Chapter 14
Data Mining Applications in Computer-Supported Collaborative Learning

Rosanna Costaguta
Universidad Nacional de Santiago del Estero (UNSE), Argentina

ABSTRACT
Our own data mining techniques allow us to discover non-explicit knowledge from a large amount of data. Currently, Computer-Supported Collaborative Learning systems generate a wealth of data, derived from the stored interactions and product of collaborative work of students and teachers. Manual processing of these interactions is both costly and tedious, and practically impossible to do in real time. Because of this, there are now trends of research that attempt to achieve automatic processing using data-mining techniques. This chapter describes the phases and tasks involved in the entire process of knowledge discovery and also presents some research applying data mining to process the contributions of students and teachers in collaborative-learning environments.

INTRODUCTION
In the last decade the use of computers to automate different processes has come to be a major event. This situation caused a significant increase in the amount of digitally stored data. It is even possible to affirm that the variety of used data types has also increased. This new reality is easily evident in areas such as e-commerce, virtual education and mobile communication, just to name a few of the areas that have evolved considerably in recent years. This increased availability of digital data is positive for any organization, either public or private, since in addition to being a sort of back up or organizational memory they represent an invaluable source of knowledge for people who occupy decision-making positions in these organizations.

Currently it is observed that although most of the data sources are digital, the largest number of data processing for this issue, including analysis and interpretation, is performed manually. Organizations that use some kind of software to perform this task are rare. Examples abound. Until not long ago TV channels in our country transmitted campaigns aimed at combating dengue, responding to a policy designed by the Ministry of Health
of the Nation. It may be asked how they came to that decision. It is possible to imagine that all the provinces affected by the disease provided its capital with data regarding the sick and contagious, and that decisions at national level were taken through meetings of provincial Ministers of Health together with their counterpart at the national level. Thus, by successive meetings the current data were analyzed and interpreted (perhaps using some statistical software), other epidemics suffered in the country and in neighboring countries were noted, and as a result of dialogue and negotiation, decisions were “manually” taken that led to the implementation of measures such as communication campaigns. Of course this approach is not wrong, but is more expensive, slower and has a strong sense of subjectivity of the participants. Furthermore, when both the quantity and variety of data type are numerous, proceeding in this way is impossible if knowledge of quality and timeliness are required for proper decision-making, simply because it exceeds our human capacities for analysis and interpretation. This is where Informatics provides specific techniques and methods to overcome these difficulties, encompassed in the Process of Discovery in Databases (aka data mining).

The education community has also suffered the impact of the changes and technological advances. In the last decade the adoption of the Internet as a way to reach students has inevitably changed the way in which knowledge and information are transmitted and shared by teachers and students. This reality has been instrumental in the emergence of new tools applicable to virtual education or e-learning. One of these approaches is the Computer Supported Collaborative Learning, which allows groups of students to work together in pursuit of a goal, sharing resources, skills and knowledge, having dialogues, collaborating and coordinating its activities through the use of the computer, in a way completely independent of the time and place variables. In collaborative environments numerous interactions between actors occur, but as it was said in previous paragraph, the difficulty arises when analyzing these data to generate useful information aimed at improving the teaching and learning process. Therefore, data mining is a promising option in the field.

This chapter is organized as follows. In section 2 constituent phases of the Process of Knowledge Discovery in Databases are introduced, and descriptive and prescriptive task types are described, to which the process can be oriented. In Section 3 some research in the area of CSCL is presented, applying data mining techniques. Finally, some conclusions are set forth in section 4.

THE KNOWLEDGE DISCOVERY PROCESS

Constant progress in computer science and associated technologies, together with the growing expansion of their use in various aspects of life has turned the storage of large amounts of data into a frequent situation. Manually scan this accumulation of data in real-time could be tedious and expensive, and usually a humanly impossible task to carry out.

The Process of Knowledge Discovery in Databases (KDD) emerged as one of the possibilities to meet the need of a methodology for intelligent data analysis aimed at extracting useful knowledge (Riquelme, Ruíz, Gilbert, 2006). KDD is understood as the process of identifying, in large volumes of data, significant patterns that are valid, novel, potentially useful and understandable for a user (Hernández-Orallo, Ramírez-Quintana, & Ferri, 2006). The KDD process consists of five phases: Integration and compilation; Selection, Cleaning and Transformation; Data mining; Evaluation and Interpretation; Dissemination and Use (Han & Kamber, 2001; Hernández-Orallo, Ramírez-Quintana & Ferri 2006; Witten & Frank, 1999).

The phase of Integration and Compilation aims to gather the data required for the next phase of the KDD process. At this stage it is useful to