附件一

科技部補助專題研究計畫成果報告

（□期中進度報告/☑ 期末報告）

社會網絡的資訊擴散

計畫類別：個別型計畫
計畫編號：MOST　101－2410－H－004－013－MY2
執行期間：101 年 08 月 01 日至 104 年 02 月 01 日
執行機構及系所：國立政治大學企業管理學系

計畫主持人：　唐揆
共同主持人：　方郁惠
計畫參與人員：吳家齊　博士後研究人員
林沛盈　碩士生-兼任助理人員
張君凱　碩士生-兼任助理人員
黃彥騰　碩士生-兼任助理人員
黃秋芸　碩士生-兼任助理人員
傅思瑜　碩士生-兼任助理人員
施雲天　碩士生-兼任助理人員

本計畫除繳交成果報告外，另含下列出國報告，共 2 份：
出席國際學術會議心得報告

期末報告處理方式：
1. 公開方式：非列管計畫亦不具下列情形，立即公開查詢
2.「本研究」是否已有嚴重損及公共利益之發現：否
3.「本報告」是否建議提供政府單位施政參考:否

中　華　民　國 104 年 3 月 15 日

附件二

科技部補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現（簡要敘述成果是否有嚴重損及公共利益之發現）或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估 ☑ 達成目標 ☐ 未達成目標（請說明，以 100 字為限） ☐ 實驗失敗 ☐ 因故實驗中斷 ☐ 其他原因 說明：
2. 研究成果在學術期刊發表或申請專利等情形： 論文：☑已發表☑未發表之文稿 ☐撰寫中 ☐無 專利：☐已獲得 ☐申請中 ☐無 技轉：☐已技轉 ☐洽談中 ☐無 其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性），如已有嚴重損及公共利益之發現，請簡述可能損及之相關程度（以 500 字為限）

Two related issues were studied in the past two years. In the first year, we concentrated on developing effective models for predicting the influence of posts on the social networks using both post- and author-related features as predictors. We offer the following recommendations for using posts in developing an effective advertising strategy:

1. We recommend considering jointly the influence of the contents and the author of the post in order to avoid undervaluing or overvaluing the influence of the post. We also recommend combining predictors from several sources, including both author- and content-related features, as well as other features such as temporal information, media type tents and time of creation, and features of the author of the post.
2. Basic text mining methods and specialized tools such as the sentiment analysis can be used to produce useful predictive features from the contents of posts.
3. The relationship between the influence score and predictors can be complicated. We recommend considering advanced learning tools to capture non-linear relationships.

In the second year, we studied how users on SNSs participate in the eWOM diffusion, which is crucial for the development of marketing strategies and the design of certain enhanced mechanisms on SNSs. Our findings reveal that receivers of SNSs are less likely to browse referral brands or product websites via the embedded hyperlink (referral visit behavior) than they are to engage in specific exploration. However, once they visit the embedded hyperlink and the referral website, they enjoy interesting or stimulating conditions so as to adopt the eWOM and even pass it to their friends. This tendency provides a promising opportunity for marketing strategy development.

1. Introduction

Social network websites (SNWs) provide users with a convenient and efficient online platform to share evaluations (reviews) of products with their contact groups. Because of the rapid proliferation of SNW participation, companies have been actively exploring the use of the vast amount of product reviews posted by SNW users to develop sustainable electronic word-of-mouth (eWOM) advertising strategies. Many companies – including, for example, Geico, Dell, and eBay – have been investing heavily in using social networks to influence consumer purchasing decisions (Kumar and Mirchandani, 2012; Evans and McKee, 2010). Empirical evidence has been established recently by academic researchers, which show the significant influence of online product reviews on consumer purchase decisions (Dellarocas, Zhang, and Awad, 2007). A survey by Gartner (2013) shows that "content creation" is the most important task for an e-marketer, and a survey by PowerReviews (2013) indicates that user-generated consumer reviews have the most influence on purchasing decisions among all the advertising contents (tools) on SWNs.

In order to utilize user-generated product reviews (posts) to develop eWOM advertising strategies, Facebook, the largest SNW, has launched several tools to engage users in sharing their own product ratings and reviews (Harris and Dennis, 2011). For example, a service is offered to allow e-marketers to use the posts shared by Facebook users to automatically generate advertisements. PowerReviews has developed a sophisticated and aggressive service to allow post writers to combine their profile data and posts on e-commerce sites (Wonham, 2010). When the service is used, a post can appear on the writer's Facebook Wall as well as in his or her friends' newsfeeds. It is expected that the trend of using self-generated product reviews in online advertising will expand rapidly in the future.

A practice of this nature relies on three postulations: (1) posts (information) shared by a "friend" could be very influential because they come from a trustworthy source and from first-hand experience; (2) certain "friends" are more influential than others; and (3) the quality and contents are important to the influence of a post (Yu et al., 2011). In order to develop effective advertising strategies in SNWs, it is essential to identify factors contributing to a successful eWOM process and to develop sound methods for predicting the influence of posted information.

In general, the influence of a post is determined jointly by the features of the post, such as contents and time of creation, and by the features of the author of the post. Although there is a rich literature on influence in social networks, these two sources of influence have been defined and studied individually. Existing studies on author-related features focus on identification of influential users (influencers) or opinion leaders (Bakshy et al., 2011; Cha et al., 2010; Kim and Han, 2009; Kiss and Bichler, 2008; Li et al., 2010; Li et al., 2011; Li and Du, 2011), and studies on post-related features investigate methods for identifying important predictive features and effective models for predicting the influence of the information shared among users (Adamic et al., 2008; Bian et al., 2009; Cao et al., 2011; Hong et al., 2011; Ratkiewicz et al., 2010; Suh et al., 2010; Yu et al., 2011). Furthermore, the influence of a post is often measured by the number of users who respond favorably to the post, such as "like" clicked counts (Yu et al., 2011), browser counts (Ratkiewicz et al., 2010), and forward counts (Hong et al., 2011).

In order to ensure the effectiveness of using a post in advertising on SNWs, we need a model for predicting the post's influence with a satisfactory level of accuracy. We discuss three components associated with developing an effective predictive model: the definition of the target variable, the selection of predictors, and the selection of a functional form for linking the target variable and predictors.

It is evident that the influence of a post would be greatly undervalued if the variation in influence among users is not considered, when it receives a favorable response from a smaller number of influential users, or vice versa. To define the target variable of the model, which can adequately reflect the influence of a post, we first assign a weight to each user according to the number of users linked to him or her. Then, we define the "influence score" of a post as the sum of the weights of the people who respond favorably to the post.

In order to identify predictors for the predictive model, we compile a list of potential predictive features from two sources. The first is a comprehensive literature review to identify influencers and predict the influence of the contents of a post. The second source is our own investigation, based on the definition of the influence score. Using these predictive features, we propose two predictive models. The first is multiple-regression analysis, a useful and popular statistical tool. It is known that regression is based on a linear pattern for describing the relationship between the target and predictors. To allow for more flexible patterns, we also propose an ensemble classification model based on five techniques, including four data mining methods – Neural Networks, Decision Trees (C5.0), Naive Bayes, and Support Vector Machines (SVM) – and a classical statistical method, Logistic Regressions. It should be noted that, in the second model, the influence score is divided into several levels (intervals). We formulate the problem as a multi-label classification problem (as in Adamic et al., 2008; Hong et al., 2011; Yu et al., 2011), and propose a framework to predict the post's influence level. To improve the overall accuracy of prediction, we exploit an ensemble model that aggregates the predictions made by multiple classifiers. From the evaluation of the two predictive models, we identify important predictive features and develop models for accurately predicting the influence of a post.

The remainder of this paper is organized as follows. In Section 2, we review the related literature, and, in Section 3, we present the definition of the influence score, predictive features, and the two predictive models. An empirical analysis of the predictive features and evaluations of the two predictive models are given in Section 4. A conclusion with a discussion for future studies is given in Section 5.

2. Literature Review

We review the literature separately on finding influencers on SNWs and assessing the influence of the contents of a post. We identify five categories of predictive features for influence on SNWs; namely, content, authorial, temporal, exogenous, and topological. We provide a detailed literature review of content and authorial features and introduce briefly the three remaining types of features. These features, along with additional features proposed in this study, will be used to develop the predicted models in the next section.

In general, authorial features are characteristics of the author. A list of recent and representative studies on finding influencers is given in Table 1. Bakshy et al. (2011) studied the influence of Twitter users by analyzing users' attributes and diffusion data. Cha et al. (2010) investigated the dynamics of the influence of users across different topics based on three measures: in degree (the number of followers), re-tweets, and mentions (the number of times a user's name is mentioned). Kim and Han (2009) proposed a method for identifying influencers by analyzing users' activities on a social graph. They empirically validated positive effects of influencers on spreading social games among Facebook users. Kiss and Bichler (2008) adopted centrality measures in selecting influencers in a customer network. They selected initial sets of customers from a telecom company with different centrality measures and evaluated these measures according to the number of other members in the network reached by these initial customers. Li et al. (2010) developed a framework to estimate the influence capability of reviewers online. Their framework aggregates modified the point-wise mutual information (PMI) and adaptive recency, frequency, and monetary (RFM) models by using an artificial neural network (ANN). Li et al. (2011) analyzed three types of blog characteristics (network-based, content-based, and activeness-based), and used an artificial neural network to develop a framework for evaluating the influential strength of bloggers and discovering potential bloggers on Wretch. Li and Du (2011) proposed a framework, called BARR, to identify opinion leaders on social blogs. Using the information retrieved from blog content, as well as the author properties, reader properties, and relationships between them, the BARR framework evaluates blogs by a modified technique for order preference by similarity (TOPSIS) first, and then identifies opinion leaders according to the quality and quantity of their published blogs.

Table 1. Representative Studies on High-Influence Users

Authors	Platform	Features
Bakshy et al. (2011)	Twitter	Number of followers, friends, and tweets Past influence of users

Cha et al. (2010)	Twitter	Number of followers, re-tweets, and mentions
Kim and Han (2009)	Facebook	Degree centrality, number of groups belonged to, pages linked, and updated per day
Kiss and Bichler (2008)	Call records	Centrality (degree, hubs, page rank)
Li et al.(2010)	Review Website	Number of subjective terms, review recency and frequency
Li et al. (2011)	Wretch Blog	Network-based factors, content-based factors, and activeness-based factors
Li and Du (2011)	My Space	Author properties, reader properties, relationship, content, and number of published blogs

Content features are characteristics associated with a post, such as the length and the number of sentiment words. For assessing the influence of a post, or an article on SNWs, we compile a list of representative studies in Table 2. We briefly describe these studies as follows.

Table 2. Representative Studies on High-Influence Content

Authors	**Platform**	**Features**	**Definition of Influence**
Adamic et al. (2008)	Yahoo! Question	Content characteristics User attributes Patterns of interaction among users	Scores from users
Bian et al. (2009)	Yahoo! Question	Question features Answer features User features	Scores from users
Cao et al. (2011)	online software market	Basic characteristics Stylistic characteristics Semantic characteristics	Number of helpfulness votes
Hong et al. (2011)	Twitter	Content of tweets Topological features Temporal information Metadata of tweets and users	Number of forwards
Ratkiewicz et al. (2010)	Wikipedia	Exogenous factors (news, temporal information)	Number of browsers
Suh et al. (2010)	Twitter	Content features Contextual features	Number of forwards
Yu et al. (2011)	Facebook	Content of the messages Media type of the post	Number of “likes” clicked

Adamic et al. (2008) studied knowledge-sharing activities on Yahoo! Answers and characterized the entropy of the users’ interests. They combined both user attributes and answer characteristics to predict whether a particular answer would be chosen as the best answer by the person who posted the question. Bian et al. (2009) argued that existing supervised approaches for estimating content quality in a community question answering (CQA) environment require a large amount of manually labeled data. Based on the mutually reinforcing relationship between user reputation and the quality of the content produced by them, they developed a semi-supervised learning method to identify high-quality content and users in Yahoo! Answers. Cao et al. (2011) argued that, in most related studies, users were required to vote on whether a review was helpful, in order to define or measure the helpfulness of a review. However, most online reviews never receive votes, which does not suggest that they are not helpful. They extracted three types of review characteristics (basic, stylistic, and semantic), and then used text mining and ordinal logistic regression to examine the impact of the three characteristics on the number of helpfulness votes a review received. Hong et al. (2011) investigated the problem of predicting the popularity of tweets on Twitter using the number of re-tweets as the measure. They formulated the task as a classification problem using a wide variety of features, such as the content of the tweets, temporal information, metadata and users, as well as the structural properties of the social graph of the users. Ratkiewicz et al. (2010) evaluated the dynamics and popularity of Wikipedia topics and web pages. They proposed a model that combined exogenous factors, such as the publication of news and temporal information. The popularity of a document was measured by the number of clicks and the number of hyperlinks pointing to it. Suh et al. (2010) focused on re-tweeting on Twitter. They examined a number of features that might affect re-tweetability of tweets, including content features and contextual features, and built a model to predict the re-tweetability of a user's tweet. Yu et al. (2011) investigated the characteristics of social marketing messages that contributed to different levels of popularity. They analyzed the messages posted by restaurants with the most Facebook fans and used the number of "likes" to measure the popularity of a message. They examined the characteristics of a message from keywords in the content and also considered the media type, whether "status," "link," "video," or "photo," of the message.

Temporal features are time-related properties, such as the time of creation temporal (Bian et al., 2009; Cao et al., 2011; Hong et al., 2011). Exogenous features describe external events that trigger collective attention (Ratkiewicz et al., 2010). They could be events in the news or topical subjects such as an actor winning a prize, political elections, etc. Topological features include global and local user-related structures, such as page rank, degree distribution, local clustering coefficient, and reciprocal links (Hong et al., 2011).

3. Predictive Models

In this section, we present two models for predicting the influence score of a post. Since the target, the influence score, is continuous, multiple-regression analysis is an appropriate method for developing a prediction model and analyzing the relationship between the features and influence score. Regression analysis is a useful and popular tool for investigating the relationship between the response variable and a set of potential predictors, but it is limited in the (linear) pattern used to describe the relationship. In order to allow more flexible patterns, we also propose an ensemble classification model based on five techniques; namely, Neural Networks, Decision Trees (C5.0), Naive Bayes, Support Vector Machines (SVM), and Logistic Regressions.

All of the studies included in Table 2 adopt the number of accesses as the estimate for the influence of a post. However, the studies included in Table 1 suggest that individual users may have different levels of influence, suggesting individuals' weights should be considered when measuring the influence of a post. Note that the weight of a user can be measured in different ways; for example, by the number of friends, the number of messages he/she has posted, or the number of "like" clicks he/she has received. We use the number of friends as the weight of a user in defining the influence score of a post. Let d_j be the number of friends of user j. The weight of user i, w_i, is defined by:

$$w_i = \frac{d_i}{\max\limits_{\forall \text{user } j}(d_j)}.$$

Accordingly, the influence score of a post is determined by

$$\sum_{\text{All click the post}} w_i.$$

In the next subsection, we discuss the features used in the two proposed predictive models. The regression model and the ensemble model are presented in the two subsequent subsections.

3.1. Predictive Features

The features considered in developing our prediction models are listed in Table 3. Note that those features marked with an "*" have not been considered in previous studies. In our framework, we include only features directly related to the target post, i.e., author and content information. Exogenous and topological features are not included because of the difficulty of collecting them in practice, which could reduce the practicality of the proposed framework.

Table 3. List of Features Considered in Developing Our Prediction Models

Category	Type	Features	Definition
Content	Length	length	Number of words in the post
	Sentiment words	num_positive	Number of positive words in the post
		ratio_positive	Ratio of positive words
		positive 1/0	Whether the number of positive words in the post exceeds the average number of positive words in all posts made by the same author
		num_negative	Number of negative words in the post
		ratio_negative	Ratio of negative words in the post
		negative 1/0	Whether the number of negative words in the post exceeds the average number of positive words in all posts made by the same author
	Punctuation	num_question	Number of question marks in a post
		num_exclamation	Number of exclamation marks in a post
		ratio_punctuation	Ratio of punctuation marks in the post
	Media type	link	Whether the post contains photos, videos, or links
	Stop words	num_stpword	Number of stop words in the post
		ratio_stopword	Ratio of stop words to total words in the post
	Tags	post_tags*	Number of tags for people's names the post contains
Temporal		day*	Whether the post was made on a weekend or a week day (0: weekend; 1: week day)
		time*	The time the post is created
Authorial	Personal	gender	The gender of the author
		age	The age of the author
	Structural	friendCount	Number of friends of the author
		friends_total_weight*	Total weight of the author's friends
		ratio_of_high_weights_friends*	Ratio of friends with high weights to total friends of the author
		avg_like*	Average weighted sum of "likes" for the author
		avg_comment*	Average weighted sum of comments on the author
		photo_tags	Number of times the author is tagged in photos
	Activity	postCount	Number of posts the author has made
		photos	Number of photos the author has posted
		groups	Number of groups to which the author belongs

The feature set includes four new authorial features; namely, friends_total_weight, ratio_of_high_weights_friends, avg_like, and avg_comment. These features are used to reflect the real size of the author's contact group and to measure his/her "popularity" among the contacts. We believe these features are essential in predicting the influence of a post. Two temporal aspects are also considered: day and time. It is expected that the day and the time of a day that a post is published affect its influence because of the nature of information flow in Facebook. Finally, we also include the number of tags for users' names a post contains.

3.2. Multiple Regression

Multiple regression is a useful statistical method for analyzing the relationship between the target variable, i.e., the influence score, and predictive features, and for developing models to predict the influence score. It is expected that the features are highly correlated; so, it is difficult to evaluate the importance of a feature by a single regression model. In order to investigate the importance of the features in predicting the influence score, we use the best subset regression procedure with the R^2 selection criterion to select regression models for given numbers of predictive features.

3.3. Ensemble Model

We develop an ensemble model for predicting when the influence score is divided into ranges. The model is shown in Figure 1, in which multiple classifiers are aggregated using a voting method. The ensemble model uses a voting method to integrate results from five classifier techniques: Neural Networks, Decision Trees (C5.0), Logistic Regressions, Naive Bayes, and Support Vector Machines (SVM). In an ensemble model, a single classifier predicts the influence score of a post, and votes for its prediction. Each post is predicted to the influence level that receives the most votes. To deal with the condition where two levels receive the same number of votes, we construct additional models. For example, suppose there are three possible levels of influence for a post, Level 1, Level 2, and Level 3; four models are constructed as shown in Table 4. It can be seen that Model 1 involves all classes, whereas the other three models, 2, 3, and 4, consider only two of the three influence scores. After voting in Model 1, if there are two levels that stand out, we use the model corresponding to these two levels to determine the final result.

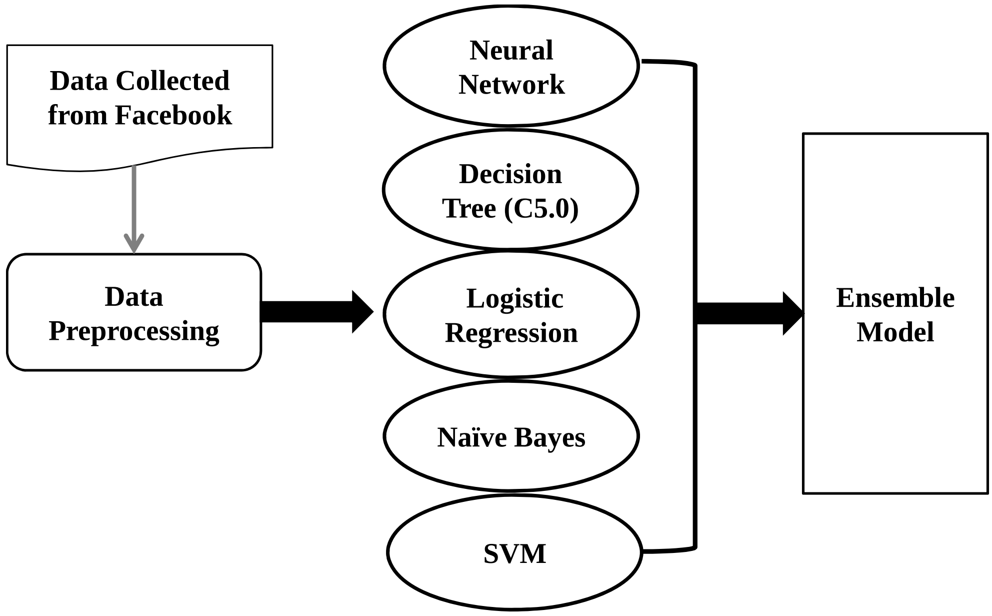

Figure 1. Prototype of Our Framework

Table 4. Ensemble Models for Three Classes

Model	Class
1	Level 1, Level 2, Level 3
2	Level 1, Level 2
3	Level 1, Level 3
4	Level 2, Level 3

4. Evaluation

We conduct an empirical evaluation to investigate (1) the relationship between the predictive features and the influence score, and (2) the effectiveness of the features in predicting the influence score of a post. For the evaluation, we selected a Facebook user and recorded all the posts from her contacts (friends). In the data preparation step, we eliminated advertisements and the posts artificially generated by computer software. As a result, we obtained, in total, 510 posts from 31 friends. The age range of the friends was from 21 to 32 years, with an average age of 24.9 years. The number of groups a friend belonged to ranged from 1 to 254, with an average of 48.6. Of the 510 posts, 244 were made by females and 266 by males. A total of 139 of the 510 posts were created on weekends, and the others were created on week days. The number of posts per person ranged from 2 to 47, with an average of 24, and the number of friends ranged from 153 to 651, with an average of 307.

The posts collected were written in Chinese. We use the Chinese knowledge information processing (CKIP) laboratory system, which is an on-line remote Chinese word-segmenting service offered by the Academia Sinica in Taiwan, to apply word segmentation. The data preparation process is illustrated by Figure 2. The NTUSD Chinese opinion dictionary (Ku et al., 2006), containing 2,812 positive and 8,276 negative opinion words (as also adopted in (Ku et al., 2007; Tan et al., 2008)), is used to identify positive and negative words. In order to analyze the relation between numbers of stop words and the influence score of a post, we count the number of common stop words according to the standard Chinese stop-words list. The number and ratio of question and exclamation marks in each post are also calculated. Finally, we compute the influence score for each post from the weighted sum of "likes" clicked.

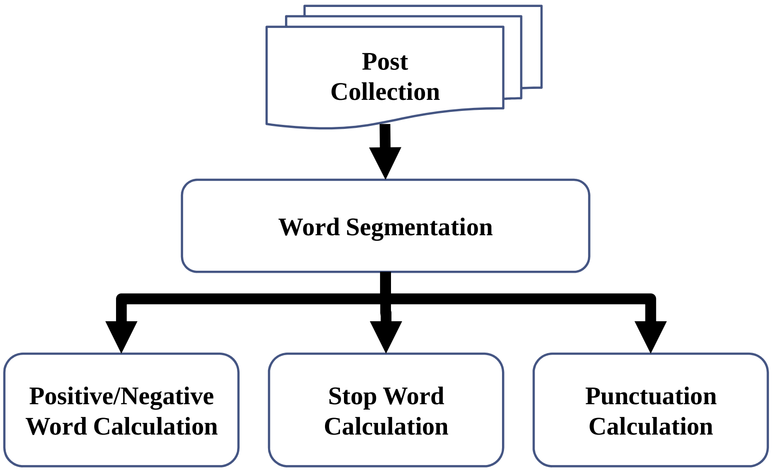

Figure 2. Data Preparation Process

For evaluating the ensemble model, we divide the 510 influence scores into several intervals by k-mean analysis (MacQueen, 1967). The boundaries between these intervals are adjusted slightly to clearer divisions, as shown in Table 5. The numbers of posts associated with the three levels, [0, 2.0), [2.0, 6.0), and [6.0, ∞), are 310, 155, and 45, respectively. For the division with four levels, the numbers of posts associated with the levels, [0, 1.5), [1.5, 4.5), [4.5, 6.5), and [6.5, ∞), are 265, 157, 51, and 37, respectively.

Table 5. Definition of Influence Levels

Influence Level	Division into Three Levels	Division into Four Levels
1	0 ≦ influence score < 2.0	0 ≦ influence score < 1.5
2	2.0 ≦ influence score < 6.0	1.5 ≦ influence score < 4.5
3	6.0 ≦ influence score	4.5 ≦ influence score < 6.5
4	N/A	6.5 ≦ influence score

Using the data, we produce a scatter plot for the influence score and the like count. As shown in Figure 3, the two variables are highly correlated, but their relationship is far from perfect. Therefore, using the like count as an influence measure may undervalue or overvalue the true influence level of a post.

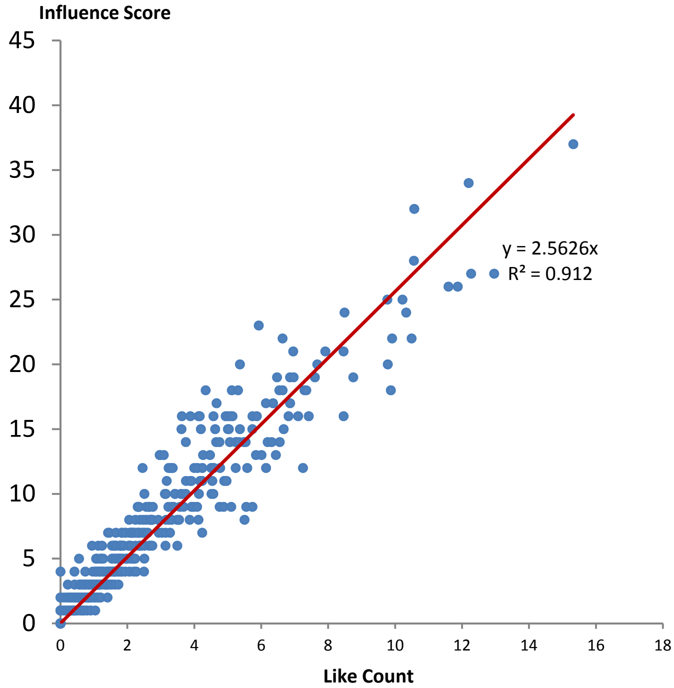

Figure 3. Scatter Plot of Influence Score and Like Count

4.1. Correlation Analysis

The Pearson’s correlation values for each feature are shown in Table 6. The Pearson's correlation coefficient (expressed as a value between +1 and −1) is a widely used measure for the linear association between two variables.

Table 6. Pearson’s Correlation Values Between Features and Influence Level

Features	**Influence Score**	**3 Levels**	**4 Levels**
Link	-0.097*	-0.110*	-0.091*
num_positive	0.076	0.099*	0.068
num_negative	0.148**	0.134**	0.129**
num_stopword	0.152**	0.162**	0.150**
num_exclamation	0.053	0.052	0.051
num_question	-0.006	-0.032	-0.035
Length	0.181**	0.161**	0.158**
ratio_positive	0.004	-0.008	-0.005
ratio_negative	-0.030	-0.028	-0.043
ratio_stopword	-0.014	0.001	0.009
positive 1/0	0.085	0.085	0.076
negative 1/0	0.096*	0.085	0.074
ratio_puntuation	-0.043	-0.047	-0.052
Age	-0.057	-0.031	-0.053
Gender	0.011	-0.001	-0.009
Groups	-0.073	-0.085	-0.062
Photos	0.163**	0.131**	0.149**
photo_tags	0.016	0.004	0.001
friendCount	0.276**	0.223**	0.263**
postCount	-0.109	-0.082	-0.096*
Day	0.040♦	0.065§	0.076§
Time	0.095♦	0.219§	0.156§
post_tags	0.078	0.061	0.100*
avg_like	0.563**	0.482**	0.519**
avg_comment	0.210**	0.203**	0.225**
friends_total_weight	0.218**	0.169**	0.208**
ratio_of_high_weights_friends	0.137*	0.113*	0.148**

Note: *$p < 0.05$; **$p < 0.01$; § phi coefficient; ♦ square root of R^2

Several features, such as “avg_like,” “avg_comment,” “friends_total_weight,” and “ratio_of_high_weights_friends,” are first considered as potential predictive features in this study to reflect variation in influence among different users. In addition, we also introduce temporal and tag features. The Pearson’s correlation values related to these features are grouped in the bottom of Table 6. Most of the features have significant correlations with the influence score with the exception of “post_tags,” “day,” and “time.” Note that an un-significant correlation does not suggest that the corresponding feature is not useful. We will further examine the relationship between these features and the influence score in the next two subsections.

4.2. Regression Models

The R^2 of the multiple-regression model with all the features included as the predictors is 0.3832. We find that some of the features are highly correlated, as indicated by large variance inflation factors. Consequently, it is difficult to evaluate the importance of a feature by a single regression model. In order to investigate the importance of the features in predicting the influence score, we use the SAS regression procedure to select regression models with 7 to 12 features and R^2 larger than 0.37. In total, 198 regression models are found. We use the percentage that each feature is included in the 198 models as a measure for the relative importance of a feature in predicting the influence score. The result is reported in Figure 4.

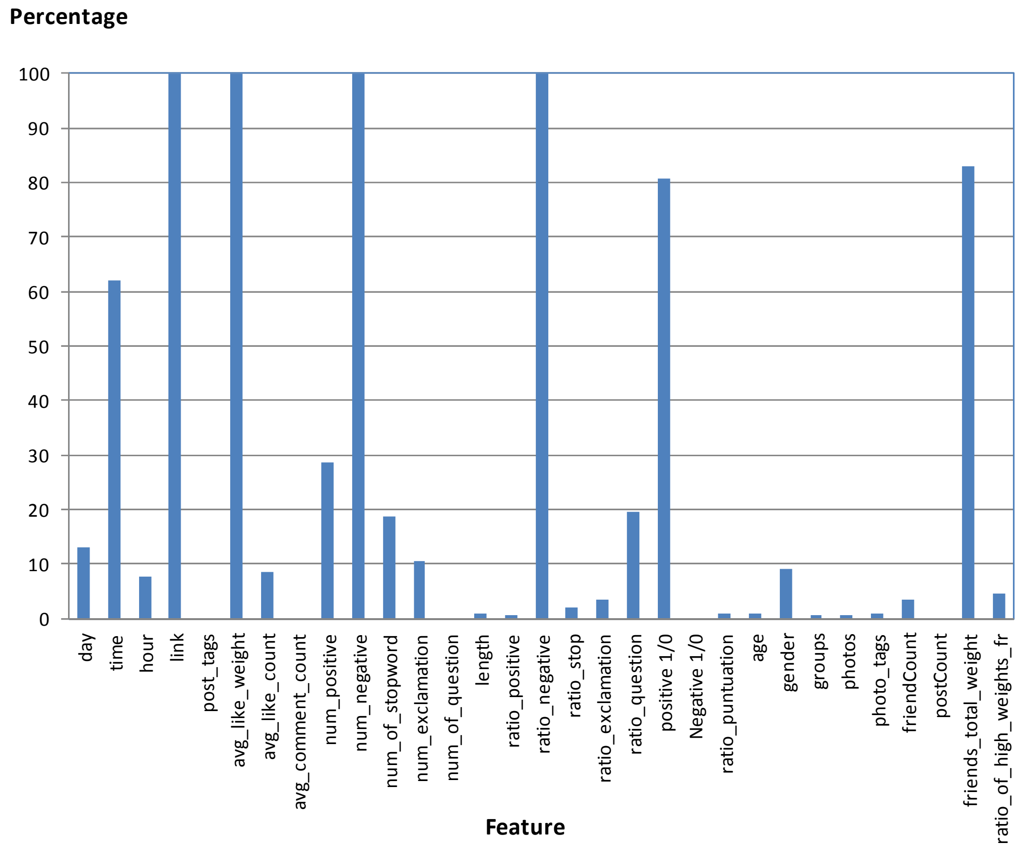

Figure 4. Percentage of a Feature Included in the Regression Models

Four features appear in all the models: "link," "average_like_weight," "number_negative," and "ratio_ negative." Three other features appear in at least 60% of the models: "time," "positive 1/0," and "friends_total_weight." In addition, two features, "number_positive" and "ratio_questions," appear in 20% or more of the models. Generally speaking, these features cover sentiment words, temporal, and friends related statistics. In addition to multiple regression, we also try using regression tree to produce a predictive model. The tree contains only "average_like_weight," "number_negative," and "ratio_ negative." The R^2 value of the tree is 0.3644, which is slightly smaller than those of the models reported in Figure 4.

4.3. Ensemble Model

For comparisons, we use three feature sets shown in Table 7 to evaluate the predictive features used in predicting the influence levels. The first set contains only the features in Table 3 that have been discussed in the literature. The features of the second set are identical to those in the first set with two exceptions. First, we use user weights for several features in the second set, whereas user counts are used in the first set. That is, "average count of 'likes' clicked for the author's posts (28)" and "average count of comments on the author's posts (29)" are in the first set, and "average weighted sum of 'likes' of the author's posts (22)" and "average weighted sum of comments on the author's posts (23)" are in the second set. Second, two additional features, "total weight of friends of the author (20)" and "ratio of friends with high weights to total friends of the author (21)" are added to the second set. We expand set 2 by including several additional features to create set 3. These additional features are the published time, the tags of people's name, "post_tags (14)," "day (15)," and "time (16)." Note that set 3 consists of all the features from existing studies (1-13, 17-19, 24-27) and all the features first proposed in this study (14-16, 20-23).

Table 7. Feature Sets Used in the Evaluation

	Set 1	Set 2	Set 3
(1) Length	●	●	●
(2) Number of positive words	●	●	●
(3) Ratio of positive words to total characters	●	●	●
(4) Whether the number of positive words exceeds the average number of positive words in posts made by the same author	●	●	●
(5) Number of negative words	●	●	●
(6) Ratio of negative words to total characters	●	●	●
(7) Whether the number of negative words exceeds the average number of negative words in posts made by the same author	●	●	●
(8) Number of question marks	●	●	●
(9) Number of exclamation marks	●	●	●
(10) Ratio of punctuation marks to total characters	●	●	●
(11) Whether a post contains photos, videos, or links	●	●	●
(12) Number of stop words	●	●	●
(13) Ratio of stop words to total words	●	●	●
(14) Number of tags of people's name contained in a post			●
(15) Whether a post was made on a week day or a weekend			●
(16) Time a post was created			●
(17) Gender of the author	●	●	●
(18) Age of the author	●	●	●
(19) Number of friends of the author	●	●	●
(20) Total weight of friends of the author		●	●
(21) Ratio of friends with high weights to total friends of the author		●	●
(22) Average weighted sum of "likes" for the author's posts		●	●
(23) Average weighted sum of comments on the author's posts		●	●
(24) Number of times the author is tagged in photos	●	●	●
(25) Number of posts the author has made	●	●	●
(26) Number of photos of the author posted	●	●	●
(27) Number of groups the author belongs to	●	●	●
(28) Average count of "likes" clicked for the author's posts	●		
(29) Average count of comments on the author's posts	●		

The ensemble model is trained using the three feature sets. We use accuracy and F1-measures (Yang, 1999) to evaluate the performance of the models. The accuracy is the ratio of the number of correct classifications to the total number of classifications. The F1-measure is the harmonic mean between precision and recall. Assume the recall of class i is r_i and the precision of class i is p_i. The F1-measure of class i, f_i, is defined as $f_i = 2p_ir_i/(p_i + r_i)$. A larger value for the F1-measure corresponds to a higher classification quality. The overall classification performance can be evaluated by the micro-average F1-measure or the macro-average F1-measure. The macro-average F1-measure is more appropriate when the class distribution is skewed. Considering the pattern of the data, we use the macro-average F1-measure to evaluate the overall classification performance.

The experimental results are shown in Table 8. The results were obtained by 10-fold cross-validation. The original dataset was randomly partitioned into ten subsets. In each test, nine of the ten subsets were used as the training data to produce the model, and the remaining data was used to evaluate the model. The process was repeated ten times, and each subset was used as test data exactly once.

Table 8. Accuracy and F1-measure for Predicting with Three Feature Sets

a. Three Influence Levels

Set	Accuracy	f_1	f_2	f_3	f_4	macro-average F1
1	0.625490	0.762946	0.195429	0.220496	N/A	0.392957
2	0.643137	0.777618	0.32163	0.262185	N/A	0.453811
3	0.692157	0.802539	0.396531	0.587279	N/A	0.621495

b. Four Influence Levels

Set	Accuracy	f_1	f_2	f_3	f_4	macro-average F1
1	0.519608	0.680509	0.158012	0.107974	0.149589	0.274021
2	0.537255	0.707007	0.194559	0.110518	0.235462	0.311887
3	0.605882	0.742532	0.353499	0.297511	0.386266	0.652363

As shown in Table 10, set 3 always produced the best results, followed by set 2. This suggests that the features proposed in this study significantly improve the prediction results.

The ensemble model integrates results from five classifiers, i.e., Neural Networks, Decision Trees (C5.0), Logistic Regression, Naive Bayes, and SVM, by a voting method. We compare the performance of the ensemble model with that of the five individual classifiers. Feature set 3 is used in this experiment. The results are shown in Table 9. The accuracy and F1-measure of the ensemble model are always better than the others. The results show that our ensemble model gives better and more robust prediction than any of the individual classifiers even though the voting step of the ensemble model costs a little extra computation time.

Table 9. Accuracy and F1-measure of the Ensemble Model and Five Individual Classifiers

a. Three Influence Levels

	Neural Network	C5.0	Logistic Regression	SVM	Naïve Bayes	Ensemble
Accuracy	0.647059	0.643137	0.65098	0.54902	0.621569	0.692157
f_1	0.784373	0.777446	0.773722	0.649087	0.759516	0.802539
f_2	0.322371	0.304712	0.365534	0.372147	0.323543	0.396531
f_3	0.315158	0.423580	0.460148	0.498513	0.429171	0.587279
macro-average F1	0.511370	0.523962	0.547989	0.506996	0.512041	0.621495

b. Four Influence Levels

	Neural Network	C5.0	Logistic Regression	SVM	Naïve Bayes	Ensemble
Accuracy	0.560784	0.545098	0.552941	0.437255	0.523529	0.605882
f_1	0.731360	0.704712	0.696767	0.556256	0.684000	0.742532
f_2	0.200572	0.304520	0.322003	0.340776	0.335605	0.353499
f_3	0.116086	0.240503	0.244726	0.211333	0.202610	0.297511
f_4	0.282139	0.329457	0.357939	0.254545	0.293627	0.386266
macro-average F1	0.507674	0.558217	0.561685	0.456854	0.515904	0.652363

5. Discussion

In this study, we consider a situation where posts shared by users on SNWs is used in eWOM for advertising. We propose two models for predicting the influence of a post using both sources of influence, post- and author-related features, as predictors. The major results, contributions, limitations, and future research follow.

5.1. Summary of Results

As discussed, we believe the effectiveness of using posts in advertising depends on the influence of post-related characteristics and those of the individual who posts the information. We also define the "influence score" of a post by considering the variation in influence among users of SNWs. Under this definition, we propose potential predictive features from two sources. The first is a comprehensive review of the literature for identifying influencers and for predicting the influence of the contents of a post. The second is from our own investigation.

We consider two scenarios for developing predictive models. In the first, the influence score is used as the target variable. Since the influence score is continuous, multiple-regression analysis is an appropriate method for developing a prediction model and analyzing the relationship between the features and influence score. Since some of the predictive features are expected to be highly correlated, we use all regression with the R^2 selection criterion to identify important features for predicting the influence score. In the second scenario, the influence score is divided into several ranges (intervals). We formulate the problem as a multi-label classification problem. To improve the overall accuracy of prediction, we exploit an ensemble model that aggregates the predictions made by multiple classifiers.

In the empirical study, the results of regression analysis indicate six of the seven new predictors proposed in this paper have a significant correlation with the influence score. The only exception is photo_tags. In addition, three proposed predictors (average_like_weight, time, and friend_total_weight) are included in either all or 60% of the regression models for predicting the influence score. For the ensemble model, the third feature set, which includes all the features from existing studies and all the features proposed in this study, leads to much higher overall prediction accuracy and quality than the other sets. For each influence level, the results associated with the third feature set are consistently and significantly better than those associated with the first and second sets. These empirical evidences provide a strong support for the proposed new predictive features.

5.2 Practical Implications

Based on the findings of this study, we offer the following recommendations to companies for developing an effective advertising strategy based on posts in SNWs:

1. Selecting a post for advertising can be facilitated by a highly sophisticated mathematic model for predicting the influence of the post with a satisfactory level of accuracy. In order for an effective predictive model to be developed, the target variable must adequately reflect the influence of the post. We recommend considering jointly the influence of the contents and the author of the post in order to avoid undervaluing or overvaluing the influence of the post. When the proposed target variable is defined, a weight may be assigned to each user according to the number of other users linked to him or her to reflect the user's influence. Then, the influence of a post may be defined as the sum of the weights of the people who respond favorably to the post. We also recommend combining predictors from several sources, including both author- and content-related features, as well as other features such as temporal information, media type tents and time of creation, and features of the author of the post. As the evaluation suggests, the proposed framework consisting of the target variables and predictors can significantly improve the accuracy of the prediction, thus resulting in a successful advertising strategy.
2. Basic text mining methods and specialized tools such as the sentiment analysis can be used to produce useful predictive features from the contents of posts. Our evaluation in Section 4.2 shows that several features generated from the sentiment analysis – i.e., "number_negative," "ratio_negative," "positive 1/0," and "number_positive" – were identified as useful predictors. Although this exact finding may not be directly applicable in other applications, it provides strong support for using these new methods in analyzing the eWOM process and developing effective online marketing strategies. This finding is especially important for developing a highly accurate predictive model, as text mining techniques are becoming more sophisticated and computationally efficient.
3. The relationship between the influence score and predictors can be complicated. We recommend considering advanced learning tools to capture non-linear relationships. For the continuous influence score, popular methods include Neural Networks, Regression Trees, and Support Vector Machines. For the discrete influence score, the four methods considered in this study may be used; i.e., Decision (Classification) Tree, Naïve Bayes, and Support Vector Machines. These methods have been incorporated in several data mining packages, such as SAS Enterprise Miner and IBM-SPSS Clementine. Open Source codes such as Weka (Hall, Frank, Holmes, Pfahringer, Reutemann, Witten, 2009) are also widely available. An ensemble model may be used to combine the results of these methods in producing the final prediction.

5.3 Limitations and Future Research

This study has several limitations that can be improved in future studies to strengthen its results and theoretical implications. First, the data size and the diversity of the users may be insufficient to completely support the conclusions. We may utilize a tool such as Facebook API to acquire a large dataset with diversity in cultural, racial, ethnic, and socioeconomic backgrounds. Second, besides the number of those clicking "like," the number of those replying to or forwarding a post might also be indicative of its influence. It is possible to consider additional types of interactions between the author and the reader as a reflection of the influence of the content. Finally, we may add flexibility to the ensemble model by assigning different weights to different classifiers and may use the Neural Network to find the weights for achieving the best performance.

Using eWOM strategies in social networks has become a new paradigm for advertising. Related issues will be topics for researchers for many years to come. In the near term, we plan to investigate different types of interactions between the author of a post and its readers in affecting the influence of a post. Furthermore, we will investigate additional forms of response to a post (including forwards, comments, or replies with a favorable or unfavorable statement, which represent different degrees of support for a post).

References

Adamic, L., Zhang, J., Bakshy, E., Ackerman, M., 2008. Knowledge sharing and Yahoo answers: Everyone knows something, Proceedings of the 17th International Conference on World Wide Web. 665-674.

Bakshy, E., Hofman, J., Mason, W., Watt, D.J., 2011. Everyone's an influencer: Quantifying influence on Twitter, Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. 65-74.

Bian, J., Liu, Y., Zhou, D., Agichtein, E., Zha, H., 2009. Learning to recognize reliable users and content in social media with coupled mutual reinforcement, Proceedings of the 18th International Conference on the World Wide Web. 51-60.

Cao, Q., Duan, W., Gan, Q., 2011. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. Decision Support Systems 50, 511-521.

Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P., 2010. Measuring user influence in Twitter: The million follower fallacy, Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. 10-17.

Dellarocas, C., Zhang, X., Awad, N.F., 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive Marketing 21, 23-45.

Evans, D., McKe, J., 2010. Social media marketing: The next generation of business engagement, Wiley Publishing, Indianapolis, Indiana.

Gartner 2013. Gartner's Social Marketing Survey 2013. Accessed from http://blogs.gartner.com/adam-sarner/2013/04/01/gartners-social-marketing-survey-2013/.

Golbeck, J., Robles, C., Turner, K., 2011. Predicting personality with social media, Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems. 253-262.

Hong, L., Dan, O., Davison, B.D., 2011. Predicting popular messages in Twitter, Proceedings of International Conference on the World Wide Web. 57–58.

Harris, L., Dennis, C, 2011. Engaging customers on Facebook: challenges for e-retailers, Journal of Consumer Behaviour 10, 338-346.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update, SIGKDD explorations 11, 10-18.

Kiss, C., Bichler, M., 2008. Identification of influencers: Measuring influence in customer networks, Decision Support Systems 46, 233–253.

Ku, L.W., Liang, Y.T., Chen, H.H., 2006. Opinion extraction, summarization and tracking in news and blog corpora, Proceedings of the American Association for Artificial Intelligence. 100–107.

Ku, L.W., Wu, T.H., Lee, L.Y., Chen, H.H., 2007. Using polarity scores of words for sentence level opinion extraction, Proceedings of the Sixth NTCIR Workshop. 316-322.

Kumar, V., and Mirchandani, 2012, Increasing the ROI of Social Media Marketing, MIT Sloan Management Review. 54, 55-61.

Lakkaraju, H., Rai, A., Merugu, S., 2011. Smart news feeds for social networks using scalable joint latent factor models, Proceedings of the 20th International Conference Companion on World Wide Web. 73-74.

Li, N., Wu, D.D., 2010. Using text mining and sentiment analysis for online forums hotspot detection and forecast, Decision Support Systems 48, 354-368.

Li, Y.M., Lin, C.H., Lai, C.Y., 2010. Identifying influential reviewers for word-of-mouth marketing, Electronic Commerce Research and Applications 9, 294-304.

Li, Y.M., Lai, C.Y., Chen, C.W., 2011. Discovering influencers for marketing in the blogosphere, Information Sciences 181, 5143-5157.

Li, F., Du, T.C., 2011. Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs, Decision Support Systems 51, 190-197.

MacQueen, J., 1967. Some methods for classification and analysis multivariate observation, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. 281–297.

McAndrew, F.T., de Jonge, C.R., 2011. Electronic person perception: What do we infer about people from the style of their e-mail messages? Social Psychological and Personality Science 2, 403-407.

Paek, T., Gamon, M., Counts, S., Chickering, D.M., Dhesi, A., 2010. Predicting the importance of newsfeed posts and social network friends, Proceedings of the Twenty-fourth AAAI Conference on Artificial Intelligence.1419–1424.

PowerReviews Inc. 2011, The 2011 Social Shopping Study. Accessed from www.powerreviews.com/.../Social_Shopping_2011_Brief1
Ratkiewicz, J., Menczer, F., Fortunato, S., Flammini, A., Vespignani, A., 2010. Characterizing and modeling the dynamics of online popularity, Physical Review Letters 105.
Silva, C., Ribeiro, B., 2003. The importance of stop word removal on recall values in text categorization, Proceedings of the International Joint Conference on Neural Networks. 1661-1666.
Suh, B., Hong, L., Pirolli, P., Chi, E., 2010. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network, Proceedings of IEEE Second International Conference on Social Computing. 177-184.
Tan, S., Wang, Y., Cheng, X., 2008. Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples, Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 743–744.
Welser, H.T., Gleave, E., Fisher, D., Smith, M., 2007. Visualizing the signatures of social roles in online discussion groups, The Journal of Social Structure 8 (2).
Wonham, L.2010. The power of reviews meets the reach of Facebook, Website Magazine. Accessed from
http://www.websitemagazine.com/content/blogs/posts/archive/2010/09/22/the-power-of-reviews-meets-the-reach-of-facebook.aspx
Yang, Y., 1999. An evaluation of statistical approaches to text categorization, Information Retrieval of Computer Science 1 (1-2), 69-90.
Yu, B., Chen, M., Kwok, L., 2011. Toward predicting popularity of social marketing messages, Lecture Notes in Computer Science 6589, 317–324.
Zhou, Y., Cao, Z., 2011. Research on the construction and filter method of stop-word list in text preprocessing, Proceedings of Fourth International Conference on Intelligent Computation Technology and Automation. 217-221.

已發表論文

Y. L. Chen, K. Tang, C. C. Wu, J. Y. Jheng, (2014) “Predicting the Influence of Users’ Posted Information for eWOM Advertising in Social Networks,” Electronic Commerce Research and Applications, 13, 431-439.

Kao, H. P. and K. Tang (2014), “Cost-Sensitive Decision Tree Induction with Label-Dependent Late Constraints,” INFORMS Journal of Computing, 26(2), 238-252.

另有一篇論文正在國外期刊審查中:

On Electronic Word-of-Mouth Diffusion in Social Networks: Curiosity and Influence.