Data Mining for Business Analytics in Retail

Min Gan
Deakin University, Australia

Honghua Dai
Deakin University, Australia

INTRODUCTION

Large volumes of retail transaction data have been accumulated since the use of point of sale (POS) machines and the application of information technology in retail in the 1970s. The retail transaction data are normally organised in relational databases. As shown in Table 1, each record in the database represents one transaction with four major fields; TID (Transaction ID), transaction time, the details of items contained in the transaction, and Total. Transactions of customers using loyalty cards can be further organised as databases of transaction sequences as shown in Table 2. Each row in Table 2 records transactions of one customer in time order. The first column represents Customer ID. The second column contains the transaction sequence of the customer, where each transaction is distinguished by a transaction ID.

The accumulated big retail business data has motivated the application of data mining techniques to retail business analytics research and practice. In the data mining community, researchers have conducted intensive research in this area, and developed a number of retail data mining approaches for tacking a variety of retail business analysis and optimization problems.

However, we argue that existing data mining approaches are not effective and robust enough to meet requirements of real-life retail business practice as we identified that (1) retail knowledge discovered by existing data mining techniques may be not effective and precise enough to reflect regularities hidden in the retail data sets, i.e., the knowledge is not necessarily reliable, and (2) existing retail data mining approaches do not apply to some new types of retail business such as e-retail over the Internet.

Main objectives of this chapter are (1) to explore state of the art data mining techniques for retail business analysis and optimisation, (2) to examine reliability of retail knowledge discovered

Table 1. A sample database of retail transactions

TID	Time	Items' Details				Total ($)
		Item	Unit Price	Quantity	Subtotal ($)	
10000001	09:15, 12/03/2012	Beer	$3.58	2	$7.16	$31.33
		Apple	$2.45	1.8 kg	$4.41	
		Diaper	$9.88	2	$19.76	
10000002	09:20, 12/03/2012	Sugar	$1.26	1.5kg	$1.89	$15.73
		Beef	$8.65	1.6kg	$13.84	
…	…	…	…	…	…	…
99999999	16:34, 09/11/2012	Toothpaste	$2.9	2	$4.80	$8.05
		Toothbrush	$3.25	1	$3.25	

DOI: 10.4018/978-1-4666-5202-6.ch057

Table 2. A sample database of retail transaction sequences

Customer ID	Transaction Sequence
C2000001	10000001, 10033018, 10632501
C2000002	10004001, 10033018, 10435513, 11475902
…	…
C2009999	60204341, 79608387

by existing retail data mining approaches and identify deficiencies of these approaches, (3) to propose solutions/ recommendations for improving the reliability, and (4) to identify and discuss new challenges and future research directions in this field.

BACKGROUND

Since the application of association rule mining to market basket analysis (MBA) (Agrawal, Imielinski & Swami, 1993), researchers have proposed a variety of retail data mining approaches for retail business analysis and optimization. These approaches are based on three major data mining techniques, association rule mining (Agrawal, Imielinski & Swami, 1993; Agrawal, Imielinski & Swami, 1994), classification (Cumby, Fano, Ghani & Krema, 2004) and clustering (Rajagopal, 2011). This section explores the three categories of retail data mining approaches.

Association-Based Approaches

Association rule mining is a data mining technique that has been widely used in retail knowledge discovery (Han & Kamber, 2001). The first application of association rule mining in retail knowledge discovery is market basket analysis (Agrawal, Imielinski & Swami, 1993). In MBA, cross-selling relationships between items are normally modelled as association rules (Agrawal, Imielinski & Swami, 1993). A famous association rule discovered in transaction databases is

$$\text{Diaper} \Rightarrow \text{Beer } [s\%, c\%] \quad (1)$$

D

This rule indicates that customers who purchase diapers tend to buy beer. The support s% indicates that diapers and beer were purchased together in s% of the transactions in the database. The confidence c% indicates that c% of the customers who purchased diapers also bought beer. MBA focuses on modelling cross-selling relationships and revealing customers' buying habits by mining association rules in transaction databases (Agrawal, Imielinski & Swami, 1993; Agrawal, Imielinski & Swami, 1994).

Based on MBA, association rule mining has been applied to address a number of retail marketing and operation problems such as Maximal-Profit Item Selection (Brijs,winnen,Vanhoof, & Wets 1999; Brijs, Goethals, Swinnen,Vanhoof & Wets, 2000; Wang& Su, 2002; Wong, Fu &Wang, 2003; Wong & Fu, 2004), product assortment and shelf-space allocation (PASA) (Brijs,Swinnen,Vanhoof, & Wets, 1999; Brijs, Goethals,Swinnen,Vanhoof, & Wets, 2000; Chen, & Lin, 2007; Nafari & Shahrabi, 2010), direct marketing (Ling & Zhang, 1998; Wang & Zhou, 2005), and analysis of effects of discount promotions (Gan & Dai, 2011).

In recent years, MBA has been extended by integrating more transaction information, e.g., product quantity and product price (Chen, Huang, & Chang, 2008), considering MBA in more complex situations such as in a multiple-store environment (Tang, Chen & Hu, 2008).

Classification-Based Approaches

Classification (Fix, & Hodges 1951; Kass 1980), as a major technique in data mining and machine learning, has been applied in various predictions in retail marketing. Four representative predictions are addressed below.

The first is the prediction of likely buyers among current non-buyers. Ling et al. (Ling & Zhang, 1998) discovered patterns by learning in databases of customers using learning algorithms

Related Content

Business Intelligence 2.0: The eXtensible Markup Language as Strategic Enabler
Rubén A. Mendoza (2010). *International Journal of Business Intelligence Research (pp. 63-76).*
www.irma-international.org/article/business-intelligence-extensible-markup-language/47196

Technology Development and Assessment to Market Using TRIZ
Zulhasni bin Abdul Rahimand Nooh bin Abu Bakar (2016). *International Journal of Business Analytics (pp. 83-97).*
www.irma-international.org/article/technology-development-and-assessment-to-market-using-triz/165012

Depicting Data Quality Issues in Business Intelligence Environment Through a Metadata Framework
Te-Wei Wang, Yuriy Verbitskiyand William Yeoh (2019). *Applying Business Intelligence Initiatives in Healthcare and Organizational Settings (pp. 291-304).*
www.irma-international.org/chapter/depicting-data-quality-issues-in-business-intelligence-environment-through-a-metadata-framework/208104

Improving Business Intelligence: The Six Sigma Way
Dorothy Miller (2010). *International Journal of Business Intelligence Research (pp. 47-62).*
www.irma-international.org/article/improving-business-intelligence/47195

Identifying Technical Requirements for a Mobile Business Analytics Application
Tuncay Bayrak (2021). *International Journal of Business Analytics (pp. 91-103).*
www.irma-international.org/article/identifying-technical-requirements-for-a-mobile-business-analytics-application/288060