

Use of a CSCW platform by trainers and trainees; Trace analysis: multimodal analysis vs data mining approach

Jean Simon, Henri Ralambondrainy

▶ **To cite this version:**

Jean Simon, Henri Ralambondrainy. Use of a CSCW platform by trainers and trainees; Trace analysis: multimodal analysis vs data mining approach. International Conference of the Learning Sciences (ICLS 2012), Jul 2012, Sydney, Australia. pp.n.c. hal-01187215

HAL Id: hal-01187215

https://hal.univ-reunion.fr/hal-01187215

Submitted on 5 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Use of a CSCW platform by trainers and trainees
Trace analysis: multimodal analysis vs data mining approach

SIMON Jean, RALAMBONDRAINY Henri
LIM
Université de la Réunion, Avenue René Cassin
97400, Saint Denis de la Réunion, FRANCE
{jean.simon ; henri.ralambondrainy}@univ-reunion.fr

Abstract: The purpose of this paper is to show that data mining tools can help refine the understanding of what happens on a platform of collaborative work. For this, we compare a multimodal analysis of traces with an analysis of these same traces performed by data mining tools. The second analysis confirms the results of the first one on some points but also gives more information and reveals some "dysfunctions" in the use of the platform. So, it seems interesting to use such tools to allow a formative evaluation of the training devices.

Introduction

The Reunion Island teacher training school trains the future teachers. Since 2005, trainers and trainees use a platform for computer supported collaborative work (CSCW) to be formed. Trainees and trainers create there and share folders where they deposit and draw various resources which must help them to make class. The objectives for the trainers are: to deposit documents and to serve as "collective memory", to improve lesson plans proposed by the trainees, to facilitate preparation of workshop of practice analysis, to pool and share within the framework of the dissertation, to help online and at distance trainees during the training period when they are in charge of a class, to validate the certificate C2i2e which confirms that the trainee is able to use ICT in education. In this paper we compare two analyzes of the 77 folders shared by 15 trainers and 277 trainees, in 2006-2007.

Multimodal analysis of the folders

In 2009 (Simon, 2009), we proposed a first classification of those folders described by 14 numeric variables: number of members, of trainees, of trainers, total of producers, of trainer producers, of trainee producers, of trainer readers, of trainee readers, total of documents, of documents produced by the trainers, of documents produced by trainee, total of readings, of readings made by the trainers, of readings by trainee.
This was done through successive analyses of the data collected: Step 1, volume of exchanges (statistical analysis), Step 2, number of producers (social network analysis), Step 3, type of producer: trainer or trainee (social network analysis), Step 4, analysis of the titles of the folders (text analysis). This multimodal analysis has proposed 6 categories:

- Category 1, 7 folders, corresponds to a set of test or error folders because they don't contain any documents.
- Category 2, 21 folders, corresponds to documents put at the disposal of the group by one trainee producer. They contain lessons plans or sequences or dissertations put on line and subject to the approval of the trainer.
- Category 3, 14 folders, (one producer) and Category 4, 13 folders, (many producers) are documents put at the disposal of the trainees most often by the trainer: the titles of the documents refer to the discipline and/or the level but also to the group of trainees at the IUFM.
- Category 5, 20 folders, is the set of folders that must allow trainees to pass the c2i2e. As mentioned, the c2i2e certifies that the student is able to use ICT in the classroom. Normally, all these folders must have the same activity: more or less the same average number of documents, of readings…
- Class 6, 2 folders, is singular by its intense activity. It corresponds to the folders that have been set up to accompany students during their training course and are intended to answer questions "just in time" and "just enough".

It is observed that, from category 1 to category 6, the activity in the folders is increasing.
We wanted to compare this approach to a more formal approach that uses data-mining methods.

Analysis of the folders by data-mining methods

In a first step we proceed to a quantitative analysis of the folders described by the numerical variables seen above by using data-mining methods (Correspondence Analysis, Ward and k-means algorithms, with the software SPAD). This first study has revealed the salient values for each variable. Some had already been identified by the multimodal analysis, for example one trainer producer by folder, but others not, in particular the size of the groups

associated to folders. After that, to proceed to a qualitative analysis of the folders we defined for each variable six modalities, interval of values, based on the salient values discovered previously. Then we applied again a data-mining approach on the data recoded in those modalities. By this way we obtain 7 classes:
The classes 1 and 7 cover categories 1 and 6 of the multimodal analysis,

- Class 1 "empty folders", 7 folders, which is characterized by a null value on all variables except members.
- Class 7 "accompaniment during training course", 2 folders, which is characterized by the higher modality on all variables

The classes 2,3,4,5 and 6 cover the categories 2, 3, 4 and 5, but there is no one-to-one correspondence between classes and categories:

- Class 2 "individualized accompaniment", 11 folders, is characterized by the fact that there are only two members per folder: one trainer and one trainee and one single producer, the trainee. These folders correspond to a work requested by the trainer to the trainee or trainee's request for assistance to the trainer.
- Class 3 "weak cooperation", 12 folders, is disparate. As the folders have mostly several producers, they do not fall within the "dissemination" classes below, but if we talk about cooperation it should be mentioned that this is low cooperation. It is characterized only by a homogeneous and relatively small number of documents deposited in folders and, to a lesser extent, a small number of members. *It contains some folders of the category 5.*
- Class 4 "dissemination to one group", 22 folders, includes on average one producer. Here folders are used to disseminate information to trainees of the group. In fact, as the average number of readings is relatively small, 18, we may question the effectiveness of this dissemination.
- Class 5 "dissemination to several groups", 4 folders, contains folders that are characterized by a high number of trainees and concern several groups of them. For 3 folders, there is only one producer. They are used to disseminate information.
- Class 6 "strong cooperation", 19 folders, consists mainly of ICT folders, but *it doesn't cover the entire category 5*. These folders are used to validate the trainee's c2i2e what implies a certain number of exchanges and in particular certain homogeneity in the number of readings by the trainers, in the number of trainee producers and in the number of documents that they deposit. This is a class where there are really production and reading and therefore a strong cooperation, but, however, definitely less than in Class 7.

This shows that the analysis performed data-mining methods provides a relevant vision in terms of training and allows distinguishing the different approaches adopted by trainers on the platform and the various uses they make of it. It has led to highlight interesting patterns of behavior as class 2 and class 5 absent from the categorization obtained by the multimodal analysis. For class 2, we see that a CSCW platform can compete with the email usually used in this case. For class 5 the interest is in its opposition with class 4 (diffusion to several groups vs to one group). The diffusion to one group corresponds to a trainer giving documents to his trainees, documents that he has used or will use in his course face-to-face. The diffusion to several groups is a little bit different: it corresponds to documents deposited by a trainer for the trainees for a possible use. In this case the folders are used like databanks. *The use of the data-mining tools has also showed that some folders that should have been in "strong cooperation", class 6, are in "weak cooperation", class 3, and, thus, do not satisfy the expected contract. This is the case for some folders used to validate the c2i2e.*

Conclusion

From our point of view, one essential contribution of data mining tools is to "objectify" our observation. While the multi-modal analysis approach is more a top-down approach, the data-mining approach is more a bottom-up approach. Romero & Ventura speak about an approach of "discovery driven" "in the sense that the hypothesis is automatically extracted from the data" (Romero & Ventura, 2007). The software proposes classes and the researcher must find what the relevance of these classes in his field is. In doing so, he questioned the discrepancy between what it should be and what it is really. Thus, data-mining can be used as a tool for evaluating the work done on the platform. It can lead to improve this device and so to improve the results. It can serve also as feedback to trainers to improve their practice and administrators to improve the system (Romero & Ventura, 2007).

Bibliography

Romero, C., Ventura, S. (2007) Educational Data Mining: a Survey from 1995 to 2005. Expert Systems with x Applications. Elsevier 1:33 (2007) 135-146.

Simon J, (2009) Three years of use of a CSCW platform by the preservice teachers and the trainers of the Reunion Island teacher training school, ICALT 09, *Proceedings of the 2009 Ninth IEEE International Conference on Advanced Learning Technologies - Volume 00,* pp 637-641, Riga, 2009