GRINREY

Advanced Research in Computer Engineering

Sandip A. Kale
Editor

Research Transcripts in Computer, Electrical and Electronics Engineering

Research Transcripts in

Computer, Electrical and Electronics Engineering

Volume 01

Advanced Research in Computer Engineering

Book Series by Grinrey Publications

- Research Transcripts in Energy
- Research Transcripts in Engineering
- Research Transcripts in Materials
- Research Transcripts in Computer, Electrical and Electronics Engineering

www.grinrey.com

Research Transcripts in

Computer, Electrical and Electronics Engineering

Volume 01

Advanced Research in Computer Engineering

Sandip A. Kale

Editor

GRINREY

Advanced Research in Computer Engineering

Sandip A. Kale, editor (Technology Research and Innovation Centre, Pune, India)

Series: Research Transcripts in Computer, Electrical and Electronics Engineering

ISBN: 978-81-948951-2-1

Published by: Grinrey Publications

511, Ganesh Nakashtram, Dhayari, Pune, Maharashtra, India

www.grinrey.com

Email: contact@grinrey.com

Contents

1 Enhanced Text Clustering Approach using Hierarchical Agglomerative Clustering with Principal Components Analysis to Design Document Recommendation System

Gauri Chaudhary[a,*] and Manali Kshirsagar[b]

[a]Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India
[b]Rajiv Gandhi College of Engineering and Research, Wanadongri, Nagpur, India

*Corresponding author: chaudhary_gauri@yahoo.com

ABSTRACT

Considering the increased usage and our increasing dependency in today's world on electronic data, substantial part of which is in textual form, it becomes necessary to devise scientific methods to infer and extract knowledge from such abundant electronic documents for strategic decision making in any target domain under consideration. The purpose of this study is to develop a common platform where all the similar text from multiple source documents from internet can be fetched and grouped using text mining and document clustering techniques. This chapter elaborates the method of hierarchical agglomerative text clustering approach to identify similar groups within documents. The method of Principal Components Analysis on text data is also further elaborated. Further combination of the two methods is proposed to find suitable clusters in text data and the results obtained show better quality clusters. For the purpose of experiments, plot summaries of movies from Wikipedia are used as the source document corpus. Various document pre-processing techniques are also explained and applied to the documents. The proposed method to get suitable clusters of similar movies can be used for recommendation to users. R programming is used for implementation of algorithms and visualization of the results.

Keywords: Data Mining, Hierarchical Agglomerative Clustering, Principal Components Analysis, Text Clustering

Advanced Research in Computer Engineering | ISBN: 978-81-948951-2-1
Book Series: Research Transcripts in Computer, Electrical and Electronics Engineering | Vol: 02

1. INTRODUCTION

With large abundance of electronic data available in today's era of internet, data mining becomes a very sought for field to extract information from such huge data for enhancing business and making business decisions [1]. A specialized field within data mining is text mining which focuses on mining information from textual data. Typical applications of text mining would be in the areas of document classification, sentiment analysis, document clustering, entity relation modeling and others [2].

Textual data is available in huge size in the form of web pages, digital libraries, articles, blogs, emails etc. In order to analyze this data, the data must be first converted into numeric form. Various pre-processing and data transformation techniques need to be applied to convert the textual data into equivalent numeric form. Data in the form of documents from various sources under the target study are collected. Document collection, also called Document corpus, is then subject to various pre-processing techniques. The focus of this study is to study and apply text mining techniques for clustering similar texts.

Clustering is one of the popular data mining functionalities that aims at identifying natural groupings within a data population based on similarity. Every data point is compared to every other data point in the population and based on some similarity measure chosen, they are assigned to groups. Document clustering is the application of clustering algorithms to collection of documents to derive meaningful clusters of documents [3]. This is a form of unsupervised learning since actual cluster labels in the data are not known in advance [4, 5].

The aim of this case study is provide a platform for choosing movie according to one's taste from plethora of options available and without having to search multiple sources. The platform should fetch similar movies based on any criterion like genres, artists, date etc. so that one can choose a movie of his taste almost immediately without having required searching multiple sources.

2. FOUNDATIONS

2.1 Text Mining

Text mining is an area which deals with processing of textual data which is unstructured like free flowing text on web pages, digital libraries, articles etc. to derive some underlying hidden knowledge [6]. These huge document collections which are not structured need to be cleaned and converted into a form which can be used for machine learning. Various pre-processing techniques mentioned in [1, 2, 6] that need to be applied to the documents are listed below:

2.2.1 Tokenization

Each document is subject first tokenization i.e. breaking of the document into collection of words or tokens. This is done typically using some delimiters such as whitespaces.

2.2.2 Remove Stop Words

Majority words in documents are actually noise i.e. words that may not really add any value to the analysis. For e.g. articles, prepositions, conjunctions etc. These are typically called stop words and need to be removed from the documents.

2.2.3 Stemming

Another major activity is stemming i.e. reducing the words to their root form. For e.g. words “running” or “ran” both have the same root word viz. “run”. So they are both replaced with the word “run”. Various stemming algorithms are available to reduce the words to their stems like Porter’.

2.2.4 TF-IDF Conversion

The documents, thus cleaned, need to be converted into numeric form such that they can be fed as an input to the mining algorithms. This can be done using TF-IDF (Term Frequency – Inverse Document Frequency) conversion. In this technique, each document is converted into a vector of words, where each word is represented by a number depicting its importance in the document collection. In the TF-IDF representation, the frequency of each word is normalized by its IDF.IDF reduces the weight of very frequent words that occur in maximal number of documents in the corpus thereby reducing the significance of commonly occurring words in the corpus. Words that occur frequently in a document but are rare across the document corpus are of interest as they would add value when comparing the documents for similarities in the clustering context. This method does so by assigning weighted scores to each word in the documents. For every document, each term is replaced by TF-IDF score explained in [7-9] which is calculated as follows (“doc” in the equations 1 and 2 indicates a document in the corpus):

$$\text{TF ("term")} = \frac{\text{frequency ("term") in a doc}}{\text{total no.of terms in the doc}} \tag{1}$$

$$\text{IDF ("term")} = \frac{\log(\text{total no.of docs})}{\text{no.of docs containing the term}} \tag{2}$$

$$TF\text{ - }IDF\ (\text{"term"}) = TF\ (\text{"term"}) * IDF\ (\text{"term"}) \tag{3}$$

2.2 Document Clustering

Document clustering aims at identifying segments within a document collection based on similarities [10]. It is a form of unsupervised learning since the groups within the documents are not known beforehand [11]. The model is trained based on the existing documents and its characteristics (typically words) so that groups within the data can be discovered based on similarities. There are various clustering approaches available. In this section we provide an overview of hierarchical clustering technique and Principal Components Analysis.

2.3 Hierarchical Clustering Technique

This technique starts with every document in a single cluster and goes on merging the clusters that are most similar till getting a single cluster. The clustering process and the result are displayed in the form of a tree or a "dendogram". The tree portrays the complete merging process showcasing the intermediate clusters at each level.

There are two different approaches within the hierarchical techniques: agglomerative and divisive. Agglomerative is a bottom-up approach that assumes every document in a single cluster at the bottom and goes on merging the clusters till single cluster at the top. Divisive approach is a top down approach starting with all data points in a single cluster at the top and then these are split on some similarity criterion recursively until every document is in separate cluster. In this section we describe the hierarchical agglomerative clustering (HAC) algorithm as in [12, 13, 14].

Hierarchical Agglomerative clustering algorithm:

i. Initially every document is treated as a separate cluster.
ii. Calculate a distance matrix which depicts the pair-wise similarities between clusters.
iii. Find the pair of clusters that are closest (most similar), remove the pair from matrix and merge them.
iv. Update the distance matrix to reflect the distances between the new cluster and other clusters.

Repeat steps (3) and (4) above until the distance matrix is reduced to a single element.

2.4 Principal Components Analysis

Principal Components Analysis is a method for reducing dimensions in such a way that the information loss is minimal [20, 21]. The basis of PCA is to identify the most significant of the components that capture maximum variance in the data [22]. One of the main objectives is to reduce the redundancy of the data.

Document term matrix used in text mining consists of rows as the document names and columns are the terms in the documents. This matrix is sparse due to the large number of terms. These terms are the dimensions in the case of document corpus and our objective is to reduce this number of terms and identify the most important or principal terms.

New set of terms is obtained by combining the old terms such that the new set of terms is either less than or equal to the number of old terms, indicate the maximum spread in the data and are representative of the original terms as they are derived from original terms. These new terms are called as principal components and are representative of more significant of the terms in the document corpus. Thus with PCA, the less significant terms of the corpus may be dropped thus reducing the dimensions for any text analysis.

3. RESEARCH METHODOLOGY

3.1 Data Collection And Transformation

Based on the motivation of getting movies according to one's taste, the dataset chosen for the purpose of this case study is reading Wikipedia pages of several movies of different genres. The document corpus was formed by adding the plot summaries of various movies from their Wikipedia pages. Movies from following 5 different genres were collected:

(1) Action thriller
(2) Comedy
(3) Animation
(4) Extra terrestrials
(5) Fantasy

The document corpus is shown in Fig. 1.

```
> summary(corpus)
                                Length Class             Mode
Alice in wonderland.txt         2      PlainTextDocument list
Alice through the looking glass.txt 2      PlainTextDocument list
Avengers Age of Ultron.txt      2      PlainTextDocument list
Avengers Infinity War.txt       2      PlainTextDocument list
Despicable Me.txt               2      PlainTextDocument list
Die Hard 2.txt                  2      PlainTextDocument list
Die Hard with a Vengeance.txt   2      PlainTextDocument list
Die Hard.txt                    2      PlainTextDocument list
Finding dory.txt                2      PlainTextDocument list
Finding Nemo.txt                2      PlainTextDocument list
Little big soldier.txt          2      PlainTextDocument list
Little White Lies.txt           2      PlainTextDocument list
The Avengers.txt                2      PlainTextDocument list
Trail of the pink panther.txt   2      PlainTextDocument list
```

Fig. 1. Document Corpus

Tokenization, stop words removal and stemming are performed and document term matrix is formed. In this matrix the rows indicate the documents and the columns are all the distinct words. Each cell in the matrix indicates the frequency of the word in the corresponding document. The document term matrix is shown in Fig. 2 . The sparsity is 90% which is high indicating that there are distinct terms in the corpus which are not common, meaning not present in all of the documents. This is good when we compare the documents for similarity in the clustering algorithms.

```
> inspect(dtm)
<<DocumentTermMatrix (documents: 14, terms: 2069)>>
Non-/sparse entries: 2980/25986
Sparsity           : 90%
Maximal term length: 15
Weighting          : term frequency (tf)
Sample             :
                                  Terms
Docs                               alice dory gruber loki marlin mcclane nemo queen stark thanos
  Alice in wonderland.txt             22    0      0    0      0       0    0    17     0      0
  Alice through the looking glass.txt 22    0      0    0      0       0    0     9     0      0
  Avengers Age of Ultron.txt           0    0      0    3      0       0    0     0    10      1
  Avengers Infinity War.txt            0    0      0    1      0       0    0     0     6     23
  Die Hard 2.txt                       0    0      1    0      0      21    0     0     0      0
  Die Hard with a Vengeance.txt        0    0      2    0      0      27    0     0     0      0
  Die Hard.txt                         0    0     22    0      0      30    0     0     0      0
  Finding dory.txt                     0   23      0    0      8       0   10     0     0      0
  Finding Nemo.txt                     0   14      0    0     17       0   15     0     0      0
  The Avengers.txt                     0    0      0   19      0       0    0     0    11      0
```

Fig. 2. Document Term Matrix

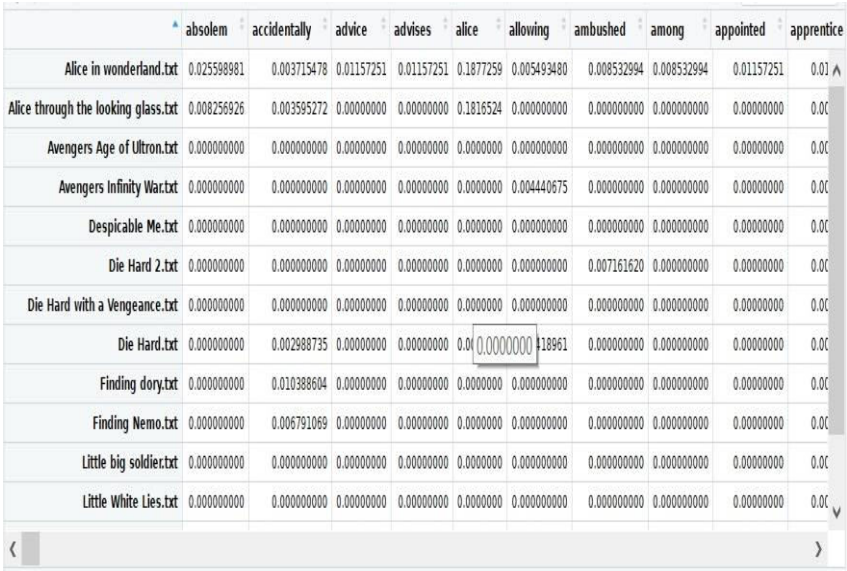
3.2 Conversion to vector space model

After cleaning the document using the pre-processing steps like tokenization, stop words removal and stemming, the corpus is converted to Vector Space Model (TF-IDF representation) as explained in [17]. TF-IDF value for every term in the document is calculated using equation (3) above. This document corpus is transformed into vector of values. Each document vector consists of TF-IDF value for each of the term in the document. Fig. 3 shows a portion of the TF-IDF representation for the documents.

3.3 Similarity Measure

Since clustering is based on grouping of similar documents, some similarity measure must be chosen to compare the documents. There are a number of possible measures for computing the similarity between documents. One of the most widely used is the cosine measure, which is explained in [8, 12, 17, 18] and defined as:

$$cosine(doc1, doc2) = \frac{doc1 * doc2}{|doc1||doc2|} \quad (4)$$



	absolem	accidentally	advice	advises	alice	allowing	ambushed	among	appointed	apprentice
Alice in wonderland.txt	0.025598981	0.003715478	0.01157251	0.01157251	0.1877259	0.005493480	0.008532994	0.008532994	0.01157251	0.01
Alice through the looking glass.txt	0.008256926	0.003595272	0.00000000	0.00000000	0.1816524	0.000000000	0.000000000	0.000000000	0.00000000	0.00
Avengers Age of Ultron.txt	0.000000000	0.000000000	0.00000000	0.00000000	0.0000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00
Avengers Infinity War.txt	0.000000000	0.000000000	0.00000000	0.00000000	0.0000000	0.004440675	0.000000000	0.000000000	0.00000000	0.00
Despicable Me.txt	0.000000000	0.000000000	0.00000000	0.00000000	0.0000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00
Die Hard 2.txt	0.000000000	0.000000000	0.00000000	0.00000000	0.0000000	0.000000000	0.007161620	0.000000000	0.00000000	0.00
Die Hard with a Vengeance.txt	0.000000000	0.000000000	0.00000000	0.00000000	0.0000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00
Die Hard.txt	0.000000000	0.002988735	0.00000000	0.00000000	0.0(0.0000000	18961	0.000000000	0.000000000	0.00000000	0.00
Finding dory.txt	0.000000000	0.010388604	0.00000000	0.00000000	0.0000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00
Finding Nemo.txt	0.000000000	0.006791069	0.00000000	0.00000000	0.0000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00
Little big soldier.txt	0.000000000	0.000000000	0.00000000	0.00000000	0.0000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00
Little White Lies.txt	0.000000000	0.000000000	0.00000000	0.00000000	0.0000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00

Fig. 3. TF-IDF representation of the document corpus

4. EXPERIMENTAL RESULTS

4.1 Optimal number of clusters

There are various methods to find the optimal number of clusters of which few popular ones are listed below:

4.1.1 Elbow method

In this method as explained in [19], the number of clusters is plot as a function of within cluster sum of square distances. The value of K is selected as the point in graph where there is noticeable decrease in sum of square distances i.e. position of a bend in the plot. Fig. 4 shows the plot of elbow method applied to the movie dataset under consideration for this case study. The formation of elbow can be seen at k= 5 or k =6.

4.1.2 Average Silhoute Criterion

Silhoute is a value indicating similarity of a document to its own cluster as against to other clusters. In this method, the average silhouette of documents for different values of K is plot. The desired value of K is the one for which the average silhoute value is maximum as explained in [15]. Fig. 5 shows the average Silhoute plot against the number of clusters. The possible values of good K from the plot could be 5, 7 and above.

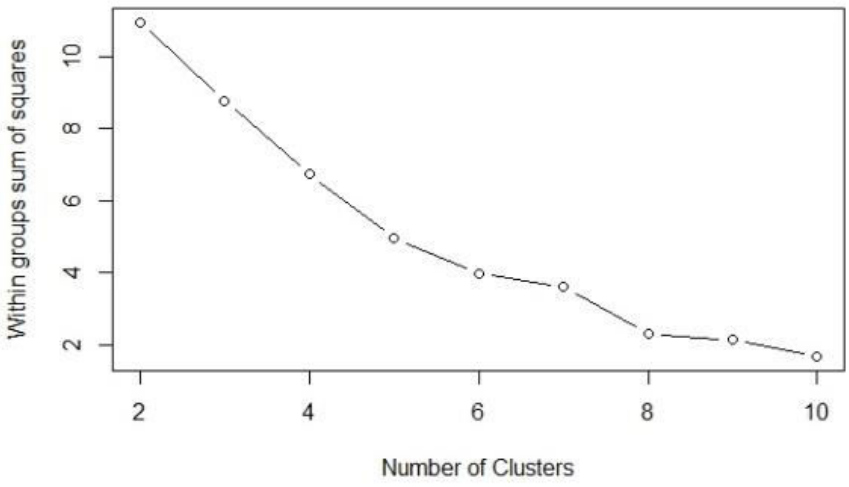


Fig. 4. Elbow method applied to the movie dataset

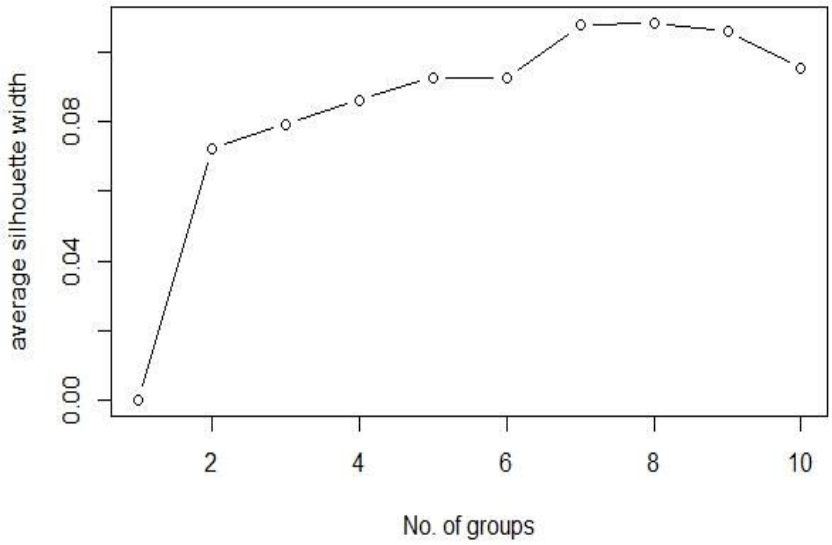


Fig. 5. Average Silhoutte Plot for the movie dataset

4.1.3 Calinski Harabasz

The Calinski Harabasz criterion is also based on low within cluster sum of squares and high between cluster sum of square distances as explained in [15]. The desirable value of K from the plot is the one that shows highest value for the Calinski Harabasz index. Fig. 6 shows the Calinski Harabasz index plot against the number of clusters for the movie dataset considered. Value of K = 5 and above look to be a good choice.

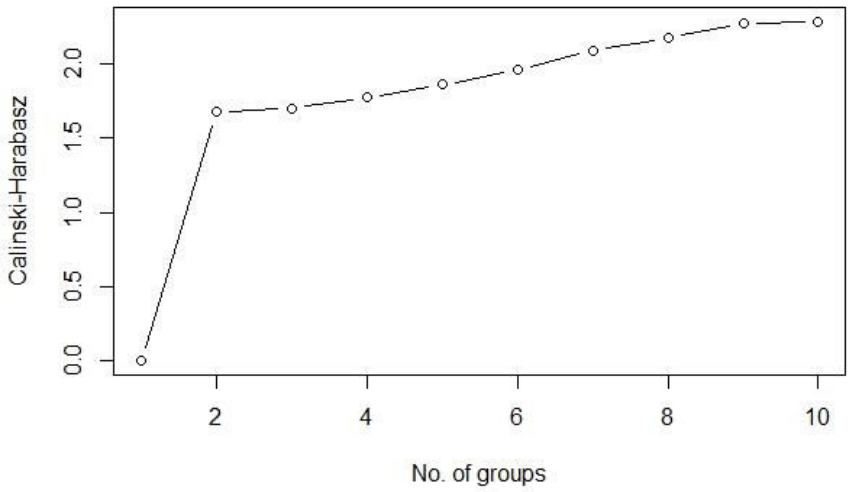


Fig. 6. Calinski Harabasz index for the movie dataset

Based on the above methods for selecting the optimal number of clusters, 5 or higher value is suitable.

4.2 Implementation of hierarchical clustering algorithm (HAC)

HAC is implemented by considering each document in one cluster to start with and then merging the clusters based on cosine similarity.

	Alice in wonderland.txt	Alice through the looking glass.txt	Avengers Age of Ultron.txt
Alice through the looking glass.txt	35.34119		
Avengers Age of Ultron.txt	48.51804	50.86256	
Avengers Infinity War.txt	53.38539	55.29014	45.16636
Despicable Me.txt	45.62894	47.50789	42.49706
Die Hard 2.txt	50.28916	52.95281	47.40253
Die Hard with a Vengeance.txt	58.96609	60.99180	56.50664
Die Hard.txt	58.08614	60.32412	55.91064
Finding dory.txt	50.47772	52.02884	47.83304
Finding Nemo.txt	51.83628	54.03702	49.44694
Little big soldier.txt	38.19686	41.20680	34.48188
Little White Lies.txt	37.38984	40.48456	33.79349
The Avengers.txt	50.87239	53.00000	37.89459
Trail of the pink panther.txt	45.16636	47.46578	42.21374

Fig. 7. Distance Matrix

Distance matrix is used to depict the similarity between two clusters whose ij^{th} element expresses the distance between the i^{th} and j^{th} cluster. Fig. 7 shows the distance matrix applied to the movie dataset.

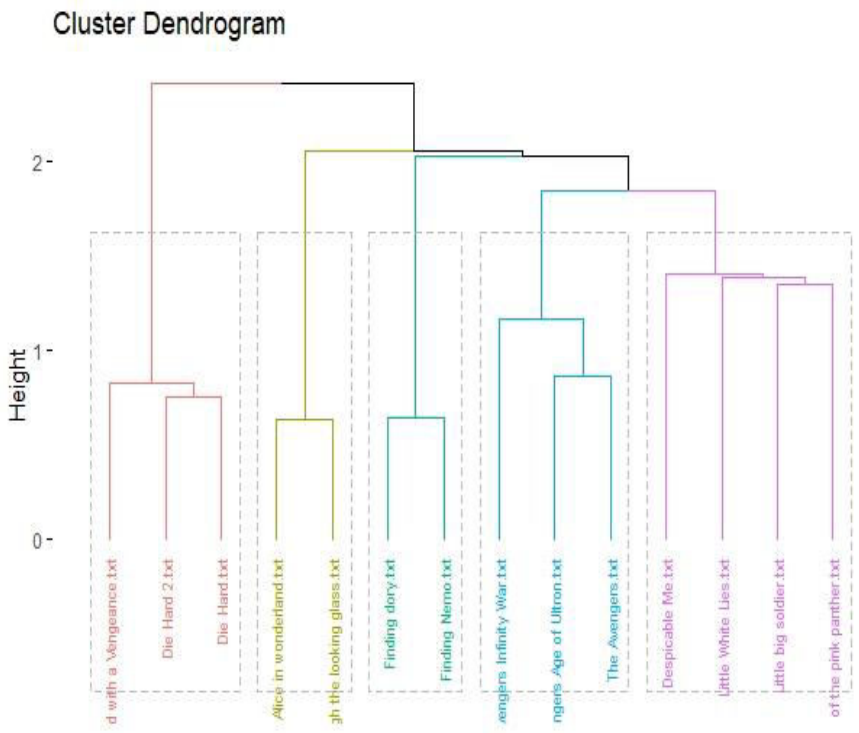


Fig. 8. Cluster dendogram using hierarchical agglomerative clustering

At each step the nodes are merged and the matrix is updated until process is complete. The output of this method is displayed in the form of a hierarchical structure called dendogram. Based on choice of optimal number of clusters, the dendogram can be cut at the desired level. Fig. 8 shows the application of hierarchical agglomerative clustering on the movie dataset resulting in the cluster dendogram.

4.3 Application of Principal Components Analysis

PCA is used to reduce the large number of variables, which in our case are terms, from the sparse document term matrix formed. The objective is to keep only the most significant terms that indicate the maximum spread in the data and use it for clustering implementation and analysis.

In order to obtain meaningful results from PCA, it is important to normalize the data first since it depends on the count of terms in a document but the length of documents is varying in the corpus. The number of principal components is chosen as (number of documents – 1) or (number of terms - 1) whichever is lesser. In our case, the number of terms is too large, so we choose the number of principal components as 13.

Fig. 9 shows the cumulative proportion of variance explained plotted against the Principal Components which increases sharply first and then gradually indicating that the maximum variance is depicted by the first few principal components. Fig. 10 indicates the proportion of variance explained against the principal components. It shows that after 5th component the proportion of variance explained is more or less steady. Hence we fixed the number of Principal Components as 5 for further analysis.

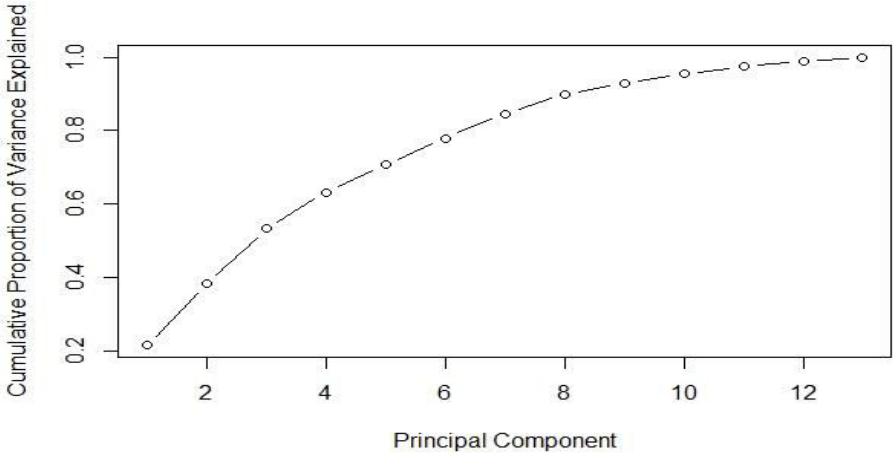


Fig. 9. Cumulative proportion of variance explained against Principal Components

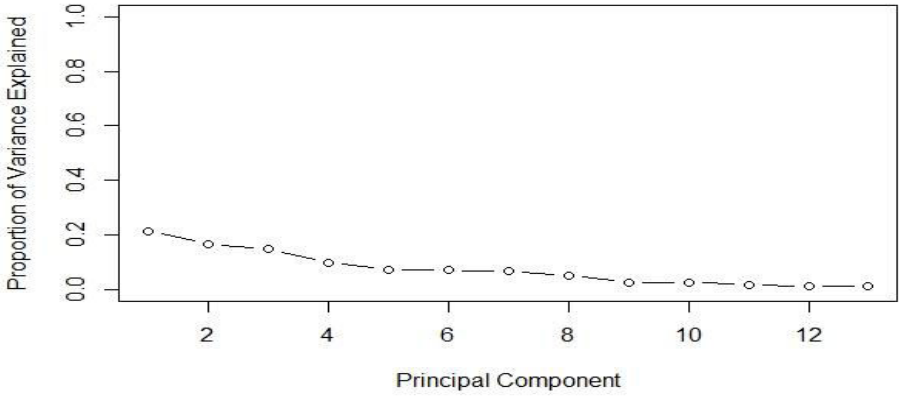


Fig. 10. Proportion of variance explained against Principal Components

Next we find which terms are the most significant in each component. Fig. 11 shows the contribution of each term in each Principal Component.

```
> tidied_pca
# A tibble: 10,680 x 3
   Tag          PC    Contribution
   <chr>        <chr>        <dbl>
 1 absolem      PC1       -0.00102
 2 accidentally PC1       -0.0243
 3 advice       PC1       -0.00114
 4 advises      PC1       -0.00114
 5 alice        PC1       -0.000656
 6 alices       PC1       -0.000656
 7 allowing     PC1        0.0251
 8 ambushed     PC1       -0.0170
 9 among        PC1       -0.0381
10 appointed    PC1       -0.00114
# ... with 10,670 more rows
```

Fig. 11. Contribution of terms in each Principal Component

Fig. 12 shows the graphical representation of contribution of terms in each of the 5 Principal Components.

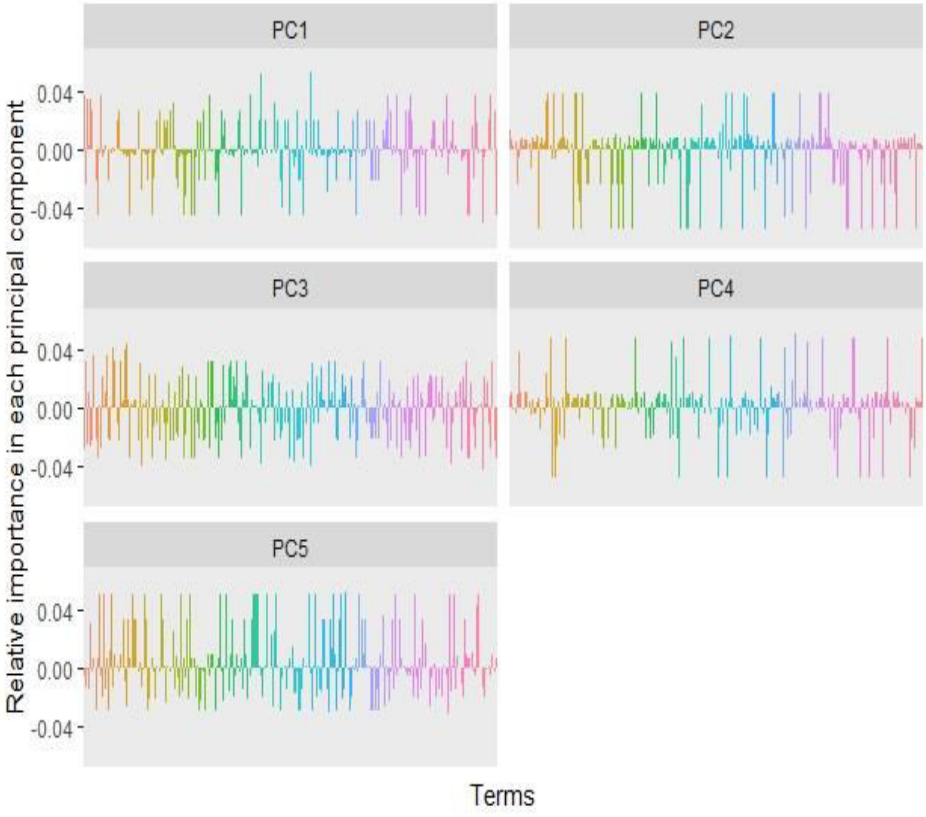


Fig. 12. Contribution of terms in all of the 5 Principal Components

We can zoom each component contribution to study in detail the highest contribution or highest absolute loadings of terms in each of the principal component. Fig. 13 depicts the contribution of terms in PC1 indicating highest positive as well as negative loadings on PC1. Fig. 14 shows the percentage variation as of the complete document corpus indicated by each principal component. As can be seen from Fig. 14, PC1 indicates the highest value at 21.58%, PC2 at 20.55% and so on.

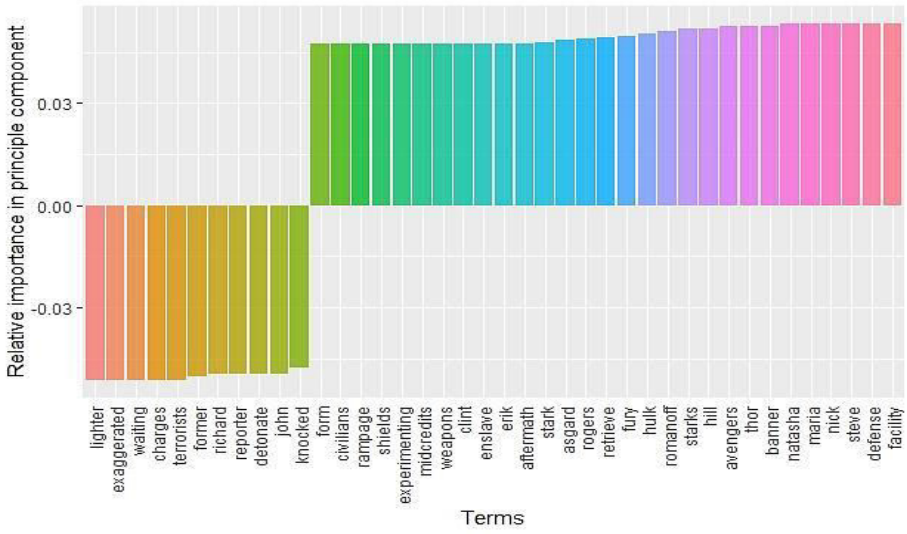


Fig. 13. Contribution of terms in PC1

```
> percent_variation
[1] 0.2168365 0.2055533 0.1982060 0.1913610 0.1880433
```

Fig. 14. Percentage Variation of the complete document corpus explained by each PC

4.4 Application of Hierarchical Agglomerative Clustering on first 2 Principal Components

PC1 and PC2 indicate the maximum variance in the entire document corpus as indicated in Fig. 14. Hence next we applied hierarchical agglomerative clustering on the first 2 principal components. Fig. 15 shows the dendogram obtained on application of HAC on PC1 and PC2.

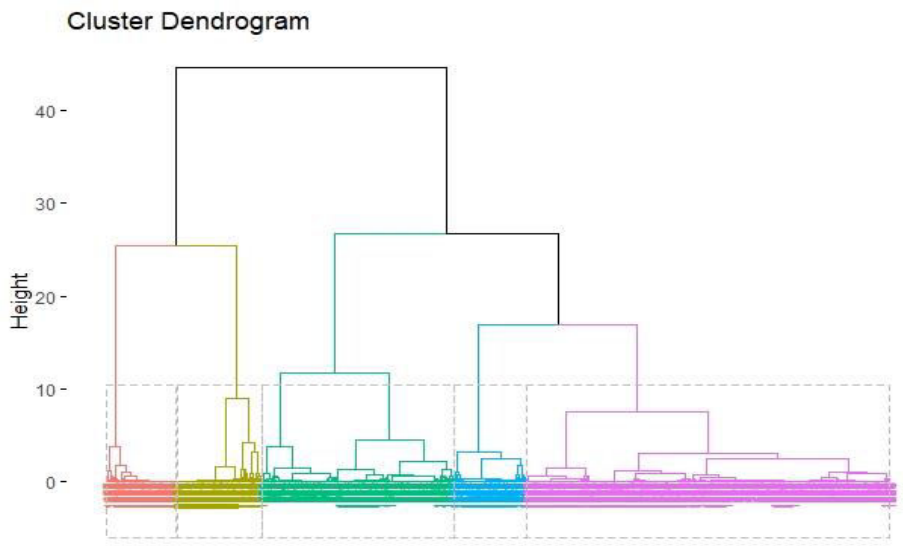


Fig. 15. Dendogram from application of HAC on principal components PC1 and PC2

5. EVALUATION

Silhouette coefficient measures how well an observation is clustered and it estimates the average distance between clusters (i.e. the average silhouette width). Observations with negative silhouette are probably placed in the wrong cluster. Fig. 16 shows the average silhouette width of 0.34 when HAC is applied to the complete document corpus.

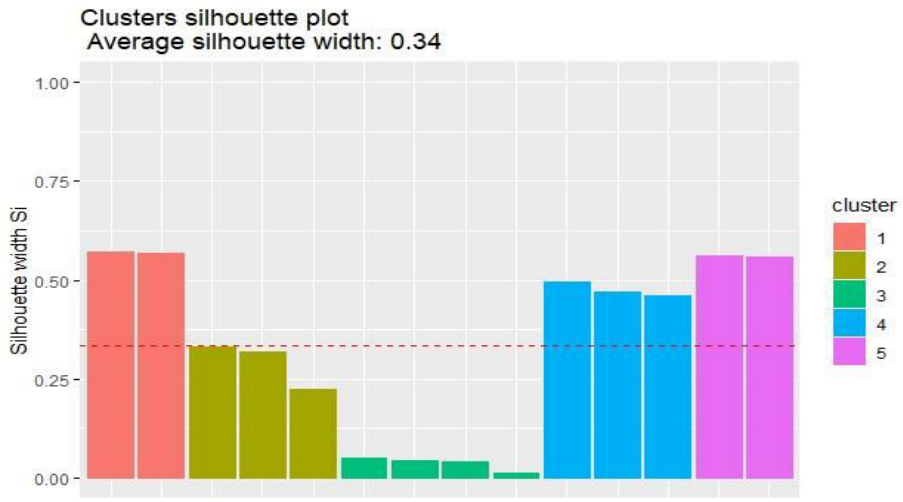


Fig. 16. Average silhouette width on application of HAC

Fig. 17 shows the average silhouette width of 0.68 when HAC is applied to the first 2 Principal Components obtained upon applying PCA to the document corpus first. As we can see the cluster quality is improved as an average silhouette width of 0.68 is obtained which is higher than that obtained on application of HAC without performing Principal Components Analysis.

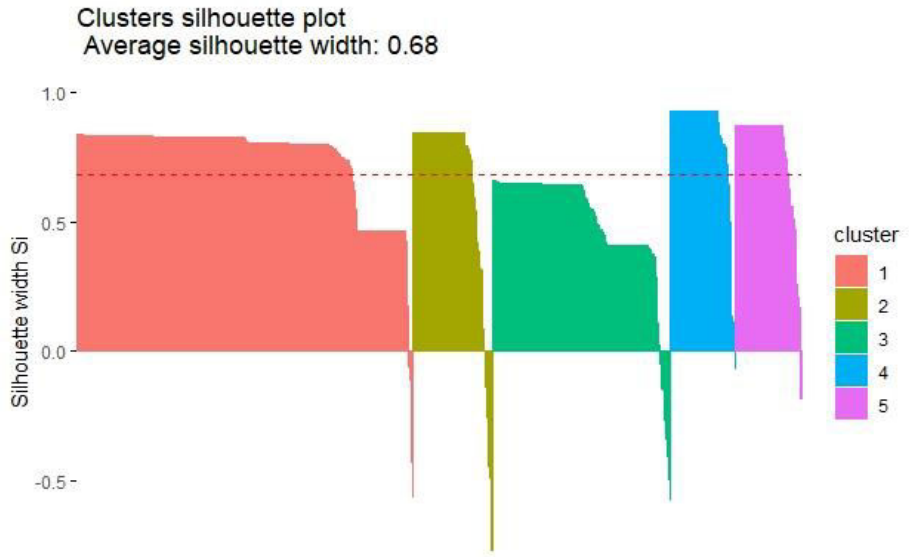


Fig. 17. Average silhouette width on application of HAC on Principal Components

6. CONCLUSION

Clustering approach for building a recommender system for movies to users based on document similarity has been discussed in this chapter. The data used was downloaded by crawling Wikipedia for movie plot summaries by different genres. Text mining methods have been applied to clean and pre-process the movie texts. Application of hierarchical agglomerative clustering and combination of hierarchical agglomerative clustering with Principal Components Analysis was studied on these move texts. When applying HAC, the optimal number of clusters has been determined using various available methods. In future, the same study can be carried out on larger dataset. Combination of other text mining and clustering approaches such as partitional approach, different similarity measures can be implemented to get still better quality clusters on larger dataset also. This can be then applied real time for building movie recommender system based on all online plot summaries available.

REFERENCES

1. G. Chaudhary, M. Kshirsagar, "Overview and application of text data pre-processing techniques for text mining on health news tweets,"*Helix*, vol. 8, issue 5,pp. 3764-3768, 2018
2. S. Vijayarani et al, "Preprocessing techniques for text mining – an overview," *International Journal of Computer Science & Communication Networks*, vol. 5, issue 1, pp. 7-16, 2015
3. Y. Wang, W. Xu, H. Jiang, "Using text mining and clustering to group research proposals for research project selection," *Proceedings of the 201548th Hawaii International Conference on System Sciences*, 2015, pp. 1256-1263.
4. X. Jiang, Y. Shi, S. Li, "Research of correction method in the feature space on text clustering," *Proceedings of the 2012 International Conference on Computer Science and Service System*, 2012, pp. 2030-2033.
5. K. Zhou, S. Yang, "Effect of cluster size distribution on clustering: a comparative study ofk-means and fuzzy c-means clustering," *Pattern Analysis and Applications*, Springer-Verlag London Ltd., 2019,pp. 1-12.
6. K. Rajput, N. Kandoi, "An ontology-based text-mining method to develop intelligent information system using cluster based approach," *Proceedings of the IEEE International Conference on Inventive Systems and Control*, Coimbatore, India, 2017, pp. 1-6.
7. G. Kou, Y. Peng, "A new hierarchical document clustering method," Proceedings of the 2009 Fifth IEEE International Joint Conference on INC, IMS and IDC, 2009, pp. 1789-1792.
8. X. Pan, J. Cheng, Y. Zia, X. Zhang, H. Wang, "Which feature is better? TF*IDF feature or topic feature in text clustering," *Proceedings of the 2012 Fourth IEEE International Conference on Multimedia Information Networking and Security*, 2012, pp. 425-428.
9. T. Pitchayaviwat, "A study on clustering customer suggestion on online social media about insurance services by using text mining techniques," *Proceedings of the 2016 IEEE Management and Innovation Technology International Conference (MITicon)*, Bang-San, Thailand, pp. 148-151.
10. Y. Zhu, B.C.M. Fung, D. Mu, Y. Li, "An efficient hybrid hierarchical document clustering method," Fifth International Conference on FuzzySystems and Knowledge Discovery, Shandong, China, 2008, pp. 395-399.
11. M. Carullo, E. Binaghi, I. Gallo, N. Lamberti, "Clustering of short commercial documents for the Web", *Proceedings of the 200819th International Conference on Pattern Recognition*, 2008, pp. 1-4.

12. M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," *Proceedings of the Text Mining Workshop*, KDD, 2000, pp. 1-20.
13. O. Gascuel, A. McKenzie "Performance analysis of hierarchical clustering algorithm", *Journal of Classification*, vol. 21, issue 1, pp. 3-18, 2004.
14. T. Jo, "String vector based AHC for text clustering," *Proceedings of the 2017 19th IEEE International Conference on Advanced Communication Technology (ICACT)*, Bongpyeong, South, 2017, pp. 673–678.
15. M.M.-T. Chiang, B. Mirkin, "Intelligent choice of the number of clusters in k-means clustering: An experimental study with different cluster spread," *Journal of Classification*, vol. 27, issue 1, pp. 3-40, 2009.
16. P. Hansen, E. Ngai, B. K. Cheung, N. Mladenovic, "Analysis of global k-means, an incremental heuristic for minimum sum-of-squares clustering," *Journal of Classification*, vol. 22, issue 2, pp. 287-310, 2005.
17. S. Yuan, G.Wenbin, "A text clustering algorithm based on simplified cluster hypothesis," *Proceedings of the 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA)*, Toronto, ON, Canada, 23-24 Dec. 2013, pp. 412-415.
18. Y.Li, C.Luo, S.M. Chung, "Text clustering with feature selection by using statistical data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, issue 5, pp. 641-652, May 2008.
19. D. Marutho, S.H.Handaka,E. Wijaya, Muljono, "The determination of cluster number at k-mean using elbow method and purity evaluation on headline news," *Proceedings of the 2018 IEEE International Seminar on Application for Technology of Information and Communication*, Semarang, Indonesia, 29 November 2018, pp. 533-538.
20. Xiaolin Xiao and Yicong Zhou, "Two-Dimensional Quaternion PCA and Sparse PCA", *IEEE Transactions on Neural Networks And Learning Systems*, Vol. 30, No. 7, pp. 2028-2041 July 2019
21. Xiaoxu Han, "Nonnegative Principal Component Analysis for Cancer Molecular Pattern Discovery," *IEEE/ACM Transactions on Computational Biology And Bioinformatics*, Vol. 7, No. 3, pp. 537-549, JULY-September 2010.
22. Jun Yan, Ning Liu, Shuicheng Yan, Qiang Yang, Weiguo (Patrick) Fan, Wei Wei, and Zheng Chen,"Trace-Oriented Feature Analysis for Large-Scale Text Data Dimension Reduction", *IEEE Transactions on Knowledge And Data Engineering*, Vol. 23, No. 7, pp. 1103-1116, July 2011

Cite this article

Gauri Chaudhary and Manali Kshirsagar, Enhanced text clustering approach using Hierarchical Agglomerative Clustering with Principal Components Analysis to design Document Recommendation System, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 1-18

2 Elucidation of Bangla Language Classification Using Neural Network Approach

Pritom Sarker[a, *] and Jannatul Ferdous[a], Nakib Aman Turzo[a]

[a]Computer Science & Engineering Department, Varendra University, Rajshahi, Bangladesh

*Corresponding author: me.pritom@gmail.com

ABSTRACT

Bangladesh has two principal languages called Sadu and Cholit. In the early times, Sadhu was operational and was composed of Sanskrit components but the current era has shifted to Cholit language, which is now being used most commonly. Sadhu was mostly used for formal documentation purposes and it is the need of the hour to translate them to Cholit language because it is more speaker friendly and can be easily understandable. Therefore, in this chapter efforts were done to familiarize the current era with the Sadhu language by creating software. Few sentences were selected and the final dataset was obtained by Principal Component Analysis (PCA). Python is used for different machine learning algorithms. Maximum work was done on Scikit Learn which is Term Frequency-Inverse Data Frequency (TF-DF) Vectorizer's class. The best performance was given by Neural Network with high precision. Speed was also anticipated and values were determined through graphs. The results showed that it translated all words from Sadhu to Cholit efficiently and in a well-oriented way. Therefore, Sadhu's complexity has been removed in this era.

Keywords: Machine learning, Neural network, Principal Component Analysis, Python, Term Frequency

Advanced Research in Computer Engineering | ISBN: 978-81-948951-2-1
Book Series: Research Transcripts in Computer, Electrical and Electronics Engineering | Vol: 02

1. INTRODUCTION

Sadhu language is an elaborated bygone register related to Bengali dialect, which was used most commonly used in the course of the Bengali Renaissance from the 19th to 20th century. It differs in vocabulary structure; verb form and its main components are Sanskrit and Tasama. Unlike Cholit language, Sadhu was utilized mostly for penmanship, which is indigenous but used in longhand and also in verbalized form. Diglossia is the main section under which the two languages lie. Most writings now come under the Cholit language. In some parts of Bangladesh like Chittagong is superficially identical to Cholit Bangla. In ancient times, the Sadhu language was used informal documentation but now it is outdated in the current situation. The origin of the Sadhu language in the literature was because of the intellectuals of Gour, which first used it. Due to this reason, it is called the Sadhu Gouriyo language. Cholit language used presently is the most speaker friendly language and is the most common one. It is the most common bond for communication and understanding for Bengali speakers. In this chapter, our focus is to distinguish between Bengali sentences whether it is Sadhu and Cholit. This effort will open doors for the conversion of Sadhu to Cholit and can translate both languages for the ease of any speaker. It may lead to the creation of software, which can detect whether a sentence is in Sadhu or Cholit and can translate it. The chief goal is to increase the recognition of our generation to ancient literature through these languages by interconversion.

2. LITERATURE REVIEW

Differences in grammar can be revealed by classifiers but cognition can't be determined by it. Five sortal classifiers are utilized by Cantonese than Mandai. 40% of nouns appear without classifier and sortal was taken by 18% of Cantonese and by 3% of Manian [1].

Components of an NP consist of just a classifier using semantic criteria to override their syntactic distributor in Mandarian and Cantonese [2].

A major part of Natural Language Processing is machine translation for the translation of language. Translation consists of three parts which are translation models, language models, and a decoder. For the translation of English to Hindi, a statistical machine translation system was developed. In a Linux environment, the model is developed by using this software [3].

Categorization of speech and language processing systems can be done for predefined linguistic information use and it is driven by data. It causes the use of methods of machine learning for extraction automatically and processes corresponding units of information which are indexed as appropriate. Hence,

ALISP (Automatic Language-Independent Speech Processing) was used to exploit an idea with a focus on speech processing [4].

In an ignition, it was depicted that context-free grammar was the main problem in many speech understanding systems. Computationally augmented phrase structure grammar is very demanding. Most efficient finite set grammar can't represent the relation of sentence meaning. It was elaborated by developing an APSG coupled with language analysis and then derived automatically. An efficient translation system was developed using it which was fast compared to others [5].

A paper discussed the integration of natural language and speech processing in Phi DM-Dialog and also kept into consideration a cost-based scheme of ambiguity resolution. With the help of an incremental pairing and generation algorithm, the simultaneous interpretation capability was made [6].

Conversion of language is the most difficult task and a case study was done for this purpose. This involves the conversion of clients system in proprietary language to programming language. There are many factors that were taken into account and affect language conversion at the automation level [7].

CJK Dictionary Publishing Society has launched an investigative project in 1996 for the development of sophisticated segmentation by collaborating with basic techniques for making an elaborated simplified Chinese and traditional Chinese database with 100% precision [8].

Speech to text conversion was done in few pieces of researches for people with impairments in hearing. Software was created which aids using phonetics to correct pronunciations. This helps in the potential of recognition in English hearing [9].

A generic method was introduced to translate to discretized Modern Standard Arabic from a written Egyptian colloquial sentence. It can easily be applied to other dialects of Arabic. A lexical acquisition was done for colloquial Arabic for the conversion of written Egyptian Arabic to MSA [10].

For recognition of two speakers in each of Spanish and English, a system was also developed. It was limited to 400 words. Analysis of language and speech recognition are coupled strongly by using the same language model [11].

In an effort by utilizing neural network conversion of text which was in Hindi to speech was done. It has many applications in the case of blind people and can also be used for educating students. A document containing Hindi text was used as input and a neural network was used for recognizing characters [12].

Grammatical errors have restrictions in variability and function in the historical periods of the English language. They become more productive in the 19th and 20th century along with major extensions in functions, range of lexical associations, and variants [13].

For conversion of Hindi text to speech in Java Swings, a Graphical User Interface was developed. Because it consists of a variety of languages spoken in different areas [14].

Progress was made recently in the synthesis of speech by developing speech synthesizers that have very high intelligibility but there are some problems like lack of naturalness and sound quality. But quality has reached a suitable level in many applications [15].

With embedded spelled letters there are also investigations done for recognition accuracy of speech. There are various methods proposed for localization of spelled letter segments, and to again classify them with a specialized letter recognizer [16].

For the translator software which offsets the absence of educational tools to some extent a development report was prepared because this tool is the need of hearing-impaired person for communication. For enhancing written language skills this tool can be utilized [17].

By using lexicon, rule-based and, data-driven techniques words can be converted to triplets with the aid of a software system which causes the conversion between graphemes and phonemes. In a hybrid system, a shotgun integrates these techniques and adds educational and linguistic information regarding graphemes and phonemes [18].

A speech-to-text engine was developed online in real time for the transfer of speech into written language and also it requires special techniques [19].

Translation dilemmas examination was done in a qualitative research. By the considering implications of the same problem's medium of spoken and written languages was critically challenged. The translation was centered and the way of dealing with issues raised by a representation would be the main concern for all researchers [20].

Sadhu and Cholit languages were also classified using machine learning techniques in which Linear Discriminant Analysis (LDA) had given the most precise and accurate results and has opened doors for further research by using different approaches [21].

Language identification of similar languages was done using Recurrent Neural Networks and its improved accuracy in recognizing different languages by combining word vector representation and Long Short-Term Memory [22].

No significant study has been found on the classification of Sadhu and Cholit language sentences and conversion. As a result, this study is conducted based upon similar works conducted in other languages.

3. METHODOLOGY

For amassing sentences related to Sadhu and Cholit data was gathered from ancient Literature books. From five significant literature books 2483 Sadhu sentences and from six important literature books 3508 Cholit sentences were collected for this purpose. The detailed procedure is depicted in Fig 1.

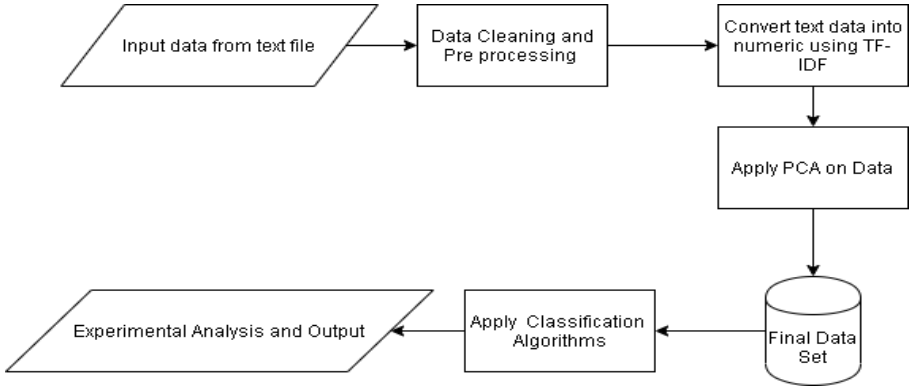


Fig. 1. Work Flow of the classification process

1. Data Clean

Non- English words are (before or after processing of natural language data it got filtered) in the data set. Extraction of all Non-English words was performed first. The natural language toolkit information center of Python is being used in this regard. The information set contains all non-English sentences. After going through the whole process 1983 data set was derived. For numeric categorization, Sadhu is termed as numeric 0 and Cholit as numeric 1.

2. Short Term Frequency–Inverse Data Frequency

An analytical statistic is a numerical or scientific form of statistic which is being contemplated to mirror the principle of a word in a docket or corpus and is called Short Term Frequency-Inverse Data Frequency (TF-IDF). This factor has weightage in retrieving information, text mining, and user modeling through the hunting of this data.

3. **Term Frequency (TF)**

Frequency of a word that pops up in a docket divided by the gross number of words in the document. Every document has its term frequency.

4. Inverse Data Frequency (IDF)

The log of the document numbers is divided by word w containing documents. Inverse data frequency determines the weight of rare words across all documents in the corpus.

Most work is being done from Scikit-Learn which is TF-DF Vectorizer's class. Text data was taken by it and converted to a numeric information set. After this conversion, the data has 3394 features. Features extraction was performed using PCA.

5. Principal Component Analysis

Through orthogonal linear transformation, data is converted to a new coordinate system so in an ordered manner each coordinate has the greatest variance by scalar projection

and so on hence, known as principal component analysis. The Principal component analysis is a section that comes under the class of Scikit-learn. First coordinate is known as the first principal component with higher variance in it and lower variance comes under the second coordinate. After application of principal component analysis, our information has 1678 traits. Data quality got lost when applications of dimensions of principal component analysis got reduced.

95% caliber of data was being maintained in the principal quality analysis case. By setting the value of 'n' components as 0.95 the amount of quality of real data preserved is 95%. 1678 characters remained in our data after the application of principal component analysis.

'তিনি সাহিত্যিক সুতরং টাকার কথা তুলিয়া তাঁহার সারস্বত সাধনার অমর্যাদা করিতে চাই না	Sadhu
'[হায় রে পোড়াকপালে সাহিত্যিক!] কিন্তু যদি তিনি নাম ধাম বদল করিয়া এই হাঁরা হরণের গল্পটা লিখিতে পরেন তাহা হইলে ত	Sadhu
'শ্রদ্ধা ও নমস্কার গ্রহণ করিবেন	Sadhu
'ওভারহেড প্রোজেক্টরের সুইসটি অফ করে তিনি হলভর্তি দর্শকদের দিকে তাকালেন	Cholit
'বিজ্ঞানীদের কনফারেন্সে বক্তব্য শেষ হবার পর সাধারণত ছোট একটি সৌজন্যমূলক করতালিদেয়া হয় কিন্তু এবারে একটি বি	Cholit
'এই সেশনটির সভাপতি সেন্টজন বিশ্ববিদ্যালয়ের বৃদ্ধ অধ্যাপক বব রিকার্ডো প্র মে করতালি দিতে শুরু করলেন এবং গ্যালারি	Cholit
'দেখতেদেখতে করতালির প্রচণ্ড শব্দে হলঘরটি ফেটে যাবার উপ্রম হল কিন্তু তবুও সেটি থেমে যাবারকোনো লক্ষণ দেখা গে	Cholit
'বিজ্ঞানীদেরকনফারেন্সে সাধারণত সাংবাদিকরা থাকেন না কিন্তু জেনেটিক ইঞ্জিনিয়ারদের এই বার্ষিককনফারেন্সে আলবার্তে	Cholit
'ফটো তোলা সম্পূর্ণ নিষিদ্ধহওয়া সত্ত্বেও সাংবাদিকদের ক্যামেরা ফ্ল্যাশ জ্বলতে শুরু করল এই ঐতিহাসিক মুহূর্তটি ধরেরাখার	Cholit
'বৃদ্ধ অধ্যাপক বব রিকার্ডো শেষ পর্যন্ত উঠে দাঁড়ালেন তাকে নির্দিষ্ট সময়ের মাঝে সেশনটিশেষ করার দায়িত্ব দেয়া হয়েছে	Cholit
'যদি এখনই তিনি নিয়ন্ত্রণটুকু হাতে না নিয়ে নেন সেটি সম্ভবহবার কথা নয়	Cholit

Fig. 2. Data Set before TF-IDF and PCA

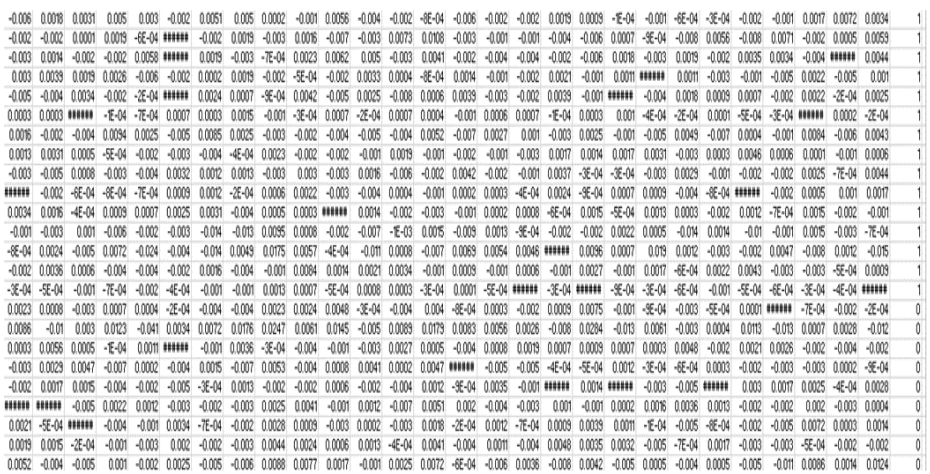


Fig. 3. Processed Data After TF-IDF and PCA

The processed data has 1042 different fields of numeric data in which the last field signifies 1 for the Cholit and 0 for the Sadhu.

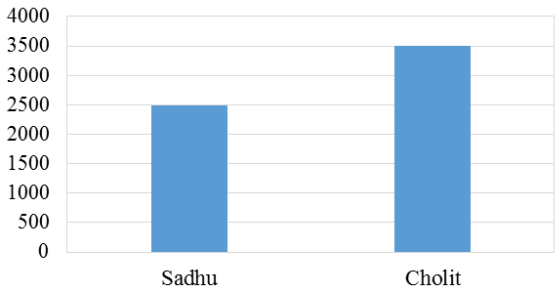


Fig. 4. Number of sentences used

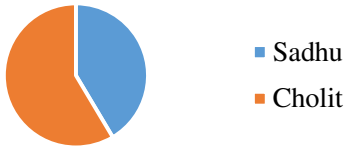


Fig. 5. Relative measurement of the number of sentences used

4. RESULTS AND EXPERIMENTAL ANALYSIS

Fifteen algorithms were used for classification. For this operation, scikit learn and Keras library was used and chosen eight best models best on the cross-validation score by doing it about 10 folds.

Table 1: Accuracy Chart of the top eight performers.

Classifier Name	**Accuracy (%)**
Logistic Regression	73.83
SVM (Linear)	74.04
Ridge Classifier	72.01
Linear discriminant analysis	75.07
AdaBoost	72.11
Standard NN Model	97.76
Smaller NN Model	98.16
Larger NN Model	98.75

Among all classifiers, Neural network functions best as depicted in the graph. In the case of LDA, it is expressed as a dependent variable as a linear combination of other features or measurements and has a resemblance with variance (ANOVA) and regression. As principal component analysis (PCA) and factor analysis, both look for linear combinations of variables which elaborate the data so LDA has also resembled them. When for each observation independent variables are continuous quantities, DA also works there. Discriminant correspondence analysis is an equivalent technique to categorical independent variables.

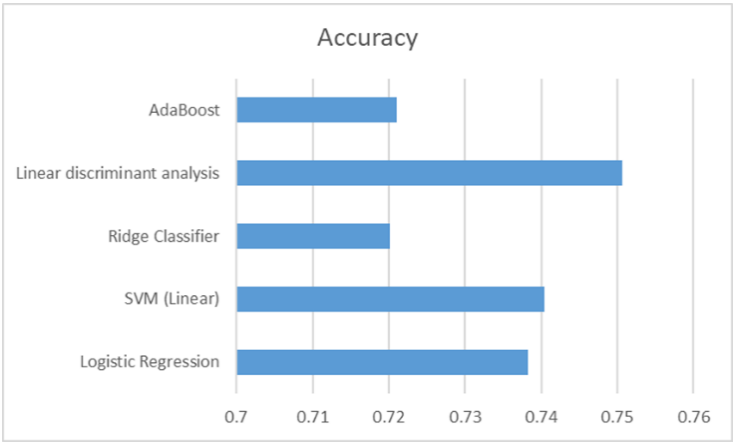


Fig. 6. Accuracy comparison of top 5 performers

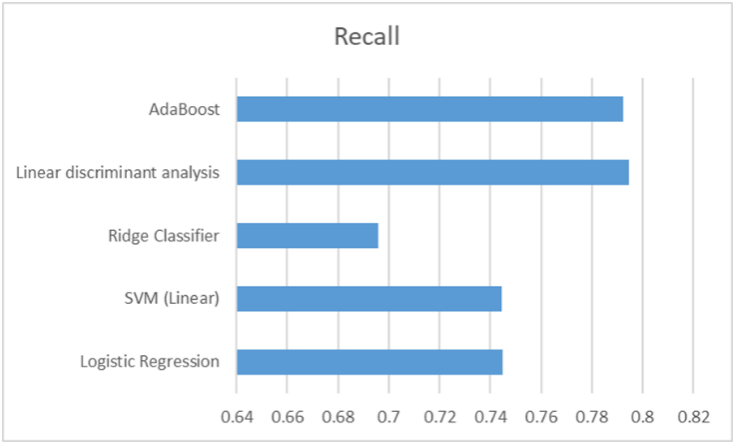


Fig. 7. Recall comparison of top 5 performers

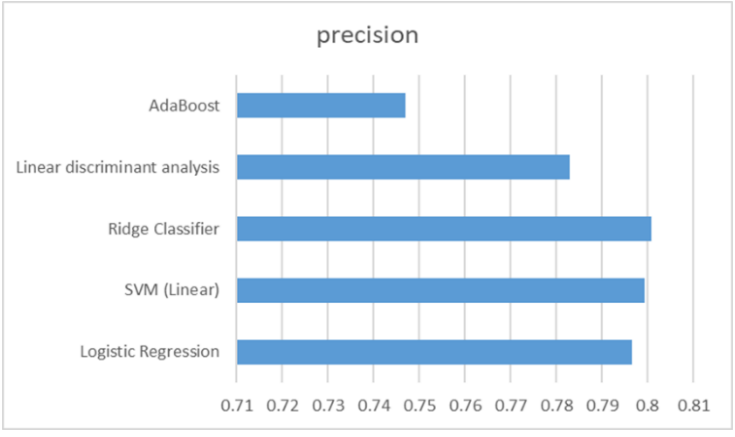


Fig. 8. Precision comparison of top 5 performers

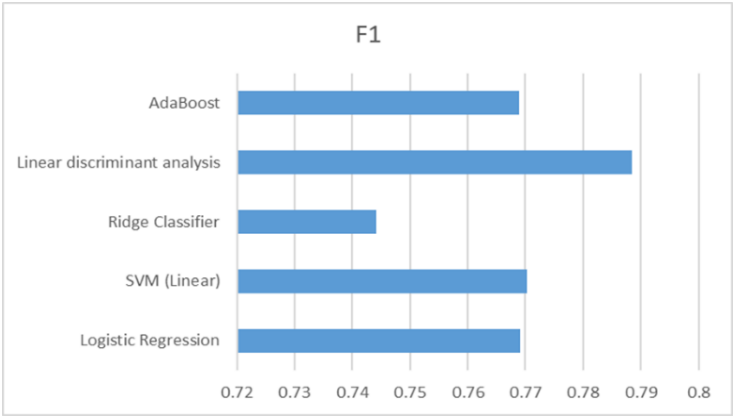


Fig. 9. F1 score comparison of top 5 performers

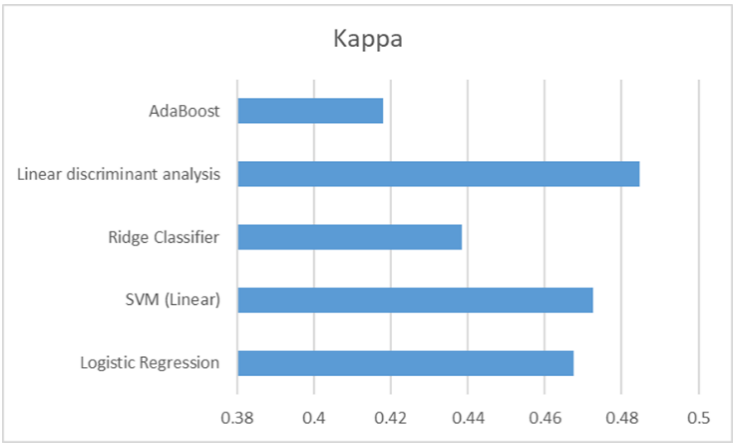


Fig. 10. Kappa comparison of top 5 performers

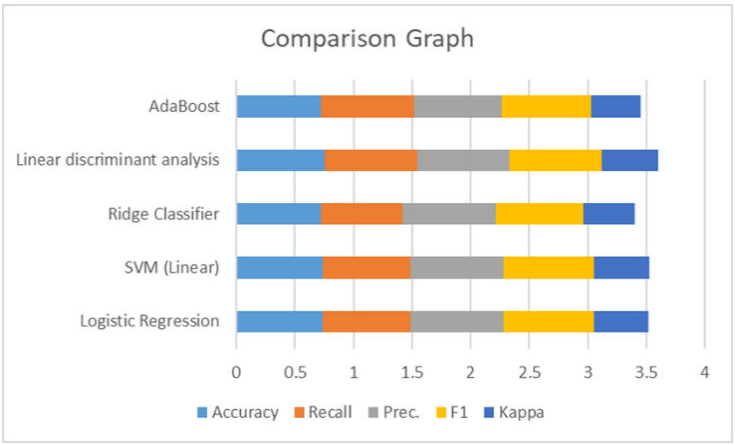
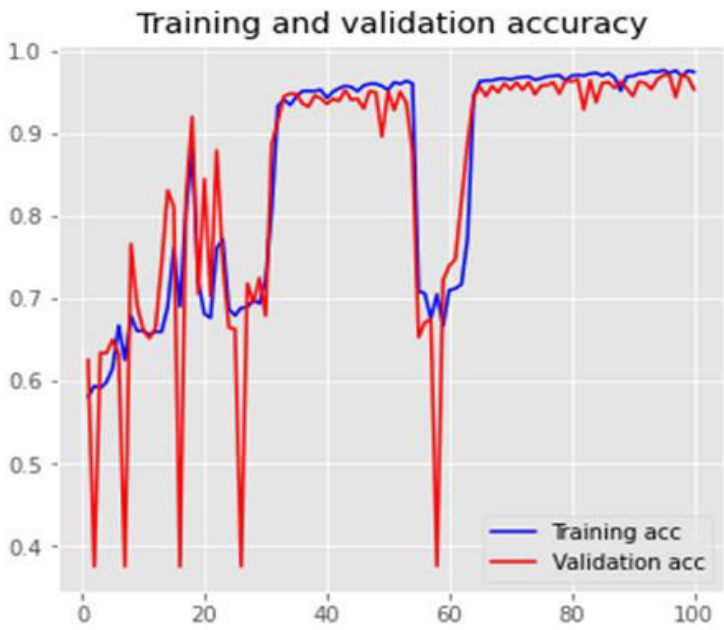
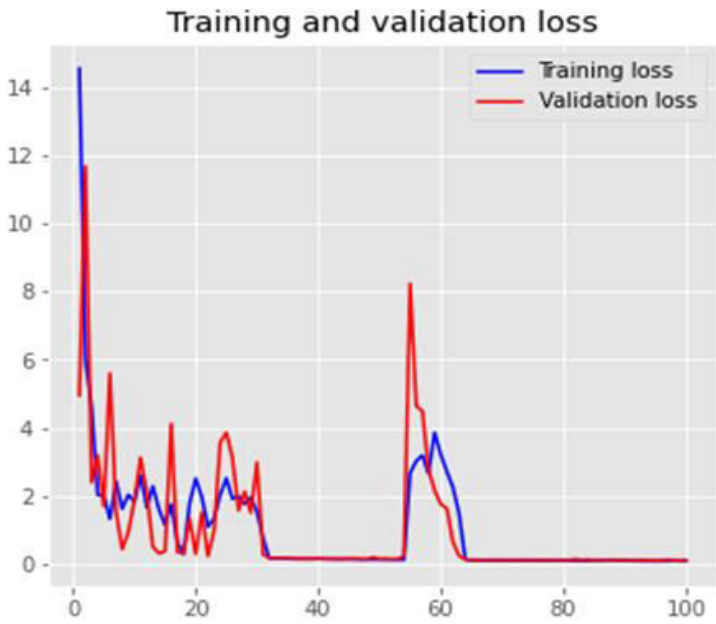


Fig. 11. Overall performance comparison of top 5 performers

4.1. Standard NN Model



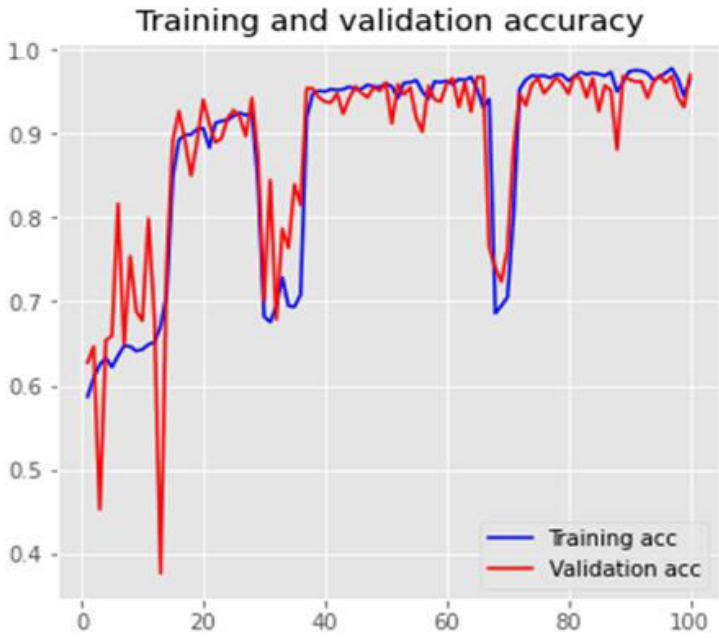
(a) Training Accuracy: 0.9776



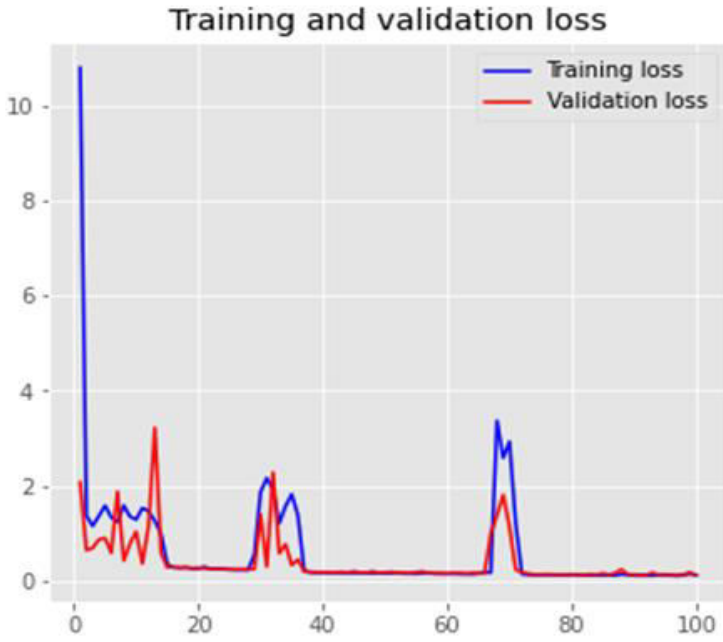
(b) Testing Accuracy: 0.9661

Fig. 12. (a) Accuracy curve using Standard Neural Network (b) Loss curve using Standard Neural Network

4.2. Smaller NN Model



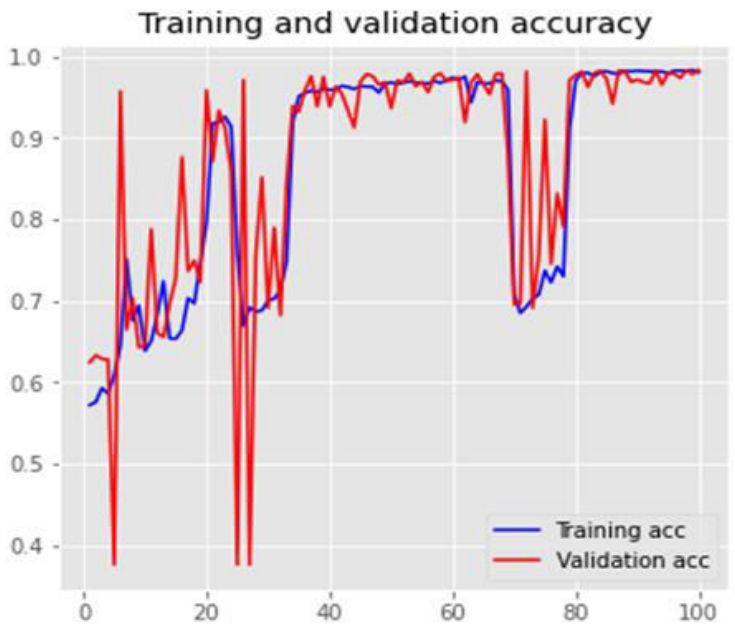
(c) Training Accuracy: 0.9816



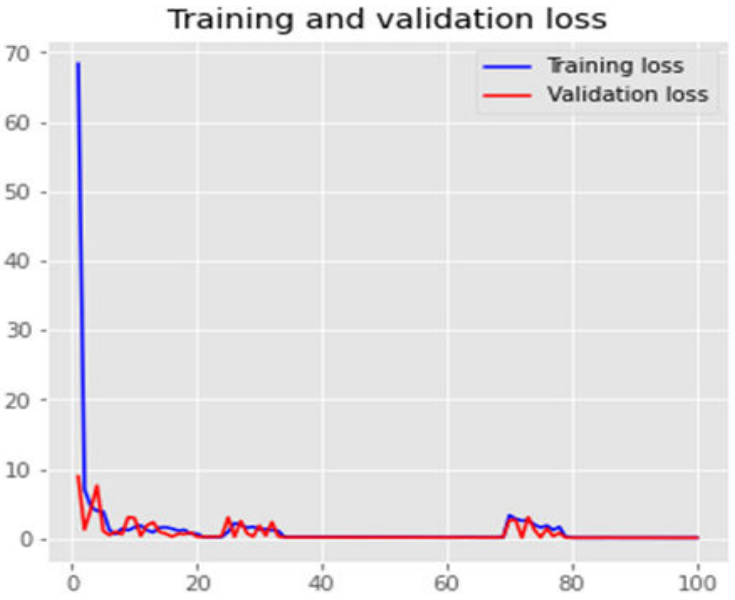
(d) Testing Accuracy: 0.9708

Fig. 13. (a) Accuracy curve using Smaller Neural Network (b) Loss curve using Smaller Neural Network

4.3. Larger NN Model



Training Accuracy: 0.9875



Testing Accuracy: 0.9840

Fig. 14. (a) Accuracy curve using Larger Neural Network (b) Loss curve using Larger Neural Network

5. CONCLUSION

By taking into account the whole process precise results were given by Neural Network in comparison to traditional machine learning approaches and gave 98.75% of accuracy than other models. The classifier can help in cataloging of Sadhu and Cholit languages of Bangladesh. As Bangladeshi literature is enriched with Sadhu language so most of the novels present from the ancient eras are in Sadhu language and this approach can help in their translation more accurately as it's highly uncommon for the present generation. That is why it can be converted to Cholit so that people can get familiar to old literature. RNN based model gave high accuracy than machine learning-based approaches. In future conversion will be tried to convey sadhu into Cholit so that the huge documents and literature can be realized.

REFERENCES

1. M. S. Erbaugh, "Classifiers are for specification: Complementary Functions for Sortal and General Classifiers in Cantonese and Mandarin," *Cahiers de Linguistique Asie Orientale,* vol. 31, no. 1, pp. 36-69, 2002.
2. S. Y. Killingley, Cantonese classifiers: Syntax and semantics, Newcastle upon Tyne: Grevatt & Grevatt, 1983.
3. N. V. p. S. Sharma, "English to Hindi Statistical Machine Translation System," *TIET Digital Repository,* 2 August 2011.
4. G. C. M. Petrovska-Delacrétaz, "Data Driven Approaches to Speech and Language Processing," in *Springer*, Heidelbergh, 2004.
5. D. Roe, F. Pereira, R. Sproat, M. Riley, P. Moreno, and A. Macarron, "Efficient grammar processing for a spoken language translation system," in *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, CA, USA, USA, 1992.
6. H. Kitano, "Phi DM-Dialog: an experimental speech-to-speech dialog translation system," vol. 24, no. 6, pp. 36-50, 1991.
7. A. Terekhov, "Automating language conversion: a case study (an extended abstract)," in *Proceedings IEEE International Conference on Software Maintenance. ICSM 2001*, Florence, Italy, Italy, 2001.
8. J. a. J. K. Halpern, "Pitfalls and Complexities of Chinese to Chinese Conversion." *international Unicode Conference*, Boston, 1999.
9. M. S. H. Nuzhat Atiqua Nafis, "Speech to Text Conversion in Real-time," *International journal of innovation and scientific research,* vol. 17, pp. 271-277, 2015.
10. H. A. K. S. a. I. Z. Bakr, "A hybrid approach for converting written Egyptian colloquial dialect into diacritized Arabic."," *international conference on informatics and system,* 2008.
11. RileyAlejandroMacarrón, "A spoken language translator for restricted-domain context-free languages" vol. 11, no. 2-3, pp. 311-319, 1992.

12. P. S. Rathod, "Script to speech conversion for Hindi language by using artificial neural network," in *2011 Nirma University International Conference on Engineering*, Ahmedabad, Gujarat, India, 2011.
13. B. Gray, "Grammatical change in the noun phrase: the influence of written language use," *Cambridge University Press,* vol. 15, no. 2, pp. 223-250, 2011.
14. K. a. R. K. Kamble, "A review: translation of the text to speech conversion for Hindi language."," *International Journal of Science and Research,* vol. 3, 2014.
15. N. a. K. A. Swetha, "Text-to-speech conversion," *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 2, no. 6, pp. 269-278, 2013.
16. A. W. Hermann Hild, "Integrating Spelling Into Spoken Dialogue Recognition," in *European Conference on Speech Communication and Technology*, Germany Carnegie Mellon University, USA, 1995.
17. B. Sarkar, K. Datta, C. D. Datta, D. Sarkar, S. J. Dutta, I. D. Roy, A. Paul, J. U. Molla and A. Paul, "A Translator for Bangla Text to Sign Language," in *2009 Annual IEEE India Conference*, India, 2009.
18. A. N. 1. a. J. Z. Merijn Beeksma 1, "shotgun: converting words into triplets: A hybrid approach to grapheme-phoneme conversion in Dutch," *John Benjamins,* vol. 19, no. 2, pp. 157-188, 2016.
19. P. Khilari, "A Review on Speech to Text Conversion Methods," *Computer Science,* 2015.
20. B. temple, "Qualitative Research and Translation Dilemmas," *Sage Journals,* vol. 4, no. 2, 2004.
21. P. S. B. K. Nakib Aman Turzo, "Interpretation of Sadhu into Cholit Bhasha by Cataloguing and Translation System," *International Journal of Trend in Scientific Research and Development,* vol. 4, no. 3, 2020.
22. M. R. a. M. S. Ermelinda Oro1, "Language Identification of Similar Language using Recurrent Neural Networks," in *10th International Conference on Agents and Artificial Intelligence*, 2018.

Cite this article

Pritom Sarker, Jannatul Ferdous and Nakib Aman, Elucidation of Bangla Language Classification Using Neural Network Approach, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 19-32

3 Facial Expression Recognition Using Neural Network

Md. Forhad Ali[a,*], Mehenag Khatun[a] and Nakib Aman Turzo[a]

[a]Computer Science and Engineering Department, Varendra University, Rajshahi, Bangladesh

*Corresponding author: cseforhad1997@gmail.com

ABSTRACT

Human emotions are states of mental health that resolve spontaneously rather than through conscious exertion, and are accompanied by physiological changes in the facial muscles that signify expressions. Nonverbal communication methods such as expressions, eye movements, and gestures are used in many applications of human-computer interaction. Identifying emotions is not an easy task because there is no difference between the emotions of a face, and there is also a lot of complexity and variability. The machine learning algorithm uses some open features to model the face. In this work, convolutional neural networks (CNNs) were developed to identify the expression of facial emotions. Facial expressions play an important role in the nonverbal communication that takes place in a person's inner emotions that are reflected on his or her face.This work has been used the Viola-Jones algorithm to detect the eye and lips region from a face and then with the help of the neural network. Also, Machine Learning techniques, Deep Learning models, and Neural Network algorithms are used for emotion recognition. This work will be proposed as an effective way to detect anger, contempt, disgust, fear, happiness, sadness, and surprise.

Keywords: Emotion Recognition, Emotions, Feature Extraction, Neural Network, Viola-Jones

Advanced Research in Computer Engineering | ISBN: 978-81-948951-2-1
Book Series: Research Transcripts in Computer, Electrical and Electronics Engineering | Vol: 02

1. INTRODUCTION

Facial expressions play a key role in understanding and recognizing emotions. Even the term “interface” suggests the importance of the face in communication between two entities. Studies have shown that reading facial expressions can dramatically alter the interpretation of what is being said and control the flow of conversation. A person's ability to interpret their emotions is very important for effective communication. The proportion of up to 93% of communication used in a normal conversation depends on the emotion of an entity. For ideal human-machine interfaces (HCI), it would be desirable for machines to be able to read human emotions. This research focuses on how computers can correctly detect the emotions of their various sensors. This experience was used as a face image as a means of reading human emotions. Research on human emotions dates back to Darwin's pioneering work and has since attracted many researchers to the field. Seven basic emotions are universal for humans. Namely, neutral, angry, disgusted, fearful, happy, sad and surprised, and these basic emotions can be identified from a person's facial expression. This study suggests an effective way to identify these four emotions using the neutral, happy, sad and surprising frontal facial emotions. Various methods of recognizing emotions have been proposed in recent decades. Many algorithms have been proposed to develop system applications capable of very well detecting emotions. Computer applications could communicate better by altering reactions in various interactions depending on the emotional state of human users. A person's emotions can be determined by the tongue, face, or even gesture. The work presented in this article examines the recognition of facial expressions.

For facial emotion recognition, the traditional approaches usually consider a face image that is distinguished from an information picture, and facial segments or milestones are recognized from the face districts. After that, different spatial and worldly highlights are separated from these facial segments. At last dependent on the separated highlights a classifier, for example, Keras library, random forest, is trained to produce recognitions results. This work is an applied, deep learning model. Deep learning is a well-set model in the pattern recognition domain. It uses a Convolutional Neural Network (CNN) algorithm using Keras library. CNN is a specific sort of artificial neural network that uses a machine-learning unit. CNN applies to objects detections, face recognition, image processing, etc. Deep convolutional neural network (DCNN) composition of many neural network layers. Which is also can be able to extract significant features from the data.

2. LITERATURE REVIEW

In a research field of emotion detection, there is a contribution of several domains like machine learning, natural language, neuroscience, etc. In previous works, they individually rummaged facial expressions, voice features, and textual data as universal indicators of emotions. Emotion can be classified into several static classifications like happiness, sadness, disgust, anger, fear, and surprise. In later works are improved by combining the image, voice, and textual data. The fusion of this data gives the maximum accurate result. This type of fusion can be done in three ways early, late, or hybrid. Other ethos features the elements of emotion and the collaborations between emotional processes and other intellectual procedures.

2.1. Emotion Detections Through Facial Feature Recognition [5]

This work deals with the emotion recognition with the Machine learning using support vector machine (SVM). Some principles are work to detection, extraction, and evaluation of facial expressions of image. These are:
i) Viola-Jones cascade object detectors and Harris Corner key points to extract faces and facial features from images.
ii) Histogram Of Oriented Gradient (H OG) Feature Extraction.
iii) Support vector machines (SVM) to train a multi-class predictor to classify the seven basic human facial expressions such as: (anger, contempt, disgust, fear, happiness, sadness, surprise).

Computers can easily recognize facial expressions and discover a person's motive, including entertainment, social media, content analysis, criminal justice, and healthcare. Here we mainly discuss two approaches such as: (Zhang approach and Gabor wavelet coefficients). Zhang has shown that a lower resolution (64x64) is sufficient. We're going to change the size of the extracted areas to 100x100 pixels.
If this use only the HOG and SVM classifier, the detection accuracy is 81%, which is much better than that of a fisherman. Only approach. When using the double classifier method, the accuracy is only 81% that of HOG, but the testing process is 20% faster.

2.2. SVM Point-based Real-time Emotion Detection [2]

This thesis deals with the detection of emotions in machine learning using a cascade of a multi-class support vector machine (SVM) and a binary SVM. This algorithm is designed to extract emotions based on the movement of 19 characteristic points. These characteristic points are located in different areas of the face such as the mouth, eyes, eyebrows and nose. He mainly works on rigid,

unchangeable points in the nose. Its division into face recognition and action unit (AU). Computers can easily recognize facial expressions and discover a person's subject, including entertainment, social media, content analysis, criminal justice, and healthcare.

One final suggestion for improvement is the fact that the user in the real-time app should keep the same distance from the camera from which the neutral frame was captured. Otherwise the theory behind the displacement reports is no longer valid. Rescaling neutral distances based on user movements can be a solution to this problem.

3. RESEARCH METHODOLOGY

This work considers the leading challenge faced by machine learning and the entire system is the training part. Where the system has to train by using real data of human face reactions. For example, if the system has to detect an angry face then the first system has to be acquainted with the angry face. Also if the system has to detect a happy face then the first system has to be acquainted with the happy face. To antecedents the system with this emotion types, the re-training process has been used. The re-training data were collected from the real world. The hardest part of this system was the re-training part. There are also many other parts of the system. Machine learning is a strong tool that enables data analysis of large databases more proficiently and fleetly. This enables the capability of detection emotion more accurate. It gives feedback in real-time. The system did not wait for the result for the future, not the image has to be stored. With help of modern-day computers, neoteric data mining techniques can analyze thousands of data within a very short amount of time saving lots of hours. Besides, using and installing such programs costs significantly less. If properly optimized these data mining techniques can give perfect outcomes than a human. This work resented a general and feasible framework for emotion data mining to identify emotion patterns using machine learning. This chapter proposed the program based on the Deep learning model and computer vision emotion recognition. This proposed method uses the CNN algorithm for this chapter. This proposed a more advanced method than the one that recognized only seven emotions with CNN. Their emotion recognition method using deep learning followed four steps, as follows.

(1) Public face database training with CNN

(2) Extract seven probabilities for each face frame.

(3) Aggregation of frame probabilities into fixed length image descriptors for each image in the data set.

(4) Classification of all images using a support vector machine (SVM) trained on image descriptors from the competition training set.

3.1 Emotion Database

In the data collection steps, this is used both in real-world media and online media to collect as much data as that could. Real-world includes different types of emotional pictures of friends and family members, relatives, some known unknown people's different kinds of facial expressions. They culled data was initially stored for future analysis. From online media, the data is collected data set from kaggle.com. This site uploaded this data set 6years ago. This site most trusted data set of emotions. This converted the data into 48×48 pixel grayscale images of faces. It contains two sections pixels and feelings. The feeling section contains a numeric code which runs from 0 to 6 in figure 1. What's more, the pixel section contains a string incorporated in statements for each picture. Furthermore, the picture should be only the picture of a face. So the collected pictures are resized and cropped picture of a face.

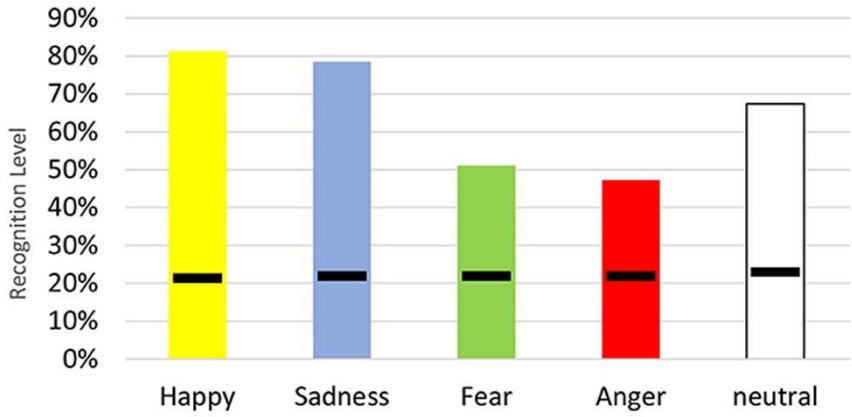


Fig. 1. Level of Emotion

3.2 Training phase using deep learning

A great way to use deep learning to categorize images is to create a convolutional neural network (CNN). The Keras library in Python makes it easy to create a CNN. Computers display images with pixels. Pixels in images are usually related. For example, in figure 2, a particular group of pixels may represent an edge in an image or other pattern. Convolutions use it to identify images. A convolution multiplies a pixel matrix with a filter matrix or a "kernel" and sums the multiplication values. Then the convolution slides to the next pixel and repeats the same process until all pixels in the image are covered. This process is shown below.

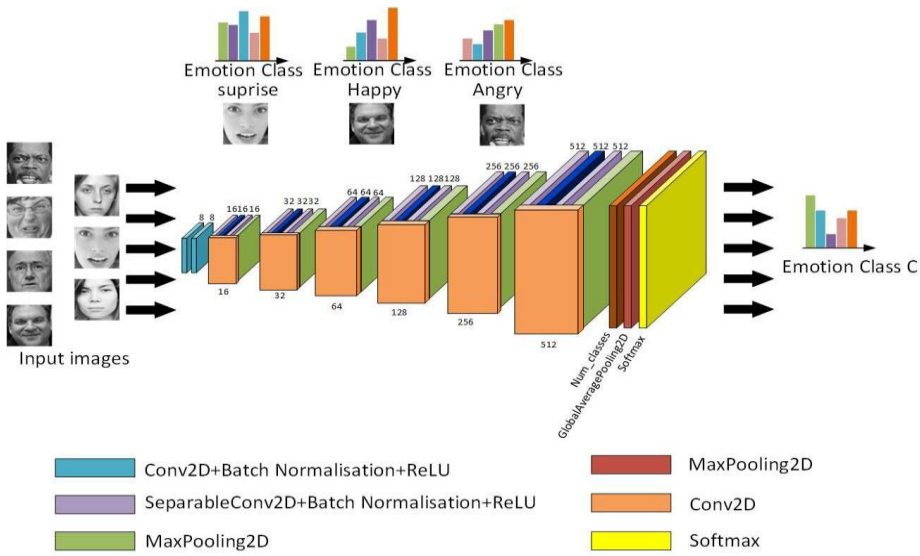


Fig. 2. Emotion detection using Convolutional Neural Network

The type of model we are going to use is Sequential. Sequential is the easiest way to create a template in Keras. You can create a model layer by layer. We use the "add ()" function to add layers to our model. Our first 2 layers are Conv2D layers. These are convolutional layers that process our input images, which are considered two-dimensional matrices. 64 in the first layer and 32 in the second layer are the number of nodes in each layer. This number can be higher or lower depending on the size of the dataset. In our case, 64 and 32 are working fine, so we'll stick with that for now. The kernel size is the size of the filter matrix for our convolution. So a kernel size of 3 means we have a 3x3 filter matrix. You will find a reminder on this subject in the introduction and in the first photo. Activation is that function. The activation function that we use for our first two layers is the ReLU, or rectified linear activation. It has been shown that this activation function works well in neural networks. Our first shift also has a registration form. This is the shape of each input image, 28,28,1 as seen earlier, where 1 means the images are in grayscale. There is a “flatten” layer between the Conv2D layers and the dense layer. Flattening acts as a link between folding and dense layers. The model will then make its prediction based on the most likely option. Then you need to compile this template. Compiling the model requires three parameters: optimizer, loss, and metrics. The optimizer controls the learning rate. It will use "Adam" as an optimizer. Adam is a good optimizer in many cases. The Adam Optimizer adjusts the learning rate during training. The learning rate determines how quickly the optimal weights for the model are calculated. A slower learning rate can result

in more accurate weights (up to a point), but the time required to calculate the weights is longer. We will use a "categorical cross-entropy" for our loss function. This is the most common classification option. A lower score indicates that the model is performing better. To make interpretation even easier, the "Precision" metric is used to display the precision score on the validation set while training the model. To train, he uses the "fit ()" function for his model with the following parameters: training data (train_X), target data (train_y), validation data and number of epochs. The test set supplied with the data record is used for the validation data and is divided into X_test and y_test. The number of epochs indicates how often the model scans the data. The more eras we run, the better the model gets up to a point. After this point, the model stops improving with each epoch. For our model we set the number of epochs to 3. After 3 epochs, it achieved an accuracy of 93% for this validation set.

3.3 Detection

K-means clustering was used with the number of clusters assumed to be two. Here, the maximum value is determined in all rows and its average is determined. Likewise, the minimum value is found in all rows and its average is determined. Using these two points as a basis, pixel values closer to the maximum average are grouped into one cluster, and pixel values closer to the minimum average are grouped into another cluster. Based on the result of clustering, the total number of components in the image is calculated. Based on the number of components, the person's eyes are first segmented using the bounding box feature. Because the eye or eyebrow forms the first element when pixel values are passed in columns, the eyes are segmented first. Using the ocular matrix, other parts of the face are segmented using a distance-based algorithm. The resulting image after running k means grouping for different expressions are shown in figure 3.

Fig. 3. K-means clustering segmentation outputs

The Viola Jones algorithm is a widely used mechanism for object recognition. The main characteristic of this algorithm is that training is slow but detection is fast. This algorithm uses the basic hair function. Hair features are the relevant features for face recognition.There are various types of features such as:

Fig. 4. Edge Features

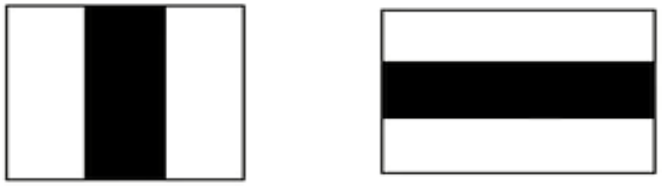


Fig. 5. Line Features

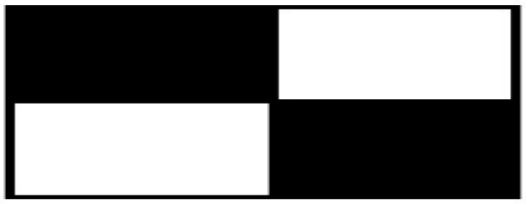


Fig. 6. Four Rectangle Features

For example, we need face detection of a person then we need at first image conversion in grayscale then second step image segmentation.

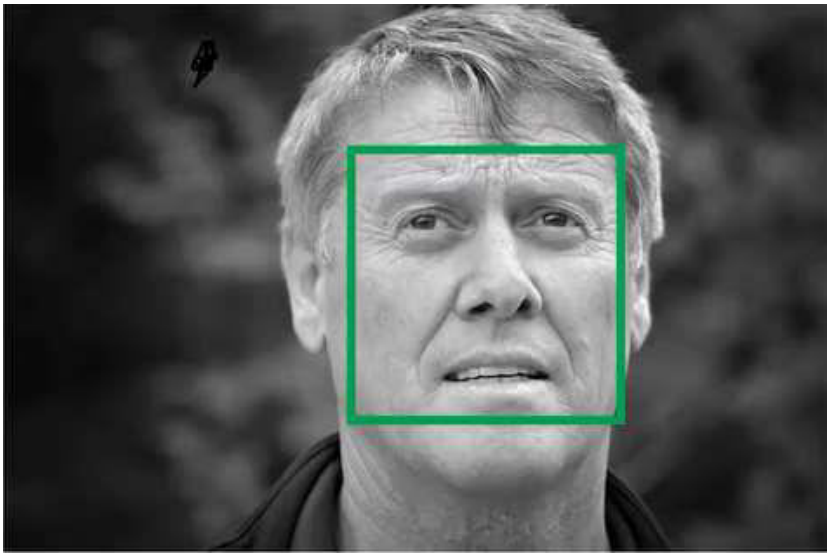


Fig. 7. Landmark image

Suppose we need to detect the eyebrow. Then we need edge features. If we want to detect nose then we need line features black-white-black in figure 7. If we want to detect teeth then we need edge features. After using these Haar features the image goes on the next feature. The ratio between these detected features is used in emotion detection.

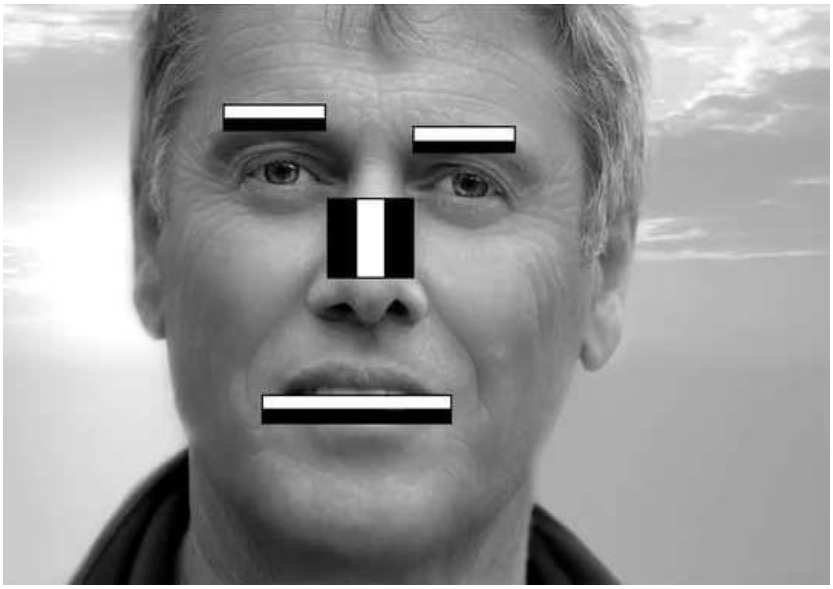


Fig. 8. Haar features image

We can calculate value by using Fourier equation,

$$\Delta = dark - white = \frac{1}{n}\sum_{black}^{n}[(x) - \frac{1}{n}\sum_{white}^{n}](x)$$

For ideal Haar features in figure 8,
The black region value is 1 and the white region value is 0. So the difference between dark and white 1-0=1.

Δ for ideal Haar features is 1

For real image,
If we calculate the black region and we average it's when we get 0.74 and the same way white region value is 0.18. So the difference between dark and white: 0.74-0.18=0.56
Δ for real image: 0.56.

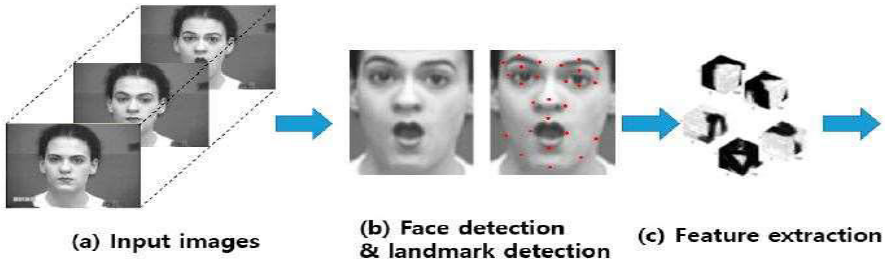


Fig. 9. Feature extraction

Neural networks are generally organized in layers. Layers consist of several interconnected nodes containing an activation function. The models are presented to the network through the input layer, which communicates with one or more hidden layers, where the actual processing is done through a weighted link system. This facial expression recognition system process is divided into three phases: image preprocessing, which includes recognition of the face and parts of the face using the Viola-Jones algorithm, extraction in Figure 9 facial features and classification of features using CNN.

Keras is an open source neural network in Python used for preprocessing, modeling, evaluation, and optimization. It is used for high level APIs because it is processed by the backend. It is designed to create a model with a loss and optimization function and a learning process with an adaptation function. For the backend, it was developed for low level convolution and computation under tensors or Tensor Flow. Importing the following Python libraries is used for preprocessing, modeling, tuning, testing, and displaying emotions with a maximum percentage. It uses sequential model and some layers like image

preprocessing, convolutional layer, pooling layer, flattening and dense layers, activation, ReLU. Image preprocessing is the first phase of the proposed system and includes facial recognition and FPs detection and extraction. It uses the Viola Jones facial recognition framework, a robust algorithm that allows images to be processed extremely quickly in real-time situations. This algorithm detects the facial area regardless of the variance in size, background, brightness and spatial transformation of the raw input image. Face FP recognition is achieved by combining classifiers in a cascade structure, which can increase recognition performance while reducing computational complexity. The final classifier is calculated by the linear combination of all weak classifiers, separating the positive and negative values in terms of weighted error (each learner's weight is directly proportional to their accuracy).

The face is first recognized, cropped, extracted and normalized to a size of 64 x 64 pixels. Then the parts of the face (eyes and mouth) are recognized, cropped and extracted from the normalized face image. The extracted face parts will be resized to the same size of 32 x 64 pixels. The reduced image scale reduces the information that must be learned from the network and speeds up training faster and with less storage costs. Convolutional layers are added to provide greater precision for large amounts of data. The data set is collected from a CSV file (in pixel format) and converted into images. Then the emotions are classified with the appropriate expressions. Here, the emotions are classified as happy, sad, angry, surprised, neutral, disgusted and afraid with 34,488 images for the training data set and 1,250 for the tests.

Each emotion is expressed in different facial features such as eyebrows, open mouth, raised cheeks, wrinkles around the nose, wide open eyelids and much more. Train the large data set for better accuracy and a result that is the feature class for an input image. Pooling is a concept of visual recognition of deep learning objects that is associated with convolution. The idea is that a convolution (or a local neural network detector) maps an area of an image onto a feature map. For example, a 5×5 array of pixels could be mapped onto oriented edge features. Flattening occurs when you flatten all of the Photoshop layers into a single background layer. Layers can increase the file size and thus tie up valuable processing resources.

To reduce the file size, you can merge some layers or even reduce the whole image to one background layer. The dense layer is the deeply connected regular neural network layer. This is the most commonly used layer. The dense layer performs the following operation on the input and returns the output. Based on the connecting forces (weight), blocking or excitation and transfer functions, the activation value is transferred from node to node. Each of the nodes adds up the received activation values. Then the value is changed according to its transfer

function. In Keras, the task can be implemented in our network architecture by adding exclusion layers. Each suppression level removes a custom hyperparameter of units from the previous level on each stack. Remember that in Keras the entry level is assumed as the first level and is not appended with add. ReLU is one of the most popular types of non-linearity in neural networks that are applied after the convolution layer and before maximum pooling. It replaces all negative pixel values on the feature map with zero. It is usually used after the fold layer.

Example: ReLU is the max (x, 0) function with input x, matrix of a folded image. ReLU then sets all negative values in the x matrix to zero and all other values are held constant. ReLU is calculated after convolution and is therefore a nonlinear activation function like als or sigmoid. Adam is an optimization algorithm that can be used in place of the classic stochastic gradient descent method to iteratively update lattice weights based on training data. It has been found that there are many endeavors have been taken using several automated techniques to analyze emotions. However, most of them are found without any establishing framework and describing how to properly use them. More specifically, understanding and maintaining the emotion analysis capability can help law-enforcement authorities effectively use machine learning techniques to track and identify emotion patterns.

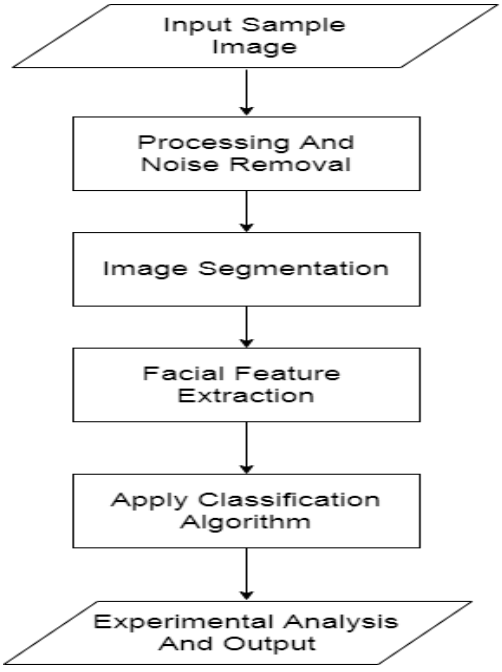


Fig. 10. Emotion Detection Data Flow Diagram

First take the picture of the user, then remove the noise in figure 10. Then just identify one person's face and apply the hair functions. Then adjust the image to match the previous training data set. Use the Python Keras library here. It works with the Convolutional Neural Network (CNN). CNN works on a sequential model. It also uses some levels like Conv2D, MaxPooling2D, AveragePooling2D, Dense, Activation, Dropout and Flatten. After the approach, these levels select the emotion from the classification set. This is the final edition.

After the preprocessing (if necessary), the normalized face image is displayed to the feature extraction part to find the key features used for classification. In other words, this module is responsible for creating a vector of features good enough to represent the image of the face. After this comparison, the image of the face is classified into one of the seven expressions (anger, contempt, disgust, fear, happiness, sadness, surprise).

4. RESULTS AND EXPERIMENTAL ANALYSIS

The first major challenge was the confined measure of information for preparing a broad framework. This needs to defeat for framework in nature. Move learning is the most prevalent response to this. In this methodology that was begun from pre-prepared strategy and calibrated this model with the put-away information which is gathered from a genuine world. A progression of starter investigations affirmed the presumption that face acknowledgment would serve better in highlight extraction in figure 11. There are models where such systems are effectively utilized.

Machine learning algorithms work well on the datasets that have a couple of hundred highlights or segment. The algorithm successfully classifies an image and classify the sentiment of the image and choose the match emotion for the image. The reason behind choosing the deep learning classifier is that the classifier runs data through several layers. And a deep learning algorithm can be useful for less unpredictable issues since they gain admittance to an immense measure of information to be compelling. For pictures, the regular benchmark for preparing profound learning models for broad picture acknowledgment approaches more than 14 million pictures. For perfect visualization of emotion detection pattern analysis, it used a decision tree. In the decision tree, the character is represented by the nodes and layers, and also the outcome of the experiment is represented by the branch.

The advantage of the decision tree is that it is very helpful and easy to visualize the emotion and interpret the result in figure 12. The working process of a decision tree is easy to understand. If it has been classified the data according to

their movement, reactions, and order which ideally different types of emotions. This also has been classified into trees and subtrees which reflects that Whether the person is sad, angry or happy, etc. if this could find something that can categorize their using these methods more simply. To do this it has been used retrain method that memorized the pattern and satisfies the condition. When any of the condition is satisfied it carry on to the end of the tree. However, if none of the conditions satisfy the intermediate condition, it will stop checking and say "The emotion cannot be identified. The emotion is unknown".

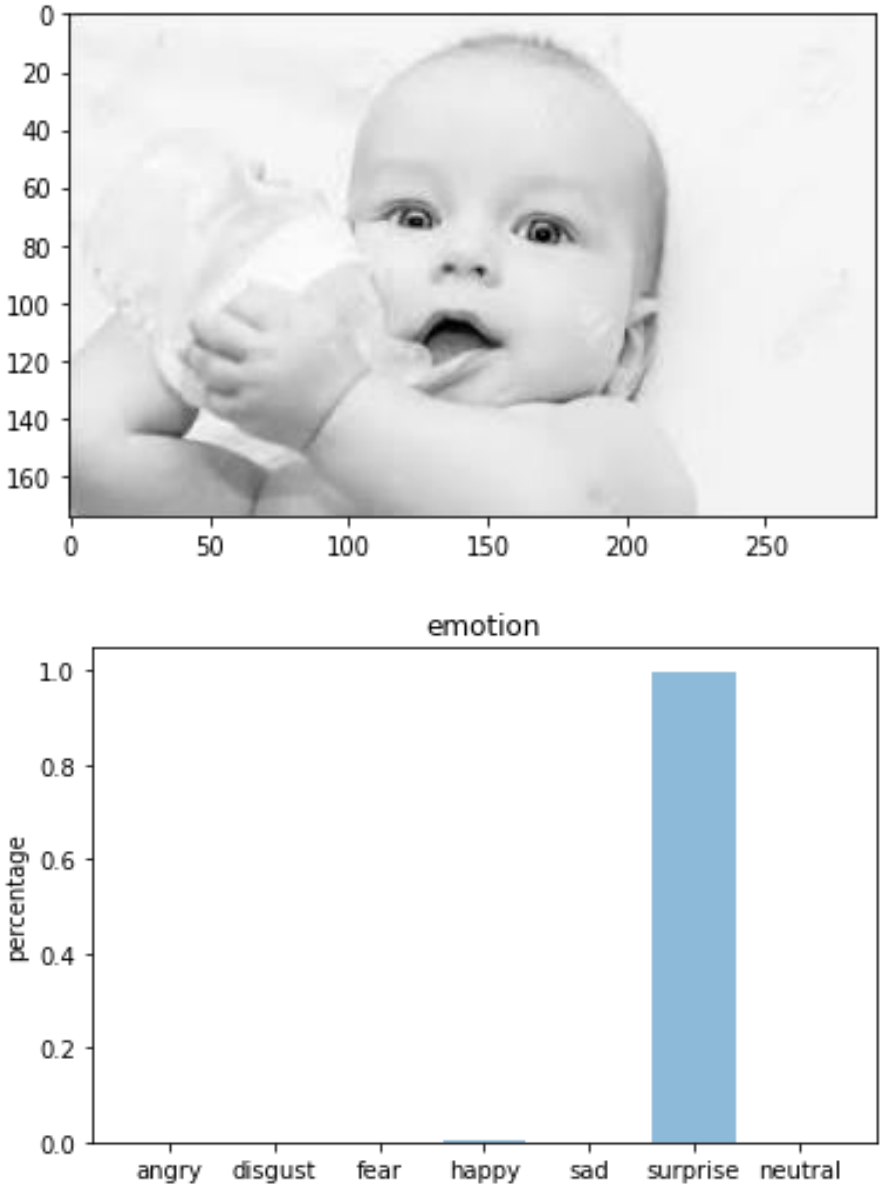


Fig. 11. Expression with surprise emotion

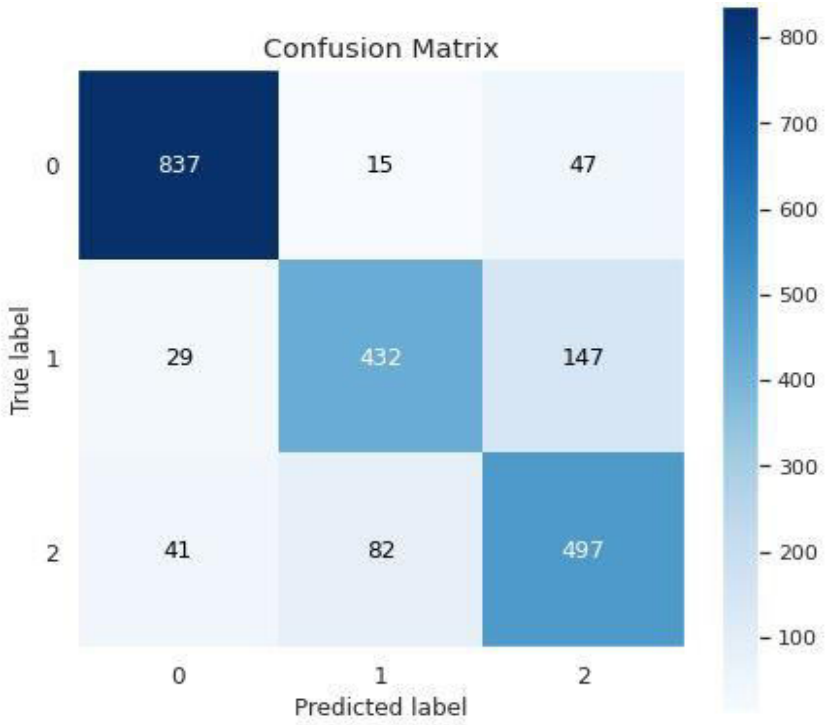


Fig. 12. Confusion matrix of CNN algorithm

Emotions are complicated to understand. There are different kinds of expression for the same emotion. Different people give different kinds of expression for the same kind of emotion. Modern-day machine learning technology can help law-enforcement authority to detect emotion so the machine can understand the emotion of humans and more behave and act like humans. This data for emotion came from different online and offline media. Such as Google, kaggel.com site. Friends and family, random people, etc. This is used Keras library to initially classify and analyze the emotion and got that data. Then with the help of Haar features and Numpy, It identifies the emotion. And with the help of platform anaconda. It generates the output from the raw data where the result is going to show in real-time.

The hierarchical data mining procedure like decision tree helps to generate probability decision by calculating various probability decisions by calculating various characteristic which is initially used to identify the emotion pattern. Along with offline and online data collection, it also conducted an effective field study to gather more people and various kinds of people and various emotional deferent expressions lots of different faces. In online data collection,

the data set is taken from kaggel.com. They provide quality data sets. They converted the images into pixel grayscale and use the numerical number of the images. So, it gives the quality data and the batter result. Both of the experts believed that this analysis of sentiment could help identify emotion more accurately and help to take accurate actions on behalf of accurate emotion identification. It would provide more knowledge about different types of expression of their sentiment as well as the percentage of each existed various kinds of emotions.

While completing this work, we found that a large quantity of test data and keywords are needed if it wants to get greater accuracy. A lack of a good quantity of raw data is also required to extend the research work. A high configuration graphics processing unit (GPU) qualified computer is also required if this wants to process a large quantity of test data in the shortest time. So, if this gets adequate data along with a high-performance computer, it will be easier for that to rise the accuracy to more than 93% in figure 13. It will also be able to use that system for a different platform for a different outcome and help to determine the emotion expression pattern in figure 14.

```
Number of instances:  35888
Instance length:  2304
28709 train samples
3589 test samples
Epoch 1/25
256/256 [==============================] - 390s 2s/step - loss: 1.7599 - accuracy: 0.2730
Epoch 2/25
256/256 [==============================] - 393s 2s/step - loss: 1.5138 - accuracy: 0.4065
Epoch 3/25
256/256 [==============================] - 392s 2s/step - loss: 1.3536 - accuracy: 0.4787
Epoch 4/25
256/256 [==============================] - 394s 2s/step - loss: 1.2466 - accuracy: 0.5215
Epoch 9/25
256/256 [==============================] - 393s 2s/step - loss: 0.9376 - accuracy: 0.6479
Epoch 10/25
256/256 [==============================] - 393s 2s/step - loss: 0.8871 - accuracy: 0.6696
Epoch 11/25
256/256 [==============================] - 392s 2s/step - loss: 0.8187 - accuracy: 0.6943
Epoch 16/25
256/256 [==============================] - 391s 2s/step - loss: 0.4975 - accuracy: 0.8158
Epoch 17/25
256/256 [==============================] - 394s 2s/step - loss: 0.4618 - accuracy: 0.8302
Epoch 23/25
256/256 [==============================] - 390s 2s/step - loss: 0.2680 - accuracy: 0.9030
Epoch 24/25
256/256 [==============================] - 389s 2s/step - loss: 0.2493 - accuracy: 0.9190
Epoch 25/25
256/256 [==============================] - 390s 2s/step - loss: 0.2389 - accuracy: 0.9372
```

Fig. 13. Output of the proposed neural set

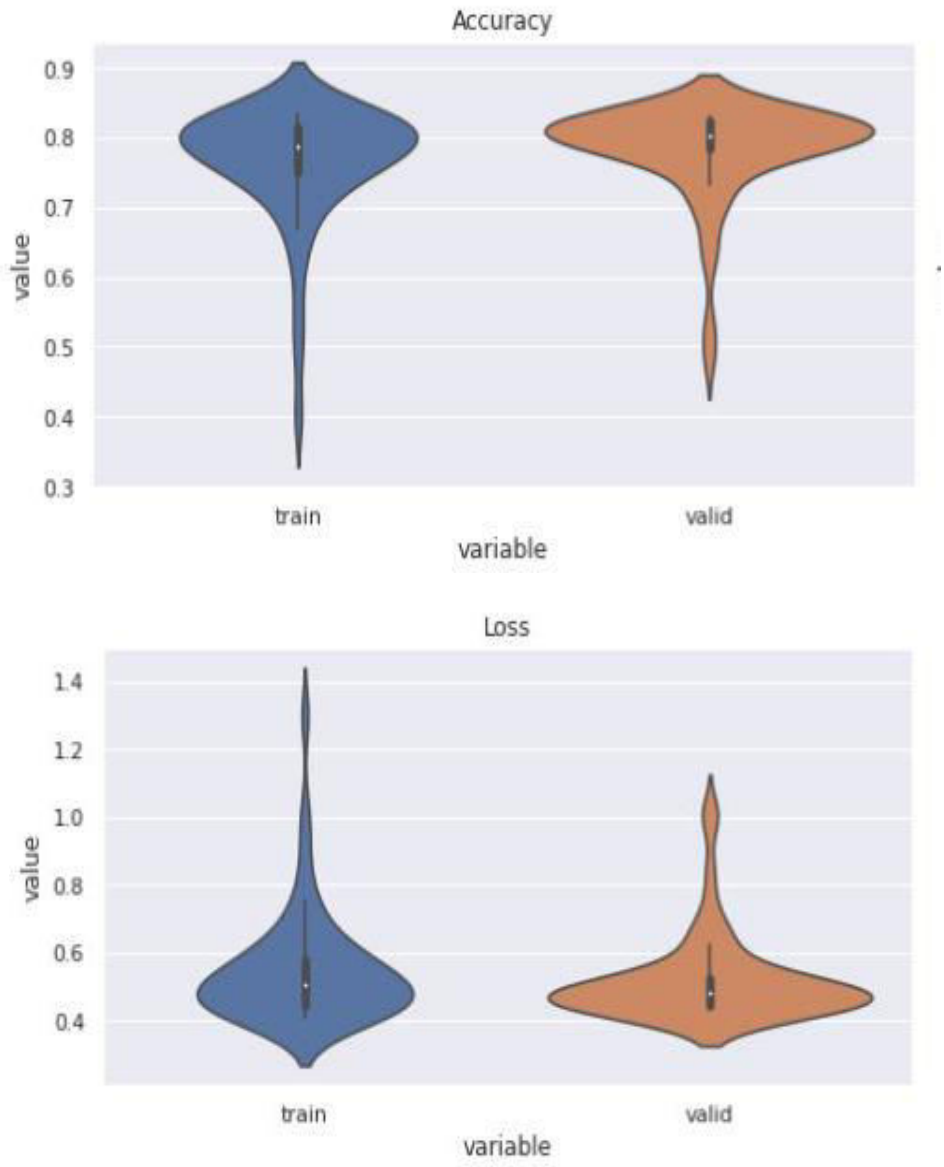


Fig. 14. Accuracy with varience of Epoch

5. CONCLUSION

By taking into account the whole process precise results were given by Neural Network in comparison to traditional machine learning approaches and gave 98.75% of accuracy than other models. The classifier can help in cataloging of Sadhu and Cholit languages of Bangladesh. As Bangladeshi literature is enriched with Sadhu language so most of the novels present from ancient era are in Sadhu language and this approach can help in their translation more accurately as its highly uncommon for present generation. That is why it can be

converted to Cholit so that people can get familiar to old literature. An experienced human can often identify another human's emotions by analyzing and looking at him or her. However, in this modern age machines are becoming more intelligent. For the time been machines are trying to act more like humans. If the machine has been trained on how to react on behalf of the human sentiment at that time. Then the machine can behave and act like a human. On the other hand, if the machine can identify the emotion it can prevent lots of occurrences too.

With increased proficiency and errorless computation emotion, data mining can facilitate accurate expression patterns enabling machines to find and act more like humans effectively. To determine the emotion expression patterns this thesis is created or framework with comprehensive research and field works. This followed the framework step by step to get the expected outcome. To follow the framework and to identify the emotion expression patterns more effectively and used deep learning CNN algorithm along with Keras, Tensorflow, and retraining concepts.

With these techniques, it was possible to identify emotions, type of emotion in the real image. To delineate the result and procedures more visually and this has also introduced decision tree techniques which helps to decide which emotions percentage is high and which emotions percentage is low. Now the high percentage of emotions get the most possible accurate emotions. And the low percentage of emotions get the low chance of existence. With this discovery, it is now possible to determine accurate emotions. And machines can identify emotion more accurately and on behalf of that, they can give a proper reaction and also can help to prevent the same unwonted occurrence. This machine can also become the replacement of a human.

REFERENCES

1. Raghav Puri, Archit Gupta, Manas Sikri, " Emotion Detection using Image Processing in Python," in IEEE Conference ID: 42835 2018 5th International Conference, 14th - 16th March 2018
2. Wout Swinkels, Luc Claesen, Feng Xiao, Haibin Shen, "SVM Point-based Real-time Emotion Detection," in IEEE, 19 October 2017
3. Chu Wang, Jiabei Zeng, Shiguang Shan, Xilin Chen, "Multi-Task Learning Of Emotion Recognition And Facial Action Unit Detection With Adaptively Weights Sharing Network," in IEEE, 2019
4. Ninad Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," 18 February 2020

5. James Pao, “Emotion Detection Through Facial Feature Recognition,” in International Journal of Multimedia and Ubiquitous Engineering, November 2017
6. Aitor Azcarate, Felix Hageloh, Koen van de Sande, Roberto Valenti, “Automatic facial emotion recognition,” January 2005
7. Dan Duncan, Gautam Shine, Chris English, "Facial Emotion Recognition in Real-Time," November 2016
8. Shivam Gupta, “Facial emotion recognition in real-time and static images,” in 2nd International Conference on Inventive Systems and Control (ICISC) IEEE, 28 June 2018
9. Aneta Kartali, Miloš Roglić, Marko Barjaktarović, Milica Đurić-Jovičić, Milica M. Janković, “Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches,” in 2018 14th Symposium on Neural Networks and Applications (NEURAL), Nov 2018
10. Jonathan, Andreas Pangestu Lim, Paoline, Gede Putra Kusuma, Amalia Zahra, “Facial Emotion Recognition Using Computer Vision,” in Indonesian Association for Pattern Recognition International Conference (INAPR) IEEE, 31 January 2019
11. Renuka S. Deshmukh, Vandana Jagtap, Shilpa Paygude, “Facial emotion recognition system through machine learning approach,” in International Conference on Intelligent Computing and Control Systems (ICICCS) IEEE, 11 January 2018
12. Hyeon-Jung Lee, Kwang-Seok Hong, “A Study on Emotion Recognition Method and Its
13. Application using Face Image,” in International Conference on Information and Communication Technology Convergence (ICTC) IEEE, 14 December 2017
14. T. K. Senthilkumar, S. Rajalingam, S. Manimegalai, V. Giridhar Srinivasan “Human Facial Emotion Recognition Through Automatic Clustering Based Morphological Segmentation And Shape/Orientation Feature Analysis,” in IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 08 May 2017
15. Piórkowska, Magda, Wrobel, Monika, “Basic Emotions,” in Springer International Publishing AG, 15 July 2017
16. Ninad Mehendale “Facial emotion recognition using convolutional neural networks
17. (FERC),” in Springer Nature Switzerland AG, 18 February 2020
18. Victor M. Alvarez, Ramiro Velázquez, Sebastián Gutierrez, Josué Enriquez-Zarate “A Method for Facial Emotion Recognition Based on Interest Points,” in International Conference on Research in Intelligent and Computing in Engineering (RICE), 25 October 2018

19. Victor M. Alvarez, Ramiro Velázquez, Sebastián Gutierrez, Josué Enriquez-Zarate "A Method for Facial Emotion Recognition Based on Interest Points," in International Conference on Research in Intelligent and Computing in Engineering (RICE), 25 October 2018
20. Dumas, Melanie, "Emotional Expression Recognition using Support Vector Machines," July 2001
21. Muzammil, Abdulrahman, " Facial expression recognition using Support Vector Machines," in 23nd Signal Processing and Communications Applications Conference (SIU) IEEE, 22 June 2015
22. Turabzadeh, Saeed & Meng, Hongying & Swash, M. & Pleva, Matus & Juhár, Jozef, "Facial Expression Emotion Detection for Real-Time Embedded Systems," January 2018
23. Byoung Chul Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," 30 January 2018

Cite this article

Md. Forhad Ali, Mehenag Khatun and Nakib Aman Turzo, Facial Expression Recognition Using Neural Network, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 33-52

4 Development Software for Preprocessing Voice Signals

Niyozmatova N[a,*] Mamatov N[a], Nurimov P[a], Samijonov A[a] and Samijonov B[a]

[a]Tashkent University of Information Technologies named after Al-Kharezmi, Tashkent, Uzbekistan

*Corresponding author: n_nilufar@mail.ru

ABSTRACT

At present, the most important task of modern science is the creation for a person of natural means of communication with a computer, where speech input of information is carried out in the most convenient way for the user. Speech recognition is one of the challenges. As practice shows, the quality of recognition depends on the properties of the preprocessing system. To improve the quality of recognition, it is necessary to develop effective and high-speed methods and algorithms for signal preprocessing. This article proposes a new approach and algorithm for extracting features of speech signals. Based on the proposed algorithm, the identification problem is solved. In addition, the chapter presents a description of the software module for each stage of the preliminary processing of speech signals. This software is a voice-based identity tool.

Keywords: *speech signal, filter, MFCC, PLP, LPCC*

Advanced Research in Computer Engineering | ISBN: 978-81-948951-2-1
Book Series: Research Transcripts in Computer, Electrical and Electronics Engineering | Vol: 02

1. INTRODUCTION

Speech is a sequence of sounds. Sound is a superposition of sound waves (vibrations) with different frequencies. From the point of view of physics, a wave is characterized by amplitude and frequency. Speech signal processing is a field of science where noise suppression, amplification, filtering, information extraction, information separation, compression, restoration and coding are carried out. It has become widespread in all areas of speech technology.

2. METHODOLOGY AND SOLUTION OF THE PROBLEM

The software for pre-processing speech signals consists of several stages; each stage includes several modules (Fig. 1).

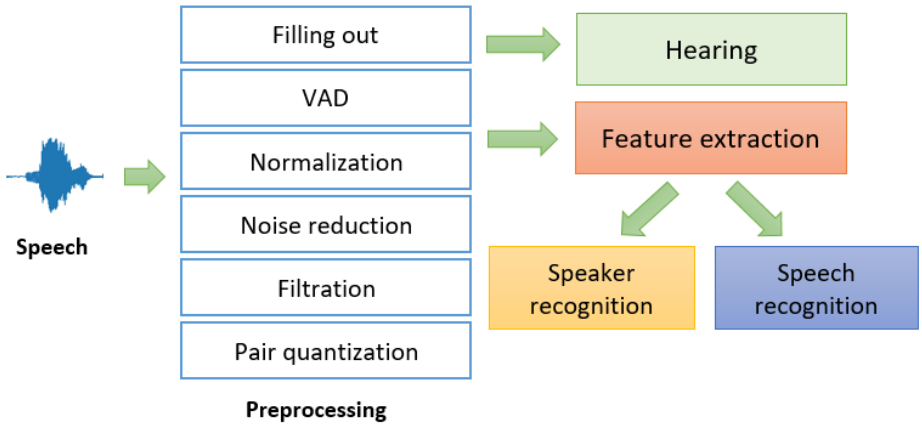


Fig. 1. The structure of the software for preprocessing speech signals

2.1. The speech signals

The first stage of the system under development is “Speech Signals”. At this stage, voice signals received from a file or through a microphone. At the first stage, speech signals received from a file or through a microphone.

2.2. Preprocessing of speech signals

At this stage, the signal are initially processed by the following modules,

2.2.1 Filling out

Gaps in speech signals are filled.

2.2.2 Voice Activation Detection-VAD

This module is used to find speech in audio. VAD is used to reliably detect speech in audio and even in background noise. In completely pure audio recordings, even elementary energy detection may be sufficient for speech recognition; but unfortunately, in the wild there are not always completely pure signals, so VAD must be noise resistant. VAD consists of the steps:

- Signal division into frames
- Formation of features for each frame
- Classifier training in active and quiet frames
- Classification of invisible frames as speech or silence

An example on the created software using the function of detecting voice activity for a speech signal is shown in Fig. 2 and Fig.3 [1].

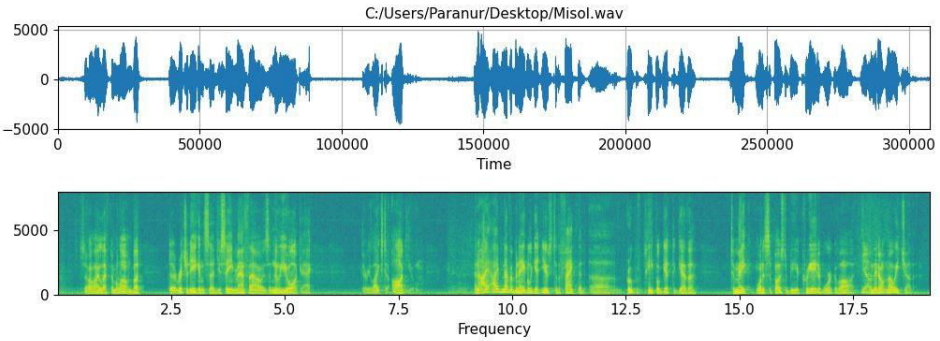


Fig. 2. Speech signal before applying VAD

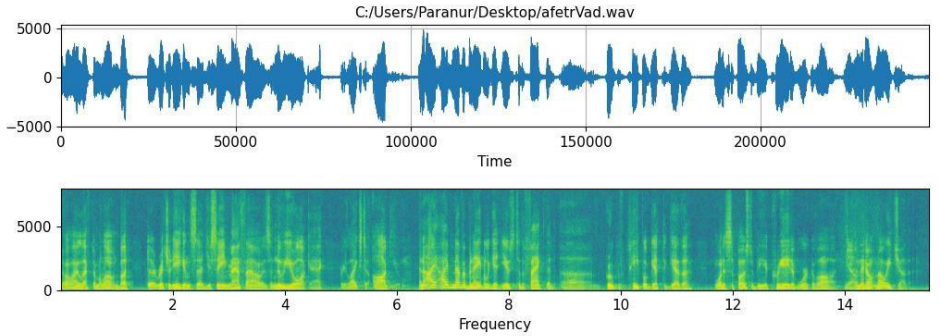


Fig. 3. Speech signal after applying VAD

2.2.3 Normalization

EBU R128 volume normalization consists of dynamic and linear modes. Support is provided in the form of single-pass (live broadcasts, files) and two-pass (files) modes. Typically, this algorithm targets IL, LRA, and the largest true peak. If the normalization mode is not linear, then the audio stream will be

sampled to 192 kHz to accurately determine the true peaks. To explicitly set the sample rate of the output, we need to use the -ar option or a sample filter [2]. The filter has the following variable parameters (Fig. 4):

- I, i - setting the built-in volume range. Here, the default values are -24.0, and the interval [-70.0; -5.0].
- LRA, lra - setting the target volume range. Here the default values are 7.0 and the interval is [1.0; 20.0].
- TP, tp - setting the maximum true peak. Here the default values are -2.0 and the interval is [-9.0; 0.0].

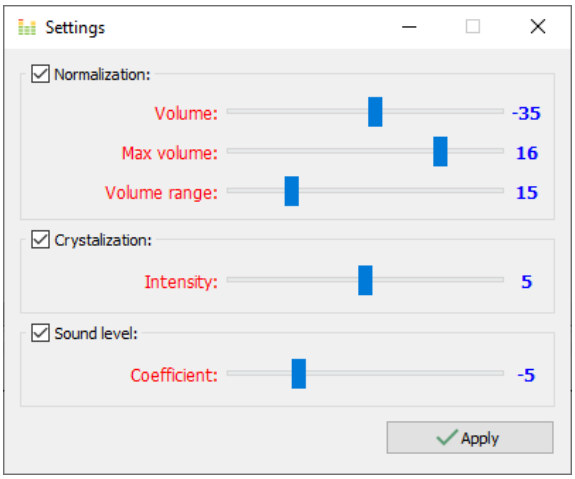


Fig. 4. Normalization parameters

2.2.4 Noise reduction

In [2], noise suppression of audio samples with FFTs is described in detail. In experimental studies, the parameters were taken as follows:

- nr is noise reduction (dB), the default value is 12 (dB), and the allowable range is nr from 0.01 to 97.
- nf is the minimum noise level in dB, where the allowable range is between -80 and -20. The default is -50 dB.
- nt is the type of noise that has the parameters: wn is white noise, Vn is vinyl noise, sn is shellac noise and cn is the user noise defined in option bn. nt parameter defaults to white noise.
- bn is the user noise range which is each of the 15 bands, where the bands are separated by '' or '|'.
- rf is the residual level (dB), where the permissible range of rf is between -80 and -20. The default is -38 (dB).
- tn - noise tracking, which takes the values 1 (true) or 0 (false). The default is false. With this power on, the noise level is automatically adjusted.
- tr - track balances, enable or disable. The default is 0 (Fig. 5).

2.2.5 Filtration

Filtering is one of the main stages of processing time or spatial series of measurements. Currently, there are many filtering methods, for example, median filtering, polynomial approximation, cosine filtering, Fourier transform and wavelet transform, etc. (Fig. 6).

The Savitsky-Golay filter is digital

Based on the Savitsky-Golay filter, a set of digital data points is carried out for data smoothing, which allows increasing the data accuracy without signal distortion. Precision is maximized in the process, that is, as in convolution, by selecting successive subsets of adjacent points using a low degree algebraic polynomial based on the least squares method. If the points are located at the same distance, then the analytical solution to the least squares equations can be found in a single set of "convolution coefficients" that can be applied to all subsets of the data to obtain smoothed estimates. The signal (or derivatives of the smoothed signal) at the center point of each subset. This method, based on established mathematical procedures, was popularized by Abraham Savitsky and Marcel J.E. (Fig. 7, Fig 8.)

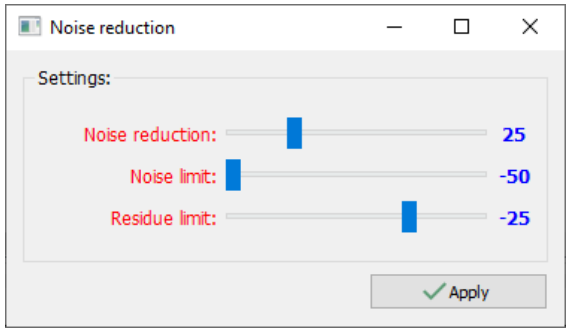


Fig. 5. Window of Noise Reduction

The created software uses the savgol_filter [3] function in the Python programming language. Filtering parameters are presented in the form below.

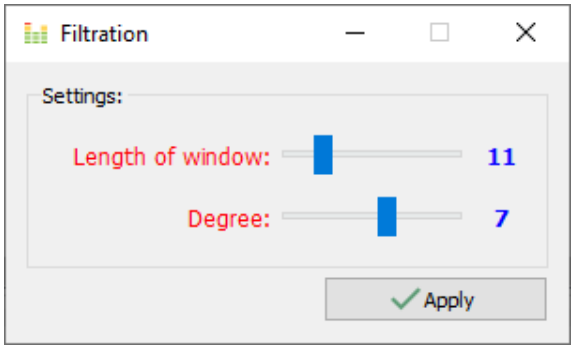


Fig. 6. Window of filtration

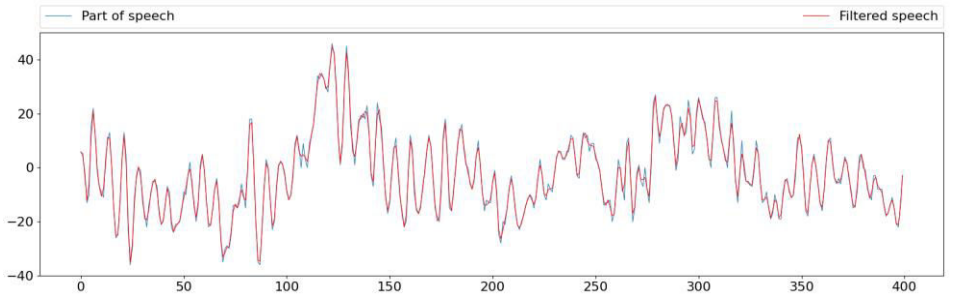


Fig. 7. An example of applying the filter of Savitsky-Golay. Win_len = 5, poly_order = 3

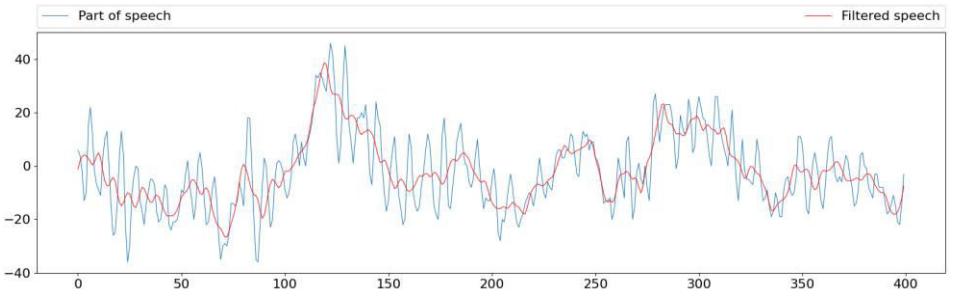


Fig. 8. An example of applying the filter of Savitsky-Golay. Win_len = 15, poly_order = 3.

2.2.6 Pair quantization

The pair quantization, carried out by reducing the number of proposed features, is implemented as follows. Features are distinguished from the speech signal based on the following formula (pattern),

$$s_i = \frac{1}{L} \sum_{j=i \cdot L}^{(i+1) \cdot L} x_j \qquad (1)$$

where x_j is a vector consisting of symbols of a given speech signal, L -step.

2.3. Hearing

At this stage, the pre-processed speech signal can be saved to a file in audio format.

2.4. Feature extraction

At this stage, the features of speech signals are formed as: MFCC, LPC, PLP and medium quantum.

The linear prediction cepstral coefficient method, the perceptual linear prediction coefficient method and robust PLP (PLP-RASTA), the cepstral

coefficient coefficient method on the chalk scale (MFCC) are the most powerful methods based on cepstral signal analysis.

LPCC is a linear prediction cepstral coefficient method. Based on the calculation of the coefficients of the autoregressive model for each frame of the audio signal. After obtaining all the model parameters, cepstral coefficients are calculated based on the recursive function.

PLP is a perceptual linear prediction coefficient method. The method differs from the LPCC method in that it takes into account the characteristics of the perception of various frequencies by a person - before calculating the parameters of the autoregressive model, the signal undergoes a certain pre-processing. The calculated instantaneous Fourier spectrum is converted into a spectrum on the barque scale, after which the operation of convolution of the masking curves of the critical bands with the obtained spectrum is performed to obtain the frequency masking effect. Next, the volume curve and cepstral processing are approximated.

The advantage of the PLP method compared to LPCC is that it allows you to suppress information related to the individual characteristics of the speaker by choosing the appropriate model order. However, this method is more sensitive to the pitch frequency.

2.4.1 Mel-coefficient coefficients (MFCC)

MFCC - based on human auditory perception, and obtained on a scale of twisted frequencies. To calculate the MFCC, a speech window is first created to divide the speech signal into frames. To get the same amplitude for all formats, high frequency formants are reduced in amplitude, compared to low frequency formants, high frequencies are emphasized. After creating a window for calculating the power of the spectrum for each frame, a fast Fourier transform is performed. After application (FFT) based on the filter base using the melting scale, it is processed by the power spectrum. To calculate the MFCCs after transforming the power spectrum to the logarithmic domain, DCT is performed in the speech signal. The following formula is used to calculate Mel for an arbitrary frequency [4, 5]:

$$mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \tag{2}$$

where $mel(f)$ -Mel frequency, a f - frequency.

MFCC is calculated using the following formula:

$$\hat{C}_n = \sum_{n=1}^{l} \left(\log \hat{S}_l\right) \cos\left[n\left(l - \frac{1}{2}\right)\frac{\pi}{l}\right] \tag{3}$$

where $l -$ number of Cepstral melting factors, $\hat{S}_l$ - output filter set, $\hat{C}_n$ are the last coefficients of MFCC.

Based on MFCC, the low-frequency region is determined more efficiently than the high-frequency region, and it can calculate formants belonging to the low-frequency range, and also describes the resonances of the vocal tract.

MFCC is universal and recognized as an interface procedure for typical identification applications [4]. In addition, MFCC is ideal when the source characteristics are stable and consistent across sounds [6]. MFCC is also able to capture information from sampled signals in which the frequency is not more than 5 kHz. Such signals cover most of the energy of sounds generated by humans.

In speech recognition, MFCC is widely used [7]. There are formants higher than 1 kHz and are not taken into account due to the large distance in the high frequency range between filters [4]. In the presence of MFCC background noise, the signs are not accurate and cannot be generalized [5].

2.4.2 Perceptual Linear Prediction (PLP)

Based on the PLP method, critical bands are combined, the intensity is compressed into loudness and a preliminary emphasis on equal loudness when extracting relevant information from speech.
PLP is based on a non-linear scale and was first used in speech recognition problems to eliminate speaker-dependent functions [8]. PLP represents a smoothed aligned and compressed, similar to human hearing, short-term spectrum, transforms it as MFCC.
Based on the PLP approach, some important hearing characteristics can be expressed, and the subsequent auditory speech spectrum is approximated using an autoregressive pan-polar model [9]. With PLP, the minimum high frequency resolution is obtained, which means the auditory filterbank approach, but produces orthogonal results like cepstral analysis. For spectral smoothing, PLP uses linear prediction, hence the name - perceptual linear prediction [10]. The combination of spectral and linear predictive analysis is PLP.
To calculate the characteristics of PLP, speech is highlighted as a window (Hamming window), the Fast Fourier Transform (FFT) is calculated and squared. This gives energy-spectral estimates. Then a trapezoidal filter is applied at a predetermined interval, usually 1 cortex.
This filter integrates overlapping responses and also effectively squeezes the highest frequencies into a narrow band. As a result of integration, high frequencies are effectively compressed into a narrow band.
Then, on the scale of distorted frequencies of the cortex, convolution is performed in the symmetric frequency domain, which allows smoothing the spectrum, where low frequencies mask with high frequencies.
Then the symmetrical frequency domain is convolved according to the scale of distorted frequencies of the cortex, which, by smoothing the spectrum, allows low frequencies to mask high frequencies. Subsequently, the spectrum is pre-emphasized to approximate the uneven sensitivity of human hearing at different

frequencies. Compression of the spectral amplitude decreases the change in the amplitude of spectral resonances. IDCT is performed to obtain autocorrelation coefficients. Spectral smoothing is carried out on the basis of solving autoregressive equations. With the help of autoregression coefficients, the values of cepstral parameters are obtained [10].

To calculate the frequency of the cortex, the following formula is used:

$$bark(s) = \frac{26.81s}{1960 + s} - 0.53 \tag{4}$$

where $bark(s)$ - core frequency, and s - frequency in Hz.,

The identification results based on PLP are much better than LPC [10]. It justifies that PLP effectively suppresses speaker-specific information. In addition, it has improved independent recognition characteristics and is resistant to noise, changes in channels and microphones.

Based on PLP, autoregressive noise components are accurately restored. In addition, PLP is more sensitive to any changes in the frequency of formants.

Linear Prediction Cepstral Coefficients (LPCC). LPCC is the cepstral coefficient, which is obtained from the calculated envelope of the LPC spectrum. The LPCC coefficients are illustrations of the Fourier transform of the logarithmic spectrum of quantities [11, 12] LPC.

Cepstral analysis ideally symbolizes speech signals and characteristics with a limited size of functions and therefore they are usually used in the field of speech processing [12].

Rosenberg and Sambur noted that adjacent predictor coefficients have a high correlation, and with less correlated characteristics the representations will be more effective, therefore LPCC is of this kind.

If the signals have a minimum phase, then the LPC is easily converted to LPCC [13].

In speech processing based on LPCC, they are likewise computed as LPC [12].

LPCC are calculated based on the following formula [13]:

$$C_m = a_m + \sum_{k=1}^{m-1} \left[\frac{k}{m} \right] c_k \, a_{m-k} \tag{5}$$

here a_m - linear prediction and C_m - cepstral coefficient.

LPCCs have a lower error rate than LPC functions [12] and have a slight noise vulnerability [11]. Higher order cepstral coefficients are mathematically limited, resulting in an extremely large amount of variation when going from lower order cepstral coefficients to higher order cepstral coefficients. Likewise, LPCC estimates are highly sensitive to quantization noise.

In high frequency speech signal, cepstral analysis gives little separability between source and filter in the field of quantity. The lowest order cepstral coefficients are sensitive to spectral tilt, and the highest order coefficients are noise sensitive.

The result of this step, i.e. the processed speech signal is transmitted to the identification step. If the identification result is lower than expected, proceed to stage 3 and the process continues.

2.5. Speaker recognition and speech recognition

At this stage, the processed speech signal is used to identify a person by voice. If the result is lower than the specified result, go to stage 3, and the process continues.

Vector quantization (VQ). VQ [15] is an efficient data compression method. It is effectively and successfully applied in various systems such as vector quantization coding and recognition.

Codebooks are created on the basis of the LBG algorithm [14, 15]. In [16, 17], the following steps of the LBG algorithm are given:

i. Development of the 1st vector codebook; which is the centroid of all training vectors.

ii. Doubling the size of the codebook is done according to the rule by dividing each current codebook c_m:

$$c_m^+ = c_m(1+\varepsilon) \tag{6}$$

$$c_m^- = c_m(1-\varepsilon) \tag{7}$$

where m ranges from 1 to the current codebook size, ε - split parameter.

iii. Finding centroids for a shared codebook

iv. Repeat stages 2 and 3 until a codebook of size M is developed.

Euclidean distance

Based on the Euclidean distance, the similarity or difference between two spoken words is calculated, arising after quantizing these words in its codebook. The new word is compared by measuring the Euclidean distance between the feature vector of the new word and the model (codebook) of known words in the base. The word with the smallest distance is chosen according to the following formula:

$$d(x,y) = \sqrt{\sum_{i=1}^{M}(x_i - y_i)^2} \tag{8}$$

where, x_i is the *i*-th vector input features, y_i is the *i*-th vector features in the codebook, d is the distance between x_i and y_i.

Based on the proposed structure (Fig. 1), software for personal identification by voice (Fig. 10) and speech recognition has been developed. The general view of the developed software is shown in Fig. 9.

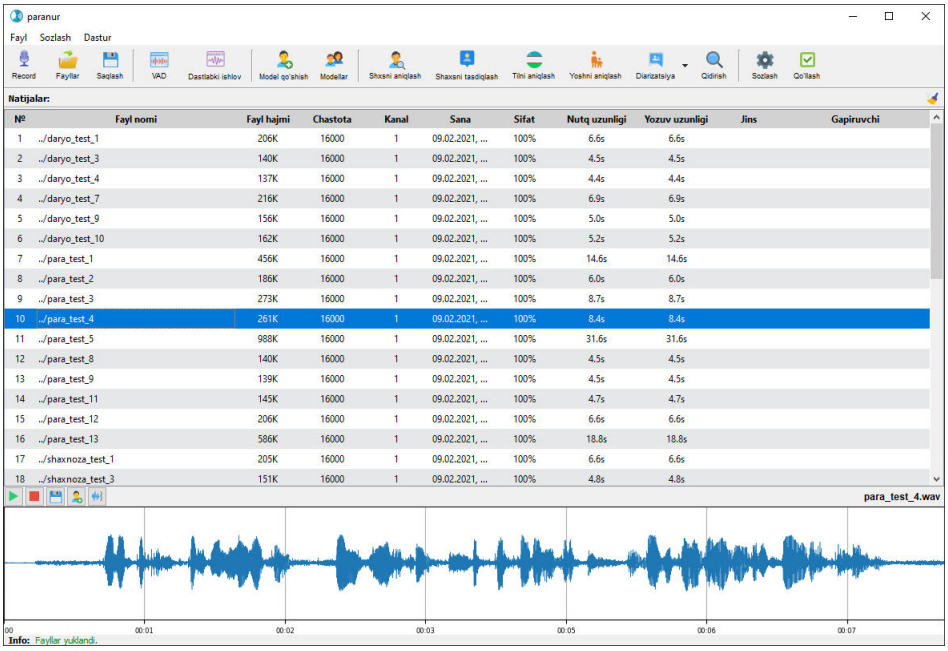


Fig. 9. Window software preprocessing speech signals

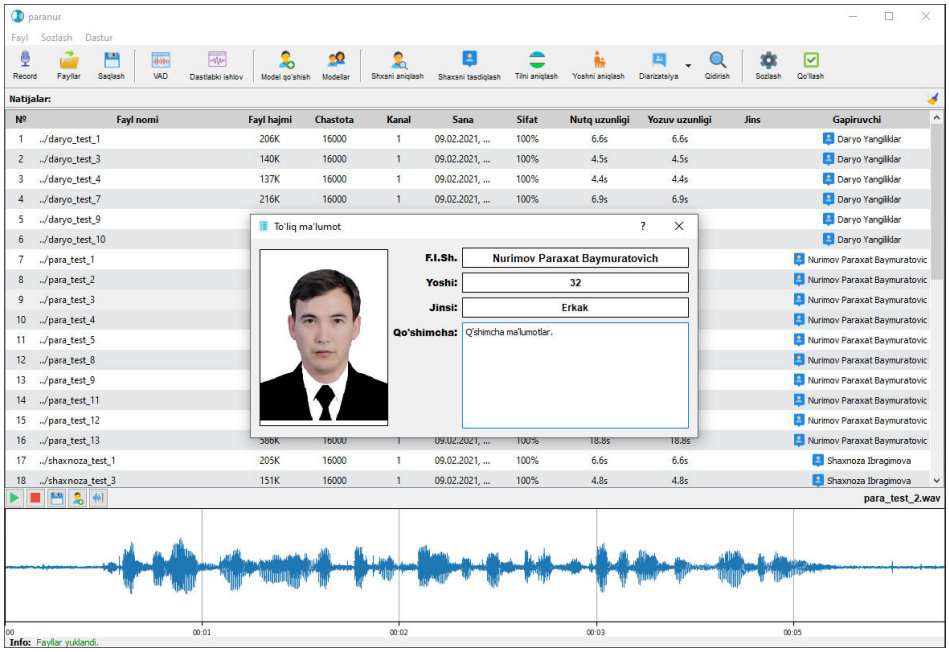


Fig. 10. Identification module

3. RESULTS

The base of speech given in [18, 19, 20] consists of 100 people, of which 74 are men and 26 are women. The database contains 10 speech files for each person, 6 of which were used for training and the remaining 4 for testing. In this case, the duration of each file is 2-5 seconds.

The experimental results were obtained on the basis of the above-mentioned speech base using the main algorithm and the proposed algorithms.

Table 1. The results of the comparison algorithms

Algorithm	Data Size (MB)	Results (%)	Time of training (sec.)	Time spent on identification (sec.)
Basic	285,3	99,6	45,3	69,3
The proposed	31,5	98,4	6,7	13,1

4. DISCUSSION

Experimental results show that the accuracy of the proposed algorithm is less than 2% than that of the existing algorithm. Experimental studies have shown that the famous algorithms do not meet the requirements of real time. The algorithm proposed in the work, by reducing the number of features used in personal identification by voice, made it possible to reduce the recognition time by up to 5 times. And also the size of the file used for recognition is significantly reduced. This ensures a reduction in the volume of the speech database and an increase in the speed of data transmission over the channels. This justifies the use of the program based on the proposed algorithm in recognition systems.

5. CONCLUSION

The chapter proposes a pair quantization algorithm. The number of signs obtained using this algorithm is at least two times less than the number of signs obtained using existing algorithms. This speeds up identification. Based on the proposed algorithm, experimental studies were carried out using an example of solving a practical problem. As a result of the work done, a software structure for preprocessing speech signals for a speech recognition and identification system is proposed. It is planned to develop a system for automatic speech recognition based on the attributes of a pair quantization algorithm.

REFERENCES

[1] https:github.com/wiseman/py-webrtcvad/

[2] https:ffmpeg.org/ffmpeg-filters.html#loudnorm

[3] https:docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.signal.

savgol_filter.html

[4] Chakroborty S, Roy A, Saha G. Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification. In: IEEE International Conference on Industrial Technology, 2006. ICIT 2006. pp. 387-390

[5] Hasan MR, Jamil M, Rabbani G, Rahman MGRMS. Speaker Identification Using Mel Frequency cepstral coefficients. In: 3rd International Conference on Electrical & Computer Engineering, 2004. ICECE 2004. pp. 28-30

[6] Chu S, Narayanan S, Kuo CC. Environmental sound recognition using MP-based features. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE; 2008. pp. 1-4

[7] Ravikumar KM, Reddy BA, Rajagopal R, Nagaraj HC. Automatic detection of syllable repetition in read speech for objective assessment of stuttered Disfluencies. In: Proceedings ofWorld Academy Science, Engineering and Technology. 2008. pp. 270-273

[8] Ravikumar KM, Rajagopal R, Nagaraj HC. An approach for objective assessment of stuttered speech using MFCC features. ICGST International Journal on Digital Signal Processing, DSP. 2009;9(1):19-24

[9] Hermansky H. Perceptual linear predictive (PLP) analysis of speech. The Journal of the Acoustical Society of America. 1990;87(4):1738-1752

[10] Kumar P, Chandra M. Speaker identification using Gaussian mixture models. MIT International Journal of Electronics and Communication Engineering. 2011;1(1):27-30

[11] El Choubassi MM, El Khoury HE, Alagha CEJ, Skaf JA, Al-Alaoui MA. Arabic speech recognition using recurrent neural networks. In: Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No.03EX795). Ieee; 2003. pp. 543-547. DOI: 10.1109/ISSPIT.2003.1341178

[12] Wu QZ, Jou IC, Lee SY. On-line signature verification using LPC cepstrum and neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 1997; 27(1):148-153

[13] Holambe R, Deshpande M. Advances in Non-Linear Modeling for Speech Processing. Berlin, Heidelberg: Springer Science & Business Media; 2012

[14] Linde Y., Buzo A. and Gray R.M. An algorithm for vector quantizer design. IEEE Trans. Communication, 1980. vol. COM-28, no. 1, pp. 84-95.

[15] A. Gersho, R. M. Gray Vector Quantization and Signal Compression. Kluwer Academic Publishers, Boston, MA, 1991.

[16] H. B. Kekre, Vaishali Kulkarni, Speaker Identification by using Vector

Quantization. International Journal of Engineering Science and Technology 2010 Vol. 2(5), pp 1325- 1331.

[17] M. Narzillo, S. Abdurashid, N. Parakhat, and N. Nilufar, “Automatic speaker identification by voice based on vector quantization method,” Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 10, pp. 2443–2445, 2019.

[18] B. Wiedecke, M. Narzillo, M. Payazov, and S. Abdurashid, “Acoustic signal analysis and identification,” Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 10, pp. 2440–2442, 2019.

[19] M. Narzillo, S. Abdurashid, N. Parakhat, and N. Nilufar, “Karakalpak speech recognition with CMU sphinx,” Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 10, pp. 2446–2448, 2019.

[20] https:www.dropbox.com/s/87v8jxxu9tvbkns/development_set.zip?dl=0

Cite this article

Niyozmatova N, Mamatov N, Nurimov P, Samijonov A and Samijonov B, Development Software for Preprocessing Voice Signals, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 53-66

5 Application of Classifiers for Assortment of Online Reviews

Biplob Kumar[a,*], Pritom Sarker[a] and Nakib Aman Turzo[a]

[a]Department of Computer Science & Engineering, Varendra University, Rajshahi, Bangladesh

*Corresponding author: kumarbiplob336@gmail.com

ABSTRACT

In Bangladesh, Ecommerce is flourishing day by day especially in the time of crisis the world is facing. There are many platforms available on these sites among which Daraz is the most successful marketplace. This online platform allowed people the ease to do shopping but a large number of reviews and comments made it difficult to opt for the best option. In this paper, the focus is on cataloguing the positive and negative reviews. For this purpose various classifiers were used by using Python. Data cleaning was done and after application of Term Frequency -Inverse Data Frequency with Principal Component analysis it was found that Ridge classifier performed best with more training time then other classifiers and depicted high accuracy. This classifier could help different businesses on different platforms to identify the positive and negative reviews and can provide customers with details about the quality of products.

Keywords: Adaboost classifier, Principal component analysis, Python, Ridge classifier, Term Frequency-Inverse Data Frequency

Advanced Research in Computer Engineering | ISBN: 978-81-948951-2-1
Book Series: Research Transcripts in Computer, Electrical and Electronics Engineering | Vol: 02

1. INTRODUCTION

Ecommerce is a short form of “electronic commerce’ which significantly elaborates any kind of exchange of currency for services online. Every business in ecommerce informs about the type of products they to their customers. There are three categories in which ecommerce businesses sell their products which are B2B, B2C and B2G. They can also be differentiated on the way they sell their services. These options are branded ecommerce stores, Ecommerce market places and conversational commerce. Popular ecommerce sites in Bangladesh are BD jobs, Clickbd, Bikroy, BoiMela, Rokomari, foodpanda, Daraz, Chaldal etc. Daraz is an online marketplace and logistics company which run in South Asian and Southeast Asian markets and was founded in 2012. It's one of the best ecommerce sites in Bangladesh. Ecommerce market in Bangladesh made a quantum jump in 2017 and developed at an incredible 70% from 2016. Ecommerce is at its peak in Bangladesh in this pandemic crisis. In this paper, a large number of Daraz product reviews were scraped and each review got labelled according to sentiment-based analysis that might be positive or negative. Total 900 reviews were amassed from Daraz and results in the building of a sentiment analyser which can be trained using data already present so that new reviews and comments can be catalogued. Presently, five ecommerce sites are growing rapidly in huge markets of Bangladesh. The principle goal of this study is to anticipate their sales and to find their customer satisfaction level.

2. LITERATURE REVIEW

Examination work was done which presents a structure for utilizing text digging for get-together client's input. This strategy is utilized to assemble the top credits related to a gathering of gadgets. An examination of properties of various gadgets and their positive and negative qualities referenced by clients were utilized to improve the group of people yet to come of items. [1].

For removal of uncertainty regarding purchases online reviews may help in clearing all doubts. A tale strategy was utilized that consolidated the Bass/Norton model and estimation examination while utilizing verifiable deals information and online audit information was created for anticipating item deals[2].

For text sentiment analysis feature extraction is one of the key method. The corresponding algorithms have important effects. A novel methodology was introduced to extract features. Summed up TF-IDF include vectors were gotten by the presentation of semantic comparability of equivalent words. The

neighbourhood examples of highlight vectors were related to OPSM biclustering calculation [3].

In an indagation utilizing film audits, it was discovered that standard technician learning strategies certainly beats human delivered baselines. Three AI techniques were utilized called Naïve Bayes, greatest entropy order, and backing vector machines performed well on conventional Topic based classification [4].

Electronic business is getting famous because of an enormous number of item surveys. Assessment mining was utilized to catch client surveys. Furthermore, it isolated audits into abstract articulations and target articulations. An epic multi-dimensional model was proposed for assessment mining which incorporated client's attributes and their conclusion about any item [5].

No conclusion has been drawn yet on the multi-domains applicability in Chinese. Comparison was done for ten approaches of aspect opinion extractions on Chinese corpora from seven domains. Compared methods include TF-based models plus POS, CRFs based opinion mining, and SVM bas4d opinion mining CART based opinion mining and LPM-based opinion mining [6].

Term selection methods reduce the size of vocabulary effectively in text categorization for improving quality of classifier. It focuses on identification of relevant terms for each category without affecting the quality of text categorization [7].

An approach was described to object reveal which searches and localize all the occurrences of an object in a video. The object is represented by a set of view point invariant region descriptors so that recognition can proceed successfully despite changes in view point. Efficient retrieval can be achieved by employing methods from statistical text retrieval including inverted file system. In addition, weightings of text and document frequency. The final rankings also depend on spatial layout of the regions [8].

The joint likelihood capacity of groupings of words in a language can be educated as an objective of the factual language model. It is naturally troublesome. Successful yet conventional methodologies dependent on n-grams get speculation by connecting short covering arrangements found in the preparation set. To battle against the scourge of dimensionality a circulated portrayal for words was utilized which permitted each preparation succession to become familiar with the model about an outstanding number of semantically neighbouring sentences [9].

Variations of a neural organization design for factual language displaying have been proposed and applied effectively e.g in the language demonstrating part of the discourse recognizer. They get familiar with an implanting for words in a persistent space that assists with smoothing a language and give better

speculation in any event, when the quantity of preparing models is deficient [10].

Two epic model structures for calculation of persistent vector portrayals of words from extremely enormous datasets were proposed. The quality was estimated in a word closeness task and the outcomes are contrasted with beforehand best-performing strategies dependent on various kinds of neural organizations [11].

With the quick development of internet shopping more clients share their encounters and item audits. Both huge amounts and different types of audits can carry trouble for likely customers to synopsis all heterogeneous surveys for reference. Another positioning framework was proposed through online audits dependent on different parts of elective items. Initially the loads of these angles were resolved with the LDA point model to figure the target notion estimation of the item. At that point purchaser's customized inclinations were taken. A coordinated diagram model was built and the last score was registered utilizing the PageRank calculation [12].

For processing natural language processing tasks continuous word representations were used. Popular methods which learn these representations ignore the wordings morphology. A vector representation is associated to each character n-grams. The method was fast which allowed to train models [13].

Archive feeling arrangement has become a zone of exploration. It tends to be viewed as a unique instance of effective characterization applied distinctly to emotional segments of an archive. Thus the critical errand in archive estimation order m is subjectivity. Approaches existing to remove emotionally depend on etymological assets [14].

Another troupe learning structure was utilized for assessment order of Chinese online surveys. Most importantly as indicated by convoluted attributes of Chinese online audits a grammatical form was separated. Calculation of Random Subspace dependent on data picked up by considering the issue of enormous highlights in the surveys which can improve base classifiers all the while [15].

These days clients search for highlights that help them explicitly yet among a great many audits, it's difficult to track down certain criticisms. The proposed framework followed a semantic-based way to deal with extricate highlights of items. The recursive profound model is utilized to distinguish the estimation direction of surveys [16].

The developing prevalence of conclusion rich assets, for example, survey gatherings for the items have made it hard to pick the correct item. A unique framework was proposed for highlight based synopsis of clients' sentiments for

online items. The last extremity of highlight conclusion sets was determined [17].

A plan was made for Stanford composed conditions to give a basic portrayal of linguistic connections in a sentence that can without much of a stretch be perceived. As opposed to the expression structure portrayal that has since quite a while ago overwhelmed the computational phonetic network [18].

The web has become a great hotspot for social affair client audits. The quantity of surveys got by an item develops fastly. The nature of client audits is surveyed as the most critical. An endeavour was made to survey see dependent on its quality to assist them with settling on a legitimate choice [19].

In another research semantic based approach was used to assist customers and manufacture merchants to mine different products features. And to find opinion summarisation about each extracted features [20].

3. METHODOLOGY

All the data regarding reviews and comments were scraped the website of Daraz. A sum total of 900 comments were scraped from this website and among them 593 are 0's and 307 are 1's. 0 and 1 (positive or negative) was manually set after reading the whole sentence by human. The data set was then filtered and emoticons were removed. All comments are in bangla language.

The methodological steps followed are as follows:

- First comments were gathered from the relevant website.
- Specific sentences were picked from reviews or comments.
- Stop words were anticipated from each sentence.
- By utilizing TF-IDF text sentence data was converted into numerical data.
- Final dataset was obtained by applying PCA.
- A variety of machine learning algorithms were applied on datasets of Python (pycaret).
- At the end dataset was analyse.

The fig 1 is image representation of methodological steps.

At first, all non- English words and stop words were discarded. The positive comments were labelled as numeric 0 and negative comments got catalogued as numeric 1 (you can see it on fig 2). All the non-English words got axed from it by us. Natural Language Toolkit (NLTK) information center of python is being used for this purpose. Ergo, after the moping through the process, on the norm 1983 data set were collected.

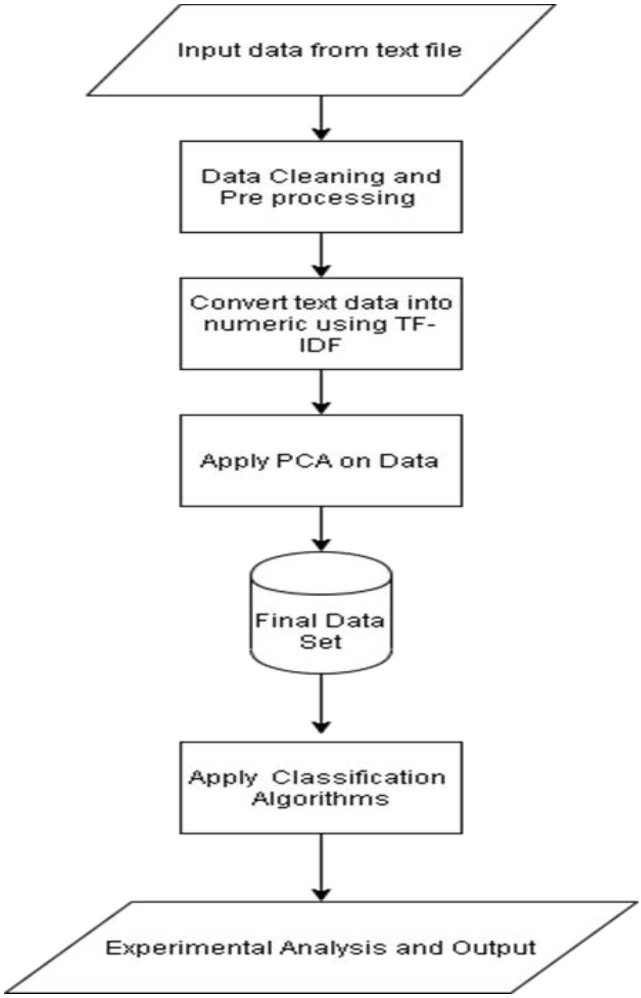


Fig. 1. Vertical Workflow

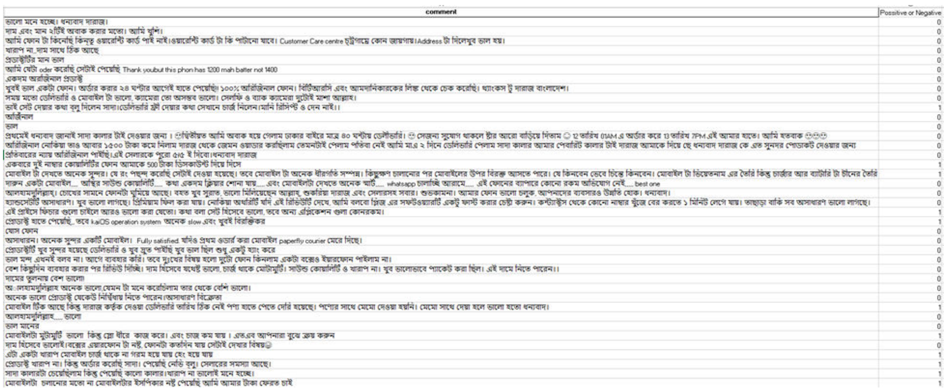


Fig. 2. Fig 2: Data before TH-IDF

An analytical statistic is a numerical or scientific form of statistic which is being contemplated to mirror the principal of word in a docket or corpus and is called Short Term Frequency-Inverse Document Frequency (TF-IDF). This factor has weightage in retrieving information, text mining and user modelling through hunting of this data.

Frequency of a word which pops up in a docket divided by the gross number of words in the document. Every document has its own term frequency.

The log of the documents number divided by word w containing documents. Inverse data frequency determines the weight of rare words across all documents in the corpus.

Most work is being done from Scikit-Learn. Text data is taken by it and converted to numeric information set. After this conversion, our data has 3394 features. Due to so many less important features, features extraction were done using PCA.

1. Principal Component Analysis

A new coordinate system is being metamorphosed from data through orthogonal linear transformation so that each coordinate has greatest variance by scalar projection of data in an ordered way and so on. This is called principal component analysis. Higher variance comes to lie in first coordinate which is called first principal component and the lower variance in second coordinate. Our information set has 1678 features after application of principal component analysis. When applications of dimensions of principal component analysis got reduced and the data quality got lost. In case of principal quality analysis, 95% calibre of data was being maintained. 95% of the quality of real data was preserved by setting value of 'n' components as 0.95. Our latest data has 1678 characteristics after application of principal component analysis.

1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	label
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Fig. 3. Data after TF-IDF

There are 1246 different fields (fig 3) of numeric data in the processed data in which last field signifies 1 for negative comments and 0 for positive comments.

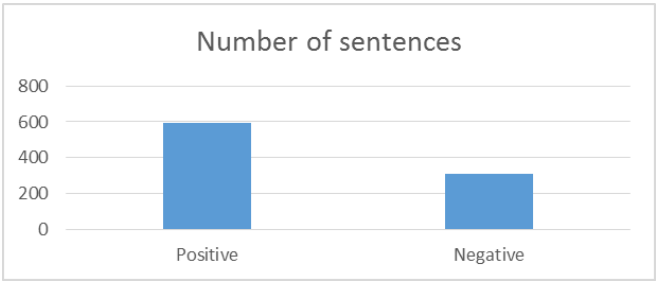


Fig. 4. Number of Positive and Negative reviews

4. EXPERIMENTAL RESULTS

Following are the result obtained after implementation of the dataset in MATLAB: Among 15 classifiers accuracy, prediction speed, training time and total misclassification of top four were used.

Table 1. Summery of Top 4 Classifiers performance

Classifier Name	**Accuracy**	**AUC**	**Recall**	**Precision**	**F1**	**Kappa**	**Time (sec)**
SVM (Linear)	0.7631	0.0000	0.5050	0.7212	0.5919	0.4323	0.1742
Extra Trees Classifier	0.7615	0.7957	0.4305	0.7856	0.5487	0.4057	0.7022
Ridge Classifier	0.7520	0.0000	0.4496	0.7283	0.5524	0.3939	0.0958
AdaBoost Classifier	0.7472	0.7535	0.5149	0.6714	0.5805	0.4045	0.7662

Ridge classifier has the fastest training speed while subspace Adaboost has the slowest as depicted in the training time graph. The Naïve Bayes decision tree was discarded due to low precision.

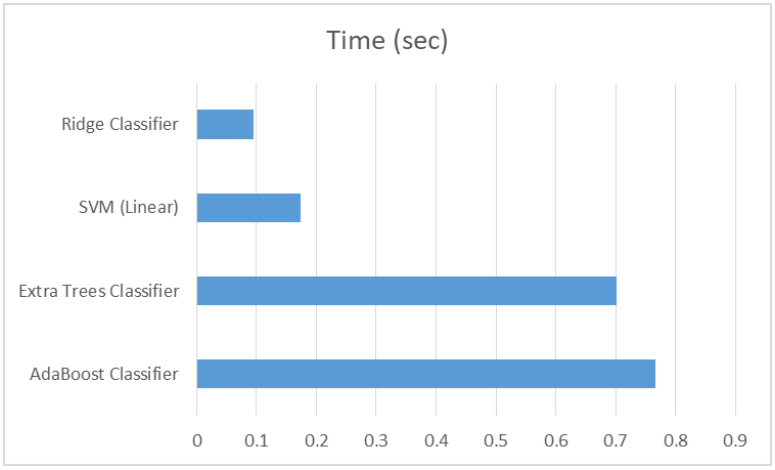


Fig. 5. Time comparison of top 4 performers

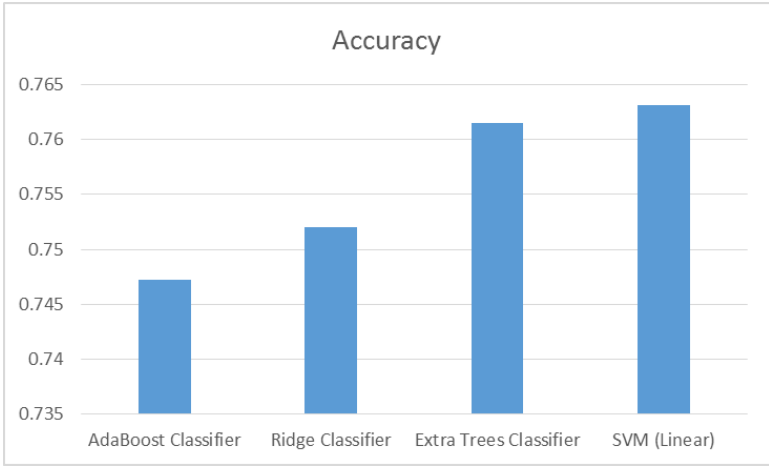


Fig. 6. Accuracy comparison of top 4 performers

Linear SVM gave the highest accuracy followed by Extra Trees Classifier.

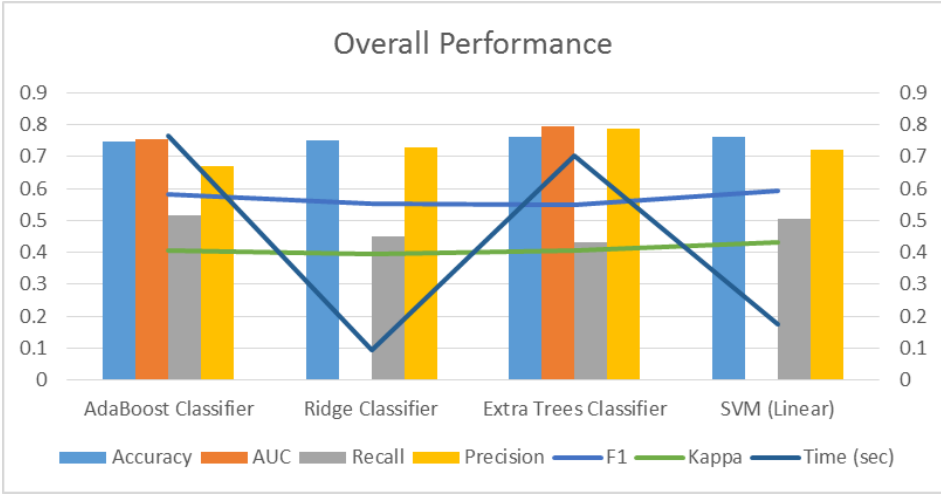


Fig. 7. Overall Performance of top 5 classifiers

Though Ridge classifier has performed best but has low training speed. Linear SVM has high precision with high anticipation speed. So in case of implementation of Python for optimization ridge classifier has done the best function for categorizing positive and negative comments or reviews of Bangla.

Here are the preset values used for the classifiers-

	Description	Value
0	session_id	1877
1	Target Type	Binary
2	Label Encoded	None
3	Original Data	(899, 1248)
4	Missing Values	False
5	Numeric Features	535
6	Categorical Features	712
7	Ordinal Features	False
8	High Cardinality Features	False
9	High Cardinality Method	None
10	Sampled Data	(899, 1248)
11	Transformed Train Set	(629, 1958)
12	Transformed Test Set	(270, 1958)
13	Numeric Imputer	mean
14	Categorical Imputer	constant
15	Normalize	False
16	Normalize Method	None
17	Transformation	False
18	Transformation Method	None
19	PCA	False
20	PCA Method	None
21	PCA Components	None
22	Ignore Low Variance	False
23	Combine Rare Levels	False
24	Rare Level Threshold	None
25	Numeric Binning	False
26	Remove Outliers	False
27	Outliers Threshold	None
28	Remove Multicollinearity	False
29	Multicollinearity Threshold	None
30	Clustering	False
31	Clustering Iteration	None
32	Polynomial Features	False
33	Polynomial Degree	None
34	Trignometry Features	False
35	Polynomial Threshold	None
36	Group Features	False
37	Feature Selection	False
38	Features Selection Threshold	None
39	Feature Interaction	False
40	Feature Ratio	False
41	Interaction Threshold	None
42	Fix Imbalance	False
43	Fix Imbalance Method	SMOTE

Fig. 8. Preset values of the classifiers

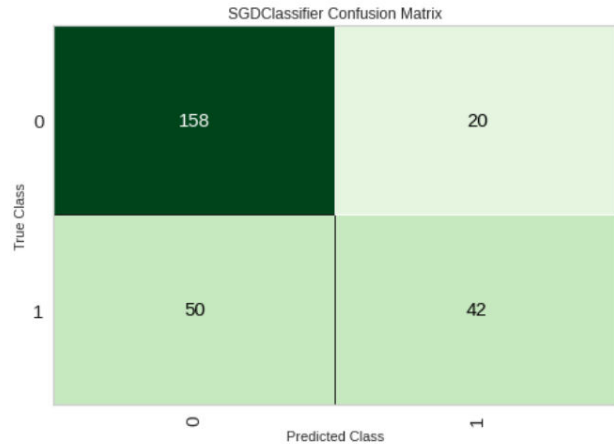


Fig. 9. Confusion Matrix of Linear SVM

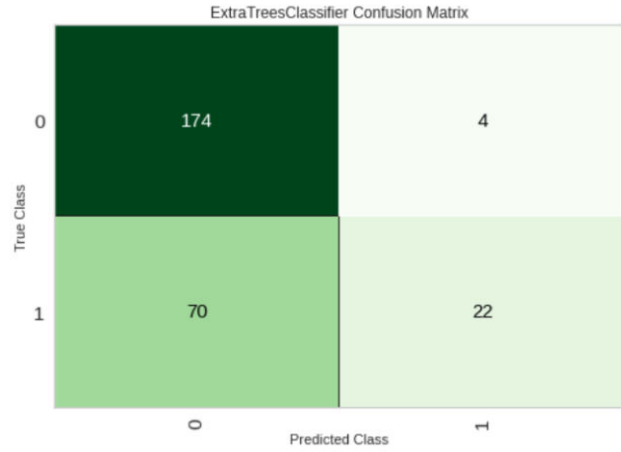


Fig. 10. Confusion Matrix of Extra Trees Classifier

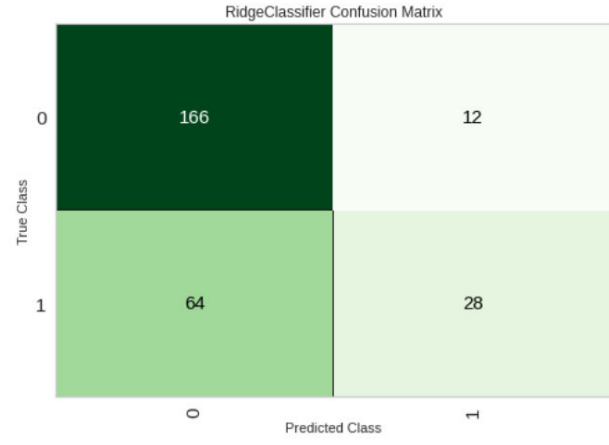


Fig. 11. Confusion Matrix of Ridge Classifier

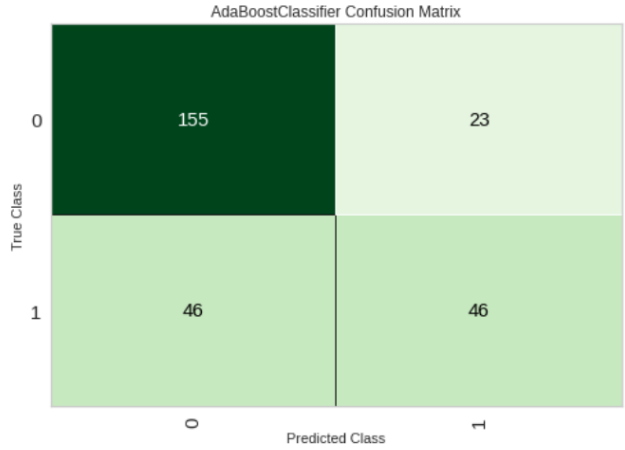


Fig. 12. Confusion Matrix of Adaboost Classifier

From the confusion matrix, its clarified Ridge classifier has high value as well as for extra trees classifier. ROC curve of various classifiers are given in Fig 13.

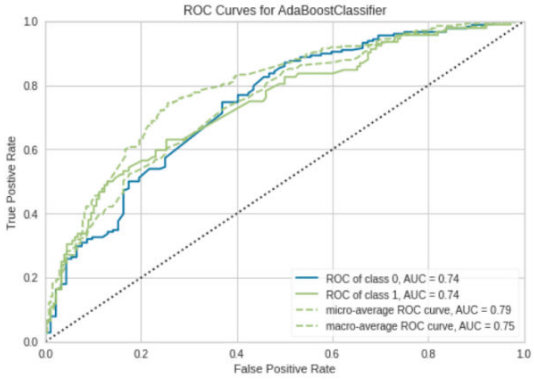


Fig. 13. Adaboost Classifier

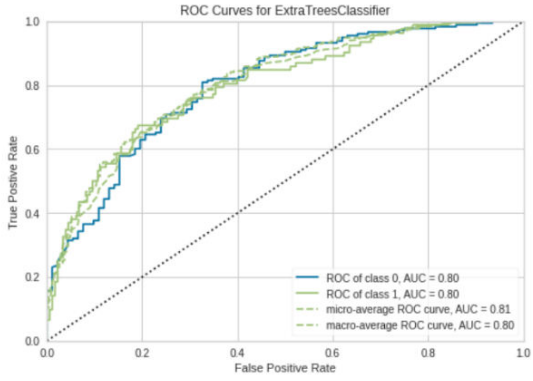


Fig. 14. Extra Trees Classifier

For ROC curves a steeper curve depicts better output and Extra trees classifier has steeper curves.

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7302	0.0000	0.4286	0.6429	0.5143	0.3377	0.3510
1	0.6984	0.0000	0.3810	0.5714	0.4571	0.2597	0.2700
2	0.7460	0.0000	0.5238	0.6471	0.5789	0.4000	0.4046
3	0.7619	0.0000	0.4762	0.7143	0.5714	0.4156	0.4320
4	0.6032	0.0000	0.1905	0.3333	0.2424	0.0000	0.0000
5	0.7619	0.0000	0.5455	0.7059	0.6154	0.4470	0.4548
6	0.7460	0.0000	0.5909	0.6500	0.6190	0.4292	0.4303
7	0.7143	0.0000	0.5000	0.6111	0.5500	0.3438	0.3475
8	0.7619	0.0000	0.4091	0.8182	0.5455	0.4075	0.4525
9	0.7419	0.0000	0.4286	0.6923	0.5294	0.3649	0.3848
Mean	0.7266	0.0000	0.4474	0.6386	0.5224	0.3405	0.3527
SD	0.0457	0.0000	0.1059	0.1200	0.1038	0.1249	0.1295

Fig. 15. Linear SVM

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.8095	0.8299	0.4286	1.0000	0.6000	0.5000	0.5774
1	0.7143	0.7659	0.1905	0.8000	0.3077	0.2059	0.2907
2	0.7778	0.8197	0.4286	0.8182	0.5625	0.4324	0.4730
3	0.7460	0.8724	0.2857	0.8571	0.4286	0.3143	0.3929
4	0.6825	0.6383	0.1905	0.5714	0.2857	0.1429	0.1786
5	0.7460	0.8492	0.3636	0.8000	0.5000	0.3604	0.4107
6	0.7460	0.7772	0.2727	1.0000	0.4286	0.3280	0.4429
7	0.7619	0.7561	0.4091	0.8182	0.5455	0.4075	0.4525
8	0.6825	0.7882	0.1364	0.7500	0.2308	0.1382	0.2189
9	0.7258	0.8159	0.3810	0.6667	0.4848	0.3165	0.3395
Mean	0.7392	0.7913	0.3087	0.8082	0.4374	0.3146	0.3777
SD	0.0378	0.0619	0.1032	0.1248	0.1194	0.1147	0.1158

Fig. 16. Extra Trees Classifier

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7143	0.0000	0.3333	0.6364	0.4375	0.2703	0.2957
1	0.7302	0.0000	0.2857	0.7500	0.4138	0.2817	0.3371
2	0.6984	0.0000	0.3810	0.5714	0.4571	0.2597	0.2700
3	0.7302	0.0000	0.2857	0.7500	0.4138	0.2817	0.3371
4	0.6667	0.0000	0.2857	0.5000	0.3636	0.1600	0.1715
5	0.7778	0.0000	0.4091	0.9000	0.5625	0.4404	0.5019
6	0.7143	0.0000	0.4091	0.6429	0.5000	0.3136	0.3293
7	0.7619	0.0000	0.5000	0.7333	0.5946	0.4345	0.4504
8	0.6984	0.0000	0.3182	0.6364	0.4242	0.2495	0.2770
9	0.6935	0.0000	0.4286	0.5625	0.4865	0.2737	0.2789
Mean	0.7186	0.0000	0.3636	0.6683	0.4654	0.2965	0.3249
SD	0.0313	0.0000	0.0696	0.1108	0.0678	0.0799	0.0890

Fig. 17. Ridge Classifier

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7302	0.8475	0.5714	0.6000	0.5854	0.3855	0.3858
1	0.6667	0.6944	0.3810	0.5000	0.4324	0.2025	0.2063
2	0.7302	0.7942	0.5238	0.6111	0.5641	0.3704	0.3727
3	0.8254	0.7999	0.5238	0.9167	0.6667	0.5600	0.6002
4	0.6508	0.6349	0.3810	0.4706	0.4211	0.1750	0.1770
5	0.8095	0.8049	0.7273	0.7273	0.7273	0.5809	0.5809
6	0.6825	0.6336	0.4545	0.5556	0.5000	0.2708	0.2738
7	0.6984	0.7306	0.5000	0.5789	0.5366	0.3148	0.3167
8	0.6984	0.7284	0.4091	0.6000	0.4865	0.2837	0.2941
9	0.6774	0.5929	0.4286	0.5294	0.4737	0.2448	0.2477
Mean	0.7169	0.7261	0.4900	0.6090	0.5394	0.3389	0.3455
SD	0.0557	0.0816	0.1003	0.1225	0.0940	0.1316	0.1373

Fig. 18. Adaboost Classifier

The tuning of results showed that having low training time as in the case of Ridge classifier its variant can give 77.78% accuracy and this is the reason Ridge classifier is the best.

5. CONCLUSION

As the results depicted that after gathering all data different algorithms were applied among which Ridge classifier has the fast training time than subspace Adaboost while Naïve Bayes was discarded due to its inferior performance. Moreover the tuning of these classifiers also inferred that regardless of the training time Ridge classifier can give the precision of up to 77.8% and this makes it the best classifier to be used for categorizing comments or reviews in any business.

REFERENCES

1. L. J. a. Y. Tsai, "Using Text Mining of Amazon Reviews to Explore User-Defined Product Highlights and Issues", in Int'l Conf. Data Mining , CA,USA, 2015.
2. Y.-J. Zhi-Ping, "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis", Journal of Business Research, pp. 90-100, 2017.
3. 3. ·. Y. X. ·. H. Z. ·. X. L. ·. X. H. ·. Z. M. Xin Chen1, "A novel feature extraction methodology for sentiment analysis of product reviews", Neural Computing and Applications, 2018.
4. B. P. a. L. Lee, "Thumbs up? Sentiment Classification using Machine Learning Techniques", in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, 2002.

5. l. R. Y. Y. L. Zhang, "Integration of Sentiment Analysis into Customer Relational Model: The Importance of Feature Ontology and Synonym", Procedia Technology, vol. 11, pp. 495-501, 2013 .
6. G. T. &. H. W. Wei Wang, "Cross-domain comparison of algorithm performance in extracting aspect-based opinions from Chinese online reviews", International Journal of Machine Learning and Cybernetics, pp. 1053-1070, 2016.
7. T. B. &. C. A. Murthy, "A supervised term selection technique for effective text categorization", International Journal of Machine Learning and Cybernetics, vol. 7, pp. 877-892, 2015.
8. J. Sivic and A. Zisserman, "Efficient Visual Search of Videos Cast as Text Retrieval", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp. 591-606, April 2009, doi: 10.1109/TPAMI.2008.111.
9. Y. Bengio, "A Neural Probabilistic Language Model", Journal of Machine Learning Research, Vol. 3, 2003.
10. F. Morin, "Hierarchical Probabilistic Neural Network Language Model", in Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics, 2005.
11. K. C. G. C. J. D. Tomas Mikolov, "Efficient Estimation of Word Representations in Vector Space", Cornell university, vol. 3, 2013.
12. C. G. Z. D. X. Kou, "Products Ranking Through Aspect Based Sentiment Analysis Of Online Heterogeneous Reviews", vol. 5, no. 27, pp. 542-558, 2018.
13. E. G. A. J. T. M. Piotr Bojanowski, "Enriching Word Vectors with Subword Information", 2017.
14. K. Sarvabhotla, "Sentiment classification: a lexical similarity based approach for extracting subjectivity in documents", Information Retrieval, vol. 14, pp. 337-353, 2011.
15. Y. X. Jiafeng Huang, "Sentiment analysis of Chinese online reviews using ensemble learning framework", Cluster Computing, vol. 22, pp. 3043-3058, 2018.
16. N. Devasia and R. Sheik, "Feature extracted sentiment analysis of customer product reviews", International Conference on Emerging Technological Trends (ICETT), Kollam, 2016, pp. 1-6, doi: 10.1109/ICETT.2016.7873646.
17. K. B. DurgaToshniwal, "Feature based Summarization of Customers' Reviews of Online Products", Procedia Computer Science, vol. 22, pp. 142-151, 2013.
18. M.-C. d. M. a. C. D. Manning, "Stanford typed dependencies manual", 2008.

19. S. S. Prakash Hiremath, "Cluster Analysis of Customer Reviews Extracted from Web Pages", Journal of Applied Computer Science & Mathematics, vol. 9, no. 4, 2010.
20. R. k. V. a. K. Raghuveer, "Web User Opinion Analysis for Product Features Extraction and Opinion Summarization", International Journal of Web & Semantic Technology (IJWesT), vol. 3, no. 4, 2012.

Cite this article

Biplob Kumar, Pritom Sarker and Nakib Aman Turzo, Application of Classifiers for Assortment of Online Reviews, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 67-82

6 The Method of Depth Map Calculating Based on Soft Operators in Multi-Agent Robotic Stereo Vision Systems

M. V. Bobyr[a,*], N. A. Milostnaya[a], S. V. Gorbachev[b], S. Bhattacharyya[c] and J. Cao[d]

[a]Department of Computer Science, Southwest State University, Kursk, Russia
[b]National Research Tomsk State University, Tomsk, Russia
[c] Rajnagar Mahavidyalaya, Birbhum, India
[d]Southeast University, Nanjing, China and Yonsei University, Seoul, Korea

*Corresponding author: maxbobyr@gmail.com

ABSTRACT

The fuzzy method of depth map calculating using stereo images obtained on the path of mobile robots (agents) is considered. The method is based on SAD (sum of absolute difference) algorithm composition and fuzzy inference. Its special feature is soft arithmetic operators with fuzzy implication usage. The accuracy of the constructing depth map method is estimated by RMSE (root mean square error). The bestest soft operator has minimum RMSE. The method of depth map calculating which has seven steps is presented. The proposed method has showed that the accuracy of the SAD algorithm increases by 20% when soft operators is used. This conclusion is confirmed by the simulation results presented in the chapter.

Keywords: Mobile Robots, Stereo Vision, Depth Map, Soft Computing, SAD, Fuzzy Logic

Advanced Research in Computer Engineering
Book Series: Research Transcripts in Computer, Electrical and Electronics Engineering

1. INTRODUCTION

The development and development of mobile robotic systems (MRS) and the system is going on very intensively in today's time. Navigation problems are one of the main problems encountered by mobile robots. Localization, mapping, and route planning are important parts of the autonomous device navigation task [1]. Localization is responsible for accurately determining the current position of the robot in the surrounding space. The maps are built by collecting the robots ' sensory data and storing it in a form that is convenient for subsequent processing. Using the data obtained during the construction of the map, it becomes possible to plan a route that allows you to reach the whole from the starting position, without allowing a collision with obstacles. The Simultaneous Localization and Mapping (SLAM) method combines the two processes to provide more efficient offline navigation. SLAM algorithms are able to cope with various hardware or external constraints, for example, SLAM methods for monocular systems [2], SLAM methods using timely lidar and visual sensors [3].

In many cases, the capabilities of an individual robot will not be sufficient to solve the tasks set [15]. There is a need to use mobile robotic groups. One of the important tasks that is performed in the management of such a multi-agent system is the formation and support of robots of a given order. Methods of group control of unmanned aerial vehicles, methods based on the use of computer vision systems, neural networks and fuzzy logic [16], and advanced robotics are being actively developed. Recklessness on a large number of existing solutions to problems of movement of groups of robots system, when implemented, problems can be with the limitation of available computing resources of the mobile robot, and each robot and groups in general.

The use of SLAM and autonomous navigation methods when controlling moving groups of MRS formations requires the addition of the motion control system of each robot with functionality related to obtaining and analyzing information from the surrounding space. In this case, the obstacle detection subsystem is of key importance, since obstacles play the role of "anchors" for linking the terrain map under construction. Solving the problems of detecting obstacles when navigating outdoors requires recalling up-to-date information about the surrounding space in the form of three-dimensional dynamic images.

A large research and publication concerns monocular vision and laser sensors. From where stereo vision technologies take on additional benefits. The hot range information preceding the stereo vision is more accurate than the information preceding the laser sensors, it can be used effectively to calculate the three-dimensional placement and configuration of the object at a lower cost.

In addition, in a complex noisy environment, stereo vision can recognize the necessary volume, reducing such cheeks sliding with background elements. Secondly, it is expected that systems using stereo vision will be more reliable in real scenes, where sudden changes in awareness often occur.

Problems associated with determining the location of moving and / or stationary obstacles on the path of a mobile robot arise during the development of a control system for it [4, 5]. To determine the relative position of objects, it is necessary to process data about the depth of the video scene and obstacles located along the path of the robot. Various methods are used to construct such scenes, for example, obtaining data from scanning lidars or using stereo vision systems [6]. It is necessary to solve a number of problems associated with finding objects located on stereo images during of developed of the stereo vision systems [7]. A depth map is a black and white image with grayscale color, in which the brightness of individual objects indicates the proportional distance to them. The main algorithms for constructing depth maps are SGM (Semi-Global Matching Algorithm) and SAD (Sum of Absolute Differences algorithm) [8]. Several passes are required to calculate the depth map using the SGM algorithm, while the second algorithm uses only one pass along the image, which reduces the processing time required to calculate the disparity value. It should be noted that 1 frame of stereo information in HD format (720p - resolution 1280 × 720 pixels) should be processed in 1/30 second, that is, in 33 ms. Therefore, it takes about 50 ns to process 1 pixel. Therefore, decision-making time is one of the main factors that must be considered when developing stereo systems for controlling mobile robots. Fuzzy logic is one of the ways to improve the accuracy of the stereo vision model [9]. However, fuzzy models have a number of system errors that reduce their accuracy [10]. It is recommended to use soft arithmetic operators during the fuzzy implication to eliminate these errors.

2. FOUNDATIONS

The obstacle detection subsystem for MRS navigation in a dynamically changing environment must meet stringent requirements. The need for a moving mobile robot to respond to changing environments in a timely manner, especially at high speeds, implies that the obstacle detection and classification system operates in real time. This implies the necessary maximum performance in conditions of operation on an on-board computer with limited computing power. Increased performance can be achieved by parallelizing data processing algorithms, so it is desirable to perform parallel execution of more resource-intensive processing steps to improve performance. Also, the dynamically

changing environment imposes the required size of the guaranteed radius of detection of dangerous obstacles. This value must exceed the length of the braking distance of the mobile robot at the maximum possible speed, including in bad weather conditions. Terrain patency at short and medium distances can be estimated using existing solutions, to minimize trajectory planning errors, the priority duty is to be obstacles at a long distance. In addition to the coordinates of the detected obstacles to plan the motion path of the robot is important additional information about the dimensions, the characteristics of the object and its overcome (for example, you can pass through low bushes, but the attempt to overcome tree trunk can result in the release of MRS of the car). The obstacle detection subsystem should allow you to get an estimate of the specified parameters for the detected objects. The sensor system used to obtain three-dimensional information about the surrounding space should ensure the functioning of the obstacle detection subsystem at your favorite time of day and have the maximum effective range performance. The sensor system should transmit the resulting three-dimensional spatial information in the video of a three-dimensional image (point clouds) - the multiplication of the coordinates of the vertices, the spatial mood that characterizes the surrounding objects. Cloud points (vertices) can carry additional information about the power characteristics of an object, such as blooming or reflectivity. To process the incoming information, it is necessary to use the methods of three-dimensional computer vision.

Often, certain methods and algorithms cannot always be successfully applied in solving traditional problems of technical vision, there are no unified methods, which are universal. This is due to the fact that the technical vision systems of robot must satisfy hard trackers, which have a decently complex implementation, besides, no sensor appears ideal for solving a certain task.

From the point of view of the implementation of the methods, there are a number of additional requirements for the technical vision system. In contrast to the algorithms for analyzing time signals, processing and analysis of images are more complex, since they represent three coordinates and time.

In the process of analyzing the literature revealed that this analysis of images is often considered as unrelated processes. It is necessary to conduct a comprehensive analysis throughout the study, while loading 6the speed of the newly developed algorithms or modifying them. In offers segmentation methods, describes the possibility of direct access to the source code. But the methods described in ra6ot are not suitable for systems that operate in real time.

Today, complex hardware and software tools (~~APS)~~ are used in three-dimensional technical vision systems, which affects the production processing methods for solving the problem in real time.

These problems can be solved by developing soft fuzzy algorithms and creating new methods of processing to extract such data as flexible in this respect. To create high-performance 3D technical vision systems, it is necessary to exclude the operation of matching stereo pairs when analyzing frames, to ensure the receipt of spatial video information by a single mobile video sensor, to increase the noise immunity and to develop a simple software o6 support – to synthesize a three-dimensional image, it is necessary to determine a strategy for images obtaining that lack projections of the object based on the results of images obtaining of the original object. To the basics of these confusions, it can be argued that the capabilities of the currently existing means of forming and image processing are limited, so when developing a video system, it is necessary to load both a trace to the completeness and quality of the information posted by them, and the possible wearing of their constituent elements.

In order to perform the task in real time, the algorithms used should be more simple and allow for efficient computational implementation. But the high range of tasks and the desire for the universality of the system force us to abandon simple algorithmic solutions and switch to complex methods of processing and analysis of video information. It is also worth noting that it is possible to organize the system's flexibility for calculating various different algorithms with a large number of variable parameters. It is also necessary to note that the operators of robots are often ordinary users, not highly qualified specialists, and often they do not fully control the mathematical methods of video information processing, as well as programming. The above-mentioned specifics of solving technical problems actualizes the development of completely new hardware and software tools, which will effectively supported all the necessary modes of operation of the robot technical vision systems.

Based on the above, we can draw the following conclusion: when developing multi-agent stereo vision systems in complex environmental conditions, the range of tasks to be solved with the help of robot is significantly complicated and, satisfactorily, the need for the development of a universal technical vision system, equipped with satisfying means for solving complex functional problems arising from the conditions of work, increases. At the same time, the defining qualities of the functional should be the speed, accuracy and flexibility of information processing methods when solving the problem in real time.

2.1. Statement of the problem

Let's set the task of improving the accuracy of the stereo vision system for constructing depth maps based on the estimate of the RMSE (root mean square error) coefficient:

$$RMSE = \sqrt{\frac{1}{w \times h} \sum_{i=1}^{h-1} \sum_{j=1}^{w-1} \left(I'_{i,j} - I_{i,j} \right)^2} \to \min, \quad (1)$$

where w, h are width and height of the image; $I'_{i,j}$ is brightness level on the reference image; $I_{i,j}$ is brightness level in the output image.
For the study, we will use soft operators with the following formulas:

Operator	Formulas for fuzzy implication	
MIN	$I = \min(a, b)$	(2)
PROD	$I = a \times b,$	(3)
MEAN	$I = \frac{a+b}{2},$	(4)
Soft-min I	$I = \frac{a+b+\delta^2 - \sqrt{(a-b)^2 + \delta^2}}{2},$	(5)
Soft-min II	$I = \frac{\sum_{i=1}^{n} a_i e^{-k a_i}}{\sum_{i=1}^{n} e^{-k a_i}},$ где $k = -100.$	(6)

where a, b are operands; δ is coefficient of softness (δ = 0.05); n is the number of operands.
The most accurate method is the one with the minimum RMSE.

3. RESEARCH METHODOLOGY

Fuzzy method for constructing a depth map from stereo images

The calculation of fuzzy method for constructing a depth map from stereo images is performed in several steps.
Step 1. The calculation of the difference in intensity levels on stereo images:

$$SAD(x, y, d) = \sum_{m=-1}^{m=1} \sum_{n=-1}^{n=1} R \sum_{m=-1}^{m=1} \sum_{n=-1}^{n=1} G \sum_{m=-1}^{m=1} \sum_{n=-1}^{n=1} B \left| I_l(x+m, y+n) - I_r(x+m+d, y+n) \right| \quad (7)$$

where I_l, I_r are color intensity on the left and right stereo images; x, y are coordinates on stereo images; R, G, B - components of the three-channel color model.

The calculation the disparity from array of the difference in intensity levels on stereo images:

$$D(x,y)_{SAD} = \arg\min_{d\in[0\ldots d_{\max}]} SAD(x,y,d). \tag{8}$$

The SAD algorithm in graphically view is presented in Figure 1.

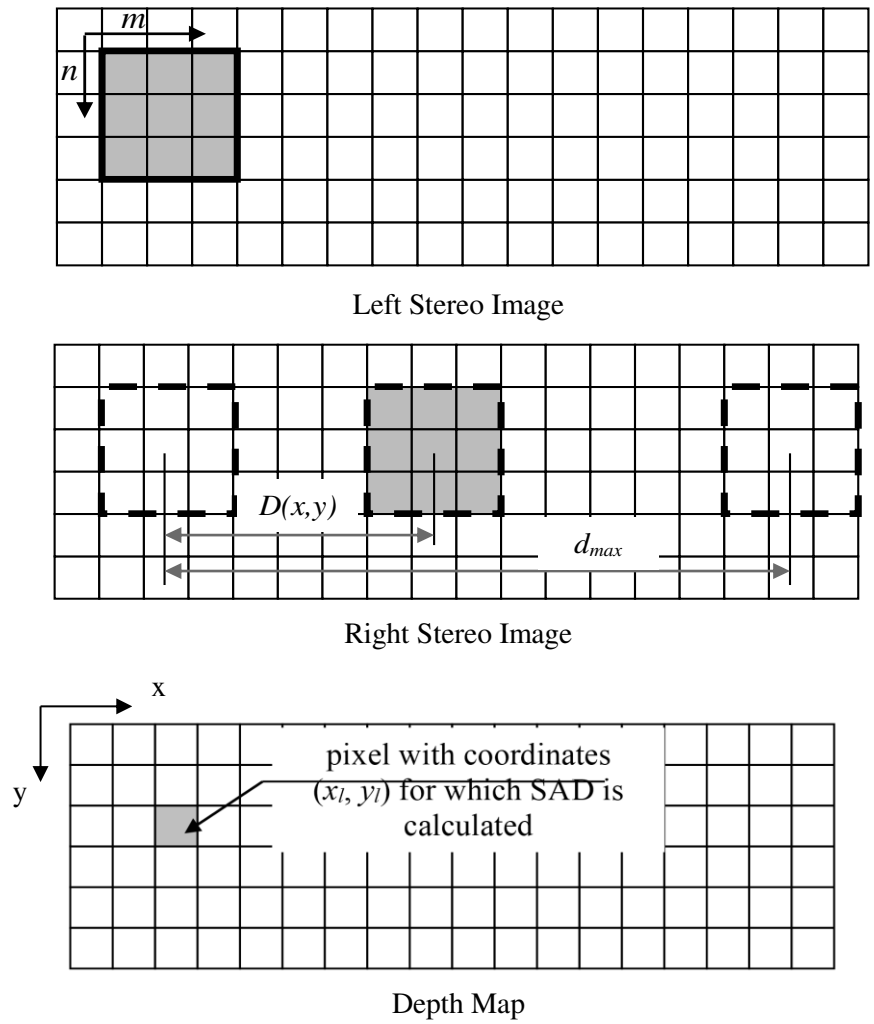


Fig. 1. Graphical interpretation of the SAD algorithm

Step 2. Calculation of the degrees of input membership functions (MF). In order to increase the speed of the fuzzy multi-criteria decision-making (FMCDM) system and implement it on the basis of Field-Programmable Gate Array

(FPGA) it is recommended to use the type of MF for calculating the degrees of input membership functions which fewer arithmetic operations are needed. For instance, to calculate triangular MF, only subtraction and division operations are required. The subtraction operation in FPGAs is simple. To implement the division operation, it is recommended to select the difference between the MF labels to be a multiple of a power of 2 (2, 4, 8, 16, 32 ...). Then the division operation is replaced by a right shift operation. Input MFs in FMCDM are shown in Figure 2, a.

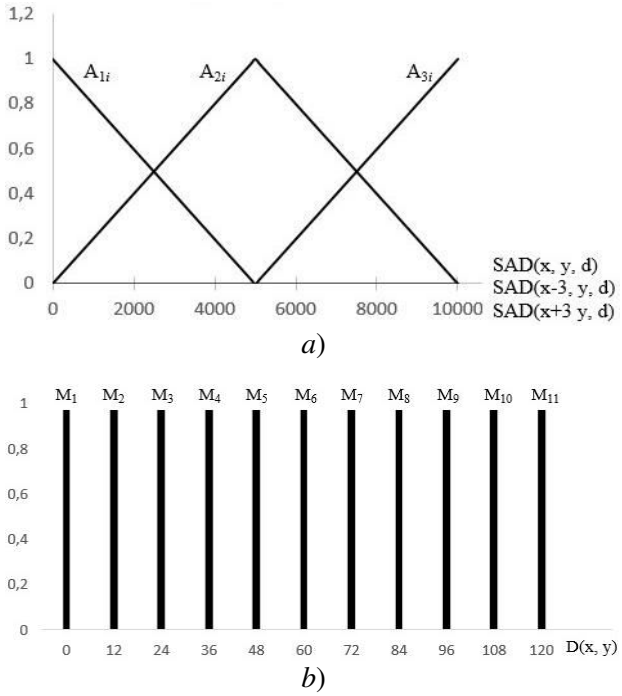


Fig. 2. Membership functions: a – input MF; b – output MF

The degrees of the membership functions of the input variables are calculated using the following formulas:

$$A_{1i} = \begin{cases} (5000 - SAD(x,y,d))/5000 \; if \;\; SAD(x,y,d) < 5000 \\ 0, else \end{cases}$$

$$A_{2i} = \begin{cases} SAD(x,y,d)/5000 \; if \;\; SAD(x,y,d) < 5000 \\ (10000 - SAD(x,y,d))/5000 \; if \;\; SAD(x,y,d) \geq 5000 \\ 0, else \end{cases} \quad (9)$$

$$A_{3i} = \begin{cases} (SAD(x,y,d) - 5000)/5000 \; if \;\; SAD(x,y,d) > 5000 \\ 0, else \end{cases}$$

where i is the number of the input variable (i = 1 ... 3).

Step 3. Calculation degree of conclusions of fuzzy rules. In general, the fuzzy rule has the form [12, 13]. The degree of conclusions will be calculated using the formulas shown in Table 1.

Table 1. Degrees of conclusions of fuzzy rules

The conclusions of fuzzy rules
$Y_1 = \Theta(A_{31}, A_{32}, A_{33})$
$Y_2 = \max([\Theta(A_{31}, A_{32}, A_{23})], [\Theta(A_{31}, A_{22}, A_{33})])$
$Y_3 = \max([\Theta(A_{31}, A_{32}, A_{13})], [\Theta(A_{31}, A_{22}, A_{23})], [\Theta(A_{31}, A_{12}, A_{33})])$
$Y_4 = \max([\Theta(A_{31}, A_{12}, A_{13})], [\Theta(A_{31}, A_{12}, A_{23})], [\Theta(A_{31}, A_{22}, A_{13})])$
$Y_5 = \max([\Theta(A_{21}, A_{22}, A_{33})], [\Theta(A_{21}, A_{32}, A_{23})], [\Theta(A_{21}, A_{32}, A_{33})])$
$Y_6 = \max([\Theta(A_{21}, A_{12}, A_{33})], [\Theta(A_{21}, A_{22}, A_{23})], [\Theta(A_{21}, A_{32}, A_{13})])$
$Y_7 = \max([\Theta(A_{11}, A_{32}, A_{33})], [\Theta(A_{21}, A_{12}, A_{23})], [\Theta(A_{21}, A_{22}, A_{13})])$
$Y_8 = \max([\Theta(A_{11}, A_{22}, A_{33})], [\Theta(A_{11}, A_{32}, A_{23})], [\Theta(A_{21}, A_{12}, A_{13})])$
$Y_9 = \max([\Theta(A_{11}, A_{12}, A_{33})], [\Theta(A_{11}, A_{22}, A_{23})], [\Theta(A_{11}, A_{32}, A_{13})])$
$Y_{10} = \max([\Theta(A_{11}, A_{12}, A_{23})], [\Theta(A_{11}, A_{22}, A_{13})])$
$Y_{11} = \Theta(A_{11}, A_{12}, A_{13})$

Note. The Θ sign means a fuzzy-logical operation of finding a hard or soft minimum using formulas (2)÷(6).

Step 4. Defuzzification.

$$D(x,y)_{d=1}^{d_{\max}} defuz = \frac{\sum_{k=1}^{11} Y_k \cdot M_k}{\sum_{k=1}^{11} Y_k}. \tag{10}$$

.

Step 5. Calculation of disparity value on the depth map.

$$D(x,y) = \arg\min_{d \in [0 \ldots d_{\max}]} \left(D(x,y)_{d=1}^{d_{\max}} defuz \right). \tag{11}$$

4. EXPERIMENTAL RESULTS

Stereo images with a size of 450÷375 pixels (Fig. 3) were used for simulate the process of constructing a depth map [14]. This makes it possible to assess visually and quantitatively the effectiveness of the proposed FMCDM. The fuzzy inference system is formed on the basis of the MFs shown in Figure 2. The maximum depth value d_{max} = 64. The disparity values (see Eq. 11) were multiplied by 4 to correspond to 256 shades of grayscale. In the MS Visual Studio 2013 on the C # programming language was developed a software model that allows building depth maps and evaluating their accuracy using formula (1). Studies were carried out for the SAD method, fuzzy operators (2)÷(6) according to the method presented in [14], and using the FMCDM from stereo images. Table 2 summarizes the values of the RMSE coefficient.

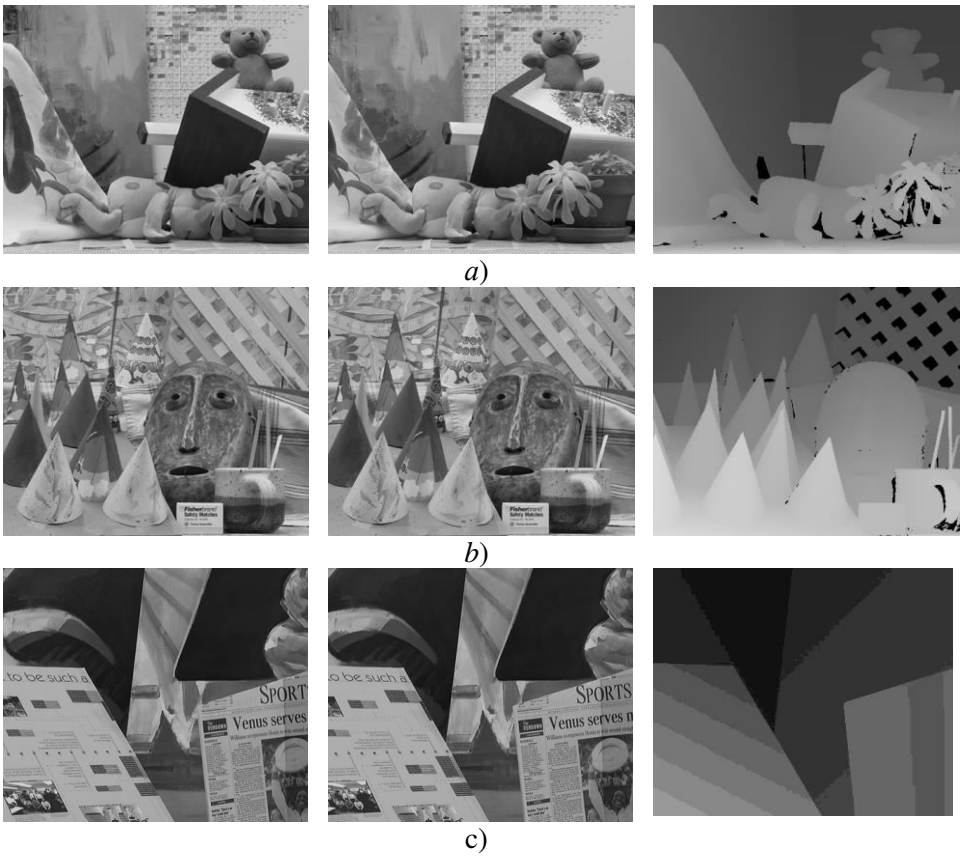


Fig. 3. Stereo images: a – Teddy; b – Cones; c – Venus

Note: The figure shows the left and right stereo images and a reference depth map.

It should be noted that in the methodology [11], only a hard operator was used to construct the depth map in the compositional rule of Zadeh.

Table 2. Value of RMSE

Image	Method	Operator	MIN	PROD	MEAN	Soft-MIN I	Soft-MIN II
Teddy	SAD	RMSE	47.77				
	[11]	RMSE	44.91	46.74	***40.98***	41.09	38.42
	FMCDM	RMSE	38.65	38.86	38.09	***37.76***	38.32
Cones	SAD	RMSE	51.79				
	[11]	RMSE	50.81	52.76	47.12	39.77	***39.7***
	FMCDM	RMSE	32.43	32.83	31.89	***31.64***	32.45
Venus	SAD	RMSE	50.07				
	[11]	RMSE	46.64	46.64	46.05	***45.92***	46.16
	FMCDM	RMSE	41.44	44.94	***38.8***	39.75	39.63

Note. The minimum RMSE values are in bold italic type.

As shown by the data presented in Table 2, when using soft operators, the accuracy of the method [11] increases. Table 3 shows video images of depth maps obtained using the proposed FMCDM. The visualization of the depth map using soft operators (2) ÷ (6) is also shown.

Analysis of Table 4 showed that all images contain artifacts in the form of white dots. White dots on the depth map show objects that are closest to stereo cameras. For example, when a mobile robot avoids obstacles, white dots will give false objects that are in front of its path.

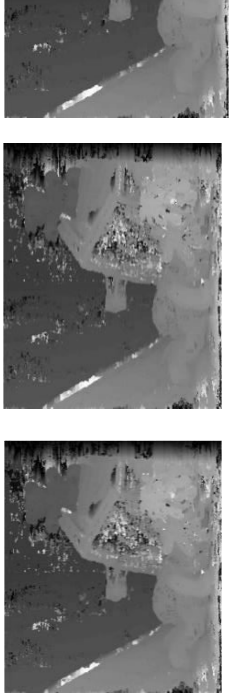
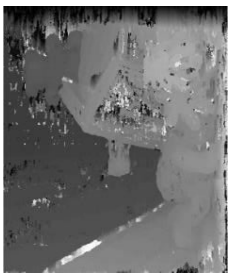
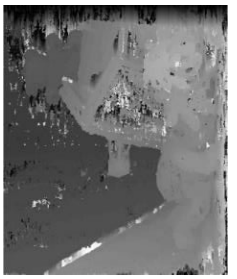
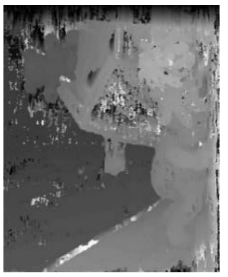
To compensate for artifacts, it is proposed to use the following method.

1. It is necessary to evaluate the data that comes to the input of the fuzzy system in areas with artifacts. In Figure 4 (a) artifact has coordinates x = 254, y = –68. The value of the disparity at this point is equal 62. This value is predicted by analyzing the input data presented in Figure 4(b). Taking into account formula (11), the disparity is equal to the value of the fourth local minimum. Scaling 62 × 4 = 248 produces a light tone.
2. It is necessary to estimate the real value of disparity on the depth map. Visual assessment (Fig. 4, b) shows that the correct value of disparity on the depth map should correspond to the first minimum and be equal to 16.
3. Exclude from the calculation the false zones of fuzzy inference triggering. To ensure this task, it is necessary to introduce an additional condition for calculating the disparity value. Taking into account the data presented in Figure 4,b, such a condition is the rule:

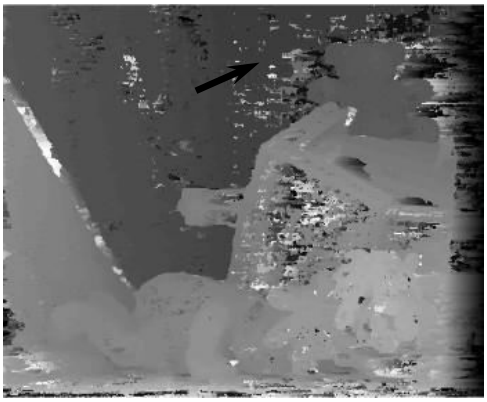
   IF ($SAD(x, y, d)$>1200) AND ($D(x,y)$≥32)

   THEN A_i = 1000

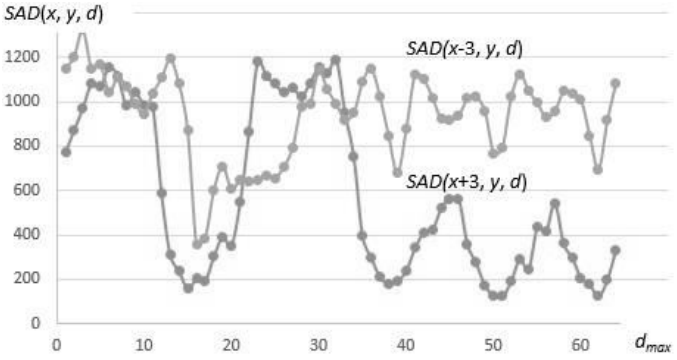
Table 3. Comparative analysis of the output images

Methodology [11]	MIN	PROD	MEAN	Soft-MIN I	Soft-MIN II
FMCDM					

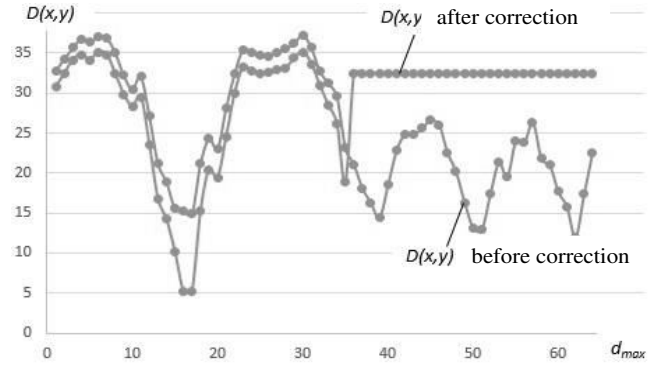
This compensating rule makes it possible to reduce the appearance of maximum values of the value D (x, y) determined by formula (11) (Fig. 4, c). Using the proposed rule a value equal to 16 is formed at the output of the fuzzy system.



(a)



(b)



(c)

Fig. 4. Depth map: a – Teddy; b – before correction D(x,y); c – after correction D(x,y)

Visualization of the depth map using the compensating rule is shown in Figure 5. Analysis of Figure 5(c) shows that there are no white dots in the resulting image.

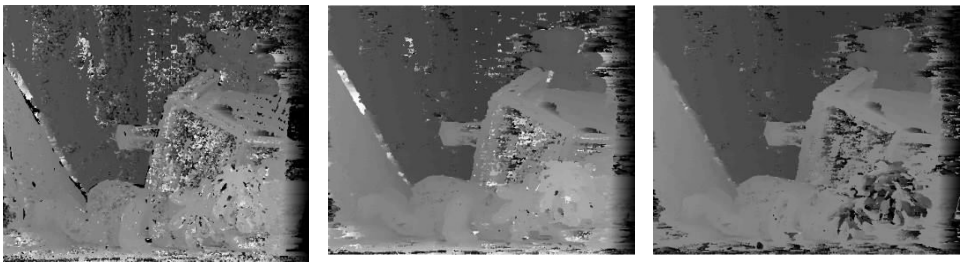


Fig. 5. Correction: a – algorithm SAD; b – operator Soft-min I without correction rule; c – operator Soft-min I with correction rule

5. CONCLUSION

Analysis of modeling the process of constructing a depth map using the proposed method showed that:

1. The use of soft operators based on the analysis of RMSE increases the accuracy of the FMCDM. Thus, the fuzzy method for constructing a depth map for the three analyzed images proposed has an advantage over the SAD algorithm by 24%, compared with the prototype method by 11%.
2. When soft operators are used in the compositional rule, the accuracy of the fuzzy model increases. The best RMSE values were obtained using the following soft operators: MEAN, Soft-min I, and Soft-min II.
3. For eliminate artifacts, it is necessary to introduce a compensating rule in the structure of fuzzy inference, which limits the possible value of input of fuzzy model.
4. The smallest operating time of the fuzzy inference system was obtained using the soft MEAN operator.
5. The developed method will increase the efficiency of mobile robot control in multi-agent robotic stereo vision systems by increasing the accuracy of determining the distance to objects and reducing the decision-making time.

ACKNOWLEDGMENT

This studies was supported by the State Assignment (Agreement № 0851-2020-0032).

REFERENCES

1. Jorge Fuentes-Pacheco, José Ruiz-Ascencio, Juan Manuel Rendón-Mancha Visual simultaneous localization and mapping: a survey, Artificial Intelligence Review, – November 2015.
2. A. Buyval, I. Afanasyev, and E. Magid Comparative analysis of ROS-based Monocular SLAM methods for indoor navigation, The 9th Int. Conf. on Machine Vision, 2016.
3. R. W. Wolcott, R. M. Eustice Visual Localization within LIDAR Maps for Automated Urban Driving, 2014.
4. Algabri, M., Mathkour, H., Ramdane, H., and Alsulaiman, M., 2015. Comparative study of soft computing techniques for mobile robot navigation in an unknown environment. Computers in Human Behavior. 50: 42–56. doi:10.1016/j.chb.2015.03.062.
5. Bobyr, M., Titov, V., Belyaev, A. 2016. Fuzzy System of Distribution of Braking Forces on the Engines of a Mobile Robot. MATEC Web of Conferences, 79, 01052. doi.org/10.1051/matecconf/20167901052.
6. Delmerico, J.A., David, P., and Corso, J.J. 2013. Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision. Image and Vision Computing. 31(11):841–852. doi.org/10.1016/j.imavis.2013.08.006.
7. Lange, M.M., Ganebnyh, S.N., and Lange, A.M. 2017. On efficiency of the hierarchical algorithm for searching approximate nearest neighbor in a given set of images. Inform. Appl. 11(3):51-59. doi.org/10.14357/19922264170306.
8. Zha, D., Jin, X., and Xiang, T. 2016. A real-time global stereo-matching on FPGA. Microprocessors and Microsystems. 47: 419–428. doi.org/10.1016/j.micpro.2016.08.005.
9. Pérez-Patricio, M., Aguilar-González, A., Arias-Estrada, M., Hernández-De León, H.R., Camas-Anzueto, J.L., and de Jesús Osuna-Coutiño, J.A. 2016. An FPGA stereo matching unit based on fuzzy logic. Microprocessors and Microsystems. 42:87–99. doi.org/10.1016/j.micpro.2015.10.011.
10. Bobyr, M.V., Emelyanov, S.G. 2020. A nonlinear method of learning neuro-fuzzy models for dynamic control systems. Applied Soft Computing Journal. 2020, 88, 106030. doi.org/10.1016/j.asoc.2019.106030.
11. Georgoulas, C., and Andreadis, I. 2011. A real-time fuzzy hardware structure for disparity map computation. Journal of Real-Time Image Processing. 6(4):257-273. doi.org/10.1007/s11554-010-0157-6.
12. Bobyr, M.V., Kulabukhov, S.A. 2017. Simulation of control of temperature mode in cutting area on the basis of fuzzy logic. Journal of Machinery Manufacture and Reliability, 2017, 46(3), p.288–295. doi.org/10.3103/S1052618817030049

13. Bobyr, M.V., Yakushev, A.S., Dorodnykh, A.A. 2020. Fuzzy devices for cooling the cutting tool of the CNC machine implemented on FPGA. Measurement: Journal of the International Measurement Confederation, 2020, 152, 107378. doi.org/10.1016/j.measurement.2019.107378
14. Middlebury Stereo Datasets. Available at: http://vision.middlebury.edu/stereo/data
15. Gorbachev, S., Gorbacheva, N., Koynov S. 2016. A synergistic effect in the measurement of neuro-fuzzy system, MATEC Web of Conferences, 79, 01065. doi: 10.1051/matecconf/20167901065.
16. Gorbachev, S., Syryamkin, V. 2018. High-Performance Adaptive Neuro-Fuzzy Classifier with a Parametric Tuning, MATEC Web of Conferences, 155, 01037. doi: 10.1051/matecconf/201815501037.

Cite this article

M. V. Bobyr, N. A. Milostnaya, S. V. Gorbachev, S. Bhattacharyya and J. Cao, The Method of Depth Map Calculating Based on Soft Operators in Multi-Agent Robotic Stereo Vision Systems, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 83-98

7

Bandwidth Control Sectoring Technique Protocol for Data Dissemination in Wireless Sensor Networks

Deepika U. Shevatkar[a, b, *], Anand A. Khatri[c, *], Yogesh K. Sharma[a] and G. M. Bhandari[b]

[a]Computer Engineering Department, Jagdishprasad Jhabarmal Tibrewala University, Rajasthan, India
[b]Computer Engineering Department, JSPM's Bhivarabai Sawant Institute of Technology and Research, Pune, India
[c]Computer Engineering Department, Jaihind College of Engineering, Kuran, Pune, India

*Corresponding author: dipikaus@gmail.com, khatrianand@gmail.com

ABSTRACT

The present chapter relates to a system for data collection in wireless sensor networks. The object is to provide an improved Quality of Services using bandwidth control sectoring technique for WSNs. To reduce the power consumption through system. The nodes are randomly deployed in the network. There are sectors formed according to the equal number of sector heads. There is one sink node, which collects data from sector heads. Common nodes are deployed randomly to transmit data packets to the respective level 1 node within a sector. Thus, bandwidth will be controlled, and congestion in the network is reduced. Bandwidth control sectoring technique (BCST) achieves the various quality parameters of Network. QoS like period of time (delay), power resource efficiency, delivery ratio, loss ration including throughput. This chapter shows the best protocol for data dissemination in wireless sensor networks. This is designed for achieving maximum QoS of wireless sensor networks.

Keywords: Bandwidth Control Sectoring Technique, BCST, Congestion Control, Sector head, Quality of Services, QoS, Wireless Sensor Networks, WSN

Advanced Research in Computer Engineering | ISBN: 978-81-948951-2-1
Book Series: Research Transcripts in Computer, Electrical and Electronics Engineering | Vol: 02

1. INTRODUCTION

Wireless sensor networks[1] can be specified to track geological and physiological factors such as temperature, noise, friction, strain, activity, etc. as self-organized and resources-less sensing devices to pass that data across a network to an end nodes or wireless device, in which the information is gathered and processed in the unit [4][12]. A sink or base station serves as a user-to-sensor connection. By entering probes and acquiring responses again from sink, one could recover the relevant data from either the network [13]. There seem to be typically several or groups of cells in a wireless sensor network. Utilizing wireless signals, the IoT devices can interact with the root station.[5][6] With signaling and computer equipment, communications equipment and power components, a wireless sensor node is configured. Constricted space, speed of information processing and processing capacity are the actual mobile devices in a wireless sensor network (WSN) which have constraints. After the node again for sensor They are liable for self-organizing an adequate system framework after the sensor nodes are configured in the network, mostly with multi-hop coordination with them [7][8]. Wireless sensor devices answer to requests given to execute particular actions from a root station [9]. Global Navigation Method and local positioning methods can be utilized to gain place and pos-related knowledge [10][11].

2. LITERATURE SURVEY

In research paper, in effect to enhance power consumption, the researcher outlined a new strategy for adapting the Minimal Power Efficient Clustering Hierarchy (OLEACH) to improve original LEACH and LEACH-C by actively sorting clusters relying on the residual energy of the nodes. In the this new method detectors node-heads evaluating the quantity of electricity staying after every round. The new cluster - based routing protocol is based on energy efficient, as the level of energy percentile for the chosen is specified in progress and consequently prevents its performance and non - stop cooperation task. In place to ensure optimal functioning of the overall system, the latest cluster - based routing system is based on a sustainable value limit, eliminating the formation of a leader of the group. The technique for effective energy clustering for WSNs was being presented. Informative tests of the local network of wearable sensors indicate that our approach is good for increasing network period and also has the likely to improve the service life of the whole network. From our perspective, O-LEACH can run on both active and passive systems. We have assessed O-LEACH only on real networks in this article. This approach also can be checked on large structures. The test results show that

compared to the initial Leeching and Leeching-C, the new scheme gains greater stability and enhances device longevity and resilience [1]. Dynamic traffic control technique is introduced by author Bharadwaj. This technique is implemented by using WSN. Sensors are used to discover the traffic congestion for dynamically manage the traffic. Dynamic traffic control has overcome the disadvantages of static traffic control. The major disadvantage of static traffic control is sometimes it may block the emergency vehicles like i.e. ambulance because of traffic congestion. In this technique, Traffic Control Unit, Monitor Unit, and Roadside Unit are used to efficiently control the traffic congestion. RFID reader is used to reading the unique RFID code for an emergency vehicle as well as it sends to the monitor unit. Sensors, proximity switch, and RFID tags are used by monitor unit to count the normal vehicles and emergency vehicles. Count information of the vehicle is sent to the traffic control unit. After receiving count information of vehicles, the signals are changed dynamically. Author has compared the results of static traffic control technique and dynamic traffic control technique. Results are taken for time used by vehicles from source to different destination as well as time is calculated by using speed and distance. Dynamic traffic control technique works for all situations. Reduces the traffic delay and saves the throughput time of travel are the advantages of the dynamic traffic control technique [2].

The author Weiqi Chen has proposed joint QoS provisioning and congestion control technique for the multi-hop wireless network. This technique is implemented by combining two techniques. One is Differentiated Queuing service and second is semi-TCP. Advantages of these techniques are provided per-packet granular QOS as well as systematic hop by hop congestion control. Using Joint technique two parameters getting by authors. These are flexible and adaptive for the dynamic multi-hop wireless network. In the joint technique, DQS is used for QoS in the logic layer and semi TCP is used for congestion control issue in the transport and MAC layer. Author has solved the different issues of the existing system. Those issues are delay estimation, overdue packet handling issue, ACK mechanism, and cross-layer design. Using joint technique author has improved the performance of total throughput and reduces the total latency in the (multi-hop) i.e. multiple hops (levels) wireless network. The results are taken for data transfer-ratio and average end-to-end delay parameter. Using joint technique author has achieved different advantages. Those are handled overdue packets, transfer ratio and decreased total-latency for the multi-hop wireless-network [3].

3. PROPOSED SYSTEM

In Figure 1 it is seen that nodes are randomly deployed in the network. There are sectors formed according to the equal number of sector heads. Due to less capacity of the communication channel and less availability of bandwidth for packet transmission, congestion over the network increases which directly affects the QoS parameters like de- lay, throughput, energy efficiency, packet delivery ratio, reliability. In the proposed system sectoring technique will reduce bandwidth problem and traffic congestion problem. There is one sink node which collects data from sector heads. Common nodes which are deployed randomly transmit data packets to respective level 1 nodes within a sector. Sector heads are nothing but level 1 node. In this way bandwidth will be controlled and congestion in network will be reduced.

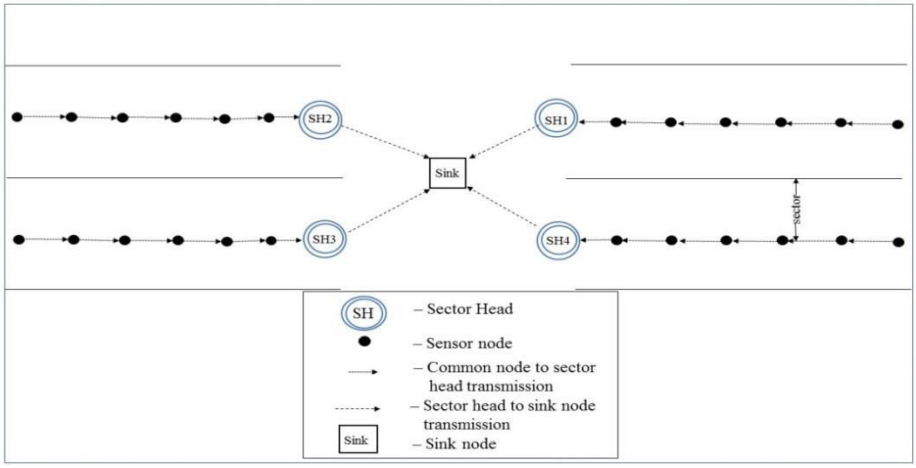


Fig. 1. Physical View (System Architecture)

BCST Algorithm:

Step 1 - START

Step 2 - Initialization of scenario.

Step 3 - Initialization of sink node equal to zero. Step 4 - Set hop count (level) to all nodes.

Step 5 - Determine the nodes which are one hop away from sink node i.e Find out level

one nodes.

Step 6 - Assign level one nodes as sector head.

Step 7 - Formation of sectors depending on the total number of sector heads.

Step 8 - Common nodes within a sector transmit the data packets to their sector head.

Step 9 - Sector heads transmit the collected information to sink node.

Step 10 - STOP

Firstly, initialization of scenario will be done and initialization of sink node = 0 will be carried out. Set hop count to all nodes to determine the levels of nodes in the network. That this next step is to figure out which points were a step ahead from the end nodes. Determine those nods as level 1 nodes. Level 1 node are near to sink node so next step is to determine level 1 nodes as sector head (SH). Due to less capacity of the communication channel and less availability of bandwidth for packet transmission, congestion over the network increases. In this case formation of sectors will control bandwidth. Form sectors equal to the total number of level 1 nodes i.e. sector head. After formation of sectors data transmission will be carried out. Data transmission is done within sector. Common nodes within a sector transmit data packets to sector head. Sink node collects data packets from level 1 nodes i.e. sector head. In this way bandwidth will be controlled and congestion in network will be reduced.

4. RESULT ANALYSIS

In this scenario total no of 30 nodes are deployed in sector form in 1000m* 1000m area for simulation. One node is assigned as data collector node i.e. sink node. Remaining 29 nodes are working as a source node. Rate of data transmission is changing from 10 to 50 packets per sec. i.e. Transmitting data transmission limit to the end nodes. Protocols like C.S.M.A ,TD.M.A and 802.15.4(Zigbee). included for performance analysis and these results are comparing with proposed MAC BCST. The routing algorithm for A.O.D.V is being included for searching short distance path for data transmission.50 bytes packet size is fixed for this scenario along with 30 nodes.

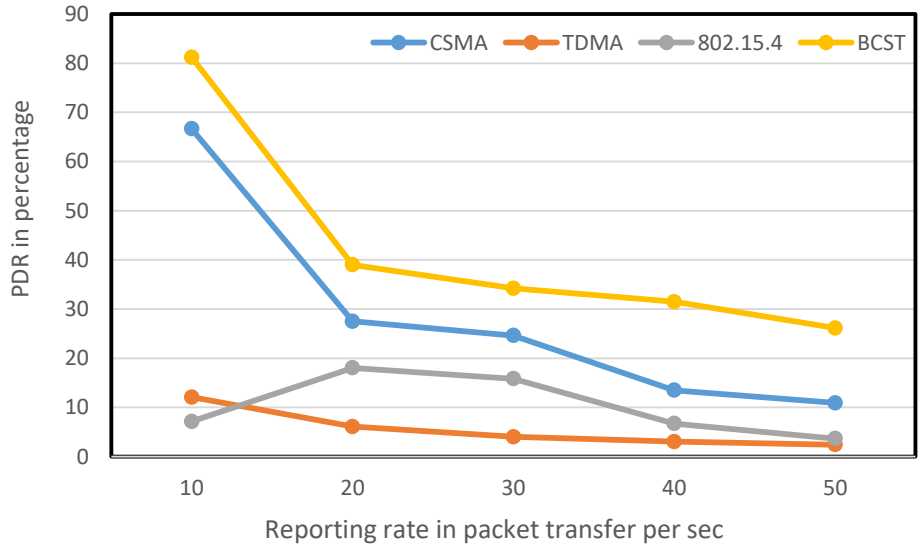


Fig. 2. Average PDR for RR

Above graph shown Average PDR (Packet Delivery Ratio) for reporting rate. Relative to modern technique, the Bandwidth Control Sectoring Protocol (BSCT) is substantially better. In case of heavy traffic like 50 packet transfer per sec it can increase the reporting rate of node but in case of less traffic like 10 packet transfer per sec it can act as a normal MAC with constant reporting rate. In normal packet delivery gives 20% better result as compare to carrier sensor multiple access protocol (CSMA) protocol, 70% as compare to Tim div multiple access system (T.D.M.A) &75% as compare to 802.15.4 protocol. Similar to CSMA, TDMA and 802.15.4, the distribution of big packets is 15 to 25 percentage points higher. as result shown the BCST protocol a maximum number of packets will be delivered and minimum number no of packet will be drop.

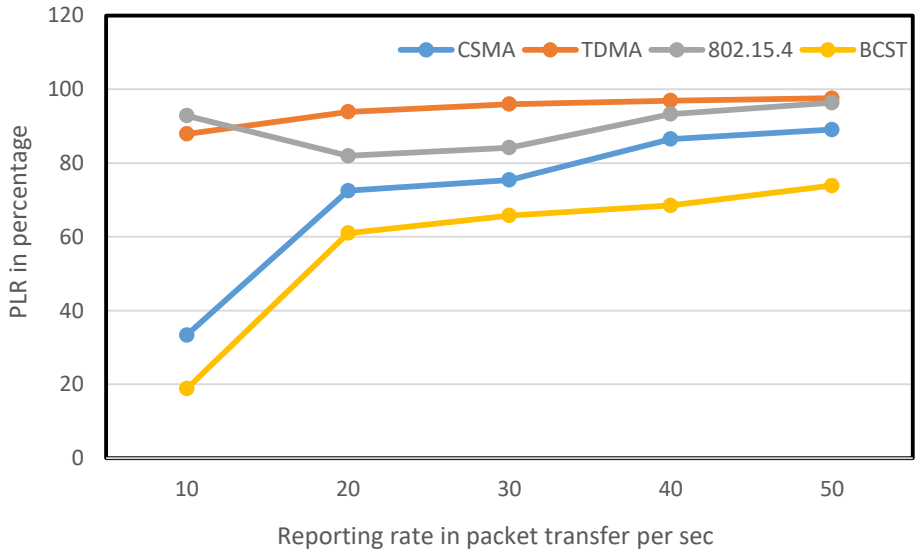


Fig. 3. Average PLR for RR

Above graph shows average PLR (Packet Loss Ratio) for Reporting rate. The packet loss ratio represents the ratio of the number of lost packets to the total number of sent packets from sensor node to base station. Congestion is one of the main factors of packet loss. In normal reporting rate CSMA protocol drops 70% packets, TDMA and 802.15.4 protocol drops 90% packets and as compare to other protocols with BCST only 20% packet drops. The transmission number is 50. Similar to certain other
methods, the BCST protocol gives 20 percent good efficiency. Tables of the PLR shows the TDMA and 802.15.4 protocols have more packet loss and CSMA and BCST has a less packet loss.

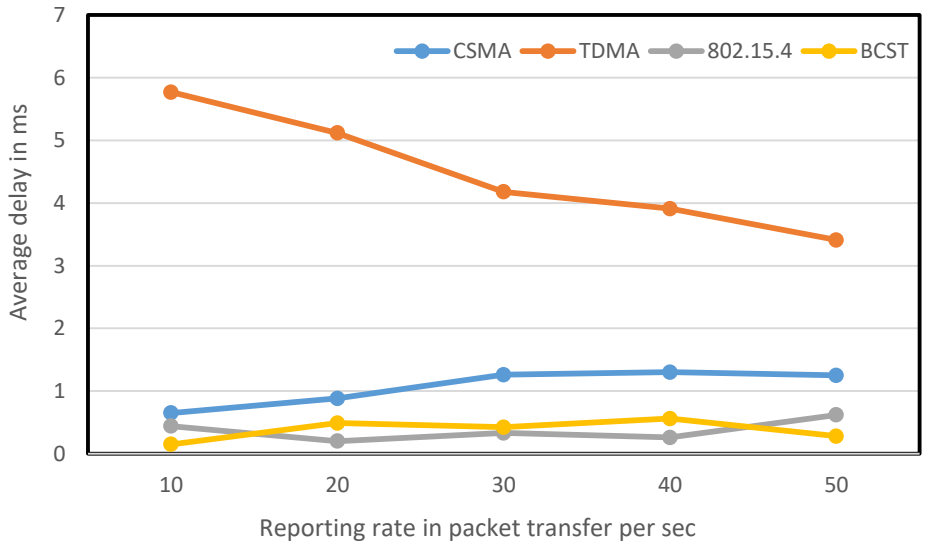


Fig. 4. Average Delay for RR

The figure described represents the typical first-to- end interval throughout the reporting count. End-to-End delay means time takes by the packet to travels to reach from source to destination node. Number of packets transfer using TDMA protocol takes more time than the other protocols. CSMA also takes more time. The number of packets transfer via BCST protocols takes less time as compare to other protocols to reach destination. The other protocol like 802.15.4 also takes less time but the number of packets is increase delay is also increases but in case of BCST protocols number of packets increase delay time will be decrease.

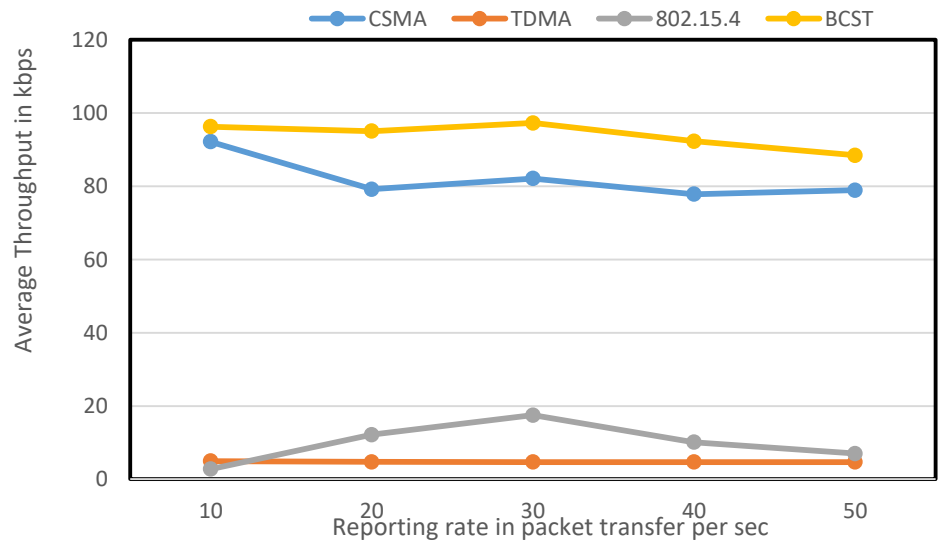


Fig. 5. Average Throughput for RR

Above graph shows Average throughput for reporting time. The BCST protocol gives better results when it is compared to CSMA. The Control overhead in BCST protocol is also less when it is matched with BCST protocol shown in table. The similar methods same like T.D.M.A and 802.15.4 as compare to BCST has less throughput. When nodes send 10 packets per sec using BCST and CSMA protocol has 90 to 95% throughput but in case of sending 50 packets BCST protocol has 90% throughput and in CSMA protocol has 70% reporting rate. In TDMA and 802.15.4 protocol have lowest throughput i.e. 0 to 20.

5. CONCLUSION

A BCST – bandwidth control sectoring protocol is a novel protocol designed for congestion control and achieving QoS parameters of wireless networks. Compared to conventional tech like C.S.M.A, T.D.M.A and 802.15.4, the said protocols perform greatly good outcomes. BCST protocol gives almost 17-20% better result for PDR, 15-20% better result for PLR, 5 to 10% better result for delay and 20 to22% increased outcome case of system throughput as compare to second best protocol CSMA. BCST is almost 40 to 50% great as similar to T.D.M.A and 802.15.4 for all QoS parameters. Sectoring technique in BCST helps to reduce heavy traffic, congestion and improve system performance. We will use the modified M.A.C method in proposed development that will blend C.S.M.A and BCST protocol for stronger system performance.

REFERENCES

1. Sepehr Ebrahimi Mood & Mohammad Masoud Javidi. Energy-efficient clustering method for wireless sensor networks using modified gravitational search algorithm. Evolving Systems (2019)
2. H.Oudani, S.Krit, M. Kabrane , K. Karimi, M. Elasikri, K. Bendaoud, H. El Bousty, L. Elmaimouni. Minimize Energy Consumption in Wireless Sensor Network Using Hierarchical Protocols. 2017 International Conference on Engineering & MIS (ICEMIS) 8-10 May 2017. 2575-1328
3. Yan-Xiao Li, Lian Qin, Qian Liang. Research on Wireless Sensor Network Security. 2010 International Conference on Computational Intelligence and Security. 11-14 Dec. 2010
4. J. Li, H. Y. Zhou, D. C. Zuo, K. M. Hou, H. P. Xie and P. Zhou. Energy Consumption Evaluation for Wireless Sensor Network Nodes Based on Queuing Petri Net. Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2014, Article ID 262848

5. Sohail Jabbar, Abid Ali Minhas, Muhammad Imran, Shehzad Khalid and Kashif Saleem. Energy Efficient Strategy for Throughput Improvement in Wireless Sensor Networks. Sensors (Basel). 2015 Feb; 15(2): 2473–2495.
6. International Journal of Engineering Trends and Applications, Volume 4 Issue 3, May-Jun 2017 Rohit Kumar, Er. Sandeep Kad. Energy Efficient Protocols for Wireless Sensor Networks -A Review. International Journal of Engineering Trends and Applications (IJETA) – Volume 4 Issue 3, May-Jun 2017
7. Gang Yang, Tiantian Xu, Mingfei Liang. Research on static clustering for wireless sensor networks. 2011 Second International Conference on Mechanic Automation and Control Engineering 15-17 July 2011. Hohhot, China
8. Deepak Goyal, Malay Ranjan Tripathy. Routing Protocols in Wireless Sensor Networks: A Survey. 2012 Second International Conference on Advanced Computing & Communication Technologies 7-8 Jan. 2012. 10.1109/ACCT.2012.98
9. Abin John Joseph, U.Hari. Hexagonally Sectored Routing Protocol For Wireless Sensor Networks. International Journal of Engineering Research & Technology (IJERT) Vol. 2, May – 2013 ISSN: 2278-0181
10. Mohamed Amine Kafi, Djamel Djenouri, Jalel Ben-Othman, Nadjib Badache. Congestion Control Protocols in Wireless Sensor Networks: A Survey. IEEE Communications Surveys & Tutorials (Volume: 16 , Third Quarter 2014). 1369 – 1390
11. [11] Viktor Richert, Biju Issac, and Nauman Israr. Implementation of a Modified Wireless Sensor Network MAC Protocol for Critical Environments. Hindawi Wireless Communications and Mobile Computing Volume 2017, Article ID 2801204
12. Mohammadreza Balouchestani, Kaamran Raahemifar, and Sridhar Krishnan. Increasing the Reliability of Wireless Sensor Network with a New Testing approach Based on Compressed Sensing Theory. 2011 Eighth International Conference on Wireless and Optical Communications Networks. 24-26 May 2011
13. Joseph Kabara and Maria Calle. MAC Protocols Used byWireless Sensor Networks and a General Method of Performance Evaluation. Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2012, Article ID 834784
14. Priyanka Sharma, Garimella Rama Murthy. Mathematical Modelling of Energy Wastage in Absence of Levelling & Sectoring in Wireless Sensor Networks. January 2017 International Journal of Communication Networks and Distributed Systems 19(1):84

15. Mingwei Li, Yuanwei Jing. Feedback Congestion Control Protocol for Wireless Sensor Networks. 2012 24th Chinese Control and Decision Conference (CCDC). 1948-9439
16. Omkar U, Nikhil T, Devanand C, Akash H, Dattatray Waghole," Smart ambulance system using IoT" 2017 International Conference on BID,PP. 171-176, 2017/12/20.
17. Sohail S, Dattatray Waghole, PKumbhar, V Kotkar, P Awaghade, "Patient monitoring system using IoT", 2017 International Conference on BID, pp. 177-181, 2017/12/20.

Cite this article

Deepika U. Shevatkar, Anand A. Khatri, Yogesh K. Sharma and G. M. Bhandari, Bandwidth Control Sectoring Technique Protocol for Data Dissemination in Wireless Sensor Networks, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 99-108

Book Series by Grinrey Publications

- Research Transcripts in Energy
- Research Transcripts in Engineering
- Research Transcripts in Materials
- Research Transcripts in Computer, Electrical and Electronics Engineering

www.grinrey.com

GRINREY

ISBN 978-81-948951-2-1