

An Enriched Frame Work for Auto Answering System for SMS using Fuzzy Classification

Manisha Khadse, Vinit Deshmukh, Gauri Joshi, Supriya Gaikwad

Manisha Khadse, Professor, Dept. of Computer Engineering, MMIT College, Maharashtra, India

Vinit Deshmukh, Student, Dept. of Computer Engineering, MMIT College, Maharashtra, India

Gauri Joshi, Student, Dept. of Computer Engineering, MMIT College, Maharashtra, India

Supriya Gaikwad, Student, Dept. of Computer Engineering, MMIT College, Maharashtra, India

Abstract - In recent years, Short Message Service (SMS) has been widely exploited in day-to-day communication. In today's world people are too busy to get time for other relative work such as to get any information about organization or to make enquiry about any government project because it is so lengthy process. They became so handy with the mobile phones to get all the relevant information. Our concept is very useful in such situation. We are presenting a concept of providing a service for the student of a college via SMS. Where a student of a college will send the SMS to the public number which is provided by the college, where the SMS is received by that mobile number of the college, then it sends the SMS to the web server. Herein the web server this SMS is replied with the desired answers which is been already entered by the respective departments, through the mobile as a gateway. Now to develop software as specified above we will require classifying the receiving SMS at the server side. We need to define some acronyms which will define the respective department such as admission, time table, sports, placement, etc. Here, SMS classification remains the important task which leads to send the SMS to respective department and will generate the correct reply. Hereby we have found that Fuzzy classification along with TF-IDF, Stemming, Stop Removal Algorithms gives good performance and yields better accuracy and time efficiency as compared to previous techniques like Naïve Bayes and Support Vector Machines in SMS Classification system.

Key Words: SMS classification, information retrieve, mobile gateway, fuzzy logic, SMS classification.

1. INTRODUCTION

Short Message Service (SMS) has been widely exploited in day-to-day communication. A general concept of Campus Short Message Service (CSMS) is to receive the query of any user and send the appropriate reply to the

same user related to that particular query about any department of the organization.

The biggest e.g. of this one is reality shows are in peak like anything and for each contestant the viewer needs to vote as their favorite with predefined code displayed. Once SMS of viewer hits the server, it classifies SMS based on contestant ids that was displayed and are predefined. So, we got motivation from this and thought the same in bigger manner, where we plan to make the software to give this SMS service to student of a collage on demand. This makes our software with higher intelligence and with huge capacity to deal with many numbers of Short Messages.

So, we are presenting a concept of providing a service for the student of a college via SMS. Where a student of a college will send the SMS to the public number which is provided by the college, where the SMS is received by that mobile number of the college, then it sends the SMS to the web server. Herein the web server this SMS is replied with the desired answers which is been already entered by the respective departments, through the mobile as a gateway.

Now to develop software as specified above we will require classifying the receiving SMS at the server side. We need to define some acronyms which will define the respective department such as admission, time table, sports, placement, etc. Here, SMS classification remains the important task which leads to send the SMS to respective department and will generate the correct reply.

The following sections consists of Literature survey, Existing System, Proposed System, Algorithms, Conclusion and Reference.

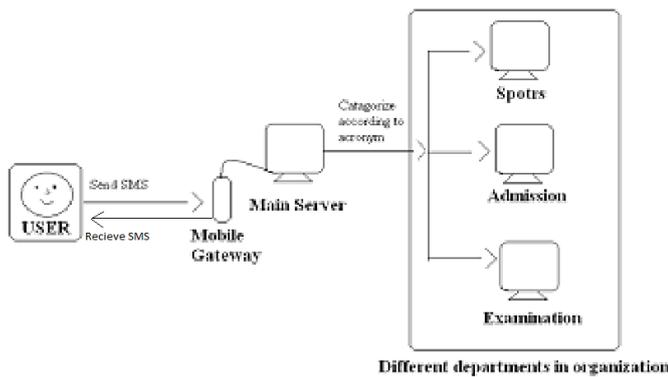


Fig1. Structural design / Architecture Diagram.

LITERATURE SURVEY

Campus SMS Service is an alternative mobile based portable solution for solving query of college students. The system is able to provide asked solution/information to students or any person situated at hostel, out of campus i.e anywhere anytime. In addition, it provides user more friendliness as user can send SMS with no predefined keywords to only one number of all departments of the collage. Earlier systems based on campus services do not provide on demand query solutions of different departments to students. If the systems do provide such facilities they are based on some limited keywords and user needs to remember them always. At the same time we also studied the mobile applications of SMS classification, which first converts the SMS language to normal one and then classifies the spam Short Messages and put them in different folders. So, we noticed and decided to make a system that will classify the SMS as well as reply the required details automatically.

1.1 Comparative Study between Naïve Bayes and Neural Network (MLP) Classifier:

Naïve Bayesian : A naive bayes classifier is a simple probabilistic classifier that is based on applying bayes theorem with strong (naive) independence assumptions. A more descriptive term for the probability model would be independent feature model.

Decision Tree: In decision tree structure, each internal node (non leaf node) denotes a test on an attribute, every branch represents an outcome of the test data, and each leaf node holds a class label. The topmost node in a tree is the root node. There are various algorithms available for making decision tree, such as CART, ID3 and C4.5 etc. These are the greedy (i.e. non-backtracking) approach in which decision tree construct in top down divide and conquer manner.

SVM : In SVM is a new method for the classification of both linear and non-linear data. SVM are supervised learning

models and it associated with learning algorithms that analyze data and recognize patterns. The basic SVM takes a set of input data, for each given input, which has two possible class forms the output making it a non-probabilistic binary linear classifier.

Neural Network : A neural network is a set of connected input or output units in which each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of the input tuples. Neural Network learning is also referred to connections between units.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

1.2 An improved method of term weighting :

In this paper, study traditional term weighting scheme tf-idf approach and analyze its drawback is given, and it presents a newly term weighting method to improve the classic tf-idf based approach and proposed a supervised term weighting scheme, which directly makes use of a kind of information ratio to judge a term's contribution for a category. The improved approach has become a new way to compute term's weights to avoid assigning higher weights to rare terms. Experimental results have proved that our approach is an effective solution to improve the performance of text classification.

1.3 An Improved TF-IDF Weights Function Based on Information Theory:

In this paper the classic TF - IDF weights function is discussed. In order to solve the disadvantage of classical TF-IDF function, an improved weight function is proposed. It uses the distribution information among classes and inside a class to describe the relationship between the terms, and it solves the shortage of former TFIDF function which only considers the relationship between the term and the whole text but neglects the relationship between different terms. In addition, it effectively enhances the capability of computing the term weighting.

1.4 Classification Of English Phrases And SMS Text Messages Using Naïve Bayes And Support Vector Machines Classifiers :

In this paper a comparative performance evaluation was done on several different types of SMS text classifiers. In particular, the classification performance of the weight enhanced Multinomial naive Bayes, Poisson naive Bayes, and

L2-loss Support Vector Machine were compared. In addition, the effects of preprocessing and incorporating additional features on the classifiers were examined. The preliminary experimental results show that there is no significant gain in classification efficiency by using the modified classification algorithms. However the feature space used by the classification methods decreased, which is beneficial for resource limited environments, due to smaller usage of space. In addition, feature space reduction decreases the calculation complexity of the Support Vector Machine approach

1.5 Construction of Concept Network from Large:

A proposed method to construct a network that describes information about the relation of things that are related to a certain keyword from electronic texts. This method has two characteristics (use of TF-IDF and use of deletion of unrelated words based on a topic). It extracts related words using a TF-IDF. By using TF-IDF, this method not only extract important words but also delete unrelated words in the networks. While expanding a network by adding words, words that are not related to the keyword are likely to be added. The proposed system deletes such unrelated words. For the deletion of unrelated words, it proposed two methods, the topic restricted method and the topic-related method. This method experimentally confirmed that the proposed TF-IDF-based related word extraction method showed better results than a method that uses conditional probabilities to extract related words. Further more, we conducted experiments to verify the effectiveness of deleting unrelated words using the topic-restricted method. It was observed that the topic-restricted method could delete most unrelated words and maintain approximately 0.8 related words from the original network. In addition, the topic-related method could delete some unrelated words and retain most related words from the original network.

1.6 Selecting fuzzy if-then rules for classification problems using genetic algorithms :

In this paper, we proposed a genetic-algorithm-based method for selecting significant fuzzy if-then rules to construct a compact fuzzy classification system with high classification power. The biased mutation probabilities played an important role in reducing the number of fuzzy if-then rules. The classification power of selected fuzzy if-then rules was also examined. It was shown by computer simulations on the iris data that the performance of selected fuzzy if-then rules was very high for training data.

That is, the classification rates for training data were above 99% in almost all the parameter specifications of genetic algorithms. The performance of the proposed method for test data was also high when the number of generations in genetic algorithms was small.

2. Existing system

Campus Connect is a text Messaging service in partnership with Globe Tele Comm. It provides an ICT solution that caters to the information and communication needs of employees, students, parents and the community through SMS. With this service, the following are possible:

- Employees and students can receive free text message such as advisories, announcements, information campaigns, notices, updates, school activity information and others.
- Employees, students and the community can inquire for selected information from the university using a Globe or TM SIM. Example: a student or a parent may inquire about the university's enrollment schedule by doing the following: - Compose a message and type BSUCAR [space] enroll then send to 2327. - In \bar{i} while, you will receive a message about the universities enrollment schedule. -Users need not register to use this service.

3. Proposed system

Campus Short Message Service System is a solution to user/student to get required information about collage anywhere anytime. The system will first receive SMS of user through a thread created in our system. The project is divided into different tasks first it will find acronyms in SMS by comparing each word in acronym database, assign priority to acronyms, relate it to respective department and find its answer, to send the answer through mobile gateway as solution. The information/solution feeding process will be done by administrators of respective departments by logging into the system every day. This project is carried out to build a automated, less complex but comprehensive system for use of collage or other related fields (by doing small changes in database). Campus SMS system is designed so that it will be able to provide required information from the organization to users out of its campus and beyond its timings with cost of only a SMS. The way this system flow is from users SMS

to server system classifying it and from there to administrator feeding information and the information again going to user as SMS. The system will always check the inbox of mobile for SMS as input. The system design will look like a complete flow from SMS receiving, classifying, information feeding and replying back to user.

4. Algorithms

In this paper we are using four algorithms :

- a. Stop Removal
- b. Stemming
- c. TF-IDF
- d. Fuzzy Logic

I. Stop Removal

1. Start
2. Input as words
3. Check preposition list
4. For each word
5. Check in preposition list
6. If present in preposition list
7. If not present then go to (4)
8. Replace word with empty character
9. Stop

II. Stemming

1. Start
2. Input as words
3. For each word
4. Check Identifying Suffixes
5. Suffixes Replacement Rules
6. Replace Suffixes
7. If not replaced then go to (3)
8. Stem words obtained
9. Stop

III. TF-IDF

1. Start
2. Input as preprocessed words
3. Term frequency in a record (TF)
4. Term Frequency across all the record of DB (DF)
5. Log (DF) = IDF
6. TF-IDF
7. Stop

IV. FUZZY LOGIC

1. Start
2. Term Weight of the SMS
3. Crisp Value Setting

4. Record Weight Calculation
5. Fuzzification
6. If- then Rules
7. SMS Classification
8. Stop

We have compared Fuzzy Classification and TF-IDF, Stemming, Stop Removal algorithms with Naïve-Bayes and Support Vector Machines (SVM) and found that our proposed system gives much better results, accuracy and efficiency in SMS Classification System.

3. CONCLUSION

Campus Short Message Service is the concept can be use in any college or any organization. It is developed to make the user friendly environment i.e. user can easily interact with the organization or take any information over their mobile phone by sending his/her query over one public number which is provided by the organization. Our system aims to provide accuracy and time efficiency in information retrieval and query classification through SMS service using Fuzzy Classification algorithm.

Acknowledgment

We would like to acknowledge the support of our project coordinator and faculty members at the Department of Computer Engineering, Marathwada Mitra Mandal's Institute of Technology Lohgaon, Pune, India for their support in developing this system. We would like to articulate my gratitude to our guide Prof. M.V.Khadse for her extremely noticeable hold up and support also to our HOD Prof. P.Daflapurkar. Their direction is strength behind the completion of this paper. We are thankful for all the suggestion and hint provided by her.

REFERENCES

- [1] Hua Jiang, Ping Li, Xin Hu and Shuyan Wang, "An improved method of term weighting for text classification", IEEE 2014.
- [2] Na Wang and Pengyuan Wang, Baowei Zhang, "An Improved TF-IDF Weights Function Based on Information Theory", *Department of Electronic and Communication zhengzhou Institute of Aeronautical Industry Management Zheng Zhou 450015, China.*
- [3] J. Maier and K. Ferens, "Classification of English phrases and SMS text messages using bayes and support vector

machine classifiers”, *Telecommunications Research Laboratories (TRLabs)*, 100-135 Innovation Drive, Winnipeg, Manitoba R3T 6A8, CANADA

[4] Yuta Doen, Masaki Murata, Ryuta Otake and Masato Tokuhisa, “Construction of Concept Network from Large Numbers of Texts for Information Examination Using TF-IDF and Deletion of Unrelated Words,” in Proc. International Conference on Management and Service Science, IEEE Press, Sept. 2009, pp. 1-4. Department of Information and Electronics Tottori University.

[5] Hisao Ishibuchi, Ken Nozaki, Naohisa Yamamoto, and Hideo Tanaka “Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms”, China.