# Analysis of the linearity relation between the tutor's actions and the learners' interactions in Distance Education courses

Ygor Amaral<sup>1</sup>, Rodrigo L. Rodrigues<sup>2,4</sup>, Fabiana Soares<sup>3</sup>, Angelo F. Dias Gonçalves<sup>3</sup>, Alexandre M. A. Maciel<sup>3</sup>

<sup>1</sup> Faculdade de Ciência e Tecnologia de Caruaru – Universidade de Pernambuco (UPE) Rodovia BR 104, Km 62 – Nova Caruaru – CEP: 55.002-971 – Caruaru/PE

<sup>2</sup> Departamento de Educação – Universidade Federal Rural de Pernambuco (UFRPE) Rua Dom Manoel de Medeiros, s/n, Dois Irmãos – CEP: 52.171-900 – Recife/PE

> <sup>3</sup> Escola Politécnica – Universidade de Pernambuco (UPE) Rua Benfica, 455 – Madalena – CEP: 50.720-001 – Recife/PE

<sup>4</sup> Centro de Informática – Universidade Federal de Pernambuco (UFPE) Avenida Jornalista Aníbal Fernandes - Cidade Universitária, Recife - PE, 50740-560

Abstract. It has been a challenge to extract relevant information that lead to a more effective understanding of the interactions between students and tutors in Distance Education (DE) courses. This work aims at presenting this relation through statistics analyses based on multiple linear regression algorithms and on Educational Data Mining (EDM). It is proposed the use of linear regression technique to estimate the level of interactions of students based in behavioral variables of the tutors. The results showed that it is possible to use the linear regression technique in order to get inferences with good precision rates with respect to the relation between the tutors' actions in the Learning Management System and the increase of students' interactions.

### 1. Introduction

Distance Education (DE) has been reaching an increasing prominence among other education modalities in Brazil. Within this context, a lot has been discussed about the processes of teaching and learning in this modality, which factors favor and which ones become obstacles in the learning processes. Bentes (2009) relates five components which should be present in the distance education systems as a premise that could guarantee good results. They are: the teacher – here responsible for the instructional material-, the tutor, the student, the instructional material, and the evaluation.

Among the actors of DE, the tutor plays a fundamental role, because he or she is responsible for keeping up with the students' progress during their learning processes. According to Preti (1996), *apud* Bentes (2009), the tutor is one of the greatest

responsible for the courses' success and he or she is constantly guiding, conducting and supervising the process of teaching and learning of the students.

One of the most important mechanisms for the evaluation of the actors in DE is the Educational Data Mining (EDM). Among the processes of EDM, the KDD (Knowledge Discovery of Databases) can be used to extract valuable information from huge data collections. According to Romero and Ventura (2007), KDD can be used, among other things, to evaluate and to improve e-learning systems by discovering useful learning information.

In Gottardo *et al.* (2012), EDM was used in order to predict the performance of students. In the same way, Santana *et al.* (2014) also predicted the students' performance, but using different attributes and therefore, proposing a complementary work to Gottardo's *et al.* (2012) it is a fact that lots of works on EDM have been done trying to understand students' behavior and consequently, predict their performance. Little has been looking at the data collected under the perspective of the tutors.

In this degree, the objective of this work is to present a research developed with tutors and students of Distance Education graduation courses and it intends to analyze the linearity relation between the tutor's actions and the learners' interactions in the discipline analyzed. Thus, to do this work, data were collected from the Moodle, the LMS (Learning Management System) most used among the universities and colleges which offer DE courses in Brazil (ABED- Censo 2013). These data were mined considering some attributes which were selected in order to answer this question: A course that has most active tutors makes the students of this course have a greater interaction? The work points to relevant factors which should be considered in order to re-think tutors' behavior and help students succeed in DE courses.

### 2. Related Technologies Study

This bibliographic review intends to support the work on the topic raised in the section 2.1: Can tutors constant action promote students' interactions? Besides that, the section 2.2 refers to works done in the field of EDM (Educational Data Mining).

### **2.1 Interactions in Distance Education**

In his article which deals with the tutorial system in DE, Faria (2010) brings information about what some scholars in the area think, regarding the role of the online tutor. For Preti (2002), the tutor has the role of a guide and should support, stimulate and keep up with the learning of the student. As for González (2005), the tutor is a mediator and should answer the issues and doubts of the students in all the learning situations proposed by the tools available in the Learning Management Systems (LMS). Gutierrez and Pietro (1994) work on the perspective of creating good communication channels and empathy through personalization, reflection and experience exchanges during the course. As for Moran (2006), he states that it is the tutor's role to guide and mediate the process of teaching and learning and highlights some characteristics of this professional, such as intellectual, emotional, managerial, communicational and ethical abilities.

Thanks to realization of the importance of the tutor's role, other researches have arisen in order to support this professional, handing in data which may contribute with this pedagogical mediation and with a better performance of the student. The tutor should assume the role of a pedagogical mediator (MASETTO, 2013), and although, he

## CBIE-LACLO 2015 Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)

or she will be a specialist of the content to be taught, he or she will also act as a consultant, as a facilitator, as a planner and as a producer of learning situations. Felicetti *et al.* (2014, p. 2629, our translation) points to the students' perception in relation to the work developed by the tutors. According to the analyses in their work, "in different moments, students claim that the education/ educators might be distant, but the constant presence of the tutors favors students' confidence".

In order to promote motivation, tutors and learners must be in constant interaction through the LMS. For Moore and Kearsley (2005) there are three distinct types of interactions in a distance learning environment, as follows: the interaction of the learner with the content, interaction with the instructor and interaction with other learners. For them, both the interactions instructor-learner and among learners stimulate students' interest in the subject and motivate them to learn.

For some authors (SILVA & FIGUEIREDO, 2012; FARIA, 2010) it is important to consider and promote human relations in distance education courses. They also believe that the establishment of a communication channel between the tutors and the students can guarantee a greater engagement of the students through the affection bonds created due to the constant interactions. Thus, positively interfere in their learning processes.

Aiming at these goals, models for learning monitoring have been created through data mining from the LMS most used, the Moodle (GOTTARDO *et al.*, 2012; LOPES & SCHIEL, 2004). These tools also point to the importance of the constant presence of the tutor as a factor for the success of the students during their learning.

#### 2.2 Educational Data Mining (EDM)

Romero (2010) says that educational data mining allows the tutor to get to know their students based on the data obtained in the learning environment and with that, validate and evaluate the educational systems and then improve some aspects of the learning processes. Romero and Ventura (2007) report that these mined data are not only transformed in knowledge, but they can also help for decision making in pedagogical strategies.

In Gottardo et al. (2012) a group of attributes was defined to produce inferences related to the performance of DE students. Experiments with Moodle's data base were executed and they tried to determine the accuracy of the selected attributes. The results showed a feasibility to perform these inferences related to the performance of the students and they also pointed to future investigations which might add intelligent resources to the LMS to help students in the development of their courses in DE. Santana *et al.* (2014) proposed a complementary work to Gottardo's *et al.* (2012) using different attributes to the ones used before and they also could cope with a good rate of accuracy, 72%. Lopes and Schiel (2004) proposed through educational data mining the development of a Decision Making Support System to monitor the learning in DE.

There are several data mining techniques in educational systems in which computer paradigms converge such as, according to Romero and Ventura (2007): decision tree construction, rule induction, artificial neural networks, instance-based learning, Bayesian learning, logic programming, statistical algorithms, etc. To complete this research, statistical algorithms were used, based on the linear regression model, in order to understand if the tutors' actions regarding the different types of communications with students in the LMS interfere in the level of students' interactions. Rodrigues *et al.* 

(2013) showed that it is possible to use linear regression techniques to obtain inferences with good accuracy rate.

## 3. Experiment Implementation

The objective of this experiment was to estimate the existing relation between tutors' actions and students' level of interactions through the LMS.

To do this experiment a database of graduation courses was used with 829 students and 68 tutors among 48 courses. The average of students per course is 78.19 (Standard Deviation 63.55), and the average of tutors per course is 3.60 (Standard Deviation 3.37).

The experiment did not consider dropouts. The second phase of the experiment consisted on the selection of the attributes. And they are as presented in the table 1 below:

Explanatory Variables				
msgEnvTutor	Number of private messages sent by tutors to their students during the course			
acessoProf	Number of accesses of the tutor to the platform			
qutPostsForumTutor	Number of posts done by the tutor in the discipline forum			
demoraCorrecaoTutor	This variable measures how long the tutor took to correct an assignment after it had been posted.			
atrasoCorrecaoTutor	This variable measures how long the tutor took to correct an assignment after the time set to answer had been expired.			
Dependent Variable				
acessosAlunos	Number of accesses of students from one group to the platform			

#### Table 1. Descriptive Statistics

To perform this work a multiple linear regression modeling was used. This technique aimed at studying the relation between the explanatory variables, which are presented in the linear form, and one metric dependent variable. Thus, a general model of linear regression can be written as follows:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + u$$

Where Y is the phenomenon being studied (metric dependent variable),  $\alpha$  represents the intercept (constant),  $\beta_k$  (k= 1,2,...,n) are the coefficients of each variable (angular coefficients),  $X_k$  are the explanatory variables (metrics) and u is the error term.

### 3.1 Correlation Matrix

Initially, to have a general view of the relation among the variables in the study, it was necessary to plot a correlation matrix to visualize the strength of the relationship of the variables involved in the study.



Figure 1. Matrix of the correlation of the variables studied.

Through the Figure 1 it is possible to verify that there is a positive correlation among several variables proposed on this study. Although the highest correlation occurred between the variables *atrasoCorrecaoTutor* and *demCorreT*, this correlation makes no sense as both of the variables belongs to the tutor category. The most outstanding correlation was between the variables *acessosAlunos* and *qntPostsForumTutor* (*cor*. 0.793), which indicates a strong positive relation.

Table 2 shows the results of descriptive statistics with respect to six variables involved in the study. The descriptive summary was important to view the range of variables, as for the average and the standard deviation.

Variable	Minimum	Maximum	Average	Standard Deviation
acessosAlunos	37.000	17102.000	4264.429	3746.913
msgEnvTutor	3.000	44890.000	5768.143	11671.639
acessosTutor	25.000	3348.000	1185.786	1082.725
qntPostsForumTutor	0.000	2275.000	453.571	545.694
demoraCorrecaoTutor	0.000	37.169	10.104	9.676
atrasoCorrecaoTutor	0.000	37.066	6.554	9.485

#### Table 2. Descriptive Statistics

### **3.2** Application of multiple linear regression model

The table below summarizes the proposed model and through it, it is possible to verify that the group of variables explains 83.9% of the Y (acessoAlunos) variance. It is also noticeable that there is no problems with left overs, since the Durbin-Watson statistics is approximately equals to 2.

Model	R	Squared R	Adjusted squared R	Estimate standard error	Durbin- Watson
1	.860ª	0.839	0.808	1953.123	1.915

Table 3. Regression Model Summary

Table 4 shows that the proposed model is statistically significant, in other words, at least one of the explanatory variables included initially is significant to explain the students' behavior regarding the quantity of accesses to the Moodle platform.

Model		Sum of squares	df	M ean Square	Z	Sig.
1	Regressão	453504910.624	5	90700982.125	23.777	.000 <sup>b</sup>
	Resíduo	160216910.626	42	3814688.348		
	Total	613721821.250	47			

Observing the table of coefficients we can notice that not all the explanatory variables have significance sig.t < 0.05. Initially, we noticed that the variables *acessosTutor*, *demoraCorrecaoTutor* and *atrasoCorrecaoTutor* present Sig.t > 0.05 and, therefore, should be left out of the model.

Tabela 5. Coefficients

Model	Not standardized coefficients		Standardized coefficients	t	Sig.
	В	Standard error	Beta		
(Constant)	1263.449	512.899		2.463	.018
msgEnvTutor	.097	.026	.359	3.699	.001*
acessosTutor	.090	.241	.032	.375	.710
qntPostsForumTutor	4.384	.714	.619	6.139	.000*
demoraCorrecaoTutor	-15.893	44.138	054	360	.721
atrasoCorrecaoTutor	62.726	40.711	.234	1.541	.131

\*significant variables at 5%.

Non-significant variables were dropped out of the model. The following equation summarizes the model which has only significant parameters at 5%. Equation of the regression model:

acessosAlunos (Y) = 1532.6+0.11\*msgEnvTutor+4.61\*qntPostsForumTutor

This model can be useful to managers of courses in Distance Education, as from the same can be prepared forecasts of the quantity of students' interactions (accesses) in a group, based on the action of the tutors regarding a larger amount of private messages changed with students (msgEnvTutor) and a greater interaction in the discussion forums (qntPostsForumTutor).

#### 4. Results and Discussions

Some important points to highlight in the descriptive analysis phase were the evidence that the posture of the tutors regarding his tasks on the platform would promote more participation and interaction on the students. This can be seen in Figure 1.

About the inferential part of the experiment, we can say that the multiple linear regression modeling explains the relationship between tutors actions' data and the level of interaction of students with the Moodle platform. This linear model was considered a good model to explain that there is a relation in the amount of interactions (accesses) of students in a class, in the amount of private messages exchanged with students and interactions in discussion forums.

We can also say, with a 95% confidence, that 83.9% of tutors who have a good level of interaction through private messages and posts on forums promote an increase in students' interaction with the platform. This result is satisfactory if we think we can apply models of this nature to try to predict the proportion of courses that have a good level of access and interactions among participants, based on the teaching activities of the tutors involved.

## 5. Conclusions

The Distance Education emerges as an important innovation for education, but some characteristics of the modality contribute to the necessity of greater human relations. As observed in our results, the relationships tend to create an improvement in engagement and motivation of students, by becoming more participative.

All in all, we can conclude that the regression analysis is useful in areas where future planning activities is critical because decision-making during the course by tutors, teachers and managers is a challenge. Thus, it is evident the important role that the EDM brings to DE. It also proves statistically what scholars have been saying about the importance of human relations in LMS platforms. The more integrated and involved students are, the less they are prone to give up their courses.

As a suggestion for future works we could actually draw a relationship between the analyses regarding the interactions and the number of dropouts of the courses, using statistical algorithms such as the linear regression model.

### 6. References

Associação brasileira de educação a distância. Censo EaD.br: relatório analítico da aprendizagem a distância no Brasil 2013. 1 ed. Curitiba: Ibpex, 2014.

Bentes, Roberto de F. A avaliação do tutor. In: Litto, Frederic M.; Formiga, Marcos (org.). Educação a Distância: o estado da arte. 1ed. São Paulo: Pearson Education do Brasil, 2009. p. 166-170.

#### CBIE-LACLO 2015

Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)

- Faria, Elísio V. de. O tutor na educação a distância: A construção de conhecimentos pela interação nos ambientes midiáticos no contexto da educação libertadora. Scientia FAER, Olímpia - SP, Ano 2, v. 2, p. 28-37, jan./jul. 2010.
- Felicetti, Maria de Fátima et al. O processo de ensino-aprendizagem na ead: a percepção do discente em relação ao trabalho desenvolvido pelos tutores e professores. In: ESUD-2014 XI Congresso Brasileiro de Ensino Superior a Distância, Florianópolis: UNIREDE, 2014. p. 2621-231.
- Gonzales, Mathias. O Tutor na EAD: Dimensões e funções que fundamentam sua prática tutorial. 2005. Available in: http://www.webartigos.com/articles/18336/1/o-tutor-na-ead-dimensoes-e-funcoes-que-fundamentam-sua-pratica-tutorial/pagina1.html.
- Gottardo, Ernani; Kaestner, Celso; Noronha, Robinson V. Avaliação de Desempenho de Estudantes em Cursos de Educação a Distância Utilizando Mineração de Dados. In: XXXII Congresso da Sociedade Brasileira de Computação, 2012, Curitiba. 10p.
- Gutierrez, Francisco; Pietro, Daniel. A mediação pedagógica: educação a distância alternativa. Campinas: Papirus, 1994. (Série Internacional do Instituto Paulo Freire)
- Lopes, Claudivan C.; Schiel, Ulrich. Uma Estratégia para Aplicar Mineração de Dados no Acompanhamento do Aprendizado na EaD. In: XIII SEMINCO, 2004, Blumenau. Anais do XIII SEMINCO, set./2004. Blumenau: FURB-Campus IV, 2004. p.87-98.
- Masetto, Marcos T. Mediação pedagógica e tecnologias de informação e comunicação. In: Moran, José M.; Masetto, Marcos T.; Behrens, Marilda A. Novas Tecnologias e Mediação Pedagógica. 21ed. Campinas, SP: Papirus, 2013. p. 141-171.
- Moran, José Manuel. A Avaliação do Ensino Superior no Brasil. Available in http://www2.eca.usp.br/moran/wp-content/uploads/2013/12/avaliacao.pdf.
- Preti, Oreste. Apoio à aprendizagem: o orientador acadêmico. Available in: https://www.nescon.medicina.ufmg.br/biblioteca/imagem/2495.pdf.
- Moore, Michael; Kearsley, Greg. Distance Education: a systems view. 2ed. Belmont-CA/USA: Thomson Wadsworth, 2005. 368p.
- Rodrigues, Rodrigo L.; Medeiros, Francisco P. A. de; Gomes, Alex Sandro. Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem. In: II Congresso Brasileiro de Informática na Educação (CBIE), 2013, Brasil. 10p.
- Romero, Cristóbal. Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics—Part c: Applications and Reviews, Vol. 40, n. 6, Novembro-2010. p. 601-618.
- Romero, C.; Ventura, S. Educational Data Mining: A survey from 1995 to 2005. Expert Systems with Applications.n.33, 2007. P.135-146.
- Santana, Leandro C; Maciel, Alexandre M.A.; Rodrigues, Rodrigo L. Avaliação do Perfil de Uso no Ambiente Moodle Utilizando Técnicas de Mineração de Dados.In:III Congresso Brasileiro de Informática na Educação (CBIE), 2014, Brasil. 9p.
- Silva, Camila G.; Figueiredo, Vítor F. Ambiente Virtual de Aprendizagem: comunicação, interação e afetividade na EAD. Revista Aprendizagem em EAD. Ano 2012, Vol. 1, Taguatinga-DF.