

推动社会计算学领域多学科对话

清华大学社会学系 高馨 甄园宜 罗家德

随着大数据与人工智能技术的发展,社会科学领域的研究范式也发生着新的变化,大数据与社会学结合、跨学科交流与对话已成为学术热点。由清华大学社会网络研究中心和中国人民大学信息学院共同举办的2020年第二届国际社会计算会议(2020 International Conference of Social Computing Conference)在线上成功举行,来自中国、美国、德国的18位计算社会学领域的学者就计算社会科学、大数据挖掘、金融科技、公共卫生等主题进行了学术汇报和交流。会议旨在促进计算机科学、统计学、社会科学、管理学、金融学、物理学等学科间的对话和跨界创新。

现有技术发展进一步激发了新的人工智能计算和人类的社会化之间产生的动态交互,从深度和广度上拓宽了社会计算学(Social Computing)的范畴,在社会和计算研究之间达成了更深层次的互动。芝加哥大学社会学系 James Evans 首先回顾了计算社会学历史,同时给出了新的具有时代前瞻性的 AI 定义——异质性智能(Alien Intelligences,即非某个人的创新,而是在群体中涌现或在群体与 AI 互动中涌现的智能),即不单是学习人类行为,同时可以扩大、区别于人类智能。因此,计算社会科学、人机之间的交互,包括人机互动(Human-computer interaction)和以人为中心的计算(Human-centered computing)均包含在社会计算学的范畴中。最后,Evans 介绍了一篇2019年发表在《自然》杂志上的文章,通过将已有文献中的化学物质和化学性质在向量空间中表示,可以预测未来20年研究中40%的物质及其性质的创新性联系,揭示了机器对于人类未来创新的重大影响。

强化学习将创建超出人类理解的交易主体,而这一研究成为现今和未来重要的议题。美国圣菲研究所的 David Wolpert 分享了经济2.0时代基于智能合约的自动化学主体交互(Automated Learning agents interaction)的研究。过去人为设计金融交易规则过于简单,交易主体间的合作和互动行为难以被预测和理解,会出现崩盘的情况。其研究主要通过使用图灵通用算法以及强化学习设计智能合约,可以降低诉讼风险,智能主体间可达成协调自治,动态创建或修正具有公开性质且具备约束力的合约。除此之外,中国人民大学宋华报告了基于机器学习方法进行供应链金融融资可得性的研究,天津大学沈德华介绍了金融大数据与金融市场的相关研究,中山大学韦立坚就P2P网贷风险机理的计算实验分析及其对互联网信贷的启示进行了展示,中央财经大学杨虎报告了在多元社交网络中识别风险投资产业领袖的研究,北京师范大学张伦介绍了关于市场竞争和垄断观念的动态发展研究。

社会网络领域中以复杂和动态网络为理论基础的大数据研究,逐渐成为社会学、管理学、计算机科学与系统科学结合的学术创新领域。清华大学社会学系罗家德基于组织员工社交网络大数据与结构化人力资源数据库的结合数据,从社会网络的视角研究了员工在团队中建立“桥连接”(bridging tie)或“绑带连接”(bonding tie)对于其绩效的影响。基于结构洞理论,进一步探究了员工在组织领导圈子中的位置是否会产生结构洞效果,亦即在不同圈子间建立“桥连接”,产生影响。最终发现在中国组织情境下,依照中国本土人情、关系、面子理论划分的组织领导圈子中,核心层成员、圈内成员及圈外人三种不同的成员身份在组织中建立“桥连接”应遵循不同的策略。其中,位于组织领导圈内的核心层成员和圈内成员,由于与领导之间具有较高的信任度,具备互惠和义务的原则,因此需要保持更高的忠诚度,建立“绑带连接”更有利于其获得高绩效。而位于组织领导圈外的成员,由于其受到更小的限制,可以通过建立“桥连接”来获得更高的绩效。此外,复旦大学陈阳报告了识别在线社交网络中“结构洞”的研究,电子科技大学吕琳媛报

告了识别复杂网络中圈结构的算法研究，斯图加特大学 Steffen Staab 介绍了如何利用社交大数据进行案例决策，德州农工大学 Xinyue Ye 汇报了“网络、流和人类动力学”的研究，伊利诺伊大学 Lav Varshney 分享了基于数字社区信任机制建立虚拟农业网络的研究成果。

社会科学中的不平等问题是社会计算学领域关注的重要议题。哥廷根大学计算机学院傅晓明介绍了使用大数据来预测社会经济地位的研究，其中，社会经济地位指标(SES, social economic status index)是反映一个人或者一个家庭社会和经济等级的指标。该研究使用了 7919137 张上海市民地铁卡 16 天的出行记录，通过推测样本的家庭住址，并结合链家房价数据，将样本划分为高、中、低三个等级的社会经济地位作为扎根真相，最终将这些地铁卡出行时间序列数据结合高德 POI 数据，计算出时间序列下人员的移动轨迹和规律、移动地点多样性等大数据指标来提炼深度学习模型，预测准确率最高可达 69%。还有一些学者关注着不平等和少数群体，电子科技大学周涛研究了数据驱动下的性别不平等，南京大学陈云松使用机器学习方法估计了中国青年中特殊性取向群体的人口比例。

以 2019 年新冠肺炎为主题，部分学者对疫情期间的公众情绪和公共管理等方面进行了深入的研究分析。香港科技大学计算机科学及工程学系的 Pan Hui 对新冠肺炎疫情期间社交媒体上的数据进行了情感分析及政府信息传播分析。基于 2020 年 1 月至 5 月 Twitter 上 17371148 个美国用户产生的 3700 万个疫情相关数据，提出了危机词库的构建方法，用普鲁契克情感色轮(Plutchik's Wheel of Emotions)进行了情感分析，并通过主题模型分析了 8 种基本情感对应的关键词和主题标签。基于 2019 年 11 月至 2020 年 8 月 17 个国家(地区)的 Facebook 数据分析了新冠肺炎疫情前后，各国(地区)官方账号所发布信息的类型和媒体类型等，研究表明链接社交媒体情绪和时事问题分析可以反映公众需求，社交媒体中的现场直播特性可提升公众参与度。其他介绍新冠肺炎疫情相关研究的学者有：锡拉丘兹大学 Daniel Acuna 介绍了其在 2019-nCov 传播期间，对图片中复杂模式测量的研究；华中科技大学黄宏汇报了“新闻媒体眼中的中国：新冠肺炎疫情下的个案研究”；芝加哥大学 Fengli Xu 分享了使用政策感知流动模型(Policy-Aware Mobility Model)来解释新冠肺炎病例在美国城市内增长的研究。

此外，还有其他学者就大数据驱动和理论驱动的相关研究进行了分享。在政治学议题中，清华大学孟天广使用微博社交媒体数据集探讨了数字空间的国家认同、爱国主义和民族主义的问题，芝加哥大学 Shilin Jia 介绍了其对 1946—2003 年《人民日报》的思想转型与相关修辞的演化过程的研究。另外还包括传播学、社会网络与大数据相关的研究，清华大学郑路研究了社交媒体中慈善行为的传播，芝加哥大学 Ishanu Chattopadhyay 介绍了关于信念在集体中形成与传播的研究。其他还包括心理学，南京大学李文中介绍了基于智能手机数据结合深度学习方法预测个人人格的研究。关于在线学习，法兰克福大学 Alexander Mehler 报告了使用知识网络向量化的方法来探究在线学习的限制性因素的研究。