DOI: http://dx.doi.org/10.26483/ijarcs.v10i6.6497

Volume 10, No. 6, November-December 2019

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

INTERACTION ANALYSIS OVER SPEECH FOR CALL CENTRE

Parth A Kholkute, Veerus D’mello, Rhea Kolhapurkar and Nicholas Patric
Computer Engineering Department,
St. Francis Institute of Technology, Mumbai, India

Abstract:Customer Service Centre is the second most important consideration just after the actual product. Also, customer service is one of the biggest contributors to the cost component for any firm. We aim to apply well-known data mining techniques to the problem of predicting the quality of interactions like those done in call centers and the problem of predicting the quality of service. The analysis of call center conversations will provide useful insights for enhancing Call Center Analytics to a level that will enable new metrics and key performance indicators (KPIs) beyond the standard approach. These metrics rely on understanding the dynamics of conversations by highlighting the way participants discuss topics. The main focus will be to reduce the average handling time, is a call center metric for the average duration of one transaction, typically measured from the customer’s initiation of the call and including any hold time, talk time and related tasks that follow the transaction. Get real-time solution. The main operations will be speaker diarization, speech to text, agent analysis, emotion recognition andother measures to help with the analysis. We will use RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) for emotion analysis, consisting of vocal emotional expressions in sentences spoken in a range of basic emotional states (happy, sad, anger, fear, disgust, surprise and calm). Emotion recognition is done by extracting features from the audio from its Mel-frequency cepstral coefficients (MFCCs) and passing it through a convolutional neural network. All of this will happen in real time as the call is taking place.

Keywords:CNNs, data mining, emotion recognition, speech diarization, speech to text.

I. INTRODUCTION

In our increasingly industrialized and globalized world, a large number of companies include call centres in their structures and more than $300 billion is spent annually on call centres around the world. For a customer, addressing the call center actually means addressing the company itself, and any negative experience on the part of the customer can lead to the rejection of company products and services. Hence, for the company, it is very important to ensure that a call centres function effectively and provides high quality service to its customers. Call centres collect a huge amount of data, and this provides a great opportunity for companies to use this information for the analysis of customer needs, desires, and intentions. Such data analysis can help improve the quality of customer service and lower the costs.

II. PROBLEM FORMULATION

Call center optimization is an important part of customer relationship management which consists of people, processes, technology and strategies. Service quality of a call center is a result of comparison of actual service performance and customer expectations. Evaluating the service quality which is offered by customer service agent to customer is more difficult than evaluating the product quality. Reduce total call time, i.e. the average handling time, get real time emotion of the speaker, get real time solution and correctly route the customers to respective agents to solve their needs.

III. LITERATURE SURVEY

Call centres provide services for many types of sectors such as telecommunication, finance, transportation, health, automotive etc. Several studies have proposed various approaches and solutions for the problem of evaluating agent performance. Performance evaluation in call centres is generally performed through listening randomly selected calls from recorded calls, and evaluating the words one by one in the related conversation. Obvious demand for automatic performance evaluation systems to reduce employee costs and to increase the time efficiency. Takeuchi has analyzed the recorded calls from a rental car reservation office with Trigger Segment Detection to find whether a customer has the intention of booking a car or not. Mishne has proposed a call centre monitoring system that uses text analytics and information retrieval methods. The system is used to analyze the content of call center conversations and detect the main issue addressed in the call. The project has presented speech analytics system adapted automatic speech recognition and text mining technologies.

[1] Minnucci (2004) reports that the most required metrics by call center managers are indeed the qualitative ones topped by Call Quality (100%) and Customer Satisfaction (78%). However, these performance metrics are difficult to implement with the adequate level of accuracy. For instance, the Baird study (2004) points out that for Customer Satisfaction, accuracy can be “negatively affected by insufficient number of administered surveys per agent resulting in not enough samples of individual agent’s work to constitute a representative sample. The result could be an unfair judgment of the agent’s performance and allocations of bonuses based more upon chance, good fortune than merit.” Accuracy is defined as true indication and it depends on the actual level of performance attainment, especially

with regard to statistical validity [2]. Current approaches to Call Centre Analytics are mostly based on Speech Analytics and Text Mining, which is essentially Search and Sentiment Analysis. Recorded speech is first indexed and searched against a set of negative terms and relevant topics. There are currently two main approaches for speech indexing: i) Phonetic Transcription and ii) Large Vocabulary Conversational Speech Recognition (LVCSR). Another common approach to the analysis of call centre data is that of automatic call categorization through supervised machine learning (Gilman et al. 2004; Zweig et al. 2006; Takeuchi et al. 2009). These methods failed in providing satisfactory results even in very broad categories. The problem still lies on data sparseness and that huge amount of training data is necessary to achieve reasonable discriminatory power. Getting huge training data is not an option also because training is highly influenced by domain specificity.

Transferring trained models from a domain to another would be problematic [3]. Existing state-of-the-art SER methods apply deep Convolutional Neural Networks (CNNs) trained on spectral magnitude arrays of speech spectrograms. To achieve high accuracy, complex CNN structures must be trained on a very large number which is in the order of millions of labelled spectrograms. This method called “fresh training” is computationally intense, time consuming and requires large graphic processing units (GPUs). At present the availability of large datasets of emotionally labelled speech is limited. On the other hand, in many cases, close to the state-of-the-art results can be achieved using a much simpler process of transfer learning [4].

IV. DESIGN

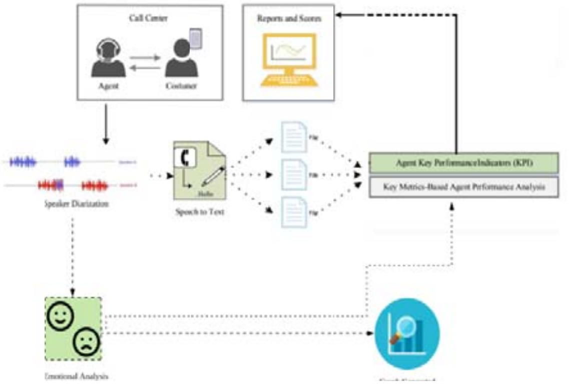


Figure 1. Architectural Design

The system will process the live call. Speaker diarization will be performed on the call to break down the call into two speaker components. Customer Speech audio and Agent Speech audio will be given for Emotional Analysis, where the detected emotions are shown with the help of a graph. Audios will then further be given for Speech to Text Transcription, where the output is stored in a text file. The output from the text file will be given as input for Agent Analysis and Hot Topic Analysis, where performance results of the agent will be generated and stored in the database.

V. RESULTS AND DISCUSSION

- ***Emotional analysis trained model performance***

Accuracy –

The accuracy of the model is 91.86%

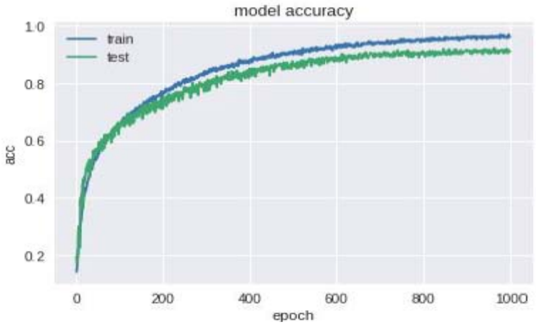


Figure 2. Emotional analysis accuracy graph

Loss –

The loss incurred by the trained model is 8.78%

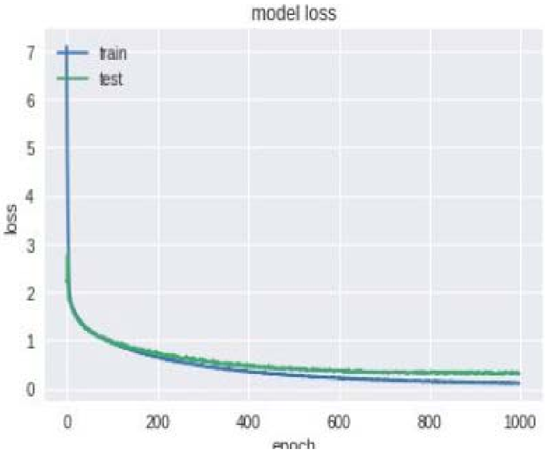
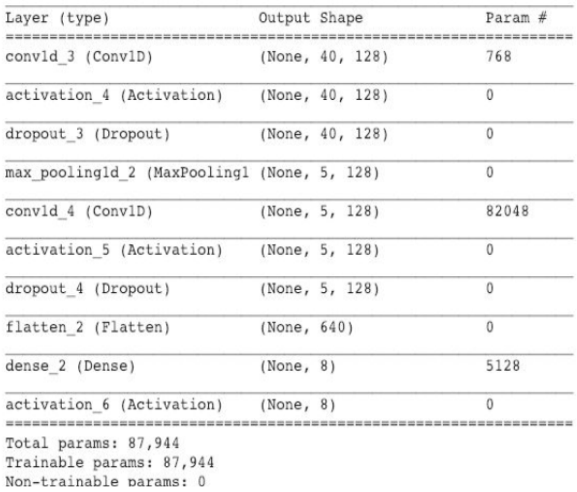


Figure 3. Emotional analysis loss graph

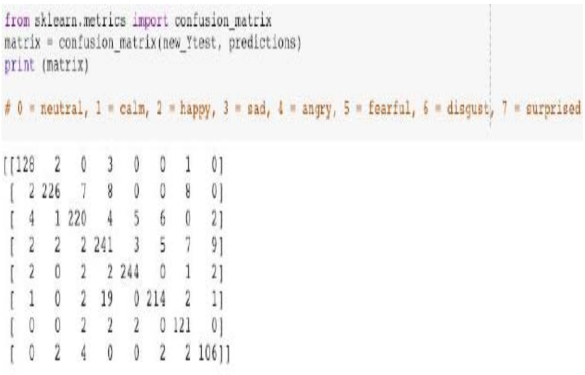
VI. MODEL SUMMARY



```
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_3 (Conv1D)            (None, 40, 128)           768
_________________________________________________________________
activation_4 (Activation)    (None, 40, 128)           0
_________________________________________________________________
dropout_3 (Dropout)          (None, 40, 128)           0
_________________________________________________________________
max_pooling1d_2 (MaxPooling1 (None, 5, 128)            0
_________________________________________________________________
conv1d_4 (Conv1D)            (None, 5, 128)            82048
_________________________________________________________________
activation_5 (Activation)    (None, 5, 128)            0
_________________________________________________________________
dropout_4 (Dropout)          (None, 5, 128)            0
_________________________________________________________________
flatten_2 (Flatten)          (None, 640)               0
_________________________________________________________________
dense_2 (Dense)              (None, 8)                 5128
_________________________________________________________________
activation_6 (Activation)    (None, 8)                 0
=================================================================
Total params: 87,944
Trainable params: 87,944
Non-trainable params: 0
```

Table I. CNN Model summary

- ***Confusion matrix of the model***



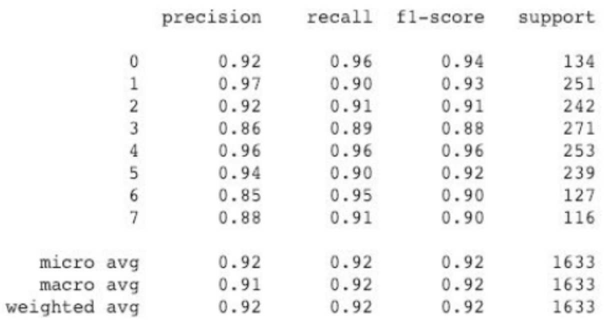
```
from sklearn.metrics import confusion_matrix
matrix = confusion_matrix(new_Ytest, predictions)
print (matrix)

# 0 = neutral, 1 = calm, 2 = happy, 3 = sad, 4 = angry, 5 = fearful, 6 = disgust, 7 = surprised
```

```
[[128   2   0   3   0   0   1   0]
 [  2 226   7   8   0   0   8   0]
 [  4   1 220   4   5   6   0   2]
 [  2   2   2 241   3   5   7   9]
 [  2   0   2   2 244   0   1   2]
 [  1   0   2  19   0 214   2   1]
 [  0   0   2   2   2   0 121   0]
 [  0   2   4   0   0   2   2 106]]
```

Fig. 5. Confusion matrix of Emotional analysis model

- ***Classification Report***



```
              precision    recall  f1-score   support

           0       0.92      0.96      0.94       134
           1       0.97      0.90      0.93       251
           2       0.92      0.91      0.91       242
           3       0.86      0.89      0.88       271
           4       0.96      0.96      0.96       253
           5       0.94      0.90      0.92       239
           6       0.85      0.95      0.90       127
           7       0.88      0.91      0.90       116

   micro avg       0.92      0.92      0.92      1633
   macro avg       0.91      0.92      0.92      1633
weighted avg       0.92      0.92      0.92      1633
```

TableII. Classification report of Emotional analysis

VII. CONCLUSION

With the help of our system, Customer-Agent call analysis can be done which will help call centres to present a reliable monitoring system resulting in accurate performance measurements by analyzing all the incoming and outgoing calls, reduce total call time, correctly route the customers to respective agents to solve their needs and increase customer satisfaction.

The future scope of our project can be enhanced by including voice activity detection which will help negate all the parts of the speech where there is no activity and also talk over analysis that will help to mine information directly from the speech.

VIII. REFERENCES

[1] BetülKarakus, Galip Aydin “Call Center Performance Evaluation Using BigData Analytics”, 2017.

[2] Baird H. “Ensuring Data Validity Maintaining Service Quality in the ContactCenter.”, Telecom Directions LLC, 2004, pp 3.

[3] SathitPrasomphan. “Detecting Human Emotion via Speech Recognition byusing Speech Spectrogram.” in IEEE International Conference on DataScience and Advanced Analytics (DSAA) Paris, France, IEEE 2015, pp 66-73.

[4] Margaret Lech, Melissa Stolar, Robert Bolia, Michael Skinner.“Amplitude-Frequency Analysis of Emotional Speech Using Transfer Learning and Classification of Spectrogram Images.” Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 4, pp. 363-371, 2018.