SDSC X DSSG DATA SCIENCE MEETUP

7 JUNE 2018 (THU)
6.30PM-9PM
SMU FUNCTION ROOM
Administration Building Level 4
81 Victoria Street
Singapore 188065

EVENT OUTLINE

Registration – 6:30PM
Networking *(Light refreshments will be provided)*

Introduction– 7PM
Singapore Data Science Consortium
Ms Toh Swee Feng, Caroline
Business Development Manager, Singapore Data Science Consortium

Technical Talk 1 – 7:15PM
**Modeling Preferences from Multi-Modal Data:
A Deep Learning Exploration**
Prof. Hady W. Lauw
*Assistant Professor of Information Systems,
Singapore Management University*

Technical Talk 2 – 8PM
SentiVec: Sentiment-Infused Word Embeddings
Mr. Maksim Tkachenko
PhD Candidate, Singapore Management University

Technical Talk 3 – 8:30PM
**C2PF: A Poisson Latent Factor Model of
User Preferences with Item Context**
Dr. Aghiles Salah
Research Fellow, Singapore Management University

For enquiries, please
contact Huimin at
65164408 or
huimin@nus.edu.sg

Jointly Organised by: **SINGAPORE DATA SCIENCE CONSORTIUM**   Data Science SG

1.   Modeling Preferences from Multi-Modal Data: A Deep Learning Exploration

Abstract: With pervasive digitization, two trends are emerging. One is the proliferation of choices in our consumption of both physical and digital goods and services. The other is the "datafication" of our behaviors, whereby users' preferences increasingly manifest themselves through multiple forms of preference signals, including consumption, ratings, text reviews, social connections, images, etc. The first points to the broad challenge of helping users to navigate the ever-expanding universe of options. The second points to the deep opportunity of leveraging data to model user preferences. While research into recommender systems has been kickstarted by modeling numerical expressions (e.g., ratings), there is now a greater recognition of the need to transcend beyond a single modality. In our journey to incorporate multi-modality, deep learning is one of the vehicles used in our exploration. In this talk, we share our recent experiences with several neural network architectures for learning representations from various modalities for expressing preferences such as visual, sequential, as well as social.

Speaker: Hady W. Lauw is currently Assistant Professor of Information Systems at Singapore Management University, as well as NRF Fellow of the Singapore National Research Foundation. Formerly, he served as postdoctoral researcher at Microsoft Research in Silicon Valley, as well as scientist at A*STAR's Institute for Infocomm Research. He received his PhD from Nanyang Technological University on A*STAR Graduate Scholarship. At SMU, he leads the Preferred.AI research project, whose research activities span data mining and machine learning, focusing on preference analytics and recommender systems. He is also active in the academic community, currently serving as the Chair of the Singapore Chapter of ACM SIGKDD. More information may be found at http://www.hadylauw.com.

2.   SentiVec: Sentiment-Infused Word Embeddings

Abstract: Distributional word embedding methods such as Word2Vec and GloVe have been critical for the success of many large scale natural language processing applications. In this talk, we explore the notion of subjectivity and how the varying levels of subjectivity in input corpora affect word embeddings for text classification (e.g., sentiment, subjectivity, topic). Through systematic comparative analysis, we discover the outsize role that sentiment words play on subjectivity-sensitive tasks, and develop a novel word embedding method, SentiVec, which is infused with sentiment information from a lexical resource and is shown to outperform baselines on such tasks.

Speaker: Maksim Tkachenko is a PhD candidate at Singapore Management University (SMU). He received his diploma in mathematics and software engineering from Saint Petersburg State University, Russia, where afterwards he served as a research engineer. He worked at Hewlett-Packard Laboratories, Russia, on natural language processing topics. At SMU, his research focuses on text mining and natural language processing methods for user preference acquisition. More information may be found at http://www.mtkachenko.info/.

3.   C2PF: A Poisson Latent Factor Model of User Preferences with Item Context

Abstract: Recommender systems are essential in guiding users as they navigate the myriads of options offered by modern applications. Typically, these systems use historical behavioural data (e.g., ratings, clicks or purchases) as the main source of information for learning. The sparsity of such data, however, makes it difficult to model and generalize user's preferences accurately. One promising direction to address this issue is to go beyond user-item interactions and consider auxiliary information, such as user's social network, item's content information, etc. In the same line, I will discuss Collaborative Context Poisson Factorization (C2PF): a recent Bayesian factor model of user preferences which integrates an important aspect into personalized recommendation, namely item relatedness. The underlying assumption in C2PF is that, items that occur within the same "context" (e.g., browsed in the same session, purchased in the same basket) are likely related in some latent aspect that guides the choices one makes, such as specification, functionality, visual appearance, compatibility, etc. By taking into account the item's context, C2PF successfully alleviates the sparsity issue by extending a user's preference to other items of similar aspects.

Speaker: Aghiles Salah is currently a research fellow at Singapore Management University (SMU). His research is in statistical learning, with a particular focus on probabilistic/Bayesian modelling of high dimensional sparse data, such as text and user preferences. Prior to joining SMU, Aghiles was a teaching and research assistant in the Department of Computer Science and Mathematics of Paris Descartes University, France. He received his PhD in Machine Learning and Data Science from Paris Descartes University in 2016. More information may be found at http://saghiles.github.io/.