

Latest way to modify Text Message classification with auditory module

Payal Kurhade¹, Sharvari Medhane², Shivanjali Kankate³, Vinita Rokade⁴, Mr. D.S.Thosar⁵

^{1,2,3,4} Final year (BE) Student ⁵ Assistant Professor

^{1,2,3,4,5} Department of Computer Engineering S.V.I.T. Chincholi, Nashik

Abstract -Short Message Service (SMS) is an integral service of the mobile phone for users to communicate with people which is faster and convenient way to communicate. However, it has some limitations like incapability of searching and categorization of SMS, scheduling, marking SMS and there is scope to improve it. To overcome various limitations, we have proposed a mobile application with title MojoText - Text Messenger which solves real time problem of text messaging. Our system provide core functionalities of text messaging and beside to that various facilities like categorization of messages based on personal, social, transactional and user defined categories with color codes, searching with customized date, scheduled text delivery, hiding of messages inside the app, reminders for due dates of billers, validity of texts, starred messages, pinned chats, signature, backup and recycle bin.

Key Words: Android, Text-to-Speech module, Natural Language Processing (NLP), Text pattern matching.

1. INTRODUCTION

We are in that world where we cannot imagine our life without mobile phone. We can handle everything from our mobile phones like our bank account, our professional work, some office work, etc. through the cell phones[6]. In our proposed system we are going to inventing the system for the handling the text messages. As we know that the number of messaging apps are provided with the multiple functions and features in android cell phones. But for the convenient and easy use of text messages there is no any system is provided[1,3]. In this resulting system we are going to add some extra feature and functionalities to the text messages where the use of text message is going to be easy. Now a days we can associate with the mobile number to our bank, various shops, PAN cards, ATM cards, Aadhar cards or in shopping malls, etc. for getting the details or important information like OTP's, educational purpose, etc.[7,8].

The best text classifier is a classifier that efficiently classify the huge number of messages in a reasonable time and short time and with the fast accuracy, and that provides helps to the human for the finding the message and easy to access[4]. Many techniques and algorithms are provided for the text categorization of the messages and the text of the message. The text classification process can be defined as classify the text into new

documents based on the knowledge and by using the algorithms provided in a classification system[2,4]. In the phase we are going to classify the text and speech it out the same time by using the text-to-speech module[5]. Classification is an important task in our system which we are going to develop and classification is an important concept of both data mining and machine learning[10], however, most of the learning techniques and algorithms comes from the machine learning community[7].

The text categorization or classification is coming from machine learning techniques and research[5,9]. A number of text classification techniques are applied through the development of the system a lot of approaches have been proposed in the our system, the automatic text categorization after the incoming message is arrive is still a major area of overview because the effective and improvement of the text message are still needs and its the time to provide some additional features to the text message[4].

3. ARCHITECTURE

The proposed system is based on the three essential parts first on is the input of the system second one is the intermediate process of the system and last one is resulting output of the system so the system architecture is given as following diagram. The implementation of the system from the very initial stage here that to avoid trouble inconvenience from the searching or finding the text messages we are developing the system.

The resulting system gets input through the incoming message or arriving message send by the sender these messages in initial stage stored like a are simple text messages and then the system application works on the text messages by using the different algorithms. The next process is shown in the architecture. In the process is the having different categories of the message like transactional, personal and company or service and so on these can be done by using the different algorithms techniques and the regular expressions. This is the implementation of the system. This can be done by using the algorithms are given as follows:

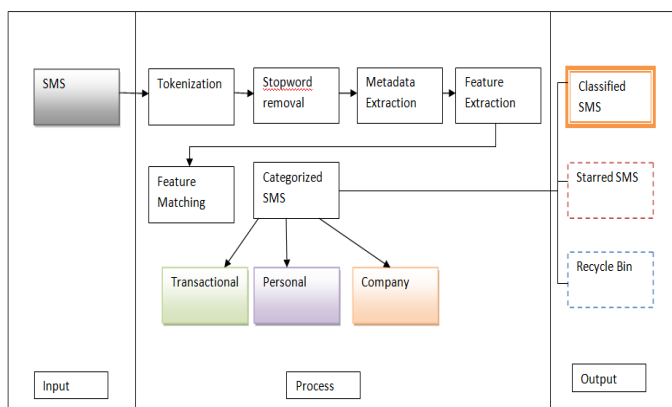


Fig : System Architecture

1. Stopword removal

2. Pattern matching (Regular expression, Token matching)

3. Decision tree (Date, category classification). Using this algorithms we can done the further process.

Input:

SMS-

The input of this proposed system is sms because we develop a system for messaging app. To sort out or to classify the messages, hence the input is important or essential part of the system.

Process:

Process is the actual flow of implementation of the system. How the system actually works on the categories of messages is comes under the process block of system architecture.

1. Tokenization:

Tokenization is the first part of the process. In which the main message coverts it into small tokens. Tokens are the part of messages which are use to classification of the message. Tokenization is nothing but dividing the message into small tokens and tokens are nothing but the sequence of characters.

2. Stopword Removal:

In this section of stopwords removal the similar or common words are get remove from the message to extract the main data of the message like is, the,etc. This is very useful for searching certain kind of message.

3. Metadata Extraction:

After stopwords removing the main data get remains. That data is get extracted for use. This is called the metadata extraction.

4. Feature Extraction:

In this the common features of the sms get extract. Means using feature matching technique the feature extraction should be done.

5. Categorized SMS:

The Classification of SMS having three important sections:

5.1 Transactional: In transactional section the messages contains the banking related information. The text message is related to bank notices or any transactions credit and debit and so on.

5.2 Personal: In this the messages will contains the message from human to human mean contact list messages. While doing the classification of this messages the personal messages get sort differently. This makes the ease of handling the application.

5.3 Company: The messages coming from company like the marketing related messages or the messages for publicity of any brand are separated in this category of classification.

3. ALGORITHM

The system provides solution to the categories of the messages and audio messages for the convenient use of the users[10,11]. When the incoming message is arrive at the receiver end then it will starting fot the pattern matching of the incoming text if the message[5]. And when its relates to the particular pattern then it will put the message into the that particular category. This is the important to the users who are going to access the system. This system implements an algorithm for the categorization of the message. The next step after this processing it will extract the some unique feature of text message and classifying the message[9,10]. This feature is used in this system for classification purpose this modern classifier belongs to the many new features in it.

Input:

SMS text

Process:

1. Read the SMS text
2. Stop word removal for the text.
3. Pattern matching for the input text
4. Regular expression is apply on the sms
5. Get the different patterns
6. Take a decision and put the sms in the related category
7. Detect the date patterns for the due dates

Output:

Classified SMS

4.EXCECUTABLE PROJECT SNAPSHOT

The following are the executable screenshot of the project. First it may contain the messages in the inbox then it will contain the service oriented messages which will useful for daily updates. Then next one contains the starred messages of the project. Then the next snapshot contain the personal category of the messages. Then next contain the side bar options of the system. Then next up we have the transactional messages or the banking messages of the resulting project.

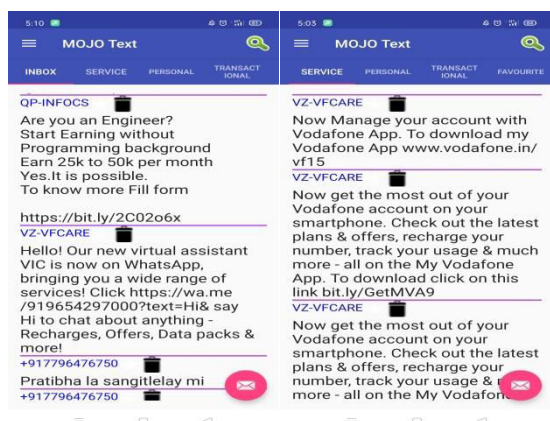


Fig: Inbox containing all the messages

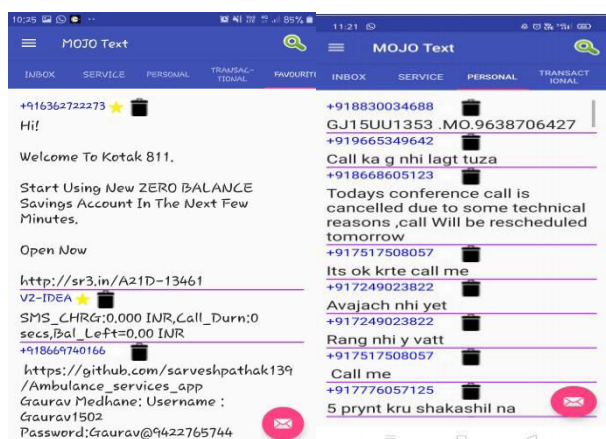


Fig: Starred messages

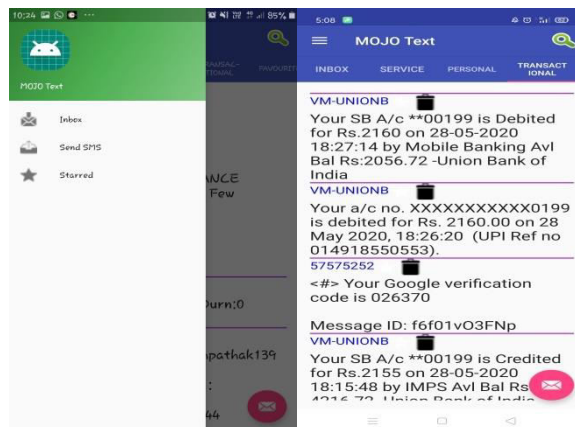


Fig: Side bar of the Messaging app.

Fig: Transactional messages.

4. PROJECT RESULT

The module information is nothing but actual convenient resulting system. In our proposed system there are two important resulting modules. These modules will show the our system is successfully implemented or not implemented. The main aim of the system is to provide the classification or categories of the text messages this categories can be a different different.i.e service, personal, transactional for the ease the daily use of the system this will very useful for the users to use it in their daily lives for the convenient use like other messaging app we are used by the text messages. The classification of the text messages is provide in the first module of the system this is very important in proposed system.

The second resulting module is to speech out the incoming message for the people who don't able to see the messages like blind peoples or when we are driving or doing some important works. This module is also very important for the people who cannot able handle the phone at any time the messages so they can listen the audio of the message. In this proposed system we are also providing the important feature to this module that if someone in any problem or need my help or the message will having the words like emergency, urgent, help, accident, etc. then this type of message will pop up the different type of tone from the normal message tone so the person who is received the message is knows that this particular person needs my help or there is some emergency.

The third resulting module of the system is going to shows or stores the starred message and recycle bin. In this module the users can able to stored their favorite messages or they can mark the particular message as starred message. This will be also shows the important dates messages. For example, my SIM card balance is going to expired soon at that time the system will shows it in the due dated message so I can easily

identify that it's a due date message. It will show the automatic reminders in the system. Then next up we add the feature of the system is not used for the text message yet is the recycle bin. In this you can able recover the messages which are deleted before but you need it for a reason so here you also access the deleted message.

Smart message classifier is the application for classification of text messages. The modules in our system are basically Categorization of messages in three categories as per personal, services and transactional. We are using stopword removal algorithm for removing common words use in that messages. This is helpful to extract the important data for further input to classification model by using regular expression. By applying regular expression on message it can identify that the message belongs from which category and do its classification.

Actually, while implementing this module it take a look on message data and scan it recognizing pattern in message content whether if message is coming related banking then it contains some banking related keywords like account number, credit, debit, loan, etc. so by identifying these things the system takes a decision and do the classification of message. The another module is of due dated messages, in which the deadline of date in message get seen and put that message into due date section. It reminds user about that data. Once the date is expires that message get remove from that section. If new due dated message coming then simultaneously it become added further in due dated section.

6. APPLICATION

1. Useful to extract important information.
2. Useful while car driving
3. Categorization of messages
4. Manages the due dated messages
5. Useful for blind people

7. ADVANTAGES

1. User Friendly:

The proposed system is user friendly because the messages are easily categorized and maintained

efficiently. User can easily handle the system and convenient for the users.

2. Speech out messages:

To speech out the incoming messages the text to speech module is used. Further mute unmute option is also provided for which message should speech out and which not.

3. Categorization of messages:

The messages are categorized like personal, transactional, services etc. because of this the Separation of messages and identification of them becomes easy.

4. Due dated messages:

The date wise important messages shown on different section and it gives us reminder. When the date has gone that time that message get expired and remove from that section.

8. DISADVANTAGE

It requires a android mobile for installation of application. We can not install it on desktop.

9. FUTURE SCOPE

In the future to make this system more superior we will also add the recycle bin facility for messages because once any message get deleted by mistakenly and we want it back then for such a kind of messages we will add the recycle bin to reading that deleted messages in our inbox. And also we are providing the starred message which will be important for the storing the favorite messages. Significant results have been obtained from this work, corresponding to which this research can be taken to real world application level for classification of text message.

10. CONCLUSIONS

In this paper, we mainly concentrated on discussing for SMS classification. We can do the classification on the different different categories of the messages. In this system we are successfully implemented the classification of the text message. This is the proposed system where the we can also recovers the deleted message after deleting it. Here the system will also able to do the storing the favorite message this was the important implementation of the system. Then there is the second part where the message will be in the audio format which can speech out the your whole message by using the text to speech model. The results obtained from our evaluation of the Classifiers of the incoming messages.

REFERENCES

- [1] 1. Castiglione, A., De Prisco, R., De Santis, A.: 'Do You Trust Your Phone?', E-Commerce and Web Technologies, Linz, Austria, September 2009 pp. 50-61.
- [2] Ho, T., Kang, H., Kim, S.: 'Graph-based KNN algorithm for spam SMS detection', Journal Of Universal Computer Science, 2013, 19, (16), pp. 2404-2419.
- [3] Church, K., & Oliveira, R.D. (2013). What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. Mobile HCI.
- [4] Silalahi M., Hardiyati R., Nadhiroh I. M., Handayani T., et al.: 'A text classification on the downstream potential of biomedicine publications in Indonesia', International Conference on Information and Communications.
- [5] 'NLP with gensim (word2vec)' <http://www.samyazaf.com/ML/nlp/nlp.html> accessed: January 2018 .
- [6] Wensen L., Zewen C., Jun W., et al.: 'Short text classification based on Wikipedia and Word2vec' 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, October 2016, pp. 1195-1200 .
- [7] Z. Cataltepe and E. Aygun. "An improvement of centroid-based classification algorithm for text classification". *IEEE 23rd International Conference on Data Engineering Workshop*, 1-2 pp. 952–956, 2007.
- [8] S. Dumais and H. Chen. "Hierarchical classification of Web content". In *Proceedings of the 23rd Annual International ACM SIGIR conference on Research and Development in Information Retrieval, Athens, Greece*, pp. 256–263, 2000.
- [9] A. McCallum and K. Nigam. "A comparison of event models for naive bayes text classification". In *AAAI-98 Workshop on Learning for Text Categorization*, 1998.
- [10] G. D. Guo, H. Wang, D. Bell, Y. X. Bi, and K. Greer."UsingkNN model for automatic text categorization". *Soft Computing*, 10(5), pp. 423–430, 2006.
- [11] R. N. Chau, C. S. Yeh, and K. A. Smith. : "A neural network model for hierarchical multilingual text categorization". *Advances in NeuralNetworks, LNCS*, 3497, pp. 238–245, 2005.
- [12] Makatolwayama and Takenobu Tokunaga. "Cluster based text categorization: a comparison of category search strategies". In *proceedings of the 18th Ann Int ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'95)*, pp. 273-281,1995.
- [13] J. Clark, I. Koprinska, and J. Poon, "A Neural Network Based Approach to Automated E-mail Classification," Proceedings of the IEEE/WIC International Conference on Web Intelligence, Canada, 2003.
- [14] X. Zhang, J. Zhao, Y. LeCun, "Character-level Convolutional Networksfor Text Classification," Proceedings of the 28th International Conference on Neural Information Processing Systems, vol. 1, pp. 649- 657, 2015.
- [15] SMS Spam Collection V.1 Dataset. Publicly available onlineat:<http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/>. Spam SMS Dataset 2011-12. Available online on request at: <http://precog.iiitd.edu.in/requester.php?dataset=smsspam>.
- [16] Bozan, Y. S., Çoban, Ö., Özyer, G. T. et al.: 'SMS spam filtering based on text classification and expert system' 23ndSignal Processing and Communications ApplicationsConference (SIU), Malatya, Turkey, May 2015, pp. 2345-2348.