Towards a Recommender System from Semantic Traces for Decision Aid

Ning Wang[1], Marie-Hélène Abel[1], Jean-Paul Barthès[1] and Elsa Negre[2]
[1]UMR CNRS 7253 HEUDIASYC, University of Technology of Compiègne, Rue Roger Couttolenc, Compiègne, France
[2]UMR CNRS 7243 LAMSADE, Paris-Dauphine University, Place du Maréchal de Lattre de Tassigny, Paris, France

Keywords: Traces of Interaction, Recommender System, Semantic Modelling.

Abstract: Collaboration allows integrating intellectual resources and knowledge from all participants in order to achieve individual or collective goals. With the help of informational environments, we can better organize, realize and record collaboration. During interactions among users in such environments, each activity produces a set of traces. Such traces are recorded and classified, based on a model of traces and can be exploited to improve collaboration. In this article, we propose a semantic model of traces and analyze classified traces by means of TF-IDF. We exploit the results to offer users recommendations and decision aid.

1 INTRODUCTION

As an important form of sharing and exchanging information, collaboration is one of the sources of power for human society development and progress (Grudin, 1994). Thanks to information technology, nowadays collaboration can be organized and managed in an informational environment. In such an environment, users achieve their purposes by realizing different actions. We are interested in the results of actions as well as in the actions themselves. For example, not only the content of a document a user has created matters, but also knowing who shares and consults this document at what time, with what kind of frequency, etc. A set of actions, step by step, is defined as a trace (Zarka et al., 2011). After proper modelling and analysis, traces could in return help to indicate the strengths and weaknesses of an individual or of a group of users (Tomaz et al., 2011). Thus, with the information exploited behind traces, we can improve collaboration, as mentioned by works focusing on the reuse of traces for different purposes such as decision aid (García-Crespo et al., 2011), or recommendation (Chang et al., 2013).

In this paper we propose a mechanism that models, records and analyzes users' traces and in return recommends and helps users making decisions which are personalized for either personal purpose or for the entire group. The following tasks are needed to achieve this objective: (i) propose a semantic structure to record the traces, (ii) evaluate the traces using TF-IDF and a semantic distance among the actions, (iii) give recommendations and provide some decision aid accordingly.

The remainder of this paper is organized as follows. We identify various limitations of the current studies on recommendation methods in Section 2. In Section 3 we propose a kind of recommender system for the need of exploiting the traces. In Section 4 we illustrate our method by giving a toy example. Section 5 provides conclusions and mentions directions for future works.

2 RELATED WORKS

Aiming at a better treatment and exploitation of traces, we need to analyze different types of traces as well as the structure of collaboration in a Collaborative Working Environment (CWE). The model of traces proposed by Li (2013) allows an elaborated analysis of interactions among users. It pays special attention to the exchanges of users in informational environment. Traces then can be exploited to feed a recommender system. Interests of recommender systems are justified by the need to

Wang N., Abel M., Barthès J. and Negre E..
Towards a Recommender System from Semantic Traces for Decision Aid.
DOI: 10.5220/0005133502740279
In *Proceedings of the International Conference on Knowledge Management and Information Sharing* (KMIS-2014), pages 274-279
ISBN: 978-989-758-050-5

manage the growing amount of information (Adomavicius 2005).

Recently, various articles were published aiming at exploiting the traces with the help of semantics. Chen et al (2013) present a mechanism for personalized knowledge search and recommendation adapting a suitable domain ontology according to the previous browsing and reading behavior of users. Vesin et al (2012) present a new approach for performing effective personalization based on semantic web technologies. Although more and more attention is focused on exploiting implicit information behind data (data mining), these recent studies did not work much on modelling users' actions. In fact a good model of organizing and storing users' actions can make future exploitation of traces more efficient.

3 OUR APPROACH

Figure 1 shows the structure of our proposed recommender system. First of all, with the model of actions we propose, users' actions are collected and modelled from an interactive platform. After being sifted by the filter of classification, we obtain classified traces, which allows a preliminary presentation back to the users. Alternatively, we apply an algorithm to calculate an index indicating the correlation between the classified traces of a certain user and a subject. These values can lead to useful information that are presented as personalized recommendation, either to a group defined as a set of users in the platform, or to an individual user.

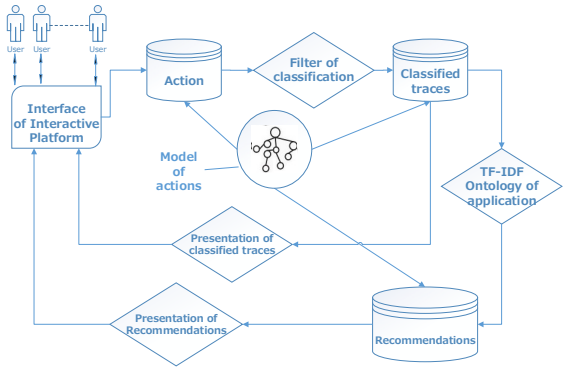

Figure 1: The structure of our proposed recommender system for the exploitation of semantic traces.

3.1 Modelling of Actions

We define the principal concepts as follows:

- Action: an interaction or an act performed by a user in a collaborative environment, e.g. sending a document to other users;
- Classified trace: a set of actions performed by a user in the informational environment classified according to the model of traces (Li, 2013);
- Set of traces: a set of classified traces.

According to our definition, an action is the most basic element forming a trace. As regarded as an important resource for our recommender system, we introduce the Resource Description Framework (RDF) to model actions (Antoniou and Van, 2004). RDF is used as a general method for conceptual description or modelling of information that is implemented in web resources. Figure 2 shows the basic structure in the RDFS graph of our model. An ellipse represents a class of resources and a rectangle represents an object property. For example, a person has the object property of "has_id_person" and the range of this property is the class called "id". In Figure 3, the class "creation" inherits all properties from "action".

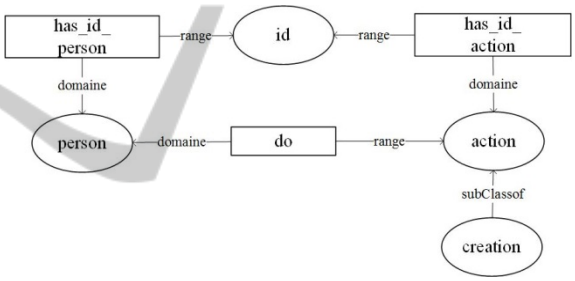

Figure 2: Basic structure in RDFS graph presenting an action.

We apply the model of actions to a web-based collaborative platform E-MEMORAe2.0 (Abel and Leblanc, 2009). Details are shown in Figure 3.

This model of actions has two main advantages compared to a traditional form of history or log of users:

- Actions are presented in a labeled, directed multi-graph. In our model the actions are represented as resources in the RDF schema and they are no longer discrete but are connected by properties. This allows a better structure of storage and usage of actions. For example, a person "Ala" chats with "Ning". This action can be presented by a RDF instance showed in the lower part of Figure 3 where "Ala" and "Ning" are two instances of the class of resource "person". "Chat_1" is an instance of the class of resource "conversation" which is linked to the action "creation".

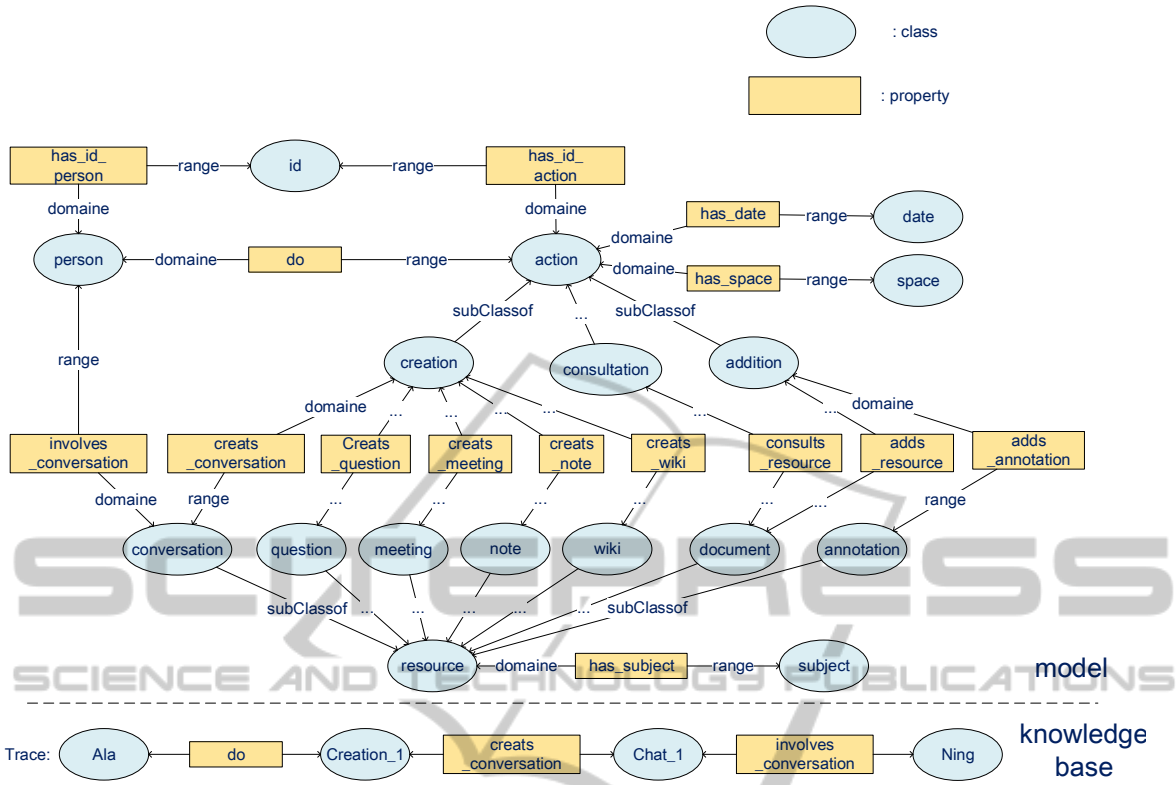

Figure 3: The model of actions in the platform E-Memorae 2.0.

- Normally different types of actions have different importance. For example creating a piece of Wiki is more important than consulting it. In our model the actions are all classified by three classes: creation, consultation and addition which enables us to treat different types of actions more efficiently.

3.2 Application of TF-IDF

In order to evaluate the importance of different traces, we apply TF-IDF as the method of evaluation. TF-IDF (Term Frequency-Inverse Document Frequency) (Jones, 1972) is a statistical method for weighting usually used for the research of information from texts. This method measures the correlation of a term in presenting a document from a corpus. The weight of a term is proportionally higher when it appears more frequently in a document. It also varies with the frequency of the word in the corpus. The values are used for evaluating the reference of a document.

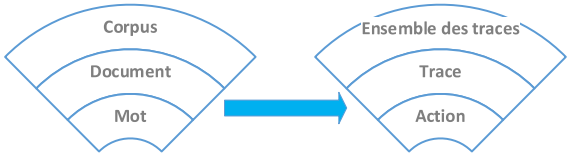

Figure 4: Relation between concepts of the method TF-IDF and the method adapted to our case.

Here we adapt this method to our case. Figure 4 illustrates the key-concepts of the method TF-IDF and those more relevant in our case. Typically TF-IDF focuses on the relation between words, documents and corpus. If a word appears more in a document and at the same time appears less in the other documents of the same corpus, it better represents this document. For our research, we are interested in evaluating the correlation between a trace of a given user and a certain subject. We propose to consider that if the actions of a user are more pertinent concerning a subject, the user has more knowledge about it. So we are able to recommend this user as an expert in this domain. In our case, we study the relation between actions, traces and the set of traces in a group of users working on the same subject.

We adapt the equation of TF-IDF:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

$$idf_i = \log \frac{|P|}{|\{p: t_i \in p\}|} \quad (2)$$

Where:

- $n_{i,j}$: the number of actions concerning the subject i performed by the user j;
- $\sum_k n_{k,j}$: the number of actions concerning all k subjects performed by the user j;
- $|P|$: the number of users in a group;
- $|\{p: t_i \in p\}|$: the number of users in a group who have performed at least one action about the subject i.

The index of TF-IDF, C_{ij} indicating the competence of user j on subject i, which can be regarded as the relevance between a subject and a user, is defined as follows:

$$C_{i,j} = tf_{i,j} \times idf_i \quad (3)$$

3.3 Application of Semantics

When we evaluate the competence of a user on a certain subject, we can also take into consideration the semantic relationship among subjects. As semantics indicates the real meaning, a smaller semantic distance between two concepts means a closer relation of them. For example if we need to find an expert on "Computer", even though a person hasn't left a weighted trace on "Computer" directly, he may have contributed to the subject "Tablet" which is quite close in a semantic view. So we can evaluate the weight of this user on the subject "Computer" by evaluating it on the subject "Tablet". We explain in detail in Section 4 by giving a toy example.

3.4 Classified Recommender System

We take the model of traces (Li, 2013) as the basis to carry out classified recommendations. According to this model, a trace is classified into 4 types: Private Trace, Collaborative Trace, Collective Trace and Personal Trace. A private trace is sent and received by the same user. A collaborative trace has one sender and at least two receivers. A collective trace has many senders and receivers while a personal trace is defined as having only one sender with no limit on its receivers.

Our recommender system takes into consideration a semantic model of the system along with the traces of recorded interactions (e.g. who has shared a document concerning the subject S with whom, e.g., who usually interacts with the expert John?). It aims at realizing recommendations for a group (improve the collaboration, identify risks, opportunities of a set of users from a group), of an individual (how to improve its efficiency, the organization of a user among his tasks), for private purposes (how to improve the private organization of a user) and for collective purposes (how to improve the communication inside a group, etc.).

4 EXAMPLE

Figure 5 illustrates an example of interaction on different subjects of two groups of users with a histogram chart. Each line represents the collaborative trace of a user, for each subject.

Now we evaluate the competence of the user "Ning" on the subject "WP". According to the histogram, "Ning" has realized 13 actions among which 1 action concerns "WP". In group 1, the number of users is 4 among which 3 have realized at least an action about "WP". According to our method of evaluation, we obtain:

$$C_{WP,Ning} = \frac{1}{13} \times \log \frac{4}{3} = 0.645 \quad (4)$$

$$C_{WP,Ala} = \frac{3}{12} \times \log \frac{3}{2} = 2.491 \quad (5)$$

Even coming from two different groups, we can still recommend that "Ala" is more competent than "Ning" on "WP" relying on $C_{WP,Ala} > C_{WP,Ning}$.

Similarly, we evaluate the competence of the two users on other subjects as shown in Table 1.

Table 1: The index of competence of "Ning" and "Ala" about the subjects involved.

	Ning	Ala
WP	0.645	2.491
Android	0	1.661
iOS	3.224	0
Tablet	1.290	0
Computer	0.0463	0.0795

Figure 6 shows a part of the domain ontology in Information Technology. It shows that the subjects "Computer" and "Tablet" are subclasses of the subject "Hardware". Also, these subjects "iOS", "Android", "WP" and "Java" are subclasses of the concept "Software". To compare the capability on the subject "Java" between "Ning" and "Ala", as we are in lack of trace of "Ning" or "Ala" on "Java", we propose that such a subject be measured by the classes (i.e., "iOS", "Android" and "WP") having the same superclass. As they are closest in the view of semantics, evidently they have closest meaning.

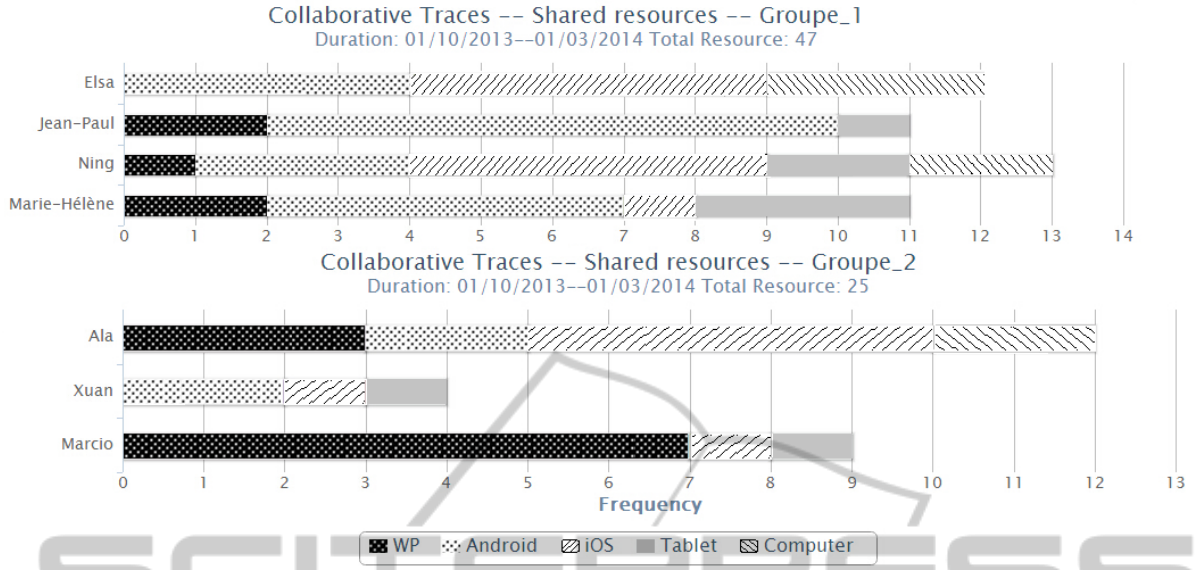

Figure 5: An example of collaborative interactions of two groups.

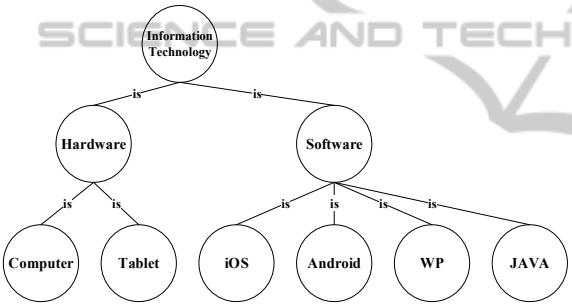

Figure 6: Semantic relationship between subjects in which "Ning" and "Ala" are involved.

We propose:

$$\sum C_{\text{sib}_S,U_1} > \sum C_{sib_S,U_2} \Rightarrow C_{S,U_1} > C_{S,U_2} \quad (6)$$

Where:

- C_{S,U_1}: the index of competence of user U_1 about subject S, in case concerned traces is not given;
- $\sum C_{\text{sib\_S},U_1}$: the sum of index of competence about user U1 concerning subjects which are siblings of subject S;

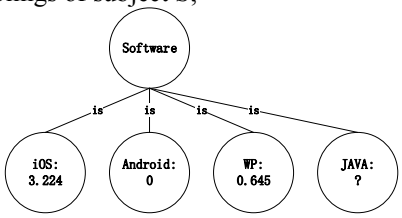

Figure 7: The index of Competence about sibling concepts of "Java" of "Ning".

Here "Java" shares the same ancestor "Software" with "iOS", "Android", and "WP". We obtain:

$$\sum C_{\text{sib\_Java,Ning}} = C_{Android,Ning} + C_{iOS,Ning} + C_{WP,Ning} = 3.869 \quad (7)$$

And,

$$\sum C_{\text{sib\_Java,Ala}} = 4.152 \quad (8)$$

As $\sum C_{\text{sib\_Java,Ala}} > \sum C_{\text{sib\_Java,Ning}}$ we can deduce that $C_{\text{Java,Ala}} > C_{\text{Java,Ning}}$ so that we recommend "Ala" be more qualified than "Ning" on the subject "Java".

5 CONCLUSIONS AND FUTURE WORK

Traces are important records in collaborative environments. Fully exploitation of such information helps us organize and improve collaboration. In this article we propose recommendations based on evaluation of traces using TF-IDF. Moreover we demonstrate that we could solve the problem when there does not exist enough relevant traces with semantics. We illustrated our method on a toy example.

Future works include implementing our proposal of recommender system. In a collaborative environment, the date and time when an action is realized are also recorded. So, it is necessary to take

into consideration the fact that a recent action has more weight than a previous action.

REFERENCES

Abel, M. H., and Leblanc, A. (2009, October). Knowledge sharing via the E-MEMORAe2. 0 platform. In *Proc of the International Conference on Intellectual Capital, Knowledge Management and Organizational Learning* (pp. 10-19).

Adomavicius, G., and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, *17*(6), 734-749.

Antoniou, G., and Van Harmelen, F. (2004). *A semantic web primer* (pp. 253-260). MIT press.

Chang, N., Irvan, M., and Terano, T. (2013). A TV Program Recommender Framework. *Procedia Computer Science*, *22*, 561-570.

Chen, Y. J., Chu, H. C., Chen, Y. M., and Chao, C. Y. (2013). Adapting domain ontology for personalized knowledge search and recommendation. *Information and Management*, *50*(6), 285-303.

García-Crespo, Á., López-Cuadrado, J. L., Colomo-Palacios, R., González-Carrasco, I., and Ruiz-Mezcua, B. (2011). Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain. *Expert systems with applications*, *38*(10), 13310-13319.

Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers.*Communications of the ACM*, *37*(1), 92-105.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, *28*(1), 11-21.

Li, Q., Abel, M. H., and Barthès, J. P. (2012). Facilitating Experience Sharing Groups-Collaborative Trace Reuse and Exploitation. In *Proceeding of International Conference on Knowledge Management and Information Sharing* (pp. 21-30).

Tomaz, L. F., Nt, J. A. R., Souza, J. M., and Xexéo, G. B. (2011, June). Bringing knowledge into recommend-dation systems. In *Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on* (pp. 246-252). IEEE.

Vesin, B., Ivanović, M., Klašnja-Milićević, A., and Budimac, Z. (2012). Protus 2.0: Ontology-based semantic recommendation in programming tutoring system. *Expert Systems with Applications*, *39*(15), 12229-12246.

Zarka, R., Cordier, A., Egyed-Zsigmond, E., and Mille, A. (2011). Trace replay with change propagation impact in client/server applications. Actes de IC2011, 607-622.