Chapter 8.15
Mining Allocating Patterns in Investment Portfolios

Yanbo J. Wang
University of Liverpool, UK

Xinwei Zheng
University of Durham, UK

Frans Coenen
University of Liverpool, UK

ABSTRACT

An association rule (AR) is a common type of mined knowledge in data mining that describes an implicative co-occurring relationship between two sets of binary-valued transaction-database attributes, expressed in the form of an ⟨antecedent⟩ ⟹ ⟨consequent⟩ rule. A variation of ARs is the (WARs), which addresses the weighting issue in ARs. In this chapter, the authors introduce the concept of "one-sum" WAR and name such WARs as allocating patterns (ALPs). An algorithm is proposed to extract hidden and interesting ALPs from data. The authors further indicate that ALPs can be applied in portfolio management. Firstly by modelling a collection of investment portfolios as a one-sum weighted transaction-database that contains hidden ALPs. Secondly the authors show that ALPs, mined from the given portfolio-data, can be applied to guide future investment activities. The experimental results show good performance that demonstrates the effectiveness of using ALPs in the proposed application.

INTRODUCTION

Investments (Bodie, Kane, Marcus, & Ryan, 2003; Cuthbertson & Nitzsche, 2001) are one of the major schools in financial research that parallels corporate finance (Damodaran, 2001), personal financial planning (Ho & Robinson, 2001), financial engineering (Neftci, 2004), and so forth. Portfolio management, aiming to minimize the overall risk while maximizing the total expected return for an investment activity, is perhaps one of

the most indispensable tools available in investments. It diversely "allocates" a given amount of assets/funds in a variety of investment-items (i.e., bonds, funds, options, stocks, etc.). In Ho and Robinson (2001) diversification (Farrell, 2006) was introduced as a principle of investments. There are three dimensions in diversification (Ho & Robinson, 2001): (1) diversity across items/assets within the same investment-security, (2) diversity across different securities of investments, and (3) diversity internationally. When addressing diversification in portfolio management, choosing to invest a portfolio that consists of a set of uncorrelated investment-items or negatively correlated investment-item pairs, noted as the correlation coefficient based portfolio theory (Ho & Robinson, 2001), is recommended.

Data mining (Bramer, 2007; Han & Kamber, 2001; Han & Kamber, 2006; Hand, Mannila & Smyth, 2001; Thuraisingham, 1999) is a promising area of current research and development in computer science, which is attracting more and more attention from a wide range of different groups of people. It aims to extract various types of hidden, interesting, previously unknown and potentially useful knowledge (i.e., rules, patterns, regularities, customs, trends, etc.) from databases, where the volume of a collected database can be measured in GBytes. In data mining common types of mined knowledge include: association rules (Agrawal & Srikant, 1994), classification rules (Quinlan, 1993), prediction rules (Han & Kamber, 2001), classification association rules (Ali, Manganaris & Srikant, 1997), clustering rules (Mirkin & Mirkin, 2005), emerging patterns (Dong & Li, 1999), sequential patterns (Wang & Yang, 2005), and so forth. In the past decade, data mining techniques have been widely applied in, for example, bioinformatics (Wang, Zaki, Toivonen & Shasha, 2005), e-commerce (Raghavan, 2005), geography (Miller & Han, 2001), marketing and sales studies (Berry & Linoff, 1997). Kovalerchuk and Vityaev (2000) systematically discussed, in the scope of computational finance, the necessi-

ties and/or possibilities of employing data mining technologies and/or methodologies in financial research.

Portfolio management, in a general prospect, refers to the overall process of creating appropriate portfolio strategies that will ensure/almost-ensure profits in future investment activities. The portfolio management process has been analysed in many literatures, but a unique scheme has not yet been agreed upon. The stages of the portfolio management process usually include:

1. **Investment-item selection:** Where a number of investment-items/assets that will be comprised in a "potential" portfolio are selected.
2. **Investment-item return prediction:** Where the expected return of each asset, selected in stage 1, is predicted.
3. **Investment-item weight determination:** Where a candidate portfolio is generated by assigning a suitable weight to each asset, based on the result of stage 2.
4. **Portfolio selection:** Where the "best" portfolio strategy is selected from a number of alternative candidate portfolios that are generated by iteratively processing stages 1, 2 and 3. Best in this case is defined according to the return and risk of the candidate portfolio.

In the past decade, research in portfolio management has demonstrated an interest in some data mining and/or machine learning concepts (Hung, Liang & Liu, 1996; Lazo, Maria, Vellasco, Aurelio & Pacheco, 2000; Tseng, 2004; Wang & Weigend, 2004; Zhang & Zhou, 2004). A number of approaches in such research are summarized as follows:

*   John, Miller, and Kerber (1996) developed a rule induction based stock selection system, namely Recon. This system marks "stocks with returns in the top 20% in a

Related Content

GR-OLAP: Online Analytical Processing of Grid Monitoring Information
Julien Gossaand Sandro Bimonte (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 716-727).
www.irma-international.org/chapter/olap-online-analytical-processing-grid/20757

Learn From the Rumors: International Comparison of COVID-19 Online Rumors Between China and the United Kingdom
Fei Liuand Meiyun Zuo (2021). *Journal of Database Management (pp. 46-68).*
www.irma-international.org/article/learn-from-the-rumors/282444

Revising the Socio-Technical Perspective for the 21st Century: New Mechanisms at Work
Louise Harder Fischerand Richard Baskerville (2020). *Journal of Database Management (pp. 69-87).*
www.irma-international.org/article/revising-the-socio-technical-perspective-for-the-21st-century/266405

A Literature Overview of Fuzzy Database Modeling
Z.M. Ma (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 105-125).
www.irma-international.org/chapter/literature-overview-fuzzy-database-modeling/7904

Understanding Business Domain Models: The Effect of Recognizing Resource-Event-Agent Conceptual Modeling Structures
Geert Poels (2011). *Journal of Database Management (pp. 69-101).*
www.irma-international.org/article/understanding-business-domain-models/49724