Abstract

There are two main research results in this project. First, high cost and uncertainty are problems of marketing. Influential online product reviews are more powerful than firm's advertisements. Our research results showed that our model outperforms two general methods in selecting influential reviewers. Our work can accurately point out which reviewer to be selected to become the virus. In the electronic commerce applications, the search of potential opinion leaders helps to target the right customers, and the marketing model like viral marketing can then be applied. Second, we propose a novel expert recommendation mechanism for knowledge sharing in the online forum communities. Accordingly, our experimental results show that with the support of the proposed recommendation mechanism, the requesters in forum can easily find the similar discussion threads to prevent spamming the same discussion. Besides, if the requesters can't find the qualified discussion threads, this mechanism provides a relatively efficient and active way to find the appropriate experts.

Keywords: Viral Marketing, Social networks computing, Recommendation system, Knowledge management, Electronic commerce

摘要

在這份期中報告裡有兩個主要成果。首先，昂貴的成本和不確定性是在行銷的困難點。我們爲了能找出網路上有力的影響者而設計出一種優異的模型。該模型不僅能幫助我們找出有力的影響者，而且也比目前常見的一般模型準確。在電子商務應用上，則可尋找隱身於社會網路之中的意見領袖，透過病毒行銷模式影響群眾。其次，我們計了一種新的專家推薦系統，可用在網路社群的知識分享。我們的實驗結果顯示出，在所設計的推薦系統的幫忙下，在網路社群中的知識請求端可以很容易找到相似的討論結果，以避免重覆的討論一再地被閱讀。另外，假如知識請求端無法找到合適的討論文章，本機制能提供有效且主動的方式來找到合適的專家。

Keywords: 病毒式行銷, 社會網路計算，推薦系統，知識管理，電子商務

1. Introduction

With the rise of the Internet, human life has changed dramatically, and different types of network applications are immediately generated. The vigorous development of Web 2.0 related applications led to another high tide, and Peer Production has realized the collaborative wisdom dream. In this research, utilizing exiting and created social computing (SC) environment, we exploit social network analysis techniques, artificial intelligence, multi-objective decision-making, information retrieval to develop social network based decision support system (SNDSS). These proposed recommendation mechanisms are further applied to social network based knowledge management (SNKM) and electronic commerce (SNEC). In the knowledge management applications, the knowledge hidden in the social network can be discovered, extracted and shared. In the electronic commerce applications, the search of potential opinion leaders helps to target the right customers, and the marketing model like viral marketing can then be applied.

Nowadays, online forums have become a useful tool for problem solving, learning discussion, and knowledge building. The most important benefit of online forums to individual users is that they can receive tailored answers from peers after formulating the problems in their own words, without using specific keywords to search online. This project proposes an innovative recommendation mechanism, which employs the role analysis, social relation, and semantic analysis to construct a more comprehensive and personalized framework for each users in the online forum space, for both discussion threads and experts in the human-expert knowledge forum. Moreover, we present a Markov Chain model to find the most possible, helpful experts while the user doesn’t find the satisfied threads. In certain knowledge forum, the help-seeker, called a requester, ask a specific problem using natural language. The proposed model applies semantic analysis with semantic expansion to find the relevant discussion threads. Nevertheless, if the searching results do not meet the user’s needs, then the model combines the role analysis and social relation to recommend the most helpful experts and provide the shortest social path (include intermediate collaborators) for the requester. This model deals with the disadvantages of traditional online forum systems. The mechanism actively recommends the most helpful experts with willingness to solve the given problem. Accordingly, questions would not be ignored and become solved in a short time via the shortest social path.

As for marketing strategy, Prior study has shown that social networks affect the adoption of individual innovations and products and the power of social network spreads information in breathtaking speed. In fact, purchase decisions made by users are usually influenced by the comments of purchased experience of their own social network. From the perspective of firms, the marketing behaviors focus only on the users who are powerful to others and willing to spread product impressions that can be expected. This strategy not only decreases costs but also increases correctness for marketing. The advancement of Internet infrastructure makes almost everyone has the ability to contribute or share information on the Internet. The sharing behaviors on the web are so-called “Web 2.0”. In other words, information flow is not purely as client/server structure but like the peer to peer architecture (P2P). The concepts of peer production and social network are also constructed by the power of Web 2.0.

2. Research Goals

In the marketing part, under current global economic structure, almost all firms have to face extreme competitions from competitors around the globe. In order to survive in such tough environment, superior marketing strategies are needed to raise sales, to gain larger the market shares as well as the loyalty of customers. Research in marketing behavior thus emerges as an important topic. In the project, the influential nodes discovery with potentials to achieve the effects of viral marketing was expected. How to measure the influence of each node is a very important topic because it decides which nodes are appropriate to be the “virus”. Enterprises can use the information to make a good marketing strategy and budget plans in order to achieve the best effects of infection.

In the knowledge-sharing part, the disadvantages of traditional user forums are as follows. First, if there is no participant interested in discussing the issue, the question will remain unanswered. Second, questions may remain in the discussion group for a long time before being answered. Third, the help-seeker usually has little guidance when given conflicting recommendations. Fourth, apart from the discussions, the advice or solution offered might be written in a form that is unclear or unreadable. Therefore, this research considers these four disadvantages and then proposes a recommendation mechanism for more efficient knowledge sharing in online forum communities.

3. Literature Review

In expert finding, as performed in TREC’s enterprise track, a system has to come up with a ranked list of experts with respect to a given topic of expertise, a corpus of enterprise documents and a list of the employees of the company as possible candidates [22]. Some other researches study the graphical structure of email communications to determine experts [3,9]. Although some recent researches [1,9] have enabled ranking function in expert search, few of them presents that ranking is not only affected by textual evidences but also by social factors. The previous works indicate semantic or textual-based analysis in recommendation domain is suitable and fruitful [14,16,27,28]. However, if only keywords in the content are used, some semantic information will be missing and some important cues may not be captured

[16]. In previous research, Role-base is usually applied in access control domain [15,23]. A fast-growing number of expert finding studies have shown that the social factor can help the researchers in understanding and analyzing certain implications and insights in knowledge network [5,25,26,28]. In knowledge sharing, a few studies apply Markov chains to model the scenario of expert finding [17,22]. Researcher start to attach importance to Markov chains in this field.

Viral marketing is a new marketing method which uses electronic communications (eg. e-mail) to trigger brand messages throughout a widespread network of buyers [7]. Dobele et al. [8] showed that "emotion" has more impact than "the expectation of recipient" in the successful message passing. Moore [18] investigated the branding influence based on viral marketing environment. Leskovec et al. [13] proposed a model to explain user behaviors in a large community. Richardson and Domingos [21] utilized probabilistic models and data from knowledge-sharing sites to design the best viral marketing plan. Zhan et al. [29] emphasize the important role of writing and referring product reviews in the internet. Hughes [11] proposed RFM (which stands for Recency, Frequency, and Monetary) analytical model in 1994 to measure the values of customers for enterprises. Newell [20] also stated that RFM method is very effective in customer segmentation. Drozdenko and Drake [10] applied the hard coding techniques on RFM weighting to assign weights to the three variables in RFM analysis. Chan [4] proposed a novel approach that combines customer targeting and customer segmentation for campaign strategies. Kuo and Chen [12] utilized fuzzy neural network to learn rules produced from order selection questionnaires in electronic commerce. Trust can also be used to indicate the strength level of relationships among people without doing detailed investigation of intention [24]. Munns [19] stated that trust is a relation from personal to individual, arises from the experiences of and influences on that individual. Brockand and Barclay [2] studied the relationships between buyers and sellers and showed that trust is based on character/motives/intentions and role competence/judgment. Dasgupta [6] states that trust is helpful in the condition where there is uncertainty about the actions that will be undertaken by others and when these actions are of consequence to those involved.

4. Methodology

To begin with, the proposed model analyzes the after-use reviews provided by online users and RFM values in each author's activity recorded to identify which authors are influential. The influential reviews represent the influence of their authors and the RFM value also indicates the infective ability of each reviewer by time segmentation. An influential ranking list of authors is expected to identify potential nodes so we hope to construct a well learned model in order to calculate each reviewer's mixed-score of two elements. Mass of data which contained complete review content and RFM attributes are needed for well-structured model training.

Subsequently, in forum network, the problem of intrinsic sparsity is also addressed as well as in the blogspace. Hence, many approaches are introduced to tackle the problems by adding implicit links between blog entries. We also focus these measures in forum as well as in blog, such as the similarity of counting the number of common tags, reply or site the same threads, etc. In this project, we utilize the multiplicity of links which takes social intimacy and popularity as a basis to calculate our score of social relation in a more comprehensive and exquisite way. In mathematics, a Markov chain is a stochastic process with the Markov property. Having the Markov property means that, given the present state, future states are independent of the past states. Therefore, combining these factors, we also apply the Markov Chain Model to predict the experts.

4.1 Discovering Influential Nodes for Viral Marketing

The high-ranking nodes are valuable targets for firms doing marketing. They are expected to spread fame of products and their manufacturers wider and stronger than other people as real virus. Firms can have some special strategies to take advantages of these potential reviewers. Figure 1 displays the concept and whole architectures about our system model simply:

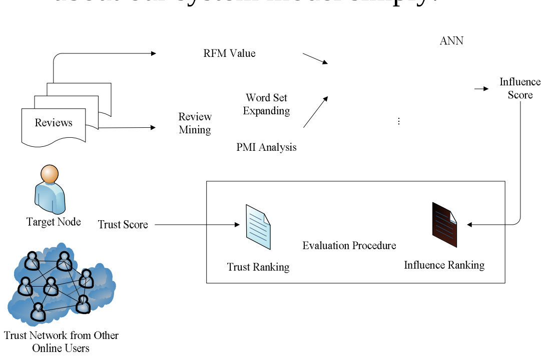

Figure 1. System concept and architecture

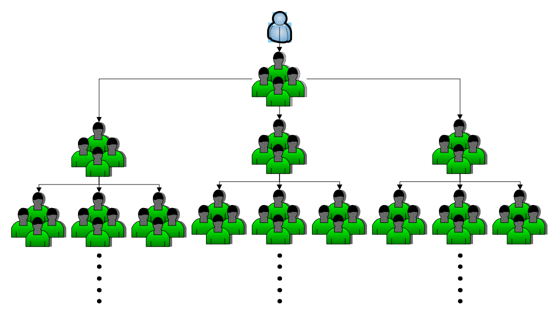

Figure 2. Trust name list expanding relations

Target node means the possible virus for viral marketing who is the product reviewer choose from an online social network environment. In this case, it is an online discussion area which provides a platform for users to write many kinds of product reviews. The model scores these viruses to decide which one is the most infective to market. The infective ability is decided by two factors: review and PMI value. The reviews was wrote by each reviewer will be analysis by text mining techniques to measure the score. The results of analysis will be quantified by our modified PMI model in six different degrees. In addition, the "RFM value" of each node acquired by recording attributes of each review (i.e. time, date, and category). The both scores will be weighted as the final virus score to decide this reviewer is valuable or not. Weighting mechanism in our proposed model is implemented by artificial neural network. It will learn the most appropriate structure of network to reflect the effects of each element by massive data training. The mechanism would be able to discover the hidden value in each review and consider verifying the effect of it at the same time. The detail statements about each unit in this architecture are described in the following subsections.

A trustable reviews must have fair attitude to reviewed and comment on products. Therefore, the positive and negative perspectives are combined to review analysis.

We define two sets of words which represent positive (S_p) and negative (S_n) sentiments respectively. Then, we expand S_{p+n} by recursive method in order to make a subjective word base to do subjective check. In order to consider the subjective of reviews, both word sets will be included in our model.

S_p={ good, nice, excellent, positive, fortunate, correct, superior }

S_n={ bad, nasty, poor, negative, unfortunate, wrong, inferior }

$S_{p+n}=S_p+S_n$

PMI (Pointwise Mutual Information) is used as a tool to calculate the strength score of each review as the basis for the results of review analysis. Turney and Littman [19] define PMI as the following equation:

$$PMI\left(t,t_i\right)=\sum\log_2\frac{\Pr_c\left(t,t_i\right)}{\Pr_r\left(t\right)\Pr_w\left(t_i\right)}$$

The key point of PMI calculation is the value of Prc(t,ti), Prr(t) and Prw(ti). We define each of them as follows:

$$\Pr_r\left(t\right)=\frac{n_{tr}}{N_r},\Pr_w\left(t_i\right)=\frac{1}{N_s},\Pr_c\left(t,t_i\right)=1$$

Many online discussion areas record impression scores for every online user that can establish their personal friend list and black list. This procedure also indicates each user's trust level to others. Figure 2 displays the trust relationships expanding among online users. The relationship can be traced deeper and deeper. The social connections of members of the discussion area could be observed. In addition, the intimate level between nodes can also be calculated in detail by leading in more trusting score. The social connections represent the influencing range of each node so called SCN (Social Connection Number). Influence power relate to the range of SCN of reviewer.

The original concept of "Recency" is the days from the last purchase date to now. In the experiment, "Recency" seems as the time range γ between current date and the latest wrote date of each node. It is measured by days. The benchmark date (i.e. current date) is set at May 20th, 2008 due to the experimental duration. Recency standardization is little different from general standardize procedures because of higher values indicate lower market values. In order to display real meaning of Recency and the convenience of later calculations, the following formula is used for Recency standardize procedures:

$$Std_{x_i} = \frac{|\gamma_i - \max \gamma_i|}{\max \gamma_i - \min \gamma_i}$$

The SCN range trace of target node can be formulated as following equation:

$$SCN = \cdots \sum_{j=1}^{n1} \sum_{i=1}^{n2} f_{ij\cdots}$$

Our purpose is to discover how large the influential range of each reviewer is, and this is a fair indicator to determine his / her influence. In other words, we want to know these reviewers are trusted by how many people. The whole Social Connection Number (SCN) of a reviewer can be constructed be recursive tracing. The data used in this experiment was collect from Epinions.com. It is an open platform which provides online users writing reviews to various products. It provides a simple trust mechanism for members to identify the effect of reviews that is good for us to retrieve related data. Dataset composition is through randomly picked up ten reviews from one product in each sub-classification respectively. In general, each review was written by one author (or node), then retrieve whole reviews written by these nodes to analysis. The process needs the techniques of web crawler or opened dataset. Training dataset include 2952 reviews which are randomly selected from Electronics sub-categories of Epinions.com. There are 715 reviews wrote by 16 reviewers are retrieved as testing dataset. Mean Absolute Percentage Error (MAPE) is adopted to reach the goal. A stands for the actual value and F is the forecasting value of data. The concept of MAPE is very simple to understand and the difference between actual value and predicting value will be displayed clearly. In addition, the reviewers in our testing data set all have some basic level of trust value so we do not need to worry about the denominator would be zero.

Time attribute of each review is needed for the calculation of Recency and Frequency value. It is convenient that the two indicators are both on the reviewer's viewpoint originally. The standardized Recency and Frequency value are displayed in Table 1. Preliminary analysis of RFM reveals that large differences exist among these reviewers' publication. Reviewers get low Recency and similarly Frequency value if reviewers write reviews continuously. This would be helpful for identifying influential nodes. The purpose of trust score calculation is to identify how large the influential range of reviewers, that is, to know the reviewers are trusted by how many people. After retrieving 2 level social relationships of each node, it grows to about 130 thousand relationships. The growing of social networks is really amazing. The result was shown in Table 2.

Table 1. Recency and Frequency value

ID / Period	Recency Standardize	Frequency <90 days	Frequency 90-365 days	Frequency >365 days
ASourdough4	20	0.090	0.260	0.650
AtlantaGreg	94	0.013	0.026	0.962
corona79	68	1.000	0.000	0.000
dkozin	35	0.207	0.272	0.522
Howard_Creech	50	0.030	0.080	0.890
hwz1	890	0.000	0.000	1.000
JIMILAGRO	1418	0.000	0.000	1.000
jvolzer	97	0.016	0.339	0.645
njpoteri	1518	0.000	0.000	1.000
porcupine1	91	0.050	0.350	0.600
readsteca	121	0.000	0.234	0.766
sarahrose12	69	0.083	0.000	0.917
theheidis	232	0.000	0.067	0.933
tucknroll	851	0.000	0.000	1.000
williamrender	1484	0.000	0.000	1.000
zan720	1079	0.000	0.000	1.000

Table 2. SCN of testing reviewers

Id / SCN	*k=1*	*k=2*	*Standardize*
ASourdough4	137	19592	0.3700896710
AtlantaGreg	4	247	0.0046899036
corona79	1	223	0.0041833940
dkozin	393	16084	0.3090834052
Howard_Creech	804	52503	1.0000000000
hwz1	759	40631	0.7764416764
JIMILAGRO	1	10	0.0001875961
jvolzer	4	85	0.0016508461
njpoteri	1	0	0.0000000000
porcupine1	1	0	0.0000000000
readsteca	1	1	0.0000187596
sarahrose12	2	2	0.0000562788
theheidis	2	52	0.0009942596
tucknroll	1	2	0.0000375192
williamrender	1	0	0.0000000000
zan720	3	198	0.0037519229

4.2 A novel recommendation mechanism for knowledge sharing in online forum communities

In this study, we propose an innovative recommendation mechanism, which employs the role analysis, social relation, and semantic analysis to construct a more comprehensive and personalized framework for each users on the entire forum space, on both discussion threads and experts in the knowledge forum. There are various important factors and dimensions we should take into consideration. We employ three underlying critical aspects: Profession and Reliability (PR), Social Intimacy and Popularity (SIP) and Semantic Similarity (SS). Moreover, we apply Markov Chain model to find the most available e and helpful experts while the user doesn't find the satisfied threads. Figure 3 depicts the architecture of the proposed Markov Chain-based recommendation mechanism. This study proposes a Markov Chain model based forum recommendation mechanism combined with the concept of role analysis, social relation, and semantic analysis. This mechanism contains the information of the forum network about profession and reliability, social intimacy and popularity, and semantic similarity respectively. The whole process of recommendation mechanism is divided into several steps as shown in Figure 4 and is described in the following sub-sections.

First of all, we choose a start point by random from the users. From friend lists and groups, we can search available and social-reachable agents. These agents are connected level-by-level by friend or friend-of relationships in the forum network. Moreover, we also take the post-reply relationship into consideration. Once the agents are decided and specified or the maximum number of searching level is reached, the members of the recommender are confirmed. Then we crawl information (such as personal file, threads post or replied, messages, etc) associated with each agent on the recommendation network. We simulate the forum network which applies the concepts of the agent and object to implement and evaluate the proposed model. In this graph-based forum network (shown in Fig. 5), m agents (users) and n objects (threads) are denoted as nodes and document-like icons, respectively. The relation edges in the network denote heterogeneous and multiplicity of links (whether explicit or implicit links). First, we clarify the existence of links and classify and annotate known links for both explicit and implicit ones to identify potential relationships in this graph. In this study, the relations are classified into following three aspects: Agent-to-Agent relation, Agent-to-Object relation, and Object-to-Object relation.

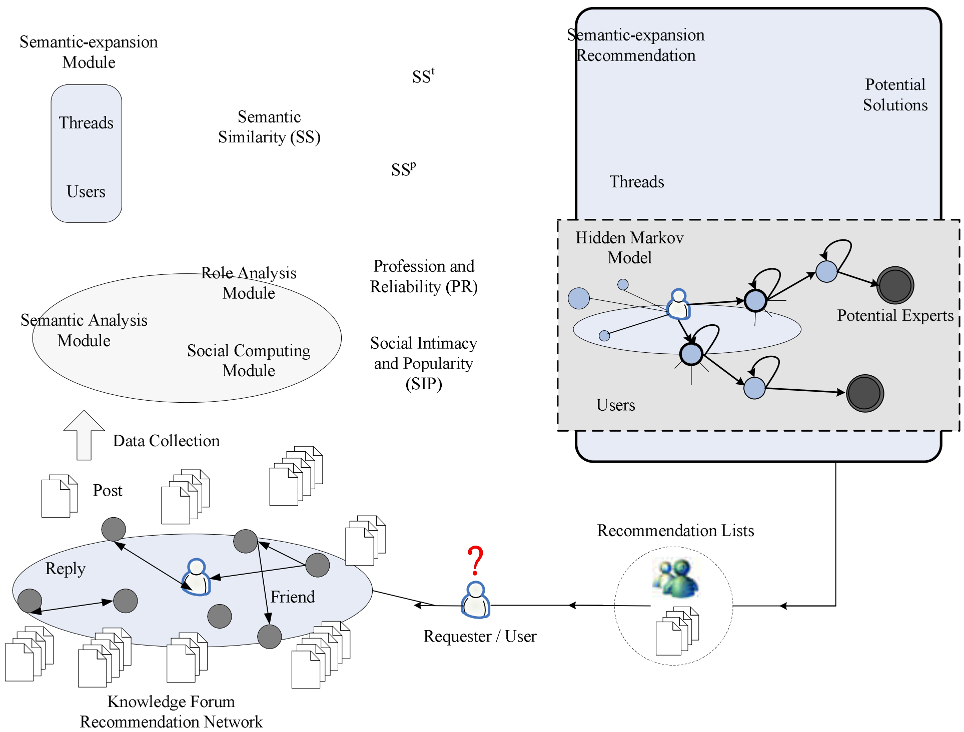

Figure 3.The architecture of the proposed recommendation mechanism

There are two kinds of relations in A-A relation. First, a friend, friend-of, or join the same group relation, reflected in the friend lists, is a hyperlink from agent to agent. We quantify the relation as a part of professional and reliability score. Second, the relation is about social similarity level which measure the strength of social intimacy and interaction in common between agents. In this part, not only explicit links in physical but also implicit similarity relations of social behaviors are taken into account, i.e. reply the same post, topic similarity, the number of same terms, cite the same threads, etc.

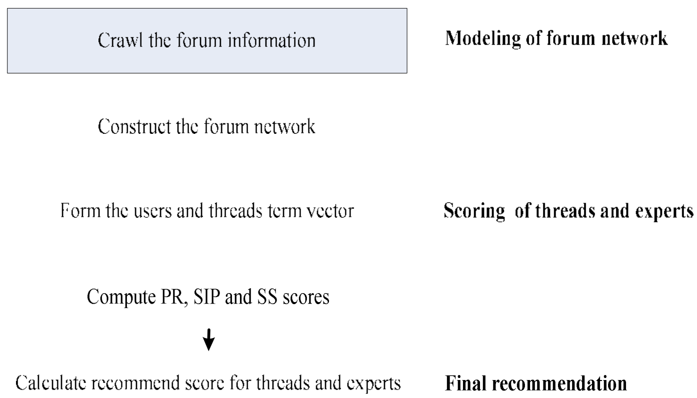

Figure 4. The whole process of recommendation mechanism

We put more emphasis on interactions occur in replying behaviors and it is the most interactive and conversational way, compared with other interactions. In accordance with this agent to object relation not only reveals the interests, profession domain and social intimacy of replier toward specific threads (topics) but also shows the popularity of requesters. It is intuitive that a certain thread (topic) gets high popularity when it has more replies and citations (in-degree links) from other agents in spite of the threads (topics) type, semantic of threads, and freshness factors. Owing to the above mentioned, replying is a crucial social behaviour to show the social importance in forum network. We can examine the SIP score associated with popularity degree. Another relation between agents and objects is possession relation and it implies that objects are post by an agent. Here is the entrance to connect agent with object layer for the purpose of inducing a requester-oriented social networking and computing mechanism.

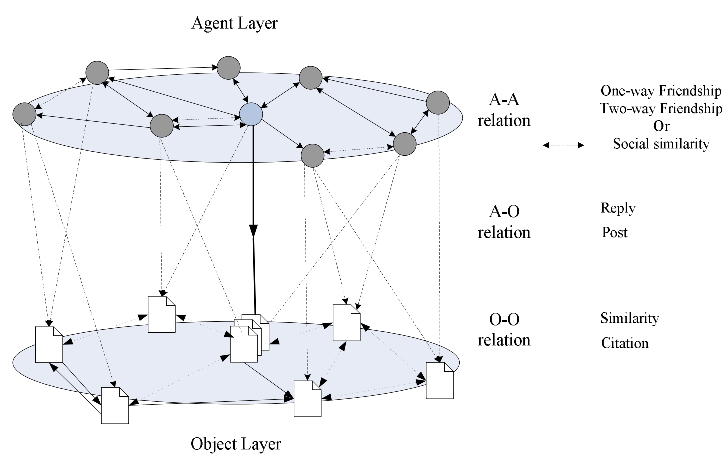

Figure 5. The definitions and classifications of links among blog network

In order to compute SIP score, citation behaviors should be brought into the model to improve the recommendation completeness. Therefore, we especially emphasize on similarity between objects. In human-expert forum context, similarity plays the same role in recommending threads and experts. The proposed approach divides the concept of similarity into two sorts: social intimacy similarity and semantic similarity of threads which associate with SIP and SS score respectively.

The $w_{i,j}$ can be interpreted as the importance of index k_i to discussion thread d_j, and it can be estimated by $w_{i,j} = tf_{i,j} * idf_i$. The equation is similar to the famous tf*idf equation, the $freq_{i,j}$ is the frequency that index term k_i appears in discussion thread d_j. The max. $freq_{l,j}$ is the number of times the most frequent index term, k_l, appears in discussion thread d_j. N is the number of discussions in D, and n_i is the number of discussions which contain index term k_i. In addition to fit the forum network, two parameters u, r are added to the equation of term frequency (tf), where

$$tf_{i,j} = u * \frac{freq_{i,j}}{\max_l \, freq_{l,j}} * (1+r), \quad idf_i = \log \frac{N}{n_i}, \text{ and}$$

$$r = \begin{cases} \dfrac{count(reply_{i,j})}{count(reply_j)}, & \text{if has reply in } d_j; \\ 0, & \text{if has no reply in } d_j. \end{cases}$$

In this study, we measure the in-degree (the number of incoming links) in our model. Moreover, we apply a variable λ, which presents the level of willing to share (reply) of an agent. Since an object (discussion thread) u belongs to an agent s, and there are n objects of s, we compute the aggregate value of u, the formula as follow:

$$Popularity(o_{ij}) = \left\{ \sum_{j=1}^{n} w_{co} \times \frac{Reply(o_{ij})}{\max Reply(A)} + w_{ci} \times \frac{Citation(o_{ij})}{\max Citation(A)} \right\} \times \sqrt{\lambda}$$

where $Reply(o_{ij})$ ($Citation(o_{ij})$) are the number of replies (citations) in object j of agent i and $\max Reply(A)$ ($\max Citation(A)$) is the maximum number of replies (citations) in our dataset. The variable λ is replies per day of an agent i. Obviously, the popularity score of an agent i take not only the global reputation but also the willingness of reply into consideration. The parameters w_{co} and w_{ci} are the weights of in-degree links from reply and citation behaviours respectively. Moreover, we apply Markov Chain model to compute and predict the most willing and capable experts. Firstly, suppose requester r selects a direct-linked friend f to ask for help. Then, r is the current state, and f is the next state. The transition probability is a condition probability generated by inter-links is social-sensitive. The higher that the transition probability that f gets, the more importance that f gets to r. The transition probability from requester r to friend f is as follow:

$$p(f_i \mid r) = \frac{S_{Reply}(f_i, r)}{\sum_{n=1}^{N} S_{Reply}(f_n, r)}$$, where N presents the total number of the direct-linked friends of r, and $\sum_{n=1}^{N} S_{Reply}(f_n, r)$ is the sum of the replies whish requester reply to his/her friends. We scoring the *SS* (*SS*^{t} *and SS*^{p}), *PR*, *SIP*, and introduce the Markov Chain model as mechanism of experts finding. When a query *q* is post, according to the *SS*^{t}, the model generate recommendation list of *k* discussion threads to the requester. If the requester doesn't find the qualified answer, the model list *k* experts which are the most helpful and willingness. The final recommendation score of experts is formulate as follow:

$$R^F(r, e_i) = p(e_i \mid r) \times [PR(e_i) + SIP(r, e_i)]$$

Table 3. The statics of the relevant discussion threads

Question	Average replies in each feedback threads (10 threads)	Relevant threads in results (10 results)
Q1	9.6	8
Q2	6.8	7
Q3	8.8	9
Q4	5	9
Q5	6.2	6
Q6	7.3	8
Q7	3.4	7
Q8	7.1	7
Q9	6	8
Q10	8.7	7
Average	6.89	7.6

This study crawls 3000 discussion threads as the testing data, collected from the Yahoo Answers (http://answers.yahoo.com/). We focus on the Sports forum. Then, we fetch another 10 discussion threads which only reserve the topics, question descriptions, requester's person profiles to request information and help from experts. The model will return 10 relevant discussion threads and 5 experts. Figure 6 shows the interface of the recommendation result. We examine the 10 feedback threads of each request query (10 query), the average replies are nearly 7 (6.89) and relevant proportion is 76 %.

Table 4. The statics of the recommend experts

Question	Average best answer rate of 5 experts in each feedback threads (10 threads)	Number of intermediate agent
Q1	36%	1
Q2	24%	1
Q3	13%	1
Q4	38%	1
Q5	28%	1
Q6	21%	1
Q7	7%	1
Q8	19%	1
Q9	13%	1
Q10	34%	1
Average	23.3%	1

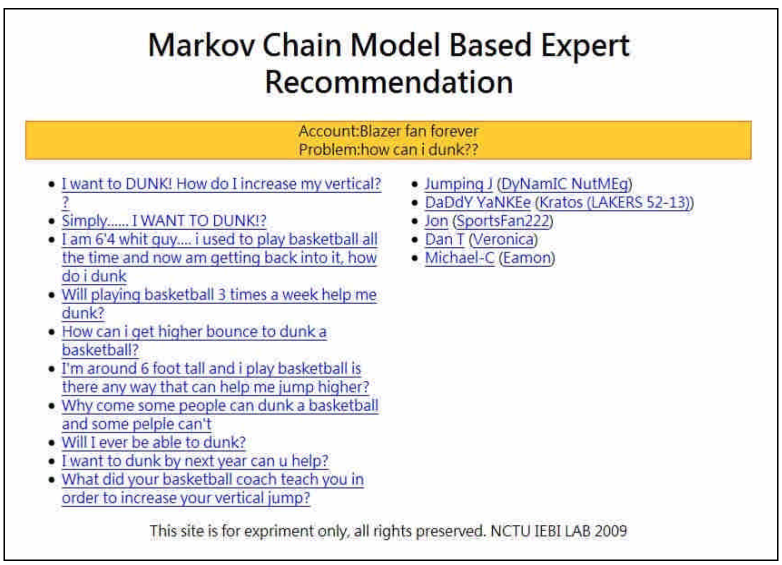

Figure 6. The recommendation result s of relevant threads and experts

The statics are list in Table 3. Hence, the best answer rate of recommended expert is 23.3% and the number of intermediate agents is equal to 1. The statics are list in Table 4. Accordingly, the results show that after the recommendation mechanism this study proposed, a requester in forum can easily find similar discussion threads to prevent spamming the same discussion. Besides, if a requester can not find the qualified discussion threads, this study provides a more efficient and active way to find the willingness experts.

5. Conclusion

As we have observed, although the advancing of IT technology and the Internet reduce the cost of marketing behaviors such as advertisement, the “uncertainty” problem still exists. Many enterprises waste much resource on invalid marketing behaviors. Viral marketing is a new and effective marketing method which is based on the power of “word of mouth” for saving much resource and troubles in mass marketing. How to find the potential nodes that are powerful to others and willing to spread positive product impressions efficiently is the key of viral marketing. Via Internet, the recommendations from other online users’ product review comments have more influential power than traditional advertising. In this work, a solution to find potential, influential nodes was proposed. The text mining techniques and the RFM analysis were combined to calculate the influential power of real online users through her/his reviews. The trust score which is composed of thousands of human connections is applied for evaluation. The final results also passed the examination of trust.

In addition, this project proposes a novel expert recommendation mechanism which combined with Profession and Reliability (PR), Social Intimacy and Popularity (SIP) and Semantic Similarity (SS) based on Markov Chain model. The process of recommendation divided into two phases, relevant discussion threads recommendation and helpful experts finding. The preliminary experiment shows the high relevance and professional results. The next steps will consist of evaluating the mechanism on several categories of forum and compare with other algorithm, such as HITS and PageRank. Hence, we’ll focus on how to maintain and enhance the human-experts’ knowledge domain and construct more complete semantic concepts than WordNet. Finally, the overall performance of the recommendation mechanism is the most importance issue where we are concerned.

Reference

[1] Balog, K., Azzopardi, L., Rijke, M.D. “Formal Models for Expert Finding in Enterprise Corpora,” Proceedings of the 29th annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, pp. 43-50.

[2] Brockand, S. J., and Barclay, D.W. “The effects of organizational differences and trust on the effectiveness of selling partner relationships”, Journal of Marketing, 61, 1997, pp. 3-21.

[3] Campbell, C.S., Maglio, P.P., Cozzi, A., Dom, B. “Expertise Identification Using Email Communications,” Proceedings of the Twelfth International Conference on Information and Knowledge Management, 2003, pp. 528-531

[4] Chan, C.C. Intelligent value-based customer segmentation method for campaign

management: A case study of automobile retailer”, Expert Systems with Applications, 34(4), 2008, pp. 2754-2762.

[5] Cooke, R.M., ElSaadany, S., Huang, X. “On the Performance of Social Network and Likelihood-based Expert Weighting Schemes,” Reliability Engineering and System Safety, 93, 2008, pp. 745-746.

[6] Dasgupta P., Trust as a commodity. Edited by Diego Gambetta. Trust: Making and breaking cooperative relations. Oxford, UK: Blackwellm, 1988.

[7] Dobele A., Toleman, D., Beverland M., “Controlled infection! Spreading the brand message through viral marketing”, Business Horizons, 48(2), 2005, pp. 143-149.

[8] Dobele A., Lindgreen A., Beverland M., Vanhamme J., Wijk R., “Why pass on viral messages? Because they connect emotionally”, Business Horizons, 50(4), 2007, pp. 291-304.

[9] Dom, B., Eiron, I., Cozzi, A., Zhang, Y. “Graph-based Ranking Algorithms for E-mail Expertise Analysis.” Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2003, pp. 42-48.

[10] Drozdenko, R.G., Drake P.D., Optimal database marketing: strategy, development, and data mining. Thousand Oaks: Sage Publications, 2002.

[11] Hughes, A. M., Strategic database marketing- The masterplan for starting and managing a profitable, customer-based marketing program, third edition. McGraw-Hill Professional, 2005.

[12] Kuo, R. J., Chen J.A. “A decision support system for order selection in electronic commerce based on fuzzy neural network supported by real-coded genetic algorithm”, Expert Systems with Applications, 26(2), 2004, pp. 141-154.

[13] Leskovec J., Adamic, L.A., Huberman, B. A. “The Dynamics of Viral Marketing”, ACM Transactions on the Web (TWEB), 1(1), 2007.

[14] Li, Y.-M., Chen, C.-W. “A synthetical approach for blog recommendation: Combining trust, social relation, and semantic analysis,” Expert Systems with Applications, 36(3), 2009, pp. 6536-6547.

[15] Li, N., Mitchell, J.C., Winsborough, W.H.“Design of a Role-based Trust-management Framework,” Proceedings of the 2002 IEEE Symposium on Security and Privacy, 2002.

[16] Liang, T.-P., Yang, Y.-F., Chen, D.-N., Ku, Y.-C. “A Semantic-expansion Approach to Personalized Knowledge Recommendation,” Decision Support Systems, 45, 2008, pp. 401-412.

[17] Lin, C., Zhou, H., Huang, Z., Wang, W., “REC: A Novel Model to Rank Experts in Communities,” The Ninth International Conference on Web-Age Information Management, 2008, pp. 301-308.

[18] Moore, R.E., “From genericide to viral marketing: on "brand"”, Language & Communication, 23(3-4), 2003, pp. 331-357.

[19] Munns, A.K. “Potential influence of trust on the successful completion of a project”, International Journal of Project Management, 13(1), 1995, pp. 19-24.

[20] Newell, F. The new rules of marketing: How to use one-to-one relationship marketing to be the leader in your industry, McGraw-Hills Companies Inc., New York, 1997.

[21] Richardson, M., Domingos, P. “Mining knowledge-sharing sites for viral marketing”, Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002.

[22] Rode, H., Serdyukov, P., Hiemstra, D., Zaragoza, H. “Entity Ranking on Graphs: Studies on Expert Finding,” Technical Report TR-CTIT-07-81, Centre for Telematics and Information Technology, University of Twente, Enschede, 2007.

[23] Sandhu, R., Ferraiolo, D., Kuhn, R. “The NIST Model for Role-based Access Control: Towards A Unified Standard,” Proceedings of the Fifth ACM Workshop on Role-based Access Control, 2000, pp. 47-63.

[24] Simmel G., and Frisby D. The Philosophy of Money. ROUTLEDGE TAYLOR & FRANCIS GROUP, 2004.

[25] Tobias, M.P., Finke, I. “SELaKT-Social Network Analysis as a Method for Expert Localisation and Sustainable Knowledge Transfer,” Journal of Universal Computer Science, 10(6), 2006, pp. 291-701.

[26] Wang, J.-C., Chiang, M.-J., Ho, J.-C., Hsieh, W.-T., Huang, I.-K. “Knowing Who to Know in Knowledge Sharing Communities: A Social Network Analysis Approach,” Proceeding of The Seventh International Conference on Electronic Business, 2007, pp. 345-351.

[27] Weng, S.-S., Lin, B., Chen. W.-T. “Using contextual information and multidimensional approach for recommendation,” Expert Systems with Applications, 36(2), 2008, pp. 1268-1279.

[28] Yang, S.J.H., Chen. I.Y.L. “A Social Network-based System for Supporting Interactive Collaboration in Knowledge Sharing over Peer-to-peer Network,” Int. J. Human-Computer Studies, 66, 2008, pp. 36-40.

[29] Zhan, J.M., Loh, H.T., Liu, Y. “Gather customer concerns from online product reviews - A text summarization approach”, Expert Systems with Applications, In Press, Corrected Proof, Available online 28 December 2007.

計畫成果自評

本計畫工作目前正在順利進行中，其主要核心為網路知識管理以及電子商務之建構與分析。本專題計畫目前完成項目與計畫初預期項目之情形如下：

1 廣泛文獻探討與理論方法研究

我們已發表的兩篇研討會論文[附錄 1－2]分別在「研究背景」與「相關文獻」都有這些內容之記載。領域涵蓋「社會網路分析」、「多目標決策」、「病毒式行銷」、「資訊萃取」、「社會網路運算」、「信任運算」等。本研究先從技術層面，在既有與建立之社會網路運算基礎服務平台，找出潛藏在網際網路之中的人類社會結構，並據以發展人類社會實際互動的各項應用。基於這些工作，我們得以建立社會決策支援系統（SN-based Decision Support Systems ,SNDSS）。另外，我們也發現病毒式行銷在電子商務上的優缺點，以作為未來進一步分析與探討的基石。

2 社群知識分享之研究

論文[附錄 1]主要研究如何發展專家推薦系統以作為社群網站知識分享的用途。在這個研究裡，我們結合了專業度與可靠度(Profession and Reliability)和以馬可夫模式為基礎的社會相似度(Social Intimacy and Popularity)，作為社群網站知識推薦系統的開發工具。這個專家推薦系統包含兩個步驟，分別為相關討論群組的推薦和專家協尋功能。這個推薦系統的優點在於它全面性地涵蓋在知識分享的觀點。因此，我們實驗的結果顯示在這個推薦系統的協助之下，在討論區的要求服務的使用者能夠很容易地找到相似的討論群組以避免重覆覽閱大量重覆的討論文章。除此之外，假如要求服務的使用者無法找到合適的的討論群組，本研究提供一個相對有效率且主動的方法以協助使用者找到正確的專家。

3 社群網路行銷方式之研究

論文[附錄 2]主在研究如何方能協助企業能有效率的在社群網路中進行病毒式行銷，發揮最大行銷效益。此研究中，我們結合了人工智慧方法，實踐企業智慧於網路環境之中，成功地在浩瀚無垠的網路社群中發掘具有影響力之使用者。令企業可進行目標式行銷，以協助企業正確快速地鎖定行銷推廣的顧客族群，在增進行銷效果的同時亦節省其行銷成本，提昇企業利益。

目前本專題研究成員將整理研究成果，並將最後成果投稿國際期刊。本研究團隊在經濟分析上的傑出部份是我們認為能夠大幅超越過去資訊管理領域在網路服務品質研究上的優勢，同時也是本研究室核心，並且是全球資訊管理學界頂尖學府如MIT、CMU目前專注的領域。因此，如受到獎勵資助的研究時程能得以延伸，我們相信能夠做出更創新、突破的研究。

附錄

已發表會議論文

1. “A Novel Recommendation Mechanism for Knowledge Sharing in Online Forum Communities”, Li, Y.-M., T.-F. Liao, Proc. 13th Pacific Asia Conference on Information Systems (PACIS 2009), Hyderabad, India, July, 2009. Forthcoming.
2. “Discovering Influential Nodes for Viral Marketing”, Li, Y.-M., C.-Y, Lai, C.-H. Lin, Proc. 42th Hawaii International Conference on System Science (HICSS-42), Manoa, Hawaii, USA, January, 2009.