Enhanced Smoothing Methods Using Naïve Bayes Classifier for Better Spam Classification

Shruti Aggarwal

Assistant Professor, Dept. of CSE, S.G.G.S.W.U., Fatehgarh Sahib (Punjab), India.

Devinder Kaur

Research Scholar, Dept. of CSE, S.G.G.S.W.U., Fatehgarh Sahib (Punjab), India.

Abstract: Text Mining has become an important research area due to the plorification of electronic documents available on web. Spam (junk-email) identification is one of the important application areas of Text Mining. A good spam filter is not just judged by its accuracy in identifying spam, but by its overall performance i.e. reduction in the cost of classifying spam without much reduction in recall. In addressing the growing problem of junk e-mail on the Internet, statistical based approach called Naive Bayesian Classifier has been used to filter unsolicited bulk e-mail due to its simplicity and superior performance. It has been found that it largely depends on the smoothing method, which aims to adjust the probability of an unseen event from the seen event that arises due to data sparseness. The aim is at enhancing the performance of Naïve Bayes Classifier in classifying spam mails by proposing a modification to Jelinek-Mercer Smoothing and Dirichlet Smoothing method shows the high performance in case of varying data set, varying number of keywords and variations in smoothing factor based on the data set used.

Keywords: Naïve Bayes Classifier, Text Classification, Smoothing Methods, Spam Classification.

I Introduction

With the explosive growth of the textual information from the electronic documents and World Wide Web, proper classification of such enormous amount of information into our needs is a critical step towards the business success. Numerous research activities have been conducted in the field of document classification, particularly applying in spam filtering, emails categorization, website classification, formation of knowledge repositories, and ontology mapping [1]. However, it is time-consuming and labor intensive for a human to read over and correctly categorize an article manually [2]. Attempts to address this challenge, a number of approaches have been developed for accomplishing such purpose, including k-Nearest-Neighbor

(KNN) classification, Naïve Bayes classification, Support Vector Machine (SVM), Decision Tree (DT), Neural Network (NN), and Maximum Entropy [1]. Text classification finds immense applications in information management tasks. Some of the applications are document classification based on defined vocabulary, sorting emails as spam or non spam, or sorting emails into various folders, documents filtering, topic identification etc [3]. Spam or Junk mail, also called unsolicited e-mail, is Internet mail that is sent to a group of recipients who have not requested it. Because of readily available bulk-mailing software and large lists of e-mail addresses harvested from web pages and newsgroup archives, unsolicited mails cause many problems such as direct marketers bombard unsuspecting E-mail boxes with unsolicited messages regarding everything from items for sale, engulfing important personal mail, wasting network band width, consuming users' time and energy to sort through it, to crash mail-servers, get-rich schemes to information about accessing pornographic Web sites[4] [5] [6]. The statistical based approach used for spam classification is more efficient than the rule based approach. The most common and simplest statistical based method for spam classification is Naive Bayes Classifier due to its simplicity and strong independent assumption. As this classifier suffers from the problem of data sparseness, Smoothing Techniques has been proved to be an efficient method. The enhanced Smoothing Techniques with Naïve Bayes Classifier have been used to concentrate not only the problem of data sparseness of classifier but also as an effort to increase the overall performance along with the reduction in cost of classifying Spam.

II Classification of Spam

In this era of rapid information exchange, electrical mail has proved to be an effective means to communicate by virtue of its high speed, reliability and low cost to send and receive [4]. Also, in recent years, the increasing popularity and low cost of e-mail have attracted the attention of direct marketers as they use to send blindly unsolicited messages to thousands of recipients at essentially no cost [7]. While more and more people are enjoying the convenience brought by e-mail, an increasing volume of unwanted junk mails have found their way to users' mail boxes [4].

This explosive growth of unsolicited e-mail, commonly known as spam, over the last years has been deteriorating constantly the usability of e-mail [8]. Unsolicited bulk e-mail, electronic and Spam messages posted blindly to thousands of recipients, is becoming alarmingly common. For example, a 1997 study by Cranor & LaMacchia, 1998 found that 10% of the incoming e-mail to a corporate network was spam [5]. Junk mail, also called unsolicited bulk e-mail, is Internet mail that is sent to a group of recipients who have not requested it [4]. The task of junk mail filtering is to rule out unsolicited bulk e-mail (junk) automatically from a user's mail stream.

Methods of Spam Filtering:

Some anti-spam filters are already available. Two types of methods have been shown to be useful for classifying email messages.

1. Rule Based Methods: Rule based methods uses a set of heuristic rules to classify emails [4]. These are mostly based on manually constructed pattern-matching rules that need to be tuned to each user's incoming messages, thus it is a task requiring time and expertise. Furthermore, the characteristics of spam for e.g. products advertised, frequent terms change over time, requiring

these rules to be maintained [5]. RIPPER rule learning algorithm is one of the algorithms worked according to rule based method. Its performance is comparable to TF-IDF weighing method [4].

2. *Statistical based approach*: This approach models the difference of statistics based on the machine learning framework. Several machine learning algorithms have been applied to text categorization. These algorithms learn to classify documents into fixed categories, based on their content, after being trained on manually categorized documents [5]. Memory Based Learner, Naïve Bayes Classifier and Boosting Tree Classifier with Ada Boost algorithm are the examples of statistical based methods [4].

While several algorithms outperform the task of classifying mail messages, Naive Bayes has several advantageous properties. First, a classifier is constructed by a single sweep across the training data and classification requires just a single table look up per token, plus a final product or sum over each token. Other approaches like Support Vector Machines, Boosting, and Genetic Algorithms require iterated evaluation; approaches like k-means require several pair wise message comparisons while decision tree building is significantly slower than Bayesian table construction. Furthermore, since Naive Bayes only need to store token counts, rather than whole messages, storage requirements are small, the classifier can be updated incrementally as individual messages are classified [9].

III Naïve Bayes classifier

A Naive Bayes classifier is a simple probabilistic classifier which is based on applying Bayes theorem with strong (naive) independence assumptions. The approach of the task of text classification is a Bayesian learning framework. The text data is generated by probabilistic model and training data is used to calculate estimates of model parameters. The parameters of the model are estimated using a set of labeled training examples and every new example is classified using Bayes rule by selecting the class with the highest probability.

Models of Naïve Bayes Classifier:

Multivariate Bernoulli model: A document is represented by a binary feature vector, whose elements (1/0) indicate presence or absence of a particular word in a given document. In this case the document is considered to be the event and the presence and absence of words are considered as attributes of the event.

Multinomial model: A document is represented by an integer feature vector, whose individual elements indicate frequency of corresponding word in the given document. Thus the individual word occurrence is considered to be events and document is considered to be collection of word events. Multinomial model is more accurate than the multivariate Bernoulli model for many classification tasks because it considers the frequency of the words too.

Bayes Theorem:

Consider $X = \{x1, x2, x3, ..., x_n\}$ to be a set of feature vectors and $C = \{c1, c2, c3, ..., c_m\}$ be set of classes labels from set C. The probability of a new example being in class c using Bayes theorem is given by:

As P(X) is independent of the class, so it can be ignored.

$$\arg \max_{c \in C} P(c/X) = \arg \max_{c \in C} P(c)P(X/c)....(2)[3] [10] [11] [2]$$

Assumption of Naïve Bayes: Naïve Bayes Classifier assumes that all the attributes (terms or words) in the example are independent of each other. Thus, the presence of any attribute of the training example does not affect the presence of other attribute in the same or other example.

The parameters of the generative model are estimated using a set of labeled training examples and every new examples is classified using Bayes rule by selecting the class with the highest probability [3].

According to Naïve Bayes Classifier, if document d is to be classified, the learning algorithm should be able to classify it in required category c from set of classes C,

$$\arg \max_{c \in C} P(c/d) = \arg \max_{c \in C} P(c)P(d/c).....(3)[3][10][11]$$

In case of text documents, a document is a combination of words or terms and from the above discussion the attributes are unrelated to each other, we have:

$$P (d/c) = P (w_1/c) P (w_2/c) P (w_3/c)...P (w_4/c)(4)[3][10] [12]$$
$$P (d/c) = \prod_{1 \le k \le n} P(w_k/c)...(5)[10][3] [12] [11]$$

Thus, from equation (3), we have,

$$\arg \max_{c \in C} P(c/d) = \arg \max_{c \in C} P(c) \prod_{1 \le k \le n} P(w_k/c) \dots (6)[3] [10] [11]$$

Where P(c) is the prior probability and P (w_k/c) is the posterior probability. P(c) is calculated as:

$$P(c) = n_i / n_i$$
 (7)[3] [10] [11]

Where n_i is the number of documents of class c and n is the total number of documents in the training set.

$$P(w_{k}/c) = \frac{count(w_{k}/c)}{\sum_{w' \in V} count(w',c_{i})}$$
(8)[3] [11]

Where $count(w_k/c)$ denotes the number of occurrences of word w_k in the documents of class c and $\sum_{w' \in V} count(w', c_i)$ denotes the number of occurrences of all the words of vocabulary in class c. Posterior Probability is also known as maximum likelihood estimator.

For NB, maximum likelihood is calculated by Laplace Smoothing as:

 $P(w_{k}/c) = \frac{1 + count (w_{k}/c)}{\sum_{w' \in V} count (w',c_{i}) + |V|} \dots (9) [10] [3] [11] [2]$

Where, V is the size of vocabulary of training set.

IV Smoothing

Smoothing is a technique which adjusts the maximum likelihood estimate so as to correct the inaccuracy due to data sparseness. It is required to assign a zero probability to unseen words [13]. The name smoothing comes from the fact that these techniques tend to make distributions more uniform, by adjusting low probabilities such as zero probabilities upward, and high probabilities downward [14].

General formulation used in smoothing is:

$$P(w | c_i) = f(x) = \begin{cases} P(w|c_i), & w \text{ is seen in } c_i \\ \alpha_d P(w|C), & w \text{ is unseen in } c_i \end{cases}$$
(10)[9] [15] [13]

Thus, one estimate is made for the words seen in c_i and another estimate is made for words unseen in c_i . In case of words unseen in the c_i , the estimate is based on the entire collection, i.e., the collection model [16].

Not only do smoothing methods generally prevent zero probabilities, but they also attempt to improve the accuracy of the model as a whole. Whenever a probability is estimated from few counts, smoothing has the potential to significantly improve estimation [14].

Smoothing Methods:

In addition, several other smoothing methods are combined into the NB model. A number of smoothing techniques have been developed in statistical natural language processing to estimate the probability of a word: [13] [17]

1. JM Smoothing: This method involves a linear interpolation of the maximum likelihood model with the collection model, using a coefficient λ to control the influence of each [18].

$$P_{\lambda}(w|d) = (1-\lambda) \frac{count (w_k/c)}{\sum_{w' \in V} count (w',c_i)} + \lambda P (w_k|C)....(11)[16] [15] [19] [13] [14]$$

Where λ is the coefficient of smoothing, P (w_k/C) is the probability of word w_k with respect to the whole collection C and *count*(w_k/c) denote the count of word w_k with respect to the class c.

Here, the main emphasis is on correctly setting the value of λ . Set λ to be a constant, independent of example and tune to optimize the bias - variance tradeoff [19].

Problem encountered with JM Smoothing is that longer examples provide better estimates (lower variance) and one can get this by with less smoothing (lower bias).

2. *Dirichlet Smoothing*: A language model is a multi-nominal distribution, for which the conjugate prior for the Bayesian analysis is the Dirichlet distribution with parameters, [18]

 $(\mu p (w1|c), \mu p (w2|c), \mu p (w3|c) \dots \mu p (w1|c))$(12)

Thus, model is given by:

 μ is a pseudo-count [19].

3. Absolute Discounting Smoothing: It decreases the probability of seen words by substracting a constant from their counts [18]. The discounted probability mass is redistributed on the unseen words proportionally to their probability in the collection model [16].

Where $\delta \in [0,1]$ and $|c_i|$ is the number of unique words in c_i .

4. Two-Stage Smoothing: This smoothing method combines Dirichlet smoothing with an interpolation smoothing [18]

$$P_{\lambda,\mu}(w|d) = (1-\lambda) \frac{count (w,d) + \mu P(w|C)}{\sum_{w' \in V} count (w,d) + \mu} + \lambda P(w/C) \dots (15)[20][16] [13] [14] [18]$$

V. Modified Smoothing Algorithm for Spam Classification

In the already existing JM and Dirichlet Smoothing methods, the probability of word w_k in collection language model is calculated as,

$$P(w_k/C) = \frac{\sum_{j=1}^{m} count(w_k, c_j)}{\sum_{k=1}^{n} \sum_{j=1}^{m} count(w_k, c_j)}....(16)[13][18][21]$$

Where m is the total number of classes and n is the total number of vocabulary words. Thus, above equation estimates total occurrences of word with respect to each class to the total number of occurrences of each vocabulary word with respect to each class. In the modified version, probability of word in collection model is not considered, rather it is considered as a function of word, which is a uniform distribution probability multiplied by the occurrence of word in collection model and is given by:

$$P(w_k/C) = P_{unif}(W_k) \sum_{i=1}^{m} count(w_k, c_i)....(17)[21]$$

Where $P_{unif}(W_k) = \frac{1}{|V|}$, |V| is the total number of vocabulary words and $\sum_{j=1}^{m} count(w_k, c_j)$ is the total number of occurrences of word w_k in all classes. So, above equation becomes:

$$P(w_k/C) = \frac{\sum_{j=1}^{m} count(w_k, c_j)}{|V|}.$$
(18)[21]

With the replacement of total word count of each vocabulary word with respect to each class, overhead for calculating the probability of with respect to whole collection has been reduced.

Algorithm for Naïve Bayes classifier with Enhanced Smoothing Technique:

1. Let V be the vocabulary set. 2. For each category $c_i \in C$, Let w_k be the vocabulary word $\in V$, Calculate: $n1 = countofword (w_k,c_i)$ For each vocabulary word w' \in V, Calculate: n2 = $\sum_{w' \in V} count(w', c_i)$ 3. for collection C, Let w_k be the vocabulary word $\in V$, Calculate: $P(w_k/C) = \frac{\sum_{j=1}^{m} count (w_k, c_j)}{\cdots}$ where m is total number of classes. 4. If (Technique is Jelinek-mercer smoothing) For each category $c_i \in C$, For each word $w_k \in V$, $P(w_k/c_i) = (1-\lambda) \frac{\text{countofword } (w_k,c_i)}{\sum_{w' \in V} \text{count } (w',c_i)} + \lambda P(w_k/C)$ Where λ is a smoothing factor lies between (0, 1) 5. If (Technique is Jelinek-mercer smoothing) For each category $c_i \in C$, For each word $w_k \in V$, $P(w_k/c_i) = \frac{\text{countofword } (w_k,c_i) + \mu P(w_k/C)}{\sum_{w' \in V} \text{count } (w',c_i) + \mu}$ Where μ is a smoothing factor lies between (0, 1)

The above modification in the smoothing techniques used for spam classification with Naïve Bayes Classifier is checked. It reduces the cost factor without much reduction in recall and is able to handle zero probability problems of unseen words over the seen words. The experiment results obtained from the modified method is shown in next section.

VI Results

Naïve Bayes Classifier with modified Smoothing methods and existing Smoothing methods is implemented for classification of spam from the legitimate mails based on the text area of the mail. The results based on the varying data set size, varying number of keywords and varying the smoothing factor with respect to the accuracy, recall and cost of classification has been discussed. The experiment is performed on training set and test set build on the basis of personal mails collected from the different email-ids.



Figure1: Spam Classification using Naïve Bayes with Enhanced Smoothing Techniques in terms of Accuracy for varying number of documents.

In figure 1, the results are shown on the basis of varying data corpus. The accuracy of the NB with Enhanced JM and Dirichlet Smoothing methods gets improved by 10% and 20% respectively as compare to Naïve Bayes Classifier with existing Smoothing methods. But we can see that with the increase in the size of the data corpus, there is always slight increment or sometimes decrement seen in the values accuracy of these classifiers.

	For 1200 words			For 900 words		
Filter used	Accuracy	Recall	Cost	Accuracy	Recall	Cost
NB-L	0.55	0.7	0.6	0.57	0.8	0.7
NB with Old JM	0.50	0.6	0.7	0.53	0.7	0.6
NB with enhanced JM	0.50	0.6	0.6	0.60	0.8	0.5
NB with old Dirichlet	0.50	0.2	0.3	0.59	0.5	0.4
NB with Enhanced Dirichlet	0.55	0.2	0.1	0.65	0.5	0.2

 Table1: Spam Classification using Naïve Bayes with Enhanced Smoothing Techniques in terms of Accuracy, Recall and Cost varying number of keywords.

As depicted in table 1, the Naïve Bayes with enhanced Dirichlet Smoothing method is proved to be best in case of precise number of keywords; also the cost of Classifying Spam is less in this method with respect to other techniques, without much reduction in recall rate. With the precise number of documents, the overall performance has been increased by almost 10% in each classifier.



Figure 2: Spam Classification using Naïve Bayes with Enhanced Smoothing Techniques in terms of Accuracy for varying Smoothing Factor μ , λ .

The graph shown in figure 2 describes the performance of already existing methods with the modified methods in terms of Accuracy and varying Smoothing factors. In case of NB with Jelinek-Mercer Smoothing, old JM method has the highest value of accuracy at λ =0.5. But the new JM method shows the same results at two values i.e. λ =0.5 and λ =0.9 which is 10% greater than existing methods. In case of NB with Dirichlet Smoothing, old Dirichlet method has highest accuracy at μ =0.5. But new Dirichlet Method achieves the high accuracy of classification at μ =0.1, μ =0.5 and μ =0.7 which is 10-15% higher than existing methods.



Figure 3: Spam Classification using Naïve Bayes with Enhanced Smoothing Techniques in terms of Recall for varying Smoothing Factor μ , λ .

The graph shown in figure 3 describes the performance of already existing methods with the modified methods in terms of Recall and varying Smoothing factors. In case of Naïve Bayes Classifier with Jelinek-Mercer Smoothing, old JM method has the highest value of Recall at λ =0.5 and λ =0.3. But the new JM method shows the same results at two values i.e. λ =0.5 and λ =0.9, which is 10-15% high than the older one. In case of Naïve Bayes Classifier with Dirichlet

Smoothing, old Dirichlet method has highest recall at μ =0.1. But new Dirichlet Method achieves the highest value of recall at μ =0.1, μ =0.5 and μ =0.7 which is 10-20% higher than older method. Naïve Bayes with new JM method and Dirichlet method achieves the highest value of recall.



Figure 4: Spam Classification using Naïve Bayes with Enhanced Smoothing Techniques in terms of Cost of Classification for varying Smoothing Factor μ, λ.

The graph shown in figure 4 describes the performance of already existing methods with the modified methods in terms of cost and varying Smoothing factors. In case of Naïve Bayes Classifier with Jelinek-Mercer Smoothing, old JM method has low cost of classification at λ =0.5 which is 0.2 and high cost of classification at λ =0.9 which is 0.9. But the new JM method takes very less cost to classify spam at λ =0.5, λ =0.7 and λ =0.9 which is 0.05. In case of Naïve Bayes Classifier with Dirichlet Smoothing, old Dirichlet method has less cost to classify spam at μ =0.1 which is 0.4. But new Dirichlet Method achieves the less value of cost at μ =0.2 and μ =0.6 which is 10-20% less than older method. Naïve Bayes with new JM method and Dirichlet method has less cost of classification than older methods.

The result showed in the above graphs states that the Naïve Bayes Classifier with Enhanced Smoothing Techniques has increased the overall performance of Spam Classification by 10-20% than the Existing Smoothing Techniques.

VII Conclusion

Naïve Bayes Classifier with Jelinek-Mercer Smoothing and Dirichlet Smoothing has been implemented for Spam Classification. With this implementation, there are numerous of issues like the zero probability due to data sparsity, cost of classifying spam from legitimate mails without much reduction in recall and increase in overall performance and problem of false positives have been studied in depth. To overcome these issues, Naïve Bayes Classifier is implemented with the modification in Smoothing techniques for calculating the collection probability for the model. The improved method shows the high performance in case of large data set, with precise number of keywords, with variations in smoothing factor. Thus, the modified method used not only increases the accuracy but also lowers down the cost of classification without much reduction in recall. On the basis of studied data set, following are the conclusions:

First, with varying data set size, the performance of classifying spam increases by 5-10%. The Naïve Bayes Classifier with modified smoothing method achieves the highest performance as compare to Naïve Bayes Classifier with already existing smoothing methods.

Second, for precise number of keywords, Naïve Bayes Classifier with the enhanced Dirichlet smoothing method achieves the highest performance. Also, the overall performance of the system increases with precise number of keywords as compare to large dictionary size.

Third, in case of varying Smoothing factors and based on the studied data-set, the Naïve Bayes with enhanced JM Smoothing shows the highest performance for spam classification at smoothing factor $\lambda = 0.5$. The results obtained by using enhanced Jelinek-Mercer Smoothing method at $\lambda = 0.5$ is same at $\lambda = 0.9$. The Naïve Bayes Classifier with enhanced Dirichlet Smoothing method shows the better result at $\mu = 0.7$, $\mu=0.5$.

Fourth, to compare both the Naïve Bayes Classifier with enhanced Jelinek-Mercer and Naïve Bayes Classifier with enhanced Dirichlet Smoothing methods, it can be said that enhanced Jelinek-Mercer Smoothing method is more accurate than enhanced Dirichlet Smoothing Method based on data-set studied.

VIII Future Work

It has been discussed that the overall performance of the task of classifying spam depends largely on the cost of classification. This states the reduction in the rate of false positives. The solution to this issue has led to the modifications in the calculating the collection probability by using uniform distribution probability. As expected this enhancement in the existing method has decreased the cost and increased the overall performance of classification. As in this solution, we have used the modified smoothing techniques, there are number of techniques that can be used as future work such as: First, there are various good Classification Algorithms other than Naive Bayes such as Support Vector Machine, Centroid Based, Nearest Neighbor, etc. Such techniques can be applied for Spam Classification task to see the improvements.

Second, the modified smoothing with Naïve Bayes Classifier can be used to classify the mails into not Just as a spam but also in number of folders.

Third, there are other various smoothing techniques Good Turing, Katz-Backoff and Witten-Bell that can be applied to Spam Classification to check the performance issues. These smoothing methods can be implemented as n-gram models, which represent the relation between the different features of the vector.

Fourth, the Naïve Bayes Classifier with modified Smoothing methods can be implemented in hierarchical manner to check the further improvements.

References

[1] S. L. Ting, W. H. Ip, Albert H.C. Tsang "Is Naïve Bayes a Good Classifier for Document Classification", International Journal of Software Engineering and Its Applications, Volume 5, No. 3, July 2011, pp. 37-46.

[2] Andrew McCallum, Ronald Rosenfeld, Tom Mitchell, Andrew Ng "Improving Text Classification by Shrinkage in Hierarchy of Classes", International Conference on Machine Learning, 1998.

[3] Hetal Doshi, Maruti Zalte "Performance of Naïve Bayes Classifier – Multinomial Model on Different Categories of Documents", National Conference on Emerging Trends in Computer Science and Information Technology, 2011, pp. 11-13.

[4] ZHANG Le, YAO Tian-shun "Filtering Junk Mail with A Maximum Entropy Model", ICCPOL, 2003, pp. 446-453.

[5] Ion Androutsopoulos, John Koutsias, Konstantinos V. Chandrinos, Constantine D. Spyropoulos "An Experimental Comparison of Naive Bayesian and Keyword-based Anti-Spam Filtering with personal e-mail messages.", In Proceedings of the 23rd Annual International ACM SIGIR conference on Research and Development in Information Retrieval, 2000, pp. 160-167.

[6] Mehran Sahami, Susan Dumais, David Heckerman, Eric Horvitz "A Bayesian approach to Filtering Junk E-mail", In Learning for Text Categorization, Volume 62, 1998, pp. 98-105.

[7] Ion Androutsopoulos, John Koutsias, Konstantinos V. Chandrinos, George Paliouras, Constantine D. Spyropoulos "An Evaluation of Naive Bayesian Anti-Spam Filtering", ArXiv, 2000. [8] Hotho, Andreas, Andreas Nürnberger, Gerhard Paab "A Brief Survey of Text Mining." Ldv Forum, Volume 20, No. 1, 2005, pp. 19-62.

[9] Khuong An Nguyen "Spam Filtering with Naïve Bayesian Classification", Machine Learning for Language Processing, April-2011.

[10] Kevin P. Murphy "Naïve Bayes Classifier", Department of Computer Science, University of British Columbia, 2006.

[11] Ajay S. Patil and B.V. Pawar "Automated Classification of Naïve Bayesian Algorithm", Proceedings of international Multi-Conference of Engineers and Computer Scientists, Volume1, March 2012, pp. 14-16.

[12] In Jae Myung "Tutorial on Maximum Likelihood Estimation", Journal of Mathematical Psychology, Volume 47, 2003, pp. 90-100.

[13] Chengxiang Zhai, John Lafferty "A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval", SIGIR, September 2001, pp. 9-12.

[14] Stanley F. Chen, Joshua Goodman "An Empirical Study of Smoothing Techniques for Language Modeling", Harvard University, Cambridge, August 1998.

[15] Mark D. Smucker, James Allan "An Investigation of Dirichlet Prior Smoothing's Performance Advantage", CIIR Technical Report IR-391, January 2005.

[16] Jing Bai, Jian-Yun Nie "Using Language Models for Text Classification", In AIRS, 2004.

[17] A. M. Jehad Sarkar, Young-Koo Lee, Sungyoung Lee "A Smoothed Naïve Bayes-Based Classifier for Activity Recognition", IETE Technical Review, Volume 27, Issue 2,2010, pp. 107-119.

[18] Quan Yuan, Gao Cong, Nadia M.Thalmann "Enhancing Naïve Bayes with Various Smoothing Method for Short text Classification", Proceedings of 21st international conference on world wide web, 2012, pp. 645-646.

[19] Victor Lavrenko "An Overview of Estimation Techniques", August 2002.

[20] Trevor Stone "Parameterization of Naive Bayes for Spam Filtering", University of Colorado at Boulder, 2003.

[21] Astha Chharia, R.K. Gupta "Enhancing Naïve Bayes Performance with Modified Absolute Discount Smoothing Method in Spam Classification", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013.