Tutorials at The Web Conference 2023

Valeria Fionda*
University of Calabria
Italy

Olaf Hartig*
Linköping University
Sweden

Reyhaneh Abdolazimi
Syracuse University
United States

Sihem Amer-Yahia
CNRS, Univ. Grenoble Alpes
France

Hongzhi Chen
AWS Shanghai AI Lab
China

Xiao Chen
The Hong Kong Polytechnic University
Hong Kong

Peng Cui
Tsinghua University
China

Jeffrey Dalton
University of Glasgow
United Kingdom

Xin Luna Dong
Meta Reality Labs
United States

Lisette Espin-Noboa
Central European University & Complexity Science Hub Vienna
Austria

Wenqi Fan
The Hong Kong Polytechnic University
Hong Kong

Manuela Fritz
University of Passau
Germany

Quan Gan
AWS Shanghai AI Lab
China

Jingtong Gao
City University of Hong Kong
Hong Kong

Xiaojie Guo
IBM T. J. Watson Research Center
United States

Torsten Hahmann
University of Maine
United States

Jiawei Han
University of Illinois at Urbana-Champaign
United States

Soyeon Han
The University of Sydney
Australia

Estevam Hruschka
Megagon Labs
United States

Liang Hu
Tongji University
China

Jiaxin Huang
University of Illinois at Urbana-Champaign
United States

Utkarshani Jaimini
University of South Carolina
United States

Olivier Jeunen
ShareChat
United Kingdom

Yushan Jiang
University of Connecticut
United States

Fariba Karimi
Vienna University of Technology & Complexity Science Hub Vienna
Austria

George Karypis
AWS AI Research and Education
United States

Krishnaram Kenthapadi
Fiddler AI
United States

Himabindu Lakkaraju
Harvard University
United States

Hady W. Lauw
Singapore Management University
Singapore

Thai Le
The University of Mississippi
United States

Trung-Hoang Le
Singapore Management University
Singapore

Dongwon Lee
The Pennsylvania State University
United States

Geon Lee
KAIST
South Korea

Liat Levontin
Technion
Israel

Cheng-Te Li
National Cheng Kung University
Taiwan

Haoyang Li
Tsinghua University
China

Ying Li
Netflix
United States

Jay Chiehen Liao
National Cheng Kung University
Taiwan

Qidong Liu
City University of Hong Kong
Hong Kong

Usha Lokala
University of South Carolina
United States

Ben London
Amazon
United States

Siqu Long
The University of Sydney
Australia

Hande Küçük McGinty
Kansas State University
United States

Yu Meng
University of Illinois at Urbana-Champaign
United States

Seungwhan Moon
Meta Reality Labs
United States

Usman Naseem
The University of Sydney
Australia

Pradeep Natarajan
Amazon Alexa AI
United States

Behrooz Omidvar-Tehrani
AWS AI Labs
United States

Zijie Pan
University of Connecticut
United States

Devesh Parekh
Netflix
United States

Jian Pei
Duke Univeristy
United States

Tiago Peixoto
Central European University
Austria

Steven Pemberton
CWI
Netherlands

Josiah Poon
The University of Sydney
Australia

Filip Radlinski
Google
United Kingdom

Federico Rossetto
University of Glasgow
United Kingdom

Kaushik Roy
University of South Carolina
United States

Aghiles Salah
Rakuten Group, Inc.
France

Mehrnoosh Sameki
Microsoft Azure AI
United States

Amit Sheth
University of South Carolina
United States

Cogan Shimizu
Wright State University
United States

Kijung Shin
KAIST
South Korea

Dongjin Song
University of Connecticut
United States

Julia Stoyanovich
New York University
United States

Dacheng Tao
The University of Sydney
Australia

Johanne Trippas
RMIT University
Australia

Quoc-Tuan Truong
Amazon
Canada

Yu-Che Tsai
National Taiwan University
Taiwan

Adaku Uchendu
The Pennsylvania State University
United States

Bram van den Akker
Booking.com
Netherlands

Lin Wang
The Hong Kong Polytechnic University
Hong Kong

Minjie Wang
AWS Shanghai AI Lab
China

Shoujin Wang
University of Technology Sydney
Australia

Xin Wang
Tsinghua University
China

Ingmar Weber
Saarland University
Germany

Henry Weld
The University of Sydney
Australia

Lingfei Wu
Pinterest
United States

Da Xu
Walmart Labs
United States

Ethan Yifan Xu
Meta Reality Labs
United States

Shuyuan Xu
Rutgers University
United States

Bo Yang
LinkedIn
United States

Ke Yang
UMass Amherst
United States

Elad Yom-Tov
Microsoft
Israel

Jaemin Yoo
Carnegie Mellon University
United States

Zhou Yu
Columbia University
United States

Reza Zafarani
Syracuse University
United States

Hamed Zamani
University of Massachusetts Amherst
United States

Meike Zehlike
Zalando Research
Germany

Qi Zhang
University of Technology Sydney
Australia

Xikun Zhang
The University of Sydney
Australia

Yongfeng Zhang
Rutgers University
United States

Yu Zhang
University of Illinois at Urbana-Champaign
United States

Zheng Zhang
AWS Shanghai AI Lab
China

Liang Zhao
Emory University
United States

Xiangyu Zhao
City University of Hong Kong
Hong Kong

Wenwu Zhu
Tsinghua University
China

ABSTRACT

This paper summarizes the content of the 28 tutorials that have been given at The Web Conference 2023.

CCS CONCEPTS

• **General and reference** → **Surveys and overviews**; *General conference proceedings.*

KEYWORDS

tutorials, the web conference

ACM Reference Format:
Valeria Fionda, Olaf Hartig, Reyhaneh Abdolazimi, Sihem Amer-Yahia, Hongzhi Chen, Xiao Chen, Peng Cui, Jeffrey Dalton, Xin Luna Dong, Lisette Espin-Noboa, Wenqi Fan, Manuela Fritz, Quan Gan, Jingtong Gao, Xiaojie Guo, Torsten Hahmann, Jiawei Han, Soyeon Han, Estevam Hruschka, Liang Hu, Jiaxin Huang, Utkarshani Jaimini, Olivier Jeunen, Yushan Jiang, Fariba Karimi, George Karypis, Krishnaram Kenthapadi, Himabindu Lakkaraju, Hady W. Lauw, Thai Le, Trung-Hoang Le, Dongwon Lee, Geon Lee, Liat Levontin, Cheng-Te Li, Haoyang Li, Ying Li, Jay Chiehen Liao, Qidong Liu, Usha Lokala, Ben London, Siqu Long, Hande Küçük McGinty, Yu Meng, Seungwhan Moon, Usman Naseem, Pradeep Natarajan, Behrooz Omidvar-Tehrani, Zijie Pan, Devesh Parekh, Jian Pei, Tiago Peixoto, Steven Pemberton, Josiah Poon, Filip Radlinski, Federico Rossetto, Kaushik Roy, Aghiles Salah, Mehrnoosh Sameki, Amit Sheth, Cogan Shimizu, Kijung Shin, Dongjin Song, Julia Stoyanovich, Dacheng Tao, Johanne Trippas, Quoc-Tuan Truong, Yu-Che Tsai, Adaku Uchendu, Bram van den Akker, Lin Wang, Minjie Wang, Shoujin Wang, Xin Wang, Ingmar Weber, Henry Weld, Lingfei Wu, Da Xu, Ethan Yifan Xu, Shuyuan Xu, Bo Yang, Ke Yang, Elad Yom-Tov, Jaemin Yoo, Zhou Yu, Reza Zafarani, Hamed Zamani, Meike Zehlike, Qi Zhang, Xikun Zhang, Yongfeng Zhang, Yu Zhang, Zheng Zhang, Liang Zhao, Xiangyu Zhao, and Wenwu Zhu. 2023. Tutorials at The Web Conference 2023. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3543873.3587713

[*] Tutorial co-chair

WWW '23 Companion, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9419-2/23/04.
https://doi.org/10.1145/3543873.3587713

1 INTRODUCTION

The Web Conference 2023 hosted 28 tutorials in Austin, Texas, United States. These tutorials were presented on the 30th of April and on the 1st of May 2023. Following the successful approach of previous years, this edition hosted two tutorial formats:

- Lecture-style (67%) tutorials cover the state-of-the-art research, development, and applications in a specific web computing and related area, and stimulate and facilitate future work.
- Hands-on (33%) tutorials feature in-depth practical training on cutting-edge systems and tools of relevance to the web conference community. These sessions are targeted at novice as well as moderately skilled users.

The duration of the 28 accepted tutorials was the following:

- 2 full-day tutorials (6 hours);
- 13 half-day tutorials (3 hours);
- 13 short tutorials (1.5 hours).

The tutorial track of The Web Conference 2023 received submissions from 56 unique institutions (30 of which are universities) located in 13 countries.

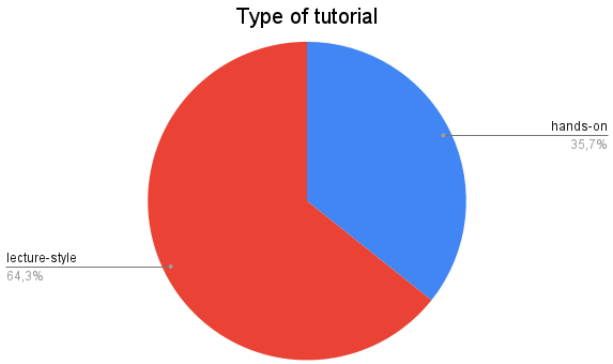

Figure 1: Distribution of tutorial's type.

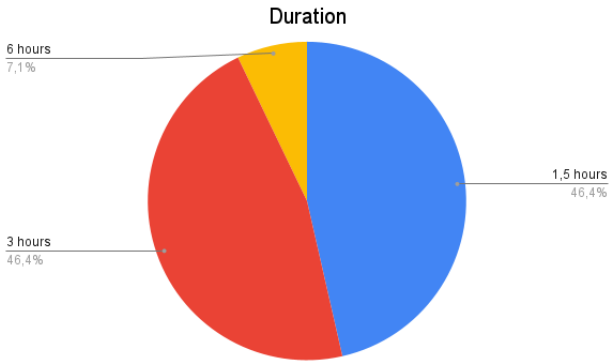

Figure 2: Distribution of tutorial's duration.

2 TUTORIAL 1: INTERNET ADVERTISING TO IMPROVE PUBLIC HEALTH: OPPORTUNITIES AND CHALLENGES

Presenters: Elad Yom-Tov, Liat Levontin, Ingmar Weber and Manuela Fritz

Internet advertising is one of the most common methods for communicating with large audiences. As such, it has been used by health authorities to improve health and wellness in a variety of areas ranging from prevention (vaccination and smoking cessation) to early disease screening and adoption of healthy living habits.

Creating effective health campaigns remains a challenge. However, recent advances in online advertising systems (e.g., algorithmic copywriting) and novel ways of using these systems have created new ways in which such campaigns can be deployed quickly and effectively.

This tutorial had the aim to give participants an overview of advertising in general and health advertising in particular, and made participants familiarize with recent advances in the area of health advertising, and provided them with tools for setting up public health advertising campaign. The goal of the tutorial was to show participants some of the areas where improvements in algorithmic approaches and deployment methods could help make a positive impact on people's health around the world.

3 TUTORIAL 2: TRUSTWORTHY RECOMMENDER SYSTEMS

Presenters: Wenqi Fan, Xiangyu Zhao, Lin Wang, Xiao Chen, Jingtong Gao and Qidong Liu

As one of the most successful AI applications, recommender systems have been applied in various user-oriented Web services. However, recent studies indicate that recommender systems can expose their untrustworthy aspects to users and society, such as spreading fake news to manipulate public opinion on social media sites, amplifying unfairness toward under-represented groups or individuals in job matching services, or inferring privacy information from recommendation results. These vulnerabilities can make unreliable recommendation results and then produce significant harmful effects in various real-world applications, such as finance and healthcare. Thus, building trustworthy recommender systems has attracted increasing attention in recent years.

The aim of this tutorial was to provide a comprehensive overview on the recent progress of **T**rustworthy **Rec**ommender systems (**TRec**) with a specific focus on six crucial dimensions: Safety & Robustness, Non-discrimination & Fairness, Explainability, Privacy, Environmental Well-being, Accountability & Auditability. For each dimension, the basic concepts and taxonomy have been introduced, then the summary of the latest and most representative works has been provided. Additionally, the potential interactions among different dimensions and possible future research directions for trustworthy recommender systems were given. In this way, academic researchers and industrial practitioners in related fields have gained a broad overview and a deep understanding of trustworthiness in recommender systems.

4 TUTORIAL 3: DATA-INFORMED INTERACTION LEARNING

Presenters: Behrooz Omidvar-Tehrani and Sihem Amer-Yahia

Users in their different roles (data owners, customers, stakeholders, advertisers) interact with data systems to seek information. Typical information seeking paradigms such as search and recommendation, depend on a crystal-clear definition of an information seeking task to operate. In reality, most data-centric tasks are subjective which require a sequence of interactions to manifest and clarify. Interacting with data systems is a tedious task and users need to receive guided recommendations from the system to optimize their interactions when dealing with subjectivity. The focus of this tutorial was on the challenge of subjective needs and multi-shot tasks, and it had the aim of reviewing (i) how the diverse and heterogeneous set of user interactions can be formalized and represented in the form of a unique model, and (ii) how data systems can leverage the interaction model to assist users in their interactions by disambiguating their needs and guiding users through their information seeking journey until landing on their ideal target. Real-world use cases of formalizing and learning interactions, both in academia and industry, were presented whose core objective was to help users interact with data systems more effectively. Future directions of interaction learning were also discussed, such as incorporating domain knowledge and multi-environment exploration.

5 TUTORIAL 4: GRAPH NEURAL NETWORKS: FOUNDATION, FRONTIERS AND APPLICATIONS

Presenters: Lingfei Wu, Peng Cui, Jian Pei, Liang Zhao and Xiaojie Guo

The field of graph neural networks (GNNs) has seen rapid and incredible strides over the recent years. Graph neural networks, also known as deep learning on graphs, graph representation learning, or geometric deep learning, have become one of the fastest-growing research topics in machine learning, especially deep learning. This wave of research at the intersection of graph theory and deep learning has also influenced other fields of science, including recommendation systems, computer vision, natural language processing, inductive logic programming, program synthesis, software mining, automated planning, cybersecurity, and intelligent transportation. However, as the field rapidly grows, it has been extremely challenging to gain a global perspective of the developments of GNNs. This tutorial of Graph Neural Networks (GNNs): Foundation, Frontiers and Applications has covered a broad range of topics in graph neural networks, by reviewing and introducing the fundamental concepts and algorithms of GNNs, new research frontiers of GNNs, and broad and emerging applications with GNNs.

6 TUTORIAL 5: MINING OF REAL-WORLD HYPERGRAPHS: PATTERNS, TOOLS, AND GENERATORS

Presenters: Geon Lee, Jaemin Yoo and Kijung Shin

Group interactions are prevalent in various complex systems (e.g., collaborations of researchers and group discussions on online Q&A sites), and they are commonly modeled as hypergraphs. Hyper-edges, which compose a hypergraph, are non-empty subsets of any number of nodes, and thus each hyperedge naturally represents a group interaction among entities. The higher-order nature of hypergraphs brings about unique structural properties that have not been considered in ordinary pairwise graphs.

In this tutorial, a comprehensive overview of a new research topic called hypergraph mining has been offered. The tutorial organizers first presented recently revealed structural properties of real-world hypergraphs, including (a) static and dynamic patterns, (b) global and local patterns, and (c) connectivity and overlapping patterns. Together with the patterns, they described advanced data mining tools used for their discovery. Lastly, they introduced simple yet realistic hypergraph generative models that provide an explanation of the structural properties.

7 TUTORIAL 6: PRACTICAL BANDITS: AN INDUSTRY PERSPECTIVE

Presenters: Bram van den Akker, Olivier Jeunen, Ying Li, Ben London, Zahra Nazari and Devesh Parekh

The bandit paradigm provides a unified modeling framework for problems that require decision-making under uncertainty. Because many business metrics can be viewed as rewards (a.k.a. utilities) that result from actions, bandit algorithms have seen a large and growing interest from industrial applications, such as search, recommendation and advertising. Indeed, with the bandit lens comes the promise of direct optimisation for the metrics we care about.

Nevertheless, the road to successfully applying bandits in production is not an easy one. Even when the action space and rewards are well-defined, practitioners still need to make decisions regarding multi-arm or contextual approaches, on- or off-policy setups, delayed or immediate feedback, myopic or long-term optimisation, etc. To make matters worse, industrial platforms typically give rise to large action spaces in which existing approaches tend to break down. The research literature on these topics is broad and vast, but this can overwhelm practitioners, whose primary aim is to solve practical problems, and therefore need to decide on a specific instantiation or approach for each project. This tutorial has taken a step towards filling that gap between the theory and practice of bandits. The goal was to present a unified overview of the field and its existing terminology, concepts and algorithms – with a focus on problems relevant to industry. The industrial perspective hopefully had helped future practitioners who wish to leverage the bandit paradigm for their application.

8 TUTORIAL 7: SAME DATA, DIFFERENT MODELS: CHOOSING AN ONTOLOGY MODELING METHODOLOGY

Presenters: Cogan Shimizu, Torsten Hahmann and Hande McGinty

As knowledge graph development (and in particular the development of their schemas) grows commensurately with the importance of knowledge graphs in industry and academia, it follows

that choosing a development methodology to fit the application scenario and domain is correspondingly important. This 3-hour, hands-on tutorial had the purposes of comparing and contrasting three distinct ontology modeling methodologies: Modular Ontology Modeling (MOMo), Domain Reference Ontology Design (DROD), and Knowledge Acquisition and Representation Methodology (KNARM). Attendees had the opportunity to execute each methodology. The tutorial culminated in a retrospective for the different subtutorials.

9 TUTORIAL 8: GRAPH NEURAL NETWORKS FOR TABULAR DATA LEARNING

Presenters: Cheng-Te Li, Yu-Che Tsai and Jay Chiehen Liao

The deep learning-based approaches to Tabular Data Learning (TDL) have shown competing performance, compared to their conventional counterparts. However, the latent correlation among data instances and feature values is less modeled in deep neural TDL. Recently, graph neural networks (GNN), which can enable modeling relations and interactions between different data entities, has received tremendous attention across application domains including TDL. It turns out creating proper graph structures from the input tabular data, along with GNN learning, can improve the TDL performance. In this tutorial, the methodologies of designing and applying GNNs to TDL have been systematically introduced. The topics covered include: (1) the foundations and overview of GNN-based TDL methods, especially on formulating TDL as various graph structures; (2) a comprehensive taxonomy of constructing graph structures and representation learning in GNN-based TDL methods; (3) the TDL model training framework that contains different auxiliary tasks and even allows open-world learning; (4) how to apply GNNs to various TDL application scenarios and tasks; (5) the limitations in current research and future directions.

10 TUTORIAL 9: CONVERSATIONAL INFORMATION SEEKING: THEORY AND APPLICATION

Presenters: Jeffrey Dalton, Filip Radlinksi, Federico Rossetto, Hamed Zamani and Johanne Trippas

Conversational information seeking (CIS) involves interaction sequences between one or more users and an information system. Interactions in CIS are primarily based on natural language dialogue, while they may include other types of interactions, such as click, touch, and body gestures. CIS recently attracted significant attention and advancements continue to be made. This tutorial followed the content of the recent Conversational Information Seeking book authored by several of the tutorial presenters. The aim of the tutorial was to introduce the CIS concepts to a broader audience as a mixture of lecture style and hands-on tutorial. The tutorial aimed to introduce CIS to newcomers to CIS in addition to the recent advanced topics and state-of-the-art approaches for students and researchers with moderate knowledge of the topic. A significant part of the tutorial was dedicated to hands-on experiences based on toolkits developed by the presenters for conversational passage retrieval and multi-modal task-oriented dialogues. This included open-source versions of the winning Alexa Prize system for the 2021/2022 Alexa Taskbot Challenge. The outcomes of this tutorial included theoretical and practical knowledge, including a forum to meet researchers interested in CIS.

Acknowledgements: This research was partially completed when Dr Trippas was affiliated with and funded by the University of Melbourne.

11 TUTORIAL 10: DECLARATIVE WEB APPLICATIONS WITH XFORMS

Presenter: Steven Pemberton

In the 50s, when the first programming languages were designed, computers cost millions, and relatively, programmers were almost free. Those programming languages therefore reflected that relationship: it didn't matter if it took a long time to program, as long as the resulting program ran as fast as possible. Now, that relationship has been reversed: compared to the cost of programmers, computers are almost free. And yet we are still programming them in descendants of the programming languages from the 50's: we are still telling the computers step by step how to solve the problem.

Declarative programming is a new approach to applications: rather than describing exactly how to reach the solution, it describes what the solution should look like, and leaves more of the administrative parts of the program to the computer. One of the few declarative languages available is XForms, an XML-based language that despite what its name seems to suggest is not only about forms. Large projects, at large companies such as the National Health Service, the BBC and Xerox, have shown that by using XForms, programming time and cost of applications can be reduced to a tenth, and sometimes even much more!

This hands-on tutorial allowed attendees to learn about the structure and workings of XForms, and gave them the opportunity to develop some surprisingly powerful and useful applications. It was a "bring your own device" tutorial. Attendees were required to install some files beforehand, and check they were working. Attendees were able to work with their choice of text editor during the tutorial.

12 TUTORIAL 11: INVISIBLE MARKUP

Presenter: Steven Pemberton

Humans are good at identifying implicit structure in notations. People can deduce the structure of a date like 30 April 2023 with no help. Computers on the other hand need extra information, and that is why we have markup languages, like

```
<date>
   <day>23</day>
   <month>May</month>
   <year>2023</year>
</date>
```

to make the structure explicit. Invisible Markup is a method of automatically discovering the structure in notations, and adding markup. It doesn't matter what the input form is, whether CSV, or JSON, CSS, bibliographic entries, family tree information, or countless other examples, the output is a consistently marked up result, currently XML being the target, as the most general of the available markup languages. For instance the input might be

```
[spec] Steven Pemberton (ed.), Invisible XML
Specification, invisiblexml.org, 2022,
https://invisiblexml.org/ixmlspecification.html
```

and while you have a lot of control over the details, the output could be

```
<biblioentry>
   <abbrev>spec</abbrev>
   <editor>
      <firstname>Steven</firstname>
      <surname>Pemberton</surname>
   </editor>
   <title>Invisible XML Specification</title>
   <publisher>invisiblexml.org</publisher>
   <pubdate>2022</pubdate>
   <link href='https://invisiblexml.org/ixml-
         specification.html'/>
</biblioentry>
```

The Invisible Markup language ixml was formally released as a standard in the Summer of 2022. This hands-on tutorial introduced the principles of the language, and took attendees through how to use it.

13 TUTORIAL 12: CATCH ME IF YOU GAN: GENERATION, DETECTION, AND OBFUSCATION OF DEEPFAKE TEXTS

Presenters: Adaku Uchendu, Thai Le and Dongwon Lee

In recent years, Natural Language Generation (NLG) techniques have greatly advanced, especially in the realm of open-ended text generation using deep learning methods. With respect to the quality of generated texts , therefore, it has become no longer trivial to tell the difference between human-written and NLG-generated texts (so-called **deepfake texts**). While this is a celebratory feat for NLG, however, it poses new security risks (e.g., generation of misinformation or phishing message at scale). To combat this novel challenge, researchers have developed diverse techniques to automatically detect NLG-generated texts. While this niche field of deepfake text detection is growing, the field of NLG is growing at a much faster rate, thus making it difficult to understand the complex interplay between state-of-the-art NLG methods and the detectability of their deepfake texts. Scaling up the problem further to the case of k NLG methods ($k \geq 2$), each generating uniquely-different yet human-quality texts, two new computational problems emerge: "Neural" Authorship Attribution (AA) and "Neural" Authorship Obfuscation (AO) problems, where the AA problem is concerned with attributing the authorship of a given deepfake text to one of k NLG methods, while the AO problem is to evade the authorship of a given text by modifying parts of the text. Both problems lie in the intersection between Machine Learning and Security/Privacy, and its importance and implications are growing rapidly on the domain of World-Wide Web, where the bulk of "information" is text-based (e.g., NLG-made Wikipedia articles with factual errors on obscure subjects may evade human curation and jeopardize the Web ecosystem). In this tutorial, therefore, the presenters called attention to the serious security risks both emerging problems posed and given a comprehensive review of recent literature on the: (1) generation, (2) detection (3) obfuscation of deepfake text authorship, (4) their utility in the Web applications and (5) critical implications in the society. This tutorial was mainly of lecture-style, together with hands-on examples of the generation and detection of deepfake texts for interactive participation from audience. The tutorial presenters selectively selectd the word "GAN", which stands for Generative Adversarial Network, as part of the title to emphasize its symbolic pioneer of modern AI generative models.

14 TUTORIAL 13: TOWARDS NEXT-GENERATION INTELLIGENT ASSISTANTS FOR AR/VR DEVICES

Presenters: Xin Luna Dong, Seungwhan Moon, Yifan Xu and Zhou Yu

Virtual Intelligent Assistants take user requests in the voice form, perform actions such as setting an alarm, turning on a light, and answering a question, and provide answers or confirmations in the voice form or through other channels such as a screen. Assistants have become prevalent in the past decade, and users have been taking services from assistants like *Amazon Alexa, Apple Siri, Google Assistant, and Microsoft Cortana.*

The emergence of AR/VR devices raised many new challenges for building intelligent assistants. The unique requirements have inspired new research directions such as (a) understanding users' situated multi-modal contexts (e.g. vision, sensor signals) as well as language-oriented conversational contexts, (b) grounding interactions on growing external and internal knowledge graphs, and (c) developing inference models with on-device constraints and privacy-secured methods. This tutorial provided an in-depth walkthrough of techniques in the aforementioned areas in the recent literature. The presenters introduced techniques for researchers and practitioners who are building intelligent assistants, and aimed to inspire research that will bring us one step closer to realizing the dream of building an all-day accompanying AR/VR assistant.

15 TUTORIAL 14: FAIRNESS IN RANKING: FROM VALUES TO TECHNICAL CHOICES AND BACK

Presenters: Julia Stoyanovich, Meike Zehlike and Ke Yang

In the past few years, there has been much work on incorporating fairness requirements into the design of algorithmic rankers, with contributions from the data management, algorithms, information retrieval, and recommender systems communities. In this tutorial, based on a recent two-part survey that appeared in ACM Computing Surveys, the presenters gave a systematic overview of this work, offering a broad perspective that connects formalizations and algorithmic approaches across subfields.

During the introductory part of the tutorial, a classification framework for fairness-enhancing interventions was presented, along which they then relate the technical methods. Next, during the main part of the tutorial, the presenters discussed fairness in score-based ranking (Part 1) and in supervised learning-to-rank (Part 2). The tutorial concluded with a set of recommendations for

practitioners, to help them select a fair ranking method based on the requirements of their specific application domain.

16 TUTORIAL 15: CAUSAL AI FOR WEB AND HEALTH CARE

Presenters: Utkarshani Jaimini, Usha Lokala, Kaushik Roy and Amit Sheth

Improving the performance and explanations of ML algorithms is a priority for adoption by humans in the real world. In critical domains such as healthcare, such technology has significant potential to reduce the burden on humans and considerably reduce manual assessments by providing quality assistance at scale. In today's data driven world, artificial intelligence (AI) systems are still experiencing issues with bias, explainability, and human-like reasoning and interpretability. Causal AI is the technique that can reason and make human like choices making it possible to go beyond narrow Machine learning based techniques and can be integrated into human decision making. It also offers intrinsic explainability, new domain adaptability, bias free predictions and works with datasets of all sizes. In this tutorial of type lecture style, the presenters detailed how a richer representation of causality in AI systems using a knowledge graph (KG) based approach is needed for intervention and counterfactual reasoning, how does one get to model based and domain explainability, how causal representations helps in web and health care.

17 TUTORIAL 16: TURNING WEB-SCALE TEXTS TO KNOWLEDGE: TRANSFERRING PRETRAINED REPRESENTATIONS TO TEXT MINING APPLICATIONS

Presenters: Yu Meng, Jiaxin Huang, Yu Zhang and Jiawei Han

Textual data are ubiquitous and massive on the web: News reports, social media posts, Wikipedia articles, etc. are being created and updated online everyday. While they contain rich information and knowledge, it has remained an open challenge to effectively leverage them in text-intensive applications. Recent developments in pretrained language models (PLMs) have revolutionized text mining and processing: By pretraining neural architectures on large-scale text corpora obtained from the web and then transferring their representations to task-specific data, the knowledge encoded in the web-scale corpora can be effectively leveraged to significantly enhance the downstream task performance. The most common adaptation approach of PLMs is the pretrain-finetune paradigm where the PLMs are further trained on downstream task labeled data. However, the major challenge of such a paradigm is that fully-supervised fine-tuning of PLMs usually require abundant human annotations, which can be expensive to acquire in practice.

In this tutorial, the presenters introduced recent advances in pretrained text representations learned from web-scale corpora, as well as their applications to a wide range of text mining tasks. The tutorial focused on weakly-supervised approaches without requiring massive human annotations, including (1) pretrained language models that serve as the fundamentals for downstream tasks, (2) unsupervised and distantly-supervised methods for fundamental text mining applications, (3) unsupervised and seed-guided methods for topic discovery from massive text corpora and (4) weakly-supervised methods for text classification and advanced text mining tasks.

18 TUTORIAL 17: THE ADVANTAGES OF ADDING NOISE

Presenters: Reyhaneh Abdolazimi and Reza Zafarani

Noise often describes undesirable disturbances or fluctuations in data, and creates problems for communication, signal detection, and neural networks. Although disruptive, noise can potentially help many information processing systems and algorithms. Noise enhancement has been utilized in image processing, signal processing, machine learning, and deep learning techniques in order to improve efficiency and performance. This tutorial presented several fields where noise is beneficial, such as (1) noise-enhanced systems, (2) noise-enhanced learning algorithms, and (3) noise-enhanced applications.

19 TUTORIAL 18: ADVANCES IN SIMULATION TECHNOLOGY FOR WEB APPLICATIONS

Presenters: Da Xu, Shuyuan Xu, Bo Yang and Yongfeng Zhang

The developments of modern Web applications, including search, recommendation, and advertisement, are highly complex and iterative processes that require a deep understanding of system dynamics and user behavior. Both academia and industry have found simulations beneficial for addressing many open challenges, including measuring the effect and bias of interventions, predicting long-term utilities such as fairness, inclusiveness, and diversity, as well as analyzing complex system properties under different human behaviors. While simulations allow researchers and developers to craft controlled environments, the delicate balance between manipulability and faithfulness, flexibility and complexity, hypothesis and reality can often cause concerns about the credibility and usefulness of simulation analysis. Indeed, the success of Web simulations depends crucially on the scope of analysis, the model assumptions, as well as the objectives and tasks of interest. Despite the fact that simulations are extensively used by the community, there was a lack of comprehensive and systematic tutorial that provided instructions and guidelines for using the right tools in a given context.

This tutorial presented the necessary background and scope, summarized the existing simulation ideas and frameworks, analyzed their main factors and pitfalls in real-world case studies, and connected Web simulations to advanced technologies from causal inference and reinforcement learning. The presenters aimed to give attendants an overview of the different aspects to consider and help them navigate through the many complex choices they may face when using simulation technologies in practice. Based on the experience of the presenters and the study of other works, this lecture-style tutorial intended to guide designing, implementing, and optimizing simulation analysis for a variety of Web applications.

20 TUTORIAL 19: MODEL MONITORING IN PRACTICE: LESSONS LEARNED AND OPEN CHALLENGES

Presenters: Krishnaram Kenthapadi, Himabindu Lakkaraju, Pradeep Natarajan and Mehrnoosh Sameki

Artificial Intelligence (AI) is increasingly playing an integral role in determining our day-to-day experiences. Increasingly, the applications of AI are no longer limited to search and recommendation systems, such as web search and movie and product recommendations, but AI is also being used in decisions and processes that are critical for individuals, businesses, and society. With AI based web applications in high-stakes domains such as hiring, lending, criminal justice, healthcare, and education, the resulting personal and professional implications of AI are far-reaching. Consequently, it becomes critical to ensure that these models are making accurate predictions, are robust to shifts in the data, are not relying on spurious features, and are not unduly discriminating against minority groups. To this end, several approaches spanning various areas such as explainability, fairness, and robustness have been proposed in recent literature, and many papers and tutorials on these topics have been presented in recent computer science conferences. However, there is relatively less attention on the need for monitoring machine learning (ML) models once they are deployed and the associated research challenges.

In this tutorial, the presenters first motivated the need for ML model monitoring, as part of a broader AI model governance, responsible web computing, and responsible AI frameworks, from societal, legal, customer/end-user, and model developer perspectives, and provided a roadmap for thinking about model monitoring in practice. They then presented findings and insights on model monitoring desiderata based on interviews with various ML practitioners spanning domains such as financial services, healthcare, hiring, online retail, computational advertising, and conversational assistants. They then described the technical considerations and challenges associated with realizing the above desiderata in practice. They provided an overview of techniques/tools for model monitoring. Then, they focused on the real-world application of model monitoring methods and tools, presented practical challenges/guidelines for using such techniques effectively, and lessons learned from deploying model monitoring tools for several web-scale AI/ML applications. The tutorial presented case studies across different companies, spanning application domains such as financial services, healthcare, hiring, conversational assistants, online retail, computational advertising, search and recommendation systems, and fraud detection. The aim of the tutorial was to inform both researchers and practitioners, stimulate further research on model monitoring, and pave the way for building more reliable ML models and monitoring tools for different web applications in the future.

21 TUTORIAL 20: CONTINUAL GRAPH LEARNING

Presenters: Xikun Zhang, Dongjin Song, Yushan Jiang, Zijie Pan and Dacheng Tao

Over the past years, deep learning on graphs has made significant progress in various areas. However, most graph learning tasks assume graphs are static, while real-world graphs may constantly grow or evolve. Therefore, it is crucial to study how to constantly adapt a graph learning model to new patterns/tasks over graphs without forgetting the previously learned knowledge. To this end, in this tutorial, the presenters introduced the newly emerging area of continual graph learning (CGL). Specifically, they (1) introduced different continual graph learning settings, (2) presented the key challenges in CGL, (3) highlighted the existing CGL techniques, and (4) discussed future directions.

22 TUTORIAL 21: NEVER-ENDING LEARNING, LIFELONG LEARNING AND CONTINUAL LEARNING IN THE ERA OF LARGE PRE-TRAINED LANGUAGE MODELS

Presenters: Estevam Hruschka

In the last decades, different Never-Ending Learning (NEL) approaches have been proposed and applied in different tasks and domains. Results from NEL approaches have not achieved the state-of-the-art in every task/application, but such approaches are becoming more frequent and more successful, thus encouraging to keep addressing the problem of how to build approaches that can take advantage of NEL principles. In addition, the variety of different names (never-ending learning, continuous learning, lifelong learning, etc.), different assumptions (single/multi-model, neural/symbolic/hybrid, external/internal memory, etc.) used to describe systems and models that can keep learning in a continuous way, as well as the new achievements on self-supervised learning and multi-task learning (which are closely related to NEL principles) brought by large pre-trained models are also relevant aspects that motivates the proposal of this tutorial for the Web community. In this tutorial, the presenters explored the Never-Ending Learning (NEL) ideas and principles, the different terminology, assumptions and approaches (and their variations) that can be found in the literature, their similarities and differences from other ML approaches (such as semi-supervised learning, reinforcement learning, etc.) and some recent applications. The tutorial also explored that it is not always straightforward to formulate a problem following a NEL approach. Thus, the presenters showed (with examples, supporting theory, algorithms and models) how to model a problem in a NEL fashion and helped attendees to become familiar with such approaches.

23 TUTORIAL 22: MULTI-MODAL RECOMMENDER SYSTEMS: TOWARDS ADDRESSING SPARSITY, COMPARABILITY, AND EXPLAINABILITY

Presenters: Trung-Hoang Le, Quoc-Tuan Truong, Aghiles Salah and Hady Lauw

Web applications frequently feature a recommender system to help users discover items (e.g., products, content articles) of interest. This tutorial focused on multi-modality, i.e., the use of side information such as text, images, or graphs to augment the preference data. In particular, the presenters covered several important aspects of multi-modality. First, they showed how models rely on the auxiliary modality to address the sparsity of preference observations in

order to better bridge users and items. These models are typically designed along modality lines, which they covered comprehensively. Second, the presenters showed how to manage comparison and cross-utilization of multi-modal models. The former is concerned with streamlining the treatment of models that share the same modality. The latter is concerned with using a model initially designed for one modality with another. Third, the presenters showed how the auxiliary modalities could act as recommendation explanations, as recipients may find textual, visual, or graphical explanations more intuitive. This was a hands-on tutorial, whereby lectures were supplemented with exercises conducted with Cornac (https://cornac.preferred.ai), a comparative framework for multimodal recommender systems.

24 TUTORIAL 23: LIFELONG LEARNING CROSS-DOMAIN RECOMMENDER SYSTEMS

Presenters: Liang Hu, Shoujin Wang, Qi Zhang and Usman Naseem

This tutorial presented the state-of-the-art and emerging studies on lifelong learning cross-domain recommender systems (RSs), including the latest and most advanced theories, methods, models, data, and applications. At the beginning of this tutorial, the background and foundation of lifelong learning and RSs were presented. Then, the 3C-principle, Complement, Composite, and Context, have been utilized to build various lifelong multidomain RSs by continuously fusing new knowledge across multiple domains. More specifically, the tutorial organizers respectively presented the details of how to model lifelong learning cross-domain RSs with (1) complementary knowledge, (2) composite knowledge, and (3) contextual knowledge over ever-evolving domains. Finally, the organizers demonstrated several representative emerging real-world applications in the e-commerce, fashion industry, finance, and multimedia.

25 TUTORIAL 24: TOWARDS OUT-OF-DISTRIBUTION GENERALIZATION ON GRAPHS

Presenters: Xin Wang, Haoyang Li and Wenwu Zhu

Graph machine learning has been extensively studied in both academia and industry. Although booming with a vast number of emerging methods and techniques, most of the literature is built on the in-distribution (I.D.) hypothesis, i.e., testing and training graph data are sampled from the identical distribution. However, this I.D. hypothesis can hardly be satisfied in many real-world graph scenarios where the model performance substantially degrades when there exist distribution shifts between testing and training graph data. To solve this critical problem, out-of-distribution (OOD) generalization on graphs, which goes beyond the I.D. hypothesis, has made great progress and attracted ever-increasing attention from the research community. This tutorial was to disseminate and promote the recent research achievement on out-of-distribution generalization on graphs, which is an exciting and fast-growing research direction in the general field of machine learning and data mining. The presenters advocated novel, high-quality research findings, as well as innovative solutions to the challenging problems in out-of-distribution generalization and its applications on graphs.

26 TUTORIAL 25: SOCIAL NETWORK MODELING AND APPLICATIONS, A TUTORIAL

Presenters: Lisette Espin Noboa, Tiago Peixoto and Fariba Karimi

Social networks have been widely studied over the last century from multiple disciplines to understand societal issues such as inequality in employment rates, managerial performance, and epidemic spread. Today, these and many more issues can be studied at global scale thanks to the digital footprints that we generate when browsing the Web or using social media platforms. Unfortunately, scientists often struggle to access to such data primarily because it is proprietary, and even when it is shared with privacy guarantees, such data is either no representative or too big. In this tutorial, the presenters discussed recent advances and future directions in network modeling. In particular, they focused on how to exploit synthetic networks to study real-world problems such as data privacy, spreading dynamics, algorithmic bias, and ranking inequalities. They started by reviewing different types of generative models for social networks including node-attributed and scale-free networks. Then, they showcased how to perform a network selection analysis to characterize the mechanisms of edge formation of any given real-world network

27 TUTORIAL 26: SPOKEN LANGUAGE UNDERSTANDING FOR CONVERSATIONAL AI: RECENT ADVANCES AND FUTURE DIRECTION

Presenters: Soyeon Han, Siqu Long, Henry Weld and Josiah Poon

When a human communicates with a machine using natural language on the web and online, how can it understand the human's intention and semantic context of their talk? This is an important AI task as it enables the machine to construct a sensible answer or perform a useful action for the human. Meaning is represented at the sentence level, identification of which is known as intent detection, and at the word level, a labelling task called slot filling. This dual-level joint task requires innovative thinking about natural language and deep learning network design, and as a result, many approaches and models have been proposed and applied.

This tutorial discussed how the joint task is set up and introduced Spoken Language Understanding/Natural Language Understanding (SLU/NLU) with Deep Learning techniques. Organizers covered the datasets, experiments and metrics used in the field. The presenters described how the machine uses the latest NLP and Deep Learning techniques to address the joint task, including recurrent and attention-based Transformer networks and pre-trained models (e.g. BERT). They then looked in detail at a network that allows the two levels of the task, intent classification and slot filling, to interact to boost performance explicitly. They did a code demonstration of a Python notebook for this model and attendees had an opportunity to watch coding demo tasks on this joint NLU to further their understanding.

28 TUTORIAL 27: SELF-SUPERVISED LEARNING AND PRE-TRAINING ON GRAPHS

Presenters: Yukuo Cen, Yuxiao Dong and Jie Tang

This tutorial gave a systematic introduction to self-supervised learning and pre-training for graph neural networks (GNNs). The presenters started with industry examples from Alibaba, WeChat, and AMiner to explain how graph mining and machine learning are benefiting from GNNs. Then they comprehensively introduced both the history and recent advances on self-supervised learning and pre-training for GNNs. Finally, they presented CogDL – a comprehensive library for graph deep learning. In CogDL, they proposed a unified design for the training loop of graph neural network (GNN) models, making it unique among existing graph learning libraries. All the GNN models, self-supervised, and pre-training techniques introduced in the tutorial are provided with reproducible code in CogDL. Uniquely, this tutorial aimed to provide the audience with 1) recent self-supervised learning techniques on graphs, 2) pre-training advances on graphs, and 3) easy-to-use CogDL APIs and examples for training self-supervised GNNs and pre-training GNNs.

29 TUTORIAL 28: WHEN SPARSE MEETS DENSE: LEARNING ADVANCED GRAPH NEURAL NETWORKS WITH DGL-SPARSE PACKAGE

Presenters: Minjie Wang, Hongzhi Chen, Quan Gan, George Karypis and Zheng Zhang

Learning from graph and relational data plays a major role in many applications including social network analysis, marketing, e-commerce, information retrieval, knowledge modeling, medical and biological sciences, engineering, and others. In the last few years, Graph Neural Networks (GNNs) have emerged as a promising new supervised learning framework capable of bringing the power of deep representation learning to graph and relational data. This ever-growing body of research has shown that GNNs achieve state-of-the-art performance for problems such as link prediction, fraud detection, target-ligand binding activity prediction, knowledge-graph completion, and product recommendations. The power of GNNs comes from its capability of collecting and aggregating information from local neighborhood using a paradigm called message passing. However, the message passing paradigm suffers from two notable issues. First, the learnt node representations quickly become indistinguishable with more layers, a phenomenon called oversmoothing. Second, generalizing it to more complex graphs such as hypergraphs is not straightforward. Therefore, many recent research work attempt to develop new GNNs beyond the message passing paradigm. Instead of defining a GNN in terms of node-wise and edge-wise operations, their models are typically defined as a series of operations over the sparse adjacency matrix of the input graph plus other dense tensor inputs like node embeddings. Previous work have shown, both theoretically and empirically, that those new architectures can be very deep and also extends well to complex graphs.

This tutorial introduced DGL-Sparse, a new package of the popular GNN framework Deep Graph Library (DGL). DGL- Sparse provides flexible and efficient sparse matrix operations for users to develop, train and apply advanced GNNs beyond the message passing paradigm. The tutorial was organized as three sections. First, the presenters gave an overview of the recent development of GNNs and its applications. Second, they introduced the basic operations of the DGL-Sparse package with an interactive notebook tutorial followed by hands-on quiz to strengthen attendee's understanding. In the last part, they walked-through an end-to-end example of a graph diffusion based GNN implemented in DGL-Sparse.