Expert Systems with Applications 39 (2012) 6279–6291

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A quantitative correlation coefficient mining method for business intelligence in small and medium enterprises of trading business

C.F. Cheung *, F.L. Li

Knowledge Management Research Centre, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Business intelligence
Data mining
Correlation coefficient mining
Product relation network
Knowledge management
Knowledge system technology

ABSTRACT

Business intelligence based on data mining has been one of the popular and indispensable tools for identifying business opportunity in sales and marketing of new products. The traditional data mining methods based on association rules may be inadequate in completely uncovering the hidden patterns of sales based on transaction records. This paper presents a qualitative correlation coefficient mining method which is capable of uncovering hidden patterns of sales and market. Hence, a prototype business intelligence system (BIS) named correlation coefficient sales data mining system (CCSDMS) has been developed and successfully trial implemented in a selected reference site. A series of experiments have been conducted to evaluate the performance of the proposed system. The results generated by the BIS are compared with a well known market available data mining system. The proposed quantitative correlation coefficient mining method is found to possess higher accuracy, better computational effectiveness and higher predictive power. With the new approach, associations for product relations and customer periodic demands are revealed and this can help to leverage organizational marketing capital to enhance quality and speed of promotions as well as awareness of product relations.

1. Introduction

Conventionally, companies desire to know what customers need and where companies should focus on the market in order to survive in today's highly competitive market. The management of market inquiries and customer relationships in organizations pay great attention to the strengthening of their organizational development. Vindevogel, Van Den Poel, and Wets (2005) studied the market basket analysis as the base for the promotion strategy for products. Buckinx, Verstraeten, and Van Den Poel (2007) make use of the internal transactional database for predicting customer loyalty for targeted marketing actions while Larivière and Van Den Poel (2005) attempt to better understand important measures of customer outcome such as next buy, partial defection and customers' profitability evolution by using random forest and regression forest techniques. Jonker, Piersma, and Van Den Poel (2004) present a joint optimization approach for the segmentation of customers and the development of the optimal policy towards homogeneous groups of customers. Burez and Van Den Poel (2007) make use of different churn-prediction models for early detection of potential churners which enables an European pay-TV company to target their customers using specific retention actions, and subsequently increase profits.

The concept of a business intelligence (BI) system is adopted in various industries in order to meet their specific business requirements. Miller, Dagmar, and Stefanie (2006) defined BI is getting the right information to the right people at the right time. BI constitutes a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help people in the organization to make better business decisions.

Nowadays, small and medium enterprises (SMEs) provide jobs for 1.36 million people and account for 60% of total employment in Hong Kong (Lau, 2007) whereas trading companies account for a large segment of SMEs. It can be seen that trading companies form the backbone of business community. "SMEs have intelligence needs and should consider seeking out relevant information" was stated by Gangadharan (2004). This advice is not only for large enterprises, SMEs also have a need to use BI systems to develop their business. Many software companies (Hyperion Solutions Corporation (2007), Cognos Company (2007) and Business Objects (2007)) attempt to provide powerful and commercial BI systems. Unfortunately, the functions of their BI systems cannot be fully applied in SMEs for trading and the large expense in their BI systems is not affordable by SMEs. One of the solutions is to develop their own BI systems based on their specific business problems.

Moreover, trading companies focus more on Marketing Intelligence (MI) and Customer Intelligence (CI) under the general heading of BI, since MI and CI are the core competency of trading companies which provide value-added services for the customers.

* Corresponding author. Fax: +852 23625267.
E-mail address: mfbenny@inet.polyu.edu.hk (C.F. Cheung).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.10.021

To overcome their particular problems, a customizable BI application of quantitative correlation coefficient mining for analyzing sales periodic ordering patterns is presented in this paper for customizing BI in SMEs, for the purpose of trading.

To realize the capability of the proposed method, a case study has been conducted for the application of proposed BI system in a selected SME electronics trading company A series of experiments have been conducted to evaluate the performance of the proposed systems. The results generated by BI System are compared with a well known market available data mining system named Weka by using a set of past sales data from selected company and a series of predefined performance indicators which are accuracy, computational effectiveness and predictive power, respectively.

2. Theoretical background for product relation network analysis and correlation coefficient mining method

2.1. Traditional association rule mining

Association rule mining is widely used in data mining for uncovering customer behaviors. According to Yen and Chen (1997), an association rule describes the association among items when they are purchased in a transaction. The main challenge for association rule mining is how to reduce the time for data mining when a large number of transactions are being considered. Various algorithms have been developed for enhancing its efficiency such as Apriori, AprioriTid (Agrawal & Srikant, 1994), All-Confidence and Bond Edward (Omiecinski, 2003) and CCMine (Lee, Kim, Cai, & Han, 2003). Basically, transactions in association rule mining may have some fields such as product name(s), transaction number, product name(s) represent(s) product(s) purchased within one transaction. The transaction number is unique for identifying different transactions.

Although traditional association rules can help to identify some patterns of product purchasing which frequently occur in transactions, they do not take into account the quantity purchased. Furthermore, in real purchasing cases, customers sometimes order a list of products using many transactions rather than only one transaction. This method may be due to minimization of inventory, complex procurement procedures and manual careless mistakes. In order to improve customer service, many companies allow their customers to make orders frequently at short notice. Because of this ordering convenience, customers frequently tend to order just one product in one transaction. In individual transactions, products lack relations and it is difficult to discern a pattern. In such situations, association rules discovered are not accurate for supporting organizational marketing strategy.

2.2. Quantitative correlation coefficient mining

Quantitative correlation coefficient mining is built based on Pearson's product-moment correlation efficient Pearson's product-moment correlation coefficient (sometimes called Pearson's correlation coefficient or sample correlation coefficient) (Easton & McColl, 2007), is a number between −1 and 1 which measures the degree to which two variables are linearly related. If there is a perfect linear relationship with a positive slope between the two variables, Pearson's correlation coefficient becomes 1. If there is a positive correlation between the two variables, whenever one variable has a high (low) value so does the other. Should there be a perfect linear relationship with a negative slope between two variables, Pearson's correlation coefficient becomes −1. This implies that there is a negative correlation between the two variables. Whenever one variable has a high (low) value the other also has a low (high) value. Pearson's correlation coefficient of value 0 implies that there is no linear relationship between the two variables. The Pearson's correlation coefficient value (r) is determined by Eq. (1). To apply Pearson's correlation coefficient to sales patterns, relationships between different sales patterns can be represented by Pearson's correlation coefficient value.

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{1}$$

where two sets of number are represented as $X = \{X_1, X_2, \ldots, X_n\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$, n is the number of measurements for X and Y.

When it comes to the adoption of Pearson's correlation coefficient for data mining, the TAPER algorithm (Xiong, Shekhar, Tan, & Kumar, 2004) is used which derives the support form of the ϕ correlation coefficient and upper bound of the correlation coefficient ϕ as shown in Eqs. (2) and (3), respectively. TAPER consists of coarse filtering and refinement processes. Coarse filtering is used for establishing the upper bound of the Pearson's correlation coefficient and it firstly filters less potential item pairs and thus reduces the rate of transaction checking. In the refinement process, it computes the exact correlation for each surviving pair determined after the coarse filtering process. Moreover, it just focuses on item pairs in individual transactions.

$$\phi_{\{A,B\}} = \frac{\sup(A,B) - \sup(A)\sup(B)}{\sqrt{\sup(A)\sup(B)(1-\sup(A))(1-\sup(B))}} \tag{2}$$

$$upper(\phi_{\{A,B\}}) = \sqrt{\frac{\sup(B)}{\sup(A)}}\sqrt{\frac{1-\sup(A)}{1-\sup(B)}} \tag{3}$$

where A and B are two items in a market basket database.

Pearson's correlation coefficient applying for organizational product relationship and the customer periodic demand are two main areas considered in correlation coefficient mining. The organizational product relationship considers the similarity of the ordering patterns for different products. By collecting the information from the customers, products with strong ordering relations can be identified and outstanding sets of products can be made visible. For example, sales patterns for product A and product B are similar and such situation occurs frequently with different customers. This infers that product A and product B tend to be used together for some specific applications. This is particularly true for electronic products which are made up of many specific electronic components. These components may not all be used in a single product. Instead, they are grouped and then operated with other specific components so as to perform some specific functions in the electronic products. By recognizing members of the product groups and by suggesting relevant products to the customers, the speed and the quality of marketing can be enhanced.

Shugan and Radas (1999) stated that customer periodic demand relies heavily on seasonal factors which are caused by some human made phenomena such as change of economic environment, change of government policy, etc. However, some of them are due to commercial reasons such as festivals, holidays, annual government actions, etc. There are many factors that affect the patterns of customer demand. The present study is not concerned with the causes of customer periodic demand. Instead, it focuses on the prediction of the customer periodic demand based on historical sales transaction records. The patterns of customer annual ordering are compared with that for previous years. By predicting this periodic customer demand, the company can play an active and important role in providing the right products to the right customers at the right time. Such promotion is more easily well received and accepted by the customers.