國立交通大學

資訊管理研究所

碩 士 論 文

在病毒行銷中尋找有影響力的節點

Discovering Influential Nodes for Viral Marketing

研 究 生：林嘉豪

指導教授：李永銘　博士

中 華 民 國 九 十 七 年 六 月

在病毒行銷中尋找有影響力的節點

Discovering Influential Nodes for Viral Marketing

研 究 生：林嘉豪　　Student：Chia-Hao Lin

指導教授：李永銘　　Advisor：Yung-Ming Li

國 立 交 通 大 學

資 訊 管 理 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Information Management

College of Management

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master

in

Institute of Information Management

June 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年六月

在病毒行銷中尋找有影響力的節點

學生：林嘉豪　　指導教授：李永銘 博士

國立交通大學資訊管理研究所碩士班

摘　　要

在傳統的行銷行為中，高昂的成本和效果的不確定性是常見的問題。我們發現對於一般人來說，由顧客所撰寫的線上產品使用心得通常比廠商的廣告更可信，尤其當這些心得是由他們的朋友們所寫的時候。社群網路的力量也使得這些產品形象就像真實的病毒一樣以驚人的速度傳播。發現那些撰寫有價值的產品評論、並且擁有廣泛人際關係的特定評論者們會是一個解決行銷上不確定性問題的好方法。

在本研究中，我們使用兩個方式來衡量每位評論者的行銷價值：改良的 PMI 和 RFM。PMI 可以量化每篇評論探勘的結果，RFM 則被用來把每位評論者的寫作情況納入影響力分數計算之中。人工智慧技術中的類神經網路被使用來為我們的模型訓練合適的網路架構。影響力的指標：信任機制被使用在模型的評估上。它包含了真實世界中數以萬計的人際關係網路。實驗結果顯示我們的模型在選擇具有影響力的評論者上比「人氣作者」和「評論分數」的排序方法有更佳的效果。本研究能指出哪些評論者在產品資訊的傳遞上是有效的，這份結果也對欲進行行銷活動的廠商具有參考的價值。

Discovering Influential Nodes for Viral Marketing

Student: Chia-Hao Lin Advisor: Dr. Yung-Ming Li

Institute of Information Management
National Chiao Tung University

ABSTRACT

High cost and uncertain effects are main problems of traditional marketing behaviors. We discover online product reviews, which are written by customers are usually more trustworthy than firms' advertisements for people, especially those written by their friends. The power of social network also makes these product impressions spread in amazing speed as real viruses. To discover influential reviewers who write valuable product reviews and have wide human relationships is a good way to solve the problem of marketing behaviors.

In this research, we propose two methods to measure the marketing value of each reviewer: revised PMI and RFM. The modified PMI quantifies each review mining result and the RFM concept is used to take each reviewer's writing status into consideration of influence calculating. The artificial neural network (ANN) is adopted to train an appropriate network structure for our model. The influence power indicator: trust is applied in the evaluation of our model and it considers thousands of human relationships among the real world. Experiment result shows that our model outperforms "popular author" and "review rating" methods in selecting influential reviewers. This research can point out which reviewers are really effective in product information spreading and the results will be valuable for companies to refer.

誌　　謝

兩年的研究生活很快的過去了。在這段時間中，經歷了許多充滿挑戰的難關，以及與研究室夥伴們的共同回憶。過程中不論是喜悅、悲傷、無奈抑或是憤怒，這本論文終究在許多人的協助與支持下一步步的完成了。在這裡，我要感謝許多曾經對這本論文的誕生付出心力的師長、同學及好友們。沒有你們，這本論文無法完成，而我也無法因為這些歷練而有所成長。

首先我要感謝我的指導教授李永銘博士。老師對於研究有著非常高的熱忱，在這兩年之中，有了老師的諄諄教誨，才能一步步地磨練出我對於研究能力的基本功，本論文以及其他眾多的研究才能順利完成。兩年的研究生涯老師引領我進入學術的領域，也帶給我許多終身受用的研究成果。此外，在為人處事上，我也從老師的身上學習到許多待人接物的道理，相信未來在社會上也能夠學以致用，真的是非常感謝您。口試委員林福仁教授、劉敦仁教授以及簡宏宇教授的指教也對論文的缺失與改善提供了非常寶貴的意見，這篇論文才得以產生更大的價值。非常感謝諸位老師的協助。

研究室的夥伴們對這篇論文的完成也有著莫大的貢獻。有了建邦，這篇論文的許多想法才能成形，實驗的進行才能夠順利。建邦也常常和我一起討論許多課業以外的問題，相信不只是我，研究室的每一位成員或多或少都是因為你的協助，才能解決許多人生與課業上的問題。在最後的半年，我們又很湊巧的成為室友，在和你相處這麼長的時間之中，真是獲益匪淺，希望能早日接到你的喜訊。在這裡我要毫不吝嗇的感謝並讚揚你對整間研究室的貢獻，謝謝你。海王子敬文是我們這屆的另一個戰友，跟你講話最大的好處就是完全沒壓力，而且你獨特的說話方式真的非常的幽默，我和建邦常常被你逗得大笑；感謝你為我的論文所提供的許多好點子，以及大家在一起時所提供的歡笑，還有宜蘭真的很好玩!相信就算幾十年後，我依然會記得管二樓下那塊我們三個放鬆和討論各種想法的地方。

這兩年的生活中，學長和學弟妹們也對我提供了極大的幫助。我們所遭遇的各種困難，易霖學長總能一派輕鬆的解答，不愧是有歷練又優秀的好學長。無尾熊是個懂得享受生活又風趣的學長，和你在一起總是非常愉快。我一定會記得和你大吃大喝和卡丁車的日子，祝你快快畢業，未來成為一位好教授。還有 Denny，在論文最緊要的

關頭提供了我最大的幫助，真的非常感謝你；恭喜你順利爭取到交換學生的資格，但是卡丁車還要再練練唷。涵文、連乃、子鳳、宗穎、正乾，謝謝你們為我們所付出的心力，我會一直記得那段和你們在研究室同甘共苦的日子。馬上就要碩二了，壓力一定很大，相信你們明年一定也能有好的成果，快快樂樂的畢業；也祝正乾在台大有新的生活、新的體驗、新的收穫。

除了自己研究室以外，和我們關係最密切的就是最佳化研究室的同學們了。賢慧的亞梅學姊總是能把研究室整理的井然有序，在我們煩惱時，也提供了最有力的協助。筱嵐學姊是我日劇的同好之一，祝妳在香港一切順利，能有好的研究成果。怡菱學姊是個有趣的人，謝謝妳常在我們忙到深夜時適時出現，帶給我們輕鬆愉快的氣氛。葵棠和昱劭的智慧，幫助我們度過許多艱難的課程，非常感謝你們。還有盈佑和總機小姐，謝謝你們總是幫大家解決民生問題。

另外還有鵝蛋，你雖然不是研究室的同學，但是論文的完成你功不可沒。在最後的一刻，你在最短的時間幫助我迅速的修改好論文，對於壓迫到你的休息時間至今我仍深感抱歉；能有你這位責任心重、可靠又有義氣的好朋友，相信會是我一生的資產；真的非常非常感謝你。

最後，我要由衷感謝家人對我的支持與栽培。有了父母的關心、慰問與支持，我才能夠度過每一個難關與困境；才能夠沒有後顧之憂的傾注全部的心力在研究上。在研究進行到最後的一段日子裡，我沒有辦法常常在身邊陪伴你們，真的非常抱歉。謝謝你們對我永遠的寬容和體諒。

這兩年的研究生活對我來說，論文不只是唯一的紀念。在這段日子裡，我得到了更高深的學識、具備了初步的研究能力、有了一些成果、得到了許多摯友，還培養了良好的適應性與抗壓性。相信在未來的日子裡，這些寶藏會讓我終身受用無窮，感謝在這段時間裡幫助我的所有人。

林嘉豪

2008 年六月　謹誌於　新竹　國立交通大學光復校區

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Research Background

Marketing is the key of commercial activities. Viral marketing is a new marketing method which spreads product information based on people's word-of-mouth. The effect of it grows in incredible speed in the Internet era. How to integrate viral marketing correctly into overall marketing strategy is very important for firms if they want to save significant marketing cost and create more business chances.

We have known that outstanding marketing strategies usually bring future revenues to enterprises because they add extra values to products (include physical, virtual products or services) and firms themselves. Future revenues not only mean current sales but also represent some business chances which are hard to predict and valuable for future business growing. As a result, marketing behaviors are absolutely needed by firms to maintain and create more profits. In facts, lots of scholars have done so many researches about marketing which cover various perspectives and domains. For example, organizational issues relevant to marketing strategy (e.g. branding, competitive behavior, positioning, and segmentation), organizational issues that span functions (e.g. quality management), the interface between marketing and business strategy, organization level phenomena that impact marketing strategy (e.g. market orientation, corporate culture), and outcomes of marketing strategy (e.g. market share, customer satisfaction) [53].

We know the purpose of marketing is getting high growing in sales, market share, and gross margin in the marketplace and the ultimate goal is enhancing the shareholder returns [46]. However, while the technological advancement reduces the manufacturing and managerial costs, the marketing cost rises rapidly [56]. The implementation of Just-in-Time

strategy and flexible manufacturing systems had reduced the manufacturing cost efficiently. Managerial costs (e.g., finance, accounting, human resource, R & D) also declined as a fraction of total corporate costs due to the adoption of new managerial and IT tools. In the cost structure of firms, only the marketing costs (include expenses such as product development, selling, distribution, advertising, sales promotion, public relations, customer service, outbound logistics, and order fulfillment) raised a lot in recent 50 years [43]. In addition, the returns on marketing are usually unpredictable, especially in advertisement. For example, firms can design delicate advertisements and spread them through various media but it is very hard to predict how many people will be attracted by the advertisements and how much revenues will be generated. The investment may be a huge success due to some special ideas in the advertisement or a catastrophic failure after spending a huge budget. These marketing behaviors are undoubtedly risky for a firm's finance, especially for business of small scale. High cost and uncertain characteristics make enterprises have to spend more and more money on marketing, but the revenues cannot be guaranteed so firms are reluctant to invest too much on marketing plans. This strategy saves some resource and cost temporarily but blocks the future growing because appropriate marketing plans are definitely needed for sustainable management in every business. Firms may also lose some profitable business chances due to conservative strategy in marketing plans.

High cost and uncertainty problems exist in the long business history of humanity no matter how excellent and cost-cutting technologies are invented in manufacturing. The marketing cost problems are still the killer for gross profits. However, the invention of the Internet changes the world and also creates new chances to traditional marketing. After 1990, the Internet brings new methods for transactions into the traditional marketplace. We can find many new business models which are born due to the Internet such as online shopping or online banking. The whole business environment also produced new transaction and

advertisement models. In the Internet era, to run a business does not need grand scale or capital. Even individuals have the ability to do business with others rapidly and accurately. By adopting simple business models, sometimes they even have better performance than traditional business models. C2C, virtual marketplace, auction brokers, and social network marketing are all outputs under this environment [38].

In the e-commerce world, information about traditional products can now be exchanged in more convenient ways now. Many kinds of new products which are different than traditional ones are also emerged because they are ideal to deliver directly in the Internet era. For example, many knowledge-based products which can be stored in digital format like movies, music, and e-books are all common in our daily life. The characteristics of the Internet make it is the most appropriate platform for knowledge-based product delivery. From the marketing's point of view, or the advertising part in specific, firm's purpose is to spread the positive impressions of products to their customers and to attract them to buy their products. Not only advertisements but also the aforementioned knowledge-based products have similar characteristics in transmitting due to their "non-physical" form. In other words, advertising is one kind of "information spreading" procedure and the Internet is the best way to achieve it. With lower cost, higher speed, and higher external effects, marketing on the Internet has more advantages than on traditional media no matter the target products are physical products or knowledge-based virtual ones.

The effects of viral marketing become larger in this scenario. Originally, viral marketing is the "word -of-mouth" action. People tell their friends about their using experience and spread the product information accidentally. The result is amplified due to the characteristics of the Internet and firms start to pay their attentions to taking advantages of viral marketing. In fact, viral marketing is not only one kind of information spreading method

but also the inclusion of the influence of friends. The information is filtered by our friends and is more trustworthy than general advertisements. If the spreading of information can be controlled, it will be a good solution for marketing problems.

1.2 Research Problem

Although marketing on the Internet has lower cost and higher influence than traditional ways, it could be the solution to only half of the problems we stated. The Internet technology definitely lowers the cost of enterprises to advertise products but cannot ensure the effects to be really achieved. In other words, online advertisements are viewed as useless message by most of the Internet users. For example, e-mail is one of the most common channels for general advertising on the web but statistics shows that more than 95% of the emails are junk mails [55]. The high garbage rate would make most people pay less attention to handle these junks and lower the effects of advertising at the same time. In addition, over-advertising even make the customers have bad impressions to firms. No matter how low the cost of advertising is on the Internet, the resources are wasted definitely. These resources may be bandwidth, server computing power, electricity or more infrastructures in order to maintaining a large scale marketing behaviors. After all, the convenience of the Internet may also lead to negative impacts on the visibility of the information we want to spread.

Lots of researches provide many methods to solve the problem. The developing of recommendation system is an important milestone. In short, the recommendation mechanisms filter most information and only send the product information the customers may be interested in. It is one kind of one-to-one marketing and achieved complete personalization. The quality of recommendations relies on the techniques used. Generally, the purchase history and personal preference will be considered as basic materials for system input by data mining techniques such as collaborative filtering, association rules, and content-based filtering [10]

[26] [59].

Effective algorithms provide product information which only the customers may want. They are also applied in many online shopping websites now. However, the Internet environment currently creates a new chance for a new type of marketing. The advancement of IT infrastructure empowers almost everyone to contribute or to share information on the Internet. The sharing behaviors on the web are so called “Web 2.0” [36]. In Web 2.0 environment, information is no longer only sent from traditional firms to customers or organizations to individuals but also passed between every node on the Internet. Individuals can share their creations with everyone in any place, anytime. In other words, information flow is not purely as client / server structure which sends and receives data in single direction but like the peer to peer architecture (P2P) which every node in the framework can play the client or server role at the same time. The concepts of peer production [5] and social network [4] are also constructed by the power of Web 2.0.

In lots of Web 2.0 community and discussion groups, people contribute their comments after using products and find comments about products they need or want to buy. More importantly, these comments may be provided by their friends. Consumers can reach the real comments of their friends more easily than ever before, and the firms can no longer control the scattered information source about their products. It makes more sense to believe the using experience provided by someone we trust rather than to buy the firms’ advertisements. In such information spreading model, the impressions of products are decided by online users’ comments and human connections, not the advertisements. It is doubtless that people now have other channels for product information with higher credibility. They are reliable, trustworthy, close to our real using habits, and the most important is: drawbacks will no longer be hidden by fancy advertisements of manufacturers. Current Internet environment

provides us another bright path to solve the uncertainty problem of marketing.

1.3 Research Objectives

The objective of this research is to solve the uncertainty problem of marketing. We have known that viral marketing is incubated based on Web 2.0 environment since it provides new chances for everyone to express their opinions. Research have shown that social networks affect the adoption of individual innovations and products [41] [47] and the power of social network spreads information in breathtaking speed [22]. In fact, people are usually affected by the purchasing decision of their friends in daily life, especially when they need to buy something expensive. From the perspective of firms, we expected only by marketing to a few people who has the ability to spread product impressions and affect their friends efficiently can accelerate business. This strategy not only saves money but also lowers the probability of consumers' complains due to annoying advertisements. By leveraging the power of social network, enterprise can achieve amazing results in lower cost and higher accuracy by marketing to fewer potential customers (or nodes). These nodes should have plenty of purchasing experiences, contribute their using experiences often, and equip with wide social networks.

In this research, we hope to find an easy way to discover influential nodes with potentials to achieve the effects of viral marketing. How to measure the influence of each node is a very important topic because it determines which nodes are appropriate to market. Among the key factors we mentioned in last paragraph which affect people to make buying decisions in current Internet market, we consider the using experience, contributing status of opinions, and social connection are the main elements to shape the influence of each node. In addition, we found that the online product review wrote by users is an excellent source to acquire these elements. The contents and basic attributes of each review can satisfy our needs

of data. As a result, we start from analyze online product reviews and RFM indexes about individuals on the professional product review website: Epinions.com. Text mining techniques, artificial intelligence model, and trust mechanisms will also be applied. By quantifying the value of each review and author, the commercial value of online users can also be identified. Enterprises can use the information to make a good marketing strategy and budget plans in order to achieve the best effects of infection. They will know who are their valuable targets and pay more attentions to these potential nodes. The results and evaluations show that our framework for finding potential nodes is effective and better than traditional "popular author" and "review rating" mechanisms.

1.4 Research Outline

The remaining part of this paper is divided into the following sections: In section 2, we survey existed literatures about our research topic. In section 3, we propose the whole system architecture and methodologies applied in this work. Next, the procedures and materials about experiments will be stated in detail in section 4. The results and evaluations will be displayed in section 5. Finally, we have a discussion and conclusion.

CHAPTER 2 LITERATURE REVIEW

This chapter reviews literatures related to our research, including viral marketing, opinion mining, RFM, trust, and artificial neural network. These concepts and methodologies will be applied in our model's construction, experiment, and evaluation design.

2.1 Viral Marketing

Viral marketing is a new marketing method which is not based on the advertising budget of firms. It uses electronic communication channels (Ex: e-mail) to propagate brand messages throughout a widespread network of buyers [15]. In fact, this new marketing way spreads the brand impressions with no partiality no matter its brand reputation. Dobele et al. [15] realized this fact and tried to find out an appropriate approach for successful marketing. They considered viral marketing a good chance for new startup firm but most firms are lack in "control" of this power. They studied several real marketing cases and analyzed why they need viral marketing, how to apply technology in it, and how to use it successfully.

Dobele et al. [14] also identified the key points about the success or failure of message passing in viral marketing. They collected and categorized many cases about the message passing behaviors in several different products. They realized that "emotion" and "the expectation of recipient" play important roles in the successful message passing. The result shows that emotional expression during message passing is important. In addition, the authors also stated that marketing to several influential people will perform better than sending message to everyone and that is what we want to achieve.

Moore [31] did a research about the branding influence based on viral marketing environment. For example, Microsoft's hotmail member increases with high speed due to its involvement with many contacts of each user. This is a famous example in viral marketing

because Microsoft spend only $50,000 on traditional marketing channel but the members of hotmail grew from zero to 12 million in 18 months [22]. In mid-2000, Hotmail owned over 66 million members and 270 thousand of new accounts are created on a daily basis. These users send their personal messages and the “Hotmail” brand at the same time are spread. In other words, Hotmail has provided a place to allow users fulfill their needs of communication and, in exchange, it is gained its reputation in cyberspace. For business operators like Microsoft, one of the most important outcomes is chances to generate profits. Thus, treating this as an opportunity, Microsoft attaches advertisement to mails. Although the advertisements every member receives are not filtered, ads do function well to spread through links among people and achieve practical effects.

Leskovec et al. [25] analyzed information about product recommendation and discovered the relationship of social network. They proposed a model to explain user behaviors in a large community where people recommend products to others in different strength according to their social network in this community. In addition, the growth and the effectiveness of the social network are identified. Richardson and Domingos [40] used probabilistic models and data from knowledge-sharing sites to design a viral marketing plan. They also tried to optimize the amount of investment for each customer and to lower the computational cost.

We can find several works about viral marketing which is based on social networks. However, most works focus on the observation of business condition or the calculation of social network spreading. In an Internet context, the effects of viral marketing become attractive for commercial application. We focus on creating a practical model which can be applied easily on business strategy making. Our works will pay much attention on applications of information technologies to help enterprise find a good solution of marketing

strategy in viral marketing. The model is constructed on the platform of Web 2.0 environment--a place where real using experiences of online users to achieve an effective marketing behavior take place.

2.2 Opinion Mining

We have known that people can get useful information about products which they want to buy in many online communities. They read product reviews which are written by other experienced users for reference to make a decision. In other words, these reviews equipped with some influencing power to the readers' decisions and the purpose of opinion mining is measuring the influencing level of them. For example, Zhan et al. [58] emphasize the important role of writing and referring product reviews in Internet environment. These product reviews are both influential for consumers' decision and firms' customer service department to do product improvement.

Empirically, there are already some social datasets distributed on the web [49] and it is helpful in simplifying the data collecting process for opinion mining. In the case of the methodologies to implement opinion mining, many scholars focus on the identification of author's attitude such as positive or negative [13] [18] [19] [37]. The semantic tendency of an article is usually decided by some specific keywords which are clear and hard to be misinterpreted. The semantic identification is helpful for review tendency judging automatically.

In this research, we intend to apply these techniques to do a detailed scoring for the value of each review. Since modified- and multi-dimensional scoring mechanisms are relevant to our problems, in this research, we focus on identifying the critical review with subjective semantics. By identifying these reviews, they are useful in understanding consumers' behavior.

2.3 RFM Measure

Hughes [20] proposed RFM analytical model in 1994. RFM stands for Recency, Frequency, and Monetary. It is a way to measure the values of customers for enterprises. For every customer, enterprises can get the three elements from his / her purchase history [39]:

Recency: What time is the last purchase of this customer or how long from last purchase to now?

Frequency: How many times of purchase happened during a specific time period?

Monetary: How much the customer spends on each purchase?

By RFM analysis, firms can understand the potential of customers easily by observing their past behaviors. Newell [35] also stated that RFM method is very effective in customer segmentation. So, the simple and direct this measure has been used in direct marketing for a decade [3]. It intends to find customers who recently purchase (Recency), the number of times they purchase (Frequency), and by how much they spend (Monetary) [30].

There are many researches based on the concept of RFM analysis. Drozdenko and Drake [16] applied the hard coding techniques on RFM weighting. They assigned weights to three variables in RFM analysis and acquired the weighted score of each person in database. This technique is also called as “judgment based RFM” due to the procedures are functions of judgments of marketers. Chan [7] presented a novel approach that combines customer targeting and customer segmentation for campaign strategies. RFM and customer life time value (LTV) are included in his model to identify and evaluate customer behaviors. Liu and Shih [28] proposed two hybrid methods which take advantage of weighted-RFM (WRFM) method and preference-based CF method to improve the accuracy of recommendations. Although RFM is not an innovative method to identify values of customers, the extensive applications have proved its importance in academic ground.

In this research, RFM will be used to derive partial values of online reviewers. Modifications will be made to accommodate data. In the end, RFM will be used to measure the potential value of customers by their past behaviors. Moreover, online product reviewers also have similar characteristics. According to the records of their writings, we can understand that these reviewers are currently active or not. Lastly, the application of RFM can be realized as the way to choose a good reviewer but what it cannot do is to create new information people need.

2.4 Trust Mechanism

Trust is a relationship of reliance [48] and is also a willingness to rely on an exchange partner in whom one has confidence [32]. Erikson [17] defines trust as "general belief in the goodness of others"; Rotter [42] describes trust as "an expectancy held by an individual or a group that the word, promise, verbal, or written statement of another individual or group can be relied upon". In brief, trust is an expectancy that the behaviors of people (or objects of trust evaluation) will follow a predetermined manner [1] and this manner is the behaviors of others they trust.

Trust can also be used to indicate the strength level of relationships among people without doing detailed investigation of intention [44]. The strength level of trust can be viewed as an indicator that the probability people will follow the behaviors of someone they trust or not. Due to the characteristics of trust, it is often used to create better working efficiency in organizations. Munns [34] stated that trust is a relation from personal to individual, arising from the experiences of and influences on that individual. Strong trust level of someone will shape a strong influence making others trust him / her. Some researches have indicated the effectiveness of trust mechanism and its implications in different academic subjects.

Trust has been described as "central to all transactions" between individual or organizational actors in economics [12]. It strengthens the motivations of people to do transactions and the benefits of each transacting target can be evaluated accordingly. Morgan and Hunt [33] theorized, modeled, and tested the success of relationship marketing and found that commitment and trust are key factors. The study of Cook and Wall [11] classifies trust between peers and trust between peers and management. They found that trust is based on "faith in intentions" and "confidence in ability". Smith and Barclay [45] studied the relationships between buyers and sellers. Their results show that trust is based on character / motives / intentions and role competence / judgment. Trust value of a supplier also influences a buyer's future interacting will with the supplier.

The concept of trust and related algorithms clarifies the intimate level between nodes or organizations. Dasgupta [12] stated that trust is helpful in the condition involving uncertainty about the actions that will be undertaken by others. Stated differently, trust mechanism is an important and effective factor for customers to make purchase decisions even customers may not be familiar with the product. There must be a higher probability to follow the recommendation from people with high trust score than the advertisements of firms. Due to this fact, the calculation of trust score of potential nodes is clearly justified. We will use this concept as the evaluation indicators to reflect the effects of our model.

2.5 Artificial Neural Network

Artificial neural network (ANN) is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionistic approach to computation [2]. The purpose of ANN is to construct an artificial model which can learn and think with a mode similar to the brain of human. ANN is appropriate for solving complex problems which includes many variables. A simple structure

of ANN is similar to Figure 2.1. By continuous lots of times of “training” and “learning”, a well-trained ANN will change its structure according to external and internal information that flows through the network. So, ANN is an adaptive and intelligent system which can vary to fit the users’ needs according to the characteristics of training data. A well-developed ANN is also expected to generate usable results from input data by following existed learning rules. Therefore, ANN is of great importance to predicate in many research areas.

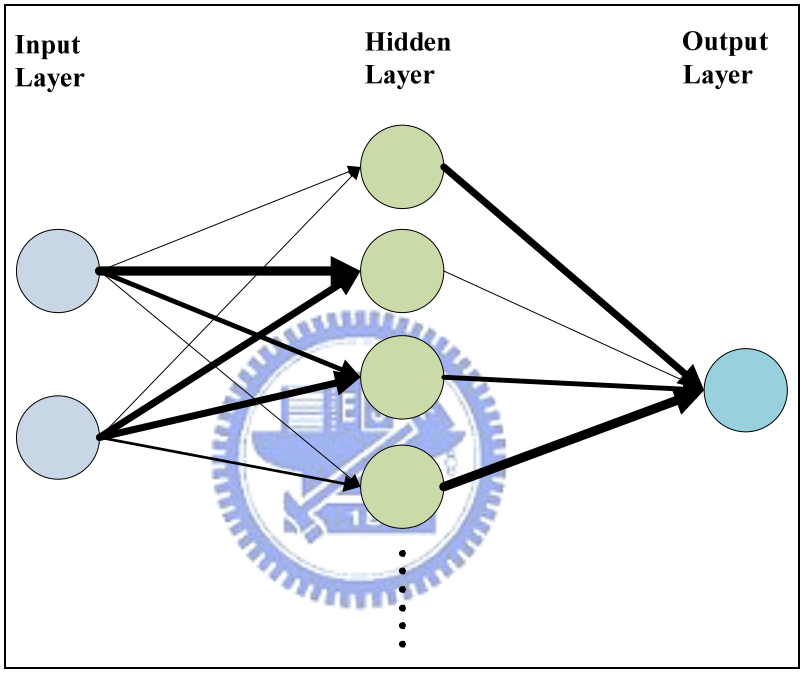


Figure 2.1 A neural network

ANN technique has been used for solving business problems extensively and can be judged as an element of business intelligence. Kuo and Chen [24] used fuzzy neural network to learn rules produced from order selection questionnaires in electronic commerce. A feed-forward ANN with error back-propagation learning algorithm is also employed to integrate different scores. Cao and Schniederjans [6] created an ANN model for a reputation agent to evaluate capabilities for selecting products and services in an e-tourism environment. Chiang et al. [9] developed an ANN model to predict and explain consumer’s choice between web and physical stores. Tsaih et al. [50] combined rule-based systems and ANN to predict the

direction of daily price changes in S&P 500 stock index futures. Li et al. [27] used ANN model and other statistical methods to forecast the final price of online auction items.

The existing works have proven the effects of ANN in solving various complex problems. In our research, we need to acquire the predicted influence score from several factors such as text mining, Recency, and Frequency. The relationships among these factors are complex and many human behaviors cannot be indicated by linear models. These kinds of complex materials are appropriate to be constructed by massive data training and learning in ANN.

CHAPTER 3 THE MODEL

In this research, we analyzed the after-use reviews provided by online users and RFM values in each author's activity recorded to identify which authors are influential. While the influential reviews represent the influence of their authors, the RFM value indicates the infective ability of each reviewer by time segmentation.

An influence ranking list of authors is generated to identify potential nodes and it is expected to construct a well-learned model in order to calculate each reviewer's mixed-score of two elements above. Data which contained complete review content and RFM attributes are needed for well-structured model training. Artificial neural network technique is then applied to achieve the training procedure for better weight measurement among these elements. The well-trained network model will be fed with selected testing nodes and the output ranking list will be compared with two existing common methods for selecting influential authors.

From enterprises' perspective, these high-ranking authors are valuable targets for their marketing behaviors. They are expected to spread products' reviews. And the impact is profound. Firms can have set up strategies to take advantage of these potential reviewers. They can provide free trial version of their newest products or special discounts to these targets in order to induce the consumption, which entails influential review. It is hoped that the behavior of spreading of product behaves as a virus. Figure 3.1 displays this concept and its architectures:

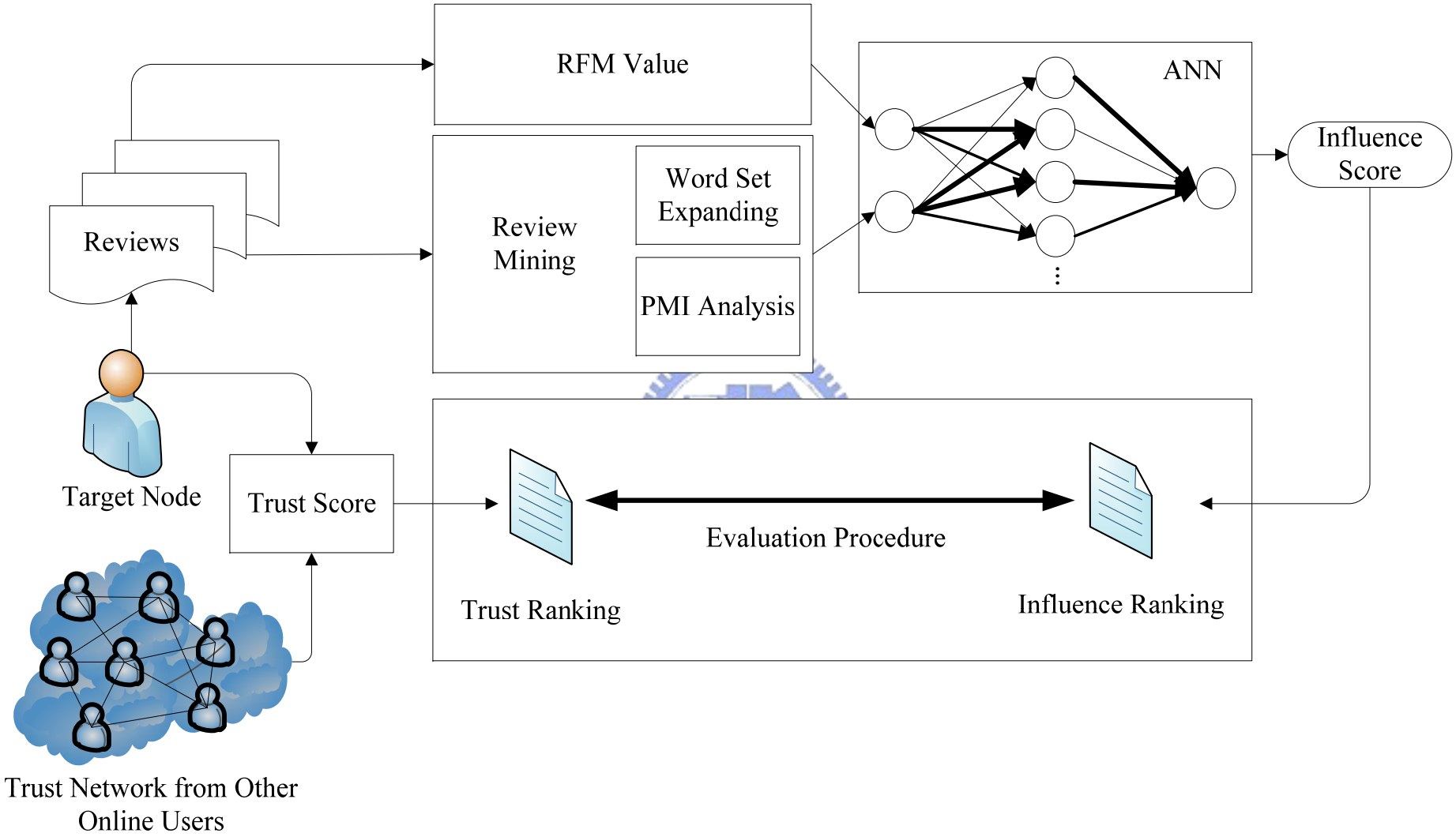


Figure 3.1 System concept and architecture

Our target nodes are chosen from an online social network environment. Online discussion platform provides users to write product reviews of any kinds. What we try to do is to assign scores to these reviewers and to decide which reviewers are the most infective to the market. Note that the infective ability is decided by two factors: reviews and RFM value.

The reviews written by each reviewer will be analyzed by text mining techniques based on their scores. The results of analysis will be quantified by our modified PMI model in six different degrees. In addition, we can also acquire the "RFM value" of each node by recording attributes of each review (i.e. time, date, and category). The both scores will be weighted by artificial neural network as the final virus score to decide the value of the reviewer. It will learn the most appropriate structure of network to reflect the effects of each element by massive data training. Our mechanism can consider both text mining methodologies and the effectiveness of each node's writings. It tries to discover the hidden value in each review and consider verifying the effect of it at the same time. We do a detail statement about each unit in this architecture in the following subsections.

3.1 Review Mining

3.1.1 Word Set Expanding

In existing research, the semantics of single article are usually classified in four categories: positive, negative, objective, and subjective. Since the reviews are not evaluated based on these classification, thus, we will not follow previous work. Instead, we focus on general attitude from reviewers and they are the nodes we want to be discovered. However, previous work indicates that reviewers who write at extreme values (e.g. all positive (hyper spam) or all negative (defaming spam) comments) are hard to be trusted [21]. In addition, objective perspectives are usually descriptions about products without additional reference value are not considered important because they are lack of "emotions," which has higher

possibility to affect the purchasing decisions of others [14]. Thus, in this study, the subjective factors are considered.

Turney and Littman [51] defined two sets of words which represent positive and negative sentiments, respectively:

$S_p = \{\text{good, nice, excellent, positive, fortunate, correct, superior}\}$
$S_n = \{\text{bad, nasty, poor, negative, unfortunate, wrong, inferior}\}$

The two sets are decided due to their lack of sensitivity to context. It means that in most situations, articles with these words represent the original meaning no matter how the structure of these articles was. Our model was analyzed based on these two sets.

We expand S_{p+n} in order to make a subjective word base. In order to consider the subjective reviews, both word sets will be included in our model. We define:

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

The composite word set S_{p+n} is the combination of the two above sets so it covers both positive and negative semantics. We believe that the ingredients of S_{p+n} will carry a complete meaning of "subjective words" which can reach our expectations. Although reviewers can express their emotion completely through subjective comments, which are valuable for other online users, firms would hope not to see any negative comments spreading on the Internet. Nevertheless, a trustable reviewer should always express their thinking fairly. A reviewer who only write positive reviews is possible to be viewed as an employee of the firm and his / her reviews will not be trusted by others. In order to achieve effective information spreading, it is necessary to select trustable reviewers. Firms should try doing the right things to please these influential reviewers who write negative comments instead

ignoring them. Any negative comments will hurt the reputations of firms, especially in the Internet world. Firms should promptly respond to negative comments, contact those reviewers, and find out what can be done to improve the products or services. This will not only remove the sources of bad reputations but also increase positive reputations of firms.

We hope to check if these words of S_{p+n} exist in each review to decide the subjective level of each review, but the number of words in the set is too few to do an accurate check. It may ends up with none being discovered. The problem can be solved by expanding original S_{p+n} set. Some online semantic lexicon such as WordNet [57] would be helpful. We plan to extract synonyms of S_{p+n} from WordNet to achieve different level of S_{p+n}. We have known that the synonyms can be traced from S_{p+n} recursively, so the size of word set could be different according to the iteration times.

We mark the word set S_{p+n}^{k} which denotes k th expansion times. For example, $k=1$ equals to original 14 items in S_{p+n}^{1} and $k=2$ equals to 14 items in S_{p+n}^{2}. The sets will grow rapidly according to k value and the number of matches will also increase due to larger word set S_{p+n}^{k}. Clearly, different value of k will lead to different matching pairs between S_{p+n}^{k} and test reviews. In our experiment, six degrees of word set expansion will be executed to observe a better expanding level. As k value becomes larger, the system will consume more resource in word matching and the whole system becomes inefficient. The six levels of word matching will be recorded and quantified in the following PMI method. The scores are used to calculate the strength of subjective of these reviews.

3.1.2 PMI Strength Level Approach

In this subsection, we use PMI (Pointwise Mutual Information) as a tool to calculate the score of strength of each review as the basis for the results of review analysis. Turney and Littman [51] define PMI in the following equation:

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

This equation can measure the semantic association between the matched term t in a review and t_i in word set S_{p+n}^{k} by calculating the emerging probability in the whole article. The key point of PMI calculation is the value of $\Pr_c\left(t,t_i\right)$, $\Pr_r\left(t\right)$, and $\Pr_w\left(t_i\right)$. We define each of them as follows:

$$\Pr_r\left(t\right)=\frac{n_{tr}}{N_r}$$

$$\Pr_w\left(t_i\right)=\frac{1}{N_s}$$

$\Pr_c\left(t,t_i\right)=1$ (i.e. term t and t_i are the same word.)

Term n_{tr} stands for the number of term t (i.e. number of matches) in target's review and N_r stands for the number of all words in this review. N_s represents the number of words in the word set while t_i was collected into it. In fact, t and t_i are the same due to the matching mechanism. Thus, the association between these two words is not our real purpose. The real effect of PMI is that it considers the number of matched words in the whole article before it reflects the subjective strength level of the target. In addition, PMI also takes the word appearance probability in each review and decreases the errors due to unequal number

of words in each review. In order to simplify the calculation and retrieve appropriate value for processing, we modify PMI in the following form:

$$PMI\left(t,t_i\right)=\log_2\left[\Pr_r\left(t\right)\Pr_w\left(t_i\right)\right]$$

The PMI score of each review is calculated by the adaptive PMI equation above. It is based on the viewpoint of each review. Every time the model processes the score of one review, $PMI(t,t_i)$ considers all matches in it, meaning that every review in our data set will have its PMI score and it comes from the sum of its all matches. It is obvious that this equation will produce a negative score and it is inconvenient for continued processing. Because of the negative characteristic of each PMI value, we will standardize every PMI score from every review before combining with other character values:

$$PMI_{i\_std}=\frac{PMI_i-PMI_{\min}}{PMI_{\max}-PMI_{\min}}$$

The modified and standardized PMI equation can help acquire the strength score of each review. Then, we can acquire the target node's score by:

$$PMI_{Avg}=\frac{\sum_{i=1}^{n}PMI_{i\_std}}{n}$$

This equation sums and averages all reviews' PMI score of every author. This step is necessary to construct a comparative standard for all authors. After the processing, every author's text results in our data set can be identified, recorded, and ranked.

Now, we have shown how to calculate reviewers' score of the target node. The review analysis scores of these target nodes are important elements when considering their influence. They will be combined and processed with other scores by later weighting mechanisms.

3.2 RFM Value

In this research, we accommodated original RFM concepts into our experimental situation. Recency and Frequency indexes are adopted in our RFM analysis only due to the characteristics of online product reviews. Monetary value is excluded because of its only inappropriateness and difficulty to measure. In the Web 2.0 environment, information contributors provide knowledge to the Internet voluntarily and their efforts have no direct relationships to pecuniary revenues.

3.2.1 Recency Model

The original concept of "Recency" is the days between the purchased date and the presence. In our work, we explain "Recency" 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.

$$\gamma_i = C - l_i$$

While l_i is the last written date of node i, C is the current date. Initial values of Recency are measured by days. They are needed to be standardized in order to combine other index values later. Recency standardization is different from general standardization procedures because higher values indicate lower market values. In order to display real meaning of Recency, the following formula to standardize Recency value:

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

The absolute value between γ_i and $\max \gamma_i$ indicate the strength of lower Recency with higher standardized value.

3.2.2 Frequency Model

Frequency represents the purchasing times in a specific time range. Similar definition is applied in our work. It indicates the number of writings in a specific time range of each author. To segment the time range is the first step to record frequency. We set three time points as the separation:

$\theta_{<90}$: The number of writings made within 90 days
$\theta_{90\text{-}365}$: The number of writings made between 90 and 365 days
$\theta_{>365}$: The number of writings made over 365 days

The three-time points are calculated. In addition, most data are electronic products and the product life cycles are shorter than general products. Divided the time range by quarter and one year at most will be appropriate Each author's writing records will be classified into the three categories and they are still in need to be standardized by the same method. However, we do not combine scores of these three categories into single Frequency score. It is apparent that writing reviews in different time range represent different level of importance and it will be reflected on the influence score of each reviewer. We put them into the ANN model for weighting because linear or static weighting of these three scores cannot represent real scenarios in life. The effects of ANN will learn appropriate distribution among the three scores automatically and the pre-combination can be ignored.

3.3 Trust Network Value Calculation

Although trust is effective in influencing the purchasing decision of others, we do not adopt it directly as an element of our model. One reason is most of the online product comment communities do not consider trust mechanisms in their website design thus make it

difficult to apply a trust-embedded system for firms to identify influential nodes. The other reason is that even online product websites do equipped with complete trust scoring mechanism, they are only useful for experienced nodes because these nodes have wider social networks for trust measuring. In order to increase the effectiveness of our model, we have added RFM measure as the factor to judge the active status of each node, and the characteristics of trust may filter some new but potential nodes out. However, many literatures indicate that the relationships between trust and influence are very tight. If nodes discovered by our model also have high trust value, the effectiveness of our model can be acceptable.

Due to the characteristics of trust, in this research, trust value will be applied in evaluation procedure. The spirit of viral marketing is to leverage the elements of social network. Thus, the output of our model must have similar effects to social networks: the power of spreading. Equipped with this character, trust value evaluation will reflect the human relationships of social network, and the effects of our model can be verified. We believe that the trust relationships among these reviewers can add more benefits to product information spreading. This architecture design of our model can make our results have strong influence, which is verified by trust, and solve bias problem for new reviewers.

We have found that many online discussion forums assign impression scores to others. Every online user can establish his / her personal friend list and black list. This procedure indicates each user's trust level which may be different among users. We will use these lists to consider the propagating ability of each user.

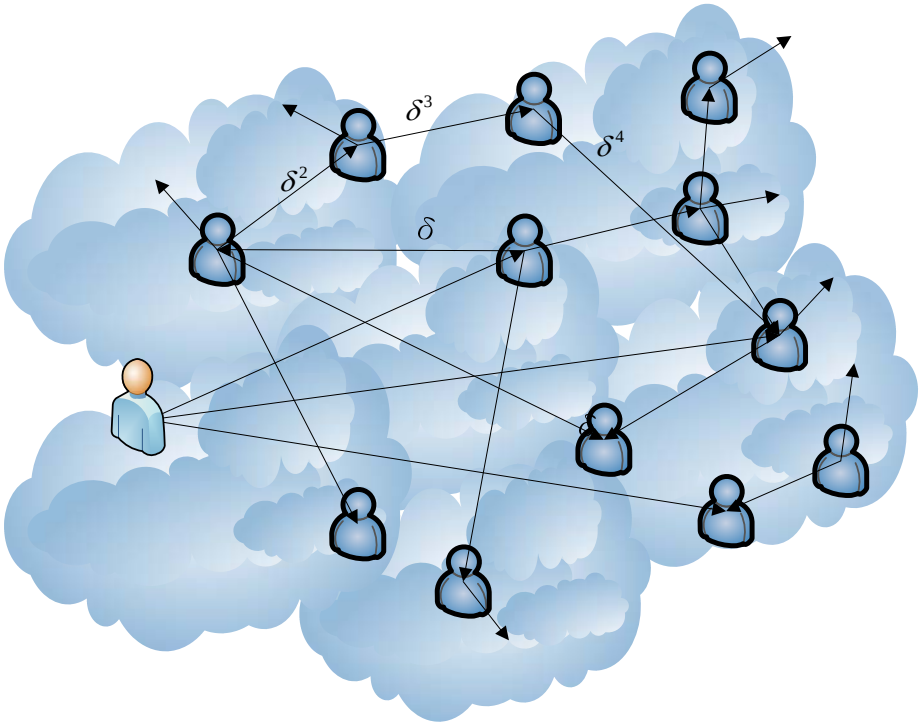


Figure 3.2 Trust name list expanding relations

Figure 3.2 displays the trust relationships among online users. For example, the target node (reviewer) is trusted by three people and his name exists in the three people's friend lists. The names of the three people would also exist in other friends' lists. The relationship can be traced more according to our need. By iterated tracing actions, we can observe everyone's social connections within the online community. The social connections represent the influencing range of each node. We call it "SCN (Social Connection Number)". If we do review mining to discover the influencing strength of a reviewer, we can measure the influential range. In brief, the tracing of social connection of each target node is our primary task and it can be formulated by the following equation:

$$SCN_i = n_i + \sum_{j \in tr_i} \delta SCN_j$$

SCN_i starts from any reviewer i in our data set. n_i is the function that counts the number of people denoted as trust node i. tr_i is the set of nodes who trust node i. We know the strength of connection would decrease as the tracing level increases. δ is added to represent the decay rate of SCN. Note that the repeated connections would happen in different tracing level because we may connect to our friends via different path. Redundant connection numbers are erased if the nodes already exist in previous tracing procedures.

Figure 3.2 shows that the tracing of SCN is a large recursive process. Friend nodes will increase at a fast speed. In order to reduce computing time and to keep the process efficient, the setting of tracing level becomes a critical topic. We explain this part in CHAPTER 4.

CHAPTER 4 EXPERIMENTS

4.1 The Data

4.1.1 Data Source

We use real data from the Internet in this experiment. In order to acquire objective results, we need to find an open platform where online users can write reviews to various products. Epinions.com (Figure 4.1) is a good platform which can satisfy our need. It not only it has a variety of product reviews but also provides a complete set of writing, rating, ranking, and trust mechanism for members to identify the effect of reviewers. The characteristics of Epinions.com are appropriate for us to retrieve related data, especially for our trust value retrieval. In addition, lots of researches have used Epinions.com as their experimenting data source [8] [29] [52] [54] so the stability and objectivity of data would not bias.

Figure 4.1 Epinions.com website

In order to acquire sufficient data for neural network training and testing, we need to prepare two independent data sources. The retrieval date of these reviews is on May 14, 2008. 2952 reviews are randomly selected from seven sub-categories under classification Electronics. They are “Home Audio”, “Video”, “Communications”, “Car Audio”, “Optics”, “Outdoor Electronics”, and “PDA & Handhelds”. The authors of these reviews all have SCN value larger than zero. The only remaining sub-category under Electronics is “Cameras and Accessories” and we will extract the testing data from it.

The selection of testing data goes through several steps. We need to have a data set which started form the reviewers’ angle due to our ranking purpose. We picked up the product “Canon PowerShot S5 IS Digital Camera” from “Cameras and Accessories” sub-category, and it has 69 different consumer reviews. All reviews wrote by the 69 reviewers are retrieved but only the latest 100 reviews of each reviewer are reserved. In addition, if the reviewers are not trusted by anyone, they are erased from our data set because we need the trust value of them to evaluate our system.

4.1.2 Data Descriptions

In above section, 941 reviews wrote by 69 reviewers are retrieved but only 16 people are reserved due to the consideration of trust. We find that 715 out of the 941 reviews are written by the 16 people. In other words, 23.19% reviewers create 75.98% reviews and they all have some level of trust. This situation can reflect the Pareto principle (or 80 / 20 principle) and prove trust is really a significant indicator of the influencing power again. The 715 reviews wrote by 16 reviewers are the confirmed testing data in our experiment. They are all wrote between December 28th, 1999 and April 30th, 2008. These reviews are classified into 16 categories and the distributions are displayed in Table 4.1, Table 4.2, Figure 4.2, and Figure 4.3.

Table 4.1 The distribution of reviews in different category

Categories	*Number of Reviews*
Electronics	276
Computers & Internet	108
Media	84
Sports & Outdoors	57
Home and Garden	51
Kids & Family	33
Hotels & Travel	22
Cars & Motorsports	17
Restaurants & Gourmet	16
Wellness & Beauty	15
Business & Technology	12
Web Sites & Internet Services	8
Games	8
Gifts	3
Personal Finance	3
Others	2

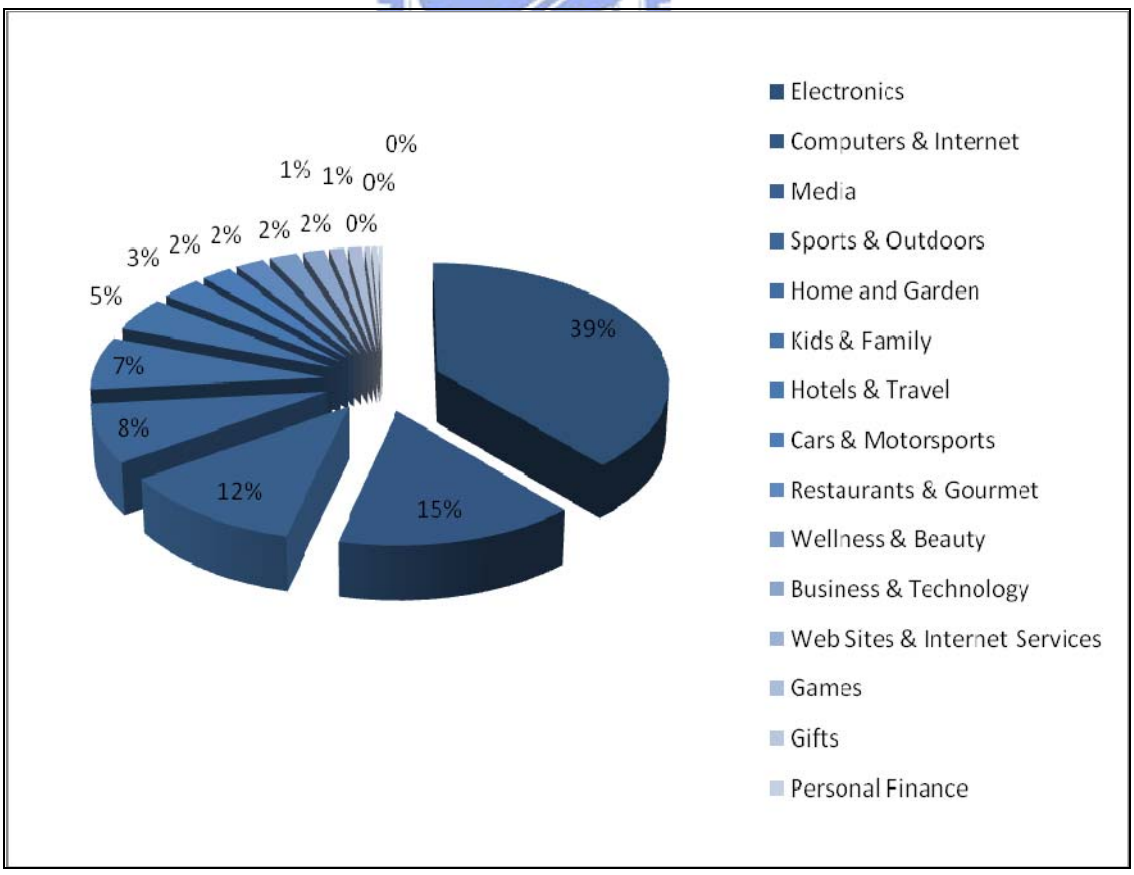


Figure 4.2 The distribution of reviews in different category

As Table 4.1 and Figure 4.2 shows, the testing reviews are centralized in “Electronics”, “Computer & Internet”, and “Media”. These three parts represent two-thirds of all reviews. We think the situation is due to our starting point in collecting data. In other words, the online platform for writing product reviews is constructed by computer and networks. It is reasonable that most users are familiar with products related to electronics and information technology. The centralized data distribution also has some advantages. It makes the main character of data can be identified easily and the applying range of our model is also cleared.

Table 4.2 The distribution of authors in different category

Categories	*Number of Authors*
Business & Technology	6
Cars & Motorsports	9
Computers & Internet	12
Electronics	16
Games	3
Gifts	3
Home and Garden	9
Hotels & Travel	8
Kids & Family	7
Media	10
Personal Finance	2
Restaurants & Gourmet	4
Sports & Outdoors	5
Web Sites & Internet Services	5
Wellness & Beauty	7
Others	2

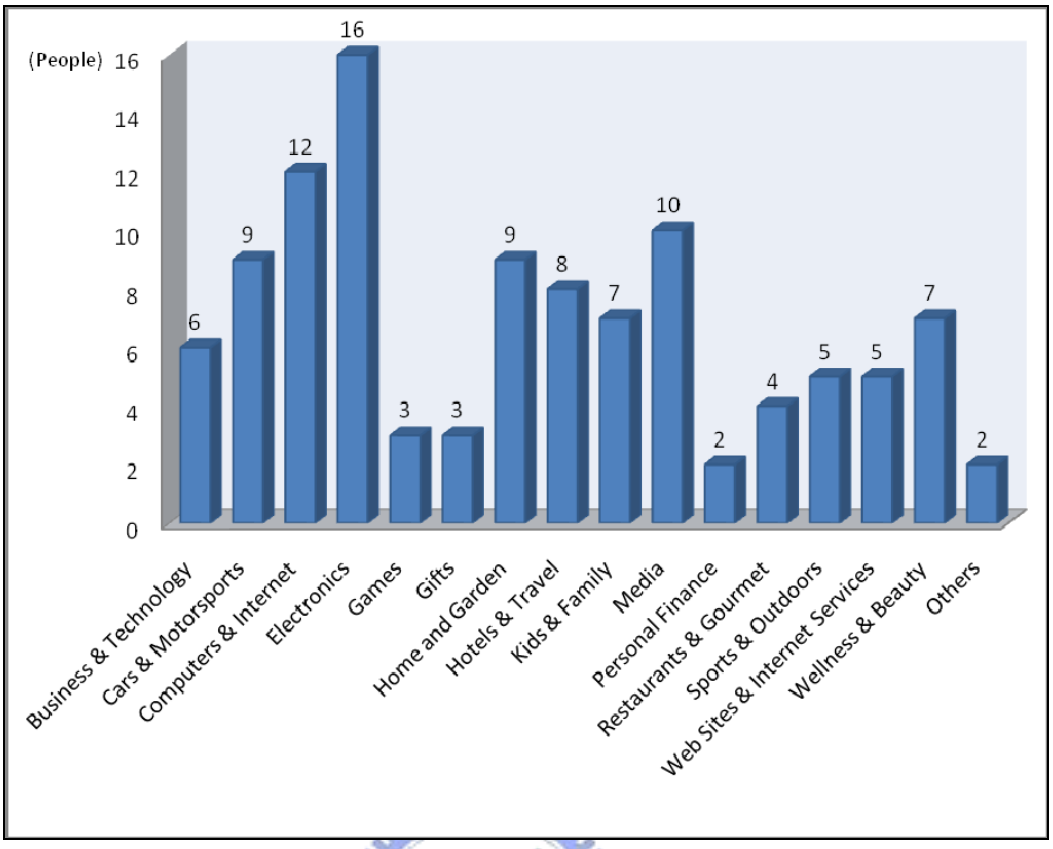


Figure 4.3 The distribution of authors in different category

Table 4.2 and Figure 4.3 state similar scenario to Table 4.1 and Figure 4.2, but it starts from the viewpoint of 16 testing reviewers. Figure 4.3 shows that how many people ever wrote reviews in each category. We can still observe that most of our 16 reviewers wrote reviews in several main categories similar to Figure 4.2. The above two figures indicates the tendency of our data set. Analysis shows that the characteristic of our reviews and reviewers are skilled mainly in electronics, computers, and media. Our experiment results are also expected to have related value for these categories.

4.2 Word Set Expansion

Training and testing reviews retrieved from Epinions.com are expected to be matched with the word set S_{p+n}^{k}. In order to carry out subjective word matching in different level, we need to expand original word set S_{p+n}^{1} gradually. We use WordNet as the benchmark to find

out the synonyms of original word set. WordNet has defined all related subjective words and listed their synonyms clearly. Many opened Application Programming Interfaces (APIs) of WordNet are helpful for simplifying this step. JWordnet interface [23] is chosen for the implementation of our synonyms retrieval.

First, our target is the adjective classification in WordNet because the contents of S_{p+n}^1 are full of adjectives and it is the most appropriate category to indicate the tendency or semantics of a review author. Next, the JWordnet API will help us extract all synonyms of S_{p+n}^1 from WordNet word base and add these words into S_{p+n}^1 to generate S_{p+n}^2. Of course, duplicate words will be ignored. The pseudo codes are listed as Figure 4.4:

```
Set S = S^1_{p+n} ;
Synset [] Wordset = new Synset;

for (S.next) {
//Feed each word in S^1_{p+n} into the loop

      String word = dictionary.lookupIndexWord(ADJ, S.i);
      /*look up the adjective synonyms of each input word from
        dictionary*/

            if (S.content != word && Wordset.content != word){
                  Wordset.add(word);
                  /*if the word is not duplicate, add the synonyms into a
                    temporary array.*/
            }
}

S.add(Wordset[j]);
//Add synonym array into S^1_{p+n} to generate S^2_{p+n} .
```

Figure 4.4 Pseudo codes of word set expanding

We construct six word set expansion levels for later word matching (i.e. Set k from 1 to 6). We plan to compare the PMI result of the six level word sets to decide which level is appropriate for our experiment. The expansion results are listed in Table 4.3.

Table 4.3 Word set expansion results

k value	1	2	3	4	5	6
Number of words in S_{p+n}^{k}	14	142	578	1241	2148	3223

It is obvious that the growing of synonyms becomes smooth gradually. The main reason is the number of duplicate words also grows larger as the k value increasing. While k becomes larger, most synonyms are erased because they already exist in S_{p+n}^{k}.

4.3 Word Matching

Word set expanding level (or k value) is the key control factor in this process. In this experiment, 920,137 words of the 715 reviews are processed. All reviews will have six different matching results, but they will be compared under the same k value only to maintain the accuracy of experiment. The executing procedures are listed in the following:

(1) Keep original word set.

(2) Execute key words matching to all words in the word set with all reviews in our data source.

(3) Recording matching counts and calculated the PMI strength score according to the model in section 3.1.2.

(4) The review score of each reviewer will be averaged and recorded in a data set.

(5) Increase k value and repeat step(2)-(4) again to acquire scores of different levels.

(6) The best score set under specific k value will pick up for upcoming process.

It is noticeable that the scoring mechanism transferred the ranking angle from “articles” to “reviewers” in step (4). This processing is necessary for our experimenting purpose because an influencing ranking list of reviewers is expected in the end.

4.4 RFM Score

Time attribute of each review is needed for the calculation of Recency and Frequency value. It is convenient that the two indicators are both based on the reviewer's viewpoint originally. The standardized Recency and Frequency value are displayed in Table 4.4 and Table 4.5.

Table 4.4 Recency value of testing reviewers

Reviewer ID	*Recency (days)*	*Standardize*
ASourdough4	20	1.000
AtlantaGreg	94	0.951
corona79	68	0.968
dkozin	35	0.990
Howard_Creech	50	0.980
hwz1	890	0.419
JIMILAGRO	1418	0.067
jvolzer	97	0.949
njpoteri	1518	0.000
porcupine1	91	0.953
readsteca	121	0.933
sarahrose12	69	0.967
theheidis	232	0.858
tucknroll	851	0.445
williamrender	1484	0.023
zan720	1079	0.293

Table 4.5 Frequency value of testing reviewers

ID / Period	*<90 days*	*90-365 days*	*>365 days*	*Total*
ASourdough4	0.090	0.260	0.650	1
AtlantaGreg	0.013	0.026	0.962	1
corona79	1.000	0.000	0.000	1
dkozin	0.207	0.272	0.522	1
Howard_Creech	0.030	0.080	0.890	1
hwz1	0.000	0.000	1.000	1
JIMILAGRO	0.000	0.000	1.000	1
jvolzer	0.016	0.339	0.645	1
njpoteri	0.000	0.000	1.000	1
porcupine1	0.050	0.350	0.600	1
readsteca	0.000	0.234	0.766	1
sarahrose12	0.083	0.000	0.917	1
theheidis	0.000	0.067	0.933	1
tucknroll	0.000	0.000	1.000	1
williamrender	0.000	0.000	1.000	1
zan720	0.000	0.000	1.000	1

Preliminary analysis of RFM reveals that large differences exist among these reviewers' publication. Some reviewers write reviews continuously. They have low Recency and similar value in three blocks of Frequency. However, there are also reviewers have not been writing for a long time and all publications are centralized in number of years ago. We think these characteristics would be helpful for identifying influential nodes in later processing because there are fewer vague scenarios in their behaviors

4.5 Trust Score

We have found that Epinions.com has equipped with a complete trust scoring mechanism for their users. Users can create their friend and black list to show whether they trust someone or not. In addition, this information is opened for everyone to query. By checking users' profile, we can know specifically whom the user trusts and who trusts this user. These data are helpful for our trust score computation.

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 pseudo codes for recursive SCN computation is showed in Figure 4.5:

```
int k; //The tracing depth

long rec(long N_k ){    //The target node is trusted by N people in level k
      if (k<2 || N_k ==0){
      return dup( N_k );
      /*If no one trusts the node or the relationship ends in level 1,
        return the number of people trust the target node in level 1.
        Redundant connections are ignored by dup().
      */
      }

      else{
            return δ^(k-1) *dup( N_k )+ rec( N_(k-1) );
      //Return cumulative SCN and erase redundant connections.
      }
}
```

Figure 4.5 Recursive SCN computation

After retrieving each node's social relationships of friends of friends, we found the growing of social networks is really amazing. In this experiment, we set the decay rate $\delta = 0.8$. While $k = 1$, SCN just indicates each node's number of friends, and there are only about two thousand relationships among the nodes. However, while $k = 2$, the number of relationships grows to about 40 thousand. Even nodes have lower SCN in level 1 exceed higher ones in level 2. We use $k = 2$ as the SCN expanding level to acquire the trust ranking of testing reviewers because we focus on the relative rankings of these reviewers. As Table

4.6 shows, the trust ranking stabilizes in level 2 and larger SCN in deeper level is not effective in our model. In addition, people usually have friendships to “friends of their friends”. Deeper relationship hardly exists in real world. The depth setting is sufficient and well-displayed the interpersonal relationship in real life. The large amounts of social connections construct a powerful influence to others. The SCN results are displayed in Table 4.6. Standardized SCN of training and testing data will be applied in neural network training and final evaluation.

Table 4.6 SCN of testing reviewers

Id / SCN	*k=1*	*Ranking*	*k=2*	*Ranking*	*k=3*	*Ranking*
ASourdough4	137	4	5786.6	4	24580.2	4
AtlantaGreg	4	5	199.2	5	3346.72	5
corona79	1	10	179.4	6	2749.64	6
dkozin	393	3	7390.6	3	27545.48	3
Howard_Creech	804	1	11418.4	1	36640.16	1
hwz1	759	2	10986.2	2	33142.36	2
JIMILAGRO	1	10	9	10	187.56	10
jvolzer	4	5	70.4	8	1430.4	8
njpoteri	1	10	1	14	1	14
porcupine1	1	10	1	14	1	14
readsteca	1	10	1.8	13	1.8	13
sarahrose12	2	8	3.6	11	9.36	11
theheidis	2	8	43.6	9	492.88	9
tucknroll	1	10	2.6	12	5.16	12
williamrender	1	10	1	14	1	14
zan720	3	7	160.6	7	2279	7

4.6 Artificial Neural Network Training

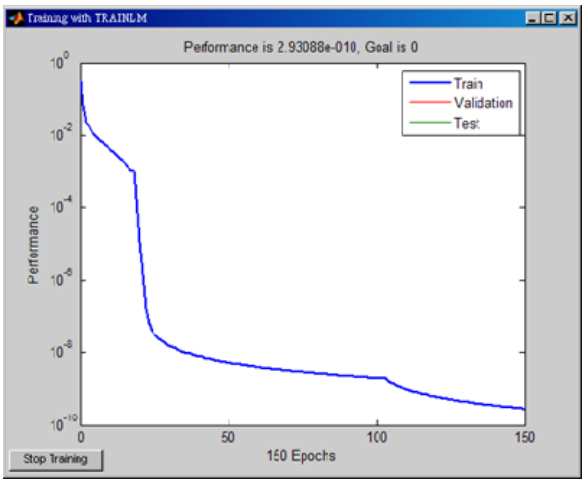
We apply artificial neural network to build a suitable model for reviewers’ rank prediction. The NNTool of MATLAB 2006 is used in this experiment for neural network constructing, modeling, training, and testing. All network types are Feed-forward back-propagation. The network is composed of three layers: input layer, hidden layer, and output

layer. Input layer have three neurons: PMI, Recency, and Frequency. Hidden layer is the main section to generate different network structure. We apply 20, 30, 40, 50, and 60 neurons respectively in this layer to observe which network structure is most fit in this experiment. Output layer only has one neuron to generate the predicted result, and it will be compared with the standardized trust score for network training. Table 4.7 lists parameters applied in network training:

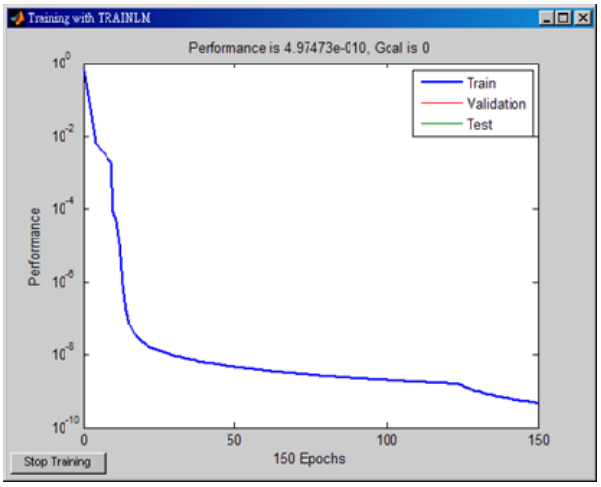
Table 4.7 Parameters for neural network training

Parameters	*Value*
Network type	Feed-forward back-propagation
Number of neurons in hidden layer	20, 30, 40, 50, 60
Training function	TRAINLM
Performance function	MSE
Epochs	150
goal	0
Mu	0.001
Mu_dec	0.1
Mu_inc	10
Mu_max	10,000,000,000
Max_fail	5

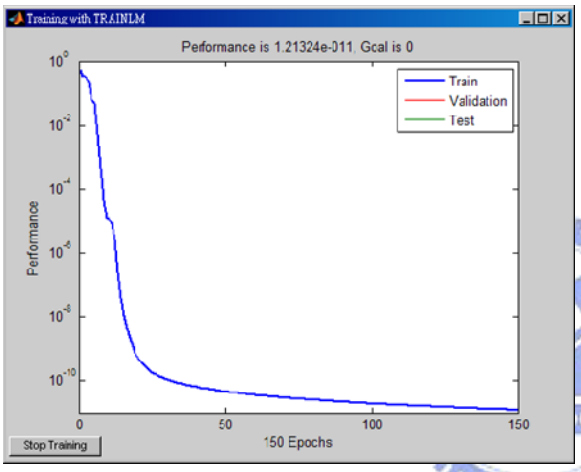
In Table 4.7, epoch stands for the number of learning cycle in each network. After several trials and errors, we found that 150 is an appropriate number because the MSE values have been close to 0. Larger epoch value does not provide apparent effects. Mu is the adaptive learning rate which changes its value according to the variation of MSE. Figure 4.6 shows the training results in different network structure:



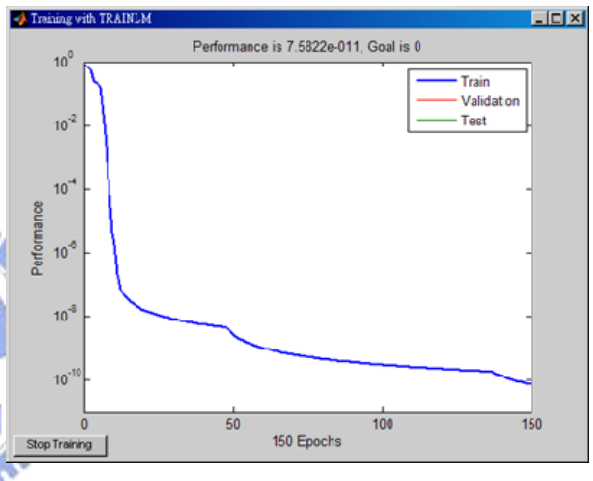
Training with 30 neurons in hidden layer



Training with 40 neurons in hidden layer



Training with 50 neurons in hidden layer



Training with 60 neurons in hidden layer

Figure 4.6 ANN Training in different structures

We found that while neurons increase, the performance functions (MSE) may decrease or increase within a tiny range. But network with 50 neurons provides lowest MSE when epoch reaches 150. It reduces the error rate with shorter time than other network structures. We choose this network model for data testing.

Neural network predicting model can be established after sufficient training of different network models. Then we input our 16 testing reviewers' data into the well-trained model to acquire the predicted trust values. The output values will be used to decide the influential ranks of them.

CHAPTER 5 RESULTS AND EVALUATION

5.1 Choose Word Set Expanding Level

According to the training results of neural network, the network model constructed with 50 neurons in hidden layer is adopted. Our input is composed of PMI, Recency, and Frequency scores of each reviewer. As we mentioned in section 3.1.1, each reviewer will have six PMI scores due to the different word set expanding level in S_{p+n}^{k}. Next, we record the results of the six different scenarios and apply a measure to acquire a better word expanding level (i.e. k value):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Mean Absolute Percentage Error (MAPE) is adopted to reach the goal. *A* stands for the actual value and *F* is the forecasting value of testing 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 being zero. Table 5.1 shows the MAPE results under the six different word expanding levels:

Table 5.1 MAPE value in different word set expanding level

Neurons	50					
k value	1	2	3	**4**	5	6
MAPE	4353.267014	3866.308969	3546.679299	**2475.813779**	4591.692716	4867.48149

Smaller MAPE represents more accurate results. In the six word expanding levels, level 4 (the bold field in Table 5.1) has the smallest differences between actual and

forecasting trust values. After that, the testing data will be processed by ANN model with 50 neurons in the hidden layer and the word set expanding level will be set to 4.

5.2 Influential Ranking Result and Evaluation Design

Table 5.2 shows the analysis result of our testing data. The ranks are decided by the predicting value:

Table 5.2 Influential ranking

Reviewer ID	*Predicting value*	*Rank*
ASourdough4	0.5101	6
AtlantaGreg	0.8826	3
corona79	0.1446	13
dkozin	0.7615	4
Howard_Creech	0.9474	1
hwz1	0.9129	2
JIMILAGRO	0.2709	9
jvolzer	0.3716	8
njpoteri	0.0013	16
porcupine1	0.0669	15
readsteca	0.1553	12
sarahrose12	0.0871	14
theheidis	0.2191	10
tucknroll	0.5523	5
williamrender	0.1579	11
zan720	0.4601	7

In order to judge whether the ranking is effective or not, we also collect the ranking results of two common ranking mechanisms to compare with our method. They are “popular author” and “review rating” approaches. We choose the two methods because they are widely applied and covered “human connections” and “value of writings”.

Popular author is a ranking mechanism applied by Epinions.com. Epinions.com chooses popular authors on a monthly basis, and the newest ranking of this month can be looked up in real time. Online users can also view the popular authors on one year basis or view the overall ranking. Epinions.com records the total number of hits to members’ reviews

to decide which authors are popular. The popular authors are also classified into different product categories to make this mechanism more complete and effective. Online users can choose the category they are interested in and view those popular reviewers related to that category.

We collect the popular author ranking of each node in 2008 for evaluation. Due to the characteristics of reviews, the ranking category should be set to “Electronics” and “Overall” for comparison. Although we found that nodes have different ranking in two categories, the relative positions of them are the same. Therefore, the popular author ranking is displayed by one ranking set only.

Review rating is another common ranking mechanism. When someone posts a review, every online member can give a rating to the review. In other words, each review has a composite score which is decided by other online users. We use the average scores of all of the reviews wrote by each author to decide his / her overall review rating. The ranking is generated from these average scores. The review rating represents the comments of their readers and we think it should reflect the feeling of people correctly. It also indicates the values of these articles for online users.

Our benchmark is the level 2 trust score of these reviewers. As we stated before, trust is a powerful indicator to represent the influence of each node. In the next section, the experiment result of our model and the two evaluating rankings will be compared with the result of trust ranking to measure their effects respectively.

5.3 Evaluation and Discussion

We also use MAPE as the standard to measure the effect of the three methods. The characters of MAPE make the error rate clear and easy to read so it is still appropriate for the final evaluation. The MAPE value here is calculated by the following equation:

$$MAPE_R = \frac{1}{16}\sum_{i=1}^{16}\left|\frac{R_{Ti} - R_{Pi}}{R_{Ti}}\right|$$

R_{Ti} stands for the trust ranking of each reviewer and R_{Pi} is the predicting ranking of the three methods. By comparing with the trust ranking, the MAPE values can be acquired respectively. We show the three sets of rankings and the MAPE results in Table 5.3:

Table 5.3 Rankings and MAPE value in different methods

id / Method	*PMI+RFM*	*Popular Author*	*Review Rating*	*Trust*
ASourdough4	6	4	12	4
AtlantaGreg	3	5	16	5
corona79	13	10	7	6
dkozin	4	2	8	3
Howard_Creech	1	1	4	1
hwz1	2	3	6	2
JIMILAGRO	9	12	13	10
jvolzer	8	6	15	8
njpoteri	16	16	1	14
porcupine1	15	11	10	14
readsteca	12	7	14	13
sarahrose12	14	13	3	11
theheidis	10	8	9	9
tucknroll	5	9	5	12
williamrender	11	15	2	14
zan720	7	14	11	7
MAPE	0.24829164	0.2739394	1.014912	0

As Table 5.3 shows, our proposing method has lower error rate than others. It proves that the composite method really works. We think the amazing result has several reasons:

First, the PMI + RFM method is composed of key factors which may decide someone's influence. By opinion mining, the values of reviews can be identified easily. We do not focus on judging reviews are positive or negative but pay attentions to "quantity of emotional expression" in them. We think this mechanism can quantify the parts which are worthy to refer in reviews. In addition, the RFM analysis helps us pick up reviewers who are really active. Without RFM, the system would have high probability to choose someone was productive in the past or writing a few good reviews only but does not catch the market trend.

Second, the ANN training process shapes the model closing to real trust value. By thousands of training, the system learned the patterns extracted from real data. Sufficient training makes the model become closer to our expectation. When the testing data is processed by a well-trained network which considered several dimensions, better predictions are produced.

Third, the reasons why other two methods have larger error rates are also our concerns. We can find that there are some drawbacks in them. If we use "popular author rating", it considers the hit numbers of reviews only. In other words, the contents of reviews are not analyzed and the qualities of them are not guaranteed. It is possible to have a situation that a reviewer with high hit numbers but cannot provide the information we need. For example, we are interested in reading reviews of a reviewer who is the popular author last month. However, we found this review is not as good as before and the hit number is still recorded by the website. More people are still attracted by the "popular author" fame, feel disappoint, and increase the hit number continuously. This vicious cycle will lead to a failure in popular author judgment in the end.

As Table 5.3 shows, “review rating” method has larger error rate than “popular author”. The result implies that some problems exist in this mechanism. We sort the ratings of our testing reviews and find an astonishing fact. The rating data is listed in Figure 5.1:

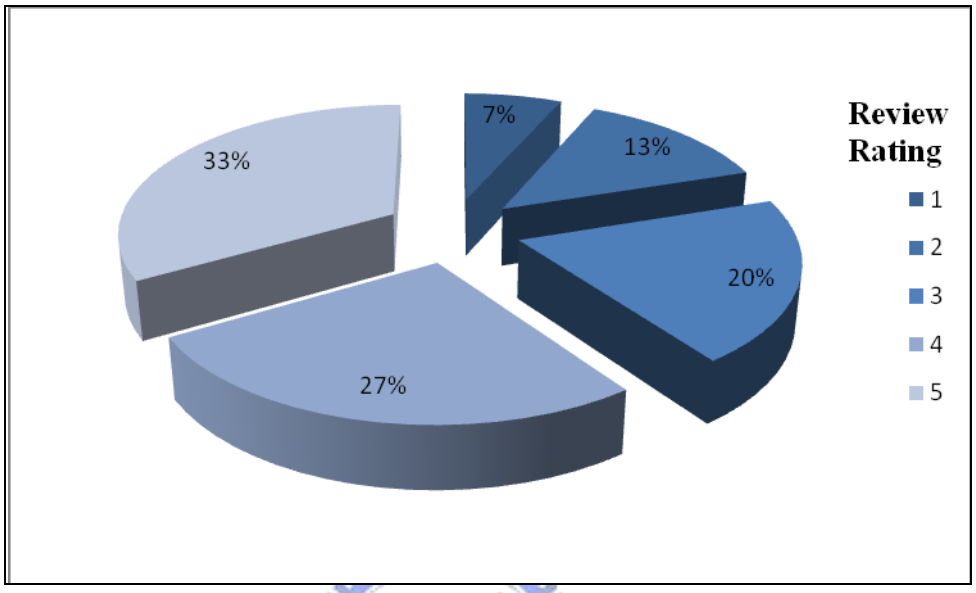


Figure 5.1 Review rating distributions among 5 rating levels

As Figure 5.1 shows, 60% reviews acquire rating 5 and 4. Online users read these reviews and scoring them to decide their ratings. We do not doubt that the ratings reflect the reviews’ real values or not because they are the judgments of people. However, the most serious problem in this mechanism is that it does not consider the RFM characteristic of reviewers. Anyone who just wrote one or two high rating reviews will get a high averaged rating. We have stated that 23.19% reviewers create 75.98% reviews in our testing data set. If we only use review rating to decide the influential nodes, there will exist high probability that top reviewers who wrote hundreds of reviews have similar or even lower scores than general ones who only wrote one or two reviews. It is all because most online reviews have high ratings and the characteristics of reviewers’ RFM are ignored.

Jindal and Liu [21] also state the phenomenon about review rating in their research. They use the data comes from Amazon.com and 59% of members have an average rating of five. This character is considered as a factor for spam opinion detection. We think the review rating mechanism for finding valuable reviewers is incomplete and may lead to bias. It is not fit to be used solely, especially while the data are not filtered or classified before.

CHAPTER 6 CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

The advancement of IT technology and the Internet reduces the cost of marketing behaviors such as advertisement but the “uncertainty” problem still exists. Many enterprises waste most of their marketing resources on inefficient marketing behaviors. Viral marketing is a new and effective marketing method which is based on the power of “word of mouth” and saves much resource and troubles in mass marketing. How to find potential nodes which spread product information efficiently in viral marketing is crucial to future business growth.

In current Internet environment, we know the recommendations from other online users are more powerful than firms’ official advertisements, and online product reviews are the source of their influence. In this research, we propose a solution which combines the text mining techniques and RFM analysis to calculate the influential power of online users and reviews. The trust score which composed of thousands of social relationship is also applied for evaluation. The result indicates that our method performs better than “popular author” and “review rating” ranking mechanisms.

For most firms, the influence of each node can be measured clearly and which node is worthy to market is easy to be identified by our model. This method provides firms a simple and helpful name list to execute marketing behaviors. It not only avoids inefficient marketing behaviors but also requires less nodes (or costs) to achieve rapid advertising effect. The high score nodes are like real viruses and infect members in their social network automatically.

This methodology also saves a lot of resource in mining online users’ activities. For example, trust collection is a deep recursive procedure and needs significant time and computing resources. Mining complete trust connections among the Internet is very difficult,

especially for small business or large data set. In addition, the word-of-mouth marketing also touches many business chances which are hard to be discovered by general recommendation systems. Recommendation systems collect customers' data from purchasing history or members' profiles, and their marketing targets are also limited in these groups. However, viral marketing can spread product information to potential customers which we do not know. It widens the range of marketing and provides more opportunities to enterprises.

6.2 Limitation of This Research

There are also some limitations in our research. First, our model is constructed based on product reviews from several different categories, especially from electronic products. We know that different products have different characteristics in using experiences and life cycles. Unrelated product reviews and reviewers may lead to inaccurate results in current model. Adjusting training data source and RFM time range are needed for different kinds of products.

Second, our model reflects the relative influence among each node, but the correct influence value is not indicated and measured in current model. In SCN tracing process, the decay rate of social connections (parameter δ) is also not considered precisely. Deeper trust network tracing leads to larger influence difference among nodes, but it is hard to be indicated in current model. Precise design of δ provides trust value closer to the real world. Accurate influence value of each node needs complete trust network tracing and lots of computing resources.

6.3 Future Work

Although we have a better method to find out influential nodes, there are still some parts in this work can be improved:

1. As the data description shown in section 4.1.2, the data is not composed of pure single category. Electronics products, Computers, and Media are the three main partitions in testing data set. It also means that the result would be more appropriate for finding potential nodes in these categories. However, there are still some other categories of reviews in the data set which have little effects to the final results. Current model can only reflect a general scenario in the Electronic products category. A better solution is to classify the reviews precisely in advance. Retrieving related data according to the need of the enterprise would lead to more accurate ranking. We plan to create different models for different categories to find better influential nodes.

2. Weighting mechanism is another factor which can be applied in current model. Current weights are decided by neural network training so it is based on online users' behavior. In fact, the weight can be tuned to fit the need of firms. For example, if enterprises hope to raise the detail and correct statements about their products in the market, they would prefer to find out reviewers who wrote really valuable comments. In other words, they give more weights about the result of PMI score. In opposite, if the enterprises just want to spread some product messages quickly (like sale information), they will pay more attention to reviewers with higher RFM score or ignore PMI calculation. Flexible weighting mechanism not only makes this model fitter to enterprises' needs but also saves computing resources.

3. In our RFM model, the Monetary value is not considered due to the characteristic of reviews. We have known that knowledge-like data is hard to be measured by currency. However, we can consider the Monetary value in another angle. For example, the price of the product commented can be measured easily, and we can map it to the review's Monetary value. This hypothesis is feasible because influential reviews of expensive

products will create high revenue for firms. We hope to integrate this element in future system architecture and observe the effects.

4. Our result provides an influence ranking based on trust evaluation. Due to the power of trust, these nodes must equip with some potential influence which affects the decisions of other buyers. However, these nodes are not verified by an information spreading test in a real world. More accurate spreading effects are expected in a viral marketing environment. We plan to observe these potential nodes for a period to verify their real effects. Questionnaires can be used to investigate the feeling of people to understand that people are affected by these reviewers or not. Comments which are replied to these reviews are also a possible way to identify the effects of these potential nodes.

5. The characteristic of trust also provides a possible addition for our model. Because of its high influence power, trust can be embedded as one element of current model after it becomes a common element of every online product review website. The composite model considers the influence of reviews, activity of reviewers, and the social connections. Stronger influential nodes are expected to be identified by the new mechanism.

REFERENCES

[1] Ammeter Anthony P., Ceasar Douglas, Gerald R. Ferris, Heather Goka. A social relationship conceptualization of trust and accountability in organizations. Human Resource Management Review, Volume 14, Issue 1, March 2004. pp. 47-65.

[2] Artificial neural network. http://en.wikipedia.org/wiki/Neural_network. Access on May 31st, 2008.

[3] Baier Martin, Kurtis M. Ruf, Goutam Chakraborty. Contemporary database marketing: concepts and applications. Evanston: Racom Communications, 2002.

[4] Barnes J A. Class and committees in a Norwegian island parish. Human Relations, 7, 1954. pp. 39-58.

[5] Benkler Yochai. Coase's Penguin or Linux and The nature of the firm. The Yale Law Journal, Vol. 112, Issue 3, December 2002. pp. 369-446.

[6] Cao Qing, Marc J. Schniederjans. Agent-mediated architecture for reputation-based electronic tourism systems: A neural network approach. Information & Management, Volume 43, Issue 5, July 2006. pp. 598-606.

[7] Chan Chu Chai Henry. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. Expert Systems with Applications, Volume 34, Issue 4, May 2008. pp. 2754-2762.

[8] Chen Mao, Jaswinder Pal Singh. Computing and using reputations for internet ratings. EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce. October 2001.

[9] Chiang Wei-yu Kevin, Dongsong Zhang, Lina Zhou. Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression. Decision Support Systems, Volume 41, Issue 2, January 2006. pp. 514-531.

[10] Cho Yoon Ho and Jae Kyeong Kim. Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce. Expert Systems with Applications, Volume 26, Issue 2, February 2004. pp. 233-246.

[11] Cook, J., & Wall, T. New work attitude measures of trust, organizational commitment and personal need non-fulfillment. Journal of Occupational Psychology, 53, 1980. pp. 39–52.

[12] Dasgupta Partha. Trust as a commodity. Edited by Diego Gambetta. Trust: Making and breaking cooperative relations. Oxford, UK: Blackwell, 1988. pp. 49–72.

[13] Ding Xiaowen, Bing Liu, Philip S. Yu. A Holistic Lexicon-Based Approach to Opinion Mining. Proceedings of the international conference on Web search and web data mining. February 2008.

[14] Dobele Angela, Adam Lindgreen, Michael Beverland, Joëlle Vanhamme, Robert van Wijk. Why pass on viral messages? Because they connect emotionally. Business Horizons, Volume 50, Issue 4, July-August 2007. pp. 291-304.

[15] Dobele Angela, David Toleman, Michael Beverland. Controlled infection! Spreading the brand message through viral marketing. Business Horizons, Volume 48, Issue 2, March-April 2005. pp. 143-149.

[16] Drozdenko Ronald G., Perry D. Drake. Optimal database marketing: strategy, development, and data mining. Thousand Oaks: Sage Publications. 2002.

[17] Erikson, Erick H. Childhood and society. New York, NY: Norton. 1950.

[18] Hu Minqing, Bing Liu. Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (Seattle, WA, USA, August 22 - 25, 2004). KDD '04. ACM, New York, NY, pp. 168-177.

[19] Hu Minqing, Bing Liu. Mining opinion features in customer reviews. In Proceedings of the nineteenth national conference on artificial intelligence, sixteenth conference on innovative applications of artificial intelligence AAAI 2004. San Jose. pp. 755–760.

[20] Hughes Arthur Middleton. (1994). Strategic database marketing- The masterplan for starting and managing a profitable, customer-based marketing program, third edition. McGraw-Hill Professional, 2005.

[21] Jindal Nitin, Bing Liu. Opinion spam and analysis. WSDM '08: Proceedings of the international conference on Web search and web data mining. February 2008.

[22] Jurvetson S. What exactly is viral marketing? Red Herring, 78, 2000. pp. 110–112.

[23] JWordnet. Pure Java API and browser interface to Princeton's Wordnet database. http://jwn.sourceforge.net/. Access on June 6th, 2008.

[24] Kuo R. J., J. A. Chen. 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, Volume 26, Issue 2, February 2004. pp. 141-154.

[25] Leskovec Jure, Lada A. Adamic, Bernardo A. Huberman. The Dynamics of Viral Marketing. ACM Transactions on the Web (TWEB), Volume 1, Issue 1, May 2007.

[26] Li Yu, Liu Lu and Li Xuefeng. A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce. Expert Systems with Applications, Volume 28, Issue 1, January 2005. pp. 67-77.

[27] Li Xuefeng, Liu Lu, Wu Lihua, Zhang Zhao. Predicting the final prices of online auction items. Expert Systems with Applications, Volume 31, Issue 3, October 2006. pp. 542-550.

[28] Liu Duen-Ren, Ya-Yueh Shih. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. Journal of Systems and Software, Volume 77, Issue 2, August 2005. pp. 181-191.

[29] Massa Paolo, Paolo Avesani. Trust-aware recommender systems. RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems. October 2007.

[30] McCarty John A., Manoj Hastak. Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. Journal of Business Research 60, 2007. pp. 656–662.

[31] Moore Robert E. From genericide to viral marketing: on "brand". Language & Communication, Volume 23, Issues 3-4, July-October 2003. pp. 331-357.

[32] Moorman Christine, Rohit Deshpandé, Gerald Zaltman. Factors Affecting Trust in Market Research Relationships. Journal of Marketing, Vol. 57, No. 1, January 1993. pp. 81-101.

[33] Morgan Robert M., Shelby D. Hunt. The Commitment-Trust Theory of Relationship Marketing. Journal of Marketing, Vol. 58, No. 3, July 1994. pp. 20-38.

[34] Munns A K. Potential influence of trust on the successful completion of a project. International Journal of Project Management, Volume 13, Issue 1, February 1995. pp. 19-24.

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

[36] O'Reilly Tim. What Is Web 2.0What Is Web 2.0- Design Patterns and Business Models for the Next Generation of Software. O'Reilly Network. September 30, 2005. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html. Access on March 5th, 2008.

[37] Pang Bo, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the Acl-02 Conference on

Empirical Methods in Natural Language Processing - Volume 10 Annual Meeting of the ACL. Association for Computational Linguistics, Morristown, NJ. 2002. pp.79-86.

[38] Rappa Michael. Business Models. Managing the Digital Enterprise. Chapter 5. http://digitalenterprise.org/index.html. Access on May 7th, 2008.

[39] RFM. Wikipedia. http://en.wikipedia.org/wiki/RFM. Access on May 30th, 2008.

[40] Richardson Matthew, Pedro Domingos. Mining knowledge-sharing sites for viral marketing. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. July 2002.

[41] Rogers Everett M. Diffusion of Innovations. Free Press, New York, fourth edition, 1995.

[42] Rotter Julian B. A new scale for the measurement of interpersonal trust. Journal of Personality Volume 35 Issue 4, December 1967. pp. 651–665.

[43] Sheth JN, Sisodia RS. 1995. Feeling the heat: marketing is under fire to account for what it spends. Marketing Management, 4, Fall 1995. pp. 8– 23.

[44] Simmel Georg, David Frisby. The Philosophy of Money. ROUTLEDGE, TAYLOR & FRANCIS GROUP, March 2004.

[45] Smith J. Brockand, Donald W. Barclay. The effects of organizational differences and trust on the effectiveness of selling partner relationships. Journal of Marketing, 61, 1997. pp. 3–21.

[46] Srivastava RK, Shervani TA, Fahey L. Market-based assets and shareholder value: a framework for analysis. Journal of Marketing, 62, January, 1988. pp. 2–18.

[47] Strang David and Sarah A. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. Annual Review of Sociology, 24, 1998. pp. 265–290.

[48] Trust (social sciences). Wikipedia. http://en.wikipedia.org/wiki/Trust_%28social_sciences%29. Access on June 1st, 2008.

[49] TrustLet, a free, collaborative project for collecting and analyzing information about trust metrics. http://www.trustlet.org/wiki/Trust_network_datasets. Access on May 30th, 2008.

[50] Tsaih Ray, Yenshan Hsu, Charles C. Lai. Forecasting S&P 500 stock index futures with a hybrid AI system. Decision Support Systems, Volume 23, Issue 2, June 1998. pp. 161-174.

[51] Turney Peter D. and Michael L. Littman. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. ACM Transactions on Information Systems, 21(4), 2003. pp. 315–346.

[52] Turney Peter D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002. pp. 417-424.

[53] Varadarajan, P. R., & Jayachandran, S. Marketing strategy: An assessment of the state of the field and outlook. Journal of the Academy of Marketing Science, 27, 1999. pp. 120–143.

[54] Victor Patricia, Chris Cornelis, Ankur M. Teredesai, Martine De Cock. Whom should I trust? : the impact of key figures on cold start recommendations. SAC '08: Proceedings of the 2008 ACM symposium on Applied computing. March 2008.

[55] Ward Mark. More than 95% of e-mail is 'junk'. Technology correspondent, BBC News website. Thursday, 27 July 2006. http://news.bbc.co.uk/2/hi/technology/5219554.stm. Access on June 6th, 2008.

[56] Weber John A. 2002. Managing the marketing budget in a cost-constrained environment. Industrial Marketing Management, Volume 31, Issue 8, 1, November 2002. pp. 705-717.

[57] WordNet. http://wordnet.princeton.edu. Access on June 6th, 2008.

[58] Zhan Jiaming, Han Tong Loh and Ying Liu. Gather customer concerns from online product reviews – A text summarization approach. Expert Systems with Applications, In Press, Corrected Proof, Available online 28 December 2007.

[59] Zhang Yiyang and Jianxin (Roger) Jiao. An associative classification-based recommendation system for personalization in B2C e-commerce applications. Expert Systems with Applications, Volume 33, Issue 2, August 2007. pp. 357-367.