SERVICE FAILURE COMPLAINTS IDENTIFICATION IN SOCIAL MEDIA: A TEXT CLASSIFICATION APPROACH

Research-in-Progress

Jiahua Jin
School of Management, Harbin Institute of Technology
Harbin, China
jiahuaking@gmail.com

Xiangbin Yan
School of Management, Harbin Institute of Technology
Harbin, China
xbyan@hit.edu.cn

You Yu
School of Management, Harbin Institute of Technology
Harbin, China
yyou@hit.edu.cn

Yijun Li
School of Management, Harbin Institute of Technology
Harbin, China
liyijun@hit.edu.cn

Abstract

The emergence of social media has brought up plenty of platforms where dissatisfied customers can share their service encounter experiences. Those customers' feedbacks have been widely recognized as valuable information sources for improving service quality. Due to the sparse distribution of customer complaints and diversity of topics related to non-complaints in social media, manually identifying complaints is time-consuming and inefficient. In this study, a supervised learning approach including samples enlargement and classifiers construct was proposed. Applying small labeled samples as training samples, reliable complaints samples and non-complaints samples were identified from the unlabeled dataset during the sample enlargement process. Combining the enlarged samples and the labeled samples, SVM and KNN algorithms were employed to construct the classifier. Empirical results show that the proposed approach can efficiently distinguish complaints from non-complaints in social media, especially when the number of labeled samples is very small.

Keywords: Social media, text mining, service failure complaints, service quality management

Introduction

Service failure happens when service falls below customer expectation. It will directly result in customer dissatisfaction. The severer service failure is, the lower customer satisfaction is, and the higher possibility to lose customer. Consequently, it is critical to recognize service failure and take remedial measures. A key factor in service success is the handling of customer complaints. Homburg and Fürst (2005) pointed out two ways to improve future service quality and reduce the likelihood of further complaints: recovering from a specific or immediate service failure that caused customer complaint, and identifying organizational-level work processes and individual worker practices. Thus, complaint messages are essential in service management. However, approximately 70 percent of the consumers who experience product or service problems do not complain, because they feel that complaining is unworthy of their time, or they don't think that complaining will result in a favorable outcome, or they simply do not know where and how to complain (Tarp 1986). As most dissatisfied customers keep silent, companies can't hear their voices and may mislead the design of service recovery and improvement strategies.

With the development of information technology and wide spread of Internet, a large number of customer-centered social media has emerged, such as Web forums, Web blogs, and micro blogs. Through these platforms, consumers can discuss any topic and share experience with others. Plenty of contents about customer's service encounter are generated through social media. Compared with traditional customer feedback channel, social media provides a better channel for companies to know their customers and to learn from service failure incidents(Kaplan and Haenlein 2010). On the other hand, due to the chaotic nature of the Internet and the openness of social media, online complaint messages can be easily and rapidly distributed to a large volume of potential consumers (Yang et al. 2011). Previous research suggests that passive word-of-mouth has negative impact on financial analyst's ratings of the firm (Luo 2007) and customer's switching behavior (McCollough et al. 2000). Thus, identifying service failure complaints in social media in an accurate and timely manner is vital for service quality management.

Due to the high volume of messages and topic diversity in social media, it is a time-consuming and inefficient task to identify service failure complaints manually. Our research endeavors to propose a process to scan large amounts of text-based data from social media for signals of interest in an automated mode. The proposed approach is able to construct a classifier with a small number of samples and mostly unlabeled samples. Experiment was conducted within a famous Chinese hotel's online customer community. The results indicate the proposed method is helpful for firms to identify service failure incidents, especially when only a small number of samples are labeled.

The remainder of the paper is organized as follows. First, we display the motivation for identifying service failure complaints on social media, and make a summarization of the current research and clarify objectives of this research. Next, we present a method to identify customer service failure complaints spreading on social media. Then experiment is conducted with an online customer community of a Chinese popular hotel. Finally, we discuss the research result, draw some conclusions and propose to future research.

Background and Related Work

According to GAP model (Grönroos 1984), service failure is an enterprise's service delivery behaviors falling below the customer's expectations or "zone of tolerance" (Zeithaml et al. 1993). Service failure may cause significant costs to service provider, such as negative reputation and passive word of mouth, loss of customers (Bitner et al. 2000). In addition, Service failure is a driving factor in customer switching behavior (McCollough et al. 2000). To gain a better understanding of issues affecting customer' service perception, it is vital for firms to gather service-relevant information. It is widely accepted that consumer feedback, especially service failure complaints, provides a valuable information source for improving service designs and market strategies (Finch 1999; Tax et al. 1998). Social media platform, characterized by vast volumes of user generated content with various quality and diverse topics, has recently received substantial attention. Research indicates dissatisfied and frustrated consumers prefer posting and sharing their experience in social media (Kaplan and Haenlein 2010). So user generated content on social media is a valuable source for service quality management. However, for the huge volume of messages in social media and its diverse topics, navigation of those contents is a challenge that may require filtering, semantic and sentiment analysis, topic mining, or other techniques.

Text mining techniques is widely used to identify relevant content in mass text-based messages. Researches have devoted substantial attention to the mining of emails (Park and An 2010), new articles (Ong et al. 2005), discussion forums (Abrahams et al. 2012; Li and Wu 2010), and customer reviews (Hu and Liu 2004; Ye et al. 2009) for useful knowledge. For the purposes of customer complaint management, Coussement and Van den Poel (2008) proposed a method to improve complaint handling through an automatic email-classification system that distinguishes complaints from non-complaints. Abrahams *et al.* (2012) proposed a new process and decision support system for auto-motive defect identification and prioritization, and developed a novel Vehicle Defect Discovery System (VDDS) that provides robust and generalizable defect discovery and classification.

Previous studies about text mining mainly employ machine learning technique. Thus, sample labeling is a critical step for text mining. The quantity and quality of sample labeling largely determines the mining result. However, with the rapid growth of information available in social media, to accurately label all the training documents as positive and negative examples becomes more and more difficult (Fung et al. 2006). In order to overcome this bottleneck, some researchers attempt to divide the training documents into three sets (a small set of positive examples, a small sets of negative examples, and a large set of unlabeled examples), and build a classifier using these three documents (Ghani 2001; Levi and Weiss 2004; Zhang and Oles 2000). Existing approaches that target to solve this problem generally employ two-step heuristics. In the first step, some negative examples are extracted from the unlabeled examples. In the second step, the classifier is built by using the given positive examples and the extracted negative examples.

However, messages spreading through social media are always short-text, oral-presentation, and with diverse topic. A small set of positive and negative examples may not be possible to represent the future distribution of all positive examples and negative examples. In this study, we endeavor to identify customer complaints from social media based on a small set of complaint examples and a small set of non-complaint examples.

Research Design

As illustrated in Fig. 1, supervised learning is applied in customer complaint identification in social media. The learning process includes data preparing and classifier building. Data preparing is aimed to acquire enough complaint examples and non-complaint examples for classifier building. Based on few manually labeled examples, Algorithm2 is applied to enlarge the sets of complaint examples and non-complaint examples. After enough labeled examples have been prepared, supervised learning method is adopted to construct classifiers. The output of algorithm 2 and the results of manually labeling process together constitute the training set. Data mining techniques, support vector machine and k-Nearest Neighbor, will be used to build classifiers. Then precision value, recall rate and F_1 score will be computed to validate the performance of these classifiers. The entire process of service failure identification could be described in algorithm1.

Data preparing

Data preparing process consists of 4 steps: data collecting, data cleaning, texts labeling, and enlargement of complaint sets and non-complaint sets. Data are collected from the social media, specifically from the online customer community. Those data are unstructured and related to different topics. The purpose of texts labeling is to acquire reliable complaint examples and non-complaint examples. However, it is impossible to label all complaint examples and non-complaint examples manually, as the number of texts is huge~~,~~ and the non-complaint texts are related to diverse topics. In addition, user generated contents tend to be short. When those texts are represented by vector space model, the distribution of features is very sparse in text representation. Thus, the features extracted from small manually labeled examples cannot reflect the true distribution of all complaint examples and non-complaint examples. Even if we enlarge the complaint example set and non-complaint example set by only few examples, the entire space occupied by the complaint and non-complaint examples may change significantly(Zhang and Oles 2000). In brief, enlarging complaint examples set or non-complaint negative examples set will improve the quality of classification, as the decision boundary between the complaint and non-complaint will be changed dramatically.

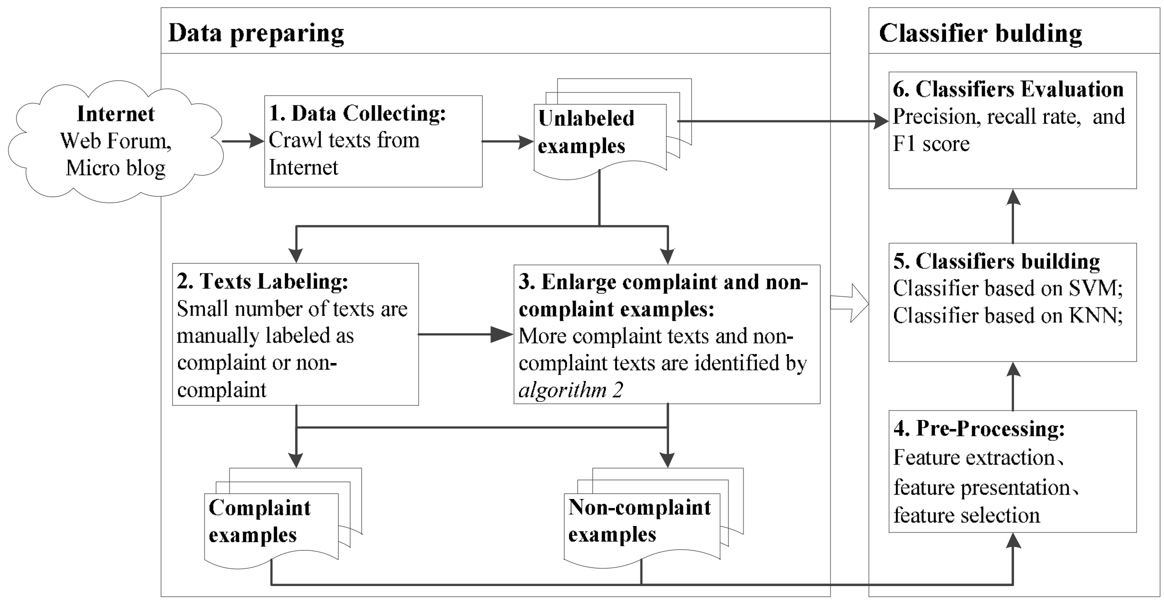

Figure 1. Research framework

Algorithm 1. Classifier for service failure identification	
Input: U (unlabeled examples) **Output:** C (classifier) **Description:** P represents complaint examples; N represents non-complaint examples.	
1. (P, N) ← manually labeled from U; 2. U' = U-(P∪N); 3. (P' , N') = **Algorithm 2**(P,N,U'); 4. P=P∪P';	5. N=N∪N'; 6. C ← build classifier by (P, U); 7. **Return** C;

In order to enlarge complaint examples and non-complaint examples, each document *d* is represented as a vector, $\vec{d} = (q_1, q_2, ..., q_n)$. Each element q_i in $\vec{d}$ represents a word w_i and is calculated as the combination of term frequency (*tf*) and inverse document frequency (*idf*): $q_i = tf_i * idf_i$. tf_i is the number of times that word w_i occurs in *d*, while $idf_i = \log(|D|/df(w_i))$, here |D| is the total number of labeled examples and $df(w_i)$ is the number of documents where word w_i occurs at least once.

According to profile-based approach (Rocchio 1971) and equations (1-2), we firstly construct a profile for each class, P and N, which denotes centroids of complaints examples and non-complaints examples, here $\| \vec{d} \|$ is the Euclidean distance of vector $\vec{d}$ and the origin. Then we use cosine coefficient to measure the similarity between a document, d, and a profile, C_i for $i \in \{P, N\}$, as is shown in equation(3).

$$C_P = \frac{1}{|P|}\sum_{d \in P}\frac{\vec{d}}{\|\vec{d}\|} \quad (1); \qquad C_N = \frac{1}{|N|}\sum_{d \in N}\frac{\vec{d}}{\|\vec{d}\|} \quad (2); \qquad S(d, C_i) = \frac{d \bullet C_i}{\|d\| \bullet \|C_i\|} \text{ for } i \in \{P, N\} \quad (3)$$

It should be emphasized that we are not trying to identify all complaint examples and non-complaint examples in U by simply using the similarities. The enlargement process aims at identifying some complaint examples and non-complaint examples that are reliable. The effect of unreliable samples in the learning phase of a supervised classifier is likely to bring distorted information and introduce a bias in the definition of the decision regions, thus, decreasing the accuracy of the final classification map. Therefore, we cannot simply compare the similarities between d and C_P, also between d and C_N by using formula (3). The purpose is to get documents that are significantly similar to either P or N. Documents that are similar to both P and N are regarded as ambiguous and will be ignored.

Thus, a document d in U belongs to P if it meets the following requirements:

$$S(d, C_P) > \mu_P \quad (4); \qquad M_P(d) = S(d, C_P) - S(d, C_N) > \Upsilon_P \quad (5)$$

where μ_P is the average similarity of documents within P and Υ_P is the average difference between documents in P and N, as is shown in formula (6) and (7) .

$$\mu_P = \frac{1}{|P|}\sum_{d \in P} S(d, C_P) \quad (6); \qquad \Upsilon_P = \frac{1}{|P|}\sum_{d \in P}(S(d, C_P) - S(d, C_N)) \quad (7)$$

Without loss of generality, a document d in U belongs to N if and only if the following condition holds:

$$S(d, C_N) > \mu_N \quad (8); \qquad M_N(d) = S(d, C_N) - S(d, C_P) > \Upsilon_N \quad (9)$$

where

$$\mu_N = \frac{1}{|N|}\sum_{d \in N} S(d, C_N) \quad (10); \qquad \Upsilon_N = \frac{1}{|N|}\sum_{d \in N}(S(d, C_N) - S(d, C_P)) \quad (11)$$

Based on the above analysis, the process of enlarging positive examples and negative examples can be described in algorithm 2.

Algorithm 2. Enlarge the set of positive and negative examples

Input: *P* (complaint examples), N (non-complaint examples), *U* (unlabeled examples)

Output: *P'* (enlarging complaint examples), *N'* (enlarging non-complaint examples)

1. $P' = \varnothing$, $N' = \varnothing$;
2. Turn all documents ($P \cup N \cup U$) into vectors
3. (C_P, C_N) ← construct a centroid for each class by using equation (1), (2)
4. **for all** $d \in U$ **do**
5.   **if** d satisfies the conditions in (2) and (3) **and** does not satisfy the conditions in (6) and (7) **then**
6. $P' = P \cup d$;
7.   **continue**;
8.   **end if**
9.   **if** d satisfies the conditions in (6) and (7) **and** does not satisfy the conditions in (2) and (3) **then**
10. $N' = N \cup d$;
11.   **continue**;
12.   **end if**
13. **end for**
14. **return** *P'*, *N'*;

Classifier building

Vector space model, which was proposed by Salton (1968), is widely used in text classification (Aasheim and Koehler 2006). The vector space approach converts text-based information to numerical vectors based on the weighted term frequencies. Each document is represented by a vector, and each element of the vector reflects the importance of the corresponding term. Following these steps, all documents are converted from the original format to word vectors. We have to stress that our data corpus was collected from a Chinese social media. Therefore the following steps are based on Chinese context.

Pre-processing

The purpose of pre-processing is to convert unstructured texts into numerical word vectors. The pre-processing includes raw text cleaning, tokenization, part-of-speech tagging and term filtering.

As the dataset was crawled from online communities, text cleansing is required to eliminate HTML codes and other unrelated contents from raw text, such as navigation of website, advertisements, etc. Cleaned texts are separated into tokens and words during the tokenization step. We remove characters other than Chinese words and use the ICTCLAS (Zhang et al. 2011) to apply Chinese punctuation. Furthermore, all words are tagged as part of speech by the ICTCLAS based on their syntactic category. Word whose part-of-speech is nouns verbs, adjectives or adverbs is regarded as informative part-of-speech; otherwise it is regarded as non-informative part-of-speech. After being punctuated, each document can be represented by a bag of words. Document d can be converted to $d = \{(w_1, f_1), (w_2, f_2)...(w_i, f_i)\}$, where each w_i represents a word contained in the document, and f_i denotes its frequency appearing in the document. Since the number of different words appearing in the collection may be very large and contain many topic irrelevant words. Word filtering is implemented to deduct dimensions and remove topic irrelevant words. Therefore, rare words, whose TF-IDF values are less than given threshold, are left out from further analysis because they are helpless in future classification. And words that are non-informative part of speech can also be eliminated for they contain less information needed for classification. The result of pre-processing is a term-by-document matrix, where each row of the matrix represents a document, and each cell in the matrix represents the frequency of appearance of term in a document

Term-vector weighting

After pre-processing, the values in the term-by-document collection are simply the raw frequencies of appearance for a term in a document and cannot indicate the importance of the term in the entire dataset. Term-vector weighting is employed to compute the importance of the term in the document and the whole dataset. Term-vector weighting can be computed in different ways, using methods such as information gain (Billsus and Pazzani 1998) and mutual information (Peng et al. 2005). As pointed by Combarro et al. (2005), the TFIDF method is simple but performs well in many situations. In this paper, we employ TFIDF to compute the weight of each term of the matrix.

Classification technique

Document classification is to assign a document to one or more class. There are many automatic document classification algorithms, such as Rochio classification (Lewis et al. 1996), decision tree classification (Quinlan 1986), support vector machine (SVM) (Cortes and Vapnik 1995), k-Nearest Neighbor rule (KNN) (Cover and Hart 1967)and Naïve-Bayesian classifier (John and Langley 1995). Prior studies found that SVM and KNN have excellent performance in text classification in most cases (Dumais et al. 1998; Li and Wu 2010). These two classification algorithms have their own advantages. In this paper, we will compare SVM and KNN classification algorithms to find a more suitable method for our research problem.

Evaluation criteria

Precision, recall and F_1 measures are widely used to assess the result of information retrieval, text mining and other machine learning fields. In text classification area, the precision for a class is the number of true positives divided by the total number of texts labeled as positive class. Recall in this context is defined as the number of true positives divided by the total number of texts that actually belong to the positive class. F_1 score considers both precision and recall in assessing the result of classification. In this paper, we compute the values of precision, recall and F_1 to assess the performance of the proposed approach.

Experiments and Results Analysis

Data corpus

The data corpus for the empirical studies are crawled down and compiled from the Internet by an automatic crawling Java program written by us, which consists of two major modules: the target URL list generating module and the HTML page parsing module. We choose to conduct the experiments on HanTing's online customer community. HanTing is committed to providing travelers with high-quality, conveniently-located, and reasonably priced hotel. Up to December 31, 2011, more than 4.4 million prestige guests have registered in HanTing club. Registered users can post complaints, advices, reviews and other user generated content through the customer community. The data corpus, which includes texts posted by customer from 2009-2012, consists of 1688 complaint examples and 3827 non-complaint examples. The non-complaint examples are diverse, including advertisements, advices, experiences and other user generated content. Texts with less than 50 Chinese characters are eliminated from the data corpus.

The positive and negative examples enlargement

Based on the proposed enlargement approach, we conducted experiments by using data with different percentage of labeled examples. The results is shown in table 1. The *Rate* column shows the percentage of labeled examples out of the total number of examples. The *Number of Example* column includes 3 sub-columns, namely *Positive*, *Negative* and *Unlabeled*, which respectively represent the number of complaint examples, non-complaints examples and unlabeled examples before examples enlargement(data in the bracket denotes the number after enlargement). The *Enlargement Evaluation* column also includes three sub-columns, *P*, *R*, and F_1**,** representing the precision value, recall value and F_1 score respectively, criterions for evaluating the enlargement algorithms.

Table 1. Complaint and non-complaint examples enlargement

Rate	Number of examples			Enlargement evaluation		
	Positive	Negative	Unlabeled	P (%)	R (%)	F_1 (%)
1%	18[a](85[b])	38[a](564[b])	5459a(4810[b])	95.3[c] (98.8 [d])	8.3[c] (33.3[d])	15.2[c] (49.8[d])
2%	34(73)	76(759)	5405(4573)	97.3(97.5)	9.7(19.8)	17.6(32.9)
3%	51(115)	114(566)	5350(4669)	93.9(97.7)	11.8(14.9)	21.0(25.9)
4%	68(332)	152(821)	5295(4142)	90.7(96.6)	18.6(21.3)	30.8(34.9)
5%	85(245)	191(924)	5239(4070)	89.8(96.6)	13.7(24.6)	23.8(39.2)
7%	119(331)	267(931)	5129(3867)	86.5(95.7)	18.2(25.0)	30.1(39.6)
11%	186(452)	422(1189)	4907(3266)	84.1(94.0)	25.4(32.8)	39.0(48.7)
15%	254(643)	575(997)	4686(3046)	79.0(93.3)	35.4(28.6)	48.9(43.8)

a *- the number of labeled examples* ***b*** *- the number of enlarged examples*
c *- precision, recall and F_1 of the result of complaint examples enlargement.*
d *- precision, recall and F_1 of the result of non-complaint examples enlargement.*

It is obvious that the enlargement algorithm is able to identify complaint examples and non-complaints example from the unlabeled corpus efficiently. On the whole, with the increase of labeled examples, class feature becomes more obvious and comprehensive, and its discriminating ability improves, consequently showing a rising trend in recall and F_1. Though the overall trend of non-complaint examples is similar to that of complaint examples, the former is more diverse in topic, because the representativeness of manually labeled samples impose a great influence on the final classification. This explain the instability of variation trend of recall value and F_1.

Different classification algorithms comparison

In this section, we evaluated the effectiveness of proposed approach by using SVM and KNN algorithms. *SVMCLS 2.0* is open-source software *and* includes SVM, KNN and other text classification algorithms. We employed *SVMCLS 2.0* to fulfill SVM and KNN algorithms and turned related parameters to their optimal values given by the software. Given a small set of labeled complaint examples and non-complaint examples, and a large set of unlabeled examples, we compare two cases: enlargement versus without enlargement. As is shown in table 2, the *Rate* column shows the percentage of complaint examples. The following six columns show accuracy, recall and F_1 of two classifiers when enlargement algorithm is not applied. The last six columns show the corresponding value of SVM and KNN when we apply enlargement algorithm. For example, when the percentage of labeled examples is 1%, the precision, recall and F_1 for SVM is 87.9%，32.8% and 47.8% respectively if the enlargement algorithm is not considered. When we apply enlargement algorithm, the accuracy drops to 76.1%, the recall rises to 69.9% and F_1 rises by 25.1%, showing obvious enhancement in classification performance after the enlargement. As more examples are labeled, more errors can possibly be introduced, resulting in a worse classification effect. The accuracy, recall and F_1 decrease as the percentage of labeled examples increases. Figure 2 shows the F_1 value of SVM and KNN classifiers based on training set of manually labeled examples (SVM-W, KNN-W) and that based on training set combining manually labeled examples and enlarged examples (SVM-E, KNN-E). We observed that F_1 value in non-enlargement condition exceeds what it is in enlargement condition when the percentage of labeled examples reaches 5%.

The results of the experiment can be summarized as follows: (1) SVM algorithm is better than KNN algorithm in complaint text classification. (2) The proposed enlargement approach can improve the classifier performance significantly, especially when the number of labeled examples is very small. (3) As the number of labeled examples increases, the enlargement approach would induce noise examples in building classifier, and lead to decrease in the classifier performance.

Table 2. Classifiers building and comparing

Rate	Without enlargement						Enlargement					
	SVM			KNN			SVM			KNN		
	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)
1%	87.9	32.8	47.8	0	0	0	76.1	69.9	72.9	68.5	62.4	65.3
2%	87.4	35.5	50.5	67.3	27.5	39.0	80.7	68.5	74.1	78.0	39.6	52.5
3%	86.0	51.3	64.3	70.7	34.6	46.5	80.0	68.0	74.8	66.1	66.8	66.4
4%	77.2	63.6	69.7	77.1	57.3	65.7	78.3	65.6	71.4	76.5	60.0	67.3
5%	80.5	62.6	70.4	75.8	58.4	66.0	77.4	64.3	70.2	60.3	71.7	65.5
7%	80.3	69.1	74.3	84.6	57.7	68.6	74.0	57.5	64.7	57.3	77.4	65.9
11%	79.0	71.8	75.2	82.2	59.3	68.9	70.1	65.1	67.5	49.3	86.8	62.9
15%	80.3	76.5	78.4	83.9	61.1	70.7	67.8	63	65.3	47.1	85.4	60.7

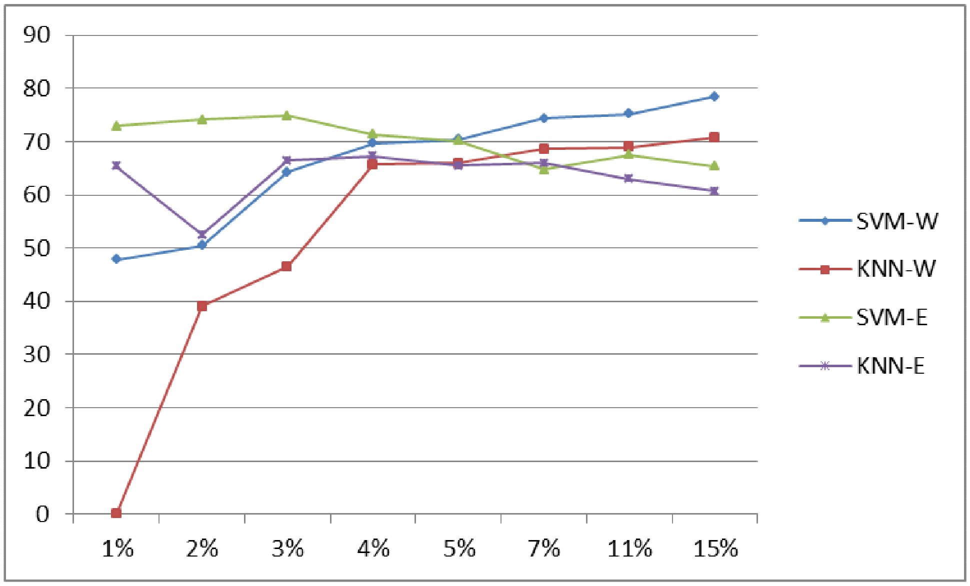

Figure 2. F1 values in different combination of algorithms

Discussions and Conclusion

Service quality management is vital for companies who want to success in cut-throat competition environment. It is widely accepted that consumer feedbacks, especially complaint messages, are valuable sources of service-related business intelligence. As the development of information technology and social media, more and more customer prefers to post their service experiences in online customer community. Thus, social media (such as forum, blog, etc.) provides a platform for company to listen to and learn from customer. However, message in social media are always huge and of different topics, making it impossible to identify customer complaint manually.

To overcome this bottleneck, we proposed a two-step approach based on text mining techniques. Given a small set of labeled complaint examples *P*, a small set of non-labeled examples *N* and a large set of unlabeled examples *U*, we proposed an enlargement algorithm to get reliable P' and N' from unlabeled examples. Then classifiers were built based on the combination of labeled examples and enlarged examples. We evaluated the performance of this method on dataset collected from an online customer community. Experimental result indicated the enlargement algorithm increase the performance of text classification significantly, especially when the number of labeled examples is small. This study has significant implication on the field practitioners. The service provider can use the proposed framework to automatically identify the service failure complaint from numerical unrelated messages in online customer communities. Especially during samples enlargement process, the precision of enlargement algorithm is quite high when the number of labeled examples is small. However, as the enlargement process may introduce misclassified samples when the number of labeled examples is large, the performance of classification based on the combination of manually labeled examples and enlarged examples is not significantly improved. In future work, we will study the dimensions of service failure and linguistic features of complaint text, and use them to compute the centroids of the labeled examples in enlargement and features selection for classification.

Acknowledgments

This research is supported by the Natural Science Foundation of China (Grant #71171067 and #71328103)

References

Aasheim, C., and Koehler, G.J. 2006. "Scanning World Wide Web Documents with the Vector Space Model," *Decision support systems* (42:2), pp. 690-699.

Abrahams, A.S., Jiao, J., Wang, G.A., and Fan, W. 2012. "Vehicle Defect Discovery from Social Media," *Decision Support Systems* (54:1), pp. 87-97.

Billsus, D., and Pazzani, M.J. 1998. "Learning Collaborative Information Filters," *Proceedings of the fifteenth international conference on machine learning*, p. 46-54.

Bitner, M.J., Brown, S.W., and Meuter, M.L. 2000. "Technology Infusion in Service Encounters," *Journal of the Academy of marketing Science* (28:1), pp. 138-149.

Combarro, E.F., Montanes, E., Diaz, I., Ranilla, J., and Mones, R. 2005. "Introducing a Family of Linear Measures for Feature Selection in Text Categorization," *Knowledge and Data Engineering, IEEE Transactions on* (17:9), pp. 1223-1232.

Cortes, C., and Vapnik, V. 1995. "Support-Vector Networks," *Machine learning* (20:3), pp. 273-297.

Coussement, K., and Van den Poel, D. 2008. "Improving Customer Complaint Management by Automatic Email Classification Using Linguistic Style Features as Predictors," *Decision Support Systems* (44:4), pp. 870-882.

Cover, T., and Hart, P. 1967. "Nearest Neighbor Pattern Classification," *Information Theory, IEEE Transactions on* (13:1), pp. 21-27.

Dumais, S., Platt, J., Heckerman, D., and Sahami, M. 1998. "Inductive Learning Algorithms and Representations for Text Categorization," *Proceedings of the seventh international conference on Information and knowledge management*: ACM, pp. 148-155.

Finch, B.J. 1999. "Internet Discussions as a Source for Consumer Product Customer Involvement and Quality Information: An Exploratory Study," *Journal of Operations Management* (17:5), pp. 535-556.

Fung, G.P.C., Yu, J.X., Lu, H., and Yu, P.S. 2006. "Text Classification without Negative Examples Revisit," *Knowledge and Data Engineering, IEEE Transactions on* (18:1), pp. 6-20.

Ghani, R. 2001. "Combining Labeled and Unlabeled Data for Text Classification with a Large Number of Categories," *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*: IEEE, pp. 597-598.

Grönroos, C. 1984. "A Service Quality Model and Its Marketing Implications," *European Journal of marketing* (18:4), pp. 36-44.

Homburg, C., and Fürst, A. 2005. "How Organizational Complaint Handling Drives Customer Loyalty: An Analysis of the Mechanistic and the Organic Approach," *Journal of Marketing*), pp. 95-114.

Hu, M., and Liu, B. 2004. "Mining and Summarizing Customer Reviews," *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*: ACM, pp. 168-177.

John, G.H., and Langley, P. 1995. "Estimating Continuous Distributions in Bayesian Classifiers," *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*: Morgan Kaufmann Publishers Inc., pp. 338-345.

Kaplan, A.M., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business horizons* (53:1), pp. 59-68.

Levi, K., and Weiss, Y. 2004. "Learning Object Detection from a Small Number of Examples: The Importance of Good Features," *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*: IEEE, pp. II-53-II-60 Vol. 2.

Lewis, D.D., Schapire, R.E., Callan, J.P., and Papka, R. 1996. "Training Algorithms for Linear Text Classifiers," *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*: ACM, pp. 298-306.

Li, N., and Wu, D.D. 2010. "Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast," *Decision Support Systems* (48:2), pp. 354-368.

Luo, X.M. 2007. "Consumer Negative Voice and Firm-Idiosyncratic Stock Returns," *Journal of Marketing* (71:3), Jul, pp. 75-88.

McCollough, M.A., Berry, L.L., and Yadav, M.S. 2000. "An Empirical Investigation of Customer Satisfaction after Service Failure and Recovery," *Journal of service research* (3:2), pp. 121-137.

Ong, T.-H., Chen, H., Sung, W.-k., and Zhu, B. 2005. "Newsmap: A Knowledge Map for Online News," *Decision Support Systems* (39:4), pp. 583-597.

Park, S., and An, D.U. 2010. "Automatic E-Mail Classification Using Dynamic Category Hierarchy and Semantic Features," *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*

(27:6), pp. 478-492.

Peng, H., Long, F., and Ding, C. 2005. "Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (27:8), pp. 1226-1238.

Quinlan, J.R. 1986. "Induction of Decision Trees," *Machine learning* (1:1), pp. 81-106.

Rocchio, J.J. 1971. "Relevance Feedback in Information Retrieval," in *The Smart Retrieval System: Experiments in Automatic Document Processing,* G. Salton (ed.). Prentice-Hall, Englewood Cliffs NJ, pp. 313-323.

Salton, G. 1968. "Automatic Information Organization and Retrieval."

Tarp, T.A.R.P. 1986. "Consumer Complaint Handling in America: An Update Study," White House Office of Consumer Affairs, Washington DC.

Tax, S.S., Brown, S.W., and Chandrashekaran, M. 1998. "Customer Evaluations of Service Complaint Experiences: Implications for Relationship Marketing," *The Journal of Marketing*), pp. 60-76.

Yang, M., Li, Y., and Kiang, M. 2011. "Environmental Scanning for Customer Complaint Identification in Social Media," *ICIS 2011 Proceedings*, Shanghai.

Ye, Q., Zhang, Z., and Law, R. 2009. "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches," *Expert Systems with Applications* (36:3), pp. 6527-6535.

Zeithaml, V.A., Berry, L.L., and Parasuraman, A. 1993. "The Nature and Determinants of Customer Expectations of Service," *Journal of the academy of Marketing Science* (21:1), pp. 1-12.

Zhang, H., Wang, M., Lv, Z., and Zhang, X. 2011. "Chinese Lexical Analysis System." ICTCLAS.

Zhang, T., and Oles, F. 2000. "The Value of Unlabeled Data for Classification Problems," *Proceedings of the Seventeenth International Conference on Machine Learning,(Langley, P., ed.)*: Citeseer, pp. 1191-1198.