

國立臺灣大學管理學院資訊管理學研究所

碩士論文

Department of Information Management

College of Management

National Taiwan University

Master Thesis

運用短文件分類技術改良微網誌政府服務之研究

An Efficient Classification Framework for Micro Blog-based

Government Services



Chih Yu Lei

指導教授：陳建錦 博士


Advisor : Chien Chin Chen, Ph.D.

中華民國 101 年 7 月

July, 2012

運用短文件分類技術改良微網誌政府服務之研究

An Efficient Classification Framework for Micro Blog-based  
Government Services



本論文係提交國立台灣大學  
資訊管理學研究所作為完成碩士  
學位所需條件之一部份

研究生：雷智宇 撰  
中華民國一百零一年七月

國立臺灣大學碩士學位論文  
口試委員會審定書

運用短文件分類技術改良微網誌政府服務之研究

An Efficient Classification Framework for Micro Blog-based  
Government Services



本論文係雷智宇君（學號 R99725001）在國立臺灣大學資訊管理學系、所完成之碩士學位論文，於民國 101 年 7 月 9 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

陳建鄴

戴碧如

陳通彰

李正帆

所長：

李瑞庭

## 致謝

時光飛逝，兩年的時間就這樣過去了。依稀還記得當初選這實驗室的憧憬與期望，到現在依舊是如此。首先，非常感謝指導教授建錦老師這幾年來的教導與提拔。和其他同學相較下較幸運的我，在大二時就先認識了建錦老師，也提早體會到了老師那溫文儒雅且充滿熱忱的教學態度。不論是在大二課程、大三大四專題、以及碩士班的指導，老師總是不吝於分享資源給學生，不厭其煩的指導我們重點所在，更能在我撰寫論文時畫龍點睛的點出問題所在及可能的解決之道。讓我在這幾年的學生生涯中可以很順利的完成課程與研究。跳脫研究內容，讓我更加值得學習的是老師的研究態度。老師總是可以針對一項問題盡可能的去做各種層面的思考，並且最後找出適合的解決辦法。這樣的研究態度對我來說不僅是在課業上非常有幫助，以後出社會工作深造希望我可以保持這種心態來完事。可以在學生時期和建錦老師相遇真是一種福氣與緣分，真的非常謝謝老師這五年來的教導！

我想要感謝WEAL Lab所有的夥伴們。感謝兩年來的好戰友璧華的陪伴。謝謝她聽愛抱怨的我發牢騷並且可靠的她總是會很有效率的安排許多事情並且完成。也謝謝經常提早準備的她總是提醒了我許多課業以及實驗室的大小事。這兩年有妳真好！其次，想要感謝仲詠及詠淳學長。學長們在我實驗室的這兩年總是對研究提出了許多不一樣的觀點，讓我可以以更加宏觀的角度去思考實驗的內容與細節。最後，當然要感謝碩一的羽君與孟潔，總是告訴我許許多多有趣的事情讓有時為實驗所苦惱的我感到開心無比，有你們在的WEAL Lab永遠都是歡樂開心度過每一天。

再來，我想要感謝我親愛的家人。縱使長大以後比較少和父母說話，但每當晚回家時總是可以發現父母替我多保留了一份餐點或消夜，這種關心不曾間斷過。感謝爸媽的精神以及經濟上的支持，讓我可以無憂無慮的完成我的碩士學業。更感謝爸媽對我的不耐心的包容與寬恕。感謝姐姐、妹妹與Ruby總是跟我分享生活大小事。

要感謝的人太多了，那就謝天吧！這兩年我所得到的東西，豈是短短一面致謝可以道盡的。還記得大學時畢業典禮上蔡老師給我們的兩個字：累積。累積自己的交友、累積自己的人脈、累積自己的實力，讓自己可以逐件的發光發熱。台大校訓也告訴我：敦品、勵學、愛國、愛人。這六年來我緊遵守著這四句話在累積自我。今日我以台大為榮、希望有朝一日台大以我為榮。

雷智宇 謹識  
于台大資管研究所  
民國一百零一年七月

# 論文摘要

論文題目：運用短文件分類技術改良微網誌政府服務之研究

作者：雷智宇

101 年 7 月

指導教授：陳建錦 博士

Web2.0 的興起提供了政府以及市民們一條直接溝通的管道。近年來此項技術中，微網誌服務如推特(Twitter)已經成為了當今最熱門的應用之一。許多政府服務也在一些議題中利用微網誌從社會大眾蒐集各式各樣的見解。一般來說，市民們對於可以利用這類媒體來訴求他們的意見感到滿意進而提升人民對政府的信任程度。然而，隨著近年來微網誌的數量呈現指數型成長，因此勢必需要運用文字探勘相關技術來有效率的分析這些大量短文章。在這篇研究中，我們對於微網誌政府服務提出了一個有效率的分類架構。為了去解決微網誌的資料稀疏性問題，運用外部知識庫以及微網誌本身提供的時間資訊來修改原始貝式分類模型中的事前以及條件機率。利用 311NYC 資料集的實驗顯示出我們所提出的分類架構可以準確的分類市民對於政府服務的見解，並且對貝式分類模型有顯著的改善。

關鍵字：電子化政府、文字探勘、分類

# THESIS ABSTRACT

## **An Efficient Classification Framework for Micro Blog-based Government Services**

By Chih Yu Lei

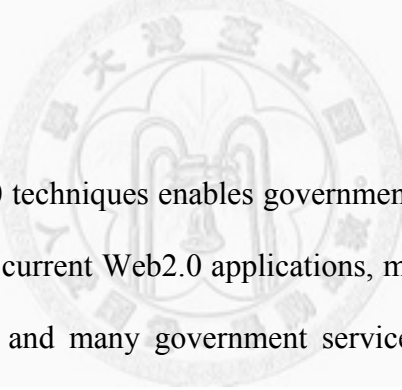
MASTER DEGREE OF BUSINESS ADMINISTRATION

DEPARTMENT OF INFORMATION MANAGEMENT

NATIONAL TAIWAN UNIVERSITY

JULY 2011

ADVISER: Chien Chin Chen, Ph.D.



The prevalence of Web2.0 techniques enables governments and citizens to communicate in a direct manner. Among the current Web2.0 applications, micro-blogging services, such as Twitter, are the most popular and many government services now exploit micro-blogs to collect opinions from the public on a range of issues. Generally, citizens are satisfied with this medium for expressing their opinions; however, as the number of micro-blogs is increasing exponentially, text mining is needed to analyze the opinions efficiently. In this paper, we propose an efficient classification framework for micro blog-based government services. To address the text sparseness problem of micro-blogs, an external knowledge base and the temporal information of micro blogs are used to modify the prior and conditional probabilities of the Naive Bayes classification model. Experiments based on the 311NYC dataset show that the proposed framework classifies citizens' opinions about government services correctly, and it achieves a significant improvement over the Naive Bayes model.

Keywords : E-Government, Government2.0, Text Mining, Classification

# Table of Contents

## 內容

致謝 .....	1
論文摘要 .....	2
THESIS ABSTRACT .....	3
Table of Contents .....	4
List of Figures .....	5
List of Tables .....	6
1. Introduction .....	7
2. Related Works .....	11
2.1 Government Services and Web2.0 Techniques .....	11
2.2 Short-text Classification .....	12
2.3 Using WordNet to Improve Classification .....	14
3. Methodology .....	15
3.1 The Naive Bayes Classification Model .....	16
3.2 Using an External Knowledge Base .....	17
3.3 Temporal Information about Tweets .....	19
3.4 Smoothing Techniques for Temporal Information .....	22
3.5 User Content Enrichment .....	24
3.6 Using the WordNet Synonym Dictionary .....	25
4. Performance Evaluation .....	27
4.1 Dataset Collection and Evaluation Metrics .....	27
4.2 Performance Evaluations .....	29
5. Conclusion and Future Works .....	35
References .....	36

## List of Figures

Figure 3-1. The Framework Architecture .....	16
Figure 3-2. Environment Distribution .....	20
Figure 3-3. Transportation Distribution.....	21
Figure 3-4 Environment Distribution after Exponential Smoothing.....	23
Figure 3-5 Transportation Distribution after Exponential Smoothing.....	23
Figure 3-6. User Content Retrieving Process .....	25



## List of Tables

Table 1-1 The most popular social network communities growth rate in 08-09.....	8
Table 4-1. The statistics of the evaluation dataset .....	28
Table 4-2. The results of manual annotation .....	28
Table 4-3. The statistics of the knowledge base .....	29
Table 4-4. The effect of $\alpha$ on the system performance .....	30
Table 4-5. The effect of $\beta$ on the system performance under $\alpha = 0.6$ .....	31
Table 4-6. The comparison between $\beta = 0$ and $\beta = 1$ .....	32
Table 4-7. The performance of add-one and exponential smoothing .....	32
Table 4-8. The performance of user content enrichment.....	33
Table 4-9. The performance of WordNet Synonym Dictionary .....	34



# 1. Introduction

Because of the prevalence of Web2.0 techniques, the World Wide Web(WWW) has become a huge information source. In contrast to the traditional Web, where users passively read content created by website owners, Web2.0 enables individuals to serve as information providers and express opinions about different issues. The Web has thus become a knowledge base for various applications (Craven et al. 2000), and governments have found that they can exploit it to facilitate communications with the public about government services (Wigand 2010). In the past, there was a lack of efficient communication channels, but various Web2.0 applications, such as micro-blogs and social media, can now provide instant and efficient communications between governments and citizens. The techniques have also motivated the development of the *Government2.0* paradigm, which refers to the use of Web2.0 technologies to publicize and commoditize government services (Nam 2011). Specifically, Government2.0 aims to establish direct and instant communication channels between governments and citizens through online comments, live chats, and message threads. Depending on the characteristics of specific services, governments can choose the technique that is most convenient people to use (Nam 2011).

Among current Web2.0 techniques, micro-blogging is the most popular application. Micro-blogging is a new weblog formalism that enables users to exchange information through short messages (Java et al. 2007). In Nielsen's 2009 survey on the size of the Internet community<sup>1</sup>, Twitter<sup>2</sup> achieved the largest community growth rate, and it is now the most

---

<sup>1</sup> [http://blog.nielsen.com/nielsenwire/online\\_mobile/twitters-tweet-smell-of-success/](http://blog.nielsen.com/nielsenwire/online_mobile/twitters-tweet-smell-of-success/)

<sup>2</sup> <http://twitter.com>

famous micro-blogging platform. As shown in Table 1-1, Twitter’s growth rate between 2008 and 2009 was 1382%, which was far greater than that of other Internet communities.

Table 1-1 The most popular social network communities growth rate in 08-09.

<b>Rank</b>	<b>Site</b>	<b>Feb 08</b>	<b>Feb 09</b>	<b>% growth</b>
1	Twitter.com	475,000	7,038,000	1382%
2	Zimbio	809,000	2,752,000	240%
3	Facebook	20,043,000	65,704,000	228%
4	Multiply	821,000	2,394,000	192%
5	Wikia	1,381,000	3,758,000	172%

The main reason for Twitter’s popularity is its availability (Sankaranarayanan et al. 2009). Since mobile computation is ubiquitous, users can send tweets (i.e., short messages) anytime anywhere through various mobile devices (e.g., cell phones). In addition, the restricted length of tweets (i.e., no more than 140 characters) forces users to express opinions in a clear and concise manner. Twitter’s friendly interface also motivates users to form virtual communities. Because the communities are quite different from people’s normal social circles in the real world, users generally acquire different information from Twitter. This subculture and various properties explain the huge success of Twitter (Sankaranarayanan et al. 2009). Java et al. (2007) investigated the topological and geographical properties of Twitter, and found that users rely heavily on tweets to communicate their daily activities and share the latest information.

Many governments are aware of the merits of tweeting and now utilize it in various government services (Wigand 2010). The increasing usage of Twitter appears to be associated with the fact that it is free and it provides instant communications (Wigand 2010). For instance, New York City extended its 311 system with Twitter<sup>3</sup> so that citizens can submit questions about the city’s facilities or request help from a facility. City officials provide

<sup>3</sup> <https://twitter.com/#!/311NYC>

instant feedback and address users' requests in a timely manner. NASA also utilized Twitter to publicize the MarsPhoenix landing project<sup>4</sup>. The service attracted nearly 160,000 followers who helped NASA share information about the progress of the mission.

Although there are many advantages in using Twitter to communicate about government services, there is a risk of information overload. As the membership of Twitter is growing exponentially, the number of daily tweets about government services may be too large for officials to handle manually. Thus, there is an urgent need for text mining techniques to manage the large volume of tweets. Text classification is a classic text mining method for managing a vast number of documents (Manning et al. 2008). Given a set of categories and the corresponding training documents, a classification algorithm learns a classifier, which then assigns new documents to the appropriate categories automatically. Many effective classification algorithms have been proposed for classifying documents, but they may not be suitable for tweet classification because tweets are short (Phan et al. 2008; Chen et al. 2011). Hence, the text sparseness of tweets would reduce the classification accuracy.

In this paper, we modify the Naive Bayes classification model (Manning et al. 2008) to classify the daily tweets of government services. To resolve the text sparseness problem, we extend training tweets with Wikipedia<sup>5</sup>, the biggest online encyclopedia. In addition, we model the temporal patterns of tweets to refine the prior probability of our naive Bayes classifier. Besides, we also apply two kinds of different smoothing techniques to leverage the temporal distribution. In addition to using Wikipedia as the external knowledge base, we take user contents (e.g., the tweets posted by users) which are not directly reported to 311NYC into consideration. Finally, we apply wordNet<sup>6</sup>, which is a large lexical database of English, to reformulate the conditional probabilities of Naïve Bayes model. Evaluations based on the NYC 311 Twitter service show that the proposed approach classifies daily tweets accurately.

---

<sup>4</sup> <https://twitter.com/#!/MarsPhoenix>

<sup>5</sup> <http://www.wikipedia.org/>

<sup>6</sup> <http://wordnet.princeton.edu/>

It also alleviates the text sparseness problem.

The remainder of the paper is organized as follows. In Section 2, we discuss the concept of Government2.0 and review related works on short text classification methods. The proposed classification approach is presented in Section 3, and its performance is evaluated in Section 4. Then, in Section 5, we present our conclusions.



## 2. Related Works

In this section, we consider the theories that explain why governments are eager to exploit Web2.0 techniques in government services. We also review a number of short text classification methods and many studies of using WordNet to improve text classification performance.

### 2.1 Government Services and Web2.0 Techniques

Two theories support the adoption of Web2.0 techniques in government services, namely, the theory of *social influence* and the theory of *collective intelligence*. Social influence exists when an individual's thoughts, feelings, and actions influence his acquaintances and vice versa. Kelman (1958) defined three kinds of social influence: compliance, internalization, and identification. Compliance means that people appear to agree with others, but they have dissenting opinions that they keep to themselves. By contrast, internalization means that people agree with others totally. Identification means that individuals are influenced by people they like or respect. Basically, social influence brings people together to form communities. As Web2.0 techniques encourage the development of social communities, adopting them in government services helps citizens form strong focus groups to monitor the government's performance.

Collective intelligence (also known as the wisdom of crowds) is a phenomenon that enables a collection of people to construct valuable knowledge through collaboration and competition (Schuler 2001). Since Web2.0 techniques provide a convenient platform for information exchange, adapting them for use in government services helps governments

gather information about issues that concern citizens. The effectiveness of government services then can be improved by remedying the concerns (Wigand 2010). Based on the theories, exploiting Web2.0 techniques to improve government services is a win-win situation for both citizens and governments. Web2.0 techniques help people form communities to monitor governments, while governments enhance their services through the collective intelligence produced by Web2.0 communities. According to the Web2.0 guideline<sup>7</sup> defined by the Web2.0 technology advisory group of Federal Web Manager Council, Twitter is an effective Web2.0 tool for constructing direct communication channels between governments and citizens. Since Twitter disseminates information instantly, it makes official agencies transparent to the public. Twitter is now used by many government services and it has become an essential tool for disseminating information about government facilities (Wigand 2010).

## **2.2 Short-text Classification**

As mentioned earlier, short-text classification is difficult because of the text sparseness problem (Phan et al. 2008). Many classification methods employ external knowledge bases to classify short text correctly. Basically, the methods can be classified as search engine-based and corpus-based approaches. Search engine-based methods regard a short text item as a query to search for relevant documents in a database. The content of the returned documents then enriches the short text to output correct classification results. Sahami and Heilman (2006) computed the similarity between short texts by using a web-based kernel. Given two short-text segments, the kernel models the segments as expanded term vectors and computes their similarity based on the vectors' inner-product. Yih and Meek (2007) extended the kernel model by submitting text segments to search engines. The words in the returned documents are used to adjust the similarity calculation. Shen et al. (2006) applied short text classification techniques to the query enrichment problem. Because the queries submitted to search engines are usually short and ambiguous, search engines have difficulty finding documents that are

---

<sup>7</sup> [http://www.usa.gov/webcontent/documents/Web\\_Technology\\_Matrix.pdf](http://www.usa.gov/webcontent/documents/Web_Technology_Matrix.pdf)

relevant to a user's interests. Classifying the topic of a query helps identify the user's information need. Shen et al.'s method examines the returned snippets, titles, and URLs of a query and extracts representative terms to construct a query-to-target-category classifier that can correctly identify the intention of the query. The method won first prize in the 2005 ACM Knowledge-discovery and Data Mining Competition (ACM KDDCUP 2005).

Corpus-based methods utilize an expert-compiled document corpus to classify short text. Phan et al. (2008) considered that search engine-based methods are not suitable for on-line applications because querying search engines is time-consuming. Instead, they compiled a large collection of documents, called the universal dataset, and extracted hidden topics from it to construct an accurate short-text classifier. In general, compiling a document corpus is laborious, so a manually compiled corpus may not be comprehensive enough for short-text classification. To obtain a larger corpus, Zelikovitz and Hirsh (2000) integrated an expert-compiled corpus with a large unlabeled document set in order to improve the similarity measure of short text. The authors showed that the unlabeled document set significantly reduces the number of classification errors, particularly when the classification training set is small. Chen et al. (2011) also derived latent topics from a document set, and proposed using Latent Dirichlet allocation (LDA) to generate a topic space from the document set. The topic space enriches the content of short text and achieves superior short-text classification performance. Hu et al. (2009) used Wikipedia to resolve the text sparseness problem. They developed two approaches, called exact match and relatedness-match, to map a document to Wikipedia concepts and categories. The degree of overlap between the mapped concepts and categories is used to compute the similarity between text-sparse documents.

Our approach differs from existing short-text classification methods because, in addition to employing a knowledge base, we consider the temporal information of tweets to enhance short-text classification.

### **2.3 Using WordNet to Improve Text Classification**

Many researches have exploited WordNet as a knowledge base to refine their classification models. Rodríguez et al. (1997) integrates WordNet information with two training approaches through the Vector Space Model(Salon & McGill 1983). The training approaches they employed are the Rocchio (i.e., relevance feedback)(Rocchio Jr, J.J. 1971) method and the machine learning-based Widrow-Hoff algorithm(Widrow & Sterns1985). The results obtained from their evaluations show that integrating WordNet clearly improves the training approaches, and the integrated technique effectively address the classification of low frequency categories. Scott and Matwin (1998) described experiments in machine learning for text classification using a novel representation of text based on WordNet hypernyms. Their experiments show that for difficult classification tasks the hypernym density representation can form significant reductions in error rates. Liu et al. (2007) use the hood algorithm(Voorhees, E. 1993) to remove word ambiguities so that each word is substituted by its sense in the context. The nearest ancestors of the senses of all the non-stopwords in a given document are selected as the classes for the given document. The effectiveness of the algorithm is evaluated by comparing its classification result with that using the manual disambiguation offered by Princeton University.

### **3. Methodology**

In this section, we introduce the proposed tweet classification framework, which can be applied to various Twitter-based government services. For ease of explanation and without loss of generality, we apply the framework to the Twitter-based 311NYC system. Figure 3-1 shows the framework architecture. The core of the framework is the Naive Bayes (NB) classification model. Basically, the framework consists of a training part and a test part. For the training part, human experts prepare a set of training tweets and label their classes. The tweets are used to construct an NB classifier that classifies the daily tweets of the 311NYC system in the test part. To resolve the text sparseness problem, we use a knowledge base to enrich the training tweets. Then, the temporal information about tweets is used to reformulate the classification model. Finally, we apply two smoothing techniques to leverage temporal distribution, consider user content (i.e., the tweets posted by users) to enrich the training tweets, and employ the WordNet lexical database to create a synonym dictionary to improve the classification performance. We discuss each system component in the following sub-sections.

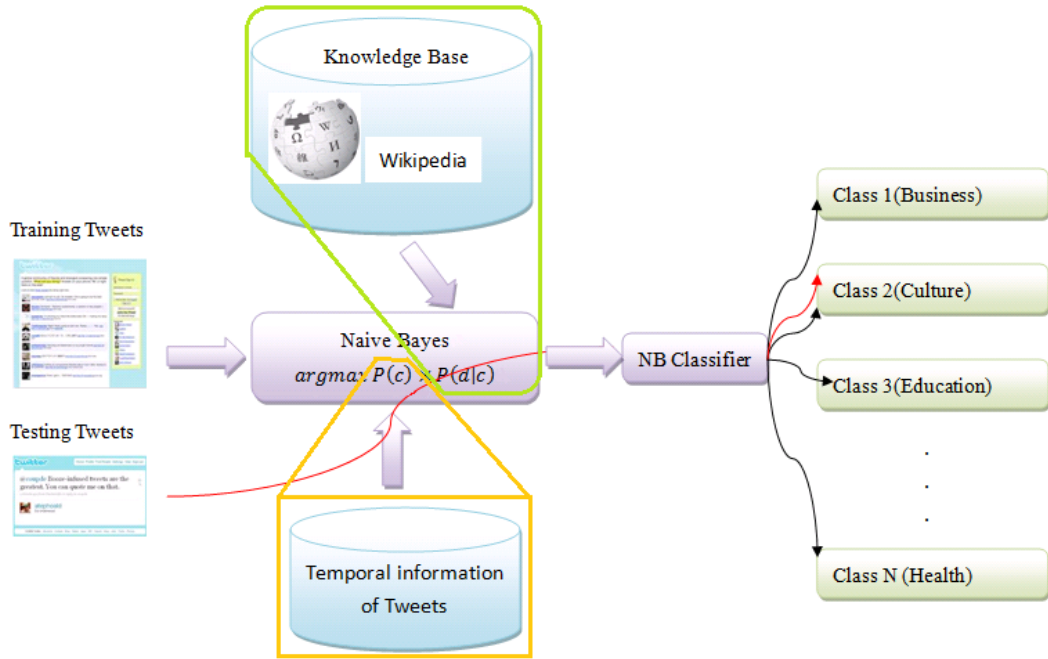


Figure 3-1. The Framework Architecture

### 3.1 The Naive Bayes Classification Model

To classify daily tweets related to the 311NYC system, we employ the Naive Bayes model, which is a probabilistic machine learning model. It is widely regarded as an effective document classification method (Manning et al. 2008). The 311NYC system categorizes its services into several classes (e.g., transportation and environment). Let  $C = \{c_1, c_2, \dots, c_K\}$  be the set of service classes provided by the 311NYC system; and let  $\underline{d} = \langle w_1, w_2, \dots, w_M \rangle$  be the content vector of a tweet  $d$ , where  $w_i$  is the  $i$ th word in  $d$ . Given the content vector, the most likely service class of  $d$  would be:

$$\hat{c} = P(c_k | \underline{d}) \quad (1)$$

By the Bayes theorem (Ghaharhani 2004), the above conditional probability can be expanded as follows:

$$\begin{aligned}
\hat{c} &= \operatorname{argmax}_{1 \leq k \leq K} \frac{P(c_k) P(\underline{d} | c_k)}{P(\underline{d})} \\
&= \operatorname{argmax}_{1 \leq k \leq K} P(c_k) P(\underline{d} | c_k) \\
&= \operatorname{argmax}_{1 \leq k \leq K} P(c_k) P(\langle w_1, w_2, \dots, w_M \rangle | c_k)
\end{aligned} \tag{2}$$

In information retrieval and text mining, it is common to assume that words are independent of each other under the given the class (Manning et al. 2008). Based on the bag-of-words assumption (Manning et al. 2008), Equation 2 can be re-formulated as follows:

$$\hat{c} = \operatorname{argmax}_{1 \leq k \leq K} P(c_k) \prod_{i=1}^M P(w_i | c_k), \tag{3}$$

where  $P(c_k)$  is the prior probability of class  $c_k$ , and  $P(w_i | c_k)$  is the probability of observing word  $w_i$  in a tweet that belongs to class  $c_k$ . Given the training tweets and the corresponding classes labeled by human experts, the probabilities  $P(c_k)$  and  $P(w_i | c_k)$  can be acquired by the following formulae:

$$P(c_k) = N_c / N \tag{4}$$

$$P(w_i | c_k) = W_{i,k} / W_k, \tag{5}$$

where  $N$  is the number of training tweets,  $N_c$  is the number of training tweets in class  $c_k$ ,  $W_k$  is the number of words in the training tweets of class  $c_k$ , and  $W_{i,k}$  is the frequency of  $w_i$  in the training tweets of class  $c_k$ .

### 3.2 Using an External Knowledge Base

The NB model may not be appropriate for short-text classification (Androutsopoulos et al. 2000). As tweets are usually short, their content is limited. Consequently, the training tweets would not represent the service classes comprehensively. In such cases, the text sparseness problem would bias the learned  $P(w_i | c_k)$  and could lead to zero probabilities when

computing  $P(c_k|\underline{d})$ . For instance, if the word  $w_j$  never occurs in the training dataset, the learned  $P(w_j|c_k)$  would be zero for all classes. Then, any test tweet that contains the word  $w_j$  would yield a zero probability for  $P(c_k|\underline{d})$  such that the most likely class would be unpredictable.

Many studies (Phan et al. 2008; Chen et al. 2011; Yih & Meek 2007; Shen et al. 2006 ; Zelikovitz & Hirsh 2000) have investigated the text sparseness problem and proposed using external knowledge bases to enhance short-text classification. A knowledge base must possess two important characteristics: (1) it must be large enough to cover all the classes (concepts) in a question; and (2) it must provide detailed knowledge. To ensure that the knowledge is comprehensive, we adopt Wikipedia, one of the largest on-line encyclopedia, as our knowledge base. According to statistics reported by Alexa<sup>8</sup>, a well-known network traffic statistics company, Internet users rely heavily on Wikipedia. Currently, Wikipedia's network traffic is ranked number six in the world. Furthermore, Wikipedia contains diversified knowledge maintained by the general public. Because every user serves as a knowledge provider and a maintainer, the knowledge is usually up-to-date and correct (Gabrilovich & Markovitch 2007). The knowledge in Wikipedia is presented as wiki pages, each of which introduces a concept mentioned in its title. To acquire the knowledge related to the services in question, we devised a wiki page retrieval method. For each service category, the 311NYC system presents a service introduction. We segment the introduction into words, and remove stop words, such as *is*, *a*, and *of*, from the segmented text to construct a service dictionary  $D_c = \{k_{c,1}, k_{c,2}, \dots, k_{c,V}\}$  where  $k_{c,i}$  is the  $i$ th keyword of class  $c$ . Then, we treat each keyword as a query to retrieve relevant wiki pages via the tool Jwiki<sup>9</sup>. Since the keywords introduce service class  $c$  precisely, the retrieved wiki pages are highly relevant to class  $c$ . Interestingly, not every keyword matches a wiki page. To obtain comprehensive knowledge of the service classes, we utilize Wikipedia's redirection mechanism. If a keyword cannot match a wiki page, the redirection mechanism returns the pages named by the synonyms of the keyword. The set

---

<sup>8</sup> <http://cn.alexa.com/siteinfo/wikipedia.org/>

<sup>9</sup> <http://jwikiapi.sourceforge.net/>

of retrieved wiki pages  $B$  is used to modify  $P(w_i|c_k)$  of our classification model as follows:

$$P(w_i|c_k) = (1 - \alpha) W_{i,k} / W_k + \alpha B_{i,k} / B_k, \quad (6)$$

where  $B_k$  denotes number of words in the wiki pages of class  $c_k$ , and  $B_{i,k}$  is the frequency of  $w_i$  in the wiki pages of class  $c_k$ . The parameter  $\alpha$ , whose range is within  $[0,1]$ , weights the influence of the knowledge base when we calculate the probability. A large  $\alpha$  value means that the knowledge base is important for the classification. In the experiment section, we examine the effect of  $\alpha$  on tweet classification.

### 3.3 Temporal Information about Tweets

Studies of temporal data mining have demonstrated the benefit of using temporal information for data classification. For instance, Hao and Hao (2008) proposed a temporal-driven Naive Bayes classification model and showed that the temporal information in data enhances data classification significantly. Revesz and Triplet (2011) also recognized the importance of temporal information. They developed a temporal-based NB classifier that utilizes the time of data for kernel regression parameter estimation. We observe that existing short-text classification methods ignore the temporal information about text. When preparing the data for our experiment, we found that requests made to government services exhibited a temporal regularity. Figures 3-2 and 3-3 show the frequency of 311NYC tweets for the Environment and Transportation classes in our experimental dataset. As shown in the figures, the tweets about Transportation were posted before and after working hours because traffic was often jammed during those periods. By contrast, most tweets about the Environment were posted around noon and midnight. This may be because people discussed public facilities during their lunch breaks, and complained about noisy neighbors when they were trying to sleep. To exploit the temporal regularity, we incorporate the temporal information about tweets into the Naive Bayes model. Specifically, we consider the time that tweets are posted

in our proposed model. Let  $t_d$  denote the time that tweet  $d$  is posted, Then, the most likely class of  $d$  published at time  $t_d$  can be computed as follows:

$$\begin{aligned}
 \hat{c} &= \operatorname{argmax}_{1 \leq k \leq K} P(c_k | \underline{d}, t_d) \\
 &= \operatorname{argmax}_{1 \leq k \leq K} \frac{P(c_k, \underline{d}, t_d)}{P(\underline{d}, t_d)} \\
 &= \operatorname{argmax}_{1 \leq k \leq K} \frac{P(c_k, t_d)P(\underline{d} | c_k, t_d)}{P(\underline{d} | t_d)P(t_d)} \\
 &= \operatorname{argmax}_{1 \leq k \leq K} \frac{P(c_k | t_d)P(\underline{d} | c_k, t_d)}{P(\underline{d} | t_d)} \\
 &= \operatorname{argmax}_{1 \leq k \leq K} P(c_k | t_d)P(\underline{d} | c_k, t_d) \\
 &= \operatorname{argmax}_{1 \leq k \leq K} P(c_k | t_d)P(\langle w_1, w_2, \dots, w_M \rangle | c_k, t_d) \\
 &= \operatorname{argmax}_{1 \leq k \leq K} P(c_k | t_d) \prod_{i=1}^M P(w_i | c_k, t_d)
 \end{aligned} \tag{7}$$

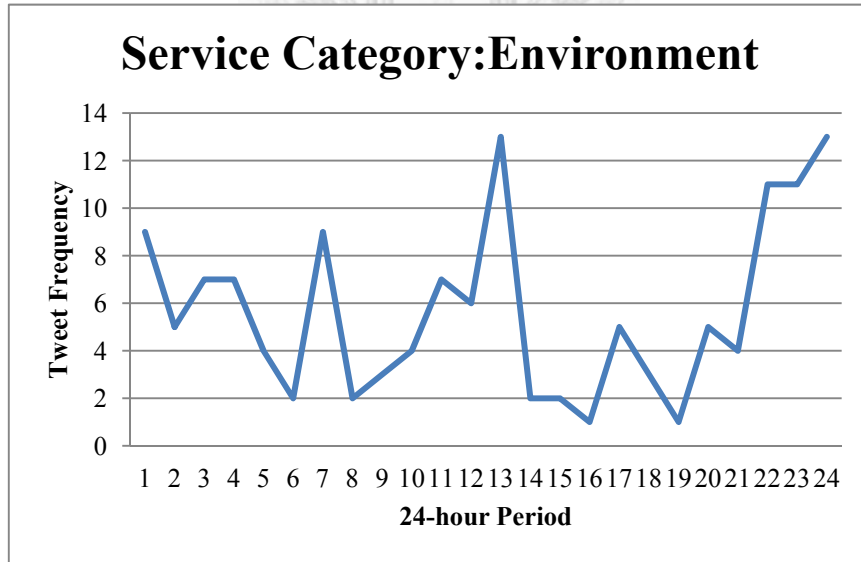


Figure 3-2. Environment Distribution

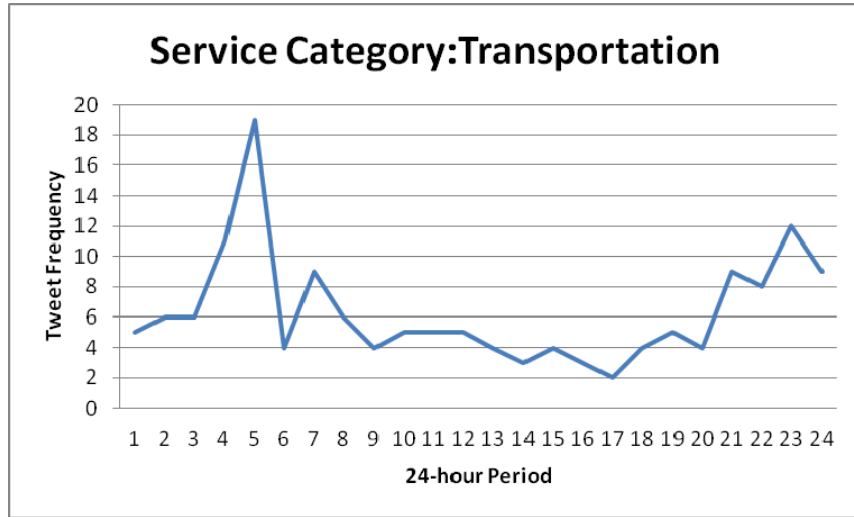


Figure 3-3. Transportation Distribution

Generally, people use a limited vocabulary to request a government service. For instance, they use words like *accident*, *jam*, and *block* to complain or make traffic assistance requests when there is a traffic jam. The phenomenon supports the assumption that the occurrence of a word in a tweet of class  $c_k$  is independent of the time of publication. In other words,  $P(w_i|c_k, t_d) = P(w_i|c_k)$ . Thus, Equation (7) can be converted as follows:

$$\hat{c} = \operatorname{argmax}_{1 \leq k \leq K} P(c_k|t_d) \prod_{i=1}^M P(w_i|c_k), \quad (8)$$

where  $P(w_i|c_k)$  is computed by Equation (6) and  $P(c_k|t_d)$  is the probability of class  $c_k$  given that a tweet is posted at time  $t_d$ . To compute the most likely class, we need to estimate  $P(c_k|t_d)$ . The publication time  $t_d$  can be modeled by different scales. Here, we consider the hour that a tweet is posted. Let  $t_d$  be an integer in the range  $[1, 24]$ . Then,  $P(c_k|t_d)$  is computed as follows:

$$P(c_k|t_d) = (1 - \beta) N_c / N + \beta T_{t_d, k} / T_{t_d}, \quad (9)$$

where the first term  $N_c/N$  is the same as in Equation (4),  $T_{t_d, k}$  is the number of training tweets

in class  $c_k$  published at time  $t_d$ , and  $T_{td}$  is the total number of training tweets published at time  $t_d$ . Parameter  $\beta$ , whose range is within  $[0,1]$ , weights the influence of the temporal information when calculating the probability. A large  $\beta$  value means that the temporal information is important for the classification. In the experiment section, we examine the effect of  $\beta$  on tweet classification.

### 3.4 Smoothing Techniques for Temporal Information

We consider two smoothing techniques to leverage the temporal distribution: Add-one smoothing(Manning et al. 2008) and exponential smoothing(Keller, G. 2005). In order to avoid zero probabilities when computing Equation (9), we add one to each time interval. Then, Equation (9) can be computed as follows:

$$P(c_k|t_d) = (1 - \beta) N_C/N + \beta (T_{td,k} + 1)/(T_{td} + 24), \quad (10)$$

where 24 is the number of time intervals. In short, add-one smoothing integrates a uniform prior with the probabilities estimated from the training data to update  $P(c_k|d_t)$  (Manning et al. 2008). Exponential smoothing is a frequently used technique used applied to time series data to either produce smoothed data for presentation or to make forecasts<sup>10</sup>. Assume that raw data sequence is presented as  $\{T_{td}, 1 \leq td \leq 24\}$  and the output sequence is  $\{E_{td}, 1 \leq td \leq 24\}$ , the exponential smoothing form is given by the following formula:

$$\begin{aligned} E_1 &= T_1 \\ E_{td} &= \gamma T_{td-1} + (1 - \gamma) E_{td-1}, 1 < td \leq 24, \end{aligned} \quad (11)$$

where  $\gamma$  is a smoothing factor, and  $0 < \gamma < 1$ . Figure 3-4 and Figure 3.5 show the result of

<sup>10</sup> [http://en.wikipedia.org/wiki/Exponential\\_smoothing](http://en.wikipedia.org/wiki/Exponential_smoothing)

exponential smoothing. Compared with Figure 3-2 and Figure 3-3, the curves are much smooth after applying the techniques. In the experiment section, we will show the classification results of applying both smoothing techniques.

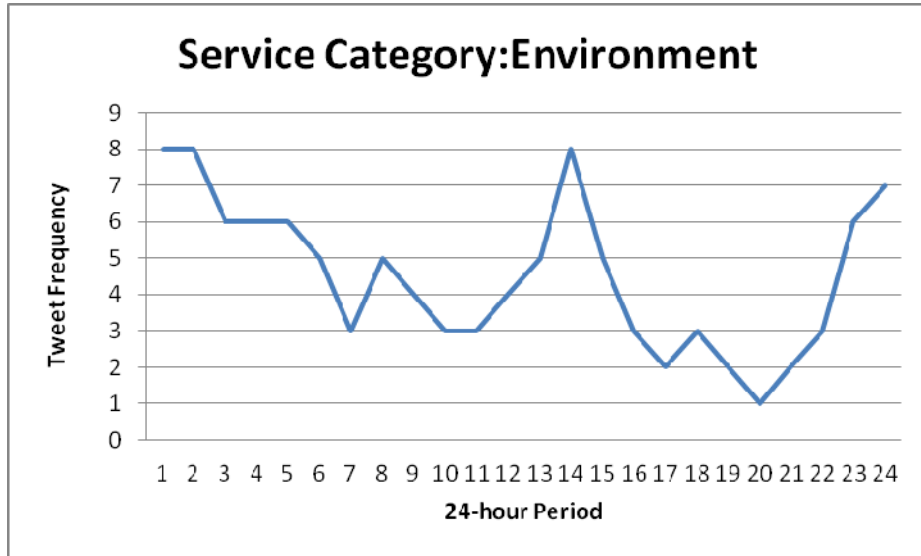


Figure 3-4 Environment Distribution after Exponential Smoothing

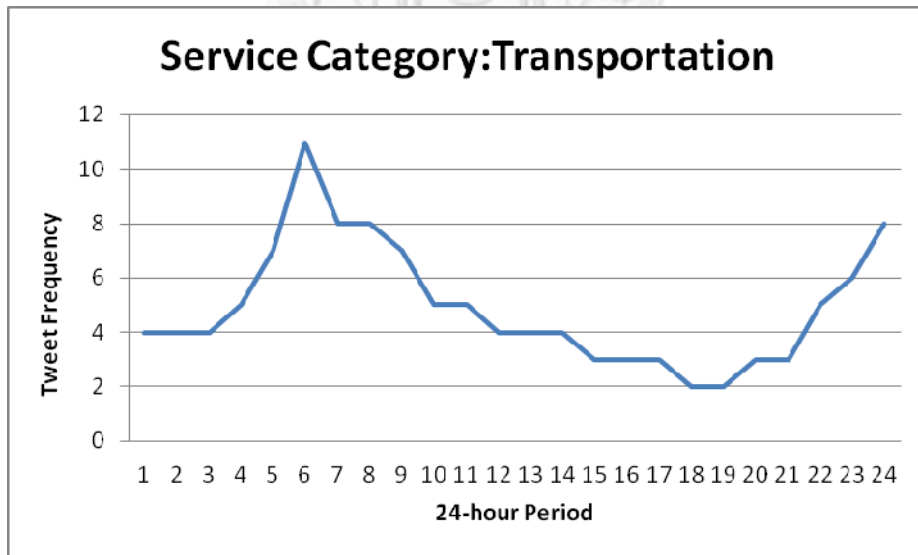


Figure 3-5 Transportation Distribution after Exponential Smoothing

### 3.5 User Content Enrichment

A Tweet, with its nature limitation of less than 140 characters, restricts the content users want to present. When complaining government services, users may not only post a tweet to the 311NYC service, but argue it with their friends simultaneously. Therefore, in addition to using Wikipedia as the external knowledge base, we take user contents which are not directly reported to 311NYC into consideration. Figure 3-6 shows how we utilize user content and an example of retrieving user content. Specifically, we extract the author list from training dataset and using Twitter API<sup>11</sup> to get all available user contents on their public timelines. Given a training tweet posted by user account  $s$  on time  $t$  and the time interval  $\Delta ti(hour)$ , we retrieve all tweets posted by  $s$  during the interval  $[t + \Delta ti, t - \Delta ti]$  to enrich the original tweet. For instance, given a training tweet posted on "2011-12-23 19:19:48" and 1 hour time interval, we retrieve one tweet entitled "Can u park in front of schools today?" from user *ms8kane* published on "2011-12-23 19:24:26", which falls in the interval  $[19:19:48 + 1, 19:19:48 - 1]$ . In the experiment section, we will show the classification result of user content enrichment. The probability will be reformulated as follows:

$$P(w_i|c_k) = (W_{i,k} + U_{i,k}) / (W_k + U_k) , \quad (12)$$

where  $W_{i,k}$  and  $W_k$  is the same as Equation (6),  $U_k$  is the number of words in the user content tweets of class  $c_k$ , and  $U_{i,k}$  is the frequency of  $w_i$  in the user content tweets of class  $c_k$ .

---

<sup>11</sup> <https://dev.twitter.com/>

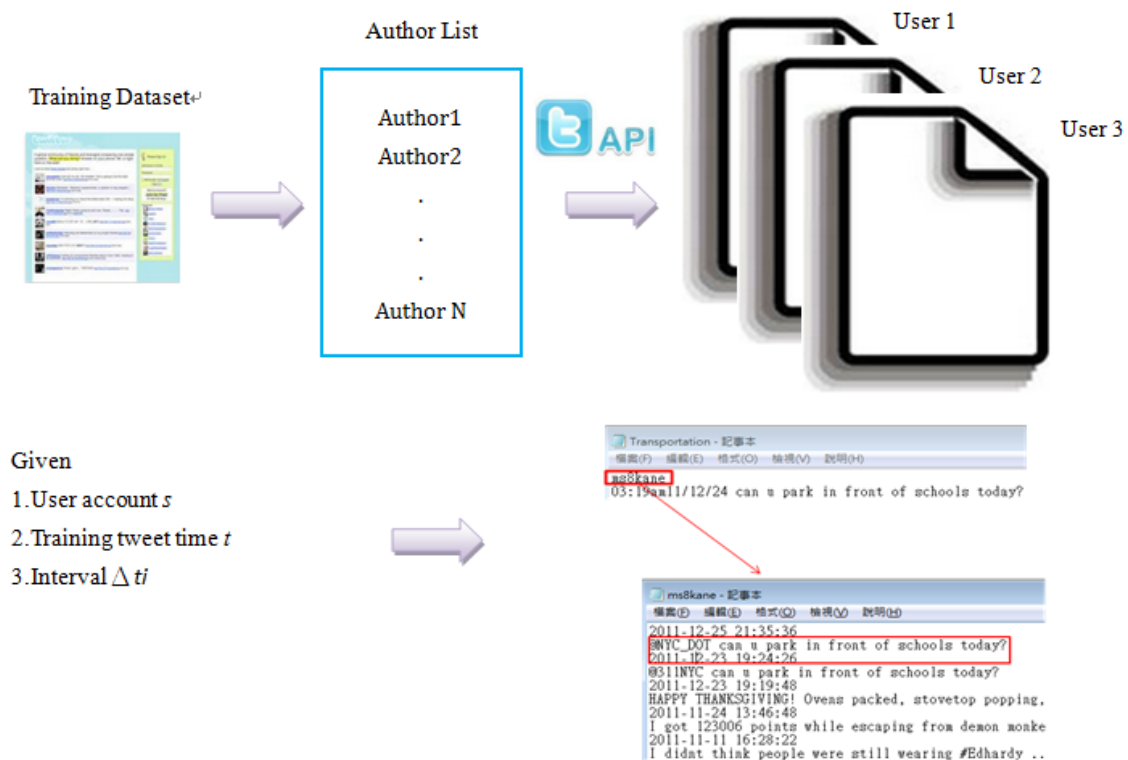


Figure 3-6. User Content Retrieving Process

### 3.6 Using the WordNet Synonym Dictionary

As we mentioned in related work, WordNet has been widely exploited as a knowledge base into classification models. In addition to using original tweets as our experiment basis, we employ the WordNet lexical database to analyze the contents of the tweets to get content synonyms and integrate these synonyms into our classification model. We first use java API for WordNet Search<sup>12</sup> to transform all text into their synonyms and compute the frequency of each synonym set under each class. There are two difficulties in using the WordNet synonyms: First, some words have more than one synonym set. For example, the word "apple" be be a fruit or a technology brand, and the synonym set that apple belongs to depends on the the word's context. To simply the difficulty, we only consider the first synonym set of each word according to different hash code in different synonym sets because the first synonym set represents the main meaning of the word. The second difficulty is that not all words have a

<sup>12</sup> <http://lyle.smu.edu/~tspell/jaws/index.html>

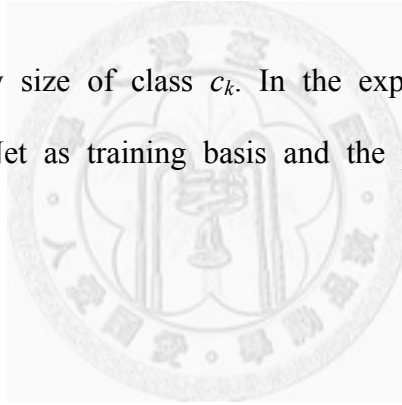
synonym. To tackle this problem, we manually create a dummy synonym set with size 1 for each unmatching word. Then, equation (6) can be reformulated as follows:

$$P(w_i|c_k) = (1 - \alpha) S_{i,k}/S_k + \alpha B_{i,k}/B_k, \quad (13)$$

where  $S_{i,k}$  is the frequency of  $w_i$  in synonym set of class  $c_k$  and  $S_k$  is the frequency of synonym of class  $c_k$ . Finally in order to avoid the zero probability problems, we employ add-one smoothing (Manning et al. 2008) in this equation as follows:

$$P(w_i|c_k) = (1 - \alpha) (S_{i,k} + 1)/(S_k + |V_k|) + \alpha B_{i,k}/B_k, \quad (14)$$

where  $|V_k|$  is the vocabulary size of class  $c_k$ . In the experiment section, we show the performance of using WordNet as training basis and the performance of using add-one smoothing technique.



## 4. Performance Evaluation

### 4.1 Dataset Collection and Evaluation Metrics

We collected tweets for evaluation from the 311NYC system. Table 4-1 details the statistics of the evaluation dataset. The 311NYC system supports 10 classes of services. A service request is made by sending a tweet to Twitter account “@311NYC.” After reading the request, system agents send a response tweet. As it is hard to resolve a problem using 140 characters (i.e., the length restriction of tweets), a hyperlink to the solution of the request is included in the response. We annotated the class of a request tweet based on the content indicated by the hyperlink. The dataset consists of 381 request tweets sent by 328 users and 381 response tweets sent by system agents between September 30, 2009 and January 17, 2012. As shown in the table, the request tweets are very short (97.95 characters on average), so classifying them is very difficult. The number of tweets in each service class is shown in Table 4-2. Note that some tweets are annotated as “Off-topic” because their requests are beyond the scope of the 311NYC system. For instance, users may send tweets to @311NYC to ask for rescue services, which are handled by government safety units. Table 4-3 shows the statistics of the knowledge base, where 361 wiki pages are retrieved by using 361 keywords extracted from the 311NYC system. Finally, in the user content enrichment process, we totally retrieved 237 user tweets.

Table 4-1. The statistics of the evaluation dataset

Dates	2009/9/30-2012/1/17
# of request tweets	381
# of response tweets	381
# of users	328
Avg. characters of request	97.95
Avg. characters of response	79.84

Table 4-2. The results of manual annotation

<b>Service Class</b>	<b># of Tweets</b>
Business	5
Culture	25
Education	7
Environment	112
Government	33
Health	4
Property	33
Public Safety	16
Social Service	9
Transportation	128
Off-topic	9
Total	381

Table 4-3. The statistics of the knowledge base

<b>Service Class</b>	<b># of keywords</b>	<b># of wiki pages</b>
Business	17	17
Culture	42	42
Education	19	19
Environment	32	32
Government	55	55
Health	43	43
Property	30	30
Public Safety	35	35
Social Service	41	41
Transportation	47	47
Total	361	361

To evaluate the performance of our framework, we use the conventional leave-one-out procedure (Manning et al. 2008). We evaluate the classification performance over multiple runs. For each evaluation run, one request tweet is selected as the test data; and the remaining request tweets and the corresponding classes are used to train an NB classifier. The results of all the evaluation runs are averaged to obtain the overall classification performance. The evaluation metric is the precision rate (Manning et al. 2008), which measures the proportion of tweets classified as belonging to a service class that are actually in that class.

## 4.2 Performance Evaluations

In this section, we examine the effect of the knowledge base. We also compare the performance with and without using the temporal information, temporal smoothing techniques, user content enrichment, and WordNet synonym dictionary method for classification performance. To evaluate the effect of the knowledge base, the parameter  $\alpha$  is set between 0 and 1, and increased in increments of 0.1. When  $\alpha = 0$ , the framework ignores the knowledge base and simply uses the training tweets to construct an NB classifier. As

shown in Table 4-4, using the knowledge base (i.e.,  $\alpha > 0$ ) improves the classification precision. In addition, the improvement under  $\alpha = 0.1$  to  $0.9$  is significant. The low precision of  $\alpha = 0$  reflects the difficulty of classifying tweets. Using the knowledge base alleviates the text sparseness problem in tweets effectively and improves the classification precision significantly. It is noteworthy that setting  $\alpha = 0.6$  achieves the best precision, and the precision decreases gradually when  $\alpha > 0.6$ . The decline implies that the weight of the training tweets should be larger than that of the knowledge base. Meanwhile, the inferior precision of  $\alpha = 1$  indicates that the training tweets cannot be ignored because they are very close to the service classes in question.

Table 4-4. The effect of  $\alpha$  on the system performance

$\alpha$	Precision
0.0	29.65%
0.1	39.10%*
0.2	38.84%*
0.3	38.84%*
0.4	39.37%*
0.5	<b>40.16%*</b>
0.6	<b>40.16%*</b>
0.7	39.63%*
0.8	38.58%*
0.9	37.79%*
1.0	29.92%

The results marked with \* show improvements over the precision when  $\alpha = 0.0$  with a 99% confidence level based on the Z-statistic for two proportions (Keller, G. 2005)

Table 4-5 shows the effect of using temporal information. Again, the parameter  $\beta$  is set between 0 and 1, and increased in increments of 0.1. Parameter  $\alpha$  is set at 0.6 because of its superior precision score in the previous experiment. As shown in the table, using temporal information (i.e.,  $\beta > 0$ ) produces good classification precision. However, the improvement

over  $\beta = 0$  is small. After investigating this result, we concluded that the small improvement is due to the shortage of training data. As only 380 training tweets and their publication times are available in each validation run to produce 24 frequency distributions over the 10 services classes, the proposed framework would be biased by noisy temporal samples. However, the shortage problem will be resolved in the future as more request tweets are published on the 311NYC system.

Table 4-5. The effect of  $\beta$  on the system performance under  $\alpha = 0.6$

$\beta$	Precision
0.0	40.16%
0.1	40.42%
0.2	40.68%
0.3	<b>40.94%</b>
0.4	40.42%
0.5	40.42%
0.6	40.16%
0.7	39.63%
0.8	39.63%
0.9	39.90%
1.0	40.68%

To analyze the use of temporal information further, we examine the performance of  $\beta = 0$  (i.e., without using any temporal information) and 1 (i.e., using temporal information only). Table 4-6 compares their performance under various settings of  $\alpha$ . The table also lists the best performance under each setting of  $\alpha$ . We observe that setting  $\beta$  at 1 outperforms  $\beta = 0$ . In addition, seven out of the ten best results are obtained by setting  $\beta$  at 1. Hence, using the temporal information about tweets ensures a superior classification performance.

Table 4-6. The comparison between  $\beta = 0$  and  $\beta = 1$ 

<i>A</i>	$\beta = 0.0$ (i.e., without temporal)	$\beta = 1.0$ (i.e., only temporal)	The best performance under $\alpha$
0.1	39.11%	40.94%	40.94%
0.2	38.85%	40.68%	40.68%
0.3	38.85%	40.94%	40.94%
0.4	39.37%	40.94%	40.94%
0.5	40.16%	40.42%	40.42%
0.6	40.16%	40.68%	40.94%
0.7	39.63%	39.63%	40.42%
0.8	38.58%	39.63%	39.63%
0.9	37.80%	37.80%	37.80%
1.0	29.92%	33.86%	34.38%

Table 4-7 shows the effect of the two temporal smoothing techniques under  $\alpha = 0.6$ . Compared with table 4-5, the smoothing techniques does not improve classification performance significantly. Although these two smoothing techniques make the whole distribution more smooth, the data sparseness problem cannot solved totally by this method.

Table 4-7. The performance of add-one and exponential smoothing

<i>B</i>	$\alpha = 0.6(\text{add-one})$	$\alpha = 0.6(\text{exponential})$
0.1	40.15%	39.89%
0.2	39.63%	40.15%
0.3	39.63%	40.41%
0.4	39.89%	40.41%
0.5	40.15%	40.15%
0.6	39.63%	39.37%
0.7	39.63%	39.10%
0.8	38.84%	38.58%
0.9	38.84%	38.58%
1.0	38.58%	38.05%

Table 4-8 shows the effect of the user content enrichment method under  $\beta = 0.5$ . Compared with the previous result, we observe that using the enrichment method hardly improves system performance. Surprisingly, it decreases system performance a little bit. The decrease may be owing to the following two reasons: First, due to privacy concerns we only retrieve 62% user contents. The low proportion of user contents decreases the effect of the method. Second, we observed that only a small proportion of users argue inferior government services to their friends. Consequently, we include a lot of noise data into our training data that decreases classification precision.

Table 4-8. The performance of user content enrichment

<i>A</i>	$\beta = 0.5$
0.1	33.59%
0.2	33.33%
0.3	33.33%
0.4	33.85%
0.5	34.12%
0.6	34.38%
0.7	34.12%
0.8	34.12%
0.9	33.33%
1.0	30.18%

Table 4-9 shows the performance of using the WordNet synonym dictionary under  $\beta = 0.6$ . As shown in the table, using WordNet improves system performance significantly. The result indicates that using the synonyms of tweet contents enhances short text classification. It also means the data sparseness problem of short text can also be reduced by transforming the words in tweets into their synonym levels. In original training text, two words with similar meaning are different unless they are exactly the same word. However, in their synonym levels, they represent the same thing and furthermore fortify the power of their meaning

although these two words are different lexically. This is the benefit of transforming words into their synonyms and the reason why data sparseness problem can be reduced by this method. It is noteworthy that the precision decreases gradually when  $\alpha > 0.6$ . The decline implies that the weight of the synonym set should be larger than that of the knowledge base. This also implies that we cannot ignore the synonym of the training tweets.

Table 4-9. The performance of WordNet Synonym Dictionary

<i>A</i>	$\beta= 0.6(\text{without add-one})$	$\beta= 0.6(\text{add-one})$
0.1	54.07%	56.69%
0.2	54.33%	57.22%
0.3	54.86%	56.69%
0.4	55.38%	56.17%
0.5	53.81%	54.33%
0.6	51.97%	52.23%
0.7	49.87%	50.13%
0.8	43.31%	41.47%
0.9	34.12%	31.23%
1.0	13.65%	10.76%

To summarize, the inferior precision (i.e., 29.65%) of the Naive Bayes model demonstrates that tweet classification is difficult. By using the temporal information about tweets, the knowledge base, and the WordNet dictionary, the proposed framework achieves a superior classification precision (57.22%), which represents a 93% improvement over the precision of the Naive Bayes model.

## 5. Conclusion and Future Works

An increasing number of government services are exploiting Web2.0 applications to obtain instant feedback from the public. Generally, micro-blogs (e.g., Twitter) can provide seamless communications between governments and citizens. However, the problems of information overload and text sparseness make micro-blog management a difficult task. To solve the problems, we have proposed a classification framework that incorporates an external knowledge base and the temporal information of tweets into the Naive Bayes classification model. We further employ two smoothing techniques to leverage the temporal distribution, consider user content to enrich the content of the training tweets, and incorporate the WordNet synonyms into our classification model. Experiments based on the 311NYC dataset demonstrate that the proposed framework classifies tweets correctly and achieves a significant improvement over the Naive Bayes model.

This paper focuses on the publication time of tweets. In a future work, we will investigate different metadata of tweets to enhance the proposed framework. For instance, the social network of Twitter users could be investigated to identify opinion leaders for a certain service class. Analysing the tweets published by opinion leaders would help governments identify the subjects that citizens regard as the most important. We believe that by incorporating text mining techniques into Government2.0, government services will become more transformative, available, and interactive.

## References

- Androutsopoulos, I., Koutsias, J., Chandrinou, K.V., and Spyropoulos, D. (2000). Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach. In *Proceedings of the 4th PKDD's Workshop on Machine Learning and Textual Information Access*. pp.1-13.
- Chen, M., Jin, X. and Shen, D. (2011). Short Text Classification Improved by Learning Multi-Granularity Topics. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. Spain, Barcelona.
- Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., and Slattery, S. (2000). Learning to Construct Knowledge Bases from the World Wide Web. *Artificial Intelligence*.118(1-2):69-114.
- Gabrilovich, E. and Markovitch, S. (2007). Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp.1606-1611. India, Hyderabad.
- Ghaharhani, Saeed. (2004) *Fundamentals of Probability*. 3rd Edition. Pearson - Prentice Hall.
- Hao, L. and Hao, L. (2008). Temporal Data Driven Naive Bayes Text Classifier. *The 9th International Conference for Young Computer Scientist*. pp.699-702. China, Hunan.
- Hu, X., Zhang X., Lu C., Park E. K., and Zhou X. (2009). Exploiting Wikipedia as External Knowledge for Document Clustering. In *Proceedings of KDD*, pp.389–396. France, Paris.

- Java, A., Song, X., Finn, T., and Tseng, B. (2007). Why we Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop*. pp.56-65. USA, California, San Jose.
- Keller, G. (2005). *Statistics for Management and Economics*. 7<sup>th</sup> International Edition, International Thomson.
- Kelman, H. C. (1958). Compliance, Identification, and Internalization: Three Processes of Attitude Change, *Journal of Conflict Resolution*, Vol. 2. 51-60.
- Liu, Y., Scheuermann, P., Li, X., and Zhu, X., (2007). Using WordNet to Disambiguate Word Senses for Text Classification *ICCS 2007, Part III, LNCS 4489*, pp. 780–788.
- Manning, C. D., Raghavan, P. and Schütze, H. (2008). *Introduction to Information Retrieval*, Cambridge University Press.
- Nam, T. (2011). New Ends, New Means, but Old Attitudes: Citizens' Views on Open Government and Government 2.0. In *Proceedings of the 44th Hawaii International Conference on System Sciences*. pp.1-10.
- Phan, X., Nguyen, L. and Horiguchi, S. (2008). Learning to Classify Short and Sparse Text & Web with Hidden Topics from large-scale Data collection. In *Proceedings of the 17th International World Wide Web Conference*. pp.91-100. China, Beijing.
- Revesz, P. and Triplet, T. (2011). Temporal Data Classification using Linear Classifiers. *Information Systems Journal* Volume: 36, Issue: 1, Publisher: Elsevier, pp.30-41
- Rocchio Jr, J.J. (1971). Relevance feedback in Information Retrieval. In G. Salton, editor, *The SMART Retrieval System: Experiments in Automatic Document Processing*. Prentice-Hall.
- Rodríguez, Manuel de Buenaga; Gómez-Hidalgo, José María; and Díaz-Agudo, Belén. (1997). Using WordNet to complement training information in text categorization. *Recent Advances in Natural Language Processing-97*. 150-157.

- Sahami, M., and Heilman, T. (2006). A Web-based Kernel Function for Measuring the Similarity of Short Text Snippets. In *Proceedings of the 15th International World Wide Web Conference*. pp.377-386. Scotland, Edinburgh.
- Salon, G. and McGill, M. 1. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill Book Co., NewYork.
- Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, M. D., and Sperling, J. (2009). TwitterStand: News in Tweets. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp.42–51, USA, Washington, Seattle.
- Schuler, Doug (2001). Cultivating Society's Civic Intelligence. *Journal of Society, Information and Communication*, Vol. 4 No. 2.
- Scott, Sam and Matwin, Stan. (1998). Text classification using WordNet hypernyms. *Usage of WordNet in Natural Language Processing Systems: Proceedings of the Workshop*. 45-51.
- Shen, D., Pan, R., Sun, J. T., Pan, J. J., Wu, K., Yin, J., and Yang, Q. (2006). Query Enrichment for Web-query Classification. *Journal of ACM Transaction on Information Systems*. 24:320–352.
- Voorhees, E. (1993). Using WordNet to Disambiguate Word Senses for Text Retrieval. In *Proceedings of the 16th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval.*, pp171-180. Pittsburgh, USA
- Widrow, B. and Sterns, S. (1985). *Adaptive Signal Processing*. Prentice-Hall.
- Wigand, F. D. L. (2010). Twitter Takes Wing in Government: Diffusion, Roles, and Management . In *Proceedings. Digital Government Society of North America*, pp.66–71.

Wigand, F. D. L. (2010). Twitter in Government: Building Relationship One Tweet at a Time. In *Proceedings of the 7th International Conference on Information Technology*. pp.563–567. USA, Nevada, Las Vegas.

Yih, W. and Meek, C. (2007). Improving Similarity Measures for Short Segments of Text. In *Proceedings of Association for the Advancement of Artificial Intelligence*. pp.1489–1494. Canada, Vancouver.

Zelikovitz, S. and Hirsh, H. (2000). Improving Short-Text Classification Using Unlabeled Background Knowledge to Assess Document Similarity. In *Proceedings of the 17th International Conference on Machine Learning*, pp.1183–1190 USA, Stanford University.

Zelikovitz, S. and Hirsh, H. (2003). Integrating Background Knowledge Into Text Classifiers. In *Proceedings of the International Joint Conference of Artificial Intelligence*. Mexico, Acapulco.

