Е. А. Попов, М. Е. Семенкина, Л. В. Липинский

ЭВОЛЮЦИОННЫЙ АЛГОРИТМ ДЛЯ АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ НЕЙРОСЕТЕВЫХ СИСТЕМ ПОДАВЛЕНИЯ ШУМА

Предлагается применять нейронные сети в качестве систем подавления шума в информационных сигналах. Нейронные сети создаются и настраиваются автоматически при помощи эволюционных алгоритмов. Показано, что нейронные сети являются надежным средством для подавления шума.

Ключевые слова: генетический алгоритм, алгоритм генетического программирования, нейронные сети, системы подавления шума.

I. S. Ryzhikov

ABOUT MULTIAGENT SYSTEM APPLICATIONS FOR SPEECH RECOGNITION PROBLEM*

In this paper we suggest two different multi agent systems for speech recognition problem. The multi agent systems (MAS) are becoming very popular because of their flexibility and applicability to complex problems. The system is based on functioning of different agents that forms the system and interacts with each other. The main profit of using multi agent approach is that every agent can be described as a simple subsystem and the whole initial task can be solved with automatic and autonomous agent actions, interactions and decision making. So the main problem can be reduced to behavior rule base tuning.

Keywords: multi agent systems, speech recognition, intelligent agents.

Due to the increasing tasks complexity nowadays it is a common task to choose and modify the one from the variety of classification, modeling and control cybernetic techniques. Since the problems are related to new applied fields with uncommon properties the modifications of methods for every distinct task or even seeking a new way to solve the problem become the main problem for the researcher. The speech recognition itself touches upon classification, optimization, modeling problems and many others; it means that success can be achieved via using the complex recognition system that deals with all the properties of every task and the main problem. There are some different ways to define the problem of speech recognition, different paradigms and theories already exist. Since the speech recognition problem complexity and dependence of the current language it was designed for there are still no any complete solutions for the general problem. Though there are plenty of different techniques to solve every occurred task that is related with speech recognition problem, the speech recognition system, actually, is not able to achieve the desired goals. The complex system consists of different parts and every on them requires a great amount of calculation resources to solve its own task with given accuracy. If the accuracy of the current step is not achieved, the error is going to increase with every following step and the output becomes far from the one it should be. The dependence of processing quality for every element in the system of the previous one's output requires a lot of resources for every current task and the special modifications for every distinct technique for appearing task and every distinct properties of the problem. But also there is another way to solve the complex problem: to create the interaction between the different elements with different goals and make their real time communication be possible. The system that is based on interaction of different components with different goals in some cases can be the multi agent system.

That is why the MAS can satisfy the needs of the complex system since its agents can be intelligent, they can communicate and their goals are to find the solution of recognition task. The benefit of MAS usage consists of three aspects. Firstly, the interacting intelligent agents can be simple, rather simpler than the task they are to deal with. Only because of the interaction, the group of simple agents can automatically solve complex tasks. Secondly, if within the time new better techniques appear or problem definition changes we do not need to rebuild the system, we just need to change the related agent or build new agents for the new goals. Thirdly, no matter what approach of recognition we use, there would always be the task similar to every approach: noise suppressing, wave representation or modeling, classification and etc. In this paper we suggest to use MAS as a decision support system.

The using of MAS in speech recognition problem was described in general at [1] and [2], and for some specific tasks at [3].

```
┌─────────────────────────┐      ┌─────────────────────────────┐
│  SENTENCE IN QUANTIZED  │─────▶│  NOISE REDUCTION, SMOOTHING │
└─────────────────────────┘      └─────────────────────────────┘
                                               │
                                               ▼
                                 ┌─────────────────────────────┐
                                 │    ADAPTATION TO SPEAKER     │
                                 └─────────────────────────────┘
                                               │
                                               ▼
                                 ┌─────────────────────────────┐
                                 │       SPLIT TECHNIQUE        │
                                 └─────────────────────────────┘
                                         │         ▲
                                         ▼         │
┌─────────────────────────┐      ┌─────────────────────────────┐
│    ETALONS DATABASE     │─────▶│       CLASSIFICATION        │
└─────────────────────────┘◀─────└─────────────────────────────┘
                                         │         ▲
                                         ▼         │
                                 ┌─────────────────────────────┐
                                 │         AGGREGATION         │
                                 └─────────────────────────────┘
```
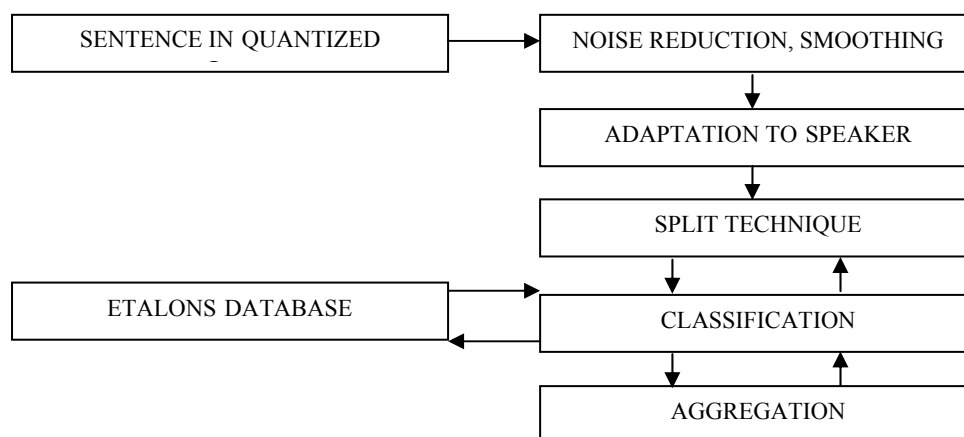
Fig. 1. The scheme of speech recognition

The main structure of the system itself depends on the researcher and problem definition, the basics of multi agent system theory is considered in [4; 5]. In [4] different types of intelligent agents are defined, due to given classification we would use only "model-based reflex agents" and "learning agents" in observing the probable structure of the speech recognition system.

Let the y be the quantized input from the entry device in a vector form. Without loss of generality let y be the digital form of a pronounced sentence. The first thing to do with input signal is to reduce or suppress the possible noise and have the new vector y' that is assumed to be the smoothed input. Then it is necessary to split the vector on different parts to make possible the recognition of every part. In case of give basis the parts should be words, syllables and vowels or consonants. It is clear, that noiseless signal of common well pronounced speech can be easily split into words. After that, we can classify the every part using the database of the examples. Example database consists of etalons: words, syllables and vowels or consonants and is set by the researcher. Then we will know that every part belongs to some class and can make aggregation in the form of meanings or speech and check if it would have any sense in the current situation.

The presented scheme fits only the ideal conditions: perfect speaker and environment, so the data is not distorted by any noise; and the case of superior database, which includes any of the possible examples with different speakers. It is clear, that the recognition task success depends on speaker. That means that if there is no database that would fit every speaker, there should be a process that adapts the system for every new speaker. The structure of this approach is given in Figure 1.

Let us describe the agents of MAS:

– noise suppressor. The noise suppressor agent's goal is to reduce the level of noise. This agent receives the information y in a vector form. Here we can use the model that is built in form of Fourier series, neural network smoothing, filtering or the other smoothing techniques, such as median smoothing or nonparametric estimation of the measured input signal.

– classification agent. This agent uses the relation to the etalon database. Classification algorithm can be based on neural network classification, Bayes classification and classification based on dynamic programming algorithm for speech recognition. Also, there are a lot of different classification techniques which can be used also. The main property of that agent is that the classification agent is a learning agent, so if the sentence is classified correctly, the y' parts form new etalons. The every set of new etalon models rebuilds the etalons set and increases the accuracy of classification problem.

– speaker agent. The adaptation to the speaker via the identification of speech characteristics. Its goal is to identify the parameters of speaker or satisfy the classification agent needs. The basically it should be related with identification of speaker's characteristics.

– splitting agent. This agent goal is to split the input sentence in form of y', so the different parts can be classified with the help of classification agent. It means that to estimate how good the identification with the current split was, the agent must receive information about classification results from the classification agent and its recommendations about how to split and follow the received recommendations to achieve the better results of classification itself.

– supervising agent. The special supervisor controls the recognition process in special cases; its goal is to control the agent community when they cannot succeed. Actually, it takes part in the process only when other agents cannot bring the success to the problem. This agent is a learning agent, that accumulates the decisions and points out what instructions and in what cases were successful staying the observer for the cases the system being working correctly.

Every agent follows its own rule base and can communicate with other agents, the only question is how we would make the structure of the interactions inside the system. The rule base and so the agent's behavior can be described with fuzzy output system or, if it is necessary, with neural networks, which can be tuned with evolution-based global optimization algorithm or any other reliable global optimization technique.

The main agent that estimates how good the cooperative work was is the classification agent. Actually, if there is some needs to make an self-adaptation system there is a need to implement some special technique to identify was the system output correct. The main structure is the fol-

lowing: if the input part cannot be classified, it starts a dialogue with the splitting agent or noise suppressor to remake the parts so the system becomes more flexible. The main dialogue is supposed to be between classification agent and splitting agent with recommendations of the first one, that includes its hypothesis of what part is going better or worse. If there is still no success, agents give all the information to the supervisor, which is going to decide what to do: add some split points, change the classification agent or filter (if there are some different ones), rebuild the speaker agent output or make a decision to ask the speaker pronounce the sentence once again. Supervisor agent also collects the information about previous problems that appeared for this system and have the probabilities of which agent was the most effective in that situation.

The scheme of MAS for speech recognition problem is shown on Figure 2.

Agents of classification and noise reduction can be multiple, and the system includes the different noise reduction and classification agents can increase the efficiency of MAS, but there would be a special need for supervisor on every level, which would coordinate the agents and learn what agent or a combined solution of them is the most perspective. Also the interaction between the agents with the similar goals can be included in agents' properties, so that these relations must be included in agents' decision making base.

Also, there is a possibility to make other MAS only for split and classification tasks. If every agent is an etalon or a set of etalons: syllable, vowel or consonants and, as far as it is a part of the speech – a pause. Then agents could be in interaction between each other to fit every part

of received sentence. The degree of how the agent fits the current part can be defined with dynamic programming method [6], for example. It means that some agents would appear to check if they are similar to different parts of sentence or not. The only question is how to form the boarders. They can increase or decrease their interval asking another agent to move forwards or backwards, relying on their own estimation of a sector and the estimation of their neigbours and check if there would be any success or any possibilities to reach it. Of course, there should be some kind of supervising system too, that would control the classification process and learn. Or the structure that forms the interaction between the agents must be adaptive and includes the special algorithm to avoid the looping.

So, for example, if the system is built on the vowels/consonants identification problem it will have the following agents:

– Vowel agents. These agents are representation of different vowels and have access to their own database of etalons. The input for the agent is the part of the quantized sentence and its goal is to fit this part.

– Consonant agents. These agents are representation of different consonants and also related to their own databases. Agents have the same goal and input as vowel agents.

Pause agent. This agent is the representation of pause or silence.

It seems to be useful to consider the decision making base for every agent for both of the systems as fuzzy network or fuzzy output system, which can be generated with global optimization techniques, or even simpler "if-else-then"-structured agents behavior base.
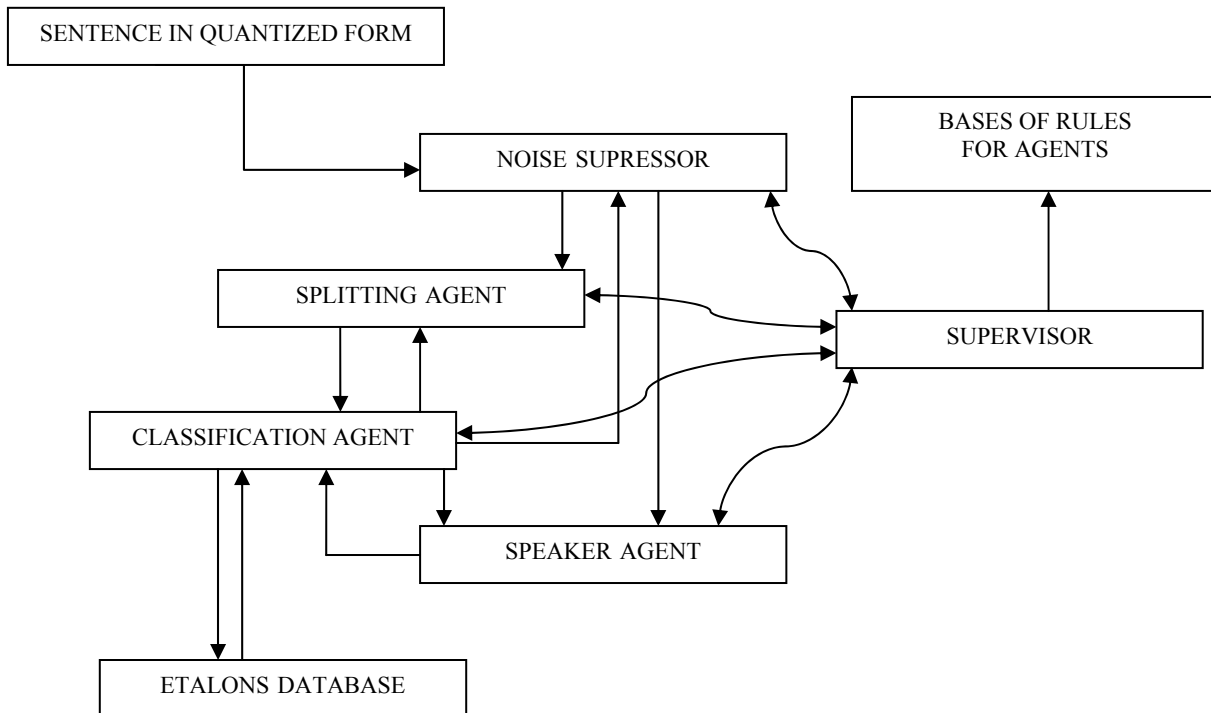

Fig. 2. Multi agent system in speech recognition

In future research the interaction of 3 different agents: splitting agent, speaker agent and classification agent would be tested. The given system can be extended to the system of speech and emotions recognition if we add some agents for emotions recognition. Also, some aggregation agents can be added to provide the relation between classification and complete meanings, if there is a special need that would appear for classification based on syllables and vowels or consonants. If the further investigation shows that dynamic programming method in classification of speech parts is fast enough and reliable then it should be a good alternative to complex classification technique for real-time speech recognition or play a special role in forming the classification algorithm for multi agent classification systems those were presented in the second part of the article.

It is also important to point out that the programming of MAS with its given properties are possible with using the one of the special programming software, which is available to free download and use all over the internet. What is also important to point out is that some platforms for MAS programming works on Java and on different devices, such as, for example, cellar phone.

All in all we should highlight that the speech recognition problem is the complex problem that requires a lot of different algorithms to carry out the different tasks. These task, as it was pointed out earlier are also complex and the accuracy of every one's depends on the quality of the solution for the previous problem.

References

1. Walsh M., Kelly R., O'Hare G. M. P., Carson-Berndsen J., Abu-Amer T. A Multi-Agent Computational Linguistic Approach to Speech Recognition // IJCAI'03 Proceedings of the 18th Intern. joint Conf. on Artificial intelligence. 2003. P. 1477–1479.

2. Walsh M., O'Hare G. M. P., Carson-Berndsen, J. An agent-based framework for speech investigation // INTERSPEECH 2005. 2005. P. 2701–2704.

3. Taha M., Helmy T., Alez R. A. Multi-agent Based Arabic Speech Recognition // Proceedings of the 2007 IEEE/WIC/ACM Intern. Conf. on Web Intelligence and Intern. Conf. on Intelligent Agent Technology Workshops, Silicon Valley, CA, USA. 2007. P. 433–436.

4. Russell S. J., Norvig P. Artificial Intelligence: A Modern Approach. 2nd ed. : Upper Saddle River. New Jersey : Prentice Hall, 2003.

5. Wooldridge M. An Introduction to MultiAgent Systems : John Wiley & Sons. 2002. P. 366.

6. Furtuna T. F. Dynamic Programming Algorithms in Speech Recognition // Informatica Economica. Vol. XII. Issue 2. 2008. P. 94–98.

И. С. Рыжиков

О ПРИМЕНЕНИИ МУЛЬТИАГЕНТНЫХ СИСТЕМ ДЛЯ ЗАДАЧ РАСПОЗНАВАНИЯ РЕЧИ

Предлагаются две различных многоагентная системы для решения задачи распознавания речи. Многоагентные системы (МАС) становятся достаточно популярными благодаря их функциональности и применимости к решению сложных задач. В основе таких систем лежит функционирование каждого ее элемента – агента, и их активное взаимодействие. Основным преимуществом подобного подхода является возможность использовать в качестве агентов простые подсистемы, гораздо более простые, чем решаемая задача. Таким образом, решение задачи сводится к настройке взаимодействий между агентами.

Ключевые слова: многоагентные системы, распознавание речи, интеллектуальные агенты.

I. S. Ryzhikov

AUTOMATIC LINEAR DIFFERENTIAL EQUATION IDENTIFICATION IN ANALYTICAL FORM

In this paper we suggest a reduction of linear dynamics identification problem to the global optimization task. The current approach allows automatic determining the structure and parameters of a linear differential equation via the usage of the modified hybrid evolutionary algorithm for extremum seeking. The a priori information algorithm needed is only the dynamic system initial point or an estimation of the initial point and the sample of measurements: system output and, if there is one, system input.

Keywords: evolutionary strategies, identification, structure and parameters, differential equation.

There are many different approaches of linear differential equation parameters estimation. But since there is no a priori information about the system itself most of them become useless. For different tasks there are some special techniques that allow solving the problem. In this paper we consider the situation when the data can be noised and the system structure is unknown. We can use, for example, stochastic difference equations [1], and build a model using the output observations. But there are some restrictions in using this approach: we still need the