Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2013 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

6-18-2013

The More the Worse? Mining Valuable Ideas with Sentiment Analysis for Idea Recommendation

Hanjun Lee
Korea University, hjlee1609@gmail.com

Keunho Choi
Korea University, keunho@korea.ac.kr

Donghee Yoo
Korea Military Academy, donghee.info@gmail.com

Yongmoo Suh
Korea University, ymsuh@korea.ac.kr

Guijia He
Soongsil University, twofirst@hotmail.com

See next page for additional authors

Authors
Hanjun Lee, Keunho Choi, Donghee Yoo, Yongmoo Suh, Guijia He, and Soowon Lee

THE MORE THE WORSE? MINING VALUABLE IDEAS WITH SENTIMENT ANALYSIS FOR IDEA RECOMMENDATION

Hanjun Lee, Business School, Korea University, Seoul, Republic of Korea, hjlee1609@gmail.com

Keunho Choi[*], Business School, Korea University, Seoul, Republic of Korea, keunho@korea.ac.kr

Donghee Yoo, Department of Electronics Engineering and Information Science, Korea Military Academy, Seoul, Republic of Korea, donghee.info@gmail.com

Yongmoo Suh, Business School, Korea University, Seoul, Republic of Korea, ymsuh@korea.ac.kr

Guijia He, School of Computer Science and Engineering, Soongsil University, Seoul, Republic of Korea, twofirst@hotmail.com

Soowon Lee, School of Computer Science and Engineering, Soongsil University, Seoul, Republic of Korea, swlee@ssu.ac.kr

Abstract

Many firms have an interest in an open innovation community, recognizing its business value. They can collect and analyze the ideas of their customers from the community to get valuable ideas which can lead to innovation such as a new product or service. However, such a community overloaded with too many ideas from customers cannot make use of them at the right time because of the limited time and human resources to deal with them. Therefore, it would be a great help to those firms if they have a recommendation system which recommends top n ideas for innovation. MyStarbucksIdea (MSI) is such an open community, created by Starbucks. To build such an innovative idea recommendation system for Starbucks, we analyzed a dataset collected from MSI, utilizing data mining and sentiment analysis techniques. Experimental results show that our recommendation system can help firms identify prospective ideas which can be valuable enough for their innovation among a large amount of ideas, efficiently.

Keywords: Open innovation, MyStarbucksIdea, Recommendation system, Sentiment analysis, Data mining.

[*] Corresponding author
Tel: +82 2 3290 1945
Fax: +82 2 922 7220
E-mail: keunho@korea.ac.kr

1 INTRODUCTION

Innovation can be described as the creation of new ideas and knowledge for meeting various customer requirements and thus developing new business values by improving internal business processes and structures (Gloet and Terziovski 2004). The need for innovation has been a top priority among CEOs (Andrew et al. 2010; Jaruzelski and Dehoff 2010) and one of the critical issues among academic researchers (Krishnan and Ulrich 2001, Hauser et al. 2006). Traditionally, firms have generated ideas for innovation in invention, development and design, depending on internal staffs of professional inventors and those ideas have been considered as competitive advantages. Most firms have taken this "closed" approach for their innovations, and accordingly, the innovations have been restricted to internal resources (March 1991, Ahlstrom 2010, Wyld 2010).

Rapid change and intense competition in today's business environment, however, has changed the way how firms initiate and develop their innovations. Since it is almost impossible for a single firm to obtain all of the relevant resources in the field, firms have begun to acquire external knowledge to complement their limited internal resources (Von Hippel 1988, Cassiman and Veugelers 2006, Beamish and Lupton 2009). Consequently, innovation-related activities such as concept generation, needs-finding, and idea generation which usually have been done in a top-down approach in traditional business, now are regarded to be done better in a bottom-up approach, together with customers. In light of this trend, the concept, "open innovation", represents a paradigm that emphasizes the use of knowledge from outside an organization as well as inside (Chesbrough 2003). This paradigm advocates organizations to take more open strategies to complement internal innovation processes (Laursen et al. 2006, Chesbrough 2003a). Accordingly, firms across diverse industry such as Dell, Starbucks and P&G, have taken the open strategies by managing their own online open innovation communities. Such community can be a strategic asset for organizational innovation in the aspects of: 1) providing external expertise, 2) generating valuable ideas, and 3) supporting innovation development (Dahlander and Wallin 2006). Starbucks, for instance, is gathering more than 1,000 ideas per week, and has gathered over 150,000 ideas from its online open innovation community up to date. These ideas are discussed, developed and adopted to contribute to Starbucks in innovating itself.

In most of such communities, as far as the number of ideas collected is concerned, more is better. Large number of ideas will increase their diversity and it is likely to result in more options for firms to adopt innovation. However, this can bring a serious information overload problem, too. As discussed above, open innovation necessarily requires discussion among customers and review on each idea, and consequently an overflowing number of ideas will make the processes difficult and ineffective. Ironically, such active ideation can even hinder effective idea generation. Therefore, the objective of this paper is to mitigate the information overload problem in an open innovation environment by recommending top-N ideas with the highest adoption probability. To that end, we analyzed the dataset collected from MyStarbucksIdea (MSI), utilizing data mining and sentiment analysis, while considering both term-based features and other features of each idea. Based on the result, it will be possible for firms to get recommendations on prospective ideas which can be valuable enough for their innovation in initial ideation stage.

The rest of this paper is organized as follows. In the next section, we review previous researches on open innovation and sentiment analysis. In Section 3, our research context is explained. Section 4 presents our proposed method including research framework and model building. Section 5 presents an experimental design including dataset, preprocessing and feature selection. In Section 6, we compare the experiment results and discuss their implications. The last section concludes our study with summary, implications, limitations, and future research.

2 LITERATURE REVIEW

2.1 Open Innovation and Information Overload

Open innovation is an emerging trend in many firms across industries. The term, open innovation, was coined by Chesbrough (2003). He described its concept as a paradigm where valuable ideas can come from inside or outside of the firm and also can go to markets from inside or outside of the firm. He also described that open innovation process leads to a critical amount of external knowledge exploration and exploitation. From a simple viewpoint, the open innovation can be explained as the flow of ideas into and out of a firm.

Prior researches recommend firms to acquire external knowledge to complement their internal innovation processes. Laursen et al. (Laursen and Salter 2006) emphasized the importance of the network of relationships between a firm and its external environment in terms of leveraging firm's performance. Chesbrough (2003) advocated that firms should adopt open innovation approach and seek to combine internal and external components of an organization to accomplish innovations successfully, especially in today's boundary free world. Some firms today, implement this open approach by launching customer-driven open innovation communities where customers freely share and discuss their opinions relevant to a firm's products or services. In these communities, customers are encouraged to participate in ideation and interaction for innovation. That is, customers share their own ideas relevant to the products or services of the firm and discuss together to develop an idea in the better shape. It is known that these communities are effective, because customers tend to have specialized knowledge about their own problems with existing products and services, and they are motivated intrinsically to share their ideas to contribute to the firms (von Hippel 2005, Fuller 2010). Furthermore, under the right conditions, customers can generate ideas which are worthy of implementation (Kavadias and Sommer 2009, Magnusson 2009, Poetz and Schreier 2010, Girotra et al. 2010). Thus, exploiting an open innovation community can bring business values by leveraging an organization's capacity to continuously advance its competencies and better adopt themselves to the changing business environment (Teece and Pisano 1994).

There exist various challenges within idea generation communities that can affect their performance, including usefulness and effectiveness. In case the community member's participation is not vibrant enough, motivating them to engage proactively can be a key issue. Willingness to engage, which serves as a motivation, has been a critical issue as in other ordinary communities, because innovative ideas will not possibly be generated without members' ideation and discussion. Thus, many researchers have paid attention to this motivation issue (Ardichvili et al. 2003, Wang and Fesenmaier 2003, Füller et al. 2008). On the other hand, information overload resulting from too many ideas can also cause serious problem to occur. That is, sufficient interaction for idea development, which is necessarily required for open innovation, cannot be expected due to too much information to be processed, even if many ideas might be collected. In our study, focusing on this problem of information overload, we propose a recommendation system aiming to mitigate the problem. Our system provides top idea list in terms of their adoptability to firm.

2.2 Sentiment Analysis

Sentiment analysis also known as opinion mining makes it possible to analyze texts containing opinions and sentiment automatically (Thelwall et al. 2011). Traditionally, sentiment analyses are conducted by extracting entities and classifying whether their opinions are positive or negative (Lerner and Keltner 2000, Zhuang et al. 2006). Sentiment detection which dates back to late 1990s (Kessler et al. 1997, Spertus 1997, Argamon et al. 1998) has drawn a growing attention in information management discipline since 2000s (Pang et al. 2002, Nasukawa and Yi 2003, Li and Wu 2010). Sentiment analysis is now one of the most major research areas in Natural Language Processing and many industries are also showing a strong interest in sentiment analysis (Pang et al. 2002, Nasukawa and Yi 2003, Li and Wu 2010, Liu 2012). Accordingly, large firms have attempted to develop their own analytical ability (Pang, Lee et al. 2002, Liu 2012).

Mainly, machine learning techniques such as Support Vector Machine (SVM), Naive Bayes and Maximum Entropy are widely used to classify texts into positive or negative categories (Cui et al. 2006, Pang et al. 2002). More recently, matrix factorization techniques have also been used to conduct sentiment classification with lexical prior knowledge (Li et al. 2009). In these researches, review texts such as product reviews or movie reviews have been the main context to which researchers generally applied sentiment analysis method (Pang et al. 2002, Whitelaw et al. 2005, Zhuang et al. 2006). Meaningful results from these researches prove that sentiment analysis can be an effective and useful way to analyze the text contents containing sentiment and emotion. In our research, therefore, we take sentiment analysis methodology to extract the sentiment contained in each idea and comment collected from MSI.com site.

3 RESEARCH CONTEXT

Our research context is MSI which is one of the most popular online open innovation communities currently. Starbucks launched this community in March 2008 for the purpose of gathering useful innovative ideas from customers directly. Up to date, more than 150,000 ideas had been posted on MSI, and Starbucks could find useful ideas among them which led to its innovation.

Starbucks provides fifteen idea categories such as Coffee & Espresso Drinks, Food and Starbucks Card as shown in Table 1. Each idea should belong to one of the categories, which is determined when the idea is posted by a customer.

Area	Category
Product	Coffee & Espresso Drinks, Frappuccino Beverage, Tea & Other Drinks, Food, Merchandise & Music, Starbucks Card, New Technology, Other Product Ideas
Experience	Ordering & Payment & Pick-Up, Atmosphere & Locations, Other Experience Ideas
Involvement	Building Community, Social Responsibility, Other Involvement Ideas, Outside USA

Table 1. MSI Idea Classification

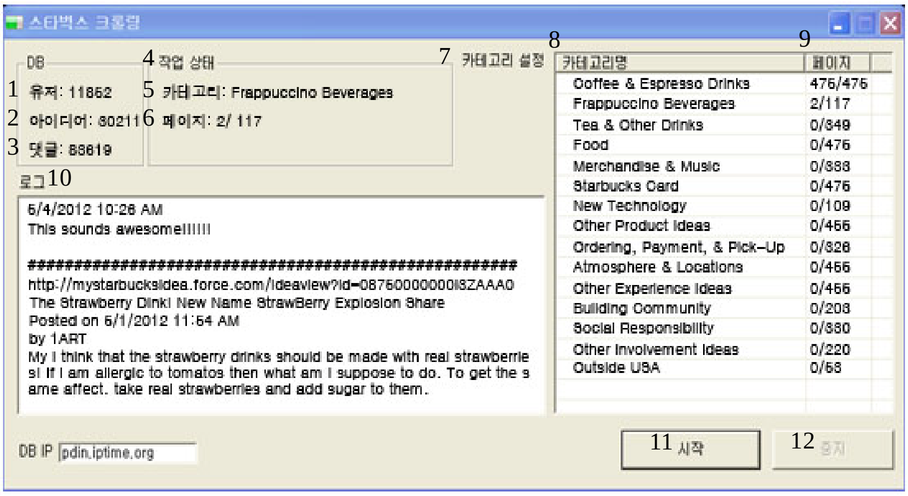
Activities in MSI consist of four stages: 1) share, 2) vote, 3) discuss, and 4) see. In share stage, customers share their ideas. Any member can share any idea in MSI by posting it. In vote stage, customers vote to promote favorable ideas and also to demote unfavorable ideas. Each idea gets its own point, increased or decreased depending on customers’ promotion or demotion, respectively. This voting can be viewed as customer-led quantified idea evaluation, and the point represents how many customers agree to it. In discuss stage, which is carried out simultaneously with vote stage, customers comment on other ideas to discuss them in more detail. Through the discussion, an initial idea is refined, further developed and evaluated. Based on each idea’s votes received and the discussion on it, ideas evaluated valuable go into the internal review process. This review process is open to all customers in see stage.

Starbucks hired MSI managers, called Idea partner, to ask them to go through initial review of all posted ideas manually. Since a large number of ideas are being posted every day, it can be a great help if there is a way of recommendation for adoptable ideas.

4 PROPOSED METHOD

4.1 Data

In order to collect ideas of customers who have enrolled with MSI, we implemented a web crawler using Visual Basic 6.0, as shown in Figure 1. From July 2, 2007 through July 15, 2007 (2 weeks), we collected using the crawler information about users, ideas of each user, comments on each idea, votes on each idea, and whether each idea had been adopted or not from all of fifteen idea categories (see the right part of Figure 1). Since our dataset was so highly skewed (i.e., only 360 adopted ideas among 84,918 collected ideas), total 720 ideas (i.e., 360 adopted ideas and 360 rejected ideas) were used in our experiments after preprocessing and balancing.



1. User, 2. Idea, 3. Comment, 4: Status, 5. Category, 6. Page, 7. Category setting, 8. Category name, 9. Page, 10. Log, 11. Start, 12. Stop

Figure 1. User interface of our web crawler

4.2 Overall Process

First, we made three datasets (i.e., Cat A[1], Cat B[2] and Cat C[3] datasets), each of which consists of ideas in categories of Cat A, Cat B and Cat C, respectively. And then, we further made two datasets (i.e., each consists of either term-based or non-term-based dataset) from each of the above three datasets. As a result, total six datasets were generated.

Using the above six datasets, classification models are built after selecting a set of variables from each dataset. And then, two adoption probabilities predicted by two classification models built from a term-based dataset and a non-term-based dataset in each of the three datasets are integrated. Finally, the top n ideas with the highest integrated adoption probability are recommended. Figure 2 depicts the overall process in case of Cat A dataset. The same process is applied to the other two datasets.

---

[1] Cat A contains these categories: 1. coffee & espresso drinks, 2. food, and 3. Starbucks card.

[2] Cat B contains these categories: 4. atmosphere & locations, 5. building community, 6. Frappuccino beverages, 7. merchandise & music, 8. new technology, 9. ordering, payment & pick-up, 10. social responsibility, 11. tea and other drinks, 12. other experience ideas, 13. other involvement ideas, 14. other product ideas, and 15. outside USA.

[3] Cat C contains categories in Cat A and Cat B

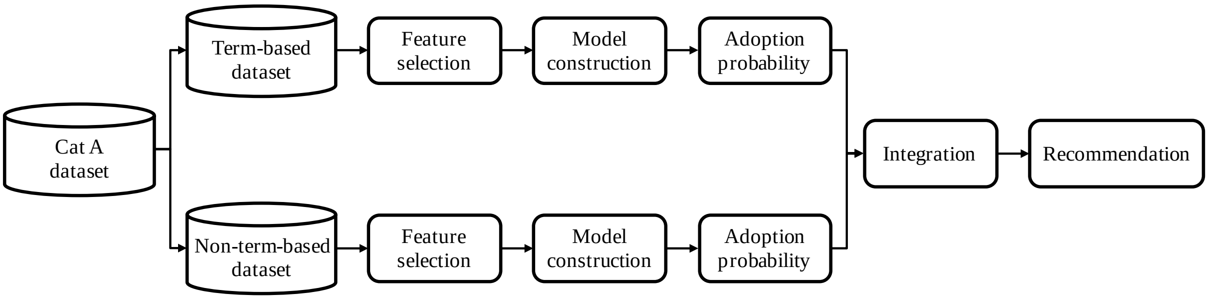


Figure 2. Overall process in case of Cat A dataset

4.3 Term-Based Dataset

To investigate the effect of feature types on classification accuracy, we made two different datasets which consist of term-based features and non-term-based features, respectively. This section explains how to make term-based dataset, and next section explains how to make the other dataset in detail.

In term-based dataset, each idea is represented as a set of weights of distinct terms extracted from the idea and term frequency–inverse document frequency is used as the weight (TF-IDF) (see Eq. (1)). Unlike news text, idea text is written with free formats, arbitrary words, abbreviations, and mistakes. Therefore, it is an important step to remove noisy words and retain useful words before building classification models. As is done in Kouloumpis et al. (2011), we applied the following steps to each idea which consists of its title and content: 1) Tokenization; 2) Normalization; 3) Recognition.

In tokenization step, URLs, emoticons, and specific symbols in an idea are removed, and the idea is split into a bag of words. Since MSI is a worldwide idea-posting website, there exist some spelling differences. Therefore, in normalization step, the British spellings of English words are transformed into American spellings. Subsequently, words are stemmed to their root formats in order to reduce feature dimensions. Finally, in recognition step, some relevant nouns related to Starbucks are recognized to be used later based on a list of products and materials in Starbucks menu, and the meaningless words (i.e., stop words) and words which do not exist in English dictionary due to noises and mistakes are removed.

Features in term-based dataset are a set of distinct words that appear in ideas, which can be different depending on the *N*-gram policy (e.g., unigram and bigram) (Pang et al. 2002). There have been some issues regarding the *N*-gram policy (Park and Paroubek 2010). If only unigram words are used, some distinct and important information can be ignored, while if only bigram words are used, they cannot represent each idea well when the number of words in each idea is not large or the length of each idea is not long as in our dataset. It should be noted that we limited the size of *N* to 1 or 2 (i.e., unigram and bigram). Therefore, we used both unigram and bigram words as features to represent each idea as follows.

$$\text{Idea } i = \{fw_1, fw_2, \dots, fw_m\}$$

, where fw_m denotes the weight of feature m in an idea. One of widely used techniques to define the weight is the TF-IDF, which is defined as follows.

$$TF - IDF_{ti} = \frac{f_{ti}}{max_{a \in T_i}(f_{ai})} \times \ln \frac{\#I}{\#I_t} \quad (1)$$

, where f_{ti} denotes the frequency of term *t* in idea *i*, *I* denotes a set of ideas, T_i denotes a set of terms in idea *i*, and It denotes a set of ideas including term *t*. The TF-IDF$_{ti}$ reflects the relative importance of a term *t* in an idea *i* referring to all ideas (Park and Paroubek 2010).

4.4 Non-Term-Based Dataset

In non-term-based dataset, each idea is represented as a set of features such as information of customers who proposed the idea (e.g., user point, number of submitted ideas, number of voting, and whether a customer is a top commenter or not), comments on the idea, votes on the idea, and sentimental information of the idea and its comments.

This study assumed that terms in an idea not only present customer's opinion but also contain his/her sentiment, which may affect the adoptability of the idea. Therefore, we thought that it would be better if the sentimental information is used when predicting the adoptability of the idea. In this study, we used a tool, SentiWordNet (Esuli and Sebastiani 2006), to obtain sentimental information (i.e., positive and negative polarity) from each idea. SentiWordNet is a lexical resource for opinion mining based on WordNet (Miller 1995). In SentiWordNet, a word can contain more than one part of speech (POS) such as noun, verb, adjective, and adverb. Moreover, a word can have one or more meanings within a certain POS.

For each meaning of a word, there are two scores each of which represents positive or negative polarity of the meaning, respectively. If the difference of the two scores (i.e., sentiment score of the meaning: SSM) has a positive value, it implies that the meaning possesses positive sentiment, while a negative value implies that the meaning possesses negative sentiment. In addition, each meaning has one index number (i.e., IndexNum) which represents the order of importance of the meaning in a word. A low IndexNum means that it is used frequently, while a high IndexNum implies that people seldom use the word to express the meaning. The sentiment score of a word (i.e., SSW) is calculated as the weighted average of SSMs of all meanings within a specific POS that the word represents in a specific sentence, where we used the inverse-IndexNum (i.e., 1/IndexNum) of each meaning as a weight on each meaning's SSM. Note that only noun, verb, adjective and adverb are considered in our experiment since each of them possesses sentiment to some extent. After calculating the SSW, the sentiment score of an idea (i.e., SSI) is calculated as the arithmetic average of SSWs in the idea in order to exclude the effect of the length of ideas on the sentiment score. The pseudo-code of sentiment analysis is described in Figure 3.

For example, when there is an idea, "*I am sad that the bagels are gone*", SSW of a word, sad, which is one of words in this idea, is calculated as follows: 1) POS of the word in the idea is identified as adjective by POS tagging method; 2) SSMs of the three meanings of the word within the POS (i.e., adjective) is calculated using SentiWordNet, as shown in Table 2; 3) SSW of the word is calculated as the weighted average of the SSMs (i.e., {(-0.625)*1 + (-0.25)*(1/2) + (-1)*(1/3)}/{(1 + (1/2) + (1/3)}). SSI of the idea is then calculated as the arithmetic average of such SSWs. The sentiment score of a comment on an idea can be calculated similarly.

Word	POS	IndexNum	Positivity	Negativity	SSM
sad	adjective	1	0.125	0.75	-0.625
sad	adjective	2	0	0.25	-0.25
sad	adjective	3	0	1	-1

Table 2. An example of SSMs of the word, sad

```
Input: words in an idea and meanings of each word with SSM
Initialize SSI = SumofWeightedSSM = Sumof Weight = 0
For each word in an Idea
    For each meaning of the word (whose POS is one of Noun, Verb, Adjective or Adverb)
        SumofWeightedSSM += SSM/IndexNum
        SumofWeight += 1/IndexNum
    End
    SSW = SumofWeightedSSM/SumofWeight
    SSI += SSW
End
SSI = SSI/Number of words in the idea
Output: SSI
```

Figure 3. Pseudo-code of sentiment analysis based on SentiWordNet

4.5 Model Construction

From the above six datasets, classification models are built using machine learning techniques (i.e., artificial neural network (ANN) and decision tree (DT)) and statistical techniques (Bayesian network (BN) and logistic regression (LR)).

Both a term-based classification model and a non-term-based classification model built from each category dataset predict adoption of each idea and yield adoption probability. And then, their adoption probabilities are integrated in the next recommendation step, so that the effect of both term-based and non-term-based features on prediction accuracy can be reflected when making better prediction.

4.6 Recommendation

In order to integrate the adoption probabilities of an idea yielded by both the term-based and the non-term-based classification models, respectively, we used weighted average of them as follows.

$$P_{Wi} = \alpha \times P_{Ti} + \beta \times P_{Ni} \quad (2)$$

, where P_{Wi} denotes the weighted average of an idea i, P_{Ti} and P_{Ni} denote the adoption probabilities of an idea i predicted by classification models built using the term-based dataset and the non-term-based dataset, respectively, and α and β denote the weight of P_{Ti} and P_{Ni}, respectively. We used the prediction accuracy (i.e., hit ratio) of each classification model as its weight. After calculating P_{Wi}, the top n ideas with the highest P_{Wi} are recommended.

5 EXPERIMENTS

5.1 Feature Selection

Prior to building classification models, we evaluated features of a term-based dataset in each category dataset using chi-square algorithm and selected more influential features when predicting the target feature. Similarly to the case of a term-based dataset, we also evaluated features of a non-term-based dataset and selected 8, 11, and 13 influential non-term-based features in Cat A, Cat B, and Cat C dataset, respectively. Table 3 shows a set of non-term-based features selected in each dataset.

No.	Non-term-based features	Cat A dataset	Cat B dataset	Cat C dataset
1	Idea point	O	O	O
2	The number of objective words	O	O	O
3	The number of positive words	O	O	O
4	Total number of words	O	O	O

5	Total number of distinct words	O	O	O
6	The number of negative words	O	O	O
7	Idea category	O	O	O
8	Top commenter	O	O	O
9	The number of non-polarity words	-	O	O
10	Deviation of comments	-	O	-
11	The number of comments	-	O	-
12	Negativity score	-	-	O
13	Positivity score	-	-	O
14	Average length of comments	-	-	O
15	Average number of objective words in comments	-	-	O

Table 3. Selected non-term-based features

5.2 Experimental Design

To construct classification models, we used Weka ver. 3.6 (open source software) as a data mining tool, which is widely used for various data mining analysis. We divided each of six datasets into training dataset and test dataset. The 80 percent of each dataset was used as training data and the rest as test data. We also used hit ratio as an evaluation measure for classification accuracy, and precision, recall, and F1 as evaluation measures for recommendation accuracy.

As mentioned in Section 4.5, ideas are sorted in descending order of adoption probability (i.e., P_{Wi}), and the top *n* ideas are recommended. However, there may be several ideas with same adoption probability, which make it difficult to rank them. In this situation, the accuracy of classification models can be different according to which ideas are recommended among them with same adoption probability. In order to mitigate this issue, therefore, we first made two policies, the best case policy and the worst case policy. The best case policy recommends really adopted ideas first among the ideas with same adoption probability, which can be identified based on test dataset. In contrast, the worst case policy recommends really not-adopted ideas first among the ideas with same adoption probability. And then, we used the average precision, recall, and F1 obtained from the two different policies to measure the final accuracy.

6 EXPERIMENTAL RESULTS

This section explains the experimental results in terms of both classification accuracy and recommendation accuracy.

6.1 Classification Results

Figure 4 shows the hit ratio of classification models in three different category datasets. As shown in Figure 4, the term-based classification models achieved better classification accuracy than the non-term-based classification models in case of ANN and LR, while the non-term-based classification models yielded better classification accuracy than the term-based classification models in case of DT and BN in all category datasets. Also, the hybrid classification models based on P_{Wi} outperformed both the term-based and the non-term-based classification models in all category datasets except BN in Cat A dataset and ANN in Cat C dataset[4]. This result shows that both the term-based features and the non-

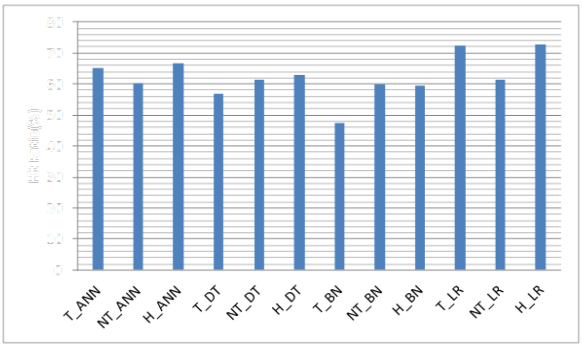
[4] T_algorithm and NT_algorithm denote classification models built from term-based dataset and non-term-based dataset, respectively, using the corresponding algorithm. H_algorithm denotes a hybrid classification model using the corresponding algorithm based on P_{Wi}.

term-based features play important roles in predicting adoptability of ideas5. In general, classification models built from Cat B dataset showed better classification accuracy than those built from the other two datasets. That may be due to the fact that Cat B dataset consists of more homogeneous categories than Cat C dataset and the number of instances in Cat B dataset is larger than that in Cat A dataset. The best classification accuracy was achieved by the hybrid classification models using LR, BN, and DT in Cat A, Cat B, and Cat C datasets, respectively.

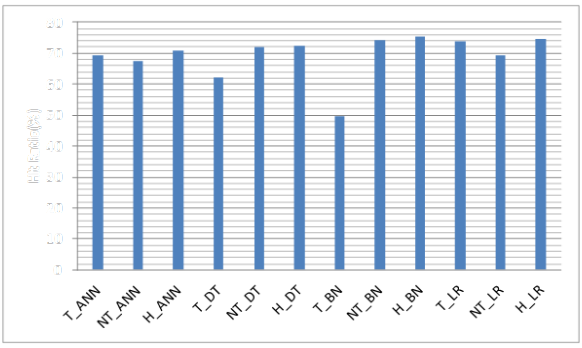
6.2 Recommendation Results

Actually, to predict whether ideas will be adopted or not cannot mitigate the information overload problem effectively, since there may be still a number of ideas predicted to be adopted. Thus, instead of just predicting the adoptability of ideas, to recommend top *n* ideas with the highest adoption probability can give more benefits to decision makers of MSI.

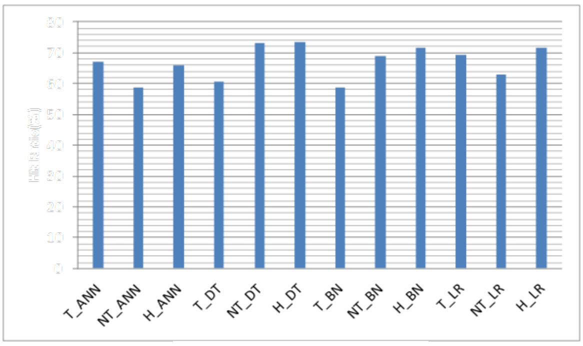
Figs. 5, 6 and 7 show precision, recall, and F1 of the hybrid classification models, respectively. As shown in Figure 5, the best precision was achieved by H_LR and H_ANN when the number of recommendations is 5 in Cat A dataset. Especially, H_LR outperformed other classification models, regardless of the number of recommendations. In Cat B dataset, the best precision was achieved by H_BN when the number of recommendations is 5. As the number of recommendations increases from 10 to 20, H_ANN showed better precision than other classification models. When the number of recommendations is 25 and 30, H_LR outperformed the others. In Cat C dataset, the best precision was achieved by H_ANN and H_LR when the number of recommendations is 5 similarly to the case of Cat A dataset. Especially, H_ANN outperformed the other classification models, regardless of the number of recommendations.



(a) Cat A dataset



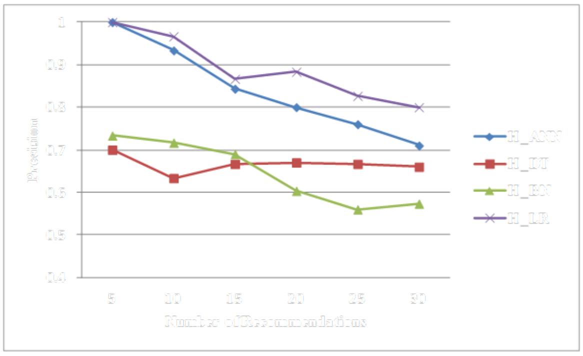
(b) Cat B dataset



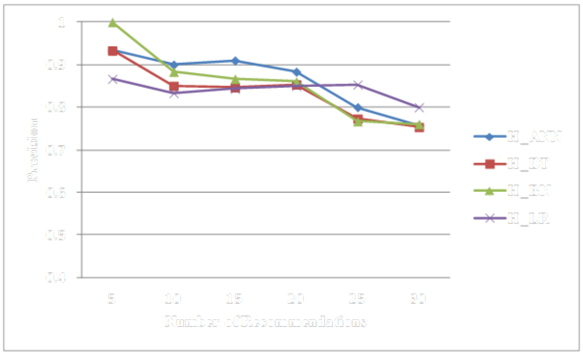
(c) Cat C dataset

Figure 4. Hit ratio of classification models

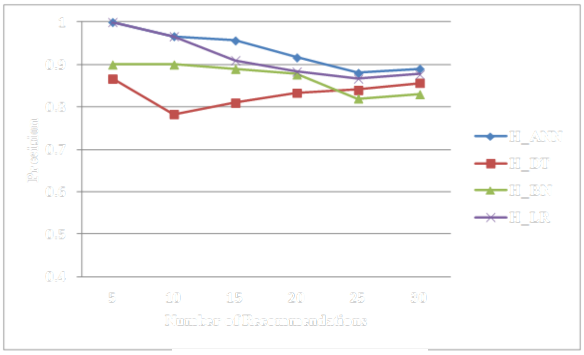
[5] From an experiment with the whole dataset including both term-based and non-term-based features, we did not obtain the better results, which may be because the characteristics of features in the two datasets are so different from each other.



(a) Cat A dataset



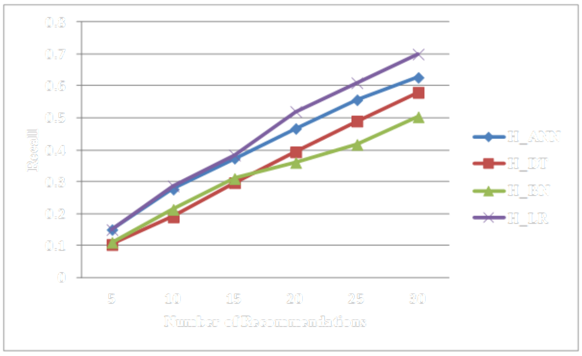
(b) Cat B dataset



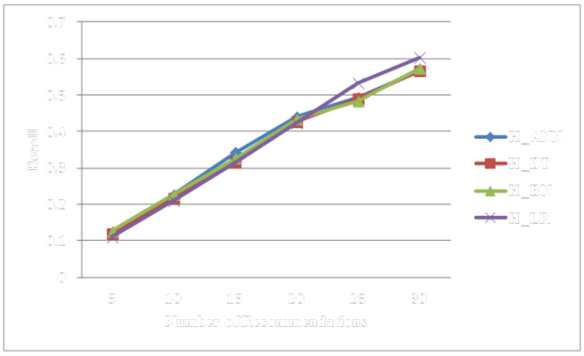
(c) Cat C dataset

Figure 5. Precision of classification models

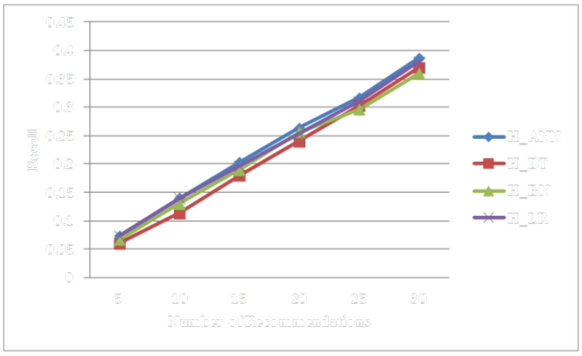
As shown in Figure 6, the best recall was achieved by H_LR when the number of recommendations is 30 in Cat A dataset. H_LR outperformed other classification models, regardless of the number of recommendations. In Cat B dataset, the best recall was achieved by H_LR when the number of recommendations is 30 similar to the case of Cat A dataset. In Cat C dataset, the best recall was achieved by H_ANN when the number of recommendations is 30. However, classification models did not show significant differences among them, regardless of the number of recommendations, in both Cat B and Cat C datasets.



(a) Cat A dataset



(b) Cat B dataset

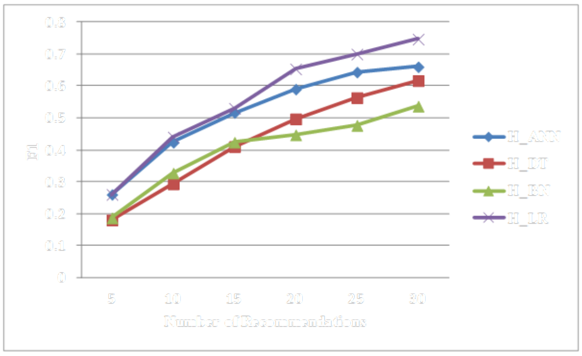


(c) Cat C dataset

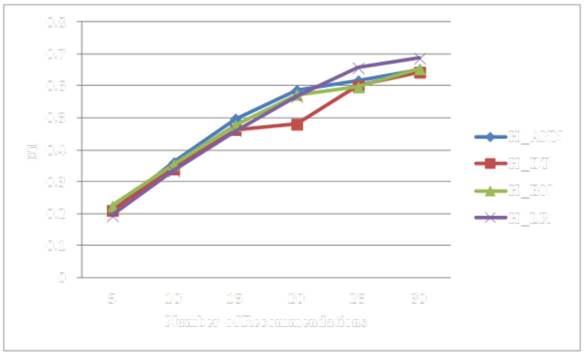
Figure 6. *Recall of classification models*

As shown in Figure 7, the best F1 was achieved by H_LR when the number of recommendations is 30 in Cat A dataset. H_LR outperformed the other classification models, regardless of the number of recommendations. In Cat B dataset, the best F1 was achieved by H_LR when the number of recommendations is 30 similarly to the case of Cat A dataset. In Cat C dataset, the best F1 was achieved by H_ANN when the number of recommendations is 30. However, similarly to the case of recall, classification models showed only slight differences among them, regardless of the number of recommendations, in both Cat B and Cat C datasets.

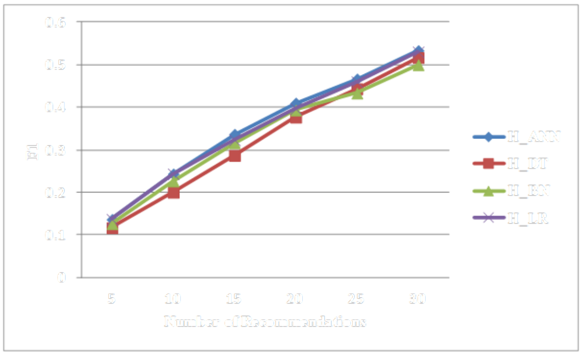
In most cases, precisions of classification models decrease as the number of recommendations increases, while recalls and F1s of them increase.



(a) Cat A dataset



(b) Cat B dataset



(c) Cat C dataset

Figure 7. *F1 of classification models*

7 CONCLUSIONS

Open innovation community is becoming a growing trend among firms across diverse industry, because it supports idea generation from customers in an effective way. Although the fact that the open innovation community can be a strategic asset for firms' innovation has proved, there is still great challenge that the firms face, information overload. The objective of this paper is to mitigate the information overload problem in an open innovation environment by recommending top n ideas with the highest adoption probability. To that end, we analyzed the dataset collected from MyStarbucksIdea (MSI), utilizing data mining and sentiment analysis, while considering both term-based features and other features of each idea.

The experimental results showed that 1) most of the hybrid classification models outperformed both the term-based and the non-term-based classification models in all category datasets, in terms of hit-ratio, 2) H_LR and H_ANN outperformed the other hybrid models in most cases, in terms of precision, recall, and F1, and 3) in most cases, precisions of classification models decrease as the number of recommendations increases, while recalls and F1s of them increase.

Our findings have several implications. First, text mining techniques for N-gram analysis and sentiment analysis are found to be useful in analyzing a large amount of text dataset. Second, both term based and non-term based features should be considered when implementing the innovative idea recommendation system. Third, using the models proposed in this study, firms can get recommendations on prospective ideas which can be valuable enough for their innovation in initial ideation stage.

However, this study has some limitations. First, the number of samples used in our experiments after preprocessing and balancing was relatively small, because there were a small number of status-confirmed ideas in our context although our initial collected dataset is large enough. Second, the data used in our study was acquired from only one open innovation community of a coffee company, Starbucks. It would bring interesting results if we conduct similar experiments using datasets from open innovation communities of other firms across diverse industry.

REFERENCES

Ahlstrom, D. (2010). "Innovation and Growth: How Business Contributes to Society." The Academy of Management Perspectives 24(3): 11-24.

Ardichvili, A., V. Page and T. Wentling (2003). "Motivation and barriers to participation in virtual knowledge-sharing communities of practice." Journal of knowledge management 7(1): 64-77.

Argamon, S., M. Koppel and G. Avneri (1998). Routing documents according to style. First international workshop on innovative information systems, Citeseer.

Beamish, P. W. and N. C. Lupton (2009). "Managing joint ventures." The Academy of Management Perspectives 23(2): 75-94.

Cassiman, B. and R. Veugelers (2006). "In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition." Management science 52(1): 68-82.

Chesbrough, H. W. (2003). "The era of open innovation." MIT Sloan Management Review 44(3): 35-41.

Chesbrough, H. W. (2003). Open innovation: The new imperative for creating and profiting from technology, Harvard Business Press.

Dahlander, L. and M. W. Wallin (2006). "A man on the inside: Unlocking communities as complementary assets." Research Policy 35(8): 1243-1259.

Esuli, A. and F. Sebastiani (2006). SentiWordNet: A publicly available lexical resource for opinion mining. Proceedings of Language Resources and Evaluation (LREC).

Füller, J., K. Matzler and M. Hoppe (2008). "Brand Community Members as a Source of Innovation." Journal of Product Innovation Management 25(6): 608-619.

Gloet, M. and M. Terziovski (2004). "Exploring the relationship between knowledge management practices and innovation performance." Journal of Manufacturing Technology Management 15(5): 402-409.

Kessler, B., G. Numberg and H. Schutze (1997). Automatic detection of text genre. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics.

Kouloumpis, E., T. Wilson and J. Moore (2011). Twitter sentiment analysis: The good the bad and the OMG!. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

Laursen, K. and A. Salter (2006). "My precious technology: the role of legal appropriability strategy in shaping innovative performance." Tanaka Business School, Imperial College London, Working Paper.

Lerner, J. S. and D. Keltner (2000). "Beyond valence: Toward a model of emotion-specific influences on judgement and choice." Cognition & Emotion 14(4): 473-493.

Li, N. and D. D. Wu (2010). "Using text mining and sentiment analysis for online forums hotspot detection and forecast." Decision Support Systems 48(2): 354-368.

Liu, B. (2012). "Sentiment analysis and opinion mining." Synthesis Lectures on Human Language Technologies 5(1): 1-167.

March, J. G. (1991). "Exploration and exploitation in organizational learning." Organization science 2(1): 71-87.

Miller, G (1995). “WordNet: A Lexical Database for English.” Communications of ACM 38(11): 39-41.

Nasukawa, T. and J. Yi (2003). Sentiment analysis: Capturing favorability using natural language processing. Proceedings of the 2nd international conference on Knowledge capture, ACM.

Pang, B., L. Lee and S. Vaithyanathan (2002). Thumbs up?: sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics.

Pak, A. and P. Paroubek (2010). Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the Seventh Conference on International Language Resources and Evaluation, European Language Resources Association (ELRA).

Spertus, E. (1997). Smokey: Automatic recognition of hostile messages. Proceedings of the National Conference on Artificial Intelligence, JOHN WILEY & SONS LTD.

Teece, D. and G. Pisano (1994). "The dynamic capabilities of firms: an introduction." Industrial and corporate change 3(3): 537-556.

Thelwall, M., K. Buckley and G. Paltoglou (2011). "Sentiment in Twitter events." Journal of the American Society for Information Science and Technology 62(2): 406-418.

Von Hippel, E. (1988). The sources of innovation. New York, Oxford University Press.

Wang, Y. and D. R. Fesenmaier (2003). "Assessing motivation of contribution in online communities: An empirical investigation of an online travel community." Electronic Markets 13(1): 33-45.

Whitelaw, C., N. Garg and S. Argamon (2005). Using appraisal groups for sentiment analysis. Proceedings of the 14th ACM international conference on Information and knowledge management, ACM.

Wyld, D. C. (2010). "Speaking up for customers: can sales professionals spark product innovation?" The Academy of Management Perspectives 24(2): 80-82.

Zhuang, L., F. Jing, X. Y. Zhu and L. Zhang (2006). Movie review mining and summarization. Conference on Information and Knowledge Management: Proceedings of the 15 th ACM international conference on Information and knowledge management.