Journal of Universal Computer Science, vol. 21, no. 1 (2015), 23-47
submitted: 16/5/14, accepted: 1/10/14, appeared: 1/1/15 © J.UCS

Dropout Prediction and Reduction in Distance Education Courses with the Learning Analytics Multitrail Approach

Wagner Cambruzzi, Sandro José Rigo, Jorge L. V. Barbosa
(Universidade do Vale do Rio dos Sinos - UNISINOS, São Leopoldo, Brazil
wagner@cambruzzi.com.br, rigo@unisinos.br, jbarbosa@unisinos.br)

Abstract: Distance Education courses are present in large number of educational institutions. Virtual Learning Environments development contributes to this wide adoption of Distance Education modality and allows new pedagogical methodologies. However, dropout rates observed in these courses are very expressive, both in public and private educational institutions. This paper presents a Learning Analytics system developed to deal with dropout problem in Distance Education courses on university-level education. Several complementary tools, allowing data visualization, dropout predictions, support to pedagogical actions and textual analysis, among others, are available in the system. The implementation of these tools is feasible due to the adoption of an approach called Multitrail to represent and manipulate data from several sources and formats. The obtained results from experiments carried out with courses in a Brazilian university show the dropout prediction with an average of 87% precision. A set of pedagogical actions concerning students among the higher probabilities of dropout was implemented and we observed average reduction of 11% in dropout rates.

Keywords: Learning Analytics, Distance Education, Virtual Learning Environments
Categories: L.3 L.3.5, L.3.6

1 Introduction

Nowadays, Distance Education courses are present in large number of public and private educational institutions. Some relevant objectives of online Distance Education courses depend on the student interaction, usually supported by digital mediation tools and digital media elements. This interaction is facilitated by wide availability of suitable technological resources [Longo, 09] and generates a context of data sets production with features regarding the learning process. These data sets constitute valuable source of information about student individual learning process.

Despite positive aspects in Distance Education proposals, dropout rates observed in these courses are very expressive and call attention of research groups, government and other public institutions. Several reports are known about this situation. Some examples are the OECD (Organization for Economic Co-operations and Development) reports on Distance Education [OECD, 13], the national reports of countries such as United States of America [Villazón-Terrazas, 11] or reports from organizations dedicated to Distance Education, such as ABED (*Associação Brasileira de Educação a Distância* - Brazilian Distance Education Association) [ABED, 13]. These reports identify a very meaningful dropout rate.

In order to mitigate dropout, the application of Educational Data Mining (EDM) and Learning Analytics (LA) techniques can be observed in several works [Romero,

10; Baker, 11; Manhães, 11; Kampff, 09; Durand, 11; Li, 11; Toscher, 10; Romero, 08; Nandeshwar, 11; Kotsiantis, 11]. The successful application of EDM resources is related to well-established Data Mining techniques and also to large amount of interaction data describing the students' actions. As a result of Virtual Learning Environment (VLE) massive use, we can observe large amount of data in digital format, originated from digital mediation tools. The major part of these data sets represents interaction aspects. As some examples, we can mention date and time information regarding student VLE access log, students' grades in exercises, textual interactions in discussion forums, or Learning Objects utilization log. Regarding Data Mining techniques, a diversity of approaches can be observed in works ranging from application of Artificial Neural Networks [Barker, 04], Association Rules [Ceglar, 06], Sequence Mining [Kay, 06], among others.

There are valuable possibilities in the application of EDM and LA techniques over data generated in Virtual Learning Environments and other educational information systems. These systems accumulate vast material and large amount of information, therefore fostering natural integration of LA initiatives in Distance Education context [Romero, 10]. The use of LA allows dealing with aspects such as visualization of students' grades and interaction, generation of behavioral patterns and creation of alternative support to learning activities [Huebner, 13]. Although results obtained with application of EDM techniques and LA systems can be considered satisfactory [Romero, 10; Baker, 11; Huebner, 13], in general their evaluation scope is delimited by some subset of attributes selected from available data. A considerable part of known experiments published are based in some planned situation in which only part of factors involved in real cases of dropout are identified. These approaches are limited in some aspects due to the fact of dealing with some subset of pertinent information. One of the reasons for this is related to the great number of possible dropout causes, which are of complex nature and refer to personal, social and institutional aspects. Therefore, this work aims to analyse a broader set of attributes related to students' behavior, in order to evaluate the obtained results in predictions, prevention and reduction of the dropout behaviour.

The necessity of interdisciplinary approach and construction of more complex scenarios to be applied in student dropout behavior analysis is supported by several studies about the dropout phenomenon, stating that there is a complex set of possible origins and correlated causes that can be related to dropout behavior [Tinto, 75; Scott, 11; Levy, 07]. Some of these studies identify the general tendencies in social or economic fields as determinant aspects of dropout. The more specific and individually based facets, such as personal characteristics are also considered relevant, together with other aspects derived from institutional policies and related to teacher attitudes. Several examples in this regard can be stated, such as the questions related to course choice, adopted course methodology, instructional material aspects, teacher interaction or even institutional support to student needs [Adachi, 09].

In this work, we propose an approach that relies on interdisciplinary team to deal with dropout phenomenon. To explore the possibilities of this approach in online Distance Education courses on university-level education, we describe an experience with university-level courses in online Distance Education. The main objective of this work was to implement a LA system to dropout prediction that could also support and integrate pedagogical actions to reverse identified dropout tendencies. The student

interaction was chosen, among several factors, as an important aspect to observe in dropout prediction. One reason for this approach is interaction importance in avoiding dropout due to the relation of interaction, participation and satisfactory results in education [Moore, 07]. Another reason are the existing facilities in data acquisition describing the student's interaction in the Virtual Learning Environment and other academic systems. Virtual Learning Environments are excellent platforms to register interaction generated by students and other complementary or supplementary records can be integrated to this initial data base, such as Academic Information Systems, Learning Objects repositories, video lectures repositories or digital social networks platforms, among other possibilities.

Due to the complex nature of dropout problem, we proposed an approach to represent and manipulate data generated from several sources, called Multitrail architecture. A LA system was developed [Cambruzzi, 12] in order to support data integration operations, data manipulation, dropout prediction and visualization. There is a component in the system dedicated to support teacher's actions, generating messages alerting teachers to some necessities of attention to be directed to specific students, due to dropout prediction results. Another relevant component of this system is a tool to record and manipulate pedagogical actions performed by teachers. Therefore, results of pedagogical actions can be analysed regarding perceived outcome.

The text is organized into five sections. Section two reviews the Distance Education, dropout and Educational Data Mining concepts, providing the basis for the rest of text. Section 3 describes the Learning Analytics system developed, mainly focusing its architecture and strategy employed to support multiple data sources and multiple applications. The fourth section describes a case study conducted in this work and the prediction mechanism implemented and evaluation of obtained results. Section 5 presents final remarks and directions for future research.

2 Background

In this section we describe Distance Education aspects, dropout concepts and approaches to deal with dropout prediction using Educational Data Mining techniques.

2.1 Distance Education and the dropout problem

The Distance Education allows learning in a flexible way, where teachers and students can be apart geographically and temporally. Mediation is carried out with resources in different media support [Carr, 12]. According to KEEGAN [Keegan, 96], what differentiates Distance Education and self-study initiatives is the possibility of multiple communication aspects, where students and teachers can interact and communicate electronically and in occasional meetings. Besides that, in most part of Distance Education courses there are also group interaction activities, related to the development of several important students´ skills. Depending on the educational activity can be observed a combination of autonomy, structure and dialogue and the balance of these aspects should be a concern in the planning and monitoring of a given course [Moore, 93].

As quoted by KAMPFF [Kampff, 09], success of Distance Education courses can be improved by effective systems for monitoring and evaluating, because these systems can help teachers to identify student's specific problems and offer appropriate support. According to Moore and Kearsley [Moore, 07], effective monitoring requires a network of indicators that provide information on student achievement. Besides that, the authors state that this should be a process to be done frequently and routinely. The high rates of dropout in distance education are a concern present in Brazil [Censo, 11], as well as in other countries [Scott, 11; Levy, 07]. In the study presented by Carr [Carr, 12] the evasion rate hovered around 50% in the United States of America. In 2004, in the study presented by Maya [Maia, 04], in which participate 22 Brazilian institutions and 51 thousand students, rates of dropout were estimated in around 30%. The work of Longo [Longo, 09] features international indexes with values above 65% of dropout. Among other organizations, the OECD [OECD, 13] maintains historical data, showing that the dropout problem is present in many countries.

The term dropout allows various interpretations. In some cases a dropout event is considered to be the course withdrawal by the student, regardless the amount of classes assisted. In other situations the dropout is differently considered according to the average period for course completion [Favero, 06]. As a consequence, some indexes measure withdrawal in a particular discipline of a course but other ones measure the withdrawal of the entire course [Adachi, 09]. Therefore, these studies consider some conceptual differences for the dropout. They consider it differently when the dropout is related to the discipline, when the dropout occurs within the course, within the institution and even when the student withdraws from the educational system itself.

Educational institutions are engaged in actions to identify variables associated with the dropout behavior. This information is afterward used in preventive actions, in order to minimize their possible effect. In some cases this identification can be made with information composed by social, motivational and educational history of student. In other situations information is very dynamic, as it is observed in relation to teaching skills or interaction and collaboration between students over the period of one academic semester. Some works identify factors influencing dropout rates and also the opposite aspects of this phenomenon that would be the presence and persistence of the student. The technical approach observed is mainly related to the use of specific forms to collect information that can establish a profile of students. Besides that approach, some initiatives apply statistical methods and the extensive use of database maintained information [Joa, 11; Araque, 09]. Once identified, the variables and general conditions related to dropout behavior are then used to help in future actions to prevent this situation. In general, the treated aspects in these approaches are more stable information, regarding the age, sex, social and historic information. In Distance Education context, some studies investigate the set of factors that can be of relevant influence in learner´s performance and satisfaction. For instance, in [Wannasiri, 12] is described a study that collected information from directly involved person in Distance Education courses, that was used to reveal critical factors for success of Distance Education courses in developing countries.

2.2 Educational Data Mining and dropout prediction in Distance Education courses

Currently the data generation capacity is much higher than the capacity of researchers to analyse the stored data [Liao, 12]. This statement is also true in the field of Education, because grows every day the volume of data generated and stored in databases, due to the widespread use of computerized systems in schools and universities. Therefore, the existence of this large volume of data has encouraged the use of data mining techniques, in order to help in searching for answers to specific education questions, related to learning processes, development of instructional materials, monitoring and prediction, among others [Baker, 11; Manhães, 11]. Educational Data mining deals with the application of Data Mining techniques [Fayyad, 96] with new sets of data obtained in different educational contexts. The nature of these data represents a potential implementation of resources that are fundamental to help in improvement of education [Romero, 10; Romero, 12]. Some examples are generation of alerts [Kampff, 09], support for recommendation systems [Durand, 11; Toscher, 10] or capture of student´s profiles [Li, 11].

There are suitable approaches to deal with data with different aspects regarding their volatility, considering sets of more stable data and more dynamic data. The first one is mainly formed by issues related to educational history and social aspects, which are observed with more static features and representing chronological aspects. In general these sets of data are treated in order to generate information that can be effective in helping institutional prevention initiatives with pedagogical actions. The second one is composed mainly by data identified in the daily interaction and activities. The main actions that these kinds of data allow are those related to the evolution of activities in short time periods, typically the ones observed in the period of one semester. The results from activities within a discipline can be used to generate diagnostics that indicate actions within a smaller time scope, dealing with possible immediate dropout situations. Distance Education courses pose an opportunity, due to their nature, that involves a high level of digital mediation and, therefore, can provide a substantial amount of data regarding students [Romero, 08].

Some works are mainly based in the acquisition of additional data that can help to explain the dropout situation [Levy, 07]. By additional data, we should understand data that is not generated normally by student action, such as the access to the Virtual Learning Environment and the accomplishment of the activities proposed. In these cases, some specific instruments, such as electronic forms, can be used in order to collect the student's opinion on some matter that can be related to dropout, such as the student satisfaction, for instance. Another approach relies only in automatically generated data and in the use of these sets to generate patterns to help in the situation understanding [Nistor, 10]. The academic practice can be used to identify collaboration patterns that are established among students with different levels of participation and with different levels of performance. For example, some works demonstrates a difference in the participation pattern of students with high grades as compared with students that did give up courses [Romero-Zaldivar, 12; Lykourentzou, 09]. Once identified, these patterns can be applied in future prevention actions.

However, some works describe an approach that integrates more broadly the preliminary study of factors to be monitored by Educational Data Mining techniques,

in order to compose scenarios that are consistent with knowledge about the general process of dropout [Romero, 10]. The main focus of important part of known works in Educational Data Mining is concentrated in activities looking for experimentation of diverse Data mining techniques and their adaptation to educational data context. Frequently the main objective expressed is the validation of a specific technique and identification of aspects to be monitored.

3 Proposed Approach

An interdisciplinary approach to Educational Data Mining and Learning Analytics was adopted in this work, with the participation of pedagogical and course management staff together with computer science professionals. There was an interaction between the project development technical team and related institutional sectors. Due to this interaction, the system conception was based on interdisciplinary studies about dropout causes, in order to identify the various possible factors of influence in this subject. Therefore, some aspects were chosen as system requisites: a) to represent several data sources; b) to allow a flexible way to manipulate different data sources; c) to simplify incorporation of future changes in data formats; d) to permit inclusion of new data sources. The situation demanding the flexibility to represent and treat interaction data is represented, in a simplified way, in Figure 1, where several possible data sources are illustrated, related to educational, professional or social aspects, among others. Each of these aspects can affect student performance and contribute to dropout events. Therefore, they need to be represented and remain available to be used as data sources in different applications.

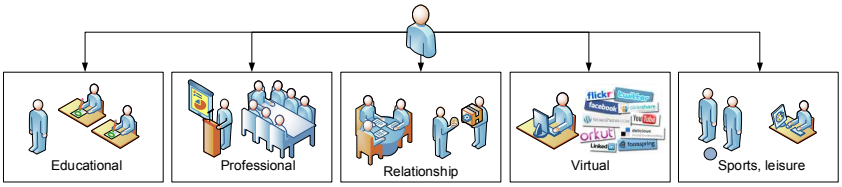


Figure 1: Student and several examples of data sources

The Learning Analytics system developed allows dropout prediction, generated with Machine Learning techniques applied over interaction data regarding students´ learning process. Results of these predictions are then used in the choice of pedagogical actions to be taken. The system includes implementation of a process with the following steps: knowledge discovery, registration of interesting patterns, identification of trends as discovered patterns, alerts generation and, finally, the recording of pedagogical actions taken and further results obtained.

In order to represent in a flexible manner the multifaceted context of students, considering that each one of these data sources (or several possible sets of combinations among them) can be useful to prediction an to other applications, the system was implemented considering a concept of multiple trails and a concept of systemic context, described in the next section.

3.1 Multiple rails and systemic context

Trails are a set of specific data representing sequences of events related to some entity [Driver, 08]. Each entity represented in a trail can be a person or device. The events represented can be any information of interest about the entity. An example of trail can be the sequence of each grade received by a student along a semester. Another example can be sequences of accesses to learning materials and time spent in each one by a student. In general, applications that rely on trails make use of their data in queries driven by a temporal dimension [Gams, 02; Levene, 02]. Also the use of trails can be very selective and explicit, regarding the data source of the trail and data format. Some applications use trails concepts in order to represent a context, that in general is associated with temporal restrictions and related to some specific type of information [Dey, 01; Driver, 08; Silva, 10].

The concept of educational trail [Schoonenboom, 12] can be observed through diverse approaches and applications, such as the support to represent and analyze parts of curriculum, personalization of instructional materials, visualization of interaction data and representation of cognitive and collaborative aspects. In the approach adopted in this work, there is the necessity to represent a considerable number of trails. Each trail is associated with some educational aspects that need to be modelled and stored. The applications implemented in the system can use one or more trails data sets, in order to fulfil their requirements.

The composition obtained by grouping trails contexts is called systemic context, in this work. The systemic context represents a set of trails that contains data of interest to specific application, given a temporal dimension. The main objective of this proposal is to accommodate multiple needs that can arise in dealing with a complex context of several data sources. This is one of the necessities stated by requisites defined to this system, such as the flexibility in data treatment and facilities to system evolution.

The schema shown in Figure 2 illustrates an example of a systemic context, where one can see several trails represented independently and sorted by the temporal dimension. The first trail represents Formal Education, the second represents Complementary Education and the last one represents Professional activities of a student. Therefore, a systemic context can be considered as the set of data from these three trails, selected between time restrictions indicated in the Figure 2 (from January of 2009 to June of 2009). Also in the figure 2 are exemplified some possible uses of trails to represent the various aspects related to a student. The example represents the 'Formal Education' trail, which is composed of records describing compulsory courses and higher education or postgraduate courses. The second trail, described by 'Complementary Education', consists mainly of extension courses. Finally, the 'Professional' trail is used to highlight different professional activities.

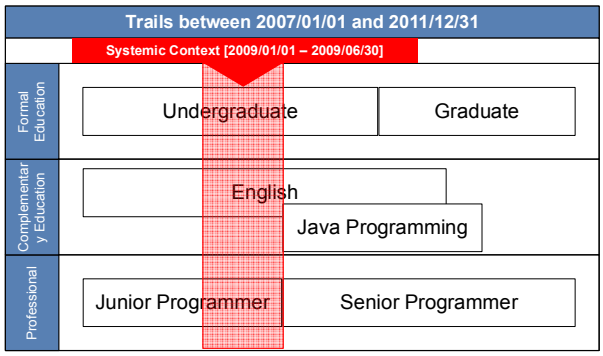
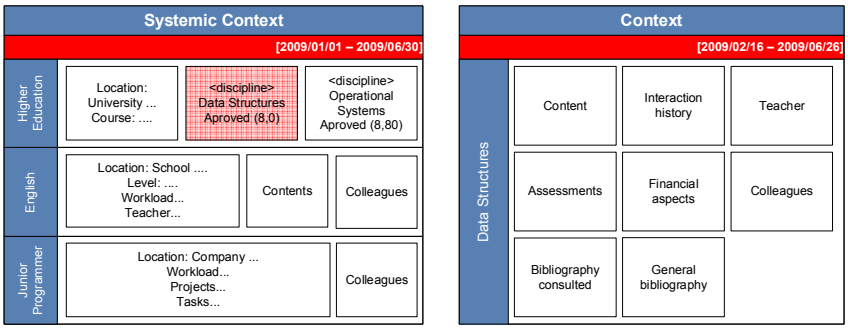


Figure 2: Example of systemic context.

The integration of new trails and the partial selection of available trails data can be of great importance to several applications. The flexibility observed in trails utilization is also extended to aspects relating to trails data acquisition, which can be made based on different devices. New sources and new devices can be included and linked to existing trails without need for adjustments in applications that are already in use. The level of detail of each trail can also be treated in this model. This is important when considering that different applications that use data from the same trail may have diverse needs related to the granularity of manipulated data. This may be necessary in cases of applications that need to deal with details of a systemic context described situation. Different applications may need to use more detailed or less detailed information.



(a) *(b)*

Figure 3: Example of different Systemic Context granularity.

Figure 3 shows an example where the trail 'Higher Education' is being accessed to perform a query in a record, which describes information about a given discipline. The item “a“ in Figure 3 indicates, in the hachured part, that ‘Data Structure’ discipline is being accessed, and the information displayed is the final grade (hich have the vaue of “8,0“). At same time, data necessities can be different to another

application, as can be seen in item ‘b‘ of Figure 3, where a much more detailed set of information is shown from the record ‘Data Structure’. In this second example, it is exemplified a situation in which the application requires access to all details of this discipline, as we can see in the figure, expressed with elements such as content, interaction history, techer, assessments, financial aspects, colleagues and bibliography.

In this way, we can illustrate the flexibility generated with different views of data stored in accordance with the needs of each application. In this sense, the Figure 3 presents the composition of information that can be obtained from various sources and using different software components. Besides that, the example shows the possibility of attending different applications demands. A broad collection of data is relevant and contributes to the composition of a more detailed and complete context for a given student. The flexibility of composition and data collection maintenance is an advantage for the implementation of applications that will use these data in different ways. This flexibility is associated with mechanisms that facilitate integration and interoperability of each trail.

3.2 Representation possibilities with ontologies

The use of ontologies provides the formal representation of human knowledge, so that it can be processed by computational entities. Some authors, such as Gruber [Gruber, 95], describe the ontology as a "conceptualization characterized by formal (explicit) properties and specific purposes". These aspects foster adoption of ontologies in this work, to support interoperability and integration. The developed system uses ontologies as a way to generate flexibility to help in operations with the information recorded in the trails. Reuse possibilities is also present with ontologies. This can be done with some descriptions in existing ontologies, such as people, places, and contexts. Examples of these possibilities are ontologies such as FOAF (Friend of a Friend) [Brickley, 13], Event [Abdallah, 13] or UbisWorld [Heckmann, 13].

There are two main necessities that are fulfilled by ontologies in this work. The first one is related to documentation of data already stored in the multiple trails. The second is related to structure and details of each trail. Data sets representing the multiple trails are physically stored in a database system, for performance reasons, considering that a huge amount of data is manipulated by the Learning Analytics system. Ontologies, therefore, provides an additional layer of metadata regarding multiple trails that is important in selection and localization of existing data and in evaluation of the needs of new applications.

The developed ontology is related to other existing ones, improving reuse options. One example of reuse is the ontology is the use of FOAF Ontology [Brickley, 13] in order to expand the description related to “Entity” concept. Another example is the use of UbisWorld ontology [Heckmann, 13] in order to expand descriptions related to elements “Context” and “Trail point”. One last example is Event ontology [Abdallah, 13], used to specialize de “Event” concept.

The ontology was described based on the methodology proposed by [Huebner, 13] and the objectives were to represent components of systemic context adopted in our approach and to provide mechanisms promoting flexibility in data integration. To accomplish the first objective the concepts in multitrail model were represented in the ontology, considering other ontologies for reuse options. Resources also used are the

query languages [Prud'hommeaux, 13] and the rule languages [Horrocks, 13], that allow the use of relations between terms in the ontologies and physical representation of represented data, stored in databases. The ontology is represented with OWL language (Web Ontology Language), in order to provide greater expressiveness and interoperability [Villazón-Terrazas, 11; OWL, 13]. The main ontology components describes the "trail" elements, and consists of a set of “Entity”, “Context”, “Time”, and “Event” concepts, which are specialized according to context. The trail elements are described in details in the “Trail_point” element and specialization or association with other representations are fostered.

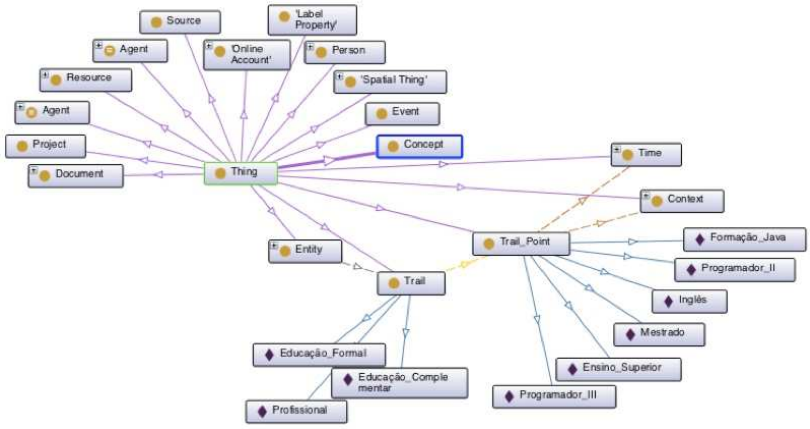


Figure 4: Example of trail and trail point representations

In order to illustrate the use of ontologies in the multitrail architecture, we describe in the figure 4 some of the ontology components that are present in the figure 2, which are the “trail” and “trail point” concepts. The concept denominated “trail” is associated with three instances shown in figure 2, that are Formal Education” (in Portuguese: “Educação Formal“), ”Complementary Education“ (in Portuguese: “Educação Complementar“) and ”Professional“ activities (in Portuguese: “Profissional“). Some advantages in using ontologies in this work are related to facilities to represent and to identify the different kind of data stored in the system. Since the set of data manipulated is diverse and can be originated in different sources, it is important to keep track of this information. Also the search for specific data in the system is facilitated with this approach.

3.3 Learning Analytics System Architecture

The Multitrail architecture, illustrated in Figure 5, is organized in three tiers: Server (MultiTrail Server), client (MultiTrail Client) and devices for data acquisition or other data sources (Client Devices). Initial system operation is associated with data sources. Some of these sources can be a range of different databases and other sources can be represented by user interaction originated in different devices. System architecture is composed by specialized components to collect and record data from databases or from devices access activities. For each device, there is a component responsible for

collecting the corresponding interaction in an architecture layer called MultiTrail Client. These Multitrail Client layer components interact with applications for registration with MultiTrail Server layer, which is responsible for organizing and providing the services required for educational applications. Client Devices Layer implements the necessary support to the activities of data collection and interaction, with flexibility for device types and data sources used. So, it is provided the interaction flexibility for users and the system through the collection of devices or data sources.

Interaction between users and applications based on this model takes place in two distinct situations, and both occur through the use of MultiTrail Client Layer components. The first of these, in the case where a device is used to record occurrences on a trail, is done through MultiTrail Client component called Collector. The second situation occurs when the system presents results to user, and this process is handled through the implementation of the specific component of the MultiTrail Client layer denominated Client Result.

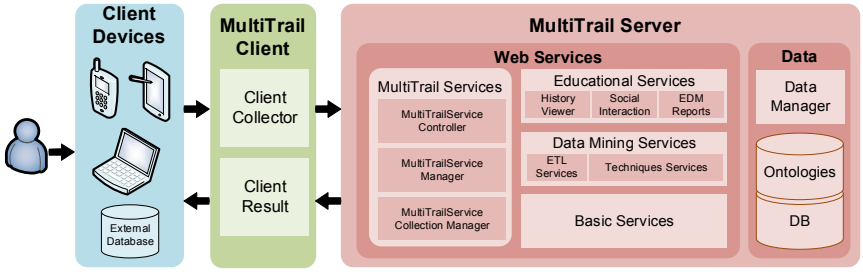


Figure 5: MultiTrail architecture

Therefore, the layer called MultiTrail Client is organized in two modules. Client Collector module is responsible for data capture and recording on the trails and in proper format. This module can be implemented in applications located on devices or software components responsible for interaction with external databases. Client Result module is responsible for interaction with user and adaptation of content to the user's device. The layer named MultiTrail Server is organized into four service modules and a repository of data. Its main goal is to provide the support required for the storage of data obtained from various sources and of multiple devices, together with the necessary integration to meet diverse services offered as support for applications. The Data Manager module is responsible for providing an infrastructure for manipulation and access to data (ontologies and databases) for other modules that compose the MultiTrail Server. Basic Services module is responsible for authentication services, application management services and the management of available services.

MultiTrail Services module is responsible for management and handling of different trails. Whenever a client application needs to manipulate data from a trail, this service will be responsible for identification and maintenance of the data. The same occurs with queries on trails, whether specific queries or queries using systemic context. MultiTrailService Controller is responsible for authentication of entities and

identification of trails. MultiTrailService Manager is responsible for the management of entities, trails and systemic contexts. Finally, MultiTrailService Collection Manager module is responsible for the recovery of trails, contexts and systemic contexts. Educational Services module is responsible for the management of educational actions. The History Viewer component is responsible for the visualization of several aspects of multiple trails. The Social Interaction component is responsible by viewing the interaction of students in diverse interaction tools, such as discussion forums or chats. The EDM Reports component is responsible by the application of techniques available in the Data Mining Services module and the display of the obtained results.

The MultiTrail architecture implementation was accomplished with ASP.NET C# language for components dedicated to collection, data manipulation, and data visualization module of mining. For the persistence of data it was used SQL Server 2008 database. The ontology used was described in OWL language and edited initially with the tool Protege [PROTéGé, 12] and it is manipulated with the API DotNetRdf [API, 13].

3.4 Implemented applications

The MultiTrail architeture facilitates the use of diverse data sets by different applications. In order to illustrate this feature of the system, five initial applications are described to illustrate the flexibility obtained with the trails and the systemic context approach.

3.4.1 Systemic Contexts Visualization

This application was originated as a case study to analyse data obtained in high school. Thus, data were collected from trails, in order to support pedagogical orientation. Teachers had access to a Web interface for monitoring student's situation within the perspective of systemic contexts.

Although the main objective of this paper is related to university students, this application is focused in data obtained from High School. The motivation for this experiment is twofold: to validate the system flexibility regarding data diversity treatment; to use this results in future actions regarding integration of the student historic, including High School data. This objective is consistent with studies about the dropout phenomenon, which indicates a complex net of possible causes, some of them going back to High School performance and events. Therefore, it is important to have access to a broad set of information regarding the student. Regarding this aspect, we consider that the high school historic is important as a way to provide information about student profile, in the earliest phases of future courses. This can allow selection of suitable actions in order to fulfil necessities or to approach expectations of students. For instance, in engineering courses with calculus disciplines can be interesting to analyse the high school mathematics performance of students. The findings of these analyses can help to identify patterns indicating needs such as previous counselling or reinforcement studies.

This application used the multiple trails concept to organize and relate historical information on three trails: "School History", "Complementary Activities" and "Educational Occurrences". Historical data from these three contexts were collected

and organized into trails. Each collector component was responsible for collecting and recording the history of a specific domain. Collector components stored information obtained in the tier MultiTrail server of Multitrail architecture. The application is related to daily activities of an educational institution in which frequent intervention is required from teachers and pedagogic coordinators in monitoring and counselling students. Often lack of information and, in particular, lack of relation among information, can lead to a lower quality of intervention. A better interaction would be possible if all systemic contexts of students were available. To meet this demand, the history recorded in trails was displayed in a timeline, allowing analysis of student achievement in a specific systemic context, according to the needs of each situation.

Figure 6 presents partial aspects of a historical student situation, where it is possible to check the age at which student attended each year of elementary and secondary education, the performance of student in each discipline, complementary activities that were carried out and other educational events. Through this Web interface is possible visual and interactive analysis of important factors, such as influence or relation of complementary activities in school performance and behavior.

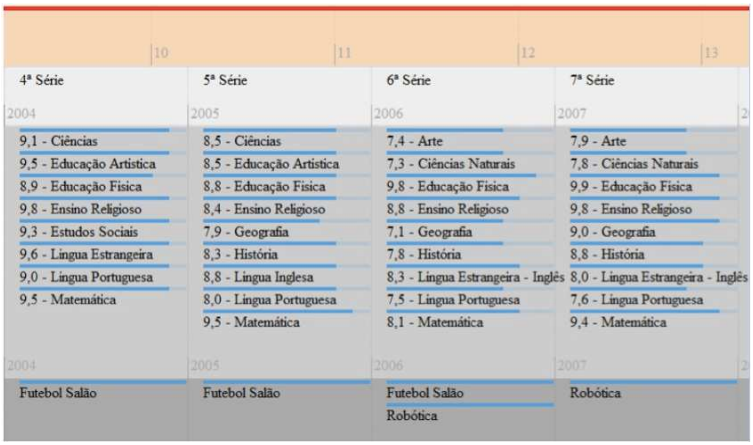


Figure 6: Visualization of Multiple Trails aiming pedagogical support

This application was developed and used in a study case involving trails of 5.921 students in a Brazilian high school course. Although an empirical evaluation of this preliminary application was done through interviews with education professionals, it was not objective of this work to evaluate obtained results. Therefore, although empiric and preliminary results were positive, indicating a good prospect about the support that this application can provide for pedagogical advisory and monitoring activities, the main objective of this application implementation was to verify the possibility of integration of diverse data trails in some specific application.

3.4.2 Interaction Visualization

The case study described here involves to collect, to process and to treat aspects of interaction together with evaluation tasks, as well as the message exchange in a

virtual learning environment. The interactions between students are important aspects to identify potential situations demanding attention. In order to deal with this situation, we developed a visualization component that exhibits, in the form of a graph, aspects related with interaction and performance, from which it is possible to analyse relationship between students' performance and participation in discussion forums.

Figure 7 represents interaction of students in a discipline of Mathematics for Administration, in a graduate course of administration, held in distance education mode. Nodes in yellow represent discipline teachers, green nodes represent students passed and nodes in brown represent students that fail. Graph edges represent interaction between the various participants in discussion tool and their thickness is proportional to interaction intensity.

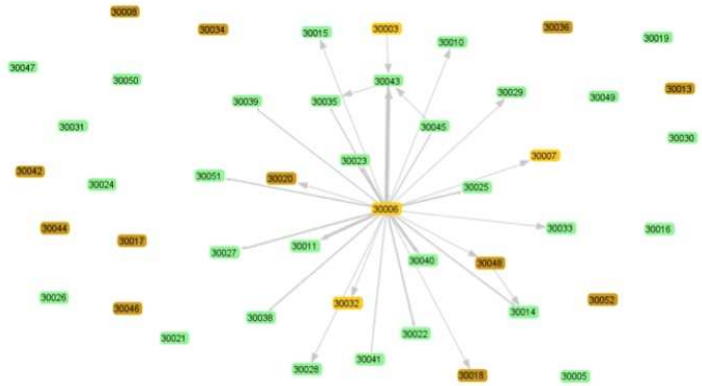


Figure 7: Visualization of interactions between students in Forum

The case study used to evaluate the application was done with data generated by 377 undergraduates during a semester of a discipline. Through visualization tools like this graph is possible to provide support for analysis relating to students' interaction in virtual learning environment and their final result in disciplines. Experiments carried out in various contexts bring evidence that active participation in forums of this discipline is related positively with performance. Available data from these classes indicate that while 15% of students who interacted with greater intensity have failed, disapproval rate increased to 45% among students that not interacted significantly. Thus, with the implementation of visualization tool is possible to track relationships between activities and performance. This application allows identification of relevant aspects of a group of students and to support pedagogical follow-up activities.

3.4.3 Textual interaction analysis and visualization

In this application the main objective is to integrate information in textual form, obtained from discussion forums and to use this information in order to identify possible students' needs. These needs are identified with analyse of text messages generated by students in discussion forums. This procedure used keywords to identify messages regarding questions or doubts. When this situation was identified in these

messages, then domain ontology was used to associate the student message to the correspondent topics being studied.

Textual information that are generated by users in virtual learning environments are of great importance for the analysis of teaching and learning context, because it represents one of the main communication resources used by students and teachers. Student messages in natural language format represent a significant information associated with student learning process in a particular content discipline. In this application the event of doubt expression is treated, in order to enable experimentation with techniques for automatic detection of messages with this broader sense.

The application is based in textual forum messages, generated by students. These messages are classified as messages relating doubts. This is done with use of a lexicon generated by linguistic experts. After that identification, domain ontology is used to relate the subjects of doubts and relate then to subjects treated in class activities. For the lexicon keyword choices, textual postings were analysed from students of Distance Education modality disciplines. The initial treatment of textual information written by users, was conducted by specialists in a manual and individual analysis of all students posts who have a discipline focused on basic concepts in systems programming, having restricted the search to the forums messages. A number of 695 posts were analysed to identify indicators of categories of text messages expressing doubts. To identify what is the subject of doubt expressed in messages it was used an ontology that represents concepts related to knowledge field presented during the course. One important aspect to notice is that this application is described in a very broad form and can be applied in any knowledge domain. The used data was not associated with the individuals that originated it, but also a dashboard with general indication concerning the entire group of students was generated.

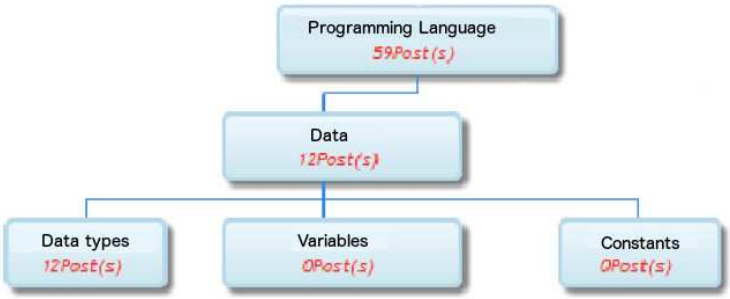


Figure 8: Visualization of an ongoing scenario regarding students´ questions

Domain ontology is used to support different operations. One is the textual information analysis, where ontology plays an important role helping to identify topics related to the class of the subject that are mentioned by students. The second is the support in visualization of the number of mentions performed to each topic, in order to show the teacher a scenario of the ongoing activities and students´ needs. Figure 8 presents part of concepts used in the ontology and for each of these concepts is shown the respective number of messages mentioning it in messages expressing

doubts or questions. Obtained results show that it is feasible and that is productive to integrate semantic and linguistic resources in order to support some actions related to teachers, such as the identification of doubts or difficulties expressed by students.

3.4.4 Dropout prediction

For dropout prediction, Artificial Neural Networks (ANNs) was chosen because it is a classic technique in mining area and presents a satisfactory performance in other similar Data Mining problems [Bishop, 07; Haykin, 01]. Artificial Neural Networks are mathematical models inspired by biological neural networks and consist of a group of artificial neurons divided into layers, interconnected through mathematical functions [Bishop, 07]. Artificial Neural Networks are used to model complex relationships between inputs and outputs or to find data standards.

Several works of educational data mining use classic model of ANNs called multi-layer perceptron (MLP). A typical structure of an RNA MLP feed-forward can be mounted with three layers [Haykin, 01]. The input layer will serve as a sample. The second layer, called hidden layer and the third, called output layer, will be responsible for conducting the mathematical calculations for both training process as for the task of classification. The output of the network consists in different levels of activations of output neurons that should be interpreted to consolidate classification. The MLP ANNs are supervised machine learning techniques and training can be carried out through Backpropagation algorithm, defined initially by [Werbos, 74], that is able to recognize complex patterns of data and use non-linear mapping functions [Freeman, 91]. The training process is to propagate the samples, calculate the error and retro-propagate by adjusting the weights.

In order to implement a dropout prediction application it was adopted a supervised machine learning technique [Alpaydin, 10], where historical data are used for the training process. Therefore, for training process, it was used data belonging to students in previous classes, split in two parts, one of them used for training and other used for test and to evaluation of obtained model. In this implementation, the main aspects considered as parameters for the ANNs are related to student's interaction in Virtual Learning Environment. The section 4.1 describes with more details the use of this application.

3.4.5 Pedagogical actions register

As mentioned before, it is very important to educational institutions that dropout prediction initiatives are implemented together with complementary actions. Here we describe an application to help in this situation. In the scope of dropout control project, mentioned in this work, we are carrying out a monitoring process regarding students' activities. The goal is to relate partial and final results obtained in disciplines with general format of interaction and with activities carried out by students in Virtual Learning environment. This monitoring process uses a series of information that was collected in several classes in previous semesters. The result enables the creation of behavior and interaction patterns that can indicate future dropout or low performance.

In this way, we compared these historical patterns with data being generated each week by current students. The teacher responsible for the class received this

information regarding if a student is identified within these standards. From this situation, the teacher carried out educational actions. These actions and the consequent interaction with this student are meant to help reverse the predicted behavior. We identified a number of actions that can be performed when student is associated with dropout profile. The main issues related to these situations are summarized in three major groups, described below, with some of possible causes.

The first group of causes is related to disincentive with the course. Some of the possible situations to be treated are the following: student does not fell to be identified with the course; student feels being without market perspective; student has difficulty with distance Education modality. The second group of causes is related to students that are discouraging with the academic activity. This can happen when student fail to make sense of the importance of the training, when student have difficulties in understanding the content, or after have been poorly evaluated. The last group of causes is related to students that are discouraging with teacher. Some of the possible situations observed are doubts not solved, or the lack of enough support to overcome their difficulties

We rely on teacher actions to help in this process, working with students to identify and revert these situations. The main way of contact is the e-mail and diary tools in virtual learning environment. Teacher should verify that these are the most appropriate interventions and also adopt other options important and possible. It is essential to have an understanding of the situation and the choice of best action. This information must be indicated on the same web form in which the teacher will receive information about the student identified with a particular pattern of behaviour. The annotation of actions carried out is important to identify the effective ones, in order to assist in future situations. Whenever possible, should be identified what is the problem or difficulty perceived by the student.

Suggestions for pedagogic actions to be adopted with students can also be associated with problems identified. If the teacher identifies that the student is discouraged with the course, for instance, he can indicate the career management program contact, to indicate other courses in the University or to indicate counselling support. Several other options can be of importance in this situation and in other possible situations, such as in cases in which the student is discouraged with the academic activity or with the teacher. The record of the situations and the pedagogic actions during some time can allow identification of possible results and therefore to guide future interventions.

Figure 9 illustrate an example of access to a form of the application developed, used in order to collect and register information regarding teacher actions. In this example, the web form is filled with information in both the standard field and the free text field. It can be observed that the last field, called “Complementary information” can be of great importance to register particular aspects not predicted in the standard fields.

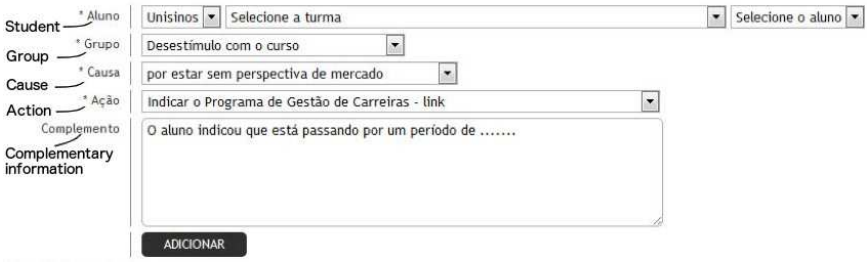


Figure 9: Example of complementary information in the form

This application allows teachers to register how and when they interact with students. Therefore, important information is obtained. With these sets of registered interactions, also teacher intervention is recorded and can be used in models to associate these interventions with future results.

4 Experiments and Results

The developed system was evaluated in a case study where the main objective was to identify students with potential risk of dropout in distance education courses, but also to implement pedagogical actions in order to mitigate these tendencies.

4.1 Dropout Prediction in distance education

For this case study, data collection was carried out with two disciplines, Mathematics for Administration and Mathematics for Computing, taught in Distance Education courses of the University UNISINOS (Universidade do Vale do Rio dos Sinos), located in south of Brazil. This case study was carried out for a period of three semesters, in the years of 2012 and 2013.

Dropout prediction was performed weekly, so at the end of each week was generated a report indicating which are the students who have been classified as a high risk of dropout. This process was carried out by the application of dropout prediction and the data were obtained from the Virtual Learning Environment and then represented in the multitrail model. This enabled the participation of teachers through pedagogical action during the discipline development, increasing the possibilities of reversion for these predicted behaviors. In this case study the objective was to identify early and individually students who have low performance or possible dropout behavior, thus enabling prevention in a personalized way.

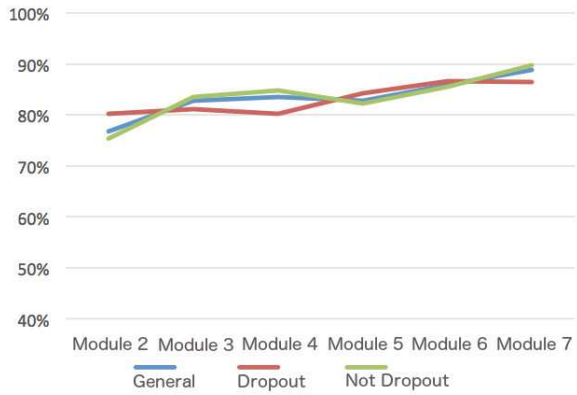


Figure 10: Dropout prediction results

The Figure 10 presents results obtained. The number of students accompanied was 2.491, belonging to eighteen classes and two disciplines. Obtained mean rate for dropout prediction was 83,6% and prediction for non-dropout behavior was confirmed in 87.3%. These numbers can be improved when the historical data sets are bigger, as it was seen in another experiment with a class in Logic, when obtained precision was around 87% of correct predictions for the dropout. Figure 10 also describes different results for predictions in every week, starting from third week of classes. This third week was selected to be the first to generate predictions in order to make sure that the necessary amount of interaction data was available. Results shows that in the third week it was possible, by using the generated predictions, to contact a relevant number of students that were associated with dropout behavior and whose predicted dropout behavior was later confirmed in the end of semester.

4.2 Pedagogical actions

In order to complement the case study with a follow-up of potential use of reports generated weekly, it was also carried out an evaluation of possibilities for reversion in dropout predicted behavior. This work was performed with a component of the solution aimed to register pedagogical actions (described in item 3.5.6) that compliments possibilities of dropout and low performance prediction tool. Thus the tools provide help to foster interaction between educators and the students.

The main goal of this module is to monitor educational actions to be taken with the students classified with a high risk of dropout. Therefore, predictions of dropout were used as support in activities of teachers involved with classes that participated in the case study. This involved a prior planning, training of participants and subsequent monitoring of pedagogical actions. In this way, every week the teachers involved with the case study had access to the report with the prediction of dropout and from this information, they started specific actions and attention to students.

These pedagogic actions have proactive character and can vary according to the needs identified and the development time of contents. Studies already carried out have confirmed that proactive actions such as those described may avoid the dropout

and improve student performance. Figure 11 provides a comparative analysis of evasion rates among groups who participated in studies and in which proactive actions have been taken and classes that did not participate in the study. Notice in Figure 11 that dropout of the classes that participated in this study is significantly lower than rates of avoidance of classes that did not participate in the study. Despite the dropout phenomenon be associated with multiple factors, as described earlier, it is considered that these results indicate a positive progress towards a significant result with the proposed process.

A set of pedagogical actions designed to help in dropout mitigation can also be considered in the context of the course design, as suggested by [Lockyer, 2013]. This aspect can play an important role as one of the driving forces related to dropout, since the course design may or may not be adequately aligned with the student cognitive profile or expectations. During the case study, some of the aspects evaluated as necessities of change to help improvements by the teachers involved were the planning of some parts of the course. The concept of transactional distance [Moore, 93] can be applied in this case, as a way to identify necessities of intervention. In the case study some specific methodological adjustments were designed in order to avoid problems associated with difficulties in the earliest phases of the courses, where the course design aspects and the individual difficulties can play an important role in dropout.

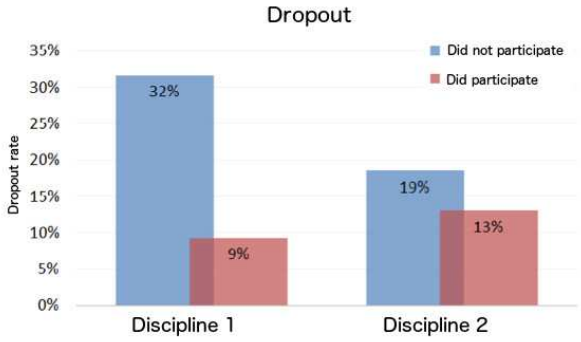


Figure 11: Results in dropout reversion

The pedagogical actions are influenced by the weekly reports, due to the fact that it provides insight to the teacher about which are the students that need some specific attention. Also is important to note that the additional data resources can improve the accuracy of the prediction, by allowing a more complete context to be described regarding the students. In this regard, both historic additional information of the discipline and the particular personal additional information can benefit the prediction, which is based in this kind of combination of historical and related data. Based on data analyses the teachers can identify specific patterns that are related to previously known tendencies or necessities of the students. The teacher experience in helping the student to overcome the associated difficulties is the main driving force towards the identification of adequate measures to foster the academic success of the students, based on the generated patterns. Therefore, these measures are defined and

implemented based on teachers experience, but are also recorded and monitored in the Learning Analytics system, to provide future insight about the effectiveness and adequacy.

5 Conclusions

This work aimed to describe possibilities to use data from different sources in order to help dealing with the dropout problem. Actions involving interdisciplinary studies and professionals involved in daily routine associated with Distance Education courses provided some important insight in the very complex nature of dropout origins. The course of action involved the use of Educational Data Mining and Learning Analytics. Beyond that, it was also decided to cope with dropout complex causes by implementing a system that allows to integrate different data sources and also to combine these sources in different applications. Therefore, the different necessities of Education professionals can be adequately addressed.

The developed Learning Analytics system presents evidences regarding relevant possibilities, as shown by the five different applications developed and tested. The first aspect to consider when evaluating the system should be related to its ability to integrate several data sources and to make them available to the applications. Due to the present variety of environments, information systems and external applications can be of interest to identify eventual dropout aspects, it is vital to the dropout prediction that the system can manipulate these data sources, as well as to accommodate new ones.

To evaluate the application of dropout prediction, this work used the developed system during three semesters in a real Distance Education set of courses, in UNISINOS University. Results showed an important and satisfactory rate of correct predictions, using in most part data generated by students interaction in Virtual Learning Environment. The flexibility of the developed system was also applied in order to promote implementation of a prediction process that can be used in a weekly base, therefore allowing the use of these predictions in pedagogical actions aiming dropout prevention.

Therefore, a new experiment was performed, in a period of one semester, to evaluate results of the integrated use of predictions and pedagogical actions, performed by teachers in the courses classes. Obtained results indicate that the amount of dropout reduction in classes involved in the case study where meaningful. Also, as important additional aspects, it can be stated the importance of wide integration with involved professionals, such as teachers and course coordinators, besides other university sectors.

The obtained results show that the area of Educational Data mining and Learning Analytics can effectively support processes aimed at detection of dropout behavior. However, a broad mapping of associated dropout factors is necessary, involving the various sectors of educational institutions, since theoretical models about dropout point to multiple causes. This kind of treatment of dropout factors is not yet perceived in most known initiatives. This analysis points to possibilities for solutions integrating historical data supplemented with more dynamic data sets obtained in the interaction of students and teachers throughout periods of school semesters. Besides the attention to multiple causes and to the use of available data sources, it is relevant to use

predictions in pedagogical actions aimed at reversion of this behavior. It is also important to highlight the need for monitoring pedagogical actions taken, to promote the use of their results in future actions, so early diagnosis and more relevant pedagogical actions can be taken.

Given the diversity of data involved, there are possibilities for the use of combined sets of Educational Data mining techniques. This exploration of algorithms, techniques and mechanisms in a broad, flexible and integrated way, along with the educational players involved, is one of the crucial points to be achieved to contribute to mitigation of dropout contexts.

References

[Abdallah, 13] Abdallah, S., Raimond, Y. (2013). Event Event Ontology – Linked Open Vocabularies. http://lov.okfn.org/dataset/lov/details/vocabulary_event.html. Acessed in 01/09/2013.

[ABED, 13] ABED. (2013). http://www.abed.org.br/site/pt/midiateca/censo_ead/. Acessed in 01/09/2013.

[Adachi, 09] Adachi, A. A. C. T., (2009). Evasão e evadidos nos cursos de graduação da Universidade Federal de Minas Gerais, Dissertação de Mestrado. Faculdade de Educação, UFMG, 2009.

[Alpaydin, 10] Alpaydin E. (2010). Introduction to Machine Learning. 2nd. ed. [S.l.]: The MIT Press.

[API, 13] API DotNetRdf. (2013). http://www.dotnetrdf.org/. Acessed in 01/09/2013.

[Araque, 09] Araque, F., Concepción Roldán, C., Salguero, A. (2009). Factors influencing university drop out rate. Computers & Education (53)3. 563-574.

[Baker, 11] Baker, R., Isotani, S., Carvalho, A. (2011). Mineração de Dados Educacionais: oportunidades para o brasil. Revista Brasileira de Informática na Educação, 19(02). 15-35.

[Barbosa, 12] Barbosa, J. L. V., Barbosa, D. N. F. ; Wagner, A. (2012). Learning in Ubiquitous Computing Environments. International journal of information and communication technology education, 8(1). 64-77.

[Barker, 04] Barker, K., Trafalis, T., Rhoads, T. R. (2004). Learning from student Data. System and Information Engeenering Desin Symposioun. 79-86.

[Bishop, 07] Bishop, C. M. (2007). Pattern Recognition and Machine Learning. 2. ed. ed. [S.l.]: Springer.

[Brickley, 13] Brickley, D., Miller, L. (2013) Friend of a Friend. http://www.foaf-project.org. Acessed in 01/09/2013.

[Cambruzzi, 12] Cambruzzi, W. L., Moraes, R. de, Leithardt, V. R. Q., Mendes, C., Geyer, C. F. R., Costa, C. A. da; Barbosa, J. L. V., Rigo, S. J. (2012). Um Modelo para Gerenciamento de Múltiplas Trilhas Aplicado a Sistemas de Apoio à Educação. In: SIMPóSIO BRASILEIRO DE INFORMáTICA NA EDUCAçãO, 2012, Rio de Janeiro.

[Carr, 12] Carr, S. (2012) As Distance Education Comes of Age, the Challenge Is Keeping the Students -technology -The Chronicle of Higher Education. http: //chronicle.com/article/As-Distance-Education-Comes-of/14334. Acessed in 01/09/2013.

[Ceglar, 06] Ceglar, A. Roddick, J. (2006). Association Mining. ACM Computing Surveys, 38(2). 1-42.

[Censo, 11] Censo EAD.BR. (2011). Relatório analítico da aprendizagem a distância no Brasil, Associação Brasileira de Educação a Distância (ABED), Editora Pearson Education do Brasil, São Paulo.

[Dey, 01] Dey, A. K. (2001). Understanding and Using Context. Personal and Ubiquitous Computing. Personal and Ubiquitous Computing. 5(1), 4–17.

[Driver, 08] Driver, C., Clarke, S. (2008).An application framework for mobile, context-aware trails. Pervasive Mob. Comput., Amsterdam, The Netherlands, The Netherlands, 4(5). 719–736.

[Durand, 11] Durand, G., Laplante, F., Kop, R. (2011). A Learning Design Recommendation System Based on Markov Decision Processes, KDD 2011 Workshop: Knowledge Discovery in Educational Data, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2011) in San Diego, CA. August 21-24. 2011.

[Favero, 06] Favero, R. V. M. (2006). Dialogar ou evadir: Eis a questão!: Um estudo sobre a permanência e a evasão na Educação a Distância, no Estado do Rio Grande do Sul. CINTED-UFRGS. Novas Tecnologias na Educação. 4(2). 15-35.

[Fayyad, 96] Fayyad, U. M. Shapiro, G. P. Smyth, P. Uthurusamy. R. (1996). Advances in knowledge discovery and data mining. Menlo Park: MIT.

[Freeman, 91] Freeman, J. A., Skapura, D. M. (1991). Neural networks: algorithms, applications, and programming techniques. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc.

[Gams, 02] Gams, E.; Reich, S. (2002). The TrailTRECer Framework: applying open hypermedia concepts to trails. Journal of Universal Computer Science, 8(10), 913–923.

[Gruber, 95] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. Int. J. Human-Computer Studies. 43(5-6). 907–928.

[Haykin, 01] Haykin, S. (2001). Redes neurais: princípios e prática. 2. ed.. ed. Porto Alegre: Bookman.

[Heckmann, 13] Heckmann, D. (2013). UbisWorld. http://www.ubisworld.org. Acessed in 01/09/2013.

[Horrocks, 13] Horrocks, I. (2013). SWRL. http://www.w3.org/Submission/SWRL/. Acessed in 01/09/2013.

[Huebner, 13] Huebner, R. A. (2013). A survey of Educational Data Mining research. Research in Higher education Journal. 10-23.

[Kampff, 09] Kampff, A., Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente, Tese de Doutorado, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre –RS, 2009.

[Kay, 06] Kay, J. Maisonneuve, N. Yacef, K. Zaiane, O.O. (2006). Mining Patterns of Events in Student´s Teamwork Data. In: Proceeding of Educational Data Mining Qorkshop. Taiwan.

[Keegan, 96] Keegan, D. (1996). Foundations of distance education. 3a. ed. London: Routledge.

[Kotsiantis, 11] Kotsiantis, S. B. (2011). Use of machine learning techniques for educational purpouses: a decision suppoort system for forecasting student's grades. Artificial intelligence review. Springer.

[Levene, 02] Levene, M.; Peterson, D. (2002). Trail Records and Ampliative Learning. In: School of Computer Science and Information Systems, Birkbeck College, University of London, Research Report BBKCS-02-10.

[Levy, 07] Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses, Computers & Education, 48(1), 185–204.

[Li, 11] Li, N., Cohen, W., Koedinger, K. R., Matsuda, N. (2011). A Machine Learning Approach for Automatic Student Model Discovery. EDM 2011: 31-40. Proceedings of the 4th International Conf on EducationalData Mining, Eindhoven, The Netherlands, July 6-8, 2011.

[Liao, 12] Liao, S. H., Chu, P. H., & Hsiao, P.-Y. (2012). Data mining techniques and applications – a decade review from 2000 to 2011. Expert Systems with Applications, 39(1), 11303 – 11311.

[Lockyer, 2013] Lockyer, L. Heathcote, E., Dawson, S. Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. American Behavioral Scientist, October 2013; vol. 57, 10: pp. 1380-1400. 2013.

[Longo, 09] Longo, C. R. J. (2009). Educação a Distância: o estado da arte. São Paulo: Pearson Education do Brasil.

[Lykourentzou, 09] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. Computers & Education (53)3. 950-965.

[Maia, 04] Maia, M. D. C., Meirelles, F. D. S., Pela, S. (2004). Análise dos Índices de Evasão nos Cursos Superiores a Distância do Brasil. ESUD 2013 – X Congresso Brasileiro de Ensino Superior a Distância. Belém, PA.

[Manhães, 11] Manhães, L. M. B., Cruz, S. M. S. da, Costa, R. J. M., Zavaleta, J., Zimbrão, G. (2011). Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados. In: XXII SBIE -XVII WIE, 2011, Aracaju. Anais. . . [S.l.: s.n.], 2011. p. 150–159.

[Moore, 07] Moore, M., Kearsley, G. (2007). Educação a Distância: uma visão integrada. São Paulo: Thomson Learning, 2007.

[Moore, 93] Moore, M. (1993). Theory of transactional distance. In Theoretical principles of distance education. Oxon: Routledge.

[Nandeshwar, 11] Nandeshwar, A., Menzies, T., Nelson, A. (2011). Learning Patterns of university student retention. Expert system with applications. 38 (2011), 14984 14996.

[Nistor, 10] Nistor, N., Neubauer, K., (2010).From participation to dropout: Quantitative participation patterns in online university courses. Computers & Education (55)2. 663-672.

[OECD, 13] OECD (2013). http://www.oecd.org/. Acessed in 01/09/2013.

[OWL, 13] OWL (2013). Ontology Web Language. W3C. http://www.w3c.org/OWL. Acessed in 01/09/2013.

[PROTéGé, 12] PROTéGé. (2012). The Protégé Ontology Editor and Knowledge Acquisition System. http://protege.stanford.edu/.. Acessed in 01/09/2013.

[Prud'hommeaux,13] Prud'hommeaux, E. (2013). SPARQL. http://www.w3.org/TR/rdf-sparql-query/. Acessed in 01/09/2013.

[Romero-Zaldivar, 12] Romero-Zaldivar V. A., Pardo A., Burgos, D, Kloos, C. A. (2012). Monitoring student progress using virtual appliances: A case study. Computers & Education, 58(4). 1058-1067.

[Romero, 08] Romero, C., Ventura, S., Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education (51)1. 368-384.

[Romero, 10] Romero, C., Ventura, S. (2010). Educational data mining: A re-view of the state of the art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 40, 601 – 618.

[Romero, 12] Romero, C., Ventura, S., Pechenizkiy, M., Baker, R. S. J. (2012). Handbook of Educational Data Mining, Ed. CRC.

[Schoonenboom, 12] Schoonenboom, J., Levene, M., Heller, J., Keenoy, K., Turcsanyi-Szabo, M. (2012). Trails in Education - Technologies that support navigational Learning. Sense Publishers, Rotterdan.

[Scott, 11] Scott, J. (2011). Distance Education Report, California Community Colleges Chancellor's Office. http:// californiacommunitycolleges. cccco.edu/ Portals/0/ eportsTB/ DistanceEducation2011_final.pdf. Acessed in 01/09/2013.

[Silva, 10] Silva, J., Rosa, J., Barbosa, J., Barbosa, D., & Palazzo, L. (2010). Content distribution in trail-aware environments. Journal of the Brazilian Computer Society, 16(1), 163 – 176.

[Tinto, 75] Tinto, V. (1975). Dropout from Higher Education: a theoretical synthesis of recent research. Washington, Review of Educational Research, 45(1). 89–125.

[Toscher, 10] Toscher, A., Jahrer, M. (2010). Collaborative filtering applied to educational data mining. KDD Cup 2010: Improving Cognitive Models with Educational Data Mining, 2010.

[Villazón-Terrazas, 11] Villazón-Terrazas, B.; Ramíresz, J.; Suárez-Figueroa, C. M., Gómez-Pérez, A. (2011). A network of ontology networks for building e-employment advanced systems. Expert Sysems with. Applications, 38(11), 13612–13624.

[Wannasiri, 12] Wannasiri B. A., Xaymoungkhoun., O., Hangjung Z. B., (2012). Critical success factors for e-learning in developing countries: A comparative analysis between ICT experts and faculty. . Computers & Education. (58)2. 843-855.

[Werbos, 74] Werbos, P. (1974). Beyond Regression: new tools for prediction and analysis in the behavioral sciences. 1974. Tese (Doutorado em Ciência da Computação) — Harvard University, Cambridge, MA, 1974.