

A pedagogical agent with embedded data mining functions to support collaborative writing

Daniel Epstein

Graduate Program of Computers in Education
Federal University of Rio Grande do Sul
(UFRGS)
Porto Alegre - Brazil
daepstein@gmail.com

Eliseo Reategui

Graduate Program of Computers in Education
Federal University of Rio Grande do Sul
(UFRGS)
Porto Alegre - Brazil
eliseoreategui@gmail.com

ABSTRACT

Internet growth has induced the development of a large number of collaborative tools for online writing and information sharing. Educators quickly realized the benefits of such tools for learners, allowing them to work online, to share their knowledge and help each other. Distance learning is a key concept in today's educational research; collaborative learning environments are becoming widespread, being more dynamic and resourceful. However, distance learning also introduced a series of problems, such as high evasion rates resulting from lack of support and personalized feedback. It has also introduced difficulties for educators to follow and reviews students' assignments. Based on this scenario, the work presented here proposes the development of a pedagogical agent supported by an intelligent tutoring system to provide students and teachers with assistance in order to minimize some of these problems. The use of a pedagogical agent allows students to have a constant feedback and guidance based on the identification of problems that may emerge from an online collaborative writing activity. The presence of this agent is intended to help students coordinate their efforts in writing a text collaboratively, and improve their work in terms of coherence. Furthermore, the pedagogical agent is also able to assist teachers, reporting problems and simplifying their tasks related to the analysis of the work produced by each student. To support our pedagogical agent we propose the use of data mining tools to extract information related to the students' writings, and a recommender system to suggest additional resources.

Keywords

Pedagogical Agent, Intelligent Tutoring System, Data Mining, Collaborative, Distance Learning

1. INTRODUCTION AND MOTIVATION

In the last few years the Web 2.0 and 3.0 have helped proliferate a large number of educationally driven tools, such

as those to support collaborative writing. These tools are meant to facilitate writing in a multiple author environment, allowing users to work on the same document concurrently, notifying them when a document is modified, and maintaining a revision history. Such features may enhance collaboration, increasing group awareness, creating a sense of consciousness in the group members about cooperative team work, and exposing authors to different aspects of writing [1].

Alongside with the internet growth and online collaborative learning environments, data mining has become more important and popular in the education field. From the student's perspective, the possibility to easily search for learning material using indexes and other reference tools increases their resources at the same time that reduces the effort needed to find relevant information. From the teachers' perspective, it can help them to summarize students' writings in a distant learning context [2], to assess the quality of posts in discussion forums [3] and to evaluate the students participation in discussion forums [4]. Furthermore, it can provide useful feedback to teachers so they can easily identify the main concepts in students' writings and the connection between these concepts [2].

The possibility to work collaboratively is attractive both for students and teachers. It allows students to exchange knowledge, to help each other and complement their work with different ideas. However, it also creates several difficulties for teachers when evaluating the work done, once it's hard and demanding to monitor each student production [5]. Besides, in a distance learning context, certain barriers may hinder the establishment and maintenance of distance learning programs, such as technical problems, infrastructure, motivational difficulties, necessary skills, social problems and time/interruptions [6]. This project's goal is to contribute with the development of a social learning environment in which collaborative writing takes place in a cohesive way, minimizing technical and interactive/social difficulties that are inherent to online collaborative work.

This work has also been motivated by the fast increase in the number of collaborative writing tools available online, and the problems that often originate from their use. For instance, the lack of interaction and actual collaboration in collaborative writing tasks, and the lack of supervision and feedback to students' work.

This project proposes a pedagogical agent to be used in a collaborative writing environment, so that it may assist teachers and students in their tasks. The agent uses data mining to identify problems in the students' writings, and based on this information it tries to guide them in improving their collaborative text production. Besides, the agent gives teachers more accurate information about the students' participation, their difficulties and interactive writing process.

2. RELATED WORK

Intelligent tutoring systems (*ITS*) are computer tools capable of providing customized instruction or feedback to learners [7]. They usually operate without the need of human intervention. They differ from traditional content-delivery computerized learning systems for their ability to improve the effectiveness of a learner's experience through the use of an artificial intelligence [8].

ITS often uses a variety of computational resources for analyzing the users interaction with the system. These are adaptive mechanisms, capable of personalizing learning according to individual student characteristics, such as knowledge on the subject, mood and emotion [9] and learning style [10]. They may be programmed to identify user's information as they interact with the system and choose from many actions the one that most likely would be beneficial to each particular user.

However, the ability to properly help the user often depends on the interface between the system and the user. Not rarely, ITS require a virtual character to interact with users. These characters are called *pedagogical agents*. The use of pedagogical agent (PA) in educational applications has demonstrated that these animated characters may improve student's engagement and learning experience [11].

A PA is a human-like virtual character that has the advantage to operate continuously and autonomously. It is capable of searching and interpreting information received or perceived through the system and provides a more natural interaction with the user. PAs are capable of adapting their actions and interventions, providing feedback and guiding problem solving, reflection, understanding and collaborative learning [12, 13, 14].

Among the many benefits of using PAs are the increase of motivation, perception of ease and comfort in the learning environment, the promotion of fundamental behaviors of learning, the realization of a need for personal relationships in learning and gains in terms of memory, understanding and problem solving [15]. Not only PAs can present contents to the users, as they may suggest additional resources, highlighting important issues and recommending new exercises and reference materials according to user's progress [16]. Studies have shown that the use of a PA with text mining features could help students bring relevant contributions to a reading discussion [17]

According to sociocultural theory, learning can be considered a regulatory process that is mediated by social interaction among individuals, cultural artifacts (computer, pedagogical agent) and speech [18]. Users' interactions help them in knowledge constructing and sharing, being internalized by

them during this process. The experiences and knowledge acquired may be reused in their future experiences [19, 20, 21, 22]. The PA may also be seen as a tool capable of mediating learning, once it may interact with learners, individualize feedbacks and foster autonomy and collaborative skills. Social interaction, according to the sociocultural perspective, is essential for the promotion of learning and development [19, 20, 21].

3. RESEARCH PROPOSAL

Based on the highlighted difficulties of the development of collaborative work online, we propose a pedagogical agent to be inserted in an online collaborative writing environment. This agent will be capable of helping students through immediate feedback about their collaborative text production, and it will assist teachers through the presentation of information/indicator regarding students' participation and progress in the assigned tasks. Our goal is to provide a full time assistance to the users of the collaborative environment and to reduce the amount of work needed to analyze their interactions and work.

In order to do that, we designed a pedagogical agent to be integrated in the intelligent tutoring system. The ITS will be responsible for collecting all information regarding the students activities in the environment. The student could be adding text, images, audio, video or any other resource to the project or simply reviewing and modifying some previous work. In any case, the ITS must keep a log of those interactions in order to determine which action to perform (when needed). Among the different information collected by the system, we may list: the student's contribution to the project (either by the addition of new contents or revision/editing of previous work); frequency (how many times each student accessed the collaborative environment, for how long and when) and a concept map summarizing what has been written by each student. Once we are considering a collaborative environment, several users may be at the system at the same time and it is important for the ITS to identify which student is responsible for which action. Whether the students access the environment simultaneously or separated, it is important for all users to have an identifier that will inform the system which user is current online and modifying any given document.

All the information collected by the ITS will be processed using data mining techniques. This allows the system to identify what type of contribution the user has made to the project and infer how cohesive and coherent the text produced collaboratively is. A data miner similar to Sobek[2] will be used to perform these tasks. Sobek is a text miner that uses statistical analysis to obtain the most relevant concepts in a text and the relationships between them. A data mining process will be used to convert multimedia resources present in the project in concepts and relations.

The system will combine the data extracted from the users' writings with the data provided by the teacher to evaluate if the students project is related to the requirements. In order to do it, the agent will compare the concepts and relationships extracted from the students' writings with the concepts and relationships extracted from the task specification and resources provided by the teacher. Breno et.

al. [4] showed that this kind of comparison could provide useful information regarding the quality/relevance of students' contributions in discussion forums. The results of this comparison will determine if the PA has to make any intervention to help students improve their text. It is particularly important for this intervention to result in positive reinforcement.

There are several types of interventions planned for both students and teacher. The most common type of interventions for students is a direct message sent by the PA inquiring about some aspect of the project. Those inquiries are intended to foster critical thinking and help students correct what the ITS identifies as a problem. The messages are sent when the students' work is incomplete or lack cohesion. In both cases, it is possible for the agent to suggest additional material that may help them correct the problem. Another possible intervention is the use of e-mail messages when students are not participating in the collaborative work, or when their contributions are not coherent with the remaining project. The last type of intervention is not an automatic answer from the PA, but an explicit request for help from a student. Specific functions are being developed to enable students to ask the PA for further information about some aspect of the project, or about its structure and coherence.

A key feature in those interactions between PA and students is the agent's ability to identify additional resources that may help the students' text production. As the students' project may include several types of media and different format of resources, the agent is able to recommend students with additional learning objects extracted from repositories as well as from the web. Learning objects are usually indexed using metadata, where keyword and object descriptions are very common, information that is used to search for learning objects that are most related to the topic at hand. Using a similar technique to the one used by Breno et. al. [4], the learning objects that present the highest similarity values are presented to the students. Learning objects selected will be separated by their format (video, audio, image, text, etc.) so that it will be easier for the student to select the most appropriate ones.

Some of the PA's functions are specific to help teachers. As it is often difficult and time consuming for teachers to analyze the individual production of each student in collaborative writing tasks, the PA will provide accurate information about each student's contributions and progress based on the information collected by the ITS. This will allow teachers to identify relevant contributions made by each student or the absence of a student in a particular subject or part of the project.

The interaction between the PA and the teachers will be set in two ways. The first one is through direct messages sent by the PA to the teacher's email or other communication method. Those messages are sent to inform about lack of participation from some student or if the system detects that one or more students need further assistance with the project (this may be identified through a constant request for help from a student, when his/her work is always identified to have coherence/cohesion problems, or when the student's

work is constantly edited by other students). The second form of interaction will be a support interface, where the teacher will be able to request specific information regarding student's activities in the environment. All those PA features are meant to provide a more personalized contact between the agent, the students and/or teachers.

Another aspect the agent is concern about is the coherence of the project. The students may not be together when writing the project and it may result in disjoint texts, unrelated or redundant information. Therefore, it is important for the agent to identify coherence problems and contact the students who produced the incoherent parts of the project. The agent may also use the information gathered from those incoherent parts to search for additional material and learning objects that could help students fill the possible gaps in the project. Although we considered using Latent Semantic Analysis to perform this task [23], it is still problematic to decide on how to interpret multimedia resources that are neither text nor learning objects (in which case it is possible to use its keywords and descriptions).

4. STATE OF THE PROJECT

In the current state of the project, we are developing a script program that will collect the data from students' writings and send it to the pedagogical agent. This is part of the development of our intelligent tutoring system. This is also one of the most challenging parts of this project, as it is important to correctly interpret and evaluate multimedia resources. Through the use of scripts without a particular user interface, we intend to create a more reusable system, allowing it to be used in many tools and environments without the need for changing the program's core.

Our data mining system based on Sobek is already being modified to provide useful information regarding resource similarity, concepts and relationships. The input for our data miner is very restrictive, but we are working on making it more general. This is most useful to investigate how to successfully mine multimedia resources. Using the learning object repository, we may conduct experiments that will assess the quality of the results and the reliability of our data mining tool.

The experiments will be carried out in projects using Google Drive ¹. The choice for this environment has been the extensive database and number of projects that are developed with this technology daily. It is one of the most known and complete environment for collaborative writing, and it also allows the input of scripts that could facilitate the integration of our pedagogical agent with the system. Google support and APIs make it a natural choice for our project to be integrated in Google Drive Documents, to ensure the dissemination of our ideas and software to a larger number of people across the world.

5. REFERENCES

- [1] S. G. Tammara, J. N. Mosier, N. C. Goodwin, and G. Spitz, "Collaborative writing is hard to support: A field study of collaborative writing." *Computer*

¹<https://drive.google.com>

- Supported Cooperative Work*, vol. 6, no. 1, pp. 19–51, 1997.
- [2] E. Reategui, D. Epstein, A. Lorenzatti, and M. Klemann, “Sobek: a text mining tool for educational applications,” 2011.
- [3] Y. Chen, X. Cheng, and Y. Huang, “A wavelet-based model to recognize high-quality topics on web forum.” in *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. IEEE, 2008, pp. 343–351.
- [4] B. F. Azevedo, P. A. Behar, and E. B. Reategui, “Qualitative analysis of discussion forums,” 2011.
- [5] A. A. Juan, T. Daradoumis, J. Faulin, and F. Xhafa, “Developing an information system for monitoring student’s activity in online collaborative learning.” in *Proceedings of the International Conference on Complex, Intelligent and Software intensive Systems, CISIS*, 2008, pp. 270–275.
- [6] F. Dabaj, “Analysis of communication barriers to distance learning: A review study,” *Online Journal of Communication and Media Technologies*, 2011.
- [7] J. Psotka, L. Massey, and S. Mutter, *Intelligent Tutoring Systems: Lessons Learned*. L. Erlbaum Associates, 1988.
- [8] P. Brusilovsky and C. Peylo, “Adaptive and intelligent web-based educational systems,” *International Journal of Artificial Intelligence in Education*, vol. 13, pp. 159–172, 2003.
- [9] S. D’Mello, R. W. Picard, and A. Graesser, “Toward an affect-sensitive autotutor,” *IEEE Intelligent Systems*, vol. 22, no. 4, pp. 53–61, 2007.
- [10] V. Yannibelli, D. Godoy, and A. Amandi, “A genetic algorithm approach to recognise students’ learning styles.” *Interactive Learning Environments*, vol. 14, no. 1, pp. 55–78, 2006.
- [11] S. D. Craig, B. Gholson, and D. M. Driscoll, “Animated pedagogical agents in multimedia educational environments,” 2002.
- [12] R. Moreno, “Multimedia learning with animated pedagogical agents,” 2005.
- [13] R. Moreno, R. E. Mayer, H. A. Spires, and J. C. Lester, “The case for social agency in computer-based teaching. do students learn more deeply when they interact with animated pedagogical agents?” 2001.
- [14] D. M. Dehn and S. van Mulken, “The impact of animated interface agents: A review of empirical research,” 2000.
- [15] A. Gulz, “Benefits of virtual characters in computer-based learning environments: Claims and evidence,” 2004.
- [16] E. B. Reategui, L. M. A. Santos, and T. L. M. R., “Pedagogical agents and the efficiency of instructional conditions in educational applications,” 2012.
- [17] I. da Costa Pinho, D. Epstein, E. B. Reategui, Y. Correa, and E. Polonia, “The use of text mining to build a pedagogical agent capable of mediating synchronous online discussions in the context of foreign language learning,” *2013 IEEE Frontiers in Education Conference (FIE)*, pp. 393–399, 2013.
- [18] D. H. Jonassen, C. B. Lee, C. Yang, and L. Laffey, “The collaboration principle in multimedia learning. the cambridge handbook of multimedia learning,” 2005.
- [19] L. S. Vygotsky, “Mind in society: the development of higher psychological processes.” Cambridge, MA, 1978.
- [20] L. S. Vygotsky, “Thought and language.” Cambridge, MA, 1986.
- [21] —, “The collected works of l. s. vygotsky. v.1. thinking and speaking.” New York, N.Y.: Plenum Press, 1987.
- [22] J. Lantolf and S. Thorne, “Sociocultural theory and the genesis of second language development. oxford university press, 2006.” *Applied Linguistics*, vol. 28, no. 3, p. 477, 2007.
- [23] M. A. Hearst, “The debate on automated essay grading,” *IEEE Intelligent Systems* Vol.15 N.5, 2000.

Daniel Epstein has Master Degree in Computer Science from Universidade Federal do Rio Grande do Sul working with robotics and artificial intelligence. He is a Ph.D. student at the Post Graduation Program in Computer Science in Education, currently researching Data Mining and Pedagogical Agent use on collaborative learning. He started the program in March 2013 and is currently writing his thesis proposal. He will complete his PhD in the middle of 2016.

Curriculum: <http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4475739T8>

Eliseo Berni Reategui has a PhD degree in Computer Science from the University of London, England, and MS in Computer Science from Universidade Federal do Rio Grande do Sul, Brazil. After finishing his PhD and working in the industry for 5 years, Dr. Reategui held a lecturer position at the University of Caxias do Sul, Brazil, for a few more years. Nowadays, he works as a lecturer and researcher at the Federal University of Rio Grande do Sul. His research interests are related to the use of computers in education, involving areas such as artificial intelligence and human-computer interaction.

Curriculum: <http://lattes.cnpq.br/9140136724972740>

The main topics on which the student would like to receive advices are:

- Data Mining
- Recommender system