Clustering of anonymized users in an enterprise intranet social network

Israel Rebollo Ruiz and Manuel Graña

[1] Israel Rebollo Ruiz at Unidad I+D empresarial Grupo I68, Computational Intelligence Group- University of the Basque Country, beca98@gmail.com
[2] Manuel Graña at Computational Intelligence Group- University of the Basque Country ccpgrrom@gmail.com

Abstract. Modern enterprises aim to implement increasingly efficient working methods. One of the trends is the use of process knowledge to guide users about the proper way of performing specific tasks, getting rid of the misleading influence of inexperienced users or veteran users with incorrect habits. But implementation of this kind of working methodos can give rise to conflicts and general stressful interdepartmental atmosphere within the company. To this end we are designing a system that allows user ordering based on the amount of daily work done so that more proficient users may have a greater influence in their group, helping all the members of the group to a reach higher levels of performance. A key feature of the proposed system is that the users are anonymized at all times. This measures seeks to avoid personal feelings to interfere on the acceptance and generation of working recommendations. Even the number of user groups and their components are unknown at all times, so that the recommendation system is intended to be fully autonomous and self-managing, increasing overall work efficiency by using the experience of some workers but without disclosing the identity of those employees. The aim is to obtain the personal alignment of the user with the proposed recommentations. We report evaluation of the approach on two test cases in a real business environment, where it has been observed that the proposed system is capable of correctly clustering users and identify the most proficient within each group.

Keywords: Anonimization, ERP recommendations, Social Networks, user clustering

1 Introduction

Even in companies where manufacturing labor is the predominant activity, bureaucratic management is becoming more important, needing tools capable of carrying out all work in the shortest time possible and with the least number of incidents [1]. To do this, the Resource Enterprise Planning (ERP) software packages offer a multitude of functionalities to perform all necessary management tasks. However, these tools are increasingly complex, the amount of data handled by the user, and the interaction with other processes makes it sometimes difficult to devise an efficient flow of actions to perform various tasks [2].

The users of these ERPs perform tasks, based on their training and their own personal experience, in patterns that are not always correct [3]. Some users even refuse to use the new tools and try to continue using as far as possible the familiar old methods [4]. There are many factors that lead users to use the ERP tools in one form or another, such as age, education level or even gender [5]. However, in many cases these usage guidelines are incorrect and may lead new employees to adopt them as working habit regardless of its effectiveness and alignment with the guidelines set by the company.

To correct this kind of behaviors, the only effective measure is to provide the correct formation and convince the users that the correct way to perform the tasks is the one set by the company [6]. Training is expensive and has a strong financial impact on ERP implementation [7]. To minimize high training costs, Grupo I68 is designing a system in which the ERP is able to identify the most common actions that a user employs to perform its assigned tasks, based on expert system recommendations and on an ERP function usage log [8]. The system learns from the user and suggests the most suitable applications at any time. However this does not solve the problem of mimicking user bad practices. Additional user clustering creates profiles based on tasks performed so that ERP programs suggested by the system are related to the overall performance of the group, and not just the user. In this article we add the user proficiency as a further factor for recommendation generation. Therefore, a skilled user will have greater influence on the group leading other users to mimick the expert actions, also diminishing the influence of less performant users.

This system can be met with strong opposition from some company employees who despite having extensive experience can not be a role model for the rest of the ERP user group. Thus the members of each user groups, and the expert(s) included in them, are anonymous so that the user does not know if the recommendation received comes from his/her own experience or that of others, eavaluating it only by increased work satisfaction [9]. The number of user groups is unknown, and are spontaneously formed according to user behavior, so that the number of groups varies from one moment to another. For experimental purposes, the identification of the experts has been done manually in the works reported in this paper, assigning a high experience value to a user by a high rank officer of the company, but always preserving anonymity or automatically using a marker with which to measure the efficiency of each user.

The rest of the paper is organized as follow: Section 2 presents the user clustering. In Section 3 we discuss the anonymization of user groups. Finally, 4 present experimental results. The last section is for our conclusions.

2 User Clustering

In order to perform user clustering, we must correctly define the common characteristics shared by two users identifying them as belonging to the same group.

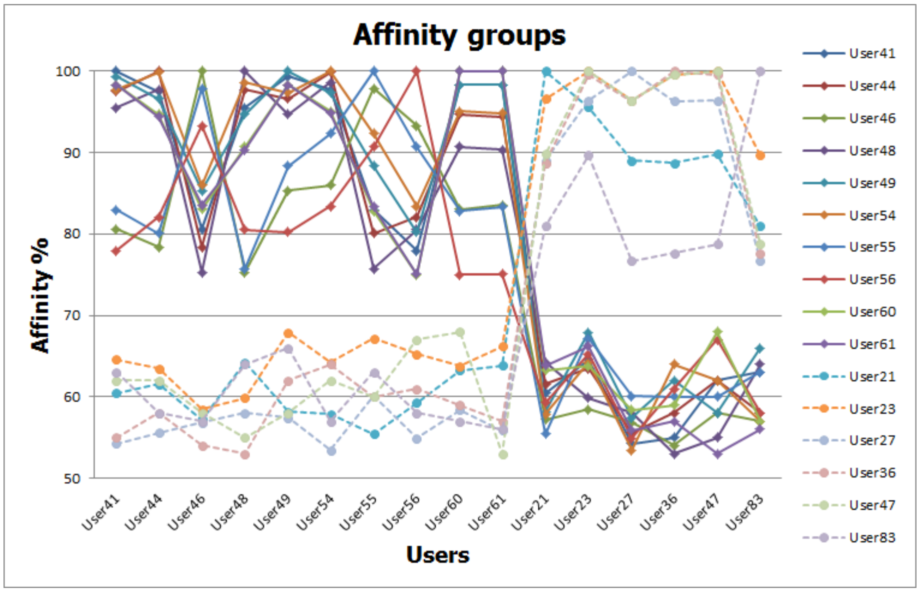

Fig. 1. Affinity Groups

We first define the affinity A between two users as the sum of the Euclidean distance between their working characteristics so that if this sum is less than a threshold U, then it will establish an affinity rapport between the two users.

$$A = \sum_{1}^{n} ||fa_i - fb_i||,$$

$$(A < U) \Rightarrow e(a, b) = 1,$$

where fa_i and fb_i are the i-th features of users a and b, and where $e(a, b) = 1$ means that there is an edge between both users in the affinity graph.

Calculating the affinity pairs, we obtain an affinity graph [10] which defines the different groups, but it may be the case that a user is related to two groups without that these groups are related to each other. For these cases, using a smaller threshold can broke the tie between groups. If lowering the threshold too much dissolves the groups, the user should choose randomly a group. A threshold too low may cause excessive fragmentation of the groups while a very large threshold can create groups in which there is too much difference between any pair of members. The number of groups in which users are divided is arbitrary, but to make a correct tuning of the system, we need to establish upper and lower bounds on the number of groups. These parameters depend on the number of users of the company. User clustering can be achieved by an algorithm for graph coloring, as we have done previously by a Swarm algorithm [11] that offers very good results compared to others in the literature [12,13].

For illustration purposes, in Figure 1 we plot the affinities between some users in one of the industrial studies performed. We can observe that users {User41, User44, User46, User48, User49, User54, User55, User56, User60} have affinity values between them in the order 75% and 100%, allowing to create an affinity group. On the other hand, users {User21, User23, User27, User36, User83, User47} also have a great affinity between them, creating another group. Moreover, this grouping is confirmed by the low affinity between the users from both groups, so that no users can belong to both groups. The data have been obtained from a company with over 400 user, and having anonymized user names for confidentiality issues, a human resources expert from the company has confirmed that the groups generated by the algorithm greatly resembled the daily work members of the same in most cases.

3 Anonymous User Groups

In the previous section we have grouped the users, taking into account various factors. These groups have been reviewed by a human resources expert to check that they are correct [14]. But the problem remains of finding the best performing user. A simple way to identify it would be to check who has the higher average affinity with other members of the group, which means that in a group of users with similar tasks, we select the user that matches in carrying out the tasks with more group members assuming that the expert will be mimicked by others. However, popular users with bad practices may adversely affect the rest of the group.

Figure 2 shows different user states, and how they can evolve from one to another [15]. Using these relationships we are able to characterize the expert user, beyond routinely performed tasks. The experience is achieved with training supported by appropriate studies. Studies and training require an effort that can sometimes be limited by age. Age is an important factor in the behavior of people and the influence of society and, in the case of a company, its working peers. Happiness is another factor that affects the behavior, or that, vice versa, is a consequence of the behavior. We must differentiate between personal and impersonal social influence, where the latter is exerted regardless of personal traits, only on the basis of working performance. Then, to ensure that system experts providing recommendations to the other users are selected on the basis of impersonal value, the system needs to emphasize the need for privacy and anonymous clustering and expert selection. Our system will hide the experience so that it is an imposition that may cause rejection of the employees but will be anonymized. This anonymization will result in increasing working satisfaction of ERP users by excellency in user recommendations.

4 Experiment Results

After consulting with the company human resources expert we have performed the following experiment. We have selected five different tasks consisting of six

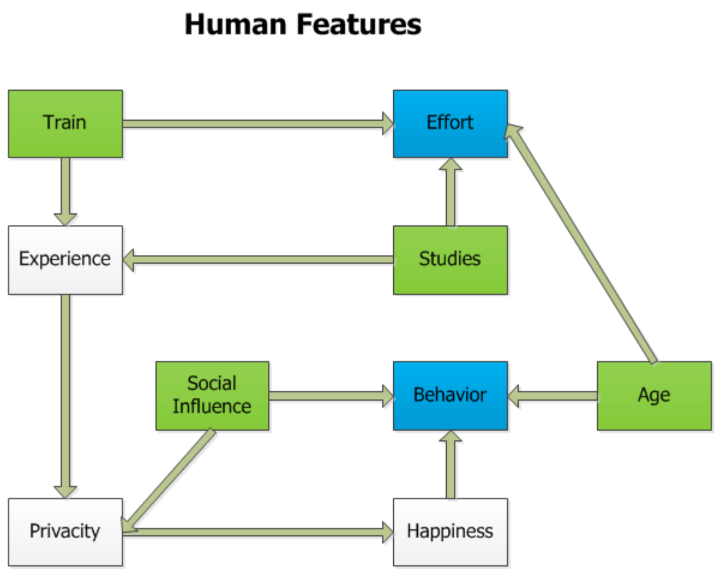

Fig. 2. Human Features relationship

steps. The work had already been done by the users in the group, so that we known *a priori* their behavior. The system has recommended certain actions taking into account the experience of each user and group ownership, taking into account the experience of the group is closely tied to that of its expert. The first steps, the overlap between what the system offers and what users actually did was low, but as the task progresses users tend to make the expert's recommendations.

In figure 3 we plot the percentage of success of the expert on each task, measured as the percentage of users that perform the same action as the expert which performing a given task. We observe that users tend to perform the same tasks as the expert. These results are encouraging to the implementation of the anonymous user clustering and expert detection system, showing that this anonymous "follow the leader" process may improve performance of the overall system by levels above 80 % adoption of expert recommendations in some task performing sequences of actions.

5 Conclusions and future work

We have already managed to create groups of users within a company on the basis of ERP logs. Evaluation by company human resource expert, these groups have been found to correspond to the consistent user categories. In this paper we argue that the user grouping information is to be kept anonymous. We have

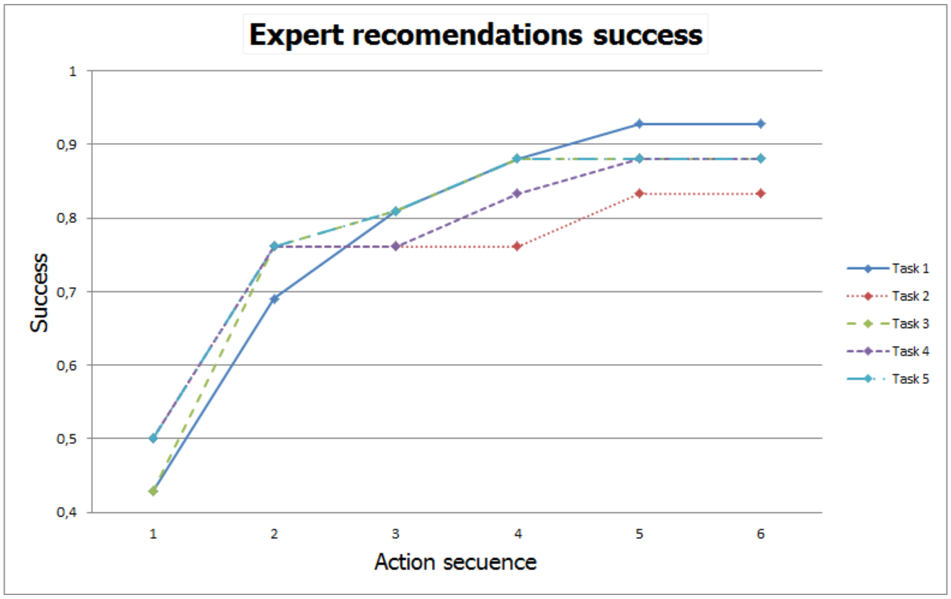

Fig. 3. Expert recommendation success for different tasks

found that even though you may give upper and lower bounds on the number of groups to generate, they are spontaneously formed without having any kind of outside interference. In this paper we have required expert opinion to validate the composition of the user group, however the system will be capable of generating autonomously user groups following the dynamic behavior variations of each user.

In addition we found that if we identify a hierarchy of experience within each group by assigning more value to their recommendations to those of others and not leaving you that will influence others, actions taken to solve various tasks are similar to those of the expert.

Future work is to define the requirements a user must meet in order to be considered an expert. In this way the system will be able to locate the experts in each group autonomously.

Acknowledgements

The Unidad I+D empresarial Grupo I68 company encourages participation in academic meetings, and academic production, within the limits of business confidentiality. The work reported here has been supported by project Gaitek IG-2012/00989 funded by the Basque Government, and also by project of the CDTI IDI-20130070 funded by Spanish Government.

References

1. Kanellou, A., Spathis, C.: Accounting benefits and satisfaction in an {ERP} environment. International Journal of Accounting Information Systems (0) (2013) –
2. Ram, J., Corkindale, D., Wu, M.L.: Implementation critical success factors (csfs) for erp: Do they contribute to implementation success and post-implementation performance? International Journal of Production Economics (0) (2013) –
3. Khan, R.Q., Corney, M.W., Clark, A.J., Mohay, G.M.: A role mining inspired approach to representing user behaviour in erp systems. In Oyabu, T., Gen, M., eds.: The 10th Asia Pacific Industrial Engineering and Management Systems Conference. (December 2009) 2541–2552
4. Gumussoy, C., Calisir, F., Bayram, A.: Understanding the behavioral intention to use erp systems: An extended technology acceptance model. In: Industrial Engineering and Engineering Management, 2007 IEEE International Conference on. (2007) 2024–2028
5. Saima Kanwal, D.I.A.M.: Evaluating erp usage behavior of employees and its impact on their performance: A case of telecom sector. Global Journal of Computer Science and Technology **10**(9) (2010) 34–41
6. Holland, C.P., Light, B.: A critical success factors model for erp implementation. IEEE Softw. **16**(3) (may 1999) 30–36
7. Poston, R., Grabski, S.: Financial impacts of enterprise resource planning implementations. International Journal of Accounting Information Systems **2**(4) (2001) 271 – 294
8. Rebollo-Ruiz, I., Graña-Romay, M.: User assistance tool for a webservice erp. In: 11th International Conference on Practical Applications of Agents and Multi-Agent Systems, Salamanca (2013)
9. Phelps, E., Zoega, G.: Corporatism and job satisfaction. Journal of Comparative Economics **41**(1) (2013) 35 – 47
10. Opsahl, T., Panzarasa, P.: Clustering in weighted networks. Social Networks **31**(2) (2009) 155 – 163
11. Rebollo, I., Graña, M.: Gravitational Swarm Approach for Graph Coloring. Volume 387 of Studies in Computational Intelligence. Springer-Verlag (2011 159-168)
12. Rebollo, I., Graña, M.: Further results of gravitational swarm intelligence for graph coloring. In: Nature and Biologically Inspired Computing. (2011) 183–188
13. Folino, G., Forestiero, A., Spezzano, G.: An adaptive flocking algorithm for performing approximate clustering. Information Sciences **179**(18) (2009) 3059 – 3078
14. Rebollo-Ruiz, I., Graña-Romay, M.: Swarm graph coloring for the identification of user groups on erp logs. Cybernetics and Systems (2013) In Press.
15. Destre, G., Levy-Garboua, L., Sollogoub, M.: Learning from experience or learning from others?: Inferring informal training from a human capital earnings function with matched employer-employee data. The Journal of Socio-Economics **37**(3) (2008) 919–938