International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9, Issue-7, May 2020

Design and Implementation of Grievance Filing Application using Automatic Text Classification

Shradha Sawant, Manjushree Shinde, Samruddhi Siddhamshettiwar, Nisha Watpade, Jayashree Jagdale

Abstract: ***This research aims to design a Grievance Filing System built using automatic text classification without any manual interruption. Various methodologies are followed to achieve it and are implemented. Performance of the different algorithms is discussed. People are less aware about the lengthy methods for lodging complaints. We propose a simplified process of enrolling grievances to ministries. The system accepts grievances in recorded voice form. The system is designed for Marathi language. Input in the form of speech will ease people's comfort for lodging grievances. We present a model where voice is first preprocessed, followed by text classification using deep learning approaches such as CNN and LSTM, the grievances will be sent to respective ministry. This system can be used by government ministries to get grievances from common people through a simplified process. User will be notified on the progress of their lodged complaint and on its successful resolution by respective ministry.***

Keywords**:** ***Deep Learning, LSTM, CNN, Text Classification.***

I. INTRODUCTION

In today's technologically fast-growing world, complaints should be lodged conveniently using online platforms. However handling existing platforms is a bit difficult for new user. Simplification in this process will lead to user's ease. There are different portals for different ministries. In our model, a single platform will handle complaints regarding all ministries. In existing platform user manually selects the department which sometimes leads to incorrect department selection. Due to this the ministries need to manually analyze whether the lodged complaint belongs to their department or it needs to be forwarded to another department.

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author
Shradha Sawant*, Dept. of Information Technology, Pune Institute of Computer Technology (PICT), Pune, India. Email: shradhasawant1711@gmail.com.
Manjushree Shinde, Dept. of Information Technology, Pune Institute of Computer Technology, Pune, India. manjushree884@gmail.com
Samruddhi Siddhamshettiwar, Dept. of Information Technology,Pune Institute of Computer Technology, Pune, India. samruddhi38@gmail.com
Nisha Watpade, Information Technology, Pune Institute of Computer Technology, Pune, India. Email:nishawatpade@gmail.com
Jayashree Jagdale, Assistant Professor, Dept. of Information Technology, Pune Institute of Computer Technology, Pune. Email: jbjagdale@pict.edu

In this project we are planning to record user's complaints in Marathi and English language which will be converted to English text. Further text classification will classify the text according to different ministries.

Grievance filing system formally means registering complaints of corresponding problems to the ministry in order to get resolved. Automatic classification of the complaint simplifies the process and reduces the tendency of sending the grievance to the incorrect ministry department. Every department can view/update processing complaint and in case of any discrepancy it can forward to another department. Text classification will be done using deep learning methods.

II. SYSTEM ARCHITECTURE

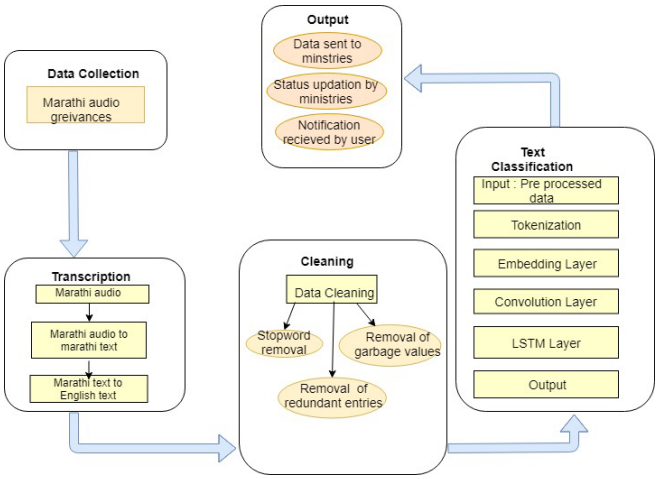

Figure 1. System Architecture

A. Data Collection Module

In this module the user will state his problem in his native language. In our system, the user will lodge his grievance in Marathi language by recording audio of his complaint. S/He will be able to see list of all her/his recorded complaints.

B. Conversion Module

Input to this language is voice of user which is Marathi voice. With Google API, we convert this Marathi voice to Marathi text, and then Marathi text to English text. Further, this English text will be given as input to the classification model.

C. Preprocessing Module

The dataset available had mixed entries of English and Marathi grievances.

The Marathi grievance entries are translated to English grievances using Google library in the preprocessing step. Text was cleansed through processes such as filling in missing values, smoothing the noisy data, stemming, tokenization, removing stop words. Now English text is termed as English grievances.

D. Classification Module

In this module we get preprocessed English grievances as input. We are using Convolutional Neural Network along with a combination of LSTM for complaint classification. The complaint related to a specific ministry will be sent to it. All ministries have their respective login credentials through which they can view their respective complaints and take corresponding action. User will be immediately notified of successful lodging of the complaint

E. Reporting and Response Module

In this module the ministry will report about the status of the complaint. The ministry has five options of status updation as per rules namely Assigned, Pending, In Process, Rejected and Closed. On every update the user will be sent a notification by the ministry department. The user is also able to see progress of complaint along with assigned department and user credentials through an android application.

III. EXISTING METHODOLOGY

Existing Grievance Filing portals include public grievance portal i.e. pgportal, aaple sarkar portal[4]. These portals facilitate grievance lodging in text format, no audio grievance lodging is available on these platforms. In existing platforms users should select the ministry and sub department, to which he wants to lodge complaint. Every user is not fully aware of all departments and grievances accepted by it, there is a chance of sending a grievance to the incorrect department and the complaint receiving department has to handle this by forwarding the complaint to the correct department. This leads to a time consuming and lengthy procedure. Moreover, at district level grievances of all the departments are received by the Collector Office and the complaints are manually classified to different departments, which can also lead to misclassification of grievances. So existing grievance lodging platforms have no support of machine learning techniques.

Existing Model:

- User has to select the ministry, to which he wants to file the complaint.
- Detail information filling is needed.
- Complaints filed are in English language.
- Traditional technologies are used.

IV. DEEP LEARNING APPROACH

Convolutional Neural Network is a multi-layer neural network which is composed of multiple convolutional layer and pooling layers in turn. Every layer consists of many two-dimensional planes and every plane consists of independent neurons. Convolution is nothing but the sliding window function applied to matrix. Convolutions is also called as feature detector. CNN is composed of some nonlinear-activation function like Relu. Unlike traditional neural network, CNN has convolutions over input layer to compute output. Each convolution layer is like filter. CNN is fully connected layer and classification targets are number of output nodes[2]. Figure 2 shows architecture of LSTM combined with CNN. CNN are quite efficient in terms of representation. Major advantage of convolutional neural network is it does not need whole vocabulary to representation. We can easily add filter with more than 5-grams with will be useful while text classification.

Pooling Layers are key aspects of CNN. By applying max operation to the pooling, we can perform pooling. Fixed size of output matrix which is typically required for classification is supplied by pooling layer. By keeping most silent information, it reduces dimensionality of output matrix. In this layer feature maps of convolutional layer are pooled. This layer helps to extract the most prominent features. Another layer of CNN is Fully Connected layer which is used to connect all the features and output values to classifier.

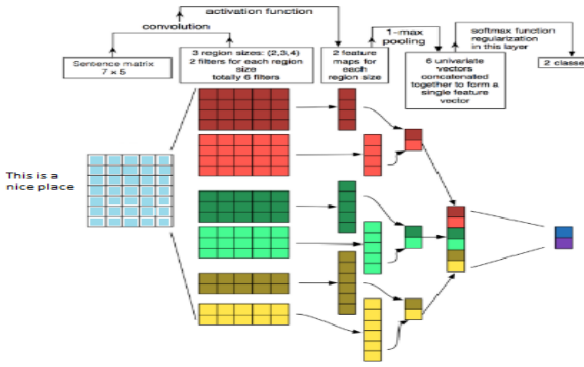

Figure 2: Combining LSTM and CNN

V. PROPOSED METHODOLOGY: DEEP LEARNING APPROACH (COMBINING LSTM AND CNN)

Combining LSTM and CNN makes the text classification more precise [7]. The LSTM based CNN output vector of CNN is fed as input vector to the LSTM model. After each feature is extracted let H be the output vector where H=[h1,h2,h3,....,hT]T .Here, length of the vector is equal to the number of hidden layers in LSTM. Here the maximum value of the feature map is obtained by max pooling operation, same as in CNN. The maximum value of the feature map represents the most important feature. Multiple convolutional filters are applied to obtain multiple feature maps from the text. Softmax function is used for classification. The structure of LSTM adds a memory cell for storing history information, and the update, deletion and output of history information are controlled by three gates respectively, they are input gate, forget gate and output gate task. On the one hand the model takes advantage of the LSTM to preserve historical information and context information in long text, and resolve the problem of vanishing gradient, on the other hand uses the CNN to further extract the local features of the text. The model combines the advantages of LSTM and CNN, and makes up for the shortcomings of CNN.

A. Data Pre-processing

This is the initial step in text classification. This step includes stemming of words for ex. 'playing' will be converted into 'play', removing stop words, tokenizing i.e. forming tokens from the sentences. All forms of a word are converted to root words in stemming.

B. Vocabulary Creation

Input is in the form of sentences i.e in text. But neural network requires numeric data. Creating vocabulary is nothing but mapping distinct words to unique number. This vocabulary will help to convert text data into numeric data so that the neural network will get numeric data. We will use tensorflow libraries like 'lookup' and 'gfile' for creation vocabulary.

C. Word Processing

In this space actual mapping of words with its respective index is done. For this tokenization of sentence is done i.e. splitting of sentence in words. Various sentences have different length and this leads to the formation of variable length word vector. To take a word vector of equal length padding is done. For example: dense= [['Some' 'title' 'ZYXW' 'ZYXW' 'ZYXW' 'ZYXW'] ['A' 'longer' title' 'ZYXW' 'ZYXW' 'ZYXW'] ['An' 'even' 'longer' 'title' 'ZYXW' 'ZYXW'] ['This' 'is' 'longer' 'than' 'doc' 'length']] numbers= [[8 4 0 0 0 0] [1 3 4 0 0 0] [9 2 3 4 0 0] [5 7 3 11 6 10]]

D. Embeddings

Word processing gives wide input. To reduce the high dimensional data to low dimensional data we need an embedding layer.

E. Convolutions

Convolution layer is used to learn about the sequence of the words. To learn sequences along with unique words, word pairs and word triplets, etc are embedded. Convolution layer applies a sliding window over input data and lets the neural network learn the weights to apply to the adjacent layer.

F. LSTM Layer

LSTM layer is used for feature extraction. Output of the convolutional layer after max-pooling is fed to the LSTM layer.

G. Training and Testing

The dataset will be divided into training and testing dataset. The model will be trained on the available training dataset. For gaining high accuracy we will train on as huge data as possible. Then, a model will be tested on testing data.
Proposed Model:

- Complaint will be automatically sent to ministry using text classification, no need of selecting ministry.
- Complaint is recorded in the form of audio.
- Complaints can be filed in native language like Marathi.
- New technologies like Machine Learning is used.

VI. IMPLEMENTATION

A. Dataset

In order to train our model we used Grievance Dataset made available from Pune Municipal Corporation. The dataset had, 1 lakh unique entries targeting 12 departments. We limited the data by selecting four departments: Solid Waste Management Electricity department, Water supply department and Road department. The dataset had each complaint correctly labelled with its respective department. All the data corresponding to the four departments has maximum unique entries. The dataset consists of mixed Marathi and English complaints. Our classification model is trained on 2983 unique entries consisting of all four department grievances. The model was fed with a testing data of 995 entries.

B. Experimental models

We have applied following four approaches for text classification:

1. CNN Approach for Text Classification: This approach includes five layers of neural networks. First layer is an embedding layer using a pre-trained Glove word vector. Next three layers of this approach consist of convolutions and max-pooling for feature extraction and the last layer is output layer which uses softmax function. Input to the first input layer is tokenized vectors of grievance text data. Accuracy of this approach was comparatively less than other approaches.

2. LSTM Approach for Text Classification: This approach includes five layers. First layer is an embedding layer using one-hot encoding. Next layer is the dropout layer which overcomes the drawback of overfitting. Further layer consists of LSTM , which performs feature extraction. Again dropout layer is applied followed by output layer

3. CNN with LSTM Approach for Text Classification: Combination of LSTM along with CNN improves the accuracy of the model. It consists of six layers namely embedding layer using one-hot vector encoding, dropout layer, convolutional layer with max-pooling, LSTM layer, dropout layer

4. CNN with LSTM Approach for Text Classification (Glove Pre-trained word-vector): In this approach Glove pre-trained word vector is used in embedding layer. It overcomes the disadvantages of one-hot vector encoding. This approach yields maximum accuracy in comparison to other implemented approaches.

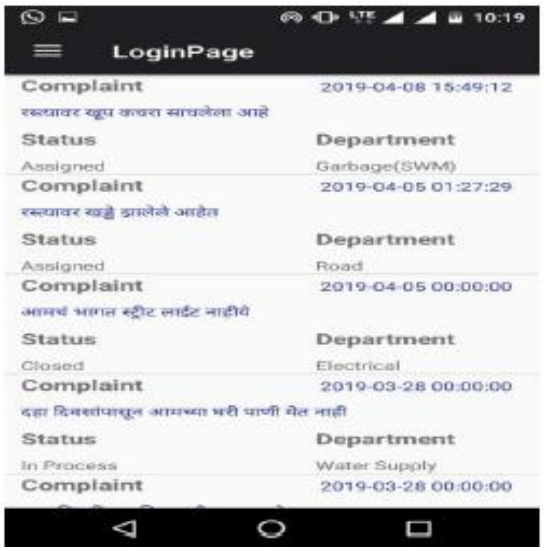

Figure 3: Grievances classified.

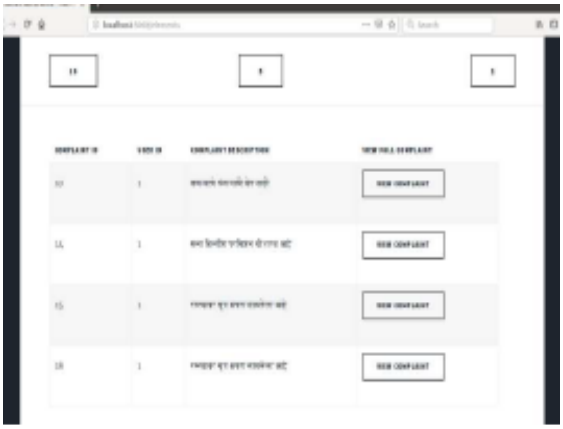

Figure 4: User Interface, ministry end.

Figure 3 shows the complaints filed are automatically classified into the designated ministry. Figure 4 is the user interface that ministry personal will be able to view in order to take further decisions on the filed grievance.

VII. RESULT AND DISCUSSION

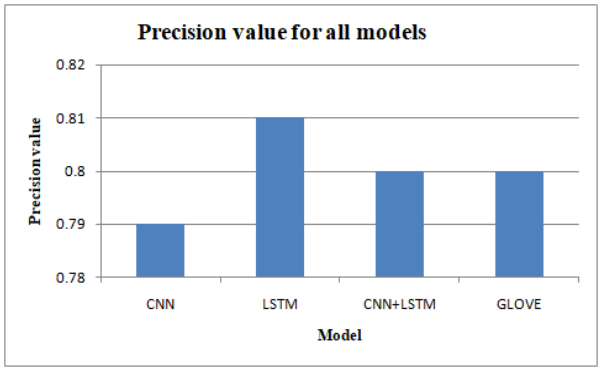

Figure 5: Precision values for classification using CNN, LSTM, CNN with LSTM and Glove models

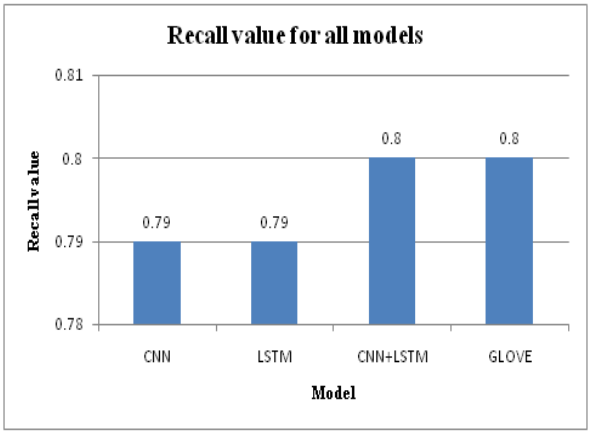

Figure 6: Recall values for classification using CNN, LSTM, CNN with LSTM and Glove models

Figure 5 shows the comparison graph of Precision values of all the models. This graph shows that the LSTM model has the highest precision value of 0.81 that is 81%. Figure 6 contains the Recall values plotted for all the models. This graph shows that the CNN combined with LSTM and CNN combined with LSTM with Glove re-trained word vector model has the highest recall values 0.8 or 81%. There is 2% rise than that of CNN and LSTM. Figure 7 contains the comparison graph of F1-Scores of all the models. Here we see, CNN combined with LSTM and CNN and LSTM with Glove Pre-trained word vector model have higher F1-Scores of 80%.

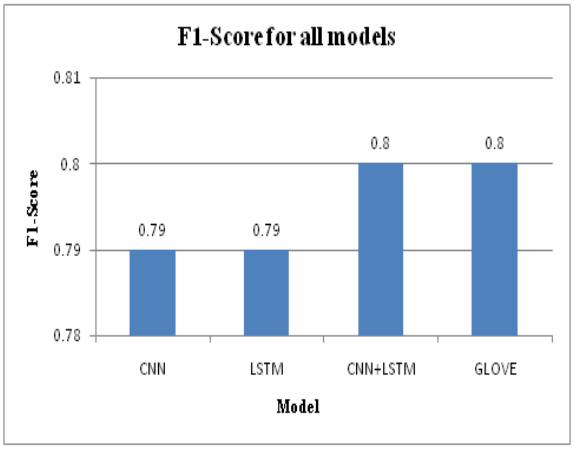

Figure 7: F1-score values for classification using CNN, LSTM, CNN with LSTM and Glove models

Table- I: Name of the Table that justify the values

Algorithm	Layers	Accuracy
CNN	First (Glove)embedding layer, 3 convolution layer	78.59%
LSTM	First embedding layer, dropout layer, LSTM layer, dropout layer	78.39%
CNN with LSTM	First embedding layer, dropout layer,1 convolutional layer, LSTM layer, dropout layer	80.10%
CNN with LSTM (Glove Pre-trained word-vector	First (Glove)embedding layer,3 convolution layer, LSTM layer	81.61%

We can say that the combination of CNN and LSTM and one with Glove Pre-trained word vector model show better performance with respect to recall and F1 score. Whereas precision of LSTM is better than all other methods as LSTM gets trained well on text data. Table 1 lays out details of implementation of every model. As seen in the table,
Accuracy of CNN and LSTM with Glove Pre-trained word vector model is 81.61% and is shown to be best as compared to other models under consideration.

VIII. CONCLUSION

The speech recognition and recording of the complaint in audio features of the system facilitates the complaint filing process for the user. Text Classification using deep learning i.e. combination of CNN and LSTM helps in improving the accuracy of the system. The proposed system involves getting complaints in the Marathi language in the form of audio and translating it to English Text. Grievance specific ministry is identified and a complaint is sent to the respective ministry. It eliminates the problem of user's selection of inappropriate department/ministry and thus smoothens the grievance filing system. Future work can be extended by including more regional languages for lodging complaints and also by identifying whether the user is giving suggestions or registering complaints.

ACKNOWLEDGEMENT

This project has been supported by Pune Institute of Computer Technology (PICT), Pune. We thank the authorities of PICT for resources, support and motivation. We thank Mr. Pawar, Deputy Collector, Panhala (now Pune) for helping us to understand the working of the system and Mr R. Jagtap, HOD, Pune Municipal Corporation for providing us the required data.

REFERENCES

1. Zhao,R. and Mao, K., 2017. Topic-aware deep compositional models for sentence classification. IEEE/ACM Transaction on Audio, Speech, and Language Processing 25(2), pp.248-260.
2. Hassan, A. and Mahmood, A., 2018. Convolutional recurrent deep learning model for classification. IEEE Access, 6, pp.13949-13957.
3. Kim, H.J., Kim, J., Kim, J. and Lim, P., 2018. Towards perfect text classification with Wikipedia-based semantic Naïve Bayes learning. Neurocomputing, DOI: 10.1016/j.neucom.2018.07.002
4. https://grievances.maharashtra.gov.in/en
5. Vijayan, V.K., Bindu, K.R. and Parameswaran, L., 2017, September. A comprehensive study of text classification algorithm. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp.1109-1113).IEEE
6. Li, C., Zhan, G. and Li, Z., 2018, October. News Text Classification Based on Improved Bi-LSTM-CNN. In 2018 9th International Conference on Information Technology in. Medicine and Education (ITME).
7. Zhang, J., Li, Y., Tian, J. and Li, T., 2018, October. LSTM-CNN Hybrid Model for Text Classification. In 2018 IEEE 3rd Advanced Information Technology, Electronics and Automation Control Conference (IAEAC) (pp. 1675-1680)IEEE.
8. Chen, G., Ye, D., Xing, Z., Chen, J. and Cambria, E., 2017, May. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 2377-2383). IEEE. J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com
9. Shu, B., Ren, F. and Bao, Y., 2018, September. Investigating Lstm with k-Max Pooling for Text Classification.In 2018 11th International Conference on Intelligent Computation Technology and Automation(ICICTA) (pp. 31-34). IEEE.
10. J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.
11. C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
12. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].
13. M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.
14. J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com

AUTHORS PROFILE

Shradha Sawant, has completed diploma in Information Technology from GRWPT, Tasgaon, Sangali. She then completed her under graduate Company, India as Application Development Associate. Her research interests are Machine Learning and Data Science. She studied BE computer degree from Department of Information Technology from Pune . She is currently working at Accenture Solutions Pvt. Ltd

Manjushree Shinde, Diploma in Information Technology from Government Polytechnic, Pune, India. She completed her Under graduate degree in Bachelor of Engineering in Information Technology from Pune Institute of Computer Technology, Pune, India. She is currently working at Amdocs, Pune, India as a Technology Integration Engineer. Her research interest areas are Data Analysis and Database Management Systems.

Samruddhi Siddhamshettiwar, Bachelor of Engineering in Information Technology, Pune Institute of Technology, Pune, India. She is currently working as a Software Engineer at Varian Medical Systems, Pune, India. Her research interest areas are Data Analytics and Machine Learning.

Nisha Watpade, Bachelor of Engineering in Information Technology, Pune Institute of Computer Technology, Pune, India. She is currently working at Mu Sigma, Bangalore, India as a Trainee Decision Scientist. Her research interest areas are Machine Learning and Data Analysis

Jayashree Jagdale, received a Diploma in Computer Technology from Pravaranagar, India and BE Computer Engineering degree from Mumbai University in 1995. She received ME in Information Technology, from MIT College of Engineering, affiliated to the Pune University, in 2011. Presently she is pursuing her Ph.D., in the field of Sentiment Analysis from PAHER Pacific University, India. She is presently working as an Assistant Professor at the Department of IT at Pune Institute of Computer Technology, Pune, India. She has more than 19 years of academics experience. Her research interests are Text Mining and Language processing. She is a Life Member of ISTE, INDIA.