Virtual Communities of Practice's Purpose Evolution Analysis Using a Concept-Based Mining Approach

Sebastián A. Ríos[1], Felipe Aguilera[2], and Luis A. Guerrero[2]

[1] Department of Industial Engineer, University of Chile
srios@dii.uchile.cl
[2] Department of Computer Science, University of Chile
{faguiler,luguerre}@dcc.uchile.cl

Abstract. Today, social networks systems have become more and more important. People have change their way to relate and communicate. Therefore, how to enhance contents and organization of a social network is a very important task. This way, we can help Virtual communities of practice (VCoP) to survive through time. VCoP are special kind of social network where the purpose is a key aspect. However, administrators are blind when trying to identify how to enhance the community. We propose a method which helps them by analyzing how purpose evolves through time. The approach has been experimentally tested in a real site with successful results.

1 Introduction

The WWW, has become a fertile land where anyone can transform his ideas into real applications to create new amazing services. Therefore, it was just a matter of time until the massive proliferation of virtual communities, social networks, etc. New social structures have been formed by massive use of new technologies. This way, people can relate to other by interests, experiences or needs.

In a scenario where WWW has become more important every day, and people is using more often the web to relate to others, to read news, obtain tickets, etc. The need to well organized web site has become one of the vital goals of enterprises and organizations. In order to accomplish such task web mining area was born about a decade ago.

Web mining are techniques to help managers (or experts) to extract information from a web site content, link structure or visitors' browsing behavior. This way, it is possible to enhance a web site, obtain visitors' interests patterns to create new services, or provide very specific adds depending on interests of visitors ([4,5]).

Today, virtual communities have experienced an exponential growth. Also, the use of web mining techniques to explode data stored in these systems has become a natural approach to obtain knowledge from them. However, a virtual community is not only a group of people accessing a web site, they establish social

J.D. Velásquez et al. (Eds.): KES 2009, Part II, LNAI 5712, pp. 480–489, 2009.

relationship through the use of Internet tools ([12]), allowing the formation of a communal identity and a shared sense of the world ([13]). In order to provide truly valuable information to help managers or web masters it is necessary to take into account the social nature of the virtual communities in web mining techniques.

Although, studies of social aspects in a virtual community are crucial, the evolution analysis of such aspects is more important yet, especially in social structures that allow to its members to define themselves different levels of participation among time, like Virtual Communities of Practice (VCoP) ([14]), provoking that community change according to member's participation.

This work focuses on the application of a novel web usage mining approach to study the social nature aspects of Virtual Communities. Specifically, we focus our work on how VCoP accomplish its purpose through the evaluation of the evolution of its goals' achievement.

2 Related Work

All virtual communities are different because of their social nature, however, there exist several common characteristics which allow to classify them ([11]). We can find interest communities, purpose communities and practice communities.

The VCoP's studies are yet immature, in spite of the studies already developed to analyze interest and purpose communities. For example, in the case of interest communities it is frequent to find recommender systems, evaluation/review systems, ranking systems, etc. Therefore, present paper is focused on VCoP.

2.1 Virtual Communities of Practice

Virtual communities of practice (VCoP) are informal, self-organizing networks of people dedicated to sharing knowledge ([14]). An important characteristic of VCoP is that very commonly some members help other to answer questions and solve problems. This way, they share and create knowledge (shared information, good practices, generate software tools and knowledge bases) ([14]).

Virtual communities of practice evolve in time passing through diverse levels of maturity ([14]). Through time, a VCoP can discover (objectives, needs, etc.) and re-invent itself ([6,14]). Therefore, to be able of study a VCoP it is not enough to perform analysis in a specific time period. It is of major importance to develop new algorithms or techniques to analyze how the VCoP evolve trough time. This way, managers or administrators will be able to discover risk situations or how the users interests are changing.

2.2 VCoP Analysis Techniques

Based on ([7]) we can identify four ways to analyze a VCoP which are commonly used. However, most of these approaches don't consider evolutionary factors in the analysis.

1. Ethnography and associated techniques: The purpose of ethnographic research is to build a rich understanding of a group or situation from the point of view of its members/participants. Disadvantage of this kind of analysis is that *in situ* studies are intrusive.
2. Questionnaires: This kind of analysis are useful for collecting demographic information and have the advantage that they can be distributed by hand to local participants, or posted via email or on the Web ([6,7]). Although, questionnaires provides useful information from virtual community members ([1]), they are not sufficient. It is recommended the use of a secondary technique, when possible, to reduce the subjectivity of members' judgements.
3. Experiments and quasi-experiments: Laboratory studies are valuable for testing the usability of the interface and users' reactions to new user interface features. To apply this approach it is needed to create the virtual community in a laboratory. This is quite complex since it is needed to have a representative group of virtual communities members, in a real (physical) experiment.
4. Data Mining & Social Network Analysis (SNA): This kind of analysis consist of using software's logs to discover useful information from the VCoP. This way, the main point is the study of communities' nature, i.e. the study of its members and the relations they establish ([6]). However, SNA analysis is designed to answer questions like: Who is the expert or experts in the community? Which sub communities or subgroups exists? But, it is not possible to answer: which is the interests of members? Which is the real goal of the community? Which topics are related with the community purpose? Which themes diverge from communities' main purpose? All these question remain unanswered.

We propose that SNA techniques are not sufficient to fully understand a virtual community of practice. Therefore, in this paper we propose a novel method to analyze a VCoP, based on a data mining approach to discover useful information from the VCoP for its better administration.

3 Proposed Methodology

The proposed methodology is based in the VCoP's purpose study. The purpose refers to a community's shared focus on an information need, interest, service. Every user has his own motivations (or interests) to use a specific VCoP and of course every community of practice can satisfy specific users' needs.

Defining the community's purpose is of major importance, since potential participants can immediately find out about the communities goals ([6]). However, VCoP's purpose is not always clear, even worst, it is not clear if all community members are aligned to the same common purpose.

But also, we must consider the evolutionary nature of purpose. Since, people change their motivations, interests, taste, etc. every day. Certainly, members of a VCoP change their purpose when using the VCoP. As example, the purpose of a newcomer are to learn the basics of a theme using the VCoP. When this newcomer becomes an expert his purpose might be to research specific and complex

aspects of such theme; or even more, his purpose might be to answer questions of new comers.

Therefore, our hypothesis is that study of evolutionary nature of purpose of VCoP's members is a vital way to understand a community and to enhance it. Until today there aren't significant studies to evaluate this evolution (as we have shown in section 2), and fulfillment of purpose through time. Also, we strongly support the idea that the study of social networks members' relations alone (SNA) it is not sufficient to obtain a good understanding of a VCoP. It is needed to analyze other social aspect, such as purpose, to fully understand and perform the necessary enhancements to allow VCoP exists through time.

3.1 Goals as a Purpose Accomplishment Measure

Of course, the key aspect of this work is to consider the purpose evolution analysis as an important tool; but, how can we measure purpose? Since, purpose is something close to the ideas, or underlying motivations of every member, it is not simple to answer this question.

Since, purpose is, from dictionary, "what something is used for". We propose to use goals as a measure of purpose accomplishment. Using this idea, we can measure if a VCoP fulfills a purpose, measuring how well members' contributions accomplish a set of goals previously defined by the owners, managers or experts of the community.

Goals definition must be performed based on interviews or surveys to community experts or administrators ([2,9,10]). Definition of goals consist of a series of phrases. These phrases respond to the question "what the community is for?". This paper only evaluates the goals from community administrators' view point. Although, same process apply for community members' goals analysis.

Afterwards, we need to select a classification or clustering algorithm in order to perform a text mining algorithm to find interesting patterns. It is expected patterns found provide useful information for administrators and experts in order to decide how to enhance the community (add new forums, erase forums, find trends, etc) based on goals fulfillment.

We selected a concept-based text mining since the adaption from concepts to goals is straightforward. This approach will be explained in next section.

As a last consideration, this technique allows us to study the goals' fulfillment through time. Therefore, we can show interesting information of how the VCoP's purpose evolves. Thus, providing useful and objective information to community experts. Without this tool, they only have an intuition on how community has evolved and if the information contained in the community forums is truly accomplishing the purpose of the community.

3.2 Concept-Based Web Text Mining

Concept-based text mining is a data mining approach based on fuzzy sets and fuzzy logic theory.

We based our work in Loh's work presented in [2] and [8]. Loh's proposal is to use a *fuzzy reasoning* model to decide wether a concept is expressed by a web document or not. This way, after the application of the reasoning model, we have classified all documents by its concepts. To do so, we compute the degree of possibility that a concept is related to a web document. In our case, web documents are posts in a VCoP's forum.

This work use goals as a way to evaluate the purpose of VCoP. Therefore, how to introduce goals in the mining process is a key issue. This is why we use the concept based web text mining, this approach allows us to use concepts as goals, then based on the goals definition, the algorithm can classify web documents by the accomplishment of such goals. In the following we will talk about goals instead of concepts.

Fuzzy logic for goals classification. In fuzzy logic, linguistic variables (LV) are not numbers but words or sentences in natural language. These variables are more complex but less precise. Let u be a LV, we can obtain a set of terms $T(u)$ which cover its universe of discourse U. e.g. $T(taste) = \{sweet, salad, acid, bitter, bittersweet\}$

In order to use LV for goals classification, we assume that a community posts can be represented as a fuzzy relation $[Goals \times Posts]$ also called $[G \times P]$. Which is a matrix where each row is a goal and every column is a post in the VCoP. To obtain such matrix we can rewrite this relation in a more convenient manner in Eq.(1) ([2]). In this expression we call "Terms" the words that can be used to define a concept and we write "WP" to refer any word inside a Web Page. In Eq.(1) the symbols "$\times$" and "$\otimes$" represent the fuzzy relation and fuzzy composition respectively.

$$[Goals \times Posts] = [Goals \times Terms] \otimes [Terms \times Posts] \tag{1}$$

As defined above, let P the total amount of posts in a VCoP and W the total number of different words among all of these posts, G the total number of goals defined for the community in study. Then we can characterize the matrix $[G \times P]$ by its membership function shown in Eq.(2), where $\mu_{G\times P} = \mu_{G\times T\otimes T\times P}$ represents the membership function of the fuzzy composition in Eq.(1). The membership values are between 0 and 1.

$$\mu_{G\times P}(x,z) = \begin{pmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,P} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,P} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{G,1} & \mu_{G,2} & \cdots & \mu_{G,P} \end{pmatrix} \tag{2}$$

The composition of fuzzy relations is performed using Nakanishi's fuzzy compositional rule Eq.(3). In Eq.(3), let $Q(U,V)$ and $Z(V,W)$ be two fuzzy relations which share a common set V. Let $\mu_Q(x,r)$ with $x \in U \wedge r \in V$ and $\mu_z(r,y)$ with $r \in V \wedge y \in W$ membership functions for Q and Z respectively. Then we can write the compositional rule as shown in Eq.(3). Where $\bigvee$ is the limited Sum $= min(1, x + r)$ and $\wedge$ is the algebraic product $= (x * r)$.

$$\mu_{Q \circ Z} = \bigvee \{\mu_Q(x, r) \wedge \mu_Z(r, y)\} \tag{3}$$

There are several alternatives to perform the fuzzy composition, [3] performed a study between six different reasoning models. One important issue that must be considered is that even if some terms are not present in a post, the degree of that post to express a specific goal should not suffer alterations. This is a reason to use Nakanishi's compositional rule. However, other rules could be used as well.

Identification and Definition of Goals. In order to apply the above proposal, we need to begin identifying the relevant goals for the study. To do so, we make use of community administrator' knowledge whom identify which are the most interesting goals to describe the VCoP's purpose. Subsequently, every goal is represented as a list of terms (assuming that a goal is a LV). We used synonyms, quasi-synonyms, antonyms,etc. using also the administrators.

We realize that several important terms are produced by slang words. For example, the word *transformator* in spanish "transformador" is used as "transformador", "trafo" or "transf" very commonly. Thus, human definition is useful to enhance goals definitions.

Afterwards, we need to define the membership values for the fuzzy relations $[Goals \times Terms]$ and $[Terms \times Posts]$. We used relative frequency of terms in a community post to represent the membership values of matrix $[Terms \times Posts]$.

More difficult is to define $[Goals \times Terms]$ values. We performed this task by asking the community experts to assign these values. For every goal which is the degree that a term has to represent that specific goal. To do so, he compared two terms each time and gave a value between 0 and 1. For example, a synonym can receive a value near 1; a quasi-synonym, may receive a value near between 0.75 and 0.95; an antonym can be set to 0, etc. This method is an indirect method with three experts.

Finally, we obtained the fuzzy relation $\mu_{G \times P}(x, z)$ applying Eq.(3). In Table 1 we present a column of matrix $\mu_{G \times P}(x, z)$, which represents the goals classification for post 4235.html from VCoP. From this Table we can say that post 4325.*html* have a strong relation with the goal 1 and goal 2, almost no relation with goals 3, 4 and 5.

Table 1. List of goals and membership values to represent post "4235.htm"

Goals	$\mu_{G \times P}$
Goal 1	0.88
Goal 2	0.72
Goal 3	0
Goal 4	0.12
Goal 5	0.01
....	

4 A Real Application

We performed the experiments into the web site of plexilandia.cl virtual community. Based on interviews to administrators of this VCoP and a preliminary study of community activity we describe the community used.

4.1 The Community

Plexilandia is a VCoP formed by a group of people who have met towards the building of music effects, amplifiers and audio equipment (like "Do it yourself" style). In the beginning was born as a community for share common experiences in the construction of plexies[1]. Today, plexilandia count more than 2000 members in more than 6 years of existence. All these years they have been shearing and discussing their knowledge about building their own plexies, effects. Besides, there are other related topics such as luthier, professional audio, buy/sell parts.

Although, they have a web page with basic information of community, most of their members' interactions are produced by the discussion forum.

In the beginning the administration task was performed by only one member. Today, this task is performed by several administrators (in 2008 they count with 5 administrators). In fact, the amount of information generated weekly makes impossible to let the administration task in just one admin.

During six years of life, this community has undergone a great sustained growth in members' contributions . The vision of administrators and experts about the community is based mostly by experience and time participating in the community. They also have some basic and global measures. For example, total number of posts, connected members, etc. However, the don't have: members browsing behavior information, members publications' quality and how they contribute to purpose of the community.

4.2 Concept-Based Text Mining Application

First, we selected data from october 2002 to june 2008 and we perform text preprocessing the text to eliminate HTML, Javascript and other programed codes.

In order to apply concept-based text mining approach we need to define concepts or in our case, goals. We defined six different goals with the help of community experts. These where used as input in the concept-based text mining approach. The algorithm took less than five minutes to finish the classification process.

4.3 Analysis of Results

Results obtained where included in a web report. This report allows administrators to understand how the purpose of plexilandia evolves. Report includes

[1] "Plexi" is the nickname given to Marshall amp heads model 1959 that have the clear perspex (a.k.a plexiglass) fascia to the control panel with a gold backing sheet showing through as opposed to the metal plates of the later models.

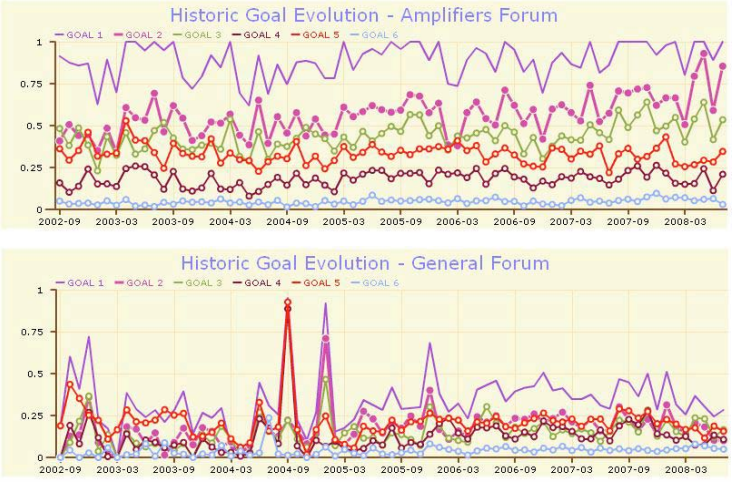

Fig. 1. Goals evolution of two forums

a graph for each forum in the community. This graph represents all goals with a different color. Then, every goal is expressed by its membership value, which means how close is the forum respect of the goals. This can be interpreted as degree of accomplishment of a goal by the forum. If a goal has a value near 1, it indicates that the posts in that forum contributes to the accomplishment of that goal. On the other hand, a value near 0, means that goal doesn't help to accomplish such goal.

For example, in Figure 1 it is possible to observe the forum of "amplifiers" and the level of accomplishment of every defined goal. It is possible to observe that: (i) this form strongly support goal 1, (ii) in last months this forum has experienced a trend to growth in the accomplishment of goal 2, and (iii) the accomplishment of other goals is much lesser. This analysis allows to discover an important conclusion from administration perspective: historically, in "amplifiers" forum the main topic was amplifiers; however, last months there is a trend to talk about related topics, such as music effects (like guitar effects). Therefore, this analysis is an objective tool to show that situation.

Moreover, the report allows to identify certain anomalies, such as, estrange peaks. Administrators, study these peaks and they found perfect sense with particular situations that happened in the forums those months.

4.4 Results' Evaluation

The main objective of these section is to evaluate the validity of experimental results. To do so, we performed interviews to community administrators who have analyzed and validated obtained results.

The importance of this evaluation is based in experts or administrators knowledge and experience about plexies and the community through 6 years. Therefore, they can validate the results but also they can quantify if a result is

expected (just by intuition) or they gain an important peace of new knowledge to understand the community.

In addition, it has been applied an usability evaluation of the generated report. This was performed to determine the administrators' satisfaction level at the moment of reading the report.

Both evaluations were applied to 3 from 5 administrators. One of them is the community founder of plexilandia.

Usability Evaluation. To perform this evaluation, community administrators had to answer a survey. We asked the level of satisfaction with the generated report. In the survey, it was measured: ease of use, ease of learn, need of help and report clarity.

Survey results show that report is easy to read and learn but requires of help when reading for first time. This is also related with administrators' previous experience using web reports (only one of them had previous experience).

Since, this is the first report generated and it is oriented only to administrators, we only focused the evaluation in the satisfaction level achieved. In the near future we expect to measure reports' efficiency and efficacy. Then, we pretend to publish results for community members.

Validity of Results. Community administrators had to quantify each result obtained and showed in the report. This is the quantification of anomalies in previous section and a quantification of identified goals evolution.

We used a three points evaluation scale: (1) not expected result; complete surprise, without a clear cause; (2) not expected result, but cause would be known; (3)expected result, known cause.

From 14 anomalies detected, 8 were expected results, and 3 represent situations completely unexpected. From the 30 identifies behavior patterns, 20 represent expected results, and only 2 are surprise situations completely.

5 Conclusion

This work has shown that community administrators are almost blind when enhancing a social network. Moreover, we have also shown that common analysis based just on social network relations are not enough.

We have proved that using the community experts or administrators combined with a data mining approach could provide much more objective and rich information. Which may be used to enhance the virtual community of practice.

Besides, we have proposed the analysis of purpose evolution of a VCoP based on goals definition as the key to the application of data mining analysis into the analysis of VCoP. This way, community experts or administrators count with objective information.

We have successfully used the purpose evolution analysis in a real VCoP with more than 2000 members and 6 years.

We think that although results are promising, it is needed more work in this direction, to find modern tools to help managers, experts or administrators to enhance their communities.

References

1. Koh, J., Kim, Y., Butler, B., Bock, G.: Encouraging participation in virtual communities. Communications of the ACM (January 2007)
2. Loh, S., de Oliveira, J., Gameiro, M.: Knowledge discovery in texts for constructing decision support systems. Applied Intelligence (December 2003)
3. Nakanishi, H., Turksen, I., Sugeno, M.: A review and comparison of six reasoning methods. Fuzzy Sets and Systems (January 1993)
4. Pal, S., Talwar, V., Mitra, P.: Web mining in soft computing framework: relevance, state of the art and future directions. Neural Networks (December 2002)
5. Perkowitz, M., Etzioni, O.: Towards adaptive web sites: Conceptual framework and case study. Artificial Intelligence (December 2000)
6. Preece, J.: Etiquette, empathy and trust in communities of practice: Stepping-stones to social capital. Journal of Universal Computer Science (January 2004)
7. Preece, J., Maloney-Krichmar, D.: Online communities: Focusing on sociability and usability. In: Handbook of Human-Computer Interaction (January 2003)
8. Ríos, S.A.: A study on web mining techniques for off-line enhancements of web sites. Ph.D Thesis, p. 231 (September 2007)
9. Ríos, S.A., Velásquez, J.D.: Semantic web usage mining by a concept-based approach for off-line web site enhancements. In: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (2008)
10. Ríos, S.A., Velásquez, J.D., Yasuda, H., Aoki, T.: Using a self organizing feature map for extracting representative web pages from a web site. International Journal of Computational Intelligence Research (IJCIR) 2, 159–167 (2006)
11. Shummer, T.: Patterns for building communities in collaborative systems. In: Proceedings of the 9th European Conference on Pattern Languages and Programs (2004)
12. Wellman, B., Salaff, J., Dimitrova, D., Garton, L.: Computer networks as social networks: Collaborative work, telework, and virtual community. Annual Reviews in Sociology (January 1996)
13. Wenger, E.: Communities of practice: Learning, meaning, and identity (January 1999), `books.google.com`
14. Wenger, E., McDermott, R., Snyder, W.: Cultivating communities of practice: A guide to managing knowledge. Harvard Business School Press (2002)