NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

C S   S E M I N A R

Title: **SEMINAR THEME: NATURAL LANGUAGE PROCESSING IN BIG DATA ANALYTICS**

Speaker: Speaker 1 :
SUN Maosong
Professor & Chair, Department of Computer Science and Technology,
Tsinghua University, China

Speaker 2 :
LIU Yiqun
Associate Professor, Tsinghua University, China

Speaker 3 :
ZHU Xiaoyan
Professor, Tsinghua University, China

Speaker 4 :
ZHANG Yue
Assistant Professor, Singapore University of Technology and Design,
Singapore

Speaker 5 :
LIU Yang
Associate Professor, Tsinghua University, China

Speaker 6 :
LI Juanzi
Professor, Tsinghua University, China

Speaker 7 :
LIU Zhiyuan
Assistant Professor, Tsinghua University, China

Speaker 8 :
ZHAO Jun
Professor, Institute of Automation, Chinese Academy of Sciences, China

Speaker 9 :
KAN Min-Yen
Associate Professor, National University of Singapore, Singapore

Speaker 10 :
LIU Kang,
Associate Professor, Institute of Automation, Chinese Academy of Sciences,
China

Date/Time: 26 November 2015, Thursday, 09:15 AM to 05:15 PM

Venue: Executive Classroom, COM2-04-02

Chaired by: Dr Chua Tat Seng, KITHCT Chair Professor, School of Computing
             (chuats@comp.nus.edu.sg)

Talk 1: 09:30 am
Research Challenges Related To Natural Language Processing

Speaker :
SUN Maosong
Professor & Chair, Department of Computer Science and Technology,
Tsinghua University, China

ABSTRACT :
Several research challenges related to natural language processing are addressed and
discussed, particularly in the context of deep learning. Some suggestions on NLP research
are further proposed.

BIODATA :
Maosong is a professor in the Department of Computer Science and Technology of Tsinghua
University. He is holding numerous key appointments such as the Chair of the Department
of Computer Science and Technology of Tsinghua University, Co-Director of NExT
Research Centre in Tsinghua, Vice President of the Chinese Information Processing Society
of China, and Director of academic committees of a number of Ministry of Education or
Provincial key research labs. He is also the Editor-in-chief of the Journal of Chinese
Information Processing, the only and the most influential Chinese journal in computational
linguistics in China.

Maosong?s research interests are computational linguistics, statistical and corpus-based
natural language processing, computational social sciences, and computational pedagogy. He
has participated as project leader or principal researcher in over 30 projects founded by the
National Natural Science Foundation of China, the National Social Science Foundation of
China, the National High-Tech R&D Program of China, the National Basic Research
Program of China as well as in projects funded by a number of international IT companies

In recent years, he led a Tsinghua University team successfully developed a MOOC
(Massive Online Open Courses) platform, XuetangX, in 2012. Now, XuetangX has become
one of the two most influential Chinese MOOC platforms, and has attracted about 1.7

million registered online learners. Serving as the vice director of Online Education Research Centre of the Ministry of Education of China (MOE) and the director of Massive Online Open Education Research Centre of Tsinghua University, he is playing a critical role in MOOC development in China.

Talk 2 : 10:00 am
Web Search With The Help of Wisdom of Crowds

Speaker :
LIU Yiqun
Associate Professor, Tsinghua University, China

ABSTRACT :
When users interact with information retrieval (IR) systems, they leave rich implicit feedback in the form of clicks, mouse movements etc. This feedback contains valuable information about users and about IR systems. Analyzing and interpreting user interactions and modeling user search behavior has become an important research topic in Web search studies. It enables us to better understand users, perform user simulations, improve search algorithms and build quality metrics. In this talk, I will at first introduce some existing efforts that are paid to understand the cognitive behavior of search users and focus on the characteristics of search users? querying, clicking-through and examination behaviors. After that, I will talk about some recent trends in search behavior modeling. Especially, IR systems become more and more heterogeneous: they deal with information of various media types, structure and semantics; run on multiple devices and support a variety of short- and long-term search tasks; serve users with different background and preferences. We will focus on how to construct user behavior modeling to support effective IR methods in such heterogeneous environments

BIODATA :
Yiqun is an associate professor at the Department of Computer Science and Technology in Tsinghua University. His major research interests is in Web Search, especially in search user behavior modeling, search performance evaluation and Web data quality estimation. He is also a Principle Investigator (PI) of NExT Research Centre in developing technologies for live media search. He has published over 30 papers at top-tier academic conferences and journals such as SIGIR, WWW, CIKM, WSDM, AAAI, IJCAI, ACM TWeb and JIR. He has received the best paper honorable mention in SIGIR 2015. He serves in the editorial board of the Information Retrieval Journal (Springer) and also as the task co-leader for the NTCIR (NII Testbeds and Community for Information access Research) IMine tasks.

Tea Break

Talk 3 : 11:00 am
Representation of Text And Knowledge

Speaker :
ZHU Xiaoyan
Professor, Tsinghua University, China

ABSTRACT:
Learning to represent data, commonly known as representation learning, has been a hot topic in recent research since it avoids heavy feature engineering and feature sparsity. How to represent word, sentence, and knowledge in KB has become fundamental problems in natural language processing. In this talk, the speaker will present a series of research how to represent text and knowledge with neural network models and embedding methods. The research shows that text classification with appropriate sentence representation, domain adaptation with enhanced auto encoders, knowledge graph embedding with Gaussian mixture models, could be largely improved.

BIODATA:
Xiaoyan is the Head of State key Lab of Intelligent Technology and Systems, Tsinghua University. Since 1993, Prof. Zhu has been the faculty member of Department of Computer Science and Technology, Tsinghua University. She received her PhD degree at Nagoya Institute of Technology in 1990.

As a visiting scientist, she spent one year working at University of California, Santa Barbara, in 2002 and half year at Cornell University in 2006. She is International Research Chair of International Development Research Centre, Canada, in Information Technology from 2009. Currently her research interests are focus on intelligent information processing, internet information acquisition, and question and answering system. She has successfully conducted the research supported by National Basic Research Program (973 program), National High Technology Research and Development Program of China (863 program), and National Natural Science Foundation of China (NSFC).

Xiaoyan received Okawa Award in Japan in 2014, Google Research Award in 2012 & 2014, Best Paper Award, COLING 2011, Best Student Paper Award, ACL 2012, and Best Student Paper Award, SDM 2014, respectively.

Talk 4 : 11:30 am
Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing

Speaker :
ZHANG Yue
Assistant Professor, Singapore University of Technology and Design, Singapore

ABSTRACT:
Neural probabilistic parsers are attractive for their capability of automatic feature combination and small data sizes. A transition-based greedy neural parser has given better accuracies over its linear counterpart. We propose a neural probabilistic structured-prediction model for transition-based dependency parsing, which integrates search and

learning. Beam search is used for decoding, and contrastive learning is performed for maximizing the sentence-level log-likelihood. In standard Penn Treebank experiments, the structured neural parser achieves a 1.8% accuracy improvement upon a competitive greedy neural parser baseline, giving performance comparable to the best linear parser.

BIODATA :
Zhang Yue is currently an assistant professor at Singapore University of Technology and Design. Before joining SUTD in July 2012, he worked as a postdoctoral research associate in University of Cambridge, UK. He received his PhD and MSc degrees from University of Oxford, UK, and his BEng degree from Tsinghua University, China.

Zhang Yue?s research interests include natural language processing, machine learning and artificial Intelligence. He has been working on statistical parsing, parsing, text synthesis, machine translation, sentiment analysis and stock market analysis intensively. Zhang Yue serves as the reviewer for top journals such as Computational Linguistics, Transaction of Association of Computational Linguistics and Journal of Artificial Intelligence Research. He is also PC member for conferences such as ACL, COLING, EMNLP, NAACL, EACL, AAAI and IJCAI. Recently, he was the area chairs of COLING 2014, NAACL 2015 and EMNLP 2015.

Talk 5 : 12:00 noon
Learning Translation Models From Non-Parallel Data

Speaker :
LIU Yang
Associate Professor, Tsinghua University, China

ABSTRACT:
While parallel corpora are an indispensable resource for data-driven multilingual natural language processing tasks such as machine translation, they are limited in quantity, quality and coverage. As a result, learning translation models from non-parallel corpora has become increasingly important nowadays, especially for low-resource languages. In this work, we propose a joint model for iteratively learning parallel lexicons and phrases from nonparallel corpora. The model is trained using a Viterbi EM algorithm that alternates between constructing parallel phrases using lexicons and updating lexicons based on the constructed parallel phrases. Experiments on Chinese-English datasets show that our approach learns better parallel lexicons and phrases and improves translation performance significantly.

BIODATA :
Liu Yang is an associate professor in the Computer Science and Technology Department at Tsinghua University. He graduated in 2007 from Institute of Computing Technology, Chinese Academy of Sciences. His work has focused on natural language processing and machine translation. He has published over 30 papers in leading NLP/AI journals and conferences and received a COLING/ACL 2006 Meritorious Asian NLP Paper Award. He served as the Tutorial Co-Chair for ACL 2014 and the Local Arrangement Co-Chair for

ACL 2015.

LUNCH

Talk 6 : 14:00 pm
Event Knowledge Learning

Speaker :
LI Juanzi
Professor, Tsinghua University, China

ABSTRACT:
In this talk, we aim to present our recent work on news event knowledge learning. With the development of the Internet, people have become overwhelmed by the increasingly large volume of online news documents. Despite that considerable work has been conducted on event extraction, event detection and tracking, which facilitates online news reading, little attention is paid to event knowledge learning from similar events and event knowledge base construction. Building a news event knowledge base is imperative not only for event knowledge representation, storage and acquisition, but also for users' understanding of events. We first introduce our work on single event knowledge learning, i.e., event topic mining and analysis, and then introduce our work on knowledge learning from similar events. Finally, we present our future work on event knowledge base construction.

BIODATA :
Juanzi is a professor in Tsinghua University. She obtained her PhD degree from Tsinghua University in 2000. Her main research interest is to study the semantic technologies by combining the key technologies of Natural Language Processing, Semantic Web and Data Mining.

She is the Vice Director of Chinese Information Processing Society of Chinese Computer Federation in China. She is also principal investigators of many key projects supported by Natural Science Foundation of China (NSFC), national basic science research program and international cooperation projects. Juanzi has published over 90 papers in many international journals and conferences such as TKDE, SIGIR, SIGMOD, SIGKDD, IJCAI, et al.

Talk 7 : 14:30 pm
Representation Learning for Large-Scale Knowledge Graphs

Speaker :
LIU Zhiyuan
Assistant Professor, Tsinghua University, China

ABSTRACT :
Knowledge graphs organize human knowledge about the world in a structured form. In a

typical knowledge base, entities are connected by multiple relations. Knowledge bases are playing an important role in most tasks in natural language processing and information retrieval. Recent years have witnessed the significant advances of distributed representation of knowledge graphs, which exhibits powerful capability in both relation extraction and knowledge inference. In this talk, I will introduce recent advances of representation learning of large-scale knowledge graphs, and outlook its research challenges and trends.

BIODATA :
Zhiyuan is an assistant research fellow in Department of Computer Science and Technology, Tsinghua University. He is interested in representation learning in NLP and social computation. He has published 20+ papers in top tier conferences in NLP and AI. He has been awarded as Excellent Doctoral Dissertation of Tsinghua University, Excellent Doctoral Dissertation of Chinese Association for Artificial Intelligence, and Excellent Postdoctoral Fellow Award of Tsinghua University. He developed user modeling system on 5 popular Chinese social media services including Weibo, and got over 3.5 million registered users with over 30 million visits.

Talk 8 : 15:00 pm
Question Answering Over Knowledge Graph

Speaker :
ZHAO Jun
Professor, Institute of Automation, Chinese Academy of Sciences, China

ABSTRACT :
Question answering over knowledge graph (KG) is a kind of deep question answering. There are mainly two kinds of approaches for QA over KG. In the symbolic logic approach, the entities and relations in KG are represented as symbols and structures, and we conduct semantic parsing to the natural language questions to generate logic semantic formulas comprised of the entities and relations in KG. In the representation learning approach, both the entities and relations in KG and the natural language questions are represented by numeric values, and answers are obtained through numeric computation. This talk will firstly give a simple survey of the two kinds of approaches, and then introduce some researches of the National Laboratory of Pattern Recognition, Chinese Academy of Sciences.

BIODATA :
Zhao Jun is a professor at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. He received his PhD degree from Tsinghua University in 1998. Before joining NLPR in 2002, he worked in the Hong Kong University of Science and Technology as a postdoctoral research fellow.

His current research focuses on natural language processing, information extraction and question answering. Prof Zhao has published over 50 peer-reviewed papers in the prestigious conferences and journals, including ACL, SIGIR, TKDE, JLMR, etc. He obtained best paper award of COLING-2014.

Afternoon Tea Break

Talk 9 : 16:00 pm
#mytweet via Instagram: Exploring User Behavior Across Multiple Social Networks

Speaker :
KAN Min-Yen
Associate Professor, National University of Singapore, Singapore

ABSTRACT :
We study how users of multiple online social networks (OSNs) employ and share information by studying a common user pool that use six OSNs ? Flickr, Google+, Instagram, Tumblr, Twitter, and YouTube. We analyze the temporal and topical signature of users? sharing behavior, showing how they exhibit distinct behavioral patterns on different networks. We also examine cross-sharing (i.e., the act of user broadcasting their activity to multiple OSNs near-simultaneously), a previously unstudied behavior and demonstrate how certain OSNs play the roles of originating source and destination sinks.

BIODATA :
Min is an associate professor at the National University of Singapore. He serves the School as an Assistant Dean of Undergraduate Studies. Min is a member of the executive committee of the Association of Computational Linguistics (ACL) and maintains the ACL Anthology, the community's largest archive of published research. He is an associate editor for the Springer "Information Retrieval" journal. His research interests include digital libraries and applied natural language processing and information retrieval. Specific projects include work in the areas of scientific discourse analysis, full-text literature mining, machine translation, lexical semantics and applied text summarization.

Talk 10 : 16:30 pm
Opinion Relation Identification: Extracting Opinion Targets/Words from Online Product Reviews

Speaker :
LIU Kang
Associate Professor, Institute of Automation, Chinese Academy of Sciences, China

ABSTRACT :
Mining opinion targets and opinion words are important tasks for fine-grained opinion mining. The one of the key components is to identify opinion relations among words. However, this task is not easy, especially from online product reviews, because there are too many informal writing styles, including grammatical errors, typo- graphical errors, and punctuation errors. This talk will briefly introduces our recent solutions of identifying opinion relations among words in informal texts, including syntactic pattern learning, exploiting alignment model, employing neural network, and so on. Moreover, it will show what we discover when the size of reviews increases. Finally, it will discuss whether only

exploiting opinion relations is enough for extracting opinion targets/words.

BIODATA :
Liu Kang received his Ph.D. from Institute of Automation, Chinese Academy of Sciences in 2010. Currently, he worked as an associate professor in Institute of Automation, Chinese Academy of Sciences. His current research interests include Natural Language Processing, Information Extraction, Question Answering and Opinion Mining etc.

He has authored/co-authored more than 30 papers in leading journals and conferences, including TKDE, JMLR, ACL, COLING, IJCAI, EMNLP, CIKM, AAAI etc. He won several honors and awards including: runner-up in KDD-CUP Track 2 in 2011, COLING 2014 Best Paper Award, ?CCF-Tencent Xiniuniao? Excellence Award (The top award) in 2014, Hanwang Youth Innovation Excellence Award by Chinese Information Processing Society of China in 2014, Google Focused Research Award in 2015.

CLOSING