

An Open-set Speaker Identification System using Genetic Learning Classifier System

WonKyung Park
Syracuse University
Syracuse, NY 13244, USA
wpark01@syr.edu

Jae C. Oh
Syracuse University
Syracuse, NY 13244, USA
jcoh@ecs.syr.edu

Misty K. Blowers
AFRL/IFEC Rome, NY 13441, USA
misty.blowers@rl.af.mil

Matt B. Wolf
Syracuse University
Syracuse, NY 13244, USA
mbwolf@syr.edu

ABSTRACT

This paper presents the design and implementation of an *adaptive open-set speaker identification system* with genetic learning classifier systems. One of the challenging problems in using learning classifier systems for numerical problems is the knowledge representation. The voice samples are a series of real numbers that must be encoded in a classifier format. We investigate several different methods for representing voice samples for classifier systems and study the efficacy of the methods. We also identify several challenges for learning classifier systems in the speaker identification problem and introduce new methods to improve the learning and classification abilities of the systems. Experimental results show that our system successfully learns 200 voice features at the accuracies of 60% to 80%, which is considered a strong result in the speaker identification community. This research presents the feasibility of using learning classifier systems for the speaker identification problem.

Categories and Subject Descriptors: I.2.6 [Natural Language Processing]: speech recognition and synthesis

General Terms: Experimentation, Design

Keywords: Classifier systems, Genetic algorithms, Linguistics, language and speech, Machine learning

1. INTRODUCTION

This paper discusses the design and implementation of an adaptive speaker identification system based-on genetic learning classifier systems (LCS). The speaker identification (SID) problem can be categorized to *text-dependent* and *text-independent*. In text-independent speaker identification, the system can identify the speakers regardless of the words spoken, while text-dependent identification requires speakers to speak predetermined words in order to be recognized.

In the *closed-set SID system*, the system can identify a fixed set of speakers. Unlike the *closed-set speaker identification problem*, the *open-set speaker problem* does not assume

the number of speakers is fixed. An open-set SID system should also be able to add new speaker profiles as needed dynamically.

Genetic rule-based learning classifier systems (LCS) [5, 6], are open-ended adaptive learning systems that can learn new rules for changing environments. This feature motivated us to use LCS to the speaker identification problem. For the LCS model, we have chosen the XCS classifier system [1] since XCS provides several improvements to the traditional LCS and it is more recently studied and updated. To our best knowledge, our system is the first LCS-based speaker identification system. We refer to our system as Speaker-Identification Learning Classifier System (SID-LCS).

One of the issues in our research was in representing real-number voice features to an appropriate classifier format. We've studied several different encoding methods including hyperrectangular, hyperspheroidal, and general hyperellipsoidal methods [3, 1]. This paper also presents an algorithm for open-set discrimination and studies new techniques for improving the speaker identification capability of SID-LCS.

We've conducted experiments to test the learning capability of the system in terms of the number of voice features the system can learn. We further tested the system's ability in the open-set discrimination task.

2. THE SPEAKER IDENTIFICATION LEARNING CLASSIFIER SYSTEM (SID-LCS)

The current SID-LCS is based on the accuracy-based XCS [2]. Since voice features are real numbers, in our first attempt, encoded them in hyperrectangular conditions [3, 8], one of the most common ways of encoding real-valued inputs for LCS. We also experimented with several different encoding methods including hyperspheres, hyperellipsoids, and general hyperellipsoids [1]. We explain the SID-LCS in detail including how voice features are sampled and how the learning components are modified for the SID problem.

2.1 Voice Feature Extraction

Voice feature extraction is done by the LPC model [4]. The speaker's voice is sampled for 4 seconds and this data goes through the LPC algorithm explained above. The re-

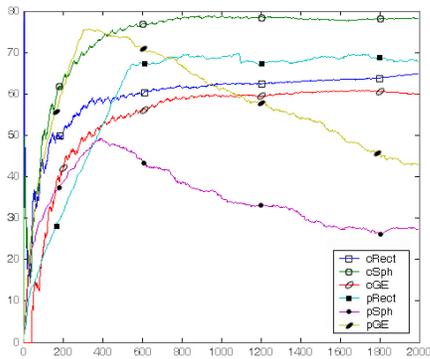


Figure 1: Comparisons between various encoding methods (4 males, 20 feature vectors for each).

result is a 14 real number vector. These 14 numbers are presented to SID-LCS’s input interface and they are matched against the condition part of the classifiers in the current population. We explain various encodings methods for real numbered inputs in the next section.

2.2 Rule Encoding Methods for Speaker Identification

Many systems traditionally use hyperrectangular conditions for real number inputs. Butz [1] introduces more expressive methods based on spheres and hyperellipsoids. We experimented with hyperspheroidal, hyperrectangular, hyperellipsoidal, and general hyperellipsoidal methods for our SID-LCS systemf which is implemented as in [1].

2.3 The Open-set Decision Algorithm

We have introduced an algorithm for the open-set decision for SID-LCS. The system makes the open-set decision based on the prediction values of the classifiers in the prediction array. (See [1] for details.) At each classification phase, the prediction array contains values from the classifiers that match with the current voice feature. The values in the prediction array are sorted according to their “best guess” for the correct action, and then the action of the highest-ranked value is chosen.

The experiments for open-set are conducted after training the system with in-set voice features. System doesn’t learn any new information from open-set data. Thus, no reward is given for the right guesses.

3. EXPERIMENTS

We have used the TIMIT voice data from DARPA [7] for training and testing our SID-LCS system. For each speaker, the data provides 20 feature vectors.

First, we’ve tested three encoding methods with the SID-LCS system: hyperrectangular, hyperspheroidal and general hyperellipsoidal encoding methods. Figure 1 shows the percent correct and the population size in tens for each encoding method over iterations. As shown in the graph labeled cSph, the SID-LCS with hyperspheroidal conditions can identify 80% of 80 feature vectors from four speakers.

For open-set classification tests, a set of ten male speaker voice vectors are chosen as *in-set speakers* for each, randomly from the TIMIT data. The system is then trained with the

ten in-set speakers and tested with 20 speakers that include the ten in-set speakers and ten additional speakers as out-set. Table 1 shows the results.

	False Reject	False Acceptance
Set1	29 (14.5%)	48 (24.0%)
Set2	25 (14.5%)	65 (32.5%)
Set3	25 (12.5%)	59 (29.5%)
Set4	29 (14.5%)	53 (26.5%)

Table 1: Open-set classification results. The results are False Reject (FR) and False Acceptance (FA) for 200 test cases.

Open-set capability can be improved in various ways. There are many possible directions of research for the matter as discussed in the next section.

4. CONCLUSIONS AND FUTURE WORK

We’ve presented a new text-independent speaker identification system using genetic learning classifier systems. Generally, LCS are used for symbolic learning domains but the speaker identification problem requires systems to deal with real-valued input voice features.

We’ve introduced several encoding methods that are suitable for the speaker identification problem: the hyperrectangular, the hyperspheroidal, hyperellipsoidal, and the general hyperellipsoidal encoding methods [1]. Experiments are conducted for the closed-set and the open-set speaker identification problems. The results are comparable to recent statistical clustering methods as discussed.

There are several future research directions in improving the system’s closed-set and open-set classification capabilities. For one, *temporal rule links* can be considered. The new LCS-based speaker identification system have advantages over common numeric-based methods in that they are capable of adding and deleting speaker profiles at any time. Integrating the LCS-based system with higher-level semantic knowledge should also be relatively easy.

5. REFERENCES

- [1] Martin V. Butz, *Kernel-based, ellipsoidal conditions in the real-valued xcs classifier system*, GECCO **June 25-29** (2005), 1835–1842.
- [2] Martin V. Butz and Stewart W. Wilson, *An Algorithmic Description of XCS*, Tech. Report 2000017, Illinois Genetic Algorithms Laboratory, 2000.
- [3] Arthur L. Corcoran and Sandip Sen, *Using real-valued genetic algorithms to evolve rule sets for classification*, IEEE Conference on Evolutionary Computation (Orlando, FL), 1994, pp. 120–124.
- [4] John Grieco, *Advanced short segment language identification*, Tech. Report AFRL-IF-RS-TR-2005-344, Air Force Technical Report, 2005.
- [5] J. H. Holland and J. S. Reitman, *Cognitive systems based on adaptive algorithms*, Pattern-Directed Inference Systems (D. A. Waterman and F. Hayes-Roth, eds.), Academic Press, New York, 1978.
- [6] S. F. Smith, *A learning system based on genetic adaptive algorithms*, Ph.D. thesis, University of Pittsburgh, Pittsburgh, PA, 1980.
- [7] DARPA TIMIT speech database, <http://www.mpi.nl/world/tg/corpora/timit/timit.html>.
- [8] Stewart W. Wilson, *Get real! XCS with continuous-valued inputs*, Festschrift in Honor of John H. Holland, Center for the Study of Complex Systems, 1999, pp. 111–121.