Summary of Tutorials at The Web Conference 2021

Smriti Bhagat[1], Paul Groth[2], Marinka Zitnik[3], Robert West[4], Francisco M. Couto[5], Pasquale Lisena[6], Albert Meroño-Peñuela[7], Xiangyu Zhao[8], Wenqi Fan[9], Dawei Yin[10], Jiliang Tang[8], Linjun Shou[11], Ming Gong[11], Jian Pei[12], Xiubo Geng[11], Xingjie Zhou[11], Daxin Jiang[11], Benjamin Ricaud[4], Nicolas Aspert[4], Volodymyr Miz[4], Jennifer Dy[13], Stratis Ioannidis[13], İlkay Yıldız[13], Rezvaneh Rezapour[14], Samin Aref[15], Ly Dinh [14], Jana Diesner[14], Alexey Drutsa[16], Dmitry Ustalov[16], Nikita Popov[16], Daria Baidakova[16], Shubhanshu Mishra[17], Arjun Gopalan[18], Da-Cheng Juan[18], Cesar Ilharco Magalhaes[18], Chun-Sung Ferng[18], Allan Heydon[18], Chun-Ta Lu[18], Philip Pham[18], George Yu[18], Yicheng Fan[18], Yueqi Wang[18], Florian Laurent[32], Yanick Schraner[20], Christian Scheller[20], Sharada Mohanty[19], Jiawei Chen[21], Xiang Wang[22], Fuli Feng[22], Xiangnan He[21], Irene Teinemaa[23], Javier Albert[23], Dmitri Goldenberg[23], Flavian Vasile[24], David Rohde[24], Olivier Jeunen[25], Amine Benhalloum[24], Otmane Sakhi[24,26], Yu Rong[27], Wenbing Huang[28], Tingyang Xu[27], Yatao Bian[27], Hong Cheng[29], Fuchun Sun[28], Junzhou Huang[27], Shobeir Fakhraei[30], Christos Faloutsos[31], Onur Çelebi[1], Martin Müller[4], Manuel Schneider[33], Olesia Altunina[4], Wolfram Wingerath[34], Benjamin Wollmer[34,35], Felix Gessert[34], Stephan Succo[34], Norbert Ritter[35], Evann Courdier[4], Tudor Mihai Avram[36], Dragan Cvetinovic[36], Levan Tsinadze[36], Johny Jose[36], Rose Howell[36], Mario Koenig[36], Michaël Defferrard[4], Krishnaram Kenthapadi[30], Ben Packer[37], Mehrnoosh Sameki[38], Nashlie Sephus[30]

[1]Facebook, [2]University of Amsterdam, [3]Harvard University, [4]EPFL, [5]LASIGE, Faculdade de Ciências, Universidade de Lisboa, [6]EURECOM, [7]King's College London, [8]Michigan State University, [9]The Hong Kong Polytechnic University, [10]Baidu Inc., [11]Microsoft Software Technology Center Asia, [12]Simon Fraser University, [13]Northeastern University, [14]School of Information Sciences, University of Illinois at Urbana-Champaign, [15]Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, [16]Yandex, [17]Twitter, [18]Google Research, [19]Aicrowd, [20]FHNW, [21]University of Science and Technology of China, [22]National University of Singapore, [23]Booking.com, [24]Criteo AI Lab, [25]Adrem Data Lab, University of Antwerp, [26]ENSAE-CREST, [27]Tencent AI Lab, [28]Tsinghua University, [29]The Chinese University of Hong Kong, [30]Amazon AWS AI, [31]CMU, [32]AIcrowd, [33]ETH Zurich, [34]Baqend, [35]University of Hamburg, [36]eyeo GmbH, [37]Google, [38]Microsoft

ABSTRACT

This report summarizes the 23 tutorials hosted at The Web Conference 2021: nine lecture-style tutorials and 14 hands-on tutorials.

ACM Reference Format:

Smriti Bhagat et al. 2021. Summary of Tutorials at The Web Conference 2021. In *Companion Proceedings of The Web Conference 2021 (WWW '21 Companion), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3442442.3453701

INTRODUCTION

To bridge the gap between research and real-world applications, The Web Conference 2021 is hosting nine lecture-style tutorials and 14 hands-on tutorials, for a total of 23 tutorials.

Lecture-style tutorials cover the state of the art of research, development, and applications in a specific Web-related area, and stimulate and facilitate future work. This includes tutorials on interdisciplinary directions, bridging scientific research and applied communities, novel and fast-growing directions, and significant applications.

Hands-on tutorials feature in-depth hands-on training on cutting edge systems and tools of relevance to the Web Conference community and are targeted at novice as well as moderately skilled users, with a focus on providing hands-on experience to the attendees.

All tutorials are part of the main conference technical program and are available free of charge to the attendees of the conference. Half-day tutorials last 3–4 hours, full-day tutorials last 7 hours.

In the remainder of this report, each tutorial is summarized in one section. Sections 1–9 describe lecture-style tutorials; Sections 10–23 describe hands-on tutorials.

1 DEEP RECOMMENDER SYSTEM: FUNDAMENTALS AND ADVANCES

https://deeprs-tutorial.github.io

Type: Lecture-style tutorial (half-day)

Organizers: Xiangyu Zhao, Wenqi Fan, Dawei Yin, and Jiliang Tang

WWW '21 Companion, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8313-4/21/04.
https://doi.org/10.1145/3442442.3453701

Abstract: Recommender systems have become increasingly important in our daily lives since they play an important role in mitigating the information overload problem, especially in many user-oriented online services. Recommender systems aim to identify a set of objects (i.e., items) that best match users' explicit or implicit preferences, by utilizing the user and item interactions to improve the matching accuracy. With the fast advancement of deep neural networks (DNNs) in the past few decades, recommendation techniques have achieved promising performance. However, most existing DNNs based methods suffer some drawbacks in practice. More specifically, they consider the recommendation procedure as a static process and make recommendations following a fixed greedy strategy; the majority of existing DNNs based recommender systems are based on hand-crafted hyper-parameters and deep neural network architectures; and they treat each interaction as a separate data instance and overlooks the relations among instances. In this tutorial, we aim to give a comprehensive survey on the recent progress of advanced techniques in solving the above problems in deep recommender systems, including Deep Reinforcement Learning (DRL), Automated Machine Learning (AutoML), and Graph Neural Networks (GNNs). In this way, we expect researchers from the three fields can get deep understanding and accurate insight into the spaces, stimulate more ideas and discussions, and promote developments of technologies in recommendations.

2 SCALING OUT NLP APPLICATIONS TO 100+ LANGUAGES

https://languagescaling.github.io/

Type: Lecture-style tutorial (half-day)

Organizers: Linjun Shou, Ming Gong, Jian Pei, Xiubo Geng, Xingjie Zhou, and Daxin Jiang

Abstract: Natural Language Processing models have achieved impressive performance, thanks to the recent deep learning approaches. However, large deep learning models typically rely on huge amounts of human labeled data. There are more than 7,000 languages spoken in the world. Unfortunately, most languages have very limited linguistic resources. Language scaling is invaluable to the advance of social welfare, and thus has attracted intensive interest from industrial practitioners who want to deploy their applications/services to global markets. At the same time, due to the huge differences in the vocabulary, morphology and syntax among different languages, scaling out NLP applications to various languages presents grand challenges to machine learning, data mining, and natural language processing.

3 LEARNING FROM COMPARISONS

https://neu-spiral.github.io/LearningFromComparisons/

Type: Lecture-style tutorial (half-day)

Organizers: Jennifer Dy, Stratis Ioannidis, and İlkay Yıldız

Abstract: Class labels generated by humans are often noisy, as data collected from multiple experts exhibit inconsistencies across labelers. To ameliorate this effect, one approach is to ask labelers to compare or rank samples instead: when class labels are ordered, a labeler presented with two or more samples can rank them w.r.t. their relative order, as induced by class membership. Comparisons are more informative than class labels, as they capture both inter- and intra-class relationships. In addition, comparison labels are often subject to reduced variability in practice. Nevertheless, learning from comparisons poses computational challenges regressing rankings features is a computationally intensive task. Learning from rankings of sample subsets of size K corresponds to inference over $O(N^K)$ labels. This requires significantly improving the performance of, e.g., maximum likelihood estimation (MLE) algorithms over such datasets. Collecting rankings is also labor intensive, and active learning algorithms need to account for the $O(N^K)$ size of potential queries. This tutorial reviews classic and recent approaches to tackle the problem of learning from comparisons and, more broadly, learning from ranked data. Particular focus will be paid to the ranking regression setting, whereby rankings are to be regressed from sample features. In particular, it covers both parametric and non-parametric models, maximum likelihood estimation and spectral algorithms, ranking regression and variational inference, sample complexity guarantees, and active learning.

4 BIAS ISSUES AND SOLUTIONS IN RECOMMENDER SYSTEM

https://lds4bias.github.io

Type: Lecture-style tutorial (half-day)

Organizers: Jiawei Chen, Xiang Wang, Fuli Feng, and Xiangnan He

Abstract: Recommender systems (RS) have demonstrated great success in information seeking. Recent years have witnessed a large number of work on inventing recommendation models to better fit user behavior data. However, user behavior data is observational rather than experimental. This makes various biases widely exist in the data, including but not limited to selection bias, position bias, exposure bias. Blindly fitting the data without considering the inherent biases will result in many serious issues, e.g., the discrepancy between offline evaluation and online metrics, hurting user satisfaction and trust on the recommendation service, etc. To transform the large volume of research models into practical improvements, it is highly urgent to explore the impacts of the biases and develop debiasing strategies when necessary. Therefore, bias issues and solutions in recommender systems have drawn great attention from both academic and industry. In this tutorial, we aim to provide an systemic review of existing work on this topic. We will introduce seven types of biases in recommender system, along with their definitions and characteristics; review existing debiasing solutions, along with their strengths and weaknesses; and identify some open challenges and future directions. We hope this tutorial could stimulate more ideas on this topic and facilitate the development of debiasing recommender systems.

5 UPLIFT MODELING: FROM CAUSAL INFERENCE TO PERSONALIZATION

https://booking.ai/uplift-modeling-f9759e3fb51e

Type: Lecture-style tutorial (half-day)

Organizers: Irene Teinemaa, Javier Albert and Dmitri Goldenberg

Abstract: 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 to select 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 at Booking.com and other industry leaders, and discuss challenges in implementing these models in production.

6 ADVANCED DEEP GRAPH LEARNING: DEEPER, FASTER, ROBUSTER, UNSUPERVISED

https://ai.tencent.com/ailab/ml/WWW-Deep-Graph-Learning.html

Type: Lecture-style tutorial (half-day)

Organizers: Yu Rong, Wenbing Huang, Tingyang Xu, Yatao Bian, Hong Cheng, Fuchun Sun, and Junzhou Huang

Abstract: Many real data come in the form of non-grid objects, i.e. graphs, from social networks to molecules. Adaptation of deep learning from grid-like data (e.g. images) to graphs has recently received unprecedented attention from both machine learning and data mining communities, leading to a new cross-domain field—Deep Graph Learning (DGL). Instead of painstaking feature engineering, DGL aims to learn informative representations of graphs in an end-to-end manner. It has exhibited remarkable success in various tasks, such as node/graph classification, link prediction, etc. Whilst several previous tutorials have been made for the introduction of Graph Neural Networks (GNNs) in TheWebConf, seldom is there focus on the expressivity, trainability, and generalization of DGL algorithms. To make it more prevailing and advanced, this tutorial mainly covers the key achievements of DGL in recent years. Specifically, we will discuss four essential topics, that is, how to design and train deep GNNs in an efficient manner, how to adopt GNNs to cope with large-scale graphs, the adversarial attack on GNNs, and the unsupervised training of GNNs. Meanwhile, we will introduce the applications of DGL towards various domains, including but not limited to drug discovery, computer vision, and social network analysis.

7 GRAPH MINING AND MULTI-RELATIONAL LEARNING: TOOLS AND APPLICATIONS

https://graph-mining-tutorial.github.io/www2021/

Type: Lecture-style tutorial (half-day)

Organizers: Shobeir Fakhraei and Christos Faloutsos

Abstract: Given a large graph, like who-buys-what, which is the most important node? How can we find communities? If the nodes have attributes (say, gender, or, eco-friendly, or fraudster), and we know the values of interest for a few nodes, how can we guess the attributes of the rest of the nodes? Graphs naturally represent a host of processes including interactions between people on social or communication networks, links between webpages on the World Wide Web, interactions between customers and products, relations between products, companies, and brands, relations between malicious accounts, and many others. In such scenarios, graphs that model real-world networks are typically heterogeneous, multi-modal, and multi-relational. With the availability of more varieties of interconnected structured and semi-structured data, the importance of leveraging the heterogeneous and multi-relational nature of networks in being able to effectively mine and learn this kind of data is becoming more evident. In this tutorial, we present time-tested graph mining algorithms (PageRank, HITS, Belief Propagation, METIS), as well as their connection to Multi-relational Learning methods. We cover both traditional, plain graphs, as well as heterogeneous, attributed graphs. Our emphasis is on the intuition behind these tools, with only pointers to the theorems behind them. The tutorial will include many examples from settings of direct interest to the Web Conference community (e.g., social networks, recommender systems, and knowledge graphs).

8 GOING FOR SPEED: FULL-STACK PERFORMANCE ENGINEERING IN MODERN WEB-BASED APPLICATIONS

https://www2021.app.baqend.com/

Type: Lecture-style tutorial (half-day)

Organizers: Wolfram Wingerath, Benjamin Wollmer, Felix Gessert, Stephan Succo, and Norbert Ritter

Abstract: Loading times are key in modern Web-based applications, because customer satisfaction and business success critically depend on the time that users have to spend waiting. But despite continuous technological advances on both the server and the client side, three developments on the Web are making fast page loads increasingly difficult to achieve. First, user demands have been rising continuously and are therefore more challenging to meet than ever before. Second, users are often not only distributed across the globe, but also predominantly relying on mobile devices with limited processing and network resources. Third, today's high degree of personalization renders traditional caching mechanisms infeasible and thereby impedes fast content delivery. Designing and implementing fast Web-based applications has consequently become a complex task that requires expertise in a variety of fields.

This tutorial presents an end-to-end discussion of latency in modern Web-based application stacks, reviewing research and engineering best practices ranging from data management over application development to user monitoring and data analytics. Our tutorial starts with a primer on why Web performance plays such a critical role for user satisfaction today and in which ways it affects business-critical metrics such as conversion rate or overall revenue. We then dissect different two- and three-tier architectures to uncover where the performance bottlenecks are located in modern Web-based application stacks, how they can be measured effectively, and what the state of the art has to offer for resolving them. A guest speaker from Google will further present a primer on the Core Web Vitals to highlight Google's perspective on web performance and its relevance for business owners everywhere. We close with a synoptic discussion of open challenges and a trajectory of possible future developments.

9 RESPONSIBLE AI IN INDUSTRY: PRACTICAL CHALLENGES AND LESSONS LEARNED

https://sites.google.com/view/ResponsibleAITutorial

Type: Lecture-style tutorial (half-day)

Organizers: Krishnaram Kenthapadi, Ben Packer, Mehrnoosh Sameki, and Nashlie Sephus

Abstract: Artificial Intelligence is increasingly being used in decisions and processes that are critical for individuals, businesses, and society, especially in areas such as hiring, lending, criminal justice, healthcare, and education. Recent ethical challenges and undesirable outcomes associated with AI systems have highlighted the need for regulations, best practices, and practical tools to help data scientists and ML developers build AI systems that are secure, privacy-preserving, transparent, explainable, fair, and accountable – to avoid unintended and potentially harmful consequences and compliance challenges.

In this tutorial, we will present an overview of responsible AI, highlighting model explainability, fairness, and privacy in AI, key regulations/laws, and techniques/tools for providing understanding around AI/ML systems. Then, we will focus on the application of explainability, fairness assessment/unfairness mitigation, and privacy techniques in industry, wherein we present practical challenges/guidelines for using such techniques effectively and lessons learned from deploying models for several web-scale machine learning and data mining applications. We will present case studies across different companies, spanning many industries and application domains. Finally, based on our experiences in industry, we will identify open problems and research directions for the Web Conference community.

10 EXPLORING BIOMEDICAL WEB RESOURCES USING SHELL SCRIPTING

http://labs.rd.ciencias.ulisboa.pt/book/WWW21.html

Type: Hands-on tutorial (full-day)

Organizers: Francisco M. Couto

Abstract: Exploring the vast amount of rapidly growing biomedical content available on the web is of utmost importance, but is also particularly challenging due to the very specialized domain knowledge. This hands-on tutorial will explain how to retrieve and process biomedical data and text using shell scripting with minimal software dependencies. The tutorial will also describe how to explore the semantics encoded in biomedical ontologies and how they address the issue of ambiguity of natural language and contextualization of biomedical entities. The tutorial will follow the examples described in the open access book "Data and Text Processing for Health and Life Sciences", including various steps that Health and Life specialists may have to perform to find and retrieve biomedical text about biomedical entities, e.g. caffeine, using publicly available web resources. This is an introductory tutorial, thus no expected prerequisite knowledge and experience in bioinformatics, text mining and ontologies is required. The participants should however have basic experience in shell scripting and pattern matching.

11 SWAPI: SPARQL ENDPOINTS AND WEB API

https://d2klab.github.io/swapi-thewebconf21/

Type: Hands-on tutorial (half-day)

Organizers: Pasquale Lisena and Albert Meroño-Peñuela

Abstract: The success of Semantic Web technology has boosted the publication of Knowledge Graphs in the Web of Data, and several technologies to access them have become available covering different spots in the spectrum of expressivity: from the highly expressive SPARQL to the controlled access of Linked Data APIs, with GraphQL in between. Many of these technologies have reached industry-grade maturity. Finding the trade-offs between them is often difficult in the daily work of developers, interested in quick API deployment and easy data ingestion. In this tutorial, we will cover this in-between technology space, with the main goal of providing strategies and tools for publishing Web APIs that ensure the easy consumption of data coming from SPARQL endpoints. Together with an overview of state-of-the-art technologies, the tutorial focuses on two novel technologies: SPARQL Transformer, which allows to get a more compact JSON structure for SPARQL results, decreasing the effort required by developers in interfacing JavaScript and Python applications; and grlc, an automatic way of building APIs on top of SPARQL endpoints by sharing queries on collaborative platforms. Moreover, we will present recent developments to combine the two, offering a complete resource for developers and researchers. Hands-on sessions will be proposed to internalize those concepts with practical exercises.

12 LARGE SCALE GRAPH MINING: VISUALIZATION, EXPLORATION, AND ANALYSIS

https://lts2.epfl.ch/reproducible-research/graph-exploration/

Type: Hands-on tutorial (full-day)

Organizers: Benjamin Ricaud, Nicolas Aspert, and Volodymyr Miz

Abstract: What happens inside social networks impacts our everyday life and is of high interest for researchers, data journalists and the general public. These networks, as well as other large online networks of pages or knowledge graphs, contain a rich but overwhelming amount of information. Due to their size and the limited API access, the extraction and analysis of information within these huge networks are challenging. In this hands-on tutorial, we propose an introduction to the data mining of large networks and the analysis of activity inside them. The tutorial is made of two parts. The first one is an overview of key concepts in (large) graph analysis, an introduction to the main exploration tools in Python and visualization using Gephi as well as a short introduction to machine learning on graphs. It covers a basic set of important tools to start exploring large graphs. During the second part, participants will form teams and focus on a particular large real-world graph either proposed by the organizers or by the participants themselves. The exploration will be guided, alternating short presentations of techniques for the exploration of large networks, using APIs, and interactions of the organizers with the teams.

13 PYNETWORKSHOP: ANALYZING THE STRUCTURE OF NETWORKS IN PYTHON - THE ESSENTIALS, SIGNED NETWORKS, AND NETWORK OPTIMIZATION

https://publish.illinois.edu/pynetworkshop/

Type: Hands-on tutorial (full-day)

Organizers: Rezvaneh Rezapour, Samin Aref, Ly Dinh, and Jana Diesner

Abstract: PyNetworkshop is a hands-on tutorial on using network libraries in Jupyter for analyzing the structure of social networks. Social network analysis is a longstanding methods toolbox used to examine the structures of relations between social entities, which can represent individuals, groups, or organizations, among other entity types. After covering general preliminaries and essentials, this tutorial focuses on different methods for analyzing the structure of signed directed networks. Existing network metrics and models are flexible in that they can detect structural dynamics that exist at three fundamental levels of analysis, namely the micro, meso, and macro levels of networks. While several open-source tools for analyzing networks are available for Python, there is a need for a pipeline that guides scholars through a multilevel analysis of networks. This tutorial is based on recent methodological advancements at the intersection of social network analysis and graph optimization.[1] The intended audience are researchers who use networks or plan to start using networks in their work. We do not assume any prior knowledge other than basic level of mathematics and basic familiarity with Jupyter Python (being able to run "Hello World!" in Jupyter).

[1] https://nature.com/articles/s41598-020-71838-6

14 IMPROVING WEB RANKING WITH HUMAN-IN-THE-LOOP: METHODOLOGY, SCALABILITY, EVALUATION

https://research.yandex.com/tutorials/crowd/www-2021

Type: Hands-on tutorial (full-day)

Organizers: Alexey Drutsa, Dmitry Ustalov, Nikita Popov, and Daria Baidakova

Abstract: Modern Web services widely employ sophisticated Machine Learning techniques to rank news, posts, products, and other items presented to the users or contributed by them. These techniques are usually built on offline data pipelines and use a numerical approximation of the relevance of the demonstrated content. In our hands-on tutorial, we present a systematic view on using Human-in-the-Loop to obtain scalable offline evaluation processes and, in particular, high-quality relevance judgements. We will introduce the ranking problem to the attendees, discuss the commonly used ranking quality metrics, and then focus on Human-in-the-Loop-based approach to obtain relevance judgements at scale. More precisely, we will present a thorough introduction to pairwise comparisons, demonstrate how these comparisons can be obtained using Crowdsourcing, and organize a hands-on practice session in which the attendees will obtain high-quality relevance judgements for search quality evaluation. Finally, we will discuss the obtained relevance judgements, point out directions for further studies, and answer questions asked during the tutorial.

15 INFORMATION EXTRACTION FROM SOCIAL MEDIA: A HANDS-ON TUTORIAL ON TASKS, DATA, & OPEN SOURCE TOOLS

https://socialmediaie.github.io/tutorials/WWW2021/

Type: Hands-on tutorial (half-day)

Organizers: Shubhanshu Mishra, Rezvaneh Rezapour, and Jana Diesner

Abstract: In this hands-on tutorial, we introduce the participants to working with social media data, which are an example of Digital Social Trace Data (DSTD). The DSTD abstraction allows us to model social media data with rich information associated with social media text, such as authors, topics, and time stamps. We introduce the participants to several Python-based, open-source tools for performing Information Extraction (IE) on social media data. Furthermore, the participants will be familiarized with a catalogue of more than 30 publicly available social media corpora for various IE tasks such as named entity recognition (NER), part of speech (POS) tagging, chunking, super sense tagging, entity linking, sentiment classification, and hate speech identification. Finally, the participants will be introduced to the following applications of extracted information: a) combining network analysis and text-based signals to rank accounts, and b) correlation between sentiment and user-level attributes in existing corpora. The tutorial aims to serve the following use cases for social media researchers: a) high accuracy IE on social media text via multitask and semi-supervised learning,

including the recent transformer based tools, b) rapid annotation of new data for text classification via active human-in-the-loop learning, c) temporal visualization of the communication structure in social media corpora via social communication temporal graph visualization technique, and d) detecting and prioritizing needs during crisis events (e.g, COVID19).

16 NEURAL STRUCTURED LEARNING: TRAINING NEURAL NETWORKS WITH STRUCTURED SIGNALS

https://www.tensorflow.org/neural_structured_learning

Type: Hands-on tutorial (half-day)

Organizers: Arjun Gopalan, Da-Cheng Juan, Cesar Ilharco Magalhaes, Chun-Sung Ferng, Allan Heydon, Chun-Ta Lu, Philip Pham, George Yu, Yicheng Fan, and Yueqi Wang

Abstract: We present Neural Structured Learning (NSL), a new learning paradigm to train neural networks by leveraging structured signals in addition to feature inputs. Structure can be explicit as represented by a graph, or implicit, either induced by adversarial perturbation or inferred using techniques like embedding learning. Structured signals are commonly used to represent relations or similarity among samples that may be labeled or unlabeled. So, leveraging these signals during neural network training harnesses both labeled and unlabeled data, which can improve model accuracy, particularly when the amount of labeled data is relatively small. Additionally, models trained with samples that are generated by adding adversarial perturbation have been shown to be robust against malicious attacks, which are designed to mislead a model's prediction or classification. NSL generalizes to both Neural Graph Learning as well as Adversarial Learning.

Neural Structured Learning is open-sourced on GitHub and is part of the TensorFlow ecosystem. The NSL website contains the theoretical foundations of the technology, API documentation, and hands-on tutorials. NSL is widely used in Google across many products and services.

Our tutorial will cover several aspects of Neural Structured Learning with an emphasis on two techniques – *graph regularization* and *adversarial regularization*. In addition to using interactive hands-on tutorials that demonstrate the NSL framework and APIs in TensorFlow, we also plan to have short presentations that accompany them to provide additional motivation and context. Finally, we will discuss some recent research in areas related to Neural Structured Learning. Topics here include using graphs for learning embeddings and several advanced models of graph neural networks. This will demonstrate the generality of the Neural Structured Learning framework as well as open doors to future extensions and collaborations with the community.

17 FLATLAND: MULTI-AGENT REINFORCEMENT LEARNING ON TRAINS

https://yanickschraner.github.io/rl-on-trains-workshop/

Type: Hands-on tutorial (full-day)

Organizers: Florian Laurent, Yanick Schraner, Christian Scheller, Manuel Schneider, and Sharada Mohanty

Abstract: This workshop investigates a real-world problem: how to schedule train traffic? This is a challenging problem. Railway networks are growing fast. The decision-making methods commonly used to schedule trains are starting to show their limits. How can we solve this problem? With machine learning, of course! In this workshop, we will use reinforcement learning to tackle this challenge. This is a real research problem on which we have been working for the past 2 years in collaboration with the national railway companies from Switzerland, Germany and France (SBB, Deutsche Bahn, SNCF). You will discover what reinforcement learning is, what it can do, and its current limitations and perspectives. They will get hands-on experience by building and tweaking railway agents and competing against each other to build the best solutions. In the web domain, news recommendation, online web systems auto-configuration and online advertising real-time bidding are practical applications for reinforcement learning. It is further suitable for simulating multi-user behavior in complex web applications like social media platforms to automatically test such platforms.

18 RECOMMENDER SYSTEMS THROUGH THE LENS OF DECISION THEORY

https://sites.google.com/view/recsys-as-decision-theory

Type: Hands-on tutorial (half-day)

Organizers: Flavian Vasile, David Rohde, Olivier Jeunen, Amine Benhalloum, and Otmane Sakhi

Abstract: Decision theory is a 100 year old science that explicitly separates states of nature, decision rules, utility functions and models to address the universal problem of decision making under uncertainty. In the context of recommender systems, this separation allows us to formalise different approaches to learning from bandit feedback. Policy approaches use an inverse propensity score estimator and directly optimise a decision rule that maps the user context to a recommendation. In contrast, value-based approaches use bandit feedback to learn a model of the reward, and considering the appropriate decision rule as a separate step. This tutorial uses the richer language of decision theory to present policy- and value-based methods in a common framework. With extensive examples we explore how these methods can be applied to recommendation problems, emphasising on situations with low probability of reward and very large action spaces. We offer side-by-side comparisons between these methods outlining their strengths and weaknesses, such as estimator variance, model mis-specification, tractability and ease-of-use. By identifying the modes of failure for every class, this can provide practical guidelines for future practitioners as to which method to apply in which types of environments. The use of bandit feedback to improve recommender system performance has become a linchpin of modern recommendation. This tutorial unifies the major classes of methods, providing a thorough overview of an actual and important topic.

19 BUILDING AN EFFICIENT TEXT CLASSIFIER FROM THE GROUND UP

https://fasttext.cc/

Type: Hands-on tutorial (half-day)

Organizers: Onur Çelebi

Abstract: Text classification is one of the most commonly studied tasks in natural language processing. In addition to its theoretical importance, the classification task is widely used in various applications including spam detection, sentiment analysis, language identification. FastText is an open-source library developed by Facebook AI Research (FAIR), that has the purpose of simplifying text classification. As often, the "secret sauce" is in the details. This tutorial will give participants more insights about these details, helping them to understand the subtleties of the model. Participants will also learn how to use a trained classifier to run predictions in the browser with web- assembly.

20 CONVERSATIONAL ARTIFICIAL INTELLIGENCE: CAN YOUR AI BEAT THE TURING TEST?

https://utanashati.github.io/conversational-ai-workshop/

Type: Hands-on tutorial (full-day)

Organizers: Martin Müller, Florian Laurent, Manuel Schneider, and Olesia Altunina

Abstract: In 1950, Alan Turing proposed his famous test to distinguish humans from machines. At the time, he probably didn't think workshop participants would attempt to beat his test with billion parameter models in real-time. But here we are! This workshop has two parts: In the first half, we will take a deep dive into conversational AI. By mastering a series of small tasks, you will discover what makes state-of-the-art models like GPT-3 so powerful and how you can build your own models. In the second half, we will run a challenge in which you will work on building the most life-like bot possible and test it in a real-life setting. You will also have the chance to evaluate other participants' bots - but with a twist! Every now and then you will actually chat with a real human. Will you be able to tell?

21 INTRODUCTION TO DEEP LEARNING WITH PYTORCH

https://theevann.github.io/webconf-pytorch-workshop/

Type: Hands-on tutorial (half-day)

Organizers: Evann Courdier

Abstract: This half-day workshop is designed for PyTorch beginners and will walk you through the basics of the PyTorch library. We will introduce the basic building blocks (Tensors, Autograd, Optimization) and also cover how to build more advanced models like CNNs in the second part. During the tutorial, we will go through a series of Jupyter notebooks which allows participants to experiment with the code.

22 MACHINE LEARNING-DRIVEN AD BLOCKING: FROM DATA COLLECTION TO DEPLOYMENT

https://eyeo.gitlab.io/machine-learning/www21

Type: Hands-on tutorial (full-day)

Organizers: Tudor Mihai Avram, Dragan Cvetinovic, Levan Tsinadze, Johny Jose, Rose Howell, and Mario Koenig

Abstract: In this hands-on tutorial, we'll give you an introduction to our journey of machine learning-based ad blocking, the research we have performed and our results. Together we'll set up a website crawling service which produces datasets for self-supervised training, we'll transform the crawled data into graphs, train a graph-based machine learning model and deploy it in a browser extension in order to run inference in the browser and block ads. The goal of this session is to give you an overview of how a machine learning project with a similar scope can be tackled and how to develop basic components ranging from data gathering to a simple extension which makes use of the data you trained your ML model with.

23 LEARNING FROM GRAPHS: FROM MATHEMATICAL PRINCIPLES TO PRACTICAL TOOLS

https://github.com/mdeff/learning-from-graphs-webconf2021

Type: Hands-on tutorial (full-day)

Organizers: Michaël Defferrard

Abstract: A graph encodes relations between objects, such as distances between points or hyperlinks between websites. You will learn how to extract information about that relational structure. This information is crucial to characterize an object through its local connectivity or an entire graph through its global connectivity. On top of that structure, a network may possess data about the objects or the relations, such as a point's color or an hyperlink's click-through rate. You will learn how to leverage a graph to analyze this data. Leveraging the structure that underlies data is an important concept, from physical symmetries dictating conservation laws to the efficiency of convolutional neural networks. The tutorial is built on deep mathematical principles but will walk you from the basics with an emphasis on intuitions and working knowledge.