Belarussian Speech Recognition Using Genetic Algorithms

Eugene I.Bovbel[1] , Dzmitry V.Tsishkou[1]

[1] Radiophysics Department, Belarussian State University
F.Skoryna Av.,4 220050, Minsk, Belarus
tishkovdv@mail.com

Abstract. One of the factors complicating activity with speech signals, is its large degree of acoustic variability. To decrease influence of acoustic variability of speech signals it is offered to use genetic algorithms in speech processing systems. We constructed program model which implements technology of speech recognition using genetic algorithms. We made experiments on our program model with database of separated Belarussian words and achieve optimal results.

1 Introduction

The essential successes on a way of construction of devices which are capable to perceive speech are recently reached. First of all this success is connected to development and intrusion of statistical methods of speech recognition [1,2]. Statistical methods, founded on HMM [3,4] are widely used for the solution of practical problems. For example, recognition of phonemes, insulated words and continuous speech, identification and verification of the speaker, diagnostic of diseases of a talking path. The flexible and potent mathematical vehicle of HMM [4,5] allows easily to use for effective simulation of variability of a speech signal. One of the factors complicating activity with speech signals, is its large degree of acoustic variability. The large degree of acoustic variability of speech signals is conditioned by a series of the causes. The first is that the speech of each separate person is determined by physical properties of a talking path, such as length of a voice channel, size of a larynx and etc. In - second, the generating of speech [2] is impossible without a verification of compliance above organs of speech, features of that are determined on learning process of the concrete person and represent a dialect or regional accent. The third, it is known, that the availability of irregularity of a voice channel is mirrored in features of speech of the person. For decreasing influencing of acoustic variability [1,2,3] of speech signals it is offered to use genetic algorithms in speech processing systems. The philosophies which were fundamentals of genetic algorithms, were formulated in activities of Mendel, abbot of a monastery of city Brno (Czech Republic). Now under the term « genetic algorithms « - not one model hides, because a broad enough class of computing algorithms [6] it contains, sometimes it is not enough similar against each other. It is necessary to mark, that the genetic algorithms have a series of properties, which one make by their rather attractive for speech signals processing. To these properties it is possible to relate:

- The genetic algorithms are based on principles of population genetics and are very friend for simulation not only acoustic variability, but also non-steady signals, to which one, the speech signal concern also.
- The genetic algorithms tend to find global or strong local optima of a given function in their search domain.
- The genetic algorithms can optimize multimodal and discontinuous functions without the need for any gradient information whatsoever.

2 Genetic Algorithms

Let separate voice expression is shown by sequence of acoustic vectors: $O = \{O_t \mid t = 1,...,T\}$ The vector of observation O_t which one is usually used for recognition of words looks like: $O_t = \{X_n \mid n = 1,..., N\}$, where X_n - some coefficients, usually it is customary cepster and deltacepster coefficients [4,5]. The genetic algorithm usually use representation of this vector by the way chromosome: $M = \{g_i \mid i = 1,...., N\}$, where each parameter X_n is described by gene g_i. Each a gene g_i consists of alleles, which one are selected from the final alphabet. It is necessary to have a function $f(X)$, which one estimates as far as the combination of parameters presented by vector X_n.

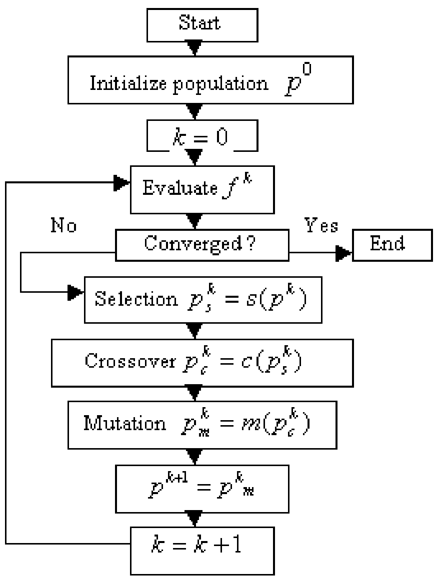

Fig. 1 The diagram of activity of genetic algorithm.

Criteria of an estimation of X_n is maximization of $f(X_n)$ .The genetic algorithm works with set of chromosomes. The set of chromosomes is called as a population and is meant: $P = \{M_l \mid l = 1,2,..., L\}$, where L - size of a population.

The procedure of transition from one population to other performs series of applications of the genetic operators:

- Selection, which implements the genetic algorithm analog of fitness-based survival
- Crossover, which hybridizes surviving designs to produce better designs
- Mutation, which randomly perturbs designs to ensure lasting population diversity

This process is fully described up in the following articles [6]. The activity of genetic algorithm on the diagram guesses availability of criteria of convergence. As the given criteria we will use an estimation of change of a average fitness on all population [6]. In cases when during some number of breeds the given criteria varies insufficiently essentially it is possible to speak about finish of genetic algorithm's work.

3 Hidden Evolution Method

At constructing systems of speech recognition it is necessary to utilize express algorithms at a stage of processing and analysis for coercion of input data in the form, which will provide the stable characteristics and will allow to reduce influence of some variable parameters. One of algorithms, which will be utilized in the system of speech recognition - method of the hidden evolution. We have developed the given method, as the procedure of analysis of the arrays of decimal numbers, and therefore it is algorithm, independent of a context. Any map can be represented as a sequence of pairs of decimal numbers. Let's consider this sequence from the point of view of fundamental bases of genetic algorithms. Let pair of numbers represents number m of the unique representatives, which have survived as a result of evolution, of generation number m . Also a pair of numbers number $m+1$ represents the unique representatives, which have survived as a result of evolution, of generation number $m+1$. Let's assume, that the pair of numbers m is the ancestors of a pair of numbers number $m+1$, within the framework of accuracy with which we can install it. Being grounded on accessible information, the method of the hidden evolution reconstructs the evolutionary process, which is hidden from us, and saves its parameters in the compact form, which has received a title a matrix of evolution. As well as the genetic algorithm, method of the hidden evolution follows from one generation to another, but thus the amount of generations is uniquely determined by parameters of the map and algorithm. The stage of reconstruction for each generation has the identical scheme, in which all possible processes of crossover and mutation are simulated and the data on the process for which connection between generations are saved is closest to actual data. We have utilized value of the operator one - point crossover and two values of operators one - point a mutation, as values of indexes of two units on string and on columns accordingly for addition to value of the given unit of unity on each generation. The data on the process of reconstruction are brought in a matrix of evolution, as additive values. After that the matrix of evolution passes normalization.

Fig. 2 Evolution Matrixes

4 Pre-classification Stage

At implementation of the system of speech recognition with usage of genetic algorithms we use some stages of analysis and processing of the map. One of them is the pre-classification stage. The given stage is intended for implementation of the procedure to optimization of clusterisation - middle of classes are better remote from each other, and the objects inside classes concentrate to middles of classes more strongly in the space of signs. Here we solve a problem of decrease of the bridging of clusters and, therefore, complexity of a problem of recognition is reduced. We implemented the given procedure being grounded on methods of transition from characters of one alphabet to characters of another. Set of possible implementations of the given method allow to encode all possible methods of transition at the fixed alphabets under the particular scheme and to utilize genetic algorithm for implementation of the procedure of optimization of clusterisation. Other essential advantage of the given approach consists that at lowering an amount of characters of the alphabet and implementation of the procedure of pre-classifications we save the unique information on each object and we reduce memory size necessary for its storage. Standard unit we obtain as average on all accessible units of matrixes of evolution with equivalent indexes. As outcome of work of the procedure of conversion from the infinite alphabet of characters in the binary alphabet of characters on all accessible indexes we obtain a binary vector. As for the binary alphabet and templates of dimension $n*n$ exists 2^{n*n} of different variants of the function of conversion F , we use genetic algorithm for search of a best value of units of the function of conversion. A measure of correspondence of solution optimum is distance between binary vectors.

5 Speech Recognition Using Genetic Algorithms

We use conditionally - independently properties of the object as streams. The procedure of learning and recognition passes on each stream irrespective of remaining streams and further integral analysis of outcomes follows. At multistream speech recognition system we have used concept of the frames, as the scheme of creation of streams of the maiden level. As a result of the preliminary analysis of an original vector of observations (containing 12 cepsters and 12 delta - cepsters at each discrete time position), we broke it into the 2d frames. At a restricted information quantity, because of a digital signature, one of ways of escalating of quantity of streams is usage of the frames with overlap. The optimum selection of a ratio of the sizes of overlap and frame allows to speak about a maximum approaching of streams to independent tags. The outcomes data confirmed by the analysis of comparative parameters multistream simulators and hardware systems of speech recognition, which one we have elaborated. Each frame serves input data for the conforming stream after normalisation stage based on an asymmetry measure. For conversion of the frame to a matrix of tags we have used a method of hidden evolution. We have elaborated the given method for speech analysis. The method of hidden evolution will use base concepts of population genetics [6] and analyse the frame as a set chromosome. Outcome of the analysis is the preservation of unique characteristics for each frame in the compact form by the way matrixes of tags of the size 8x8. After that a matrix of tags to become with input dates for the subsequent stage of classification. Further with the help of the coded bank of filters and fuzzy logic we transmit a matrix of tags of each stream in to binary vector, size 64. Using genetic algorithm, which one is specially updated for the given operation, we optimise the scheme of conversion on a grade level so, that the binary vectors of one class become closer to each other and further from the quoters of diverse classes. The binary vectors are used of input dates of each stream. We use binary classifiers as structure, which one stores the indispensable information for process of identification and quantity of computing resources on some orders smaller than for more composite discriminatory analysis's demands.

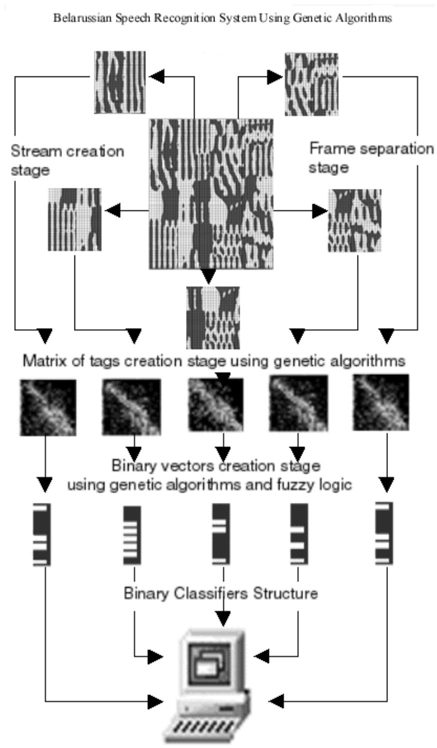

Fig. 3 Belarussian speech recognition system using genetic algorithms, where vector of observations is shown in the 2d bottom view projection.

6 Experimental Setup

For realization of experiment the database was built. 20 Belarussian words were recorded till 50 times by ten speakers with sampling rate 44100 Hz. For segmentation was used 30ms Hamming window with the size of frame 10ms. For each record of the database were constructed of a vector of the observation consisting from 12 cepstral and 12 delta – cepstral coefficients. Next we implement genetic algorithms technique for speech signals recognition system. Ten speech signals records (from different speakers) of each word were used for training and forty for testing. We get following experimental results, testing all words from the database:

- Amount of memory to store information about each record – 64 bit
- System does not need to rebuilt itself to add new word record
- Computational time to add each record to genetic algorithms speech recognition system's structure – 3 sec.
- Computational time to recognize each record – 0.1< sec.
- Recognition rate under conditions listed above (on the same database used for training and recognition) – 100%

All tests were made on pentium2-400 class workstation.

References

1. Jelinek F. «Распознавание непрерывной речи статистическими методами". ТИИЭР,1976 т.64 N 4 с 131 – 161. in rus.
2. Lawrence Robber and Biing – Hawing Jung. "Fundamentals of Speech Recognition". PTR Prentice – Hall Inc, Englewood Cliff, 1993.
3. Hero, A.Bourland and Nelson Morgan "Connectionist Speech Recognition: A Hybrid Approach" Kluwer Academic Publishers, Boston MA,1994.
4. E.I.Bovbel, P.P. Tkachova, and I.E.Kheidorov. "The usage of hidden Markov models based on autoregressive principles for isolated word recognition" Proc.of.SPIE 13th Annual Intern. Symposium AeroSense, vol.3720, April 5-9, 1999, Orlando, Florida, USA.
5. E.I.Bovbel, P.P. Tkachova, and I.E.Kheidorov."The analysis of speaker individual features based on autoregressive hidden Markov models", Proc. Of Eurospeech-99, Budapest, Hungary, September 22-26, 1999.
6. E. Michielssen, Y. Rahmat-Samii and D.S.Weile."Electromagnetic systems design using genetic algorithms". Modern Radio Science. Oxford University Press. P 91 – 123.