

DATA WEB MINING IN E-COMMERCE: PROGRESS AND PERSPECTIVES

Armand Florin BERTEA

UAIC, Romania

armandbertea@gmail.com

Abstract. *The move from traditional to online services and the spectacular increase in the number of online customers bring great challenges to the field. Analysing the Internet flow, valuable insights about the traffic coming and leaving an e-commerce site can be found, helping them to become more effective and increasing their competitiveness. This is what Web mining does, as it deals with the mining of remarkable knowledge from the Internet, studying the way users interact. Web mining is used in e-commerce to find the browsing habits of customers, helping in understanding their preferences. This paper examines the main processes of data mining (Web Content, Web Structure and Web Usage information) and the various types of web data mining technologies, emphasizing their use in the online business. An overview of the latest developments is presented and the main challenges are outlined. As a final point, possible perspectives of web data mining in the e-business domain are analysed.*

Keywords: development perspectives, e-business, web content mining, web structure mining, web usage mining

JEL classification: L81, L89

DOI: 10.12948/ie2019.03.01

1. Introduction

Before the 90's, the commerce was not web-based (because the web did not exist then). Electronic computing tools, if used, were designed to streamline the process of processing and sending commercial documents, such as invoices [1]. After this period, with the first shy, and then explosive evolution of the Internet and the proliferation of web servers, online commerce, or as it is commonly known – electronic commerce or e-commerce - is beginning to grow rapidly [2], because of its advantages over traditional business: faster speed and lower expenses [3].

In the early 2000, management expert Peter Drucker predicted that online commerce would essentially influence the way that business would be done [4], and the present days prove that its predictions are completely accurate.

E-commerce implies electronic data transmission using devices of telecommunication networks to improve business processes and implement business strategies, supplying services and commodities, providing after-sales services, supporting services/products to services, processing payment, managing business transactions, processing payment and supporting customers [5], [6].

Even more, we can talk about e-commerce as part of e-business, which refers to more than buying and selling goods and services, namely conducting online business such as servicing customers or conducting various electronic transactions [7], hence applying Internet technology in all aspects of the business world [8].

As most of the internet users have at least one profile in one of the main social networks, the web becomes the major market defining E-Commerce. Since consumers play a vital role in all

forms of business, under the conditions of fierce competition, the ability to predict consumer preferences, the characteristics of target groups, and possible market developments becomes essential. This is why the possibility to analyse given data became crucial and e-commerce commerce benefits a lot from web mining in its attempt to conceive strategies that lead to customer satisfaction [9].

A huge amount of data is collected in business-to-consumer and it can be used, if gathered, managed and interpreted correctly, to extract valuable information and knowledge. Using web data mining technologies, hidden correlations can be discovered, allowing better predictive analysis to be used in the strategy and decision-making.

Utilization of big data in e-commerce offers a better experience for the customer as it needs can be understood and satisfied, by foretelling the user's concerns and behaviours and, last but not least, by means of real time analysis prices are altered in order to compete with other vendors [10].

2. Web mining – definitions and types of web mining

Data mining (sometimes called knowledge discovery) is the method of extracting information from large sets of data and converting it into a comprehensible structure for future use by means of specific techniques and methods. The steps of the process are shown in figure 1 [11].

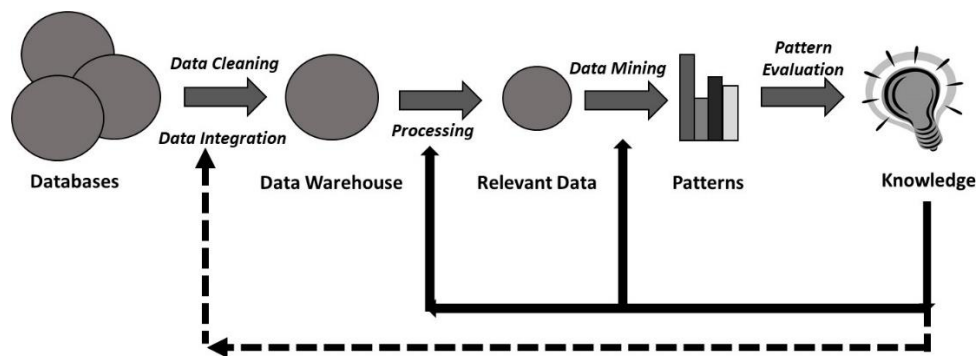


Figure 1. The steps of Data mining seen as knowledge discovery in databases

If Data mining is executed in the web, it is called Web mining. Hence, Web mining refers to the discovery and extraction of interesting information from World Wide Web documents and services, applying data mining techniques [12].

The process is based on different methods to gather, analyse and understand the extracted data. Unlike Data Mining, which inspects databases with a certain level of explicit structure to extract information and convert it into a comprehensible structure for future use, Web Mining analyses unstructured or semi-structured data as they are contained in web pages [13].

Web mining has two different methods, namely a process-based view and data-based view.

The Web mining process comprises the following stages [14]:

- Finding resource by retrieving and collecting Web documents;
- Selection and pre-processing of information to identify the specific data and convert it into an suitable form of treatment;
- Discovery of patterns.
- Checking the validity of the extracted information and its representation in a suitable manner.

From the perspective of the process-based view, data mining is a set of ordered tasks, while according to the data-based view, data mining is based on the type of data to be analysed. Based on the type of data, Web mining techniques can be categorized into three types: web-

content mining, structure mining and usage mining [15-17], as it can be seen in figure 2.

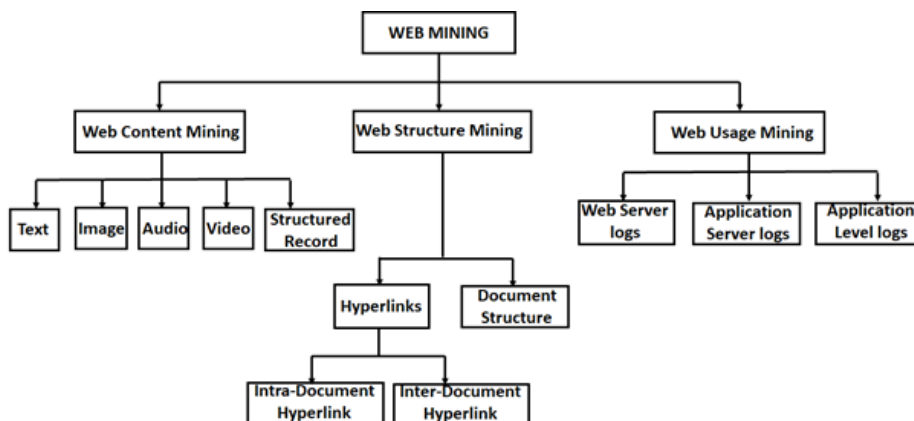


Figure 2. Types of Web Mining

2.1 Web Content Mining

Web Content Mining (WCM) is a mixture of text and multimedia mining [18], used to detect valuable information from webpages content such as text, image, audio, video, metadata, hyperlinks [19-21]. As Web content mining helps better understand customer behaviour, it consistently helps boost e-businesses. Two approaches can be identified in web content mining: agent based approach and database approach.

The agent based approach is based on intelligent search agents, which hunts for information according to a precise query using domain features and user profiles, information filtering/categorizing agent, which filter data in accordance with predefine information, and personalized web agents that acquire user inclinations and find documents correlated to the user profiles. The Web Content Mining database approach is based on unstructured data mining (such as text documents) [22] using pattern matching and tracing keywords and phrases, structured mining (data in the form of list, tables and tree) using web crawler, wrapper generation and page content mining [23], semi-structured data mining (from sources that does not enforce rigid data structure) using object exchange model, Top Down Extraction and web data extraction language, and multimedia mining using SKy Image Cataloguing and Analysis Tool (SKYCAT), colour Histogram Matching, Multimedia Miner and Shot Boundary Detection [22], [23].

2.2 Web Structure Mining

WSM (Web Structure Mining) is used to produce structural summary about the Web site and Web page, discovering the link structure model [18] and finding the relationship between the user and the web [22]. It can be said that the main goal of web structure mining is to analyse the web pages and organise them into a structured manner.

WSM analyses both the intra-document structure (within the web document) and inter-document structure (within the web itself) [17].

Web Structure Mining can be separated into two categories, according to the type of structural information used: hyperlinks and document structure [24].

2.3 Web Usage Mining

Web usage mining (WUM), which represents a way to comprehend and handle customers' web comportment [18], is considered to be the main mode of monitoring market evolution, as customers cannot essentially be seen only in content and structure mining [2], [25].

The application of web usage mining in the e-commerce include customer analysing, website

optimization, web personalization, and business intelligence [18]. The stages of Web usage mining [26] are shown in figure 3.

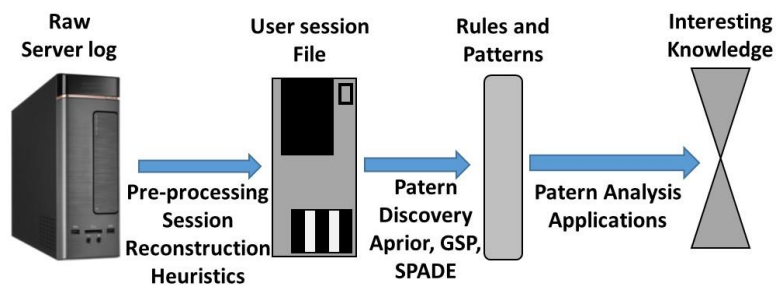


Figure 3. Phases of Web Usage Mining (Shrivastava, 2016)

The most important methods used in web usage mining, which follow a pre-processing phase, comprise: association rules, frequent pattern discovery and clustering [25], [28].

Association rule is a methods frequently used in the web usage mining, which supports web site to acquire a more efficient content organization, finding associations between pages that regularly appear next to one another in user sessions. This represents a powerful tool for adapting to market changes, and help e-businesses to analyse customer behaviours [2].

Frequent pattern discovery is used in web usage mining to find user navigation patterns which frequently occur. Two types of algorithms are used for mining sequential patterns: one based on association rules mining, and the second one uses tree structure and Markov chain to represent survey patterns.

Clustering techniques identify groups of similar items amongst high volumes of data based on distance functions which determines the level of similarity between different items [25].

3. Web mining prospects and challenges

Besides significant benefits, there are some challenges for web data mining e-commerce. In a very brief enumeration, it's about the difficulties in data transformations, scalability of data mining algorithms, making data mining models understandable to business users, support slowly changing dimensions. A challenge of data mining refers to threat analysis, hence detection of like insurance, credit cards, and telecommunications fraud by developing dynamic methods to risk recognition. In the last years problems connected to privacy invasion and ethical issues became general of interest.

Even though these problems are serious and complex, they can overcome by e-commerce companies by applying the correct techniques, beginning with the e-commerce web site, which has to be designed in a way that allows search engines to easily read it. In addition, making good use of the cloud computing technology in e-commerce decreases costs for effective data mining [9].

4. Web mining - privacy and ethical issues

Though there are numerous profits from web mining, nowadays there is a major drawback: Web mining does pose a threat to significant ethical values like privacy and individuality [28-30]. The values of privacy and individuality have to be respected, and the impact of disregarding these issues has consequences in many domains.

Web content and structure mining can affect privacy when data published on the web is mined and, after combining with other data, are used in a completely diverse perspective. Privacy concerns in web usage mining can appear when the actions of web users are traced and investigated without their awareness.

As the problem develops rapidly, a multitude of solutions to privacy-problems have appeared, none of them assuring adequate protection. A combined solution package of these solutions has to be developed in the near future to relax the tensions and fears that are growing in this area. Privacy Preserving Data Mining (PPDM) can be a partial solution to warrant the privacy of users during the data mining process [31], but future approaches, even of legislative nature, have to be taken.

5. Conclusions

Web mining has very high significance in E-commerce, as it leads to a better understanding of customer's needs and the improvement of the website performance, as its structure can be adapted to the requirement given by the customer. This paper describes the main characteristics of web mining, briefly reviewing the various Web usage mining concepts, its techniques and applications and the implication in e-commerce, highlighting the problems that need to be overcome and the prospects for future development.

References

- [1] L. Sadath, "Data Mining in E-Commerce: A CRM Platform", *International Journal of Computer Applications*, vol. 68, no. 24, pp. 32-37, April 2013.
- [2] H. Artail, A. El Halabi, A. Hachem and L. Al-Akhrass, "A framework for identifying the linkability between Web servers for enhanced internet computing and E-commerce", *Journal of Internet Services and Applications*, vol. 8, no. 1, pp. 1-19, December 2017.
- [3] H. Mohana and M. Suriakala, "An overview study on web mining in ecommerce", *International Journal of Scientific Research*, vol. 6. No. 8, pp. 175-177, August 2017.
- [4] P. Drucker, *Managing in the Next Society*. New York: Truman Talley Books, 2002, pp. 40.
- [5] Zacharoula, C. Koliouka, G. Tsekouropoulos and V. Samathrakis, "E-Commerce and Database Technology in Small-Medium Wood Enterprises in Greece", in *Proc. International Conference on Information and Communication Technologies for Sustainable Agri-production and Environment*, Skiathos, Greece, 2011, pp. 901-911.
- [6] T. Siddiqui and M. Muntjir, "A Modern Approach to Integrate Database Queries for Searching E-Commerce Product", *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 14, no. 4, pp. 523-531, April 2016.
- [7] E. Turban, J. Outland, D. King, J. K. Lee, T.-. Liang and D. Turban, *Electronic Commerce 2018 A Managerial and Social Networks Perspective*. Springer International Publishing AG, 2018, pp. 7-10.
- [8] P. Balaraman and S. Chandrasekar, "E – Commerce Trends and Future Analytics Tools", *Indian Journal of Science and Technology*, vol. 9, no. 32, pp. 1-9, August 2016.
- [9] M. Ismail, M. M. Ibrahim, Z. M. Sanusi and M. Nat, "Data Mining in Electronic Commerce: Benefits and Challenges", *Int. J. Communications, Network and System Sciences*, vol. 6, no. 12, pp. 501-509, December 2015.
- [10] B. Pavithra, M. Niranjanmurthy, S. Kamal and S. Martien, "The Study of Big Data Analytics in E-Commerce", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 2, pp. 126-131, October 2016.
- [11] E. Indarto. *Data Mining. Recommender Systems 0.7*. [Online]. Available: <http://recommender-systems.readthedocs.io/>. June, 2013, [Accessed March. 20, 2018].
- [12] Y. Thushara and V. Ramesh, "A Study of Web Mining Application on E-Commerce using Google Analytics Tool", *International Journal of Computer Applications*, vol. 149, no.11, pp. 21-26, September 2016.
- [13] S. N. Kumar, "World towards Advance Web Mining: A Review", *American Journal of Systems and Software*, vol. 3, no. 2, pp. 44-61, 2015.

- [14] I. Zaqout, A. Mahdi and M. Alhabbash, “Web Mining: A Review”, International Journal of Science and Engineering Investigations, vol. 5, no. 57, pp. 45-50, October 2016.
- [15] K. Tandele and B. Pansare, “Web Usage Mining with Improved Frequent Pattern Tree Algorithms”, International Journal of Computer Science and Information Technology Research, vol. 3, no. 2, pp. 952-958, 2015.
- [16] V. B. Aggarwal, V. B. Durgesh and K. Mishra, Big Data Analytics, Springer Nature, 2018, pp. 305-317.
- [17] K. Anu, “Web Mining Evolution & Comparative Study with Data Mining”, International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 5, pp. 1010-1014, May 2017.
- [18] S. Sharma and M. Rai, “Web Mining: Roadmap to Customer”, in Proc. Proceedings of the 11th INDIACOM IEEE Conference - Computing for Sustainable Global Development, New Delhi, India, 2017, pp. 4834-4837.
- [19] A. Kumar and R. Kumar Singh, “A Study on Web Content Mining”, International Journal of Engineering And Computer Science, vol. 6, no. 1, pp. 20003-20006, January 2017.
- [20] A. Adsod and N. Chopde, “A Review on Web Mining”, International Journal of Engineering Trends and Technology, vol. 10, no. 3, pp. 108-113, April 2014.
- [21] L. Mary and G. Silambarasan, “Web Content Mining: Tool, Technique & Concepts”, International Journal of Engineering Science and Computing, vol. 7, No. 5, pp. 11656-11660, May 2017.
- [22] S. Saini and H. M Pandey, “Review on Web Content Mining Techniques”, International Journal of Computer Applications, vol. 118, no. 18, May 2015.
- [23] C.D. Rao and G. M. Someswar, “Analysis of research issues in web data mining”, International Journal of Technical Research and Applications, vol. 2, no. 3, pp. 18-24, May-June 2014.
- [24] N. R. Satish, “A Study on Applications, Approaches and Issues of Web Content Mining”, International Journal of Trend in Research and Development, vol. 4, no. 6, pp. 41-43, November-December 2017.
- [25] N. Jokar, A. R. Honarvar, S. Aghamirzadeh and K. Esfandiari, “Web mining and Web usage mining techniques”, Bulletin de la Société des Sciences de Liège, vol. 85, no.1, pp. 321 – 328, 2016.
- [26] J. N. Shrivastava and S. P. Singh, “A Survey of Web Usage Mining: Concepts With Applications and Its Future Scope”, International Journal of Computer Science Trends and Technology, vol. 4, no. 2, pp. 1-5, March - April 2016.
- [27] A. Sidana and H. Aggarwal, “Review of web usage of data mining in web mining”, International Journal of Advanced Research in Computer Science, vol. 8, no. 5, pp. 2742-2746, May – June 2017.
- [28] S. S. Gautam and M. K. Tiwari, “Web Mining — Concepts and its Applications”, International Research Journal of Computer Science, vol. 3, no. 1, pp. 8-13, January 2016.
- [29] A. S. Prasad, V. M. K. and V. Rao, “Some Studies on Web Mining Ethical Issues and Challenges”, International Journal of Trend in Research and Development, vol. 3, no. 4, pp. 496-499, July-August 2016.
- [30] T. Nusrat Jabeen, M. Chidambaram and G. Suseendran, “Security and privacy concerned association rule mining technique for the accurate frequent pattern identification”, International Journal of Engineering & Technology, vol. 7, no. 1, pp. 19-24, January 2018.
- [31] B. Sundararajan, D. Peri, N. Radhakrishnan and M. Awasthi “An Extensive Survey of Privacy Preserving Data Mining Techniques”, International Journal of Computer Science and Network, vol. 6, no. 5, pp. 547-550, October 2017.