Chapter 8

Data Mining User Activity in Free and Open Source Software (FOSS)/ Open Learning Management Systems

Owen McGrath
University of California, Berkeley, USA

ABSTRACT

Free and Open Source Software (FOSS)/Open Educational Systems development projects abound in higher education today. Many universities worldwide have adopted open source software like ATutor and Moodle as an alternative to commercial or homegrown systems. The move to open source learning management systems entails many special considerations, including usage analysis facilities. The tracking of users and their activities poses major technical and analytical challenges within web-based systems. This chapter examines how user activity tracking challenges are met with data mining techniques, particularly web usage mining methods, in four different open learning management systems: ATutor, LON-CAPA, Moodle, and Sakai. As examples of data mining technologies adapted within widely used systems, they represent important first steps for moving educational data mining outside the research laboratory. Moreover, as examples of different open source development contexts, exemplify the potential for programmatic integration of data mining technology processes in the future. As open systems mature in the use of educational data mining, they move closer to the long-sought goal of achieving more interactive, personalized, adaptive learning environments online on a broad scale.

INTRODUCTION

At a basic level, collegial sharing of source code has been around as long as computing itself in academia (Roof & Frazier, 1962). What has varied over the years and across contexts is how these

code sharing efforts are organized. The history of open source in academic research and instruction has received prominence in recent analyses of how the open source phenomena got started. The success of open source, it is argued, has as much to do with the various open source processes and organizational arrangements as with the code itself (Weber, 2004). The four open source learning

management systems that serve as backdrops for the data mining projects described in this chapter represent a spectrum of higher education open source undertakings. As projects and products, ATutor, LON-CAPA, Moodle, and Sakai differ in terms of technology, design, size, target audiences, organizational structures, funding models, software licensing choices, extendibility, and their pace of development. In no small way, analyzing why and how educational data mining techniques are being employed in these four open source contexts also involves understanding the underlying organizations and processes, not just the technologies and system designs.

Data mining of web-based educational systems is also not particularly new. For more than a decade, exciting and visionary applications of educational data mining in web-based learning systems have been conceived, prototyped, and tested (Johnson, 1998). As Romero and Ventura show, much of the work to date has been carried out mainly within the lifecycle and setting of individual research and development projects with small populations of research subjects (Romero & Ventura, 2007). In contrast, the projects described in this chapter involve data mining technologies and techniques being adapted to widely available open source learning management systems. For many reasons, traditional techniques for studying online users have proven inadequate when applied in systems of the scale and complexity seen in these widely adopted, open technology and open learning management systems (McGrath, 2008b). As an alternative for studying usage, educational data mining techniques are being embraced because the techniques excel at uncovering usage structure and access patterns within the enormous volume and variety of loosely structured user data generated in large systems. And while exploratory in nature, the four projects described here are not only for research and development. They lean more towards developing practical approaches to building and managing distributed capture and analysis systems for large web-based production deployments in

production at many universities and colleges. The four projects involve fairly simple applications of data mining, from external usage analytics to association-rule based feedback, that only begin to tap the fuller potential educational data mining has shown in pure research prototype systems. Two are grant funded research experiments, but in all four cases practical needs have motivated the turn towards applying the algorithms of data mining and machine learning as a way of understanding users and usage.

BACKGROUND

As more teaching, research, and administrative processes are moved online, there arise both needs and opportunities for tracking, analyzing, and improving the online institutional systems involved. Important to the success of these online environments, for instance, is meeting increasing demands to document educational outcomes. Even as the distance between students, faculty, and administrators decreases online, demands for more detailed and rigorous accounting of teaching activities, curricular progress, and learning outcomes are increasing. Before the advent of large cross-institutional open source projects, e-learning tracking and reporting functionality was typically provided by software vendors. An educational technology group at a college or university could confine its efforts to meeting the mandates and needs of the local institution by using the reporting features of the commercial systems. In the case of open source applications, are often left to consider how information should be captured and managed, how to ensure that these requirements are ever included in the development process, or how to address these needs in other ways (e.g., data mining) when they're not.

Behind open and community source projects such as ATutor, LON-CAPA, Moodle, and Sakai are organizations still experimenting with management structures that might allow them better

Related Content

Bridging the Gap between Agile and Free Software Approaches: The Impact of Sprinting
Paul J. Adamsand Andrea Capiluppi (2009). *International Journal of Open Source Software and Processes (pp. 58-71).*
www.irma-international.org/article/bridging-gap-between-agile-free/2771

A Preventive Action Management Platform in Healthcare Information Systems
Hugo Peixoto, António Abelha, Manuel Santosand José Machado (2015). *Open Source Technology: Concepts, Methodologies, Tools, and Applications (pp. 447-460).*
www.irma-international.org/chapter/a-preventive-action-management-platform-in-healthcare-information-systems/120930

Software Fault Prediction Using Deep Learning Algorithms
Osama Al Qasemand Mohammed Akour (2019). *International Journal of Open Source Software and Processes (pp. 1-19).*
www.irma-international.org/article/software-fault-prediction-using-deep-learning-algorithms/242945

Internet Policy Issues and Digital Libraries' Management of Intellectual Property
Adeyinka Tellaand A. K. Afolabi (2015). *Open Source Technology: Concepts, Methodologies, Tools, and Applications (pp. 890-901).*
www.irma-international.org/chapter/internet-policy-issues-and-digital-libraries-management-of-intellectual-property/120947

Evaluating the Potential of Free and Open Source Software in the Developing World
Victor van Reijswoudand Emmanuel Mulo (2012). *International Journal of Open Source Software and Processes (pp. 38-51).*
www.irma-international.org/article/evaluating-the-potential-of-free-and-open-source-software-in-the-developing-world/101205