人工智能学术研讨会 **Academic Symposium on Artificial Intelligence (2019.11.4-8)**

报告

概念嵌入：基于概念森林的深度表达学习可解释性研究

Concept Embedding: Interpretability Study in Deep Representation Learning Based on Concept Forest

王国胤教授 | 重庆邮电大学研究生院院长

Professor Guoyin Wang | Dean of Graduate School, Chongqing University of Posts and Telecommunications

讲者介绍 Biography

王国胤教授，重庆邮电大学研究生院院长，大数据智能研究院院长，计算智能重庆市重点实验室主任，大数据智能计算示范型国合基地（科技部）负责人。是长江学者特聘教授，“万人计划”领军人才，“新世纪百千万人才工程”国家级人选，国家重点研发计划项目首席科学家。曾任国际粗糙集学会(IRSS)理事长，现任中国人工智能学会(CAAI)副理事长、重庆市人工智能学会(CQAAI)理事长、中国计算机学会(CCF)理事，是《Int. J. of Approximate Reasoning》、《Trans. on Rough Sets》、《计算机学报》等 10 余种期刊编委，是 IRSS 会士和 CAAI 会士。主要从事粗糙集、粒计算、知识发现、数据挖掘、认知计算、大数据智能等研究，出版专著 20 余部（含编著），发表 SCI/EI 收录论文 300 余篇，论著被他引 10000 多次。获国家级高等教育教学成果二等奖、重庆市自然科学一等奖、吴文俊人工智能科学技术奖科技进步一等奖等成果奖励 8 项。带领的团队获评“国家级教学团队”和“重庆高校创新团队”。

Guoyin Wang received the B.E. degree in computer software, the M.S. degree in computer software, and the Ph.D. degree in computer organization and architecture from Xi’an Jiaotong University, Xi’an, China, in 1992, 1994, and 1996, respectively. He worked at the University of North Texas, USA, and the University of Regina, Canada, as a Visiting Scholar during 1998–1999. Since 1996, he has been working at the Chongqing University of Posts and Telecommunications, Chongqing, China, where he is currently a Professor and a Ph.D. supervisor, the Dean of Graduate School, the Director of Big Data Intelligence Institute, the Director of Chongqing Key Laboratory of Computational Intelligence, and the Director of Big Data Intelligent Computing National Int. Cooperation Base. His research interests include big data intelligence, data mining, machine learning, rough set, granular computing, cognitive computing, etc. He has published over 300 reviewed papers and over 20 books. Dr. Wang was the President of the International Rough Set Society (IRSS) 2014-2017. He is a Vice-President of the Chinese Association for Artificial Intelligence (CAAI), and a council member of the China Computer Federation (CCF).

报告摘要 Abstract

深度神经网络内部知识可解释性差，已经成为了当前机器学习研究亟需解决的关键问题。本报告分析导致这一问题的原因，即传统深度神经网络研究中所存在的知识语义与数据分离表达，深度神经网络的隐特征空间表达缺乏语义，以及无语义性导致机器学习缺乏可解释性和鲁棒性。进而，提出概念嵌入的研究思路，利用概念树的可解释性和层次结构特性，建立概念森林的知识深度表达模型，将概念森林的层次知识表达结构映射到深度神经网络的特征空间表达结构，实现可解释的深度表达学习，具体包括：建立概念森林的知识深度表达模型，实现可解释的“数据→概念→概念树→概念森林”的数据与知识统一表达；提出“数据→概念”、“概念→概念树”、“概念树→概念森林”的学习算子和变换算子，建立贝叶斯云模型，构建概念森林的知识深度表达空间；建立深度神经网络特征空间与概念森林知识深度表达空间之间的映射，实现零/少样本学习的知识迁移与动态演化，以及跨层概念间的关联表示和知识推理。最后介绍一些相关研究工作情况。

The poor interpretability problem of the neural network’s internal knowledge is a key problem to be studied in machine learning now. The reasons for this problem are analyzed, that is, knowledge and data are expressed in different systems

independently in traditional neural network models, the latent feature space of deep neural networks has no semantic meaning, non-semantic causes lack of interpretability and robustness in machine learning. A research idea of concept embedding is proposed. Taking advantage of the interpretability and hierarchical structure characteristics of concept trees, it aims to develop a deep knowledge representation model called concept forest. It maps the hierarchical knowledge representation structure of a concept forest to the feature space representation structure of a deep neural network, and further realizes interpretable deep representation learning. It includes the following three key research tasks. The first, realizing a unified interpretable representation “data -> concept -> concept tree -> concept forest” for data and knowledge through establishing a deep knowledge representation model based on concept forest. Then, proposing learning and transforming operators of "from data to concept", "from concept to concept tree" and "from concept tree to concept forest", developing a Bayesian Cloud Model, and constructing a deep knowledge representation space based on concept forest. The third, establishing a mapping between the hierarchical feature space of a deep neural network and the deep knowledge representation space of a concept forest, realizing the knowledge transformation and dynamic evolution of zero-shot or small-shot learning, as well as the association representation and knowledge reasoning among cross-layer concepts. Some related researches are introduced at last.

有兴趣合作之项目 Interested topics for future collaboration

基于概念森林的深度表达学习可解释性研究