ISSN 22773061

Enhancement of Speech Recognition System by neural network approaches of Clustering

Anurag Upadhyay Associate Professor LPU, Jalandhar
Chitranjanjit Kaur M-tech LPU, jalandhar

Abstract: This paper addresses the problem of speech recognition to identify various modes of speech data. Speaker sounds are the acoustic sounds of speech. Statistical models of speech have been widely used for speech recognition under neural networks. In paper we propose and try to justify a new model in which speech co articulation the effect of phonetic context on speech sound is modeled explicitly under a statistical framework. We study speech phone recognition by recurrent neural networks and SOUL Neural Networks. A general framework for recurrent neural networks and considerations for network training are discussed in detail. SOUL NN clustering the large vocabulary that compresses huge data sets of speech. This project also different Indian languages utter by different speakers in different modes such as aggressive, happy, sad, and angry. Many alternative energy measures and training methods are proposed and implemented. A speaker independent phone recognition rate of 82% with 25% frame error rate has been achieved on the neural data base. Neural speech recognition experiments on the NTIMIT database result in a phone recognition rate of 68% correct. The research results in this thesis are competitive with the best results reported in the literature.

Keywords: Artificial neural networks, SOUL neural networks, Clustering, Speech Recognition system.

Council for Innovative Research
Peer Review Research Publishing System
Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 6, No 1

editor@cirworld.com
www.cirworld.com, member.cirworld.com

ISSN 22773061

Introduction: Evolution of Neural network in Speech recognition:

Earlier Neural Network was technique that was used only for pattern recognition, numerous researchers apply neural network for recognition of speech. In previous work basic assignments are completed by neural networks such as:

- Voice/Unvoiced
- Nasal/fricative/plosive

Fine outcome of experiment by researcher encourage them to progress on phoneme that distinguish one word from other and classification of words are also done. Neural network play vital role in this. There are two crucial approaches to speech classification:

- Static
- Dynamic

Static classification: In this classification, the neural network selects input speech at one time and makes one conclusion. Huang &Lippmann in 1988 perform uncomplicated and polished research that reveals that neural network can execute complicated resolution from speech data. They applied MLP (Multilayer Perceptron) in which 2 inputs, 50 hidden units and 10 outputs, on different vowels utter by children, women and men. After large number of iterations decision region formed that is best possible as compare to classification draw by hand. Burr in 1988 applied static network and get fine outcome from this.

Dynamic classification: Neural Network in this classification select only small window of the speech and series of local decision are making when window slides over speech data these local decision integrate and generate global result. This approach use recurrence in dynamic approach. Waibel et al (1989) using Time Delay Neural network get good results, the architecture follow in this consist of 3 and 5 delays in input and hidden layer. This design is beneficial: it make network to develop the general feature detectors that economized on weights due to compaction. This network was used to train phonemes on Japanese words, consists samples of 2000 on database of 5260 words.

SOUL Neural Network

Standard n-gram back-off language models (LMs) rely on a discrete space representation of the vocabulary, where each word is associated with a discrete index. On the contrary, Neural network language models (NNLMs) are based on the idea of continuous word representation. Distributional similar words are represented as neighbors in a continuous space. This turns n-grams distributions into smooth functions of the representations and helps to make use of hidden word and context similarities. These representations and the associated probability estimates are jointly estimated in a neural network. Neural networks, working on top of conventional n-gram models, have been introduced as a potential means to improve discrete language models. This topic has recently gained much attention in the domain of speech recognition. Both neural network approach and class-based models were shown to pertain to the few approaches that provide significant recognition improvements over n-gram baselines for large-scale speech recognition tasks. Probably the major bottleneck with NNLMs is the computation of posterior probabilities in the output layer. This layer must contain one unit for each vocabulary word. Using such a design makes handling of large vocabularies, consisting of hundred thousand words, infeasible due to a prohibitive growth in computation time. Short-list NNLMs, that estimate probabilities only for several thousand most frequent words, were proposed as a practical workaround this problem, Structured OUtput Layer (SOUL) neural network language modeling approach. It is based on a tree representation of the output vocabulary. This approach successfully combines the benefits of neural network and class-based techniques in one single framework. As opposed to standard NNLMs, SOUL NNLMs make it feasible to estimate the n-gram probabilities for vocabularies of arbitrary size. As a result, all the vocabulary words, and not just the words in the short-list, can benefit from the improved prediction capabilities of the NNLMs.

Arbitrary sized vocabularies handle by structure output layer neural network. Previous work about SOUL Neural Network can switch various languages like Arabians, Persian etc. in this work different modes of different languages are predict using structured Output Layer neural network. Speech recognition has been a dynamic part for many decades, many applications and techniques are recognizing the speech. Neural network is statistical techniques for speech recognition system. Enhancement of speech recognition: Several points are consider like

- Clustering of vocabulary
- Parallel implementation of neural network
- Improvement in Back Propagation Through Time Neural Network

In this enhancement means performance of speech recognition using artificial neural network by clustering of vocabulary. In clustering grouping of data is done by similarities in data. This is an excellent application for neural network. It perform great in mine different voice signals related to some subsets related to mode, vocabulary etc.

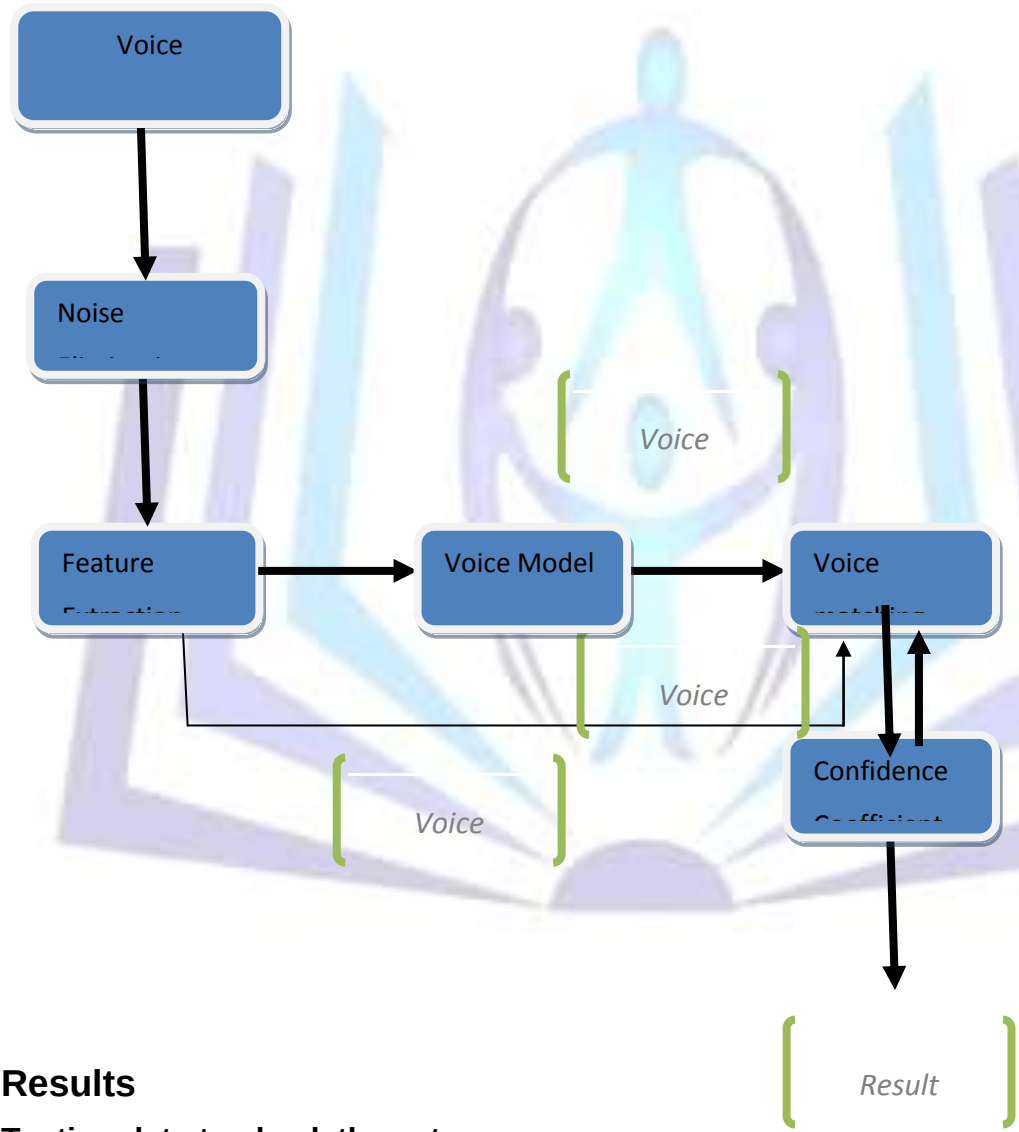
Speech Recognition system: In process of Speech recognition system any signal capture by any media convert into set of words, recognize the speaker or recognition the type of voice, or mode of voice. The first step in speech recognition is to digitize the speech data. Organize records into wise grouping is one of the most essential modes of understanding and

ISSN 22773061

learning. Cluster analysis is the proper learning methods and algorithms for grouping, or clustering. Cluster analysis does not use grouping labels that tag items with earlier identifiers. The lack of type in sequence distinguishes data clustering unsupervised simple clustering algorithms, K-means, was first published in 1955. In spite of the fact that K-means was projected over six decade ago and thousands of clustering algorithms have been available since then, K-means is still widely used because it provide semi supervised clustering, ensemble clustering and simultaneous feature extraction.

Steps in SOUL NN consist of pre-training and post training, in pre-training short-list output by vector initializations method, that estimate projection space of matrix R where R is row selection. Vector Quantization is quantization for converting large data set into small data sets in signal processing; it is basically used data compression. It represents the data in different group that are near to centroid point, various clustering methods are used but in this k-mean cluster algorithm is used. In next step principle component analysis is done on matrix with real values, to transform linear correlated variables uncorrelated variables that reduce the dimensionality of data set and indentify meaningful variables based on variables. Then k means apply on this data set and at last tree structure as output is constructed that is not binary tree but tree with many softmax layers.

Method follow



Results

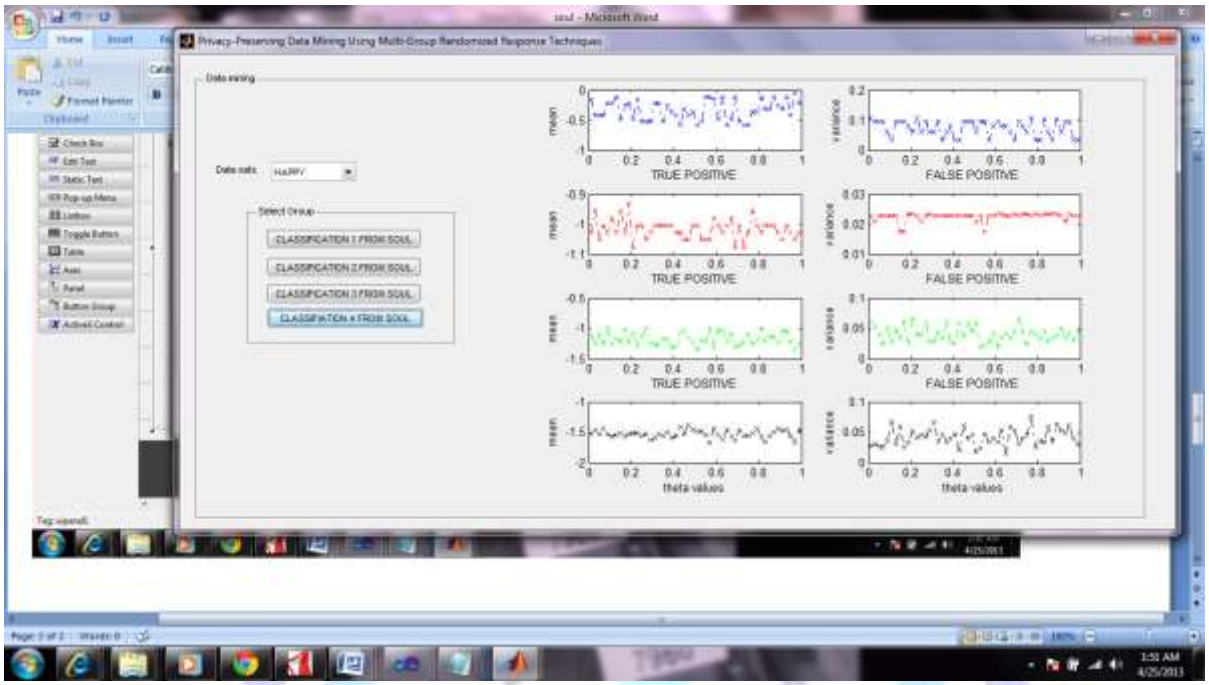
Testing data to check the category

In testing phase different trained data is predicted according to type of category we can select any kind of wave file and plot points of data near to centroid. To remove the noise in the data butter filter is used, that remove the noise in the signal as represent with syntax. In check c

Classification of data

Check the category classification around variance and the mean of data, that represent the positive false, positive true and theta values, in this data mining technique is used that mine this value calculate with SOUL NN.

ISSN 22773061



Terms in Classification

False Positive	Test results that are in incorrect, incorrect classification of speech
True Positive	Test actual positive that identify correctly
Theta values	Way the sound waves vary according to mean and variance.

Conclusion:

Contrast the SOUL NNLM and pick out NNLMs, the reduction of the speech recognition error rate is less than the confound reduction. This can be explained by relatively elevated data exposure with shortlists modes of speech. Such statistics show that comparable amount short-list do moderately sound in nuts and bolts of data coverage even for models with very different size. Thus by the experimental results, the improvements in speech recognition system is obtained by using SOUL NN. This technique carries out separate training of different parts of the structured output layer. A classify of scale more data are used to train the SOUL NNLM without any excessive enhance in computational cost and training time. Although it was experiential that the enhanced SOUL NNLM is valuable when used on its own, the enhanced training method does not seem to have much influenced on the overall performance after interpolation with. Examination of SOUL NNLM construction led to numerous conclusions about the peculiarity of the SOUL architecture. Recurrent modes should be treated discretely though the size of the shortlist can be kept minute. The number of classes and the profundity of the clustering tree do not have much manipulate on bewilderment. The use of clustering tree itself is important since it provides faster training and better perplexities as compared to other NNs.

Since my thesis focuses on phoneme recognition with neural network, I have also demonstrated that the chain learning technique can aid to get better phoneme recognition by structure outputs layers of the neural network. In this SOUL neural network working on different types of voice and predict the category that to which type of category voice belong to.

Future Work

Even though a positive level of phoneme recognition has been achieve in this thesis, the recognition rate is still as much excellent as the some models and fain previous work from commercial point of view. So in future work is needed to further

improve the identification rate. There are numerous ways to develop the recognition rate based on the model described in this thesis. In the sound preprocessing stage, instead of using the clustering by K mean other clustering technique can be used. In this thesis work only different category of speech data is test which can enhance by addition recognition system for multiple languages. Feature extraction can be done by different approaches like Back Propagation Through Time and parallel processing that will Enhance the performance of recognition system. An improved algorithm to evaluate the internal weights of the neural network also needs to be developed, since the current algorithm is not quite time efficient. The current processing time increases rapidly when the number of training samples increases, so very limited training samples have been used in this thesis. This is work on static classification of voice data which can be convert into dynamic classification of data. If a more proficient training algorithm is used, and the system saves more phoneme information and knowledge from training, the recognition rate should increase accordingly. Application of the SOUL NNLM is not confined to speech recognition but can be used for other language technology tasks. It only works for statistical machine translation.

References:

Amin Ashouri Saheli, Gholam Ali Abdali, Amir Abolfazl suratgar (March 18 - 20, 2009) “Speech Recognition from PSD using Neural Network”, Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I, IMECS 2009, , Hong Kong

A. Emami,(2006) “A neural syntactic language model,” Ph.D. dissertation, ohns Hopkins University, Baltimore, MD, USA.

Austin Marshall March 3, 2005 “Neural Network for Speech Recognition” 2nd Annual Student Research Showcase

Dr. R L K Venkateswarlu,Dr.Vasantha Kumari,A K V Nagayya, “Efficient Speech Recognition by Using Modular Neural Network ” Int. J. Comp. Tech. Appl., Vol 2 (3), 463-470

Hai Son Le, Ilya Oparin Abdel Messaudi, Alexandre Allauzen Jean Lue Gauvain(2010), Francois Yvon, Large Vocabulary SOUL neural network lan

guage mode , Universit´e Paris-Sud LIMSI CNRS, Spoken Language Processing Group B.P. 133, 91403 Orsay, cedex, France

H.-S. Le, I. Oparin, A. Allauzen, J.-L. Gauvain, and F. Yvon, (Nov 2011)“Structured output layer neural network language model,” in Proc. of ICASSP’, pp. 5524–5527.

H.-S. Le, I. Oparin, A. Messaoudi, A. Allauzen, J.-L. Gauvain, and F. Yvon (Nov 2011), “Large vocabulary SOUL neural network language models,” in Proc. of Interspeech’, pp. 1469–1472.

Judith Justin, Ila Vennila(2011)“Performance of Speech Recognition using Artificial Neural Network and Fuzzy Logic”, European Journal of Scientific Research ISSN 1450-216X Vol.66 No.1 (2011), pp. 41-47

Juraj KOSCAK , Rudolf JAKSA and Peter SINCAK (2010), Prediction of Temperature Daily Profile by Stochastic update of Back propagation Through Time Algorithm, Journal of Mathematics and System Science, ISSN 2159-5291, USA

Kerel Vesely(2010), Parallel Training of Speech Neural Network for Speech Recognition., Lukas Burget and Frantisik Greze

Lakhsmi Kumari Ranbatla and Koti Verra Naggaya Ande(2010), “A novel Approach of Speech Recognition by using Generalized Regression Neural Network,”

Leonid B. Litinskii, Dmitry E. Romanov(2010), Neural Network Clustering Based on Distances Between Objects, Institute of Optical-Neural Technologies Russian Academy of Sciences, Moscow

Martin W¨ollmer, Felix Weninger, Florian Eyben, Bjorn Schuller (2011), Acoustic-Linguistic Recognition of Interest in Speech with Bottleneck-BLSTM Nets, Institute for Human-Machine Communication, Technische University at Munchen, Germany

Martin Wollmer ¨, Florian Eyben1, Bjorn Schuller ¨, Ellen Douglas-Cowie2, Roddy Cowie2 1Technische Universitat M¨ unchen, Institute for Human-Machine Communication, ¨80290 Munchen,(2011) Germany “Clustering in Emotional Space for Affect Recognition Using Discriminatively Trained LSTM “

Mohamad Adnan Al-Alaoui, Lina Al-Kanj, Jimmy Azar, and Elias Yaacoub(September 2008) “Speech Recognition using Artificial Neural Networks and Hidden Markov Models” IEEE MULTIDISCIPLINARY ENGINEERING EDUCATION MAGAZINE, VOL. 3, NO. 3

M. M. El Choubassi, H. E. El Khoury, C. E. Jabra Alagha, J. A. Skaf and M. A. Al- aloui “Arabic Speech Recognition Using Recurrent Neural Networks”

MRA (2010), Incorporation of dynamic parameters in hybrid feature based bangle phoneme recognition using multilayer neural networks

N.Uma Maheswari, A.P.Kabilan, R.Venkatesh(DEC 2010), A Hybrid model of Neural Network Approach for Speaker independent Word Recognition

Oparin, M. Sundermeyer, H. Ney, and J.-L. Gauvain, (Dec 2012) “Performance analysis of neural networks in combination with n-gram language models,” in Proc. of ICASSP’ pp. 5005–5008.

P. Xu and F. Jelinek,(2007) “Random forests and the data sparseness problem in language modeling,” Computer Speech & Language, vol. 21, no. 1, pp. 105–152.

S. Parveen, PD Green “Speech Recognition with Missing Data using recurrent neural net ”, Speech and Hearing Research Group Department of Computer Science University of ShefÞeld ShefÞeld S14DP, UK

TomasBrno(July 2010),”Recurrent neural network based language model”, University of Technology Johns Hopkins University,

T. Mikolov, M. Karafiat, L. Burget, J. ˇCernock´y, and S. Khudanpur (oct 2010) “Recurrent neural network based language model,” in Proc. of Interspeech’ , pp. 1045–1048.

Vibha Tiwari (2010),”MFCC and its applications in speaker recognition system”, Deptt of electronics Engg, Gyan Ganga Institute of technology and management, India

Xin-gaung , Li Sch of Inf(2011) “ Speech recognition based on K-mean Clustering and neural network”, On seventh international conference. ICNC, volume-2, p-614