IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 4 Issue 1, January 2017

ISSN (Online) 2348 – 7968 | Impact Factor (2015) - 4.332

www.ijiset.com

A Study on Feature Selection, Clustering Techniques for Stuttering Identification in Automatic Speech Recognition

M.A. Josephine Sathya1*, Dr.S.P.Victor2
1Research Scholar, Mother Teresa Women's University, Kodaikanal , Tamilnadu , India.
2Associate Professor & Head, Department of Computer Science, St.Xavier's College (Autonomous), Palayamkottai, Tamilnadu , India, drspvictor@gmail.com.

Abstract

Speech Recognition technology has evolved for more than fifty years spurred on by advances in acoustic processing of signals, developing algorithms, designing architecture and hardware. Speech Recognition Technology it has gone from a laboratory level to a science, and eventually to a full fledged advanced technology that is practiced and utilized by a many people like research scientists, linguists, engineers and system designers. Many researchers enrich the technology and yet so much of speech applications are to come over the upcoming years. Recent development in automatic speech recognition (ASR) technology leads to the need for development of many sophisticated interesting and more accurate speech recognition applications. Much effort was involved to increase the accuracy of the speech recognition system but still erroneous output is generating in current speech recognition systems. This paper includes the ASR technology which is used to create the training system which was used for giving training for the children with fluency disorders. Section I describes the ASR system, Section II describes the feature selection, Section III describes the Confidence measures, and Section IV describes the Clustering Techniques.

Keywords *– Automatic Speech Recognition, feature selection, confidence measures, clustering Techniques.*

1. Introduction

Automatic Speech Recognition can be utilized as a user interface for the system device and is also responsible for data Input / Ouput modality between the user and the required application. System now –a– days are to act as an intelligent terminal that makes a smooth move for the end – user and the applications are to deal with robust speech which may include pauses , soar throat speech, silent start of the speech, hesitation, fast speech, disrupted speech (stuttering) and various other phenomena that naturally occur in any speech.

Any ASR system can have

1. Spontaneous or continous speech
2. Semantics in speech
3. Environmental disturbance in speech

1.1 Spontaneous or Continous speech

Eventhough robust, continous or spontaneous ASR is eagerly expected by many people. Current ASR systems deal with read only speech because spontaneous speech is often not based on grammar and are not properly structured.

1.2 Semantics in speech

For any specific task, the perplexity, average number of words active at any given time , can be reduced sustanially if semantic and pragmatic information is used in addition to grammar.

1.3 Environmental disturbance in speech

Environmental disturbances degardes the performance of any ASR systems thereby producing poor acoustical results. Thus it becomes a necessity to have similar acoustical environment so that the training and testing environments are the same. It would be desirable to have a system that works independently of recording conditions (using different rooms , microphones , noise levels).Quality of the input speech plays another vital role in low performance of SRS system. Speech recognition is affected by the quality of the input because of the speakers distorted voice or variation of voice the speaker or some environmental disturbances.

Any SRS is based on acoustic signal processing and it is the process of converting acoustic signal captured by a microphone to a group of words. SRS is classified into two different types namely speaker – dependent and the other is speaker – independent.

Speaker – dependent software specifically designed for unique characteristics of a single person's voice, in a way similar to voice recognition, new users must first "train" the software by speaking to it, so the computer can analyze how the user utters the word.

Spontaneous speech includes disfluencies and is become tedious to recognize than the normal speech. In some SRS, user must be previously practiced to the system. ie., the recognition engine must be trained many time regarding a particular voice.

2. Feature Selection

Feature selection is one of the important and frequently used techniques in data preprocessing. Feature selection reduces the number of features , removes irrelevant, redundant or noisy data and brings the immediate effects for applications typical feature selection process consists of four basic steps as shown in Figure 1.1 , namely, subset generation , subset evaluation , stopping criterion and result validation [1]. Subset generation is a search procedure that produces candidate feature subsets for evaluation based on a certain search strategy. Each candidate subset is evaluated and compared with the previous best one according to a certain evaluation criterion. If the best subset turns out, it replaces the previous subset. The process of subset generation and evaluation is repeated until a given stopping criterion is satisfied. Then the selected best subset usually needs to be validated by prior knowledge or different tests via synthetic and/or real world data sets.

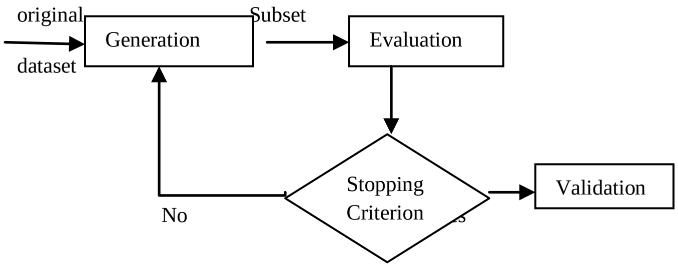

Figure 1.1 A Traditional Framework for Feature Selection

2.1 Subset Generation (SG)

It is a search procedure aimed at producing a subset of features for evaluation using specific search strategy[2]. SG is initiated with one of the 3 ways

1. No features
2. All features
3. Random subset of features

2.2 Subset Evaluation (SE)

It is measured goodness of a subset produced by the subset generation procedure and the value is compared with the previous available best. Typically, an evaluation function tries to measure the capability of a feature or a subset to distinguish among different class labels. If the generated value is better when compared with the previous best, then the new value is used to replace the previous best value of the subset [3, 4, 5].

2.3 Stopping Criterion

It involving a process of subset generation and evaluation, which is repeated until a satisfactory degree of given stopping criterion. In the absence of a suitable stopping criterion, the feature selection process may run exhaustively or forever through the space of subsets. The choice for a stopping criterion is influenced by two major factors such as

1. Generation procedures
2. Evaluation functions

Of these two, stopping criterion based on the former factor includes a predefined number of features that are selected or a predefined number of iterations made. Similarly stopping criterion based on the latter includes whether an addition of any feature does not produce a better subset or an optimal subset according to some evaluation function is obtained. The iteration loop continues until it reaches a satisfactory stopping criterion and the feature selection process halts by generating output of a selected subset of features to a validation procedure[1].

2.4 Subset Validation

Where the best subset thus generated is given as an output to the validation processs. Generally ,the fitness of a selected feature subset feature subset is evaluated based on the classification accuracy of the induction algorithm. There are several methods used for such evaluation, including Holdout Sampling , Cross Validation , Leave-One-Out, and Bootstrap Sampling[6].

2.5 Approaches in feature selection

Two approaches are there in the feature selection

1. Forward Selection
2. Backward Elimination

1) Forward Selection

It Starts with no variable and adds then one by one , at each step in such a way that the addition of the variable and adds then one by one , at each step in such a way that the addition of the variable decreases the error. This process continues till further addition of the variable does not decrease the error significantly.

2) Backward elimination

It considers all the variable initially and each variable is removed one by one at each step in such a way that the removal of a variable decreases the error.

2.6 Confidence Measures (CM)

In speech recognition, CM is used to measure reliability of recognition results. CM enables us to assess the output of SRS. CM provides us with an estimate of the probability that a word in the recognizer output is either correct or incorrect. Cm ia a quantitative estimate of a word’s correctness. CM is a number between 0 and 1 indicating our degree of belief that a unit output by a recognizer (phrase , word , phone etc.) is correct. The most important application of CM is in speech dialogue systems (e.g..) ticket booking , call routing , information provision etc.Errors can be disastrous in a recognition system , but confirmation of each content word is tedious. System can use CM to decide which words are correct and which words need to be confirmed or corrected.

Let us consider a recognizer generating a sequence of hypothesized word tokens $W_{i\,,}$ i = 1, …., n. Associated with these tokens are a sequence of class labels named as C_i , where C_i is defined as

$$Ci \;=\; \begin{cases} 1 \,;\ \text{if } W_i \text{ is correctly recognized} \\ 0 \,;\quad \text{otherwise} \end{cases}$$

Associated with the i^{th} token is its true posterior probability p_i defined as

$$p_i = P\left(C_i = \frac{1}{x} \right) \qquad (1)$$

` Where p_i is the probability that the i^{th} token is correctly recognized , given all the information about the recognition process, denoted by X. The notation $p_i(x)$ is used to indicate that p_i is a function of the feature vector x , $x \in X$.

The true posterior probability p_i is unknown a priori and must be estimated. A CM for the recognized token W_i is an estimate of the true posterior probability $p_{i.}$ Use of acoustic CM [7] can be very useful for most ASR applications. It provides the necessary information and of great help to be able to predict whether a hypothesis provided by an ASR system is correct or not. For instance, high level dialogue systems can be significantly improved if there is a good idea of recognition accuracy.

3. CLUSTERING

A clustering technique as the name implies rely upon some physical parameters of the data elements in a dataset to aggregate similar data into groups or clusters by its affinity with distance values and quality. It is widely used in scientific analysis where no previous knowledge on dataset exists and to gather an assumptive grouping based on certain similarities. Clustering technique could help to evolve in a decision system with only input information and without any decision attribute for information retrieval [8].

3.1 Types of clustering approach

Two methods of clustering approach namely

1. K – Means algorithm
2. Fuzzy C – Means (FCM) algorithm

These algorithm are used along with rough set theory for reducing feature size of the datasets. These two algorithms are implemented in original dataset without considering the class labels and further rough set theory is implemented on the patitioned dataset to generate feature subset after removing the outliers by using KNN.

3.1.1 K – Means Clustering

K – Means clustering is one of the most commonly used clustering algorithm developed by MacQueen in 1967, which is most suitable for smaller datasets to partition the dataset into K – clusters. The name K – Means originates from the means of the K clusters that are created from n objects. The term related to K – Means “associated cost function” is defined in terms of the distances between the cluster objects and the cluster center. The objective is to find k partitions that minimize the cost function, which is explained as below.

Given an initial set of k means (centroids) $m_i^{(1)}, \dots \dots \dots \dots . m_k^{(1)}$, the algorithm proceeds by alternating between two steps as in (2) , (3):

1) Assignment Step

Assign each observation to the cluster with the closest mean

$$S_i^{(t)} = \left\{ \begin{matrix} x_j : \left\| x_j - m_i^{(t)} \right\| \le \left\| x_j - m_i^{(t)} \right\| \\ for\, all\, i^{*} = 1, \dots \dots , k \end{matrix} \right\} \qquad (2)$$

2) Update step

Calculate the new means to be the centroid of the observations in the cluster.

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \in S_i^{(t)}} X_j \qquad (3)$$

The algorithm is deemed to converge when the assignment no longer change . In the research Work , it is decided to run K – Means algorithm multiple times or iteratively using different initial states and from the resultant data , the lowest error state is considered for further evaluation.

3.1.2 Fuzzy C – Means Clustering (FCM)

FCM is the most classical method for fuzzy clustering , which assigns data to multiple clusters at different degrees of membership[9].The FCM algorithm leads to minimize the following objective function:

$$J_{FCM} = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i\,j}^{m} \left\| x_i - c_j \right\|^2 \qquad (4)$$

Where $1 \le m < \infty$ is the fuzzifier , c_j is the i^{th} cluster center, U_{ij} is the degree of membership of x_i in the cluster j , x_i is the i^{th} d – dimensional measured data and the center.Fuzzy partitioning is carried out through an iterative optimization of the objective function shown in (4) , with the update of the parameters u_{ij} and the cluster center c_j by

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} \qquad (5)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \qquad (6)$$

This iteration will stop when $\left\| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right\| < \in$, where $\in$ is a termination criterian between 0 and 1 , and k is the iteration step. This procedure converges to a local minimum or a saddle point of $J_{FCM.}$

The algorithm proceeds as follows

(i) Initialize U = $u_{ij}\ matrix\ , u^{(0)}$

(ii) At k^th step calculate the center vectors
$C^{k} = [c_j]\ with\ U^{(k)}\ by\ (6)$

(iii) Update $U^{(k)}\ , U^{(k+1)}\ by\ (5)$

(iv) If $\left\| U^{(k+1)} - U^{(k)} \right\| < \in then\ stop;$
Otherwise repeat steps (ii) and (iii).

4. Conclusion

ASR generally order the hypotheses by computing scores for each of the utterance hypotheses. For application to act on speech input, the applications must be able to assess the confidence that the input has been decoded in a proper manner. CM aids in the selection of multiple hypotheses. Various means of evaluating the performance for the effectiveness of CM and the notion of characterizing CM in terms of their discrimination power and bias are described here. While for the recognition of isolated word , phone – based measures give better results ,in continuous speech recognition. In this paper it is analysed that confidence measures are used to to do the analysis of confidence annotation for spoken language systems at different levels : word , utterance and concept levels.Use of CM as heuristic to combine several hypotheses from different recognizers yield expected results in stuttering recognition.Future work leads for the utterance level the detection of , not only out of domain , but also misunderstanding utterances is to be considered because of poor recognition.

References

[1] Huan Liu and Lei Yu, " Towards Integrating Feature Selection Algorithms for Classification and Clustering ", IEEE Transactions on Knowledge and Data Engineering , Vol. 17, pp. 491-502, 2005.

[2] Langley P, "Selection of relevant features in machine learning", AAAI Fall Symposium Series on Relevance, Menlo Park: AAAI Press, Pp. 140 – 144 , 1994.

[3] Liu H , Motoda H and Yu L, "Feature Selection with selective sampling", In proceedings of the Nineteenth International Conference on Machine Learning , Pp. 395- 405, 2002.

[4] Hall M A , "Correlation- based feature selection for machine learning" , In Proceeding of the Seventeenth International Conference on Machine Learning, Pp. 359 – 366, 2000.

[5] Liu H and Setiono R, "Feature Selection and classification – a probabilistic wrapper approach", In proceedings of Ninth International Conference on Industrial and Engineering Applications of AI and ES, Fukuoka, Japan, Pp.419 – 424, 1996.

[6] Kohavi R, "A Study of cross-validation and bootstrap for accuracy estimation and model selection", In C. Mellish (Ed.) Proc of the Fourteenth International Joint Conference on Artificial Intelligence, Pp. 1137 – 1145, 1995.

[7] Kamppari. S. O and Hajen . J.J, "Word and phone level acoustic confidence scoring" , In Proceedings of ICASSP , Istanbul , Turkey , Pp. 1799 – 1802, 2000.

[8] Thangavel K , shen Q and Pethalakshmi A , "Application of clustering for feature selection based on rough set theory approach" , AIML Journal , Vol 6 , No.1, Pp.19 – 27 , 2006.

[9] Bezde JC , "A review of probabilistic , fuzzy and neural models for pattern recognition" , J Intell Fuzzy Syst, Vol.1 , No.1 , Bioinformatics, 1995.