



Review

Educational data mining: A survey and a data mining-based analysis of recent works



Alejandro Peña-Ayala*

WOLNM & ESIME Zacatenco, Instituto Politécnico Nacional, U. Profesional Adolfo López Mateos, Edificio Z-4, 2do piso, cubículo 6, Miguel Othón de Mendizábal S/N, La Escalera, Gustavo A. Madero, D.F., C.P. 07320, Mexico

ARTICLE INFO

Keywords:

Data mining
Educational data mining
Data mining profile
Educational data mining approach pattern
Pattern for descriptive and predictive educational data mining approaches

ABSTRACT

This review pursues a twofold goal, the first is to preserve and enhance the chronicles of recent educational data mining (EDM) advances development; the second is to organize, analyze, and discuss the content of the review based on the outcomes produced by a data mining (DM) approach. Thus, as result of the selection and analysis of 240 EDM works, an EDM work profile was compiled to describe 222 EDM approaches and 18 tools. A profile of the EDM works was organized as a raw data base, which was transformed into an ad-hoc data base suitable to be mined. As result of the execution of statistical and clustering processes, a set of educational functionalities was found, a realistic pattern of EDM approaches was discovered, and two patterns of value-instances to depict EDM approaches based on descriptive and predictive models were identified. One key finding is: most of the EDM approaches are ground on a basic set composed by three kinds of educational systems, disciplines, tasks, methods, and algorithms each. The review concludes with a snapshot of the surveyed EDM works, and provides an analysis of the EDM strengths, weakness, opportunities, and threats, whose factors represent, in a sense, future work to be fulfilled.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Data mining (DM¹) is a computer-based information system (CBIS) (Vlahos, Ferratt, & Knoepfle, 2004) devoted to scan huge data repositories, generate information, and discover knowledge. The meaning of the traditional mining term biases the DM grounds. But, instead of searching natural minerals, the target is *knowledge*. DM pursues to find out data patterns, organize information of hidden relationships, structure association rules, estimate unknown items' values to classify objects, compose clusters of homogenous objects, and unveil many kinds of findings that are not easily produced by

a classic CBIS. Thereby, DM outcomes represent a valuable support for decisions-making.

Concerning education, it is a novel DM application target for knowledge discovery, decisions-making, and recommendation (Vialardi-Sacin, Bravo-Agapito, Shafti, & Ortigosa, 2009). Nowadays, the use of DM in the education arena is incipient and gives birth to the *educational data mining* (EDM) research field (Anjewierden, Kollöffel, & Hulshof, 2007). As we will see in Section 2, in a sense the first decade of the present century represents the kick-off of EDM.

EDM emerges as a paradigm oriented to design models, tasks, methods, and algorithms for exploring data from educational settings. EDM pursues to find out patterns and make predictions that characterize learners' behaviors and achievements, domain knowledge content, assessments, educational functionalities, and applications (Luan, 2002). Source information is stored in repositories managed by conventional, open, and distance educational modalities.

Some of the EDM trends are anticipated here. One of them corresponds to the standard integration of an EDM module to the typical architecture of the wide diversity of computer-based educational systems (CBES). Other tendency demands that EDM provides several functionalities during three stages of the teaching-learning cycle. The first stage corresponds to the provision of EDM *proactive* support for adapting the educational setting according to the student's profile prior to deliver a lecture. During

* Address: WOLNM & ESIME Zacatenco, Instituto Politécnico Nacional, U. Profesional Adolfo López Mateos, Edificio Z-4, 2do piso, cubículo 6, Miguel Othón de Mendizábal S/N, La Escalera, Gustavo A. Madero, D.F., C.P. 07320, Mexico. Tel.: +52 55 5694 0916/+52 55 5454 2611 (cellular); fax: +52 55 5694 0916.

E-mail address: apenaa@ipn.mx

URL: <http://www.wolnm.org/apa>

¹ AIWBES: adaptive and intelligent web-based educational systems; BKT: Bayesian knowledge tracing; CBES: computer-based educational systems; CBIS: computer-based information system; DM: data mining; DP: dynamic programming; EDM: educational data mining; EM: expectation maximization; HMM: hidden Markov model; IBL: instances-based learning; IRT: item response theory; ITS: intelligent tutoring systems; KDD: knowledge discovery in databases; KT: knowledge tracing; LMS: learning management systems; SNA: social network analysis; SWOT: strengths, weakness, opportunities, and threats; WBC: web-based courses; WBES: web-based educational systems.

the student-system interaction stage, it is desirable that EDM acquires log-data and interprets their meaning in order to suggest recommendations, which can be used by the CBES for personalizing services to users at real-time. In the next stage, EDM should carry out the evaluation of the provided education concerning: delivered services, achieved outcomes, degree of user's satisfaction, and usefulness of the resources employed. What is more, several challenges (i.e., targets, environments, modalities, functionalities, kinds of data, ...) wait to be tackled or have been recently considered by EDM, such as: big data, cloud computing, social networks, web mining, text mining, virtual 3-D environments, spatial mining, semantic mining, collaborative learning, learning companions, ...

The present work extends the period described by earlier surveys, summarized in Section 2.2, that cover from 1995 up to 2009. The aim is to preserve and update the chronicles of recent EDM development. The scope of the work is limited and provides a partial image of the EDM activity published in all celebrated events and available media. In spite of this, the work provides a snapshot of the EDM labor that several members have been achieving. Inclusively, it applies the essential subject, DM, to organize, analyze, and discuss the content of the overview. Such a policy is a novelty: to preach through example.

As result of the application of such a policy, the next four contributions are offered to be used by the EDM community: a DM profile, an EDM approach profile, a pattern for EDM approaches based on descriptive models, and a pattern for EDM approaches based on predictive models. The first facilitates the description of the DM baseline that supports an EDM approach. The second is useful to define the nature and baseline of an EDM approach. The third and four are patterns to design EDM approaches, which are useful as a reference to develop similar versions of descriptive and predictive models.

In this paper a survey of EDM works fulfilled from 2010 up to 2013 1st Qtr. is presented. In addition, the method followed for producing the overview is outlined in Section 2, as well as the gathered material is stated. A *sample* of 240 EDM works is summarized in Section 3. Such a collection is organized according to typical functionalities fulfilled by CBES that were found from the material. In Section 4, an analysis of the sampled works is provided to shape the recent status and evolution of the EDM field, and some EDM approach patterns are highlighted. Finally, the conclusions

Section tailors a snapshot of the sample and a critical analysis of the EDM arena that are useful to inspire future work.

2. Method and materials

In this section, the method and the materials of the overview are described. The method is a framework devoted to gather and mine EDM works. The materials tailor the survey domain through five subjects: a reference to prior EDM reviews, the scope of the collected EDM works, a profile of DM, a summary of CBES, and the data representation of EDM approaches used for mining.

As a result of the method application, a *sample* of 240 EDM works published between 2010 and the first quarter of 2013 was gathered. It is made up of two *sub-samples*, one of 222 EDM approaches and another of 18 EDM tools (i.e., the first represents EDM applications and the second software). The sample symbolizes a valuable source that is used to provide a highlight of the EDM arena in Section 3 and a brief analysis in Section 4. Moreover, the sample is examined to produce statistics and discover some findings, which are illustrated in the following subsections as well as in Sections 3 to 5.

2.1. Framework applied for knowledge discovery of educational data mining works

The method used to carry out this survey is a framework designed to gather, analyze, and mine EDM works. It follows a workflow to lead the activities oriented to knowledge discovery in databases (KDD). The workflow is split into three stages. The development of each stage is achieved by three tasks. Thus, nine tasks compose the whole KDD workflow pictured in Fig. 1, whose purpose and outcomes are explained as follows:

The “EDM work collection” stage performs three tasks. The first task seeks source references that publish EDM works. As a result, a collection of EDM works is gathered. The second evaluates EDM works and produces an *EDM approach profile* per each chosen EDM work. The third analyzes the EDM approach profiles and organizes a *raw EDM database*.

The “data processing stage” encompasses the tasks labeled as fourth, fifth, and sixth in Fig. 1. The fourth task transforms the raw EDM database into an *ad-hoc EDM database* to facilitate statis-

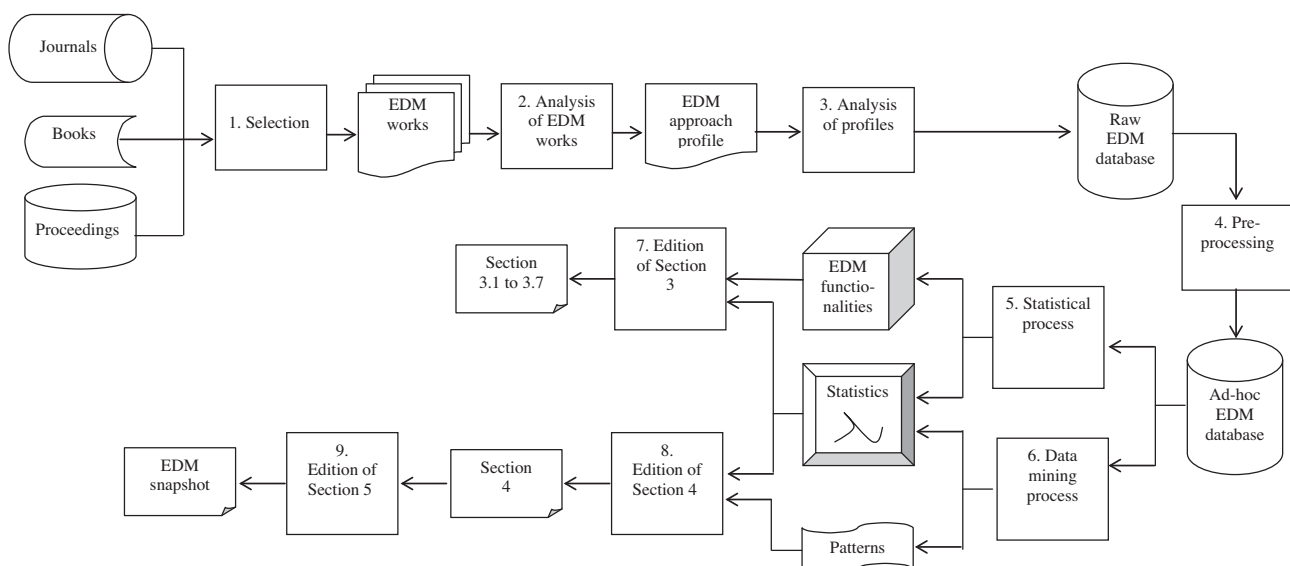


Fig. 1. Workflow of the DM approach performed to analyse, classify, represent, and mine data of the EDM related works.

Download English Version:

<https://daneshyari.com/en/article/386814>

Download Persian Version:

<https://daneshyari.com/article/386814>

[Daneshyari.com](https://daneshyari.com)