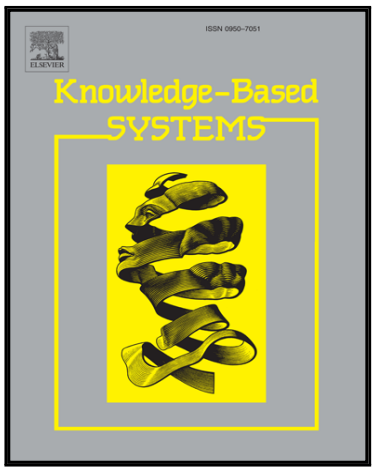

Improving Stock Market Prediction via Heterogeneous Information Fusion

Xi Zhang, Yunjia Zhang, Senzhang Wang, Yuntao Yao,
Binxing Fang, Philip S. Yu

PII: S0950-7051(17)30603-2
DOI: 10.1016/j.knosys.2017.12.025
Reference: KNOSYS 4161

To appear in: *Knowledge-Based Systems*

Received date: 15 January 2017
Revised date: 5 December 2017
Accepted date: 25 December 2017

Please cite this article as: Xi Zhang, Yunjia Zhang, Senzhang Wang, Yuntao Yao, Binxing Fang, Philip S. Yu, Improving Stock Market Prediction via Heterogeneous Information Fusion, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.12.025

Improving Stock Market Prediction via Heterogeneous Information Fusion

Xi Zhang[a,*], Yunjia Zhang[a], Senzhang Wang[b], Yuntao Yao[a], Binxing Fang[a,e], Philip S. Yu[c,d]

[a] *Key Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[b] *College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[c] *Department of Computer Science, University of Illinois at Chicago, IL 60607, USA*
[d] *Institute for Data Science, Tsinghua University, Beijing 100084, China*
[e] *Institute of Electronic and Information Engineering of UESTC in Guangdong, Dongguan Guangdong 523808, China*

Abstract

Traditional stock market prediction approaches commonly utilize the historical price-related data of the stocks to forecast their future trends. As the Web information grows, recently some works try to explore financial news to improve the prediction. Effective indicators, e.g., the events related to the stocks and the people's sentiments towards the market and stocks, have been proved to play important roles in the stocks' volatility, and are extracted to feed into the prediction models for improving the prediction accuracy. However, a major limitation of previous methods is that the indicators are obtained from only a single source whose reliability might be low, or from several data sources but their interactions and correlations among the multi-sourced data are largely ignored.

In this work, we extract the events from Web news and the users' sentiments from social media, and investigate their joint impacts on the stock price movements via a coupled matrix and tensor factorization framework. Specifically, a tensor is firstly constructed to fuse heterogeneous data and capture the intrinsic

*Corresponding author
Email addresses: zhangx@bupt.edu.cn (Xi Zhang), 2011213120@bupt.edu.cn (Yunjia Zhang), szwang@nuaa.edu.cn (Senzhang Wang), yaoyuntao@bupt.edu.cn (Yuntao Yao), fangbx@bupt.edu.cn (Binxing Fang), psyu@uic.edu (Philip S. Yu)