WEBKDD 2000 - Web Mining for E-Commerce

Myra Spiliopoulou
Otto-von-Guericke Universitaet.
Universitaetsplatz 2, D-39016
Magdeburg
myra@iti.cs.uni-magdeburg.de

Jaideep Srivastava
University of Minnesota & Yodlee, Inc
3600 Bridge Parkway, Suite 200
Redwood City, CA 94065
srivasta@cs.umn.edu

Ron Kohavi
Blue Martini Software
2600 Campus Drive, SanMateo,
California 94403
ronnyk@bluemartini.com

Brij M. Masand
Redwood Investment Systems
76 Summer Street
Boston, MA 02110-1225 USA
brij@redwood.com

ABSTRACT

In this paper, we provide a summary of the WEBKDD 2000 workshop, whose theme was 'Web Mining for E-Commerce'. This workshop was held in conjunction with the ACM SIGKDD International Conference on Knowledge Discovery in Databases (KDD-2000).

Keywords

Web mining, e-commerce, personalization, clickstream analysis.

1. THEME

The ease and speed with which business transactions can be carried out over the Web has been a key driving force in the rapid growth of electronic commerce. In addition, customer interactions, including personalized content, e-mail campaigns, and online feedback provide new channels of communication that were not previously available or where very inefficient. The Web is revolutionizing the way businesses interact with each other (B2B) and with each customer (B2C). It has introduced entirely new ways of doing commerce, including e.g. auctions and reverse auctions. It also made it imperative for organizations and companies to optimize their electronic business.

Knowledge about the customer is fundamental for the establishment of viable e-commerce solutions. Web mining for e-commerce is the application of web mining techniques to acquire this knowledge for e-commerce. Typical concerns in e-commerce include improved cross-sells, up-sells, personalized ads, targeted assortments, improved conversion rates, and measurements of the effectiveness of actions.

The WEBKDD 2000 workshop is the second workshop held in conjunction with the ACM SIGKDD International Conference on Knowledge Discovery in Databases (KDD) and dedicated to the challenges of web mining. WEBKDD'99 focused on the aspects of web mining related to user profiling; the long version of its proceedings has appeared as volume 1836 of the Lectures in Artificial Intelligence series (LNAI) by Springer Verlag. In response to call for papers, WEBKDD 2000 received 31 contributions. Each was reviewed by at least three program committee members. Seven submissions were selected for presentation as long papers, and six as short papers reporting on good ideas at a rather preliminary phase.

The URL http://robotics.stanford.edu/~ronnyk/WEBKDD2000 contains the final versions of the workshop papers and the slide presentations.

2. WORKSHOP

The KDD community responded very enthusiastically to the WEBKDD 2000 workshop, and we received far more requests for attendance (approximately 110) than there was space. About 85 people attended the workshop, which brought together e-commerce practitioners, tool vendors and data mining researchers. The paper presentation was divided into three sessions.

The first session, titled *Web personalization and recommender systems*, focused on how web mining can address one of the fundamental issues of B2C e-commerce, namely personalized customer experience. As often described by Jeff Bezos, CEO of Amazon.com, and mentioned by Joseph Pine in his The Experience Economy [1], customer experience is the key to building customer loyalty to an on-line store, since leaving the store is exactly one click away. In this session we had three long and two short papers, which presented the leading edge ideas in this important area. Mobasher, Dai, Luo, Nakagawa, Sun, and Wiltshire's paper, titled *Discovery of Aggregate Usage Profiles for Web Personalization*, described how usage data from web logs can be analyzed/mined to build user profiles, and how these could be use to enhance the user's browsing experience. Vucetic and Obradovic's paper, titled *A Regression-Based Approach for Scaling-Up Personalized Recommender Systems in E-commerce*, presented an approach to applying regression techniques to understand user preferences for recommender systems. This approach is interesting since statistical techniques have not been applied sufficiently to this problem. Sarwar, Karypis, Konstan and Riedl's paper, titled *Application of Dimensionality Reduction in Recommender Systems – A Case Study*, presented a novel application of dimensionality reduction techniques from scientific computing to the recommendation system problem. The two short papers in this session, namely Lin, Alvarez and Ruiz's *Collaborative Recommendation via Adaptive Association Rule Mining* and Chang and Yuan's *A Synthesized Learning Approach for Web-Based CRM*, presented early results in alternative approaches to the recommendation system problem. In summary, the wealth of interest in applying various techniques to the recommendation system problem shows the centrality of this problem to Web personalization.

The second session, titled *Mining frameworks and case studies*, presented experience reports from four groups on using web mining in various e-commerce applications. Any field of enquiry must have its 'proof of the pudding is in eating it' component, and the case studies presented here provide exactly that flavor. Ansari, Kohavi, Mason and Zheng's paper, titled *Integrating E-Commerce and Data Mining: Architecture and Challenges*, provided a comprehensive overview of the issues in applying data mining techniques to E-commerce. The authors bring a lot of real world perspective from their experience at Blue Martini Software, especially from the viewpoint of an E-commerce solution provider. Theusinger and Huber's *Analyzing the footsteps of your customers, A case study by ASK|net and SAS Institute GmbH* presented the experience of applying data mining in E-commerce from a solutions perspective, where SAS's tools were used to solve a problem for ASK|net. Sanford Gayle, in his paper *The Marriage of Market Basket Analysis to Predictive Modeling: The Essential Challenge in Exploiting Web-Log Files for Prediction*, presented an approach to using association and correlation analysis to extract predictive models from web logs. Coenen, Swinnen, Vanhoof and Wets' paper *A Framework for Self Adaptive Websites: Tactical versus Strategic Challenge* examined the various issues in building such sites. The essential tension seems to be between making a web site personalized to individual users - maybe even dynamically change it based on the particular user's behavior – and the information overload it can cause.

The third session, titled *Navigation analysis*, focused on how clickstream data can be analyzed to extract valuable e-commerce knowledge from it. Being able to analyze clickstream data provides an unprecedented opportunity to understand in detail the process leading up to a buy/not buy decision vs. just recording the final outcome - as is the case with point-of-sale data. Clickstream data is over 95%+ of all data collected in most large-scale e-commerce environments, and contains a wealth of knowledge embedded in it. Berendt's paper, titled *Web Usage Mining, Site Semantics, and the Support of Navigation,* provided a general overview of the issues in clickstream analysis, and how the mined knowledge can be used for supporting site navigation. Kato, Nakayama and Yamane's paper *Navigation Analysis Tool based on the Correlation between Content Distribution and Access Patterns,* presented an approach whereby mined patterns from site content can be correlated with mined patterns from site usage, and a tool based on this approach. Investigating applications of such a tool would be an interesting line of inquiry. Gaul and Schmidt-Thieme's *Mining Web Navigation Path Fragments* presented some novel algorithms for extracting navigational path fragments. Finally, Tan and Kumar's *Modeling of Web Robot Navigational Patterns* addressed the challenging and commercially important problem of separating the site visits of web robots from humans. This is crucial for at least two applications: (1) as competitive pressures increase, commerce sites would like to block robots that collect sensitive information, and (2) accurate modeling of human users' e-commerce behavior requires that web robot accesses be filtered out. While the papers in this session present some of the leading ideas, the research in this area is just beginning and we have barely scratched the surface.

3. CONCLUSION

WEBKDD 2000 turned out to be a very successful workshop by all measures. More than 110 people showed interest in the workshop and over 85 attended it. The quality of papers was excellent, the discussion was lively, and a number of interesting directions of research were identified. This is a strong endorsement of the level of interest in this rapidly emerging field of inquiry.

4. REFERENCES

[1] B. Joseph Pine, James H. Gilmore, B. Joseph Pine II. The Experience Economy. Harvard Business School Pr; ISBN: 0875848192, http://www.amazon.com/exec/obidos/ASIN/0875848192/ref=sc_b_1/103-2009916-9046229