Information Sciences 181 (2011) 5143–5157

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Discovering influencers for marketing in the blogosphere

Yung-Ming Li *, Cheng-Yang Lai, Ching-Wen Chen

Institute of Information Management, National Chiao Tung University, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:
Received 23 February 2010
Received in revised form 22 January 2011
Accepted 1 July 2011
Available online 22 July 2011

Keywords:
Influential model
Viral marketing
Social networks
Blogosphere
Artificial neural network

ABSTRACT

Discovering influential bloggers will not only allow us to understand better the social activities taking place in the blogosphere, but will also provide unique opportunities for sales and advertising. In this paper, we develop an MIV (marketing influential value) model to evaluate the influential strength and identify the influential bloggers in the blogosphere. We analyze three dimensions of blog characteristics (network-based, content-based, and activeness-based factors) and utilize an artificial neural network (ANN) to discover potential bloggers. Based on peer and official evaluations, the experimental results show that the proposed framework outperforms two social-network-based methods (out-degree and betweenness centrality algorithms) and two content-based mechanisms (review rating and popular author approaches). The proposed framework can be effectively applied to support marketers or advertisers in promoting their products or services.

1. Introduction

With the advent of online social networking, word-of-mouth (or viral) marketing is increasingly being recognized as a crucial strategy in social influence and marketing domains. The essence of word-of-mouth marketing is to reach out to a broad set of potential customers and attract considerable attention via social interactions. Unlike direct and mass marketing, which only recognize the intrinsic value of a customer, word-of-mouth marketing additionally exploits the network effect of a customer by taking the network factors into consideration to measure the real customer value [46]. Through word-of-mouth diffusion, information can spread more quickly and easily among social networks. Appropriate marketing campaigns based on social networks could generate a significant increase in the sales amount and reduction in the promotion cost.

Blogging systems have gained a great deal of attention as an emerging social media that exploits existing social networks by inspiring bloggers to share their own posts or personal information with others. The weblogs indeed provide a more open communication channel for people to read, commentate on, cite, socialize in, and even reach out beyond their social networks, make new connections, and form communities [35]. Blogging is a huge word-of-mouth engine [48] and the blogosphere has become a good platform for advertisers to promote new products or services and for customers to locate product comments and purchasing suggestions. In order to achieve the goal of word-of-mouth marketing, conceptually, we can start by targeting influential members to form the virus of the network for the purpose of diffusing information and making recommendations to their friends [32]. But how should we choose the influential seeds with the strongest virulence? And which characteristics should be taken into consideration when identifying the influential nodes? In this re-

* Corresponding author.
E-mail addresses: yml@mail.nctu.edu.tw (Y.-M. Li), cylai.iim96g@g2.nctu.edu.tw (C.-Y. Lai), kisd0955@hotmail.com (C.-W. Chen).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.07.023

search, we are particularly intrigued by the issues related to modeling the key factors that contribute to a successful marketing influencer and identifying the potential marketing influencers in the blogosphere.

A typical blog site combines texts (basic content), images or videos (multimedia content), and links (network-based linkage). In our work, these factors are categorized into three dimensions – network-based, content-based, and activeness-based factors. All these types of sources are considered to develop a more comprehensive and robust influence model and estimate the precise value of marketing influence. Besides, social influence is modeled as a graph-based representation, which is constructed by nodes and edges. The nodes in the blogosphere stand for the blog sites with some social characteristics or behaviors and have social influence values. The edges represent the relations based on the social activities that indicate direct social influence.

A common approach to identifying influential bloggers is to analyze the social network structures of the blog sites [2,3,16,20], whereas article content mining is another approach [12,13,57]. In this paper, utilizing content and network analysis techniques combined with an ANN (artificial neural network), we develop a framework to evaluate the influential strength of bloggers for the purpose of supporting marketing strategies. Our framework is further empirically experimented and verified using the data collected from Wretch – a prestigious online blogging system with the largest number of bloggers in Taiwan. Compared with the results of other social network analysis approaches (betweenness and out-degree centrality) and content-ranking mechanisms (review rating and popular author ranking), our proposed model has better performance in predicting the influential strength of the bloggers.

The remaining part of this paper is organized as follows. In Section 2, the existing literature related to our research topic is reviewed. In Section 3, we discuss the factors that contribute to the marketing influence and develop the MIV model. Section 4 describes the experimental data source, settings, and procedures. The experimental results and evaluations are discussed in Section 5. Section 6 concludes our research contributions and presents future research directions.

2. Related works

2.1. Ranking mechanisms in the blogosphere

A fast-growing number of blog studies show that the blogosphere as a social network can help researchers to understand and analyze certain implications and insights. It has been raising research issues and receiving lots of attention in various domains. Prior works have considered the factors in a social-relation-based dimension to measure the importance and relationships of web pages or blogs [2,19,31,36,40]. The concept of blog ranking is similar to that of blog recommendation to some extent. Fujimura et al. [19] assigned scores to each blog entry by weighting the hub and authority scores of the bloggers based on an eigenvector, which is similar to PageRank [4] and HITS [34]. These studies are based on the eigenvector calculation of the adjacency matrix of the links. However, Kritikopoulos et al. [36] ranked blogs according to their similarity in social behaviors by graph-based link analysis, which demonstrated an excellent paradigm of link analysis. However, there is an inherent problem of sparseness in the blogosphere that has already been noticed. Adar et al. [2] and Kritikopoulos et al. [36] coped with this problem by extending and increasing explicit and implicit links based on various blog aspects where a denser graph will result in better performance of ranking and recommending. Moreover, in order to solve the sparsity problem, the extracted communities studied by Lin et al. [40] only cover a portion of the entire blogosphere, and the ranking method only extracts dense subgraphs from highly ranked blogs.

2.2. Social influence in the blogosphere

Finding influential blog sites in the blogosphere is an important research problem, which investigates how these blog sites influence the world inside and outside the blogosphere [20]. Previous researchers have suggested several influencing factors to strengthen the reliability and robustness of the influence model. For example, the works by Agarwal et al. [3], Java et al. [29], and Subramani and Rajagopalan [50] applied various network-based parameters and dimensions to examine the influence of commentary information in online social networks. Subramani and Rajagopalan [50] highlighted two factors that play key roles in determining the nature of influence episodes in viral marketing. Java et al. [29] applied a PageRank-based heuristics for influence models to a graph derived from links between blogs and web pages. They also discussed the applicability of the proposed algorithms; this demonstrated how to justify and verify a proposed mechanism. Moreover, Agarwal et al. [3] presented a well-defined preliminary influence model to identify influential bloggers and pave the way for building a robust model that allows influencers to be found. In our model, we extend the concept and idea from Agarwal et al. [3], but with a different problem definition in the online marketing domain. Identifying the blog sites with greater marketing influence capabilities is crucial in the promotion of products/services, which will result in lower marketing costs and wider sales channels to increase the visibility of products/services.

Agarwal et al. [3] analyzed the factors of blog posts to represent bloggers, such as number referenced in, number of references to, number of comments, and length of blog posts. However, the social network of bloggers is not taken into consideration. Social network analysis and content mining are two common approaches to identifying influencers on the Internet. In our research, we combine social-network-based analysis, content-based analysis, and activeness analysis to discover the influential bloggers in the blogosphere.

2.3. Blog marketing via word of mouth

Word-of-mouth marketing is a new marketing method that uses electronic communications (e.g. blogosphere, forum, email) to spread messages throughout a widespread network of potential customers. Dobele et al. [14] analyzed the word-of-mouth marketing impact on firms to show how the technologies can be successfully applied. Besides, they also studied several real cases and suggested how the company can utilize this idea. In a social network, marketing through word of mouth is extremely powerful as people are likely to be affected by the decisions of their friends and colleagues [32]. Researchers have investigated the diffusion process of word-of-mouth and viral marketing effects in the success of new products [16,46]. Zhan et al. [57] emphasized the important role of writing and referring to product reviews on the Internet (such as the blogosphere or online communities). In the methodologies of implementing opinion-mining, many researchers focused on the identification of the author's attitude as positive or negative [16]. As for marketing applications, several probabilistic models were proposed for choosing customers with a large overall effect on the social network [16,46].

In general, discovering influential nodes from online social networks is one of the major avenues of word-of-mouth marketing research [17,33]. In our research, the issue of word-of-mouth marketing from the aspect of influential node discovery is addressed. Our work focuses on developing a framework that supports the enterprise to operate word-of-mouth marketing successfully. Specifically, we focus on the selection of appropriate bloggers with a better influence strength. Hopefully, these targeted bloggers could further diffuse and propagate information and even make recommendations to their friends [32].

Motivated by Agarwal et al. [3] and Kempe et al. [32], we examine several factors of marketing influence to construct our MIV (marketing influence value) model. Our influence model is close to those proposed by Agarwal et al. [3], Subramani and Rajagopalan [50], and Java et al. [29]. Their models aimed to identify and model the spread of influence in online social networks. However, our research focuses on identifying the nodes with marketing influence in the blogosphere and differs from that briefly discussed above.

2.4. Back-propagation neural network

Artificial neural networks (ANNs) utilize a mathematical or computational approach based on the concept of biological neural networks. The purpose of ANNs is to construct a model that could learn weightings similar to human thinking, and the major advantage is their flexible nonlinear modeling capability. They are particularly appropriate for solving complex problems with several variables. A number of works showed that the learning ability of back-propagation neural networks in conducting forecasts and predictions is appropriate in different domains [8,24,37]. They have been extensively used to solve business problems and can be used as an element of business intelligence. For example, Kuo and Chen [37] applied a fuzzy neural network to learn the rules produced from order selection questionnaires in electronic commerce. A feed-forward ANN with an error back-propagation learning algorithm was employed to integrate different scores. Chiang et al. [8] developed an ANN model to predict and explain the consumer's choice between web and traditional stores. Considering the nonlinear and complexity characteristics of blogging behaviors, our study also exploits this technique to build a feasible model to aggregate the evaluation including several factors and achieve an accurate identification of influential bloggers.

3. The model

By identifying a set of possible influential bloggers/blog sites, marketers could take them as the marketing nodes and launch marketing strategies to enhance the effectiveness and availability of blog marketing. Besides, the role of the influential bloggers is mainly to create awareness and signal benefits to others within their blog social network; they can be particularly influential in encouraging the trial and adoption of novel products and services [50]. In our work, we divide the marketing influence value into three main categories: network-based, content-based, and activeness-based values, as shown in Fig. 1. The blogosphere incorporates several important factors and properties in examining whether a blogger indeed has sufficient potential to elaborate the influence of viral marketing. These important factors are described as follows.

3.1. Network influence analysis

The blogosphere implicitly stores a great deal of social-network-related information, which can be used to evaluate the influence of customers. Among them, the network-based factors can be subdivided into two parts: "social connection" and "social interaction" [21,33]. Social connection network-based factors generally refer to the explicit relationship links or visits. A blog site with higher links and popularity generally has greater influential strength [30,41]. Social interaction network-based factors consider the evaluation or feedback-based social activity links. Specifically, the social interaction includes the amount of comments and citations, and the blogroll. A blogger with more comments and citations gains more attention. The blogroll reflects the trustworthiness of a blogger. Meanwhile, a blogger with more attention and trustworthiness has a more powerful influence [26].

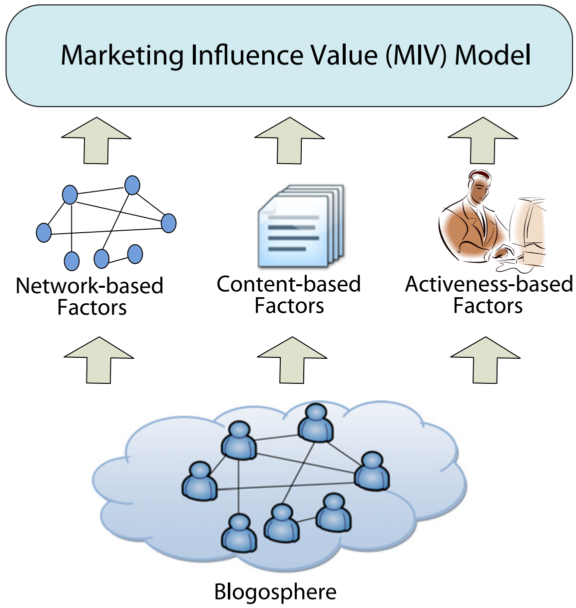

Fig. 1. Social influence and MIV factors in the blogosphere.

3.1.1. Social connection factors

In-links and out-links (PageRank). The nature of blog social networks is characterized by its linkage structure, including in-degree and out-degree links. Although there is an inherent problem of sparseness in the blogosphere, Adar et al. [2] and Kritikopoulos et al. [36] coped with it by extending and increasing the explicit and implicit links based on various blog aspects, where a denser graph will result in a better performance of ranking and recommending. According to Agarwal et al. [3], the number of out-links denotes negative utility/effect in influential value and a loss of novelty (originality) of a blog post. Brin and Page [4] also suggested that the PageRank score contributing to other nodes should be divided by the number of out-links. The in-links structure in the blog marketing domain suggests the concept of brand and recognition: (1) the more in-links, the stronger the brand power; (2) the higher the novelty of a blog post, the better recognized the blog content. Based on in-links and out-links, we can analyze the characteristics of the social network in a blogosphere.

The PageRank algorithm works by simulating a random walk with a probability q to jump to other sections of the web graph, and probability $1-q$ to follow a link. Denote $IL(a)$ as the set of in-links of web page a and $OL(a)$ as the set of out-links of web page a. The PageRank score of web page a is formulated as:

$$PR(a) = q + (1-q)\sum_{p_i \in IL(a)} PR(p_i)/|OL(p_i)|,$$

where $|OL(p_i)|$ is the number of out-links of page p_i. The in-links and out-links values respectively denote the number of friend-of relationships and the number of friends in the social network of a blogger. According to this evaluation of the network structure, we can calculate the network-based values of a blogger.

Network externalities (*visits*). The influence of network externality is widely presented in domains ranging from traditional industries such as telecommunications, hardware and software industries, and information marketing [49] to Internet service providers such as social-networking-related services or products. In the context of the marketing domain, network externality is a positive factor in measuring the marketing influence value. In addition to those factors directly derived from social network usage, network externality accrues broadly to the set of all adopters [50]. Therefore, the number of visitors (α) during a period of time on a blog site is used as an indicator of the network externality factor. These implicit network links are created by visiting behaviors. It is quite intuitive that a blog site can be more effective in promoting a product/service if it has more disseminators.

3.1.2. Social interaction factors

Blogroll relationships (*trustworthiness*). Blogroll relationships are used to measure the trustworthiness of bloggers. Friend or friend-of relationships are an especially crucial factor in referencing trustworthy and reliable information [10,11,39]. We quantify the relation as a degree of trustworthiness and reliability toward a blogger who is worthy of conducting a belief and commitment that the advertising agent will have good referral or recommendation behaviors. In the blog marketing domain, the reputation (γ) was taken to measure the trustworthiness of a blogger. Various definitions and notions of trust and reputation have been studied across diverse disciplines. Previous studies such as [1,48] indicated that reputation could be seen as a form of social control mechanism due to users' fear of having a bad reputation. Although not explicitly described, they considered reputation as a propagated notion that is passed to other agents by means of "word-of-mouth." In our context, the reputation factor has a similar definition, a perception of intentions and norms that an agent creates through her/his past actions. In the Wretch blogosphere, each member could recommend specific blog content and a member could only recommend one blog post once. A blogger/blog site gains a higher trust rating and a wider scope of influence if she/he receives more recommendations. Thus the number of members' recommendations would be used as an indicator of the blogroll relationship.

Comments and citations. Comments and citations are two of the most popular social-based interactive activities on a blog site. To measure the impacts on the viral marketing effect, we take the notions used by Agarwal et al. [3] and define the number of comments received as its capability of generating activities. In other words, few or no comments indicate little interest of fellow bloggers to take his/her advice to purchase a particular product. However, a large number of comments implies that the blogger has more impact on others such that they are willing to write feedback. In the blog marketing domain, the commenting behaviors successfully catch the attention of Internet surfers (potential customers). This research takes the number of comments from Wretch as an indicator to measure the marketing influence power in social interactions.

Citation is a kind of in-degree link that could be used as a substitute for the recognition level of a blog post. Thus, the marketing influence in adopting products/services or propagating epidemic information should be more significant when a blogger/blog post has more citations than others. People tend to cite the high-quality posts (i.e., the most suitable and critical product comments) to improve the variety and reputation of their own blog site. Additionally, a blog post has more chances of being seen by others if it is cited frequently. More chances of being seen indicate a larger influential range. Comments and citation are two major social behaviors in the blogging environment. No matter what type the received comments and citations are, the blogger attracts much more attention than others who have fewer comments and citations. As a result, we take the number of comments and the number of citations as indicators to measure the influencing power and range of social interactions, respectively, for identifying influential bloggers.

3.1.3. Network-based influence value calculation

Based on the above descriptions of network-based factors, we transform the web context into the blog social network – a representation of a marketing influence graph in which the number of friend-of relationships (IL) and the number of friends (OL) are taken into consideration for calculating the network-based value. The network-based value ($\mathrm{NV}(b)$) of blog site b is affected by the above factors – the score of PageRank $PR(b)$, the number of visitors (α), the value of reputation (γ), the number of comments (c), and the number of citations (η), which are taken into the equation to prepare the ground for developing the MIV model. The network-based value of a blog site b is formulated as

$$\mathrm{NV}(b) = \alpha_b PR(b) + \gamma_b (c_b + \eta_b).$$

The first term in the measurement represents the value from social connection factors, and the second term stands for the value from social interaction factors.

3.2. Content influence analysis

Since the blog is formed by text-based articles with reverse chronological sequences of dated entries, a content-oriented analysis is indispensable for a better understanding of the tendency, preference, and viewpoint of bloggers about the marketing/advertising of a specific product/service. To evaluate the content influence, the subjective degree, length, and living time of the blog are considered. People tend to trust and pay more attention to the blogs with subjective wordings, a longer length, and longer living time [3,5,15]. These important content-based factors are described as follows.

Subjectiveness. People generally like to follow the opinion leaders' advice and take their suggestions to purchase a product/service, and visitors are more impressed by subjective comments [15]. The tendency that all positive or negative comments are hard to be trusted has been indicated by previous work [13]. However, this factor is not used to judge the positive or negative aspect but to estimate the subjectiveness of bloggers' expressions. As for the "subjective words," we take a total of 4542 positive words and 4333 negative words defined by the HowNet Knowledge Database (http://www.keenage.com/) as our "subjective word set." HowNet is an online bilingual common sense ontology describing the semantic relations between concepts (represented by Chinese and English words) and semantic relations between the attributes of concepts [22].

The total occurrences of a subjective word w_i in all the blogs belonging to the blog post set Φ_b can be written as

$$O(b) = \sum_{w_j \in p_j} \sum_{p_j \in \Phi_b} \mathrm{occurrence}(w_i, p_j), ,$$

where occurrence(w_i, p_j) represents the number of subjective terms w_i occurring in blog posts p_j. Let B be the set of all the targeted blog sites. The subjectiveness score (s_b) of blog site b is formulated as:

$$s_b = O(b) \Big/ \max_{\beta \in B} \{O(\beta)\}$$

Length of a blog post. Aligned with the findings of Agarwal et al. [3], the length of a blog post is positively correlated with the number of comments, which means longer posts attract more people's attention. We also take the length of a blog post as a factor in determining the marketing influence. The blog length score of (λ_b) of blog site b is computed by averaging the length of all the blogs in the blog post set Φ_b and expressed as:

$$\lambda_b = \left(\sum_{p_j \in \Phi_b} |p_j|\right) \Bigg/ |\Phi_b|, ,$$

where $|p_j|$ is the length of blog post p_j and $|\Phi_b|$ is the number of blogs written by blog site b.

Living time in the blogosphere. The living time of a blog website is used to measure the influence of the content. In general, more social interactions (e.g. citation and reply) related to a blog site will occur as the living time of the blog site becomes longer. As a result, the influential strength of the posted content will increase. In this research, the living time of a blog is defined as the time period from the time of the first blog post to the time of the latest blog post.

The living time length score of (τ_b) of blog site b is calculated as:

$$\tau_b = T(b) \Big/ \max_{\beta \in B}\{T(\beta)\},$$

where $T(b)$ is the time interval between the post timestamps of the first and the latest posted blogs in Φ_b.

In the blog domain, the content-based value is especially appropriate in understanding the bloggers' preferences and affinities [39]. To measure the content-based marketing influence value of a blog site, we combine the above content-based factors such as subjectiveness (s), the length of a blog post (λ), and living time in the network (τ) together in the MIV model. The content-based value of a blog site b is formulated as

$$\mathrm{CV}(b) = \tau_b(s_b + \lambda_b).$$

3.3. Activeness influence analysis

An influential node should have fascination and activeness in his/her network community. In the blogosphere, people make social conversations with words. For example, bloggers share their experiences, users raise questions or make comments, and bloggers give responses or make comments. Intuitively, a more active blogger would show more willingness to make social conversation with other users. The following factors are used to evaluate the activeness-based value of bloggers.

Content posts. Content posts (cps) are one of the main activities of bloggers. Bloggers share their own experiences or make comments on products in a blog post. Regularly contributing content can attract users to revisit the blog site or subscribe to the blog posts. Therefore, it can help bloggers to gain more popularity for their own blog sites.

Comment replies. A blogger replies to comments representing that the blogger has interest in making social conversation with others. People might be influenced by kinds of social conversations, so-called promotional chat [42], in social media. In this research, the number of comment replies (cr) is used to reflect the level of users' willingness to share or to answer other users' questions. It is an important indicator in evaluating the influence of users in a social community.

We calculate the activeness-based value of a blogger by the counts of his/her activity records during a period of time in the blogosphere. We define $act(b) = cr + cp$ as the number of activities of blogger b during a time period T. The formula of activeness-based value (AV) is defined as below:

$$\mathrm{AV}(b) = \frac{act(b)}{T}.$$

In this formula, the time period T is chosen as 365 days because our data set was collected in a 1-year period.

3.4. Marketing influence value calculation

Combining the network-based, content-based, and activeness-based values of a blog site, we could quantify the marketing influence value (MIV) of a blog site. By adopting the MIV model, we could differentiate influential bloggers in the marketing domain to alleviate the marketing costs and promote a product/service with ease. Therefore marketers can use the MIV model to identify the influential bloggers in order to run marketing campaigns. The MIV model is developed by aggregating the weighted network-based value (NV), content-based value (CV), and activeness-based value (AV). The formulation of the MIV model is

$$\mathrm{MIV}(b) = \omega_n\mathrm{NV}(b) + \omega_c\mathrm{CV}(b) + \omega_a\mathrm{AV}(b),$$

where MIV(b) is the marketing influence value of blog site b, and ω_n, ω_c, and ω_a are the weights that can be used to adjust the contribution of network-based, content-based, and activeness-based marketing influence values. Because of the complexity of human behaviors in the blog social network, in this research, a three-layer back-propagation neural network (BPNN) is employed to deal with the uncertain weighting problem between NV, CV, and AV such as ω_n, ω_c, and ω_a for forecasting the final MIV value. Notice that while we calculate the score of each measure by aggregating these related components according to additive or productive models [3,8], the final marketing influence value (MIV) is nonlinearly adjusted by the artificial neural network. This approach is commonly used by existing works such as [8,44].

4. Experiments

In the following, we conduct an empirical study to verify our proposed MIV model. We first construct the blog social network based on a data source crawled from the blogosphere. Then, the bloggers are ranked according to their MIV scores. According to the evaluation results from the peer and official ratings, we compare the performance of the proposed MIV model with that of other blog site selection approaches.

4.1. Data source

We test our model by using a data set collected from the Wretch blog [52,54]. It is the largest weblog community in Taiwan with millions of users registered. In the community users can upload photos to their albums, write their blog, and interact with others. An experience good is a product or service whose quality and value are difficult to evaluate before it is used or consumed [43,45]. Word-of-mouth marketing is particularly appropriate for promoting experience goods as they create repurchases and attract new consumers by rewarded reputation. Since the purpose of our model is to find nodes with the power to influence the buying decisions of others, Wretch/Cate is an especially suitable candidate to conduct the experiments since food is one kind of experience good. The Cate category contains bloggers/blog sites of bons vivants. They share food information and restaurant recommendations and experiences on their own blog site. It also provides abundant interactions (such as publishing, citations, comments, friends and friend-of lists, etc.) about Cate information.

The data set was collected on 14 May, 2008 under the category of Cate. We collected all the articles and related bloggers as our base for further analysis. First, once the bloggers and neighboring/related bloggers had been crawled, we turned the focus to collecting articles and related information from their blog sites. For the sake of computational efficiency, at most 10 articles for each target author are examined for further analysis. Table 1 lists the statistics about the data collected in our experiment. The statistics illustrate the overall network and content-based characteristics of the collected bloggers (i.e., a blog site).

4.2. Experiment design

The experiment processes for our proposed framework include the influential factor analysis, the MIV score calculation, and the performance evaluation, as shown in Fig. 2.

Influential factor analysis. After the blog social network is constructed, we calculate the network-based value (NV), content-base value (CV), and activeness-based value (AV) by extracting the elements required by each influence analysis module. Notice that according to PageRank [53] the value of q is set to 0.25 as our default setting for aggregating NV.

MIV score calculation. In this study, an artificial neural network (ANN) with a three-layer back propagation neural network (BPNN) is used to deal with the uncertain weighting problem between network-based, content-based, and activeness-based values and predict the overall influential strength. The MATLAB neural network toolbox is applied to construct a three-layer ANN for training and testing the appropriate neural network model. The constructed three-layer ANN is composed of an input layer, hidden layer, and output layer. The input layer has three neurons: the aggregated network-based value (NV), content-based value (CV), and activeness-based value (AV). In the hidden layer, 50 neurons are used to adjust the weights adaptively. Only 1 neuron is included in the output layer, which is the predicted MIV value. The detailed settings about the BPNN model are listed in Table 2.

As the users/readers are the targets to be influenced by the bloggers, whether a blogger is influential or not should be evaluated by the users/readers [23,55]. To establish a trained ANN model, we first collected 70 bloggers sampled from

Table 1
Statistics data of the blogger set.

Statistics from our examined blogger set	
Number of available bloggers	382
Number of available blog posts examined	3455
Average number of friends per blogger	13.312/per blogger
Average number of friends-of per blogger	15.385/per blogger
Average number of posts per blogger	180.796/per blogger
Average live time per blogger (day)	517.215/per blogger
Average cumulative visitors per blog site	129948.259/per blog site
Average words per post	409.683/per post
Average number of comments per post	1.687/per post
Average number of citations per post	0.005/per post
Average number of user recommendations per post	123.823/per post

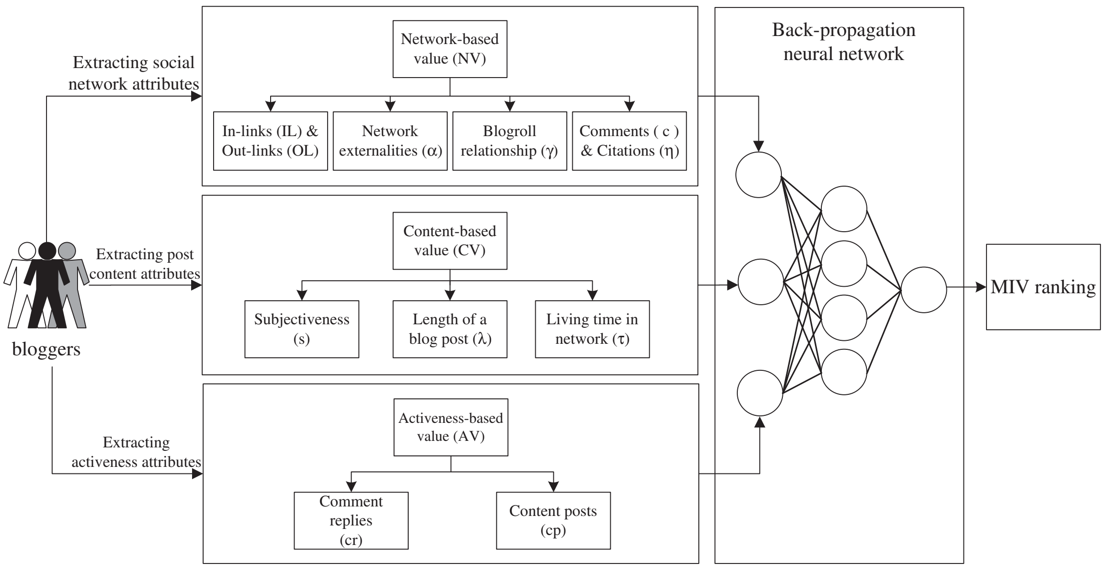

Fig. 2. MIV calculation progress.

Table 2
The settings of the BPNN model.

Parameters	Value	Parameters	Value
Number of hidden-layer neurons	50	Mu	0.001
Training function	TRAINLM	Mu_dec	0.1
Performance function	MSE	Mu_inc	10
Epochs	500	Mu_max	10,000,000,000
Goal	0	Max_fail	5

the original blogger set. Then, we invited 30 online users to score (0–100) the influential strength of the bloggers/blog sites. The evaluations are used as the data set for training the artificial neural network. Many artificial neural network algorithms learn incrementally and allow dynamic training setting [51]. So, the ANN model applied in our MIV framework can be feasibly implemented in the dynamic online learning environment. After the neural network predicting model is established, the remaining 212 testing bloggers are fed into the well-trained ANN model to acquire the predicted MIV values. Finally, the MIV values are used to rank the influential strength of the bloggers/blog sites.

Performance evaluation. In order to evaluate the effectiveness of the proposed framework, we compare our model with other benchmark approaches using two performance evaluation methods – peer voting and official recommendation. In the peer voting, a consensus influence evaluation process is conducted with the help of participants with an e-commerce background. In the official recommendation, the precision and recall rates of a generated recommendation of bloggers are evaluated by comparing them with the blog websites recommended by the official experts of Wretch.

5. Results and evaluations

We conduct experiments based on different selection algorithms and compare their corresponding results. This research takes the network, content, and activeness factor dimensions into consideration; it is differentiated from previous work [38] in which only the network and content factor dimensions were considered. Also, we compare the proposed approach with two common social-network-based influence evaluating methods – the "out-degree centrality" and "betweenness centrality" algorithms [18,33] – as well as two common content-based ranking mechanisms – the "review rating" [9] and "popular author" [7,56] approaches. The reason for choosing these four benchmark methods is that they are commonly used and represent a specific performance index of "social connections" or "value of writings." Finally, the performance of the proposed approach was compared with different dimension combinations.

Centrality approaches are the structural measurements that can be used to identify influencers in online social networks [33]. Degree centrality is one of the simplest centrality measures. The edges between two nodes in social networks are directional; it could be distinguished as in-degree centrality and out-degree centrality. However, the interactions between two non-adjacent nodes might depend on those nodes that lie on the path between them. A node is central to a community if it interconnects with many other nodes. Therefore, the node might have the potential to be an opinion leader in a community

[18]. Betweenness centrality is used to measure the degree of a node falling on the shortest paths between others. As the out-degree centrality and between centrality are simple and outperform other social centrality measurements [18,33], they are used to evaluate the performance of our proposed approach. The out-degree centrality measure can be formulated as:

$$OutDegree(b) = \sum friend(b),$$

where $friend(b)$ is the number of friends of blogger b. Between centrality is measured by

$$\text{Betweenness}(b) = \frac{\sum_{b \neq i \neq j} g_{ij}(b)}{G_{ij}},$$

where G_{jl} is the number of shortest paths linking the two nodes j and l and $g_{jl}(b)$ is the number of shortest paths linking the two nodes j and l containing node b. We utilize the Ucinet [6] software to calculate the centralities of bloggers.

"Review rating" is a common approach used to evaluate the influential strengths of the reviewers and reviews [9]. Wretch offers an online recommendation system for customers to recommend blog posts after reading. It lists the most popular bloggers and their blog posts based on the accumulated number of recommendations. In practice, the sales of products are greatly influenced by professional product reviews and the impact of review ratings on product sales has been examined by prior studies [45,47]. Having a higher ranking indicates that the reviewers/reviews have greater influence on the consumers. The review rating is measured by the average number of recommendations for the blogs written by blog b and is calculated as:

$$ReviewRating(b) = \frac{\sum_{p_j \in \Phi_b} OR(p)}{|\Phi_b|},$$

where $OR(p)$ represents the number of recommendations for blog post p.

"Popular author" is another common online ranking approach for evaluating the influential strengths of the reviewers and reviews [7]. Having a higher author rating indicates that the reviewers are popular and/or that their blog entries are helpful or may be capable of influencing customers. These reviewers may have greater influence on other readers or potential consumers [56]. In the Wretch blogosphere, the popularity of a typical blogger b is evaluated by the number of visitors and formulated as:

$$PopularAuthor(b) = \frac{\sum_{p_j \in \Phi_b} Visits(p)}{|\Phi_b|},,$$

where $Visits(p)$ represents the number of recommendations for blog post p.

While influence may be a subjective judgment, our purpose is to help businesses to find influential bloggers to support their marketing strategy. Prior studies suggest that most users only access the documents/articles that are shown in the top 20 list [9,25,27,28]. For instance, 100% of users of the Fireball search engine, 82% of users of a Spanish Web directory, 77% of users of the Excite search engine, and 89% of users of the AlltheWeb.com (a major and highly rated European search engine) viewed only the first 20 results. Therefore, in this research, we suggest a list of the top 20 bloggers according to the measured MIV values for each approach and compare their performances based on the set of selected bloggers.

5.1. Comparisons based on peer voting

It is important yet challenging to evaluate the impacts of the chosen blog sites in the blog marketing domain. It is reasonable to assume that the explicit rating or ranking scores of bloggers on the website could be used to calculate the relevance between the results of our proposed model and the real-world ranking mechanisms. Thus, we build a consensus influence evaluation process similar to the Delphi method. A total of 58 participants with a background in e-commerce were invited to evaluate our MIV model. The detailed procedures are shown in Fig. 3. The evaluation process includes the following two stages. In the first stage, the participants were invited to visit the blog sites on the recommendation list, mark the influential level, and report their comments. We asked them to visit the blog sites that are listed in the top 20 influential blogger list and evaluate their influence strengths on 5 levels – "no influence," "below average," "average," "above average," and "high influence." Here, according to the evaluated results from the first stage, we make the first selection in aggregating the marked influence strength level. In the second stage, the system lists the blog sites and the evaluation results of the consensus influence level, and these comments are collected. Then, the same participants were asked to revisit the blog sites to make a second evaluation. Lastly, the final group evaluation results are determined based on the majority of individual evaluations. Notice that a consensus influence evaluation process could include more than two stages. The reason that we adopt this two-stage approach is mainly for the sake of participation motivation of the peer users.

At the end of the consensus influence assessment, we will be able to respectively obtain a valid top 20 influential bloggers list for the three approaches shown in Tables 3–7. A blogger who is listed in these lists with a consensus peer evaluation score of "average" or better is treated as an influential blogger. In the five tables, these bloggers are marked with an asterisk ($*$).

We utilize the following formulation to calculate the recommendation accuracy rate according to the user evaluation of the selected influential bloggers.

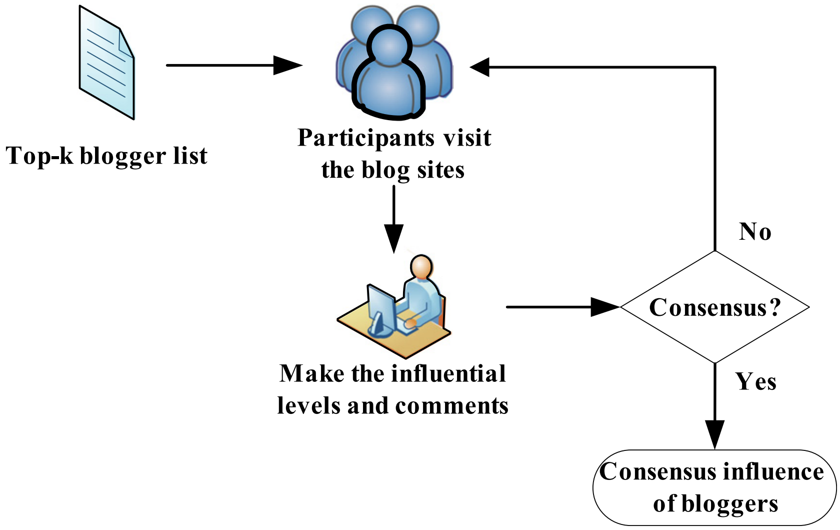

Fig. 3. Consensus influence evaluation procedure.

Table 3
Top 20 influential bloggers of the MIV approach.

Rank	Blogger#	Value	Peer evaluation	Rank	Blogger#	Value	Peer evaluation
1	Blogger360*	89.8188	High influence	11	Blogger194*	84.0265	High influence
2	Blogger020*	88.1657	High influence	12	Blogger325*	82.8223	Average
3	Blogger336*	87.3116	Above average	13	Blogger058*	82.2988	Below average
4	Blogger116*	87.1799	High influence	14	Blogger067*	81.4612	Above average
5	Blogger286*	87.1787	Above average	15	Blogger085	80.1482	Below average
6	Blogger349*	86.9033	High influence	16	Blogger193*	79.9163	Above average
7	Blogger215	86.4317	Below influence	17	Blogger379*	78.7199	High influence
8	Blogger214*	85.6799	Above average	18	Blogger378*	78.7059	High influence
9	Blogger009*	85.3481	High influence	19	Blogger104*	78.0841	High influence
10	Blogger300*	84.5937	Above average	20	Blogger080*	77.4804	Average

Table 4
Top 20 influential bloggers of the betweenness centrality algorithm.

Rank	Blogger#	Value	Peer evaluation	Rank	Blogger#	Value	Peer evaluation
1	Blogger059	1.409	No influence	11	Blogger253*	0.877	Average
2	Blogger151*	1.242	Average	12	Blogger086	0.873	Below average
3	Blogger107	1.223	No influence	13	Blogger230*	0.861	Average
4	Blogger379*	1.173	High influence	14	Blogger127	0.856	No influence
5	Blogger360*	1.053	High influence	15	Blogger067*	0.853	Above average
6	Blogger057	1.008	Below average	16	Blogger316	0.831	No influence
7	Blogger058	0.98	Below average	17	Blogger291*	0.823	Average
8	Blogger083*	0.957	Above average	18	Blogger085	0.819	Below average
9	Blogger084*	0.955	Above average	19	Blogger286*	0.808	Above average
10	Blogger378*	0.933	High influence	20	Blogger233	0.788	No influence

$$\text{Accuracy} = \frac{\text{Number of influential bloggers}}{\text{Number of total recommended bloggers}} \times 100\%.$$

We compare the recommendation accuracy according to three influence levels: “high influence,” “above average,” and “average.” We can easily observe the dominance of our proposed recommendation approach over other algorithms from Fig. 4. Specifically, approximately 45% of the bloggers discovered by our approach are those with high influence. In addition, our approach outperforms the other algorithms with an 85% improvement (higher than average level) in the accuracy rate. The proposed MIV model comprehensively considers the factors in social networks, content, and activeness dimensions to discover influencers from the blogosphere – the factors in the social network dimension catch the connection importance of bloggers in the blogosphere; the factors in the content dimension could reflect the quality of blog posts written by bloggers; the factors in the activeness dimension can measure whether the bloggers continuously interact with others.

Table 5
Top 20 influential bloggers of the out-degree centrality algorithm.

Rank	Blogger#	Value	Peer evaluation	Rank	Blogger#	Value	Peer evaluation
1	_Blogger009*_	84	High influence	11	_Blogger379*_	50	High influence
2	Blogger053	83	Below average	12	Blogger008	48	Below average
3	Blogger081*	81	Average	13	Blogger010	44	No influence
4	Blogger104*	78	Above average	14	_Blogger012*_	42	High influence
5	Blogger052	68	Below average	15	Blogger014*	41	Average
6	Blogger166	65	No influence	16	Blogger016*	40	Average
7	Blogger037	64	No influence	17	Blogger004	37	No influence
8	_Blogger018*_	63	High influence	18	Blogger021	32	Below average
9	_Blogger019*_	57	High influence	19	_Blogger025*_	26	Average
10	Blogger007	51	No influence	20	Blogger023	25	No influence

Table 6
Top 20 influential bloggers of the review rating mechanism.

Rank	Blogger#	Value	Peer evaluation	Rank	Blogger#	Value	Peer evaluation
1	_Blogger193*_	3046	Above average	11	_Blogger067*_	1118.3	Above average
2	_Blogger084*_	2998.1	Above average	12	Blogger030*	990.7	Average
3	_Blogger020*_	2525.75	High influence	13	Blogger091	867.3	Below average
4	_Blogger012*_	2017.7	High influence	14	Blogger208*	808.9	Average
5	Blogger089	1643.4	Below average	15	_Blogger009*_	806.8	High influence
6	Blogger234	1489	No influence	16	Blogger167*	764.2	Above average
7	Blogger119	1387.444	No influence	17	_Blogger214*_	741.2	Average
8	Blogger197	1185.5	Below influence	18	Blogger334	726.7	No influence
9	Blogger250	1164.3	Below influence	19	Blogger279	576.7	Below average
10	_Blogger193*_	3046	Above average	20	_Blogger379*_	548.8	High influence

Table 7
Top 20 influential bloggers of the popular author mechanism.

Rank	Blogger#	Value	Peer evaluation	Rank	Blogger#	Value	Peer evaluation
1	_Blogger349*_	12656.57	High influence	11	Blogger030*	5742.8	Average
2	_Blogger352_	7733.9	Below average	12	_Blogger089_	5696.63	Below average
3	Blogger301*	7203.92	Above average	13	Blogger208*	5356.56	Average
4	_Blogger067*_	6812.69	Above average	14	Blogger132	5289.18	Below average
5	_Blogger360*_	6479.38	High influence	15	_Blogger116*_	5279.91	Above average
6	Blogger075*	6442.83	Average	16	Blogger319	5247.03	Below average
7	_Blogger286*_	6349.95	Above average	17	Blogger250	5164.94	Below average
8	Blogger197	6213.26	Below average	18	Blogger260	5099.23	Below average
9	_Blogger009*_	5841.02	High influence	19	_Blogger378*_	5063.16	High influence
10	_Blogger194*_	5761.83	High influence	20	Blogger245	5041.93	Below average

5.2. Comparisons based on official recommendation

The contents in Wretch's official Cate blog site (http://www.wretch.cc/blog/Wretchfood) are made up of the blog posts written by the selected expert Cate bloggers. The expert Cate bloggers, the co-authors of Wretch's official Cate blog site, are selected from their members to edit the official blog site content together. Any members who are willing to join the co-author can submit an application to Wretch. Then, Wretch's evaluation team measures the quality of applicants' sites to see if they are good enough to join the co-authors of their official Cate blog site. Wretch has already identified 27 expert Cate bloggers from among their online users. These expert bloggers were invited to be the co-editors of Wretch's official Cate blog site. These expert bloggers together contribute Cate content to this official blog site.

A total of 24 Cate expert bloggers are included in our blogger set. In the previous comparison, each blogger recommendation list generated by different discovery mechanisms contains some expert Cate bloggers. For example, in Tables 3–7, the bloggers who are Wretch-identified Cate experts are displayed in italics. These expert bloggers are treated as the officially recommended targets for computing the precision and recall values. The recall rate is formulated as:

$$\text{Recall} = \frac{\text{Number of discovered expert bloggers}}{\text{Number of total official listed expert bloggers}} \times 100\%.$$

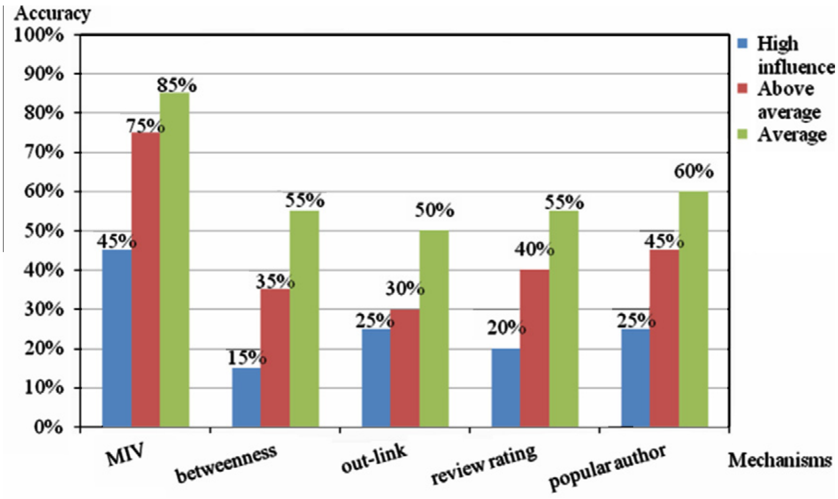

Fig. 4. Recommendation accuracy.

Table 8
Top 20 evaluation results based on official recommendation.

	MIV	Betweenness	Out-degree	Popular author	Review rating
# of Experts discovered	13	7	6	10	8
Precision value	65%	35%	30%	50%	40%
Recall value	56%	29%	23%	41%	30%

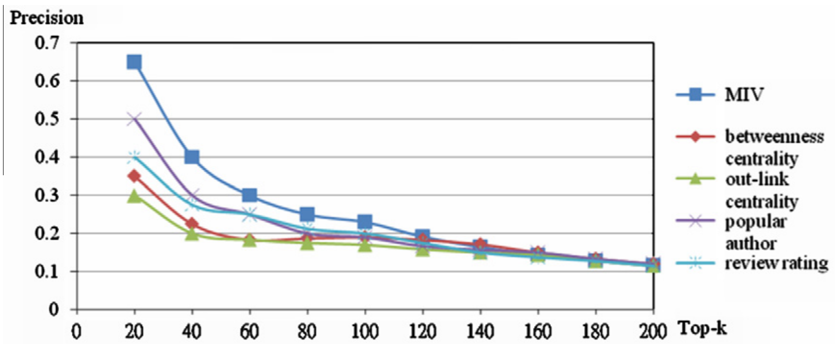

Fig. 5. Precision rate comparisons.

The precision rate is measured as

$$\text{Precision} = \frac{\text{Number of discovered expert bloggers}}{\text{Number of listed bloggers(top-}K)} \times 100\%.$$

Denote N as the number of official listed expert bloggers and n as the number of discovered expert bloggers in a top-K list for a discovery mechanism. The recall value is measured as n/N and the precision value is calculated as n/K. Table 8 lists and compares the number of discovered expert bloggers and the precision rate and recall rate with respect to these three approaches. We can observe that the performance of our proposed approach is significantly better than that of the other four.

In addition, we also construct experiments to evaluate the performance by changing the size of the recommendation list (i.e., the value of K). The 21 defined expert bloggers in Wretch are used as the recommended targets in the following evaluation process. The precision and recall comparisons are shown in Figs. 5 and 6, respectively. As we can observe in Fig. 5, when the number of experts in the recommended list is more than the top 120, these three lines will become very close. This is because almost all the defined experts have already been discovered.

The comparisons of recall values in Fig. 6 show the dominance of our proposed approach over the others. Precisely, our method could successfully discover all the expert bloggers in a top 100 list, while the betweenness centrality, out-degree centrality, popular author, and review rating approaches discovered all the expert bloggers in the top 140, top 161, top

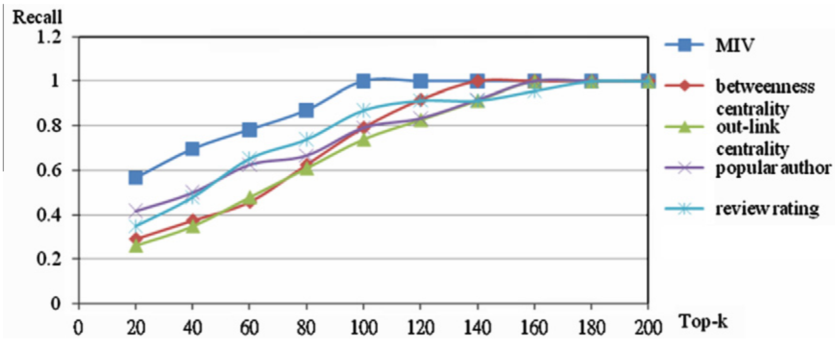

Fig. 6. Recall rate comparisons.

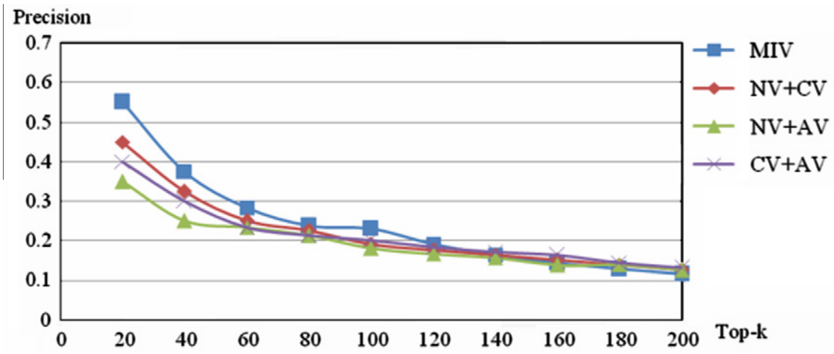

Fig. 7. Precision rate comparisons with different dimension combinations.

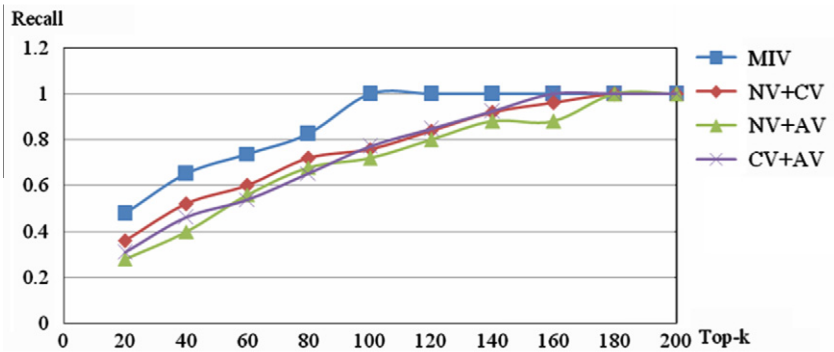

Fig. 8. Recall rate comparisons with different dimension combinations.

159, and top 171, respectively. In all the situations, our proposed system significantly outperforms the other four blog site selection approaches.

Figs. 7 and 8 compare the performance of the proposed approach with other approaches with different attribute/dimension combinations. We can observe that the NV + AV approach has the lowest effectiveness in recommendation because it completely ignores the content-based dimension. Our proposed MIV model outperforms the others in both precision and recall rates as it considers all the three factors (network, content, and activeness) at the same time.

6. Conclusion

The "uncertainty" problem in marketing-led enterprises results in the risk that resources will be wasted on inefficient marketing. Word-of-mouth marketing is a new and effective marketing method that is based on the potential nodes that are influential and powerful towards others in online social networks. It can be helpful in saving resources and marketing costs and creating more business opportunities. Blogging is a huge word-of-mouth engine and the blogosphere has also become an excellent platform for advertisers to promote new products or services and for customers to locate product comments and purchasing suggestions.

In this research, we construct a comprehensive model in which three dimensions of factors – network-based, content-based, and activeness-based factors – are considered in measuring marketing influence strength and in identifying potential and influential authors in the blogosphere. The proposed approach can support the marketers/advertisers in promoting their products/services with less effort and cost. Utilizing our proposed model, the most worthy bloggers with marketing value could be identified effectively. This study provides a feasible yet powerful way to generate a ranked list of bloggers according to their influential powers in improving the effectiveness of marketing activities. Our experimental results show that this model could significantly reduce marketing costs and uncertainty, and the proposed model could give better results than other approaches, such as the betweenness and out-degree, review rating, and popular author metrics. Besides, our approach outperforms others with an average accuracy improvement of 85%.

As mentioned, social network analysis and content mining are two common techniques for identifying influential bloggers and a single method might not accurately evaluate the influence of bloggers. We draw on the strength of each to offset the weakness of the other. In order to take advantage of the strength of word-of-mouth marketing, in the current model, the bloggers who were discovered by the proposed approach would have the strength of society, resonance, and activity at the same time. Society indicates whether a blogger has a great number of online friendships. We consider the social connection and social interaction factors to analyze the society of bloggers. Resonance reflects whether a blog post arouses the reader's interest and desire for consumption. The content-based factors are used to estimate the resonance of the blog posts of bloggers. Activity represents whether a blogger is active in sharing information through his/her blog site. The activeness-based factors are used to determine the activity of bloggers. This research builds a feasible model to aggregate these factors for estimating the possible marketing value of bloggers.

From the perspective of marketing practice, the proposed framework can help enterprises to carry out a successful word-of-mouth marketing strategy that can save a lot of resources in finding potential customers and provide more business opportunities for enterprises. For firms, the marketing influential value of each blogger can be measured clearly and the bloggers most worthy of being marketed can be easily identified by the proposed model. After the influential nodes have been appropriately identified, firms are able to develop some special marketing strategies to take advantages of these potential reviewers. For instance, enterprises can provide free trial versions of the new products or special discounts for these targeted bloggers. This proposed method provides a helpful and effective name list of bloggers to improve marketing behaviors.

There are still a few avenues for future research. First, this research focuses mainly on the discovery of influential bloggers. The blogging impact on real consumption behaviors (e.g. the number of sales or visits) of restaurants or food after the related promoting articles posed might be further examined. Second, in our framework, the influence strength is measured based on general content and network characteristics, which are domain-independent. However, it would be interesting to compare further the effectiveness of word-of-mouth marketing with respect to different application domains. Third, in addition to the articles and network characteristics, the influential power of online reviews might consider other factors. For example, the quality of embedded photographs in the comments could be considered as one of the factors of content-based value. Lastly, this research mainly focuses on the discovery of potentially influential bloggers. It is desirable that effective and efficient diffusion strategies based on these identified bloggers can be further developed.

Acknowledgment

This research was supported by the National Science Council of Taiwan (Republic of China) under the Grant NSC 97-2410-H-009-035-MY2.

References

[1] A. Abdul-Rahman, S. Hailes, Supporting trust in virtual communities, in: Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui, Hawaii, 2000, pp. 4–7.

[2] E. Adar, L. Zhang, L.A. Adamic, R.M. Lukose, Implicit structure and the dynamic of blogspace, in: Workshop on the Weblogging Ecosystem, 13th International World Wide Web Conference, New York, 2004.

[3] N. Agarwal, H. Liu, L. Tang, P.S. Yu, Identifying the influential bloggers in a community, in: Proceedings International Conference on Web Search and Web Data Mining, Palo Alto, California, USA, 2008, pp. 207–218.

[4] S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems 30 (1-7) (1998) 107–117.

[5] D. Bonhoeffer, Blog statistics – Length of stay, (http://www.livingroom.org.au/blog/archives/blog_statistics_length_of_stay.php), October 22, 2003, Accessed on 11.06.2008.

[6] S.P. Borgatti, M.G. Everett, L.C. Freeman, Ucinet for Windows: Software for Social Network Analysis, Analytic Technologies, Harvard, MA, 2002.

[7] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: Online book reviews, in: Working Paper 10148, National Bureau of Economic Research, December, 2003 (http://www.nber.org/papers/w10148) Accessed on 14.04.2009.

[8] W.K. Chiang, D. Zhang, L. Zhou, Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression, Decision Support Systems 41 (2) (2006) 514–531.

[9] F. Cacheda, A. Vina, Experiences retrieving information in the World Wide Web, in: Proceedings of the 6th IEEE Symposium on Computers and Communications, Hammamet, Tunisia, 2001, pp. 76–79.

[10] V. Carchiolo, A. Longheu, M. Malgeri, Reliable peers and useful resources: searching for the best personalised learning path in a trust- and recommendation-aware environment, Information Sciences 180 (10) (2010) 1893–1907.

[11] J. Caverlee, L. Liu, S. Webb, The social trust framework for trusted social information management: architecture and algorithms, Information Sciences (1801) (2010) 95–112.

[12] K. Dave, S. Lawrence, D.M. Pennock, Mining the peanut gallery: opinion extraction and semantic classification of product reviews, in: Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, 2003, pp. 519–528.
[13] X. Ding, B. Liu, P.S. Yu, A holistic lexicon-based approach to opinion mining, in: Proceedings of the International Conference on Web Search and Web Data Mining, Palo Alto, California, USA, 2008, pp. 231–240.
[14] A. Dobele, D. Toleman, M. Beverland, Controlled infection! spreading the brand message through viral marketing, Business Horizons 48 (2) (2005) 143–149.
[15] A. Dobele, A. Lindgreen, M. Beverland, J. Vanhamme, R. Wijk, Why pass on viral messages? Because they connect emotionally, Business Horizons 50 (4) (2007) 291–304.
[16] P. Domingos, M. Richardson, Mining the network value of customers, in: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, 2001, pp. 57–66.
[17] W. Duan, B. Gu, A.B. Whinston, Do online reviews matter? An empirical investigation of panel data, Decision Support Systems 45 (4) (2008) 1007–1016.
[18] L.C. Freemann, A set of measures of centrality based on betweenness, Sociometry 40 (1) (1977) 35–41.
[19] K. Fujimura, T. Inoue, M. Sugisaki, The EigenRumor algorithm for ranking blogs, in: Proceedings of the Second Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Chiba, Japan, 2005.
[20] K.E. Gill, How can we measure the influence of the blogosphere? in: Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, New York, 2004.
[21] S.A. Golder, D. Wilkinson, B.A. Huberman, Rhythms of social interaction: Messaging within a massive online network, in: Proceedings of the Third International Conference on Communities and Technologies, London, 2007, pp. 41–66.
[22] Y. Guan, X.L. Wang, X.Y. Kong, J. Zhao, Quantifying semantic similarity of Chinese words from Hownet, in: IEEE Proceedings of the International Conference on Machine Learning and Cybernetics, Beijing, 2002, pp. 234–239.
[23] F.Y. Guo, User experience research and management of online advertising and merchandising, in: Proceedings of the Third International Conference on Internationalization, Design and Global Development: Held as Part of HCI, San Diego, CA, LNCS, vol. 5623, 2009, pp. 457–466.
[24] C. Hamzacebi, Improving artificial neural networks' performance in seasonal time series forecasting, Information Sciences 178 (23) (2008) 4550–4559.
[25] C. Holscher, G. Strube, Web search behavior of Internet experts and newbies, International Journal of Computer and Telecommunications Networking 33 (1-6) (2000) 337–346.
[26] T. Johnson, K. Kaye, Wag the blog: how reliance on traditional media and the Internet influence credibility perceptions of weblogs among blog users, Journalism and Mass Communication Quarterly 81 (3) (2004) 622–642.
[27] B.J. Jansen, A. Spink, An analysis of web searching by European AlltheWeb.com users, Information Processing and Management 41 (2) (2005) 361–381.
[28] M.B.J. Jansen, A. Spink, T. Saracevic, Real life, real users, and real needs: a study and analysis of user queries on the web, Information Processing and Management 36 (2) (2000) 207–227.
[29] A. Java, P. Kolari, T. Finin, T. Oates, Modeling the spread of influence on the blogosphere, in: Proceedings of the 15th International World Wide Web Conference, Edinburgh, UK, 2006.
[30] A. Kale, A. Karandikar, P. Kolari, A. Java, A. Joshi, T. Finin, Modeling trust and influence in the blogosphere using link polarity, in: Proceedings of the Second International Conference on Weblogs and Social Media (Short Paper), Menlo Park, CA, 2007.
[31] P. Kazienko, M. Adamski, AdROSA – adaptive personalization of web advertising, Information Sciences (17711) (2007) 2269–2295.
[32] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 2003, pp. 137–146.
[33] C. Kiss, M. Bichler, Identification of influencers – measuring influence in customer networks, Decision Support Systems 46 (1) (2008) 233–253.
[34] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM 46 (5) (1999) 604–632.
[35] P. Kolari, T. Finin, K. Lyons, On the structure, properties and utility of internal corporate blogs, in: International Conference on Weblogs and Social Media, Boulder, Colorado, USA, 2007.
[36] A. Kritikopoulos, M. Sideri, I. Varlamis, BlogRank: Ranking weblogs based on connectivity and similarity features, in: Proceedings of the Second International Workshop on Advanced Architectures and Algorithms for Internet Delivery and Applications, Pisa, Italy, 2006.
[37] R.J. Kuo, 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 26 (2) (2004) 141–154.
[38] Y.M. Li, C.Y. Lai, C.W. Chen, Identifying bloggers with marketing influence in the blogoshpere, in: Proceedings of the 11th International Conference on Electronic Commerce, Taipei, Taiwan, 2009, pp. 335–340.
[39] Y.M. Li, C.W. Chen, A synthetical approach for blog recommendation: combining trust, social relation, and semantic analysis, Expert System and Applications 36 (3) (2009) 6536–6547.
[40] Y.R. Lin, H. Sundaram, Y. Chi, J. Tatemura, B.L. Tseng, Blog community discovery and evolution based on mutual awareness expansion, in: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, CA, USA, 2007, pp. 48–56.
[41] C.L. Lin, H.L. Tang, H.U. Kao, Utilizing social relationships for blog popularity mining, in: Proceedings of the 5th Asia Information Retrieval Symposium on Information Retrieval Technology, Sapporo, Japan, LNCS, vol. 5839, 2009, pp. 409–419.
[42] D. Mayzlin, Promotional chat on the Internet, Marketing Science 25 (2) (2006) 155–163.
[43] P. Nelson, Information and consumer behavior, Journal of Political Economy 78 (2) (1970) 311–329.
[44] H.L. Poh, T. Jasic, Forecasting and analysis of marketing data using neural networks: a case of advertising and promotion impact, in: Proceedings of the 11th Conference on Artificial Intelligence for Applications, Los Angeles, CA, USA, 1995, pp. 224–230.
[45] D.A. Reinstein, C.M. Snyder, The influence of expert reviews on consumer demand for experience goods: a case study of movie critics, Journal of Industry Economics 53 (1) (2005) 27–51.
[46] M. Richardson, P. Domingos, Mining knowledge-sharing sites for viral marketing, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, 2002, pp. 61–70.
[47] M.S. Sawhney, J. Eliashberg, A parsimonious model for forecasting gross box-office revenues of motion pictures, Marketing Science 15 (2) (1996) 113–131.
[48] R. Scoble, S. Israel, Naked Conversations: How Blogs are Changing the Way Businesses Talk with Customers, Hoboken, NJ, John Wiley & Sons, Inc., 2006.
[49] O. Shy, The Economics of Network Industries, Cambridge University Press, MA, USA, 2001.
[50] M.R. Subramani, B. Rajagopalan, Knowledge-sharing and influence in online social networks via viral marketing, Communications of the ACM 46 (12) (2003) 300–307.
[51] B. Widrow, S. Stearns, Adaptive Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1985.
[52] Wikipedia Wretch (<http://en.wikipedia.org/wiki/Wretch_%28website%29>), Accessed on 06.05.2008.
[53] Wikipedia PageRank (<http://zh.wikipedia.org/wiki/PageRank>), Accessed on 18.02.2008.
[54] WretchBlog (<http://www.wretch.cc/blog/>), Accessed on 06.05.2008.
[55] D.J. Xu, S.S. Liao, Q. Li, Combining empirical experimentation and modeling techniques: a design research approach for personalized mobile advertising applications, Decision Support Systems 44 (3) (2008) 710–724.
[56] Z. Yu, Z. Wu, H. Chen, H. Sheng, J. Ma, Mining target marketing groups from users' web of trust on Epinions, in: Proceedings of the Second International Conference on Weblogs and Social Media, Seattle, Washington, 2008, pp. 116–121.
[57] J. Zhan, H.T. Loh, Y. Liu, Gather customer concerns from online product reviews – a text summarization approach, Expert System and Applications 36 (2) (2009) 2107–2115.