

Applying Data Mining Techniques to e-Learning Problems

Félix Castro^{1,2}, Alfredo Vellido¹, Àngela Nebot¹, and Francisco Mugica³

¹ Dept. Llenguatges i Sistemes Informàtics, LSI,

Universitat Politècnica de Catalunya,

Campus Nord, C. Jordi Girona 1-3,

Barcelona 08034, España {fcastro, avellido, angela}@lsi.upc.edu

² Centro de Investigación en Tecnologías de Información y Sistemas,

CITIS, Universidad Autónoma del Estado de Hidalgo,

Ciudad Universitaria, Carretera Pachuca-Tulancingo km. 4.5,

Hidalgo, México fcastro@lsi.upc.edu

³ Instituto Latinoamericano de la Comunicación Educativa (ILCE),

Calle del Puente 45, México D. F. 14380, México

fmugica@ilce.edu.mx

Summary. This chapter aims to provide an up-to-date snapshot of the current state of research and applications of Data Mining methods in e-learning. The cross-fertilization of both areas is still in its infancy, and even academic references are scarce on the ground, although some leading education-related publications are already beginning to pay attention to this new field. In order to offer a reasonable organization of the available bibliographic information according to different criteria, firstly, and from the Data Mining practitioner point of view, references are organized according to the type of modeling techniques used, which include: Neural Networks, Genetic Algorithms, Clustering and Visualization Methods, Fuzzy Logic, Intelligent agents, and Inductive Reasoning, amongst others. From the same point of view, the information is organized according to the type of Data Mining problem dealt with: clustering, classification, prediction, etc. Finally, from the standpoint of the e-learning practitioner, we provide a taxonomy of e-learning problems to which Data Mining techniques have been applied, including, for instance: Students' classification based on their learning performance; detection of irregular learning behaviors; e-learning system navigation and interaction optimization; clustering according to similar e-learning system usage; and systems' adaptability to students' requirements and capacities.

1 Introduction

Within a decade, the Internet has become a pervasive medium that has changed completely, and perhaps irreversibly, the way information and knowledge are transmitted and shared throughout the world. The education community has not limited itself to the role of passive actor in this unfolding story, but it has been at the forefront of most of the changes.

Indeed, the Internet and the advance of telecommunication technologies allow us to share and manipulate information in nearly real time. This reality is determining the next generation of distance education tools. Distance education arose from traditional education in order to cover the necessities of remote students and/or help the teaching–learning process, reinforcing or replacing traditional education. The Internet takes this process of delocalization of the educative experience to a new realm, where the lack of presential intercourse is, at least partially, replaced by an increased level of technology-mediated interaction. Furthermore, telecommunications allow this interaction to take forms that were not available to traditional presential and distance learning teachers and learners.

This is e-learning (also referred to as Web-based education and e-teaching), a new context for education where large amounts of information describing the continuum of the teaching–learning interactions are endlessly generated and ubiquitously available. This could be seen as a blessing: plenty of information readily available just a click away. But it could equally be seen as an exponentially growing nightmare, in which unstructured information chokes the educational system without providing any articulate knowledge to its actors.

Data Mining was born to tackle problems like this. As a field of research, it is almost contemporary to e-learning. It is, though, rather difficult to define. Not because of its intrinsic complexity, but because it has most of its roots in the ever-shifting world of business. At its most detailed, it can be understood not just as a collection of data analysis methods, but as a data analysis process that encompasses anything from data understanding, preprocessing and modeling to process evaluation and implementation [16]. It is nevertheless usual to pay preferential attention to the Data Mining methods themselves. These commonly bridge the fields of traditional statistics, pattern recognition and Machine Learning (ML) to provide analytical solutions to problems in areas as diverse as biomedicine, engineering, and business, to name just a few. An aspect that perhaps makes Data Mining unique is that it pays special attention to the compatibility of the modeling techniques with new Information Technologies