

自然语言处理报告会

会议时间：2018 年 11 月 15 日（星期四）下午

会议地点：北京大学二教 515 教室

会议日程：

13:30-14:15	王厚峰 教授 (北京大学信息科学技术学院)
	自然语言处理中的数据与算法
14:15-15:00	穗志方 教授 (北京大学信息科学技术学院)
	文本挖掘与知识图谱
15:00-15:45	万小军 研究员 (北京大学计算机科学技术研究所)
	机器写作技术与应用
15:45-16:30	孙 栩 研究员 (北京大学信息科学技术学院)
	自然语言生成关键技术研究
16:30-17:15	冯岩松 副教授 (北京大学计算机科学技术研究所)
	<i>Combining Human Knowledge with Neural Network Models: A Case Study for SLU and Beyond</i>
17:15-18:00	严 睿 研究员 (北京大学计算机科学技术研究所)
	人工智能在机对话系统中的技术现状与挑战

会议摘要信息

报告一

自然语言处理中的数据与算法

王厚峰

摘要：自然语言处理的很多应用需要建立在“理解”的基础上，无论是算法上，还是数据，对自然语言的理解都存在不少制约。报告结合问答和对话，分析了自然语言处理面临的困难，特别是数据问题和算法问题导致的困难。

简介：王厚峰，北京大学计算语言学研究所所长、教授，从事自然语言处理相关的工作 20 多年，近年的研究包括问答、对话、语篇分析，发表论文 70 多篇。曾经作为首席专家负责过国家 863 项目，作为首席专家负责过国家社科重大项目，主持国家自然科学基金面上或重大研究计划项目 7 项。



报告二

文本挖掘与知识图谱

穗志方

摘要：在人工智能时代，知识图谱是智能化系统的核心基石，而文本挖掘技术又是构建大规模知识图谱的关键技术。本报告介绍课题组围绕文本挖掘与知识图谱这个研究方向，进行的一系列科学研究和技术开发，具体包括：知识图谱补全技术、文本关系提取技术、事件知识提取技术、医学文献内容挖掘以及医学知识图谱构建。

简介：穗志方，北京大学信息科学技术学院教授、博士生导师，计算语言学研究所副所长，计算语言学教育部重点实验室主任，中国中文信息学会理事。在汉语语言资源建设方面，长期以来基于汉语语言学理论，探索适合汉语的可计算的语言知识库理论体系、高效的知识库建设方法以及多层次语言知识的挖掘方法。在基于互联网的文本信息挖掘与知识图谱构建领域积累了丰富的科研经验。作为课题负责人，承担国家 973 计划课题、863 项目、国家自然科学基金、国家社会科学基金、中国出版集团项目等多项科研项目。在自然语言处理相关顶级国际会议 ACL、IJCAI、AAAI、COLING、EMNLP 上发表多篇学术论文。作为主要成员，制订国家标准 2 项。研究成果“综合型语言知识库”获 2011 年度国家科技进步二等奖和 2010 年度中国电子学会电子科学技术奖一等奖。



报告三

机器写作技术与应用

万小军

摘要：机器写作(也即文本自动生成)是自然语言处理领域的一个重要研究方向，实现机器写作也是人工智能走向成熟的一个重要标志。简单来说，我们期待未来有一天计算机能够像人类一样写作，能够撰写出高质量的文章，包括各类新闻、诗词歌赋等。本次报告将总结机器写作技术与应用现状，并介绍本团队在机器写作方面的研究进展，最后指出机器写作所面临的挑战以及可能的应对方法。

简介：万小军，北京大学计算机科学技术研究所博士生导师，语言计算与互联网挖掘研究室负责人，在北京大学获得学士、硕士与博士学位。研究方向为自然语言处理与文本挖掘，研究兴趣包括自动文摘与文本生成、情感分析与语义计算等。担任计算语言学顶级国际期刊 Computational Linguistics 编委，TAACL 常务评审，EMNLP-IJCNLP19 程序委员会主席，曾多次担任自然语言处理与人工智能领域一流与重要国际会议领域主席 (Area Chair) 或高级程序委员 (SPC)，包括 ACL、NAACL、EMNLP、IJCAI、AAAI、IJCNLP 等。荣获 ACL 2017 Outstanding Paper Award (杰出论文) 与 IJCAI 2018 Distinguished Paper Award (最佳论文) 等多项荣誉或奖励。与今日头条合作推出 AI 写稿机器人小明 (Xiaomingbot)，与南方都市报合作推出写稿机器人小南等机器写作应用系统。



报告四

自然语言生成关键技术研究

孙栩

摘要：自然语言生成作为自然语言处理的“输出端”与机器翻译、文本摘要、对话系统等多个任务紧密相关。深度学习方法的出现使得机器生成文本的质量取得了长足进步，但当前仍有诸多挑战需要克服。例如，现有生成文本表达单一、缺少多样性，语言风格僵硬、缺少性格特征，话语前后不一致、不连贯。本报告首先将简要介绍深度学习在自然语言生成中的应用，然后探讨前述问题的最新解决方案，并介绍我们在该方向的近期研究进展，主要包括：多样化评论生成、文本情绪转换以及语义一致性改进。

简介：孙栩，2012年开始担任北京大学信息学院研究员、博士生导师。2010年于日本东京大学获得计算机博士学位。先后在日本东京大学、美国康奈尔大学、香港理工大学担任研究职位。曾在微软公司美国雷蒙德研究院访问研究。研究方向为自然语言处理、机器学习，目前关注自然语言的结构化学习、自然语言生成研究。主要成果：1) 面向自然语言的结构化学习理论与方法方面的创新研究；2) 基于深度神经网络的自然语言生成方面的一系列创新研究。先后在自然语言处理、机器学习领域的国际高水平会议和期刊发表50余篇论文，包括ACL, ICML, NIPS等CCF-A类期刊和会议论文20余篇。担任EMNLP、IJCAI等旗舰国际学术会议的领域主席、高级程序委员；长期担任ACL、AAAI等本领域大部分主流会议的程序委员。2014年入选中组部第十批“千人计划”青年人才。2015年获得求是基金会的“求是杰出青年学者奖”。2017年获得大川研究助成奖。2018年获得中国计算机学会CCF-NLPCC青年新锐奖。2018年获得COLING 2018最佳论文奖。



报告五

Combining Human Knowledge with Neural Network Models: A Case Study for SLU and Beyond

冯岩松

摘要: Recent success of many neural network models in natural language processing (NLP) tasks is often bound by the number and quality of annotated data, but there is often a shortage of such training data. In this talk, I will discuss how we can combine a neural network model with human knowledge, such as regular expressions or hand-crafted rules, to improve supervised NLP solutions when the training data is not perfect. Firstly, we develop methods to exploit the rich expressiveness of regular expressions at different levels within a neural network model, showing that the combination can significantly enhance the learning effectiveness when a small number of training examples are available. In the second attempt, we encode human-crafted constraints as a semantic loss to teach a neural network model to behave as we expect, where the learned model can, to some extent, effectively work on imperfect training data that can be easily obtained.

简介: Dr. Yansong Feng is an associate professor in the Institute of Computer Science and Technology at Peking University. Before that, he worked with Prof. Mirella Lapata and obtained his PhD from ILCC at the University of Edinburgh. His current research interests include using probabilistic methods to distill knowledge from large volumes of natural language texts, and supporting intelligent human-computer interfaces, such as question answering and dialogue, which have been published in TPAMI, AIJ, ACL, EMNLP, NAACL and so on. He received the IBM Faculty Award in 2014 and 2015, and the IBM Global Shared University Research Award in 2016.



报告六

人工智能在机对话系统中的技术现状与挑战

严睿

摘要：近年来，自动人机对话系统在学术界和工业界都获得了相当大的关注度，随着微软小冰百度度秘等产品的发布，以及大量人工智能公司的建立，这些对话系统背后的技术在逐步积累，也逐步解密。随着研究者的探索愈发深入，人们看到了一个对话系统逐渐从科幻电影中走进现实生活的可能。在讲座中，我将回顾人机对话的发展历程，以及随着深度学习技术盛行之后，由数据驱动模型带来的革命性改变。讲座将从人机对话的已有应用出发，再分析现有对话系统的不足，展望下一代人机对话系统的挑战。同时会分享我们组在人机对话研究所做的努力与探索，以及相应的代表性成果。

简介：严睿，北京大学研究员，博士生导师，前百度公司资深研发，华中师范大学与中央财经大学客座教授与校外导师。主持研发多个开放领域对话系统和服务类对话系统，发表高水平研究论文 70 余篇，担任多个学术会议（KDD, IJCAI, SIGIR, ACL, WWW, AAAI, CIKM, EMNLP 等）的（高级）程序委员会委员及审稿人。

