

# Chapter 15

## Data Mining Meets Internet and Web Performance

**Leszek Borzemski**

*Wroclaw University of Technology, Poland*

### ABSTRACT

*Data mining (DM) is the key process in knowledge discovery. Many theoretical and practical DM applications can be found in science and engineering. However there are still such areas where data mining techniques are still at early stage of development and application. In particular, an unsatisfactory progress is observed in DM applications in the analysis of Internet and Web performance issues. This chapter gives the background of network performance measurement and presents our approaches, namely Internet Performance Mining and Web Performance Mining as the ways of DM application to Internet and Web performance issues. The authors present real-life examples of the analysis where explored data sets were collected with the aid of two network measurement systems WING and MWING developed at our laboratory.*

### INTRODUCTION AND MOTIVATION

Web performance is an important research area and a hot topic in Internet community. Many software resources are mirrored on hundreds of servers on the Internet. Using a nearby (in the sense of network distance) server would probably speed up the download, and reduce the load on central servers as well as on the Internet as a whole. Our ultimate aim is the development of a system that would have measure

user perceived data transfer performance between the known (observed) or unknown (non-observed) Web servers and user desktop, as well as to provide network throughput predictions on the minute-by-minute, hour-by-hour and day-by-day basis.

The particular motivation of this work is to show our approach for providing Web performance predictions by means of data mining techniques. We concentrate essential issues in the context how we apply data mining to Web performance prediction, with emphasis on the applicability to real-world problems.

DOI: 10.4018/978-1-61520-757-2.ch015

Data mining (DM) is the key process in knowledge discovery in science and engineering. Many pure theoretical and practical DM applications can be found in the literature and real-life applications. However, we notice unsatisfactory progress in the application of DM in the analysis of Internet and Web performance issues, and we hope that our contribution fills this research gap to some extent at least.

Time is the key aspect of performance evaluation of acceptable Internet and Web service levels to end users. Networks always add a delay which is more or less significant in data delivering between network nodes but it always occurs. Sometimes the delay can be estimated accurately but when a significant and changing delay occurs in the data transmission, the user application performance may suffer seriously. This degradation may cause different impacts depending on the way how the application works, including total breakdown.

The ultimate goal is to build an advanced access to Internet providing a means whereby users (whether human or machine) improve likelihood that they use network at good performance. We propose to deploy a new Internet service called Network Monitoring Broker (NMB) which would measure network performance, describe network characteristic and publish the forecasts of network behavior, especially for automatic Internet resource selection for a user domain (e.g. particular local area network).

NMB infrastructure would be established at each of the nodes of the virtual organization and can be organized in the likeness of the intermediary servers that mediate the interaction between clients and servers of the World Wide Web (Borzemski, 2007). Such service could be used in Grids and computing cloud metacomputing infrastructures to analyze current and historical data transfer performance deliverable at the application level for a set of network resources to characterize future throughput of network paths in order to find best predictions at specified action periods.

Nowadays we have developed WING (Borzemski & Nowak, 2004b) and MWING (Borzemski et al., 2007a) active measurement infrastructures to be used by the broker to measure Internet and Web in well-designed and controlled active performance experiments. We have also developed own decision-making forecasting methodology and algorithms to solve performance prediction problem formulated in our research. The broker performs Internet/Web performance forecasting using data mining methods and algorithms<sup>9</sup>. D. Barelos, E. Pitoura, G. Samaras, Mobile agents procedures: metacomputing in Java, in: Proceedings of the ICDCS Workshop on Distributed Middleware (in conjunction with the 19th IEEE International Conference on Distributed Computing Systems (ICDCS99)), Austin, TX, June 1999..

Our approach is to provide the network performance prediction service which would work similar to weather forecasting system. Network data transfer quality prediction is made in two basis steps. The first step consists of clustering the historical network performance measurements to determine classes of network behavior (the weather). These classes may be described by many measurable network parameters – in this work we use two fundamental network performance measures: *round-trip time* of data packets and *throughput* of data transfer. Each measurement is also characterized by a *time stamp* denoting the date and time at which a measurement occurred. Additional non-measurable features but also related to the event can be taken into account. Among them there is topological information such as geo-spatial parameters for users and servers, Autonomous System affiliation, routers and Autonomous Systems on user-to-server and server-to-user routes (Borzemski & Nowak, 2008).

The second step in the forecasting process consists of the determination the decision trees to descript and generalization of previously network behavior classes. The leaves represent classifications and branches represent conjunctions of

33 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/data-mining-meets-internet-web/42367](http://www.igi-global.com/chapter/data-mining-meets-internet-web/42367)

## Related Content

---

### A Review on Spatial Big Data Analytics and Visualization

Bangaru Kamatchi Seethapathy and Parvathi R (2018). *Modern Technologies for Big Data Classification and Clustering* (pp. 179-197).

[www.irma-international.org/chapter/a-review-on-spatial-big-data-analytics-and-visualization/185983](http://www.irma-international.org/chapter/a-review-on-spatial-big-data-analytics-and-visualization/185983)

### Data Mining Used for Analyzing the Bankruptcy Risk of the Romanian SMEs

Laura Giurca Vasilescu, Marian Siminica, Cerasela Pirvu, Costel Ionascu and Anca Mehedintu (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 920-946).

[www.irma-international.org/chapter/data-mining-used-analyzing-bankruptcy/73477](http://www.irma-international.org/chapter/data-mining-used-analyzing-bankruptcy/73477)

### Mining Integrated Sequential Patterns From Multiple Databases

Christie I. Ezeife, Vignesh Aravindan and Ritu Chaturvedi (2020). *International Journal of Data Warehousing and Mining* (pp. 1-21).

[www.irma-international.org/article/mining-integrated-sequential-patterns-from-multiple-databases/243411](http://www.irma-international.org/article/mining-integrated-sequential-patterns-from-multiple-databases/243411)

### Study on the Different Forms of Plagiarism in Textual Data and Image: Internal and External Detection

Frederic Jack (2019). *Advanced Metaheuristic Methods in Big Data Retrieval and Analytics* (pp. 75-90).

[www.irma-international.org/chapter/study-on-the-different-forms-of-plagiarism-in-textual-data-and-image/216094](http://www.irma-international.org/chapter/study-on-the-different-forms-of-plagiarism-in-textual-data-and-image/216094)

### Fusion Cubes: Towards Self-Service Business Intelligence

Alberto Abelló, Jérôme Darmont, Lorena Etcheverry, Matteo Golfarelli, Jose-Norberto Mazón, Felix Naumann, Torben Pedersen, Stefano Bach Rizzi, Juan Trujillo, Panos Vassiliadis and Gottfried Vossen (2013). *International Journal of Data Warehousing and Mining* (pp. 66-88).

[www.irma-international.org/article/fusion-cubes-towards-self-service/78287](http://www.irma-international.org/article/fusion-cubes-towards-self-service/78287)