

Analyzing & Clustering Enterprise Microblog Users with Supernetwork

Xi Yunjiang¹ Du Diedie¹ Liao Xiao² Zhang Xuehong¹

¹(School of Business Administration, South China University of Technology, Guangzhou 510641, China)

²(School of Internet Finance and Information Engineering, Guangdong University of Finance,
Guangzhou 510521, China)

Abstract: [Objective] This paper proposes an integrated modeling method to process multi-dimensional user interest data, aiming to examine the spectral clustering method for analyzing user interests. [Methods] First, we retrieved Weibo (Microblog) data of “Three Squirrels” and used supernetwork model to integrate the modeling of contents and user interaction data. Then, we constructed an interactive interest index and grouped the users with spectral clustering algorithm. Finally, we evaluated the clustering results with the Silhouette Coefficient and Davies-Bouldin methods. [Results] We found that the clustering DB value reached 0.57 (k was set at 15), which was evenly distributed. [Limitations] More research is needed to further explore user characteristic data and the impacts of different data dimensions on user interests. [Conclusions] This study proposes maintenance and marketing suggestions for enterprise Weibo profiles, which will help them identify user interests and improve marketing effectiveness.

Keywords: Supernetwork Enterprise Microblog User Interests Spectral Clustering

机器学习预测金融市场走势

康奈尔大学最新研究表明,机器学习可以评估用于预测金融市场走势的数学工具的有效性,该项研究基于该领域有史以来最大的数据集。研究人员开发出的模型还可以预测期货市场走势。

预测市场走势是一项艰巨的任务,因为市场信息量巨大而且波动性很强。研究人员指出:“由于数据库过于庞大,使用行业标准技术来估算市场走势变得非常棘手。而机器学习优势在于,它是分析数据的另一种方式。本研究的关键发现:在某些情况下,一个交易所展现出的微观结构特征是如此强大,以至于可以用于预测其他交易的走势。”

在这项研究中,研究人员使用随机森林机器学习算法以更好地理解一些模型的有效性。尽管机器学习在金融领域早已被使用,但通常被称为“黑匣子”。在这种情况下,人工智能算法使用大量数据来预测期货的模式,但没有揭示其做出决定的原因。研究人员认为这种方法短期内可能有效,但是对影响市场格局真正的起因却知之甚少。

“在该研究中,机器学习的应用场景是:如果我对如何推动市场发展有一个理论,那么如何进行验证?”研究人员解释,“我如何才能真正理解我的理论是否足够好?又如何利用从这种机器学习方法中学到的知识来帮助我建立更好的模型,并理解那些过于复杂而无法建模的东西?”

在该领域有大量的历史市场数据可供使用,实际上,市场记录了自1980年以来的每笔交易信息。计算能力和数据可用性的提高使得执行更细粒度和更全面的分析成为可能,但是这些数据集以及分析它们所需的计算能力对于学者而言仍然是一个不小的挑战。

(编译自:<https://www.sciencedaily.com/releases/2020/08/200811142913.htm>)

(本刊讯)