

AI 时代社会科学研究方法创新与模型 “过度拟合”问题探索

冯帅帅 张佳星 罗教讲

摘要：大数据和机器学习的有效结合推动人工智能获得重大突破，同时也为社会科学研究方法创新带来新的发展契机。传统理论假设和统计知识驱动的量化研究对模型的过度拟合问题关注不够，导致研究结论的一般化能力受限，更使社会科学研究成果的社会预测功能为人所诟病。而基于交叉验证和正则化方法的机器学习建模方法可能有效解决过度拟合问题，为开展社会预测研究提供方法支撑。本文首先讨论了过度拟合问题的发生根源和内在机制，继而对模型过度拟合问题的机器学习纾解方法进行了介绍，最后分析了机器学习建模方法的不足和限制性因素。将新的机器学习方法运用于社会科学研究，这是机遇而非威胁，研究者需要保持客观态度，努力确保自己有能力根据实际需要将经典方法和新方法的组合应用于具体研究之中。

关键词：AI；机器学习建模；量化研究；过度拟合；预测研究

①作者简介：

冯帅帅，湖南师范大学公共管理学院讲师，博士；张佳星，南京大学社会学院博士研究生；罗教讲，武汉大学社会学院教授、博士生导师。

基金项目：

本文为国家哲学社会科学基金重大项目“大数据时代计算社会科学的产生、现状与发展趋势研究”（16ZDA086）阶段性成果。

Innovation of Social Science Research Methods and Exploration of Model Overfitting in AI Era

Shuaishuai FENG, Jiaying ZHANG, Jiaojiang LUO

ABSTRACT

The effective combination of big data and machine learning techniques has promoted major breakthroughs in artificial intelligence, and brought new opportunities for the innovation of quantitative research methodology in social sciences. Quantitative research driven by traditional theoretical assumptions and statistical knowledge does not pay enough attention to model overfitting, which limits the generalization ability of research conclusions and ignores prediction in social science research. The machine learning models based on cross-validation and regularization methods may provide a way to address the problem of overfitting and offer methodological support for predictions in social research. This paper introduces the machine learning approaches to address the model overfitting with discussion on the reason and internal mechanism behind as well as its advantages and disadvantages. We argue that implementing machine learning techniques along with classic methods that fit research needs would offer an opportunity rather than a threat to social science researchers.

KEY WORDS

AI; Machine learning models; Quantitative study; Model overfitting; Prediction in social research