

# Chapter V

## Machine Learning and Web Mining: Methods and Applications in Societal Benefit Areas

**Georgios Lappas**

*Technological Educational Institution of Western Macedonia,  
Kastoria Campus, Greece*

### **ABSTRACT**

*This chapter reviews research on machine learning and Web mining methods that are related to areas of social benefit. It shows that machine learning and Web mining methods may provide intelligent Web services of social interest. The chapter reveals a growing interest for using advanced computational methods, such as machine learning and Web mining, for better services to the public, as most research identified in the literature has been conducted during the last years. The chapter objective is to help researchers and academics from different disciplines to understand how Web mining and machine learning methods are applied to Web data. Furthermore it aims to provide the latest developments on research that is related to societal benefit areas.*

### **INTRODUCTION**

The Web is constantly becoming a central part of social, cultural, political, educational, academic, and commercial life and contains a wide range of information and applications in areas that are of societal interest. Web mining is the field of data mining that is related to the discovery of knowledge from the Web. The Web can be con-

sidered as a tremendously large and rich in content knowledge base of heterogeneous entries without any well specified structure, which proportionally makes the Web at least as complex as any known complex database and perhaps the largest knowledge repository. The vast information that surrounds the Web does not come only from the content of Websites, but is also related to usage of Web pages, navigation paths and networking

between the links of Web-pages. All these properties establish the Web as a very challenging area for the machine learning community to apply their methods usually for extracting new knowledge, discovering interesting patterns and enhancing the efficiency of Websites by providing user-demand content and design.

Web mining is a relatively new area, broadly interdisciplinary, attracting researchers from: computer science fields like artificial intelligence, machine learning, databases, and information retrieval specialists; from business studies fields like marketing, administrative and e-commerce specialists; and from social and communication studies fields such as social network analyzers, pedagogical scientists, and political science specialists. Herrera-Viedma and Pasi (2006) denote that due to the complexity of Web research there is a requirement for the use of interdisciplinary approaches like statistics, databases, information retrieval, decision theory, artificial intelligence, cognitive social theory and behavioral science. As a relatively new area there is a lot of confusion when comparing research efforts from different point of views (Kosala & Blockeel, 2000) and therefore there is a need for surveys that record and aggregate efforts done by independent researchers, provide definitions and explain structures and taxonomies of the field from various points of view.

The overall objective of this chapter is to provide a review of different machine learning approaches to Web mining and draw conclusions on their applicability in societal benefit areas. The novelty of this review is that it focuses on Web mining in societal benefit areas. There exist similar work related to Web mining in (Baldi, Frasconi, & Smyth, 2003; Chakrabarti, 2003; Chen & Chau, 2004; Pal, Talwar & Mitra, 2002). Baldi et al. (2003) cover research and theory on aspects of Internet and Web modeling at the information level based on mathematical, probabilistic, and graphical treatment. Chakrabarti focuses on studies that connect users to the information they seek

from the Web providing lots of programs with pseudocode. Chen and Chau provide an extended review of how machine-learning techniques for traditional information retrieval systems have been improved and adapted for Web mining applications. Pal et al. (2002) present an overview of machine learning techniques with focusing on a specific Web mining category, the Web content mining that will be described in next section. This work is differentiated from the aforementioned related work as the chapter particularly focuses on Web mining and machine learning that may help and benefit societal areas in ways of extracting new knowledge, providing support for decision making and empowering valuable management of societal issues. This survey aims to help researchers and academics from different disciplines to understand Web mining and machine learning methods. Thus, it is aimed at a relatively broad audience and tries to provide them with a different and more open view on Web research. Therefore this work addresses researchers from both computer science and other than computer science disciplines with the intention: (a) for computer science researchers, to provide them with the latest developments on the theory and applications of Web mining, focusing also to the need for Web mining applications in societal beneficial areas, and (b) for researchers from other than computer science disciplines, to draw their attention to existing machine learning methods that may help them to seek for more effective results in their Web research.

Later in the chapter, some background to the different perspectives of Web mining has been provided with a short review on machine learning methods. Afterwards, a study on related machine learning methods applied to Web mining have been put forward, which is followed by applications related to societal benefit areas. Finally it discusses current trends and future challenges on machine learning and Web mining.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/machine-learning-web-mining/7547](http://www.igi-global.com/chapter/machine-learning-web-mining/7547)

## Related Content

---

### ASCCN: Arbitrary Shaped Clustering Method with Compatible Nucleoids

Renxia Wan, Lixin Wang and Xiaoke Su (2012). *Exploring Advances in Interdisciplinary Data Mining and Analytics: New Trends* (pp. 219-232).

[www.irma-international.org/chapter/asccn-arbitrary-shaped-clustering-method/61177](http://www.irma-international.org/chapter/asccn-arbitrary-shaped-clustering-method/61177)

### E-Government Knowledge Management (KM) and Data Mining Challenges: Past, Present, and Future

LuAnn Bean, Deborah S. Carstens and Judith Barlow (2009). *Social and Political Implications of Data Mining: Knowledge Management in E-Government* (pp. 28-41).

[www.irma-international.org/chapter/government-knowledge-management-data-mining/29063](http://www.irma-international.org/chapter/government-knowledge-management-data-mining/29063)

### A Survey of Spatio-Temporal Data Warehousing

Leticia Gómez, Bart Kuijpers, Bart Moelans and Alejandro Vaisman (2009). *International Journal of Data Warehousing and Mining* (pp. 28-55).

[www.irma-international.org/article/survey-spatio-temporal-data-warehousing/3895](http://www.irma-international.org/article/survey-spatio-temporal-data-warehousing/3895)

### Opinion Mining for Instructor Evaluations at the Autonomous University of Ciudad Juárez

Rafael Jiménez, Vicente García, Abraham López, Alejandra Mendoza Carreón and Alan Ponce (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1280-1296).

[www.irma-international.org/chapter/opinion-mining-for-instructor-evaluations-at-the-autonomous-university-of-ciudad-juarez/308544](http://www.irma-international.org/chapter/opinion-mining-for-instructor-evaluations-at-the-autonomous-university-of-ciudad-juarez/308544)

### Enhancing the Diamond Document Warehouse Model

Maha Azabou, Ameen Banjar and Jamel Omar Feki (2020). *International Journal of Data Warehousing and Mining* (pp. 1-25).

[www.irma-international.org/article/enhancing-the-diamond-document-warehouse-model/265254](http://www.irma-international.org/article/enhancing-the-diamond-document-warehouse-model/265254)