December 2005

Discovering Users' Participant Roles in Virtual Communities with the Help of Social Interaction Theories

Cheng Zhang
School of Management, Fudan University

Chenghong Zhang
School of Management, Fudan University

Discovering Users' Participant Roles in Virtual Communities with the Help of Social Interaction Theories

Zhang, Cheng
Department of IM & IS
School of Management, Fudan University
zhangche@fudan.edu.cn

Zhang, Chenghong
Department of IM & IS
School of Management, Fudan University
chzhang@fudan.edu.cn

Abstract

In today's electronic businesses, many companies apply virtual communities (VC) to support their electronic customer-relation management (e-CRM). Since participants play different roles on the activeness and success of a VC, it is valuable for companies to treat these participants differently. Considering the situation of large amount of VCs with great population of participant base, companies need some objective and effective methods to distinguish different roles of participants in a VC based on finite online observable behaviors. These roles do not simply mean posting more or less, but rather a compound summary of individual behaviors and social patterns. Integrating with knowledge in social interaction study, such as IPA (Interaction Process Analysis), we apply several quantitative methods, including classification algorithms, clustering algorithms and neural networks, to analyze participants' roles and social influence in VCs. Results demonstrate how those methods can distinguish different groups of participants from individual behaviors and social patterns in VCs. This study also provides a new trial on how advanced information technologies can efficiently understand social factors in VCs with the help of social interaction theories

Keywords: Virtual Community, Social Interaction Theory, Machine Learning

1. Introduction

Computer-mediated social groups (Rheingold 2000) in which members build and maintain their inter-personal social relationships are named as virtual community (VC, e.g., Koh and Kim 2001). In today's businesses, many companies provide vendor-supported VC as part of their e-CRM (customer-relation management) strategy. For example, Microsoft Inc. maintains worldwide newsgroups and various online knowledge bases for users and developers to seek information and to share knowledge. eBay Inc. builds a large online community to broadcast company information and to encourage its members discussing topics of mutual interests, sharing useful information, and making friends in this virtual society. These successful cases show that active VCs greatly help companies provide more efficient service and increase users' loyalty (Reichheld and Schefter 2000). Such benefit increases as customer base increases.

To develop and maintain active and successive VCs, several important factors should consider. Among them, participant is the foundational factor of a community. Participants are influenced by each other unequally (Carlson and Buskist 1997) in a community. Individuals who can significantly influence others in a community play the important role on community's effectiveness and successiveness. Borrowing the concept of mass customization from market research (e.g., Fulkerson and Shank 1999), it is valuable for companies to segment the participants and treat them differently in order to "focus on the right customers" (Reichheld and Schefter 2000). For instance, trust and feedback usually come from active members; novices need more encouragement; reward can give to active members to reinforce their behaviors and to motivate others to follow.

Since participants play different roles on the activeness and success of a VC, it is valuable for companies to treat these participants differently. Considering the situation of large amount of VCs with great population of participant base, companies need some objective and effective methods to distinguish different roles of participants in a VC based on finite online observable behaviors. These roles do not simply mean posting more or less, but rather a compound summary of individual behaviors and social patterns. By integrating with knowledge in social interaction study, such as IPA (Interaction Process Analysis), this study applies several information technologies to analyze participants' roles and social influence in VCs. The aim is to answer two important questions in VCs: could individuals be objectively distinguished into groups that are significantly different from each other based on their observable behavioral and social patterns in a VC? And if so, could the task be automatically carried out with minimum human intervention to meet the practical need?

2. Social Interactions and Participant Roles

Interactions among participants form up a social network in VCs. Since active group contributes most of postings, the network structure in a VC is usually a start type, in which most social interactions are connected with a few participants at one end. The strength of such connection can be measured by the frequency of interactions and the extent of emotional closeness (Granovetter 1973). As participants reply to others' postings and receive responses, social interactions in between naturally enhance. In another world, they become more familiar with, or aware of, each other. As a result, they become more influenced by those they interact with (Carlson and Buskist 1997). Moreover the behavioral and social patterns of a participant will determine his role in VC, i.e. how the others in the community and himself view who he is /which he belong to (Turner 1985).

Lack of exact and clear definitions of various roles in VCs in research literatures, this study cannot determine the number of roles that a VC must have. In addition, studied behavioral roles in physical world may not be applied directly into VCs because of phenomena like online infidelity (e.g. Mileham 2004). Therefore, this study focuses to classify two basic and primitive roles: active group vs. non-active group. With theoretical support on more detailed classification of roles in VCs, the members are able to be clustering into more representative groups in the future. Note that in this study, "activeness" does not simply mean posting more, but rather a compound summary of individual behaviors and social patterns. Because of the variance of leadership efficiency (Hogan et al. 1994) according to the situation (Strang 2004), we do not restrict these active participants as leaders. However, we believe the leadership exists inside this group.

3. Behavioral and Social Measurements

In VCs, individual behaviors can be mainly described as follow: one logs into a community, browses and posts messages to share his idea with others, stamped with his identification information or anonymously, and then quits. Therefore, posting messages in a VC is individual main behavior. Much valuable information hides behind the posting activity: the frequency of individual's posting and the length of posted message represent individual's immersion in a VC: a higher posting frequency and a higher posting length means greater participation in the community and greater interest in topics; the time of a message posted indicates when individual enters the community; and identification information in a message, such as nickname, email address and internet protocol (IP) address, usually represents a unique person; New topics ("threads") and replying postings in a community usually represent two different behaviors respectively: raising questions and attempts to solve them; communications via posting create social interactions (Banyard and Grayson 2000) among individuals. The repeat and the extent of interpersonal communications via posting reinforce social interactions: More interactions and greater extent usually represents a closer relation among communicators. (Strang 2004) also argued that repeated behaviors could improve leader effectiveness from project management perspective. Information mentioned above forms up behavioral and social patterns in VCs. In this way, individual can be quantified as a set of values assigned to these patterns.

To develop measurement instruments, three factors are considered: participation rate, communication content and communication network.

Participation rate is a quantitative measurement on individual behavior and contribution to a community. It has been consistently concluded that the ones talk the most in a group are likely to become leaders (e.g., Stein and Heller 1983) and contribute more to communities (Hare et al. 1996). Overall, eight indexes on behavioral patterns are developed to measure participant rate: the number and length of postings in a period; the number and length of starting a new topic in a period; the number and length of posting replies; and the number and length of being-replied postings (response). These measurements are quite common in VC survey studies (e.g., Koh and Kim 2001). The length here is counted by postings' word count.

Then we consider communication content. In VCs, verbal content is one most important factor in analysis. To classify social interactions based on postings in VCs, a method that can describe various behavioral patterns, i.e. the "meaning of meaning", is required. One well-developed system for this purpose is IPA (Interaction Process Analysis, Bales 1950) which divides interactions into four main categories: questions, attempted answers, positive reactions and negative reactions. In this way, individual behaviors and their social interactions can be understood from two perspectives: task activities and socio-emotional activities, regardless of any detailed contents of messages. Although there are three sub categories under each of four categories, we believe IPA's first-level categories have provided a clear enough map on classifying social interactions. Besides, analyzing at the first level of categories will greatly reduce the chance of misclassifying the social interactions. Measurements include the number of questions, attempted answers/information, positive reactions and negative reactions in postings. Because individuals communicate by messages in literal VCs like forums, newsgroups and emails, it is reasonable to conduct IPA on messages, to see whether an individual performs a particular social interaction with the other in postings.

Finally we consider the factor of communication network. As we discussed before, individual behaviors determine his social patterns in a VC. When individual communicates more with others, greater breadth of social relations probably creates. Because of high participant, he probably owns greater social influence of reciprocity on others and more participants like to respond his postings. Furthermore, the repeat of communications in pairs is an observable and basic measurement for the social interaction level: As two communicate more times with more content, they become more familiar with each other. Therefore, we develop $6*(n-1)$ measurements on social patterns for each individual where n is the number of participants in a VC: the number and length of postings between individual m and $n-1$ others; and the number of question/answer/positive reaction/negative reaction of postings between individual m and $n-1$ others. These measurements present individual social interactions in pairs. All social patterns form up a communication network as shown in Figure1. Based on the observation that a few members contribute most of postings in a VC (Brenner 1998), the network structure is expected to be centralized, in which most social interactions are connected with a few participants at one end.

Figure 1 goes here

Figure 1: communication network in a VC. Each dot at the end of a line stands for an individual and the black dot line represents VC's range. V_m^n *is the set of (n-1) pairs of interactions the individual m owns.* $w_{m,j}$ *is the weight, measured by repeat times and extent of communications in pairs between m and j. Black real line stands for social interactions of surface contact or mutuality. Blue dot line shows a range of active group or potential leaders.*

Social patterns generate as participants interact with others. Table 1 summarizes all measurements discussed above and shows the relation between individual behavioral patterns and social patterns.

Table 1 goes here

Table1: The summary of behavioral and social pattern measurements.

As we have developed *12+6*(n-1)* attributes to describe behavioral and social patterns in a VC, individual can be viewed as a fix-attribute set with different assigned values. Therefore, V_m^n represents a unique individual m from his multi-dimension patterns. By applying such a *12+6*(n-1)*-dimension matrix, the question of classifying participants in a community transfers to cluster them based on the similarity (or "distance") among their pattern values. If a multi-dimension hyperplane exists, i.e. participants inside hyperplane share similar patterns but significantly differ from outside, we can declare that there are different types of participants in a VC.

To validate clustering result, we will evaluate how postings from different types of participants influence the whole community postings. Because of possible delay for messages to reach participants (they may not visit VC every day), social influence may also be delayed. Since postings will lose their effect/attractiveness as time passes and three days should be long enough for most participants to respond, we evaluate the influence in three time segments separately: the situation of no-delay, one-day delay and two-day delay.

4. Research Methods

In this study, we import several machine learning technologies on classification, clustering and forecasting participants. These quantitative techniques are developed in artificial intelligence (AI) as tools to learn variable relations with build-in adapting ability to seek for better learning. Their significant advantages, such as independence of assumptions on data properties and accuracy on modeling nonlinear data patterns (Smith and Gupta 2000), promise themselves as useful and objective tools to understand individual complex behaviors in VCs. Depending on them, we can discover knowledge on participants' role and influence in VCs. Another advantage is that these methods, with high automation and good learning ability, can meet practical requirement of measuring a large number of VCs with huge custom base.

4.1. Classification Algorithm

To analyze messages' content posted in the newsgroup and classify them into IPA categories, classification algorithm is implemented in this study. As SVM has been proved to work well for text categorization, e.g., (Joachims 1998 and 1999a), we choose it to perform content classification in this study.

Before classification, postings should be transformed into valued features set. Paice-Husk stemming algorithm (Paice 1990) is used and the "stop-words" list is from CMU Text Learning Group at http://www.lb.cs.cmu.edu/~TextLearning/eriks-code/code.html. After stemming, each word corresponds to a feature, valued with the number of times it appears in the message.

Although emotional words and symbols are usually ignored in previous classifying research, they are useful to understand the social interactions in this study. Therefore one modification in this study is to keep words that may express writer's emotions, such as exclamatory mark, question mark, and thankful words. To avoid overlarge features, only the word occurring at least in two different postings in samples is chosen as the feature. Overall, a few hundreds of word features are chosen from the sample. Another modification on feature selection is to consider structural information, i.e. posting types and positions in the newsgroup. Therefore, we add two external features about the posting: whether it is the beginning of a thread of discussion or in the middle; and how many respond followed it. Finally, to improve performance, the features in the postings are normalized and the feature dimensions are scaled by IDF (inverse document frequency) approach. The classification task is carried out by SVM-Light (Joachims 1999b).

4.2. Clustering Algorithm

In our work, a high-dimensional data set which describes participants' behaviors and social interactions are used to recognize different types of participants in VCs. Cluto (Karypis 2002), a soft package implementing K-means clustering algorithm (Hartigan and Wong, 1979) at http://www-users.cs.umn.edu/~karypis/cluto/index.html, is used for this study.

4.3. Neural Network and Random Walk Model

Another applied technology is neural networks (NN), defined as mathematical models developed through learning from data. We use multilayered back-propagation NN (MBP-NN, Rumelhart et al. 1986) to analyze the impact of different individual behaviors in a VC, while random walk model is used for validation.

In our case, if relations between active participant postings and VC outcome are found by MBP-NN, we can validate clustering result and prove social influence in a VC. Otherwise, VC outcome should follow random walks. For instance, postings from potential leaders may motivate more feedback and consequently booms VC activeness, i.e. more postings are expected in the following days. If such interaction does not occur, the outcome of VC should show as random. Out of samples, 60% (N1) is used for training while 40% is for testing (N2). In this study, Neural Network Toolkit in MatLab 6.1 (The MathWorks, Inc) is used.

To eliminate the randomness from the initial seed/weight, experiments of both NN and random walk are replicated for 15 times, each with a different random seed/different set of weights. The average performance of 15 times replication is used as the estimated value of the system performance. The significance of difference is tested via ANOVA and then multiple comparison procedure is carried out based on Levene (Levene 1960)'s test of homogeneity of variances.

There is no consensus about how to evaluate the precise of a relation model, but two basic measures are employed in assessing the predictive power of our model: root mean square error (RMSE), and the sign of directional change (SIG).

In summary, classification algorithms is used for IPA in message content analysis, clustering algorithms is used for recognizing different roles of participants in VCs, and neural networks (NN) is used for validating the participants' influence on VC total postings.

5. Data Analysis

We randomly chose two newsgroups of different size from Microsoft newsgroups list, following two basic rules: first, at least eight-week data was available for longitudinal study purpose; second, the newsgroup should contain at least *100 (1000)* postings from different members during eight weeks to ensure the subject is not "dead". Because Microsoft newsgroups are to support product/developing, our VC subjects are pre-determined as a product-oriented VC mainly for individuals with the same interest in programming or software usage. In summary, we got a *187*-day data from "microsoft.public.fox.vfp.dbc", in which *229* participants contributed to *918* postings, starting from 05/03/2002 to 11/05/2002, and the other *114*-day postings from "microsoft.public.sqlserver.setup", in which *149*7 participants contributed to *3700* postings, starting from 08/05/2002 to 11/26/2002. Participants in newsgroups were unaware of this research. In the remaining part of this paper, "microsoft.public.fox.vfp.dbc" was VC1 and "microsoft.public.sqlserver.setup" was VC2.

First of all, *186* messages from VC1 and 204 from VC2 were randomly chosen for classification learning. Out of samples, *60%* postings were assigned to training set while the remaining *40%* were assigned to test group. Table 2 showed the IPA accuracy of samples' contents. This learning ability was acceptable to understand the content of the remaining postings, considering the estimated *85%* precision of human classification upper limit (Sundheim 1992) and the reported *4%-14%* error in manual classification tasks (Marsh and Perzanowski 1998). Note that there was a high accuracy of recognizing negative content in samples. We noticed that only *4% (1%)* of training samples in VC1 (VC2) contained negative expression and it was a similar situation in testing samples. We inferred that participants in chosen newsgroups were friendly and avoided negative emotions. Therefore, the negative measurement might not be useful for clustering in this study.

Table 2 goes here

Table 2: IPA accuracy of SVM learning in two newsgroups. Question / Answer / Positive / Negative were four categories in IPA. The accuracy was counted as the percentage of correctly classified.

After using SVM to recognize the content of messages, *12+6*(n-1)* behavioral and social patterns of each individual were used for clustering without manual intervention. Participants in two newsgroups were categorized into two sets: Set1 contained *51.5%* participants in VC1 (*57.7%* in VC2), while Set2 contained the remaining. Although Set2 was smaller, it contributed to *66.8%* (*73.6%*) of total posting number and *75.7%* (*84.2%*) of total posting length in VC1 (VC2).

This result gave us a brief map of how these two set of people were different from each other. Among these postings, Set1 was more active in posting new thread: it began *87.6%* (*70.0%*) of total threads that covered *82.9%* (*69.7%*) of total length of new topics, indicating that they might more like asking questions. On the contrary, the remaining set contributed more on replying, with a ratio of *82.0%* (*98.8%*) of total replies numbers and *82.4%* (*99.6%*) of total replies length, indicating that this set of participants tended to follow others' thread than to create new thread. We inferred that this small group put more effort in answering others' question, or help others, than consulting the others. Furthermore, Set2's posting got more responses in the community. For the moment, we named Set1 as non-active group and Set2 as active.

The result of IPA also showed different characteristics of social interactions between two groups. As shown in Table 4, participants in the non-active group averagely asked more for orientations / suggestions / opinions than those in active group, while active group members significantly preferred giving orientations / suggestions / opinions than others did. In other words, non-active group liked seeking for helps while the active group liked offering helps. Furthermore, it was clear that non-active participants expressed more positive emotions than active participants: As one consulted more with others, he would like to express more thanks simultaneously. Since negative emotions were rare in these VCs, they were not usable for this study.

Table 3 goes here

Table 3: Groups' social interactions in VCs. Numbers in cells were average values of individual social interactions in VCs. The numbers with grey background were the result from VC1 while the ones in bracket were the results from VC2.

After getting a brief idea of different behaviors from two groups, we analyzed groups' social structures. The result was shown in Table 5. Participants in active group interacted with others in the VC for *10.12* (*6.08*) times averagely during the period and these communications totally involved *83.4%* (*83.0%*) participants in the VC. The non-active group members interacted for only *4.77* (*1.12*) times and their communications only covered *41.5%* (*13.9%*) members. This result briefly indicated that the active group contained broader social network in the VC than the non-active group did. Furthermore, the communications of active group were not only limited inside the group (or clique): as shown in Table 5, their communications covered most of the participants in the whole VC, while the participants in the non-active group communicated more with active group members rather than with ones from the same group. In other words, the social network among active participants tended to be a "star" that connected to non-active participants. Comparing communication content between the groups, we confirm that the non-active participants liked asking questions and expressing positive emotions, while the active participants offered more people attempted answers and information.

Table 4 goes here

Table 4: Groups' characteristics of social interactions in a VC. # of Interactions was the average communications with another in a VC. Intra (Inter)-group interaction breadth was to answer the questions: how many percentages of people did participants interact with in the same group (or from the other group). Questions/ Answers/ Positive Emotions Breadth were the percentages of people in a VC that different groups interacted with Question/ Answers/

Emotions. The numbers with grey background were the result from VC1 while the ones in bracket were the results from VC2.

In summary, these two groups were significantly different from each other by measuring the degree of participating rate, social interactions and communication network.

Based on clustering result, the analysis of social influence on VC postings from different groups was carried out by MBP-NN and then the analytical result was compared with random walk model. The result in Table 6 showed that when considering participants' postings, the VC's outcome was more accurate to predict. In other words, this result validated the existing of social influences from members in VCs. Similarly, behaviors from those two groups improved accuracy on the directional change of community postings, i.e. whether more or less people would participate in the community. As we mentioned before, a few participants contained more social resources in VCs and their behaviors would motivate greater feedback than others did, i.e. motivating others to follow. Since we had examined that the non-active group tended to ask more questions or to seek more help while the active group liked replying or offering helps, behaviors from the non-active group might also play an important role on VC postings in total, i.e. motivating others to answer.

Table 5 goes here

Table 5: Groups' influence on VC postings, measured by RMSE and SIG. (*): with 5% significance in ANOVA and multiple comparison procedure. Numbers in the cells under RMSE column were the root mean square error of predicting the numbers of postings in the future. Numbers in the cells under SIG column were the accurate rate of predicting the changes of community activity. The numbers with grey background were the results from VC1 while the ones in bracket were the results from VC2.

6. Conclusions and Discussions

In today's electronic businesses, many companies apply virtual communities (VC) to support their electronic customer-relation management (e-CRM). Since participants play different roles on the activeness and success of a VC, it is valuable for companies to treat these participants differently. Considering the situation of large amount of VCs with great population of participant base, companies need some objective and effective methods to distinguish different roles of participants in a VC based on finite online observable behaviors. These roles do not simply mean posting more or less, but rather a compound summary of individual behaviors and social patterns.

In this study, we answered two important questions in VCs: could individuals be objectively distinguished into groups that are significantly different from each other based on their observable behavioral and social patterns in a VC? And if so, could the task be automatically carried out with minimum human intervention to meet the practical need? Integrating with knowledge in social interaction study, such as IPA (Interaction Process Analysis), we applied several information technologies to analyze participants' roles and social influence in VCs.

First, we developed a behavioral and social pattern matrix to understand individuals in VCs; second, we used this matrix to distinguish participant roles in VCs; third, we imported classification, clustering and neural network algorithm into VC research. The results demonstrated how information technologies could do this work and how different participant roles of members, such as leaders and followers, could be acquired in an objective and automatic way. These roles did not simply mean posting more or less, but rather a compound summary of individual behaviors and social patterns. Based on this knowledge, the organizations can develop their e-loyalty efficiently by “focusing on the right customers” (Reichheld and Schefter 2000) in VCs.

This study provided a new trial on how advanced information technologies, integrating with knowledge in social interaction, such as IPA, can efficiently understand social factors in VCs. Because of the high degree of automation and unsupervised learning ability, quantitative methods in AI is an efficient and economic approach to deal with great amount of VCs with large size of member base in practice.

We also conducted experiments with more clusters. The results were encouraging as members were categorized into groups with more “similarity” inside but more difference from each group. Therefore, with theoretical support on more detailed classification of roles in VCs, the members are able to be clustering into more representative groups in the future. However, the unsupervised learning approach is only based on the rule to discover the closest “similarity”, so sometimes it cannot produce the groups that researchers exactly want. In that situation, a few supervisions from human on describing roles will be very helpful. Furthermore, the result from these quantitative approaches can be validated from the result of surveys to test how accurate these advanced information technology can perform in VC studies.

References

Bales, R.F. *Interaction Process Analysis: A Method for the Study of Small Groups*, 1950, Reprinting by Chicago: University of Chicago Press (1976).

Banyard, P., and Grayson, A. *Introducing Psychological Research: Seventy Studies that Shape Psychology*, 2nd edition, New York: Palgrave, 2000.

Brenner, E. “*VC Conference in Bath, England*,” *Information Today* (15:8), 1998, pp. 16-19.

Carlson, N.R., and Buskist, W. *Psychology: the Science of Behavior*, 5th edition, Boston: Allyn and Bacon, 1997.

Fulkerson, B., and Shank, M. “The New Economy Electronic Commerce, and the Rise of Mass Customization,” In: Shaw, M. et al., *Handbook on Electronic Commerce*, New York: Springer Verlag, 1999, pp.411-430.

Granovetter, M. “The Strength of Weak Ties,” *American Journal of Sociology* (78:6), 1973, pp.1360-1380.

Hare, A.P., Blumberg, H.H., Davies, M.F., and Kent, M.V. *Small Groups: an Introduction*, Westport, Conn: Praeger, 1996.

Hogan, R., Curphy, G.J., and Hogan, J. “What We Know about Leadership: Effectiveness and Personality,” *American Psychologist* (49), 1994, pp.493-504.

Joachims, T. “Text Categorization with Support Vector Machines: Learning with Many Relevant Features,” In *Proceedings of the 10th European Conference on Machine Learning*, Springer, 1998, pp.17-142.

Joachims, T. “Transductive Inference for Text Classification Using Support Vector Machines,” In *Proceedings of 16th International Conference on Machine Learning*, Morgan Kaufmann, San Fran-cisco, CA, 1999a, pp. 200–209.

Joachims, T. “Making large-Scale SVM Learning Practical,” *Advances in Kernel Methods - Support Vector Learning*, Schölkopf et al, edition, MIT-Press, 1999b.

Karypis, G. *CLUTO, A Clustering Toolkit*, Technical Report: #02-017, University of Minnesota, 2002.

Koh, J., and Kim, Y.G. “Sense of Virtual Community: Determinants and the Moderating Role of the Virtual Community Origin,” *Proceedings of the Twenty-Second International Conference on Information Systems*, 2001, pp. 407-410.

Hartigan, J.A., and Wong, M.A., “A K-Means Clustering Algorithm,” *Applied Statistics* (28:1), 1979, pp.100-108.

Levene, H. “Robust Tests for Equality of Variances,” In *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 1960, pp. 278-292.

Levinger, G. “A Three-Level Approach to Attraction: Toward an Understanding of Pair Relatedness,” *Foundations of Interpersonal Attraction*, In Huston T.L., edition, New York: Academic Press, 1974.

Marsh, E., and Perzanowski, D. “Muc7 Evaluation of IE Technology: Overview of Results,” In *Proceedings of Message Understanding Conference*, 1998.

Mileham, B.L.A. “Online infidelity in Internet chat rooms: an ethnographic exploration,” *Computers in Human Behavior*, In Press, Available online April 2004.

Paice, C.D. “Another Stemmer,” *SIGIR Forum* (24:3), 1990, pp.56-61.

Rheingold, H. *The Virtual Community: Homesteading on the Electronic Frontier*. Cambridge, Mass: MIT Press, 2000.

Reichheld, F.F., and Schefter, P. “E-Loyalty: Your Secret Weapon on the Web,” *Harvard Business Review*, July-August, RN: R00410, 2000, pp.105-113.

Rumelhart, D.D., Hinton, G.E., and Williams, R.J. “Learning Representations by Back-Propagating Errors,” *Nature* (323), 1986, pp.533-536.

Smith, K.A., and Gupta, J.N.D. “Neural Networks in Business: Techniques and Applications for the Operations Researcher,” *Computers & Operations Research* (27), 2000, pp.1023-1044.

Stein, R.T., and Heller, T. ”The Relationship of Participation Rates to Leadership Status: A Meta-Analysis,” *Small Groups and Social Interaction*, In: Blumberg, H.H. et al. edition, New York: Wiley, 1983.

Strang, K.D. “Examining Effective Technology Project Leadership Traits and Behaviors,” *Computers in Human Behavior*, In Press, Available online December 2004.

Sundheim, B. “Overview of the Fourth Message Understanding Evaluation and Conference,” In *Proceedings of the Fourth Message Understanding Conference*, 1992, pp. 3-21.

Turner, J.C. “Social Categorization and the Self-Concept: a Social-Cognitive Theory of Group Behavior,” *Advances in Group Processes* (2), 1985, pp.77-121.

Figures

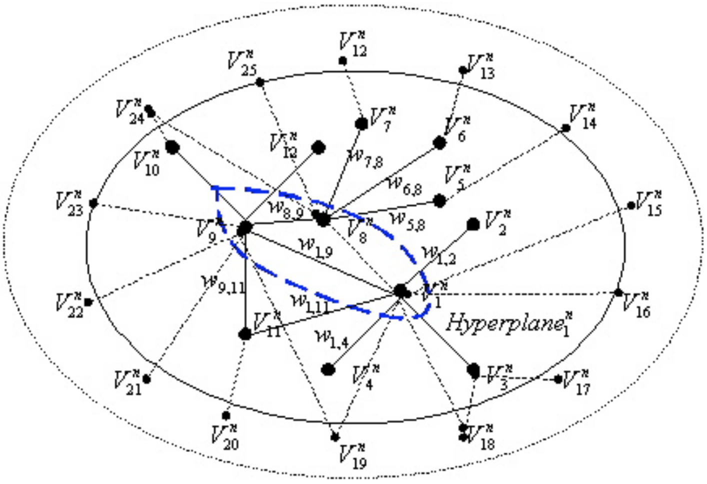

Figure 1: communication network in a VC.

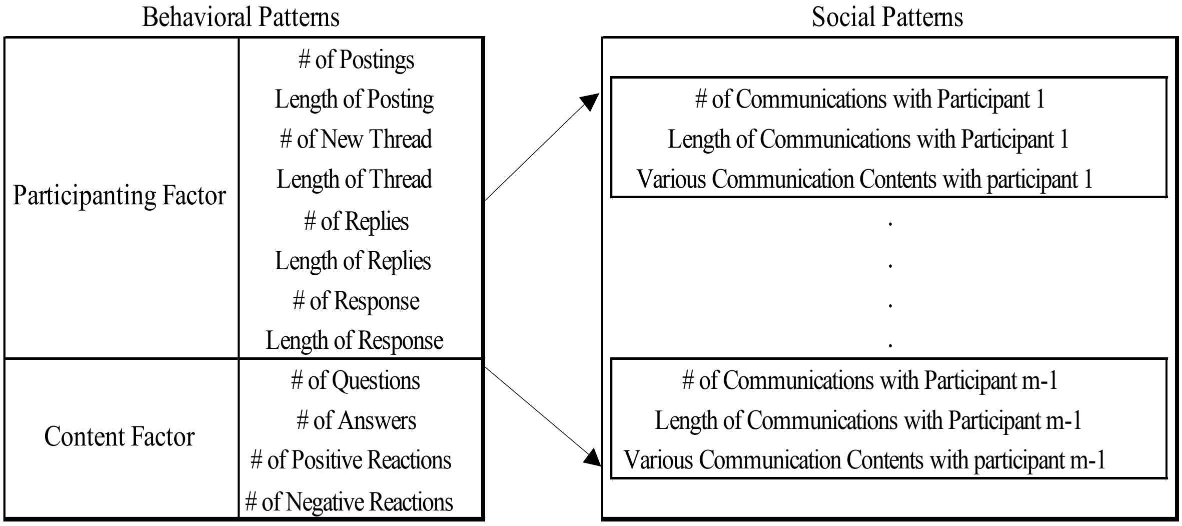

Table1: The summary of behavioral and social pattern measurements.

Accuracy Groups	Question	Answer	Positive	Negative
VC1	85.1%	85.1%	70.3%	95.9% (*)
VC2	86.6%	91.5%	70.7%	100% (*)

Table 2: IPA accuracy of SVM learning in two newsgroups.

Groups	Count	Questions	Attempted Answers	Positive Emotions	Negative Emotions (*)
Non-Active	51.5%	1.59	2.42	1.51	0.03
	(57.7%)	(1.10)	(1.12)	(2.02)	0.00
Active	48.5%	0.41	5.43	0.50	0.03
	(42.3%)	(0.69)	(4.28)	(1.08)	0.00

Table 3: Groups' social interactions in VCs.

Groups	# of Interactions	Interaction Breadth	Intra-group Interaction Breadth	Inter-group Interaction Breadth	Questions Breadth	Attempted Answers Breadth	Positive Emotions Breadth
Non-Active	4.77	41.5%	20.3%	64.0%	34.5%	36.7%	32.3%
	(1.12)	(13.9%)	(1.6%)	(30.6%)	(13.2%)	(13.4%)	(20.1%)
Active	10.12	83.4%	75.7%	90.7%	17.9%	52.4%	16.6%
	(6.08)	(83.0%)	(86.3%)	(80.7%)	(11.05)	(34.3%)	(13.2%)

Table 4: Groups' characteristics of social interactions in a VC.

Influence In VC	RMSE				SIG			
	Active Group (1)	Non-Active group (2)	Random Walk (3)	Comparison	Active Group (1)	Non-Active Group (2)	Random Walk (3)	Comparison
No-Delay	7.09	6.71	6.95	NA	56.4%	55.9%	48.2%	1,2>3 (*)
	(21.43)	21.09	28.88	3>1,2 (*)	64.1%	64.6%	46.2%	1,2>3 (*)
1-Day Delay	7.05	7.45	9.06	3>2>1 (*)	55.0%	53.9%	50.7%	1>3 (*)
	(21.44)	22.02	36.59	3>1,2 (*)	73.5%	71.3%	45.8%	1,2>3 (*)
2-Day Delay	7.33	7.44	9.45	3>1,2 (*)	55.3%	54.5%	48.9%	1,2>3 (*)
	(21.61)	(22.35)	(37.41)	3>1,2 (*)	(68.0%)	(64.8%)	(47.4%)	1,2>3 (*)

Table 5: Groups' influence on VC postings, measured by RMSE and SIG. (*).