Evolutionary Algorithms for Optimizing Speech Data Projection

A. Spalanzani
CLIPS/IMAG, BP 53, 38041
Grenoble cedex 9 FRANCE
Anne.Spalanzani@imag.fr

S.A. Selouani
LCP Institute of electronics,
USTHB, BP 32 El Alia-Algiers-
ALGERIA

H. Kabré
CLIPS/IMAG, BP 53, 38041
Grenoble cedex 9 FRANCE

Abstract

We present an approach which limits significantly the drop of performances related to Automatic Speech Recognition Systems (ASRSs) caused by acoustic environment changes by combining a Principal Component Analysis (PCA) and Evolutionary Algorithms (EA) in order to transform the noisy acoustic environment into a predefined and well-known environment. The results show that in noisy and changing environments, the proposed PCA/EA optimized systems achieve high recognition rate compared to the baseline system

Present systems of automatic speech recognition remain confronted to the problem of input conditions changes. These changes are mainly due to room acoustics, ambient noise, speaker variability and microphone characteristics. In order to face this difficulty, numerous techniques have been developed (Junka and Haton 96). Despite various efforts to address various aspects of the robustness question, the problem of the adaptation to changing environments remains one of the main obstacle to the deployment of speech recognition systems in real world applications. In a previous work (Spalanzani & Kabré 98), we applied Evolutionary Algorithms (EA) to evolve the speech recognizer, in this paper, we propose to deal with the signal itself.

Our investigated approach can be viewed as a signal transformation through a mapping transformation using a Principal Component Analysis (PCA) and EA (that are Genetic Algorithms (GA) and Evolution Strategies (ES)). This transformation attempts to achieve an auto-adaptation of Automatic Speech Recognition Systems (ASRSs) under adverse conditions. The underlying idea consists in giving ASRSs confronted to adverse conditions the ability to recover an acoustical environment which is as close as possible to those of the learning phase. This is proposed to be done by representing the data corrupted by noise (telephone ring, people speaking, etc.) in a subspace generated by principal components extracted from data acquired in a canonical environment (i.e. without noise).

The retained principal components are modified by genetic operators such as mutation and crossover. The resulted basis vectors are used to project degraded input data. The aim is to find projected noisy data as close as possible to clean data. The task for ES or GA is to search through all the rotation axes defined by PCA. Evolution is driven by a fitness function defined in term of ASRS recognition rate.

According to a classical EA algorithm, the search of the best projecting space is operated. After the creation of the clean corpus (i.e speech data wihout noise projected in a PCA space), the ASRS, based on a connectionist model, is trained until it has converged. The Principal Components used for the projection of the corpus allow the creation of a population of individuals able to evolve to adapt to the noisy environment. The best projections basis are selected for the mating. Two methods are compared: classical GA and (μ,λ)-ES with a gaussian mutation. Results obtained (Figure 1.) show an improvement of both GA and ES technics. After 500 generations, GA's performances reach 45.8% of speech recognition whereas ES's performances reach 52.2% in 100 generations only, which improves the speech recognition rate of 22% compared to the baseline system (PCA without evolution).

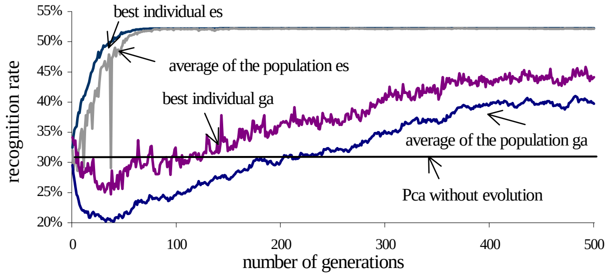

Figure 1. Evolution of PCA/EA performances

References

J.C. Junka and J.P. Haton (1996). *Robustness in automatic speech recognition*. Kluwer Academic publisher.

A. Spalanzani and H. Kabré (1998). *Evolution, Learning and Speech Recognition in Changing Acoustic Environment*. Parallel Problem Solving by Nature (PPSN'98), Springer Verlag, pages 663-671, Amsterdam, September, 1998.