

# Data Analysis Tools and Methods for Improving the Interaction Design in e-Learning

Paul Stefan Popescu  
University of Craiova  
Department of Computer Science  
Craiova, Romania  
+40724571133  
sppopescu@gmail.com

## ABSTRACT

In this digital era, learning from data gathered from different software systems may have a great impact on the quality of the interaction experience. There are two main directions that come to enhance this emerging research domain, Intelligent Data Analysis (IDA) and Human Computer Interaction (HCI). HCI specific research methodologies can be used to present the user what IDA brings after learning and analyzing user's behavior. This research plan aims to investigate how techniques and mechanisms available in both research areas can be used in order to improve learners' experiences and overall effectiveness of the e-Learning environment. The foreseen contributions relate to three levels. First is the design and implementation of new algorithms for IDA. The next level is related to design and implementation of a generic learning analytic engine that can accommodate educational data in attempt to model data (i.e., users, assets, etc.) and provide input for the presentation layer. Last and top level is represented by the presentation layer where the output of the underlying levels adapts the user interface for students and professors.

## Keywords:

Learning analytics, intelligent data analysis, interaction design, user modeling

## 1. INTRODUCTION

Standard books or their digital versions (eBooks) or standard e-Learning environments are usually just a simple presenting method of the learning material. In this digital era our day by day devices must become proactive to our needs, i.e. they have to know what we need before we even have to ask them. Considering the field of e-Learning, in order to find user's needs and to improve his learning experience we can log various activity related data as a first step in a data driven analytic engine. These actions may define learners' behavior in e-Learning environments providing IDA with raw data to be analyzed. Based on this data IDA creates a data model which is based on user's performed actions. A sample output of the IDA process may be represented by a user model that is aimed to directly influence the user interface.

Learning using on-line educational environments is getting more and more popular but the effectiveness of interaction between students or students and professors is usually poorer than the interaction in physical educational environments. Improving the interaction design process in e-Learning platforms may have a direct impact on the effectiveness of the learning and be achieved by following a data driven approach. The proposed approach is

related to several prerequisites and the learning resource that needs to be well structured and presented. Others are related to the interaction between students and the links that can be created between them, proper data visualization techniques, interpretation of results, adequate data analysis processes with specific goals regarding interface adaptation.

## 2. RELATED RESEARCH IN I.D.A.

Learning analytics and Machine Learning[2] is still one of the most interesting parts of the IDA research area. One research area of this domain is related to the classification procedures. Some of them are related to the usage of classification on text[1] and some of them are regarding to usage of classification as an user analyzing method[4].

Analysis of students' activities in the online educational systems with the goal of improving their skills and experience through the learning process has been an important area of research in educational data mining. Most of the techniques are trying to predict student's performances[5,6,7,12] based on their actions.

The work in this domain started in the year of 2005 with a workshop referred to as 'Educational Data Mining' AAAI'05-EDM in Pittsburg, USA[8] which was followed by several related workshops and the establishment of an annual international conference first held in 2008 in Montreal[9]. Before of EDM, user modeling domain was the one that was encapsulating this research area.

Several papers, journals and surveys have been written but only two books were published: the first is "Data mining in E-learning"[10] which has 17 chapters oriented to Web-based educational environments and the second is "Handbook of Educational Data Mining"[11] which has 36 chapters about different types of educational settings.

In this research proposal the goal is to combine HCI with IDA and educational research in order to improve the learners experience in digital educational environments. This domain is also related to Intelligent Interfaces research area.

## 3. RESEARCH AND DEVELOPMENT STATUS

As research status two papers have been written so far. I am a co-author of the paper Advanced Messaging System for On-Line Educational Environments[3]. This paper presents a method of using a classification procedure for retrieving a set of recommended messages that might be interesting to students.

The second paper is entitled „Building an Advanced Dense Classifier”[4], which has already been published at IDAIR 2014 and won the best paper award. This paper presents a classifier that implements several extra functionalities which can lead to better results. Its goal is to build a Decision Tree classifier that accommodates data (instances). This new data structure extends the functionality of a Decision Tree and is called DenseJ48. This new classifier implements efficiently several extra functionalities besides the core ones that may be used when dealing with data.

Based on this paper, as development background a Weka package which implements the classifier’s functionalities is under development. I am also a contributor (<http://apps.software.ucv.ro/Tesys/pages/development.php>) of *Tesys*[13], an e-Learning platform used in several faculties from Craiova, mainly focusing on the eLeTK (e-Learning Enhancer Toolkit)[14] module. This is how I found out about Intelligent Data Analysis and Information Retrieval, and the benefits these research areas can bring to the online educational environments.

As relevant training in September 2013 I applied for and obtained a scholarship for attending the 9th European Summer School in Information Retrieval, which took place in Granada, Spain. Being part of this event helped me improve my knowledge in the domain of Information Retrieval – the presentations covered most of this research area, from basics to evaluation techniques and Natural Language Processing. Later I attended Research Methods in Human-Computer Interaction between 25th and 31th of July 2014 in Tallinn, Estonia (<http://idlab.tlu.ee/rmhci>) in order to deepen my knowledge of HCI research methodologies.

## 4. RESEARCH PROBLEMS FROM PHD PROPOSAL

Problems related to this research can be structured in a three layer representation. There is a certain need for improving the interaction between the users (students, professors, etc.) and the system that provide them the learning experience. The research problems are related to closing the gap between classical and digital learning paradigms.

Development of new tools is fundamentally based on functionality provided by a generic learning analytic engine, among which there are: generic representation of learning analytics data of users, integration of various implementations of IDA algorithms, custom integration of interaction design process artifacts. All these three layers build up a learning analytics engine that is designed to run as a service along e-Learning environments in an attempt to improve the quality of the on-line educational system.

### 4.1 Layers description

#### 4.1.1 Data Representation Layer

First layer is related to the representation of the raw data that can be gathered from the log files and the database. Our desire is to find what data (features, parameters, ranges, etc) is relevant for online learning environments. Based on this data we have to extract features that can define learning resources or those features that enable us to obtain a user representation.

#### 4.1.2 Learning Analytics Layer

Based on the data gathered it is possible to employ different IDA algorithms in order to obtain custom built data pipelines. Experimenting at this level with different algorithms and different feature sets can lead to obtaining output information for solving different problems. Data aggregation and pipelining are the mainly used processes. The purpose of this layer is to offer to the next one data in a structured format which can be presented on the interface.

#### 4.1.3 Presentation Layer

The presentation of the learning material is very important, leaving a mark on the mental model created by the learning resources. In this layer the HCI component of this proposal is employed.

Taking into consideration these aspects related to both domains we can say that there is a need for new tools that could be integrated within the digital learning environments in order to provide an improved learning experience that fulfills the user’s needs.

## 4.2 Research questions & Proposed Approach

The questions that have to be addressed when we talk about research in e-Learning environments are related to the main actors that are using the on-line educational environments. Therefore, learners, teachers and administrators (which can do the data analyst job), by the generic meaning, are the ones we focus on because they are the main users of these systems. Secretaries of the learning environments only concur to configure the e-Learning environment.

The presented questions are from the business goal perspective. Answering these questions needs a close discussion about the presented underlying levels, which are the same regardless of the tackled issue, that define data driven process.

- *How IDA can be efficiently and effectively used for an on-line educational context?*

Proper usage and integration of IDA techniques can create a framework which data analysts and developers can employ for further work.

- *How can e-Learning resources be managed/aggregated in an IDA context?*

There are various types of resources that exist in on-line educational environments. Depending on how they are managed and aggregated, application developers can benefit from them.

- *Which are the common (general purpose) functionalities when dealing with educational data pipelines?*

Several functionalities exist in dealing with data but not all of them are feasible for working with educational data. In this particular case we need to find the most effective ones and adapt them to this particular case.

- *How can the student know his place among his colleagues and be motivated to study harder?*

This question is highly important from the student’s perspective. Without knowing his place among his colleagues and

without having an explicit learning path, the learner will not have the indication of his final result and will not have the motivation to maximize his potential. In e-Learning environments, students do not participate together in courses, like in a regular environment, so they are unaware of their colleagues' knowledge level. In a traditional classroom, there is always a certain level of competitiveness, so each student is constantly motivated to improve himself. Therefore, an important goal is to achieve a similar scenario in the online educational environments, although it is not the only one. Besides being competitive, the students must also be engaged in helping others and in turn receive help when they are having difficulties understanding something.

- *How can the professors know where exactly do the students have problems, so they can adapt the course material?*

From the professor's point of view, being aware of his students' progress and the difficulties they encounter in understanding the material is possibly the most important requirement. Although each student is different and has his own learning curve, common points can be found and an overall perception can be formed. The professor must be able to build a mental model regarding the overall performance of his students. By doing so, he can modify and perfect in time the content of the course. Also, taking into consideration the fact that the difficulty level of the final evaluation must be consistent with the students' level of understanding of the course, the professor needs to be aware of that level so he can make the proper adjustments.

- *Which data should be logged in order to extract relevant information about the students?*

Any e-Learning environment whose goal is to integrate an intelligent component should be able to log the necessary data and extract the values of the features. Logging the needed data is a prerequisite to the data analysis process. Logging too much can create a useless load of the server but logging not enough will make impossible the features extraction.

Features are very important in IDA because they define the entity that will be analyzed. Choosing the right features are crucial in different IDA processes. A comprehensive list of features (with proper data types, range values and significance) should be available for further analysis.

## 5. CLOSING REMARKS

On-line educational environments are here from a long enough time. This aspect brings in front of the scientists many opportunities for improving the learning process and to lower the distance from the classical educational environments to the online ones. Many research areas concur to improve the learning process but the most relevant are the user centered ones.

There are 3 different research areas that concur to bring learners several improvements. IDA is the first one bringing data mining and machine learning algorithms and generate user models, followed by HCI, which is used to optimize the interfaces and create friendly interaction environments and finally the Educational research area is where we put in practice this work.

## 6. REFERENCES

- [1] Vitomir Kovanovic, Srecko Joksimovic, Dragan Gasevic, Marek Hatala :Automated Cognitive Presence Detection in Online Discussion Transcripts, LAK Workshops, 2014.
- [2] Dragan Gasevic, Carolyn Penstein Rosé, George Siemens, Annika Wolff, Zdenek Zdráhal : Learning analytics and machine learning, LAK Conference, 2014..
- [3] Mihai Mocanu, Paul-Stefan Popescu, Dumitru Dan Burdescu, Marian Cristian Mihaescu,: Advanced Messaging System for On-Line Educational Environments Sesimbra, Portugal, 26-28 June 2013.
- [4] Paul Stefan Popescu, Cristian Mihaescu, Mihai Mocanu, Dumitru Dan Burdescu, Building an Advanced Dense Classifier, 5th International Conference on Information, Intelligence, Systems and Applications, Chania, Crete, July 2014.
- [5] Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G.: Predicting student performance: an application of data mining methods with an educational Web-based system. In: Frontiers in Education, 2003. FIE 2003 33rd Annual (Volume:1), pp. T2A--13. Westminster (2003).
- [6] Baradwaj, B.K., Pal, S.: Mining Educational Data to Analyze Students' Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Volume 2, No. 6, 63--69 (2011).
- [7] Márquez-Vera, C., Cano, A., Romero, C., Ventura, S.: Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Applied Intelligence, Volume 38, Issue 3, 315--330 (2013)
- [8] Beck, J, Proceedings of AAAI2005 workshop on Educational Data Mining, (2005).
- [9] Baker, R.S.J.d., Barnes, T., Beck, J.E. (Eds.) Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings. Montreal, Quebec, Canada. June 20-21, 2008.
- [10] C. Romero and S. Ventura, Data Mining in E-Learning, WIT Press, 2006
- [11] C. Romero and S. Ventura, Handbook of Educational Data Mining, CRC Press, 2010
- [12] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer and, W. F. Punch, "Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA" In Proceedings of ASEE/IEEE Frontiers in Education Conference, Boulder, CO: IEEE, 2003..
- [13] Marian Cristian Mihaescu: Software Architectures of Applications Used for Enhancing On-Line Educational Environments, International Journal of Computer Science and Applications (IJCSA), vol. 10, ISSUE 1 MARCH 2013
- [14] Burdescu, D.D., Mihaescu ,M.C.: Tesys: e-Learning Application Built on a Web Platform. In: Proceedings of International Joint Conference on e-Business and Telecommunications, pp. 315-318. Setubal, Portugal (2006)