12-31-2004

The Utilization of Data Mining Analytics to Enhance the Return on Investment of Online Document Repositories

Stephan Kudyba  
New Jersey Institute of Technology

Sergei Ananyan  
Megaputer Intelligence Inc.

Kenneth Lawrence  
New Jersey Institute of Technology

Follow this and additional works at: http://aisel.aisnet.org/amcis2004

Recommended Citation
http://aisel.aisnet.org/amcis2004/122

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
The utilization of data mining analytics to enhance the return on investment of online document repositories

Stephan Kudyba  
New Jersey Institute of Technology  
Stephan.P.Kudyba@NJIT.edu

Sergei Ananyan  
Megaputer Inc.  
s.ananyan@megaputer.com

Kenneth Lawrence  
New Jersey Institute of Technology  
Kenneth.D.Lawrence@NJIT.edu

Abstract

This research paper addresses the utilization of complementary technologies to more effectively identify and accommodate market demand for knowledge resources available in a document repository. More specifically, it investigates the utilization of data mining methodologies to better identify consumer purchasing patterns for engineering based documents which are available via an e-commerce platform. Document repositories often possess voluminous sources of value added information for a given marketplace, however one drawback to this resource is that consumers often do not have the time to fully and effectively search existing documentation. One way to overcome this is through the incorporation of data mining technologies that identify consumer preferences and purchasing propensities. With this information, e-commerce platforms can more effectively direct consumers to existing resources that are available within a document repository. The result is a higher volume of sales for existing customer visits and increased customer satisfaction.

Key Words:

Online Information Repositories, Data Mining, E-commerce CRM

INTRODUCTION

The opportunities for organizations to increase their operational efficiencies continue to abound given the innovation of existing and introduction of new information technologies. Many of these productivity-increasing opportunities come in the form of effective investment in complementary technologies whose operational capabilities integrate in order to increase the speed and accuracy of information communication and transaction facilitation. One such combination of technologies that offers potential positive returns to investment involves the utilization of online document repositories that provide consumers with readily accessible information corresponding to some established content along with data mining analytic applications. Given the often voluminous content residing in these repositories, consumers often are required to conduct time consuming search activities to investigate available resources, where value added documents can ultimately go unidentified. Organizations can achieve both higher customer satisfaction along with increased revenue by investing in data mining analytics that augment the operational capabilities of existing online repositories.

COMPLEMENTARY TECHNOLOGIES THAT INCREASE EFFICIENCY

The evolution of e-commerce has provided consumers an additional medium in which to acquire information and generally conduct commerce. Some of the positive attributes which e-commerce may entail over more traditional “brick and mortar” retail mechanisms involve time savings in purchasing products and services, information as to the choices of goods and services and, depending on the e-commerce platform, the cost of transacting. (Reibstein, 2002). These positive attributes however, can vary depending on the type of good or service that is offered in the online sector. One area of particular success for e-commerce initiatives has been the creation of information repositories or platforms that provide online access to documents and general resources corresponding to a particular subject matter. The online
environment in this market sector can reduce the time to transaction and enhance the choice of alternative resources within a particular content area.

Organizations however, can miss crucial opportunities to fully service the needs of consumers with their online capabilities given the complex nature of consumer behavior and the sheer volume of products (information sources) that exist. With a greater understanding of consumer propensities, organizations can better facilitate the presentation of available goods and services during the customer visit. Advanced analytic software technology, e.g. data mining can help increase the ability of organizations to identify consumer behavior in what products they are likely to purchase. With this information, they can better leverage the existing information resources available to online visitors by displaying that information during customer transactions. More specifically, data mining analytics enable organizations to increase sales through more accurate cross-selling initiatives and also enhance the level of customer satisfaction by reducing customer search and transaction time and enhance the choice of purchase through automated document suggestions.

THEORETICAL ISSUES

E-commerce platforms continue to evolve as the potential growth of online sales remains promising. Forrester Research forecasts that online retail sales will continue to see growth of up to 25% (Johnson and Delhagen, 2003). However, despite this growth potential, the online retail environment continues to suffer from some noteworthy limitations, one of these involving the failure of online platforms to understand consumer needs at specific points in their shopping process (Zeng, Reinartz, 2003). By better understanding the interests of consumers during their website visits, organizations can enhance site performance by introducing value added product and service information during critical site activities that ultimately could lead to increased sales and enhanced customer relationship building. The danger to this process of introducing online information however, lies in the potential to generate an unpleasant customer site visit, as they are interrupted with information that is viewed with little value (Xia and Sudharshan, 2002). Interruptions and general deterioration in customer satisfaction can ultimately mitigate the primary objectives behind the deployment of e-commerce.

There are two distinct initiatives behind the development of online retail capabilities. The first involves providing a low cost market mechanism for providers of goods and services to connect with the consumers in their market space. The other involves the development of e-commerce platforms to enhance customer relationship management with an increased focus on driving top line revenue through more effective content facilitation (Rust and Kannan, 2003). A primary requirement to driving revenue rests in the ability to better identify consumer propensities and optimize site designs to best facilitate them. This topic coincides with the research conducted by (Lipshitz and BarIlan, 1996) which recommended that further studies be conducted to validate the consumers’ decision making process in actual decision making situations.

IDENTIFYING ONLINE CONSUMER BEHAVIOR THROUGH DATA MINING

The field of data mining, which involves the incorporation of mathematical and statistical processes to identify patterns and relationships within data, has been used to enhance the functionality of e-commerce (Buchner, 1998). Two types of data mining techniques that enable e-commerce providers to better understand consumer behavior entail Market Basket Analysis and Collaborative Filtering. The former focuses on identifying which products and services are purchased together e.g. in bundles) along with other key drivers that lead to these sales. These key drivers generally entail the association between one product or a group of products that are purchased and the likelihood an additional product may be purchased (Brin, 1997). MBA can enhance website effectiveness by displaying key selling product combinations to site visitors. However, the volume of products and services available on particular sites e.g. information repositories) can number in the hundreds of thousands and calculations among these via the Market Basket approach become unrealistic.

Collaborative filtering based on Bayesian techniques facilitate the creation of a special matrix of conditional probabilities of product purchases in a real time environment (Resnik, 1994). This generation of near, real time product recommendations are vital to enhance the customer online shopping experience as they are presented with valued information. Valued recommendations can help consumers better navigate through sites and reduce purchase time. The online organization benefits from a higher increased value per consumer transaction through a higher amount of products purchased (Agrawal & Srikant, 1994).
CASE APPLICATION OF TECHNOLOGIES

In order to test the return on investment for complementary technologies, more specifically the addition of data mining analysis to an e-commerce platform, we will take a detailed look at potential revenue enhancements resulting from cross-sales recommendations produced by mining methods for an online information repository. The organization in this case, offers one of the broadest collections of engineering documents ranging from research and design, to manufacturing and repair.

Case Description

With over $400 million in annual sales, the company focuses on developing and implementing engineering, technical, and regulatory information solutions for customers in over 100 countries. It produces comprehensive collections of technical standards, safety publications and design guidelines and serves the vital information needs of customers via its B to C platform.

The organization recognized the potential lack of guidance for their customers during the product selection process, which could result in lost sales and dissatisfied customers who were unable to find the products of interest to them. This was a potential drawback given the existing online repository which contains over 125,000 documents in multiple categories. The diverse nature of the customer base of more than 60,000 and corresponding consumer interests, posed major technical challenges for traditional recommendation systems to deliver accurate and complete results in real time.

To help customers identify related documents at the time of their online transaction, the organization implemented a document suggestion system based on a Market Basket Mining approach. Based on historical sales data, the mining analysis returns a list of documents that are most commonly purchased in conjunction with any given document in response to online product information requests.

The collaborative filtering approach was also utilized to facilitate recommendations for very large selections of products for customers who already purchased more than one product from the platform. The system which utilizes Bayesian techniques (Vassilios, 2003) is trained periodically in a background regime on a large pool of historical purchase transactions to identify typical purchasing patterns and stores the necessary intermediate results, thus allowing real time production of recommendation rules in response to transactions performed by a customer through the website.

ROI from Web Enhanced Operations with Data Mining

Benefits resulting from the implementation of data mining analytics in conjunction with the online document platform were achieved primarily in two ways. The first enhancement to the e-commerce initiative came in the form of increased purchases (e.g. purchases of more than one product) for each customer transaction. Cross-sell transactions facilitated by the mining methods accounted for almost 6% of the total volume of sales, which were not achieved by the site prior to the introduction of mining. The second return to the organization came in the form of increased customer satisfaction and loyalty through offering customers relevant purchase recommendations during their site visit and also reducing their search time.

The organization facilitating the online document site established key parameters to determine the effectiveness of utilizing data mining by implementing software tracking tools to monitor the system performance and its impact on document sales. The business performance of the solution was evaluated by tracking the volume of additional sales of products recommended by the mining model to existing customers at the time of placing their order. The cost of the mining implementation was recouped during two months of operation as it provided a 6% increase in total sales.

The value of the recommendation system (facilitating product cross-sales) involves more effectively leveraging an organization’s resources to meet customer demand. In the case of the online platform, with 125,000 existing documents, data mining techniques exposed products that may have otherwise gone unnoticed. One key point to emphasize is the mathematical and statistical calculations involved with data mining, help identify products and services that are more likely to be viewed as attractive to customers, which goes beyond the capabilities of mere search techniques. This advances the
previous research, which asserts that automated information support may help decision makers in choosing alternatives (Simon & Egidi, 1992).

CONCLUSION

This study supports the notion that investment in the complementary technology of data mining can improve the performance of an online information repository by better leveraging existing products and services to meet the needs of consumers, which ultimately increases the value of existing transactions. This investment can also yield increased consumer satisfaction by reducing the time of transaction through reduced search.

With the continued growth in the realm of e-commerce, a major issue for website development is to consider is the creation of an efficient marketplace. Efficiency refers to making information on products and services readily available to perspective consumers. With voluminous product and service resources available at various e-commerce platforms and given the diverse nature of consumer interests, this process is tenuous. Online platforms themselves pose inefficiencies in a product/service market, however with the addition of sophisticated data mining methods, which can better combine likely interests of consumers with available products, the result is the creation of a more efficient market platform that can ultimately increase the return on investment to the corresponding technologies and increase consumer satisfaction.

Future points and the continuation of research for this study will address a more detailed analysis of data variables involved in the mining process and potentially the consideration of product profit margin in the overall cross-selling process.

REFERENCES