Chapter XXI

E-Commerce and Data Mining: Integration Issues and Challenges

Parviz Partow-Navid
California State University, Los Angeles, USA

Ludwig Slusky
California State University, Los Angeles, USA

ABSTRACT

Web mining is the application of data mining techniques to discover the usage patterns of Web data, in order to better serve the needs of Web site visitors. Web mining consists of three phases: data gathering, analysis and reporting. This chapter describes each of these phases in detail along with a discussion of electronic customer relationship management (eCRM). Several challenging research areas that need to be investigated for further enhancement of this field are also presented.

INTRODUCTION

For most companies, competitiveness in e-commerce (EC) demands a meaningful presence on the Web. Web pages are required to establish the company's image, to sell products and to offer customer support. The success of a Web site directly affects the success of the EC strategy of the company. EC is also emerging as one of the best application areas for data mining technology (Konrad, 2001).

Data dining (DM) technologies have been around for many years without notable success. While DM tools help in the identification of data patterns, DM applications are, for the most part, in the research area that focuses on high possible rewards along with high-risk business decisions. Having the following characteristics present in a DM

application can help minimize the risk of business failure (Ansari, Kohavi, Mason & Zheng, 2002; Kohavi, 2001):

1.   Data contained reach descriptions — For example, broad customer data combined with potentially useful data fields allow DM programs to look beyond plain correlations.

2.   A substantial amount of analyzed data — Having many matching records guarantees the statistical significance of the found data patterns.

3.   Controlled and reliable data collection procedures — Noisy and distorted data may fail to disclose patterns.

4.   The ability to evaluate and track tangible results — Measurable return on investment can be very convincing.

5.   Ease of integration of DM with existing operational processes — The operational system and the DM analysis system can be integrated into one closed-loop system.

EC is the ideal domain for DM. Many of the above characteristics are readily present in EC but rarely in legacy systems. Data are gathered electronically, rather than manually, thus less noise is introduced from manual processing. EC data are rich, containing information such as purchase history and demographic data. Also, some data that were difficult to gather before are now easily accessible. For instance, EC systems can keep track of the actions of the customer in the visual "shop," including what they review, what they place into their shopping carts, but do not purchase and so on. In the past, to gather such data, companies had to follow customers and record their activities, or had to conduct sophisticated analyses of store videos. Large amounts of data can be collected inexpensively in EC. To take full advantage of this process, DM must be integrated into EC systems. Such integration can significantly decrease the data preparation time. An integrated system can also offer users a standardized user interface and easy access to metadata. The goal of this chapter is to provide an overview of the challenges and issues in Web mining. We discuss Web mining phases and the inherent power in current eCRM solutions. Research areas that need to be examined for further enhancement of this area are also presented.

DM DEFINITION

DM can be defined (Berry & Linoff, 2000) as the process of exploration and analysis of large quantities of data in order to extract meaningful patterns and rules. DM tools identify patterns in data. These patterns can be used to help in the decision-making process and forecast the impact of these decisions. DM can decrease analysis time by directing attention to the most significant variables. The information resulting from DM falls into the following five categories:

•   **Classification.** This information identifies the main characteristics of a certain group (e.g., determining which telephone lines are used for Web access).

•   **Estimation.** This information helps forecast future values based on patterns within large sets of data (e.g., estimating a family's total income).

•   **Association.** This information determines the relationship between events that take place at one time (e.g., identifying what things go together in a shopping basket at the supermarket).

Related Content

Application of Big Data in Healthcare: Opportunities, Challenges and Techniques
Md Rakibul Hoqueand Yukun Bao (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1189-1208).*
www.irma-international.org/chapter/application-of-big-data-in-healthcare/150211

A Query Language for Mobility Data Mining
Roberto Trasarti, Fosca Giannotti, Mirco Nanni, Dino Pedreschiand Chiara Renso (2013). *Developments in Data Extraction, Management, and Analysis (pp. 23-44).*
www.irma-international.org/chapter/query-language-mobility-data-mining/70791

Restoration Technique to Optimize Recovery Time for Efficient OSPF Network
Pertik Gargand Ashu Gupta (2016). *Research Advances in the Integration of Big Data and Smart Computing (pp. 64-88).*
www.irma-international.org/chapter/restoration-technique-to-optimize-recovery-time-for-efficient-ospf-network/139396

P2P-COVID-GAN: Classification and Segmentation of COVID-19 Lung Infections From CT Images Using GAN
Nandhini Abirami, Durai Raj Vincentand Seifedine Kadry (2021). *International Journal of Data Warehousing and Mining (pp. 101-118).*
www.irma-international.org/article/p2p-covid-gan/290272

LF-LDA: A Supervised Topic Model for Multi-Label Documents Classification
Yongjun Zhang, Zijian Wang, Yongtao Yu, Bolun Chen, Jialin Maand Liang Shi (2018). *International Journal of Data Warehousing and Mining (pp. 18-36).*
www.irma-international.org/article/lf-lda/202996