

# Chapter 9

## The Performance Mining Method: Extracting Performance Knowledge from Software Operation Data

**Stella Pachidi**

*Utrecht University, The Netherlands*

**Marco Spruit**

*Utrecht University, The Netherlands*

### ABSTRACT

*Software Performance is a critical aspect for all software products. In terms of Software Operation Knowledge, it concerns knowledge about the software product's performance when it is used by the end-users. In this paper the authors suggest data mining techniques that can be used to analyze software operation data in order to extract knowledge about the performance of a software product when it operates in the field. Focusing on Software-as-a-Service applications, the authors present the Performance Mining Method to guide the process of performance monitoring (in terms of device demands and responsiveness) and analysis (finding the causes of the identified performance anomalies). The method has been evaluated through a prototype which was implemented for an online financial management application in the Netherlands.*

### INTRODUCTION

As the exponential increase of information is forcing us into the era of Big Data (Demirkan & Delen, 2013) and the organizations of the 21st century function in a global marketplace defined by intense and growing turbulence (Heinrichs & Lim, 2003), the need for tools that enable quick and effective extraction of business insights is more evident than ever. Business Analytics has evolved and includes techniques to access, report and analyze data to understand, analyze and forecast business performance (Delen,

DOI: 10.4018/978-1-4666-9840-6.ch009

& Demirkan, 2013). Such techniques include data mining, which has been used to support knowledge workers in generating relevant insights (Heinrichs & Lim, 2003) to support decision making in various processes at different strategic levels (Bolloju, Khalifa, & Turban, 2002; Courtney, 2001).

Knowledge Management has been recognized as an important process in software organizations for supporting core software engineering activities in order to decrease costs, increase quality and lead to better decisions (Bjørnson & Dingsøyr, 2008; Lindvall & Rus, 2002). Although a lot of research has focused on managing knowledge in software development (Moreno García et al, 2004), little attention has been drawn to the management of knowledge on the customers' experience with deployed software products (van der Schuur, Jansen, & Brinkkemper, 2010) for software maintenance purposes (Midha & Bhattacharjee, 2012). However, the rise of cloud computing (Marston et al., 2011) designates the need of Software-as-a-Service (SaaS) organizations to extract and analyze knowledge on how their software operates (van der Schuur et al., 2010), in order to efficiently manage with changing requirements (Srikanth & Jarke, 1989), unexpected performance issues (Zo, Nazareth & Jain, 2010) and increasing scalability (Delen, & Demirkan, 2013), to maintain quality of their services (Sun, He & Leu, 2007). With this paper we aim to contribute to the management of knowledge related to software operation (van der Schuur et al., 2010), with the goal to support software maintenance processes (Midha & Bhattacharjee, 2012). Specifically, we answer the research question: *How can we detect performance anomalies and their causes in software operation data?* We examine several data mining techniques applicable for the extraction of performance knowledge from software operation data, and construct a method that incorporates different techniques and helps structure the complicated process of performance mining. We evaluate the method and the selected data mining techniques by means of a prototype that was run with operation data of an established online financial management application.

## SOFTWARE PERFORMANCE EVALUATION

Performance is described by the proportion of the amount of effective work that is accomplished by a software system, over the time and resources used, in order to carry out this work (Arinze, Igbaria, & Young, 1992). In practice, software performance may be characterized by different aspects, which are evaluated in measurable terms using different performance metrics, such as response time, throughput, availability, latency, or utilization of resources (e.g. percentage of CPU or memory usage). Software Performance Evaluation is a critical process for all types of software products, indispensable at every stage of the product's life (Arinze et al., 1992). It is performed to determine that the system meets the user-defined performance goals and detect possible improvement points; to compare a number of alternative designs and select the best design; or to compare a number of different solutions and select the system that is most appropriate for a given set of applications (Jain, 1991).

A common technique for evaluating software performance is monitoring the system, while it is being subjected to a specific workload (Jain, 1991). Monitoring consists in observing the performance of systems, collecting performance statistics, analyzing the data and displaying the results. The commonest functionalities of a monitor include: finding the most frequently accessed software segments, measuring the utilization of resources, finding the performance bottlenecks, etc. A performance monitor can provide valuable information about the application and system run-time behavior, in order to carry out a dynamic analysis of the system's performance.

17 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/the-performance-mining-method/150164](http://www.igi-global.com/chapter/the-performance-mining-method/150164)

## Related Content

---

### Extraction of Medical Pathways from Electronic Patient Records

Dario Antonelli, Elena Baralis, Giulia Bruno, Silvia Chiusano, Naeem A. Mahotoand Caterina Petrigni (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 1004-1018).

[www.irma-international.org/chapter/extraction-medical-pathways-electronic-patient/73481](http://www.irma-international.org/chapter/extraction-medical-pathways-electronic-patient/73481)

### Mining Topological Patterns in Spatio-Temporal Databases

Wynne Hsu, Mong Li Leeand Junmei Wang (2008). *Temporal and Spatio-Temporal Data Mining* (pp. 130-165).

[www.irma-international.org/chapter/mining-topological-patterns-spatio-temporal/30265](http://www.irma-international.org/chapter/mining-topological-patterns-spatio-temporal/30265)

### A Survey of Spatio-Temporal Data Warehousing

Leticia Gómez, Bart Kuijpers, Bart Moelansand 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)

### Zero-Shot Feature Selection via Transferring Supervised Knowledge

Zheng Wang, Qiao Wang, Tingzhang Zhao, Chaokun Wangand Xiaojun Ye (2021). *International Journal of Data Warehousing and Mining* (pp. 1-20).

[www.irma-international.org/article/zero-shot-feature-selection-via-transferring-supervised-knowledge/276762](http://www.irma-international.org/article/zero-shot-feature-selection-via-transferring-supervised-knowledge/276762)

### Mining Top-k Regular High-Utility Itemsets in Transactional Databases

P. Lalitha Kumari, S. G. Sanjeeviand T.V. Madhusudhana Rao (2019). *International Journal of Data Warehousing and Mining* (pp. 58-79).

[www.irma-international.org/article/mining-top-k-regular-high-utility-itemsets-in-transactional-databases/223137](http://www.irma-international.org/article/mining-top-k-regular-high-utility-itemsets-in-transactional-databases/223137)