE Journal of Selected Topics in Quantum Electronics}, journal = {IEEE Journal of Selected Topics in Quantum Electronics}, pages = {1-13}, title = {Digital Electronics and Analog Photonics for Convolutional Neural Networks (DEAP-CNNs)}, volume = {26}, year = {2019} }

[1]-@Article{Wu2020TODBERTPN, author = {Chien-Sheng Wu and S. Hoi and R. Socher and Caiming Xiong}, booktitle = {Conference on Empirical Methods in Natural Language Processing}, pages = {917-929}, title = {TOD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogue}, year = {2020} }

[2]-@Article{Cui2022M6RecGP, author = {Zeyu Cui and Jianxin Ma and Chang Zhou and Jingren Zhou and Hongxia Yang}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {M6-Rec: Generative Pretrained Language Models are Open-Ended Recommender Systems}, volume = {abs/2205.08084}, year = {2022} }

[3]-@Article{Ri2021mLUKETP, author = {Ryokan Ri and Ikuya Yamada and Yoshimasa Tsuruoka}, booktitle = {Annual Meeting of the Association for Computational Linguistics}, pages = {7316-7330}, title = {mLUKE: The Power of Entity Representations in Multilingual Pretrained Language Models}, year = {2021} }

[4]-@Article{Budzianowski2019HelloIG, author = {Paweł Budzianowski and Ivan Vulic}, booktitle = {Conference on Empirical Methods in Natural Language Processing}, pages = {15-22}, title = {Hello, It’s GPT-2 - How Can I Help You? Towards the Use of Pretrained Language Models for Task-Oriented Dialogue Systems}, year = {2019} }

[5]-@Article{Ji2021MentalBERTPA, author = {Shaoxiong Ji and Tianlin Zhang and Luna Ansari and Jie Fu and P. Tiwari and E. Cambria}, booktitle = {International Conference on Language Resources and Evaluation}, pages = {7184-7190}, title = {MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare}, year = {2021} }

[6]-@Article{Qiao2022ReasoningWL, author = {Shuofei Qiao and Yixin Ou and Ningyu Zhang and Xiang Chen and Yunzhi Yao and Shumin Deng and Chuanqi Tan and Fei Huang and Huajun Chen}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Reasoning with Language Model Prompting: A Survey}, volume = {abs/2212.09597}, year = {2022} }

[7]-@Article{Li2021PretrainedLM, author = {Junyi Li and Tianyi Tang and Wayne Xin Zhao and Ji-rong Wen}, booktitle = {International Joint Conference on Artificial Intelligence}, pages = {4492-4499}, title = {Pretrained Language Models for Text Generation: A Survey}, year = {2021} }

[8]-@Article{Hao2022LanguageMA, author = {Y. Hao and Haoyu Song and Li Dong and Shaohan Huang and Zewen Chi and Wenhui Wang and Shuming Ma and Furu Wei}, booktitle = {arXiv.org}, journal = {ArXiv}, title = {Language Models are General-Purpose Interfaces}, volume = {abs/2206.06336}, year = {2022} }

[1]-@Article{Li2018DeepDE, author = {Na Li and Deyi Tuo and Dan Su and Zhifeng Li and Dong Yu}, booktitle = {Interspeech}, pages = {2262-2266}, title = {Deep Discriminative Embeddings for Duration Robust Speaker Verification}, year = {2018} }

[2]-@Conference{Sadhana2017MiningTO, author = {S. Sadhana and L. Sairamesh and S. Sabena and S. Ganapathy and A. Kannan}, booktitle = {2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM)}, journal = {2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM)}, pages = {196-200}, title = {Mining Target Opinions from Online Reviews Using Semi-supervised Word Alignment Model}, year = {2017} }

[3]-@Article{Limsopatham2016BidirectionalLF, author = {Nut Limsopatham and Nigel Collier}, booktitle = {NUT@COLING}, pages = {145-152}, title = {Bidirectional LSTM for Named Entity Recognition in Twitter Messages}, year = {2016} }

[4]-@Inproceedings{Michel2018ConstitutionCF, author = {Cécile Michel}, pages = {43-70}, title = {Constitution, Contents, Filing and Use of Private Archives: The Case of Old Assyrian Archives (nineteenth century BCE)}, volume = {11}, year = {2018} }

[5]-@Article{Jobson1982EvaporationIT, author = {H. E. Jobson}, journal = {Eos, Transactions American Geophysical Union}, pages = {1223-1224}, title = {Evaporation Into the Atmosphere: Theory, History, and Applications}, volume = {63}, year = {1982} }