Lip-Reading using Neural Networks

Abhay Bagai[†], Harsh Gandhi[††], Rahul Goyal[†††]

Ms. Maitrei Kohli[††††], Dr. T.V. Prasad[†††††]

[†]Student of Computer Science Engineering, Lingaya's Institute of Mgt. & Tech., Faridabad, 121002 India
[††] Student of Computer Science Engineering, Lingaya's Institute of Mgt. & Tech., Faridabad, 121002 India
[†††] Student of Computer Science Engineering, Lingaya's Institute of Mgt. & Tech., Faridabad, 121002 India
[††††]Faculty of Computer Science Engineering, Lingaya's University, Faridabad, 121002 India
[†††††]Head of Computer Science Engineering Dept., Lingaya's University, Faridabad, 121002 India

Summary

Lip-Reading has been practised over centuries for teaching deaf and dumb to speak and communicate effectively with the other people. In this study, the use of neural networks in lip reading is explored. We convert the video of the subject speaking different words into images and then images are further selected manually for processing. As per the research the horizontal and the vertical distance between the lips varies for each and every word considering the close proximities of similar sounding words. Based on this research we can create the database of commonly used words and our neural network model can form clusters of words based on its intelligent approach. This approach can be associated with various voice recognition softwares and help in increasing their efficiency readily even in a noisy environment, and creating new dimensions for human computer interaction..

Key words:
Lip-Reading, Neural Networks, Clustering, Predicting

1. Introduction

Lip-Reading is an area which till date is only practised by elementary student-teacher method, but in this approach lip reading has been implemented through e-learning or virtual learning technique which is a difficult and a challenging problem. Also, in a noisy environment where voice recognition softwares tend to underperform, using this approach increases the efficiency remarkably. Despite the difficulty associated with the unpredictable nature of problem domain, several researchers have attempted to develop models for the same problem but with different approaches. Most of them have tried to compare the two images by classic image comparison techniques. However the results obtained were of lower efficiency considering the efficiency downgrading constraints of image comparison technique used in the approach. In our study, we explore the use of neural networks in lip reading techniques. We convert this problem into a clustering one, where we calculate the four different distances: horizontal and vertical
viz. H-1 (horizontal distance between inner lip points), H-2(horizontal distance between outer lip points), V-1

(vertical distance between inner lip points) and V-2 (vertical distance between outer lip points) and cluster them.

The remainder of this paper is organized as follows. Section 2 briefly reviews the basics of ANN and also presents the comparison between the statistical techniques and our ANN approach. Section 3 gives the details of our methodology by specifically talking about the data, the neural network model, the experiment methodology and the performance measures used in this study. Finally, the Section 4 of the paper discusses the overall contribution of this study, along with its limitations and further research directions
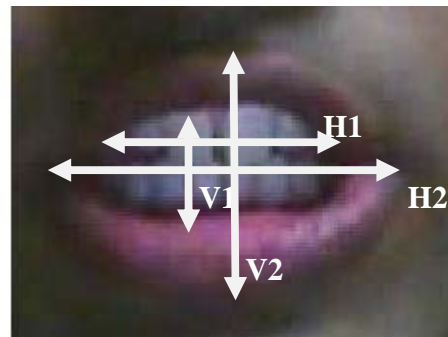

Figure 1: Image showing the four distances.

2. Theoretical Consideration

A neural networks ability to perform computation is based on hope that we can reproduce some of the flexibility and power of human brain by artificial means. Basically a neural network is machine that is designed to model the way in which the brain performs a particular task or function of interest. The network is usually implemented by using electronic components or is simulated in software on a digital computer.[6].

The property of neural network that is of primary significance is the ability of network to learn from its environment, & to improve its performance through

learning. A neural network learns about its environment through an interactive process of adjustments applied to its synaptic weights and bias level. After completing the learning process successfully the network is ready to be deployed for independent functioning.

Neural networks have been applied to an increasing number of real world problems of considerable complexity. The most important advantage is that artificial neural networks are capable of solving problems that are too complex for conventional technologies – problems that do not have an algorithmic solution or that solution is too complex to be found.

The application domain of neural networks today touches almost the entire sphere of science. These include: Association, Clustering, Classification, Pattern Completion, Regression and Generalization, Forecasting, Optimization etc to name a few.

Many application bibliographies exist. However, none of these include an application in lip reading using distance clustering approach. This study is one of the first to attempt the use of neural networks for addressing this challenging problem that has drawn the attention of many researchers in such areas of pattern recognition and image comparison science.

This study is differentiated from the others as follows. First, there is no reported study on using neural networks clustering technology in lip reading applications. This approach seems to be the first attempt of its kind in this problem domain. Another distinguishing feature of this study comes from its statistical approach. This approach is based on the most commonly used words and their intensive database comprising of just their distances making the search and processing faster and more accurate as compared to traditional image comparison techniques. This study also compares the similar sounding words and displays each one of them as an option. The results suggest that our neural networks model performs better than the ones reported in the literature.

3. Experimental Consideration

In this study, the use of neural networks in lip reading is explored. Here, the lip reading problem is converted into problem using clustering. Data clustering deals with the problem of classifying a set of N objects into groups so that objects within the same group are more similar than objects belonging to different groups. Each object is identified by a number m of measurable features, consequently, ith object can be represented as a point $X_i \in R^m$, i = 1, 2, ..., N. Data clustering aims at identifying clusters as more densely populated regions in the space R^m.[4].

In this study, we explore the use of Neural Networks in predicting the letter or word spoken by the subject. The video of the word spoken by the subject is taken and is

split in to various .jpeg images using our tool.[1]. These images are then further processed using the code in Matlab.[3]. Once the images are selected the database creation begins as follows:

3.1 Database Creation

3.1.1 The suitable images are selected from the set of split images. For achieving better results we have taken in to account minimum of two and maximum of five images for a letter or a word as suited.

3.1.2 The lip area on the image is highlighted using the spline curve [2] by marking the control points and then the grid is superimposed on the image so that the distances can be taken with better accuracy.

3.1.3 The lip area is extracted or cropped from the complete image and then the various distances i.e. H1, H2, V1, V2 as explained in section 1 are calculated and the distances are stored in the database, then these four distances are summed up for each image and the resultant sum comprises of the sums of the relevant images of the particular letter or the word and send for further clustering that takes place with the help of Neuro Solutions using Matlab.[3].

3.2 Application Testing

Once we get the clustered data through Neuro Solutions our database stands complete. Now as the database is complete the testing of the application is done predicting the letter or the word by matching it and classifying it with our database.

3.2.1 The test images passes through steps 3.1.1 to 3.1.3 from where we get the test sum which is to be compared with the original entries of the database.

3.2.2 The comparison is done by keeping in consideration a buffer of 25 units for manual error constraints

3.2.3 After the comparison takes place the prediction of the letter or the word done through the classified entries of the clusters
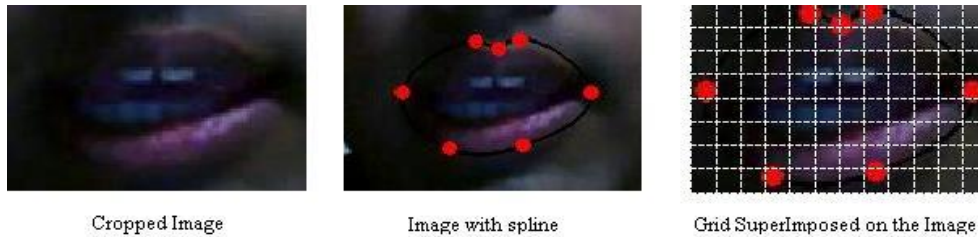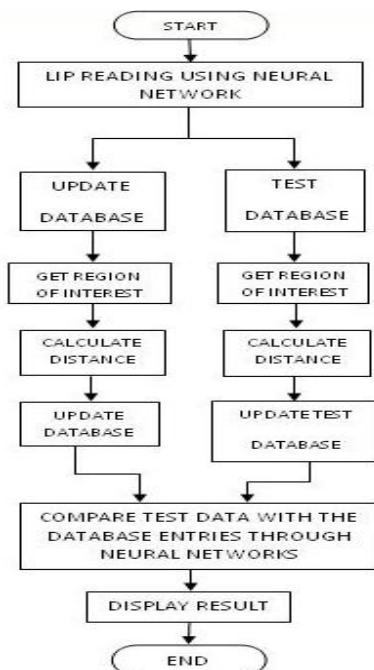
Figure 2: Images showing Cropped, Spline and Grid

Figure 3: Process Flowchart Model.

4. Conclusion and Results

The approach of using Neural Networks in this field aims to classify various letters, words of English language on the basis of distance separation of lips while speaking them & thus predicts the letter or word the next time it is spoken. We have implemented this project using an evaluation version of the software NeuroSolutions5.[4]. Since we have used the evaluation version of the software, we are therefore getting the maximum accuracy of 52%. This is mainly due to the reason that the evaluation version does not incorporate all the features required to compute accurate result. This research project is the first attempt to use Neural Networks classification (clustering) for addressing this challenging problem that combines two different application domains of Classification & Predicting & brings out the much desired output. This

model would be highly beneficial to the:

      a)   Voice Recognition Applications-

They tend to show higher accuracy and hence efficiency increases remarkably when used with the above approach model especially in the noisy environments.

      b)   Enhanced Human Computer Interaction-

It takes personalized computing to the next level with more enhanced command recognition even in the noisy environments.

      c)   Personalised systems for deaf and dumb people-

The deaf and dumb can easily personalize their systems by interacting with the system just through their lip movements.

Compared to the other models using image comparison techniques, using the same experimental conditions, as reported in Section 1 & 3, our statistical approach clubbed with neural networks clustering performed significantly better.

Beyond the accuracy of our results in reading the lips for letters and words, this neural network model could also be adapted to read sentences by altering the inputs and training the network more intensively.

This model can also be embedded in a device having a camera and a speaker which is put on a deaf and dumb person's head, where the camera captures the video of the person making the lip movements and converts those movements into speech through the speaker fitted in the device itself.

From an application perspective, once developed to a production system, such a neural network model can be made available (via a web server or as an application service provider) to the public at various places such as railway stations, bus stands, government buildings, banks, courts etc. where the system effectively interacts with the people especially those who are physically challenged. For e.g.: an automatic ticketing system at railway and bus stations can recognise the passenger and after confirming his/her identity over the web understands his command and issues him his desired ticket. In the same way similar systems can be implemented in places like banks and public convenience places.

References

[1] Advanced X Video Converter, Version 5.0.3. The World Wide Web address is www.aoamedia.com.

[2] Cubic Spline Interpolation, Sky McKinley and Megan Levine, Math 45: Linear Algebra.

[3] Matlab, The MathWorks, Inc. (Copyright 1984-2005) Version 7.1.0.246 (R14) Service Pack 3. The World Wide Web address is www.mathworks.com.

[4] Neural Network Clustering Based on Distances between Objects, Leonid B. Litinskii, Dmitry E. Romanov, Institute of Optical-Neural Technologies Russian Academy of Sciences, Moscow.

[5] NeuroDimension, Inc. (2004). Developers of NeuroSolutions v4.01: Neural Network Simulator. The World Wide Web address is www.nd.com Gainesville, FL.

[6] Predicting Admission Counselling Triumph of Colleges Using Neural Networks, Maitrei Kohli, Priti Puri, 7th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Databases (AIKED'08), University of Cambridge, UK, Feb 20-22, 2008.

Harsh Gandhi, a final year student at Lingaya's Institute of Mgt. & Tech., Faridabad, Haryana, India. His areas of interest include Image processing, Artificial Neural Networks, software development life cycles, systems & operations and project management areas.

Rahul Goyal, a final year student at Lingaya's Institute of Mgt. & Tech., Faridabad, Haryana, India. His areas of interest include Image processing, Computer architecture and Artificial Neural Networks

Abhay Bagai, a final year student at Lingaya's Institute of Mgt. & Tech., Faridabad, Haryana, India. His areas of interest include Image processing, Artificial Neural Networks, Computer organization and Operating System.

T. V. Prasad received his master's degree in Computer Science from Nagarjuna University, AP India and a doctoral degree from Jamia Milia Islamia University, New Delhi, India. He has over 14 years of experience in industry and teaching. He has worked as Deputy Director in the Bureau of Indian Standards, New Delhi. Currently he is Professor and Head of Department of Computer Sc. & Engg. At Lingaya's University, Faridabad, Haryana, India. He has 40 papers and two books to his credit. His areas of interest include bioinformatics, artificial intelligence, consciousness studies, computer organization and architecture. He is a member of reputed bodies like ISRS, CSI, APBioNet, etc.

Maitrei Kohli received her bachelor's and master's degree in Computer Science from M.D. University, Haryana, India. She has 2 years of experience in teaching. Currently, she is working as a Lecturer in the Dept. of Computer Sc. & Engg. at Lingaya's University, Faridabad, Haryana, India. She has around 10 papers to her credit. Her papers have been published in various international conferences – Univ. of Cambridge, U.K. and journals like the IJCSNS. She has one copyright to her acclaim. Her areas of interests include neural networks, bioinformatics, artificial intelligence, consciousness studies. She is a member of reputed bodies like International Association of Engineers, International Neural Network Society, WSEAS etc.