A Survey on the Use of Genetic Algorithms in Natural Language Processing

Pawel Lempa  Michael Ptaszynski  Fumito Masui
Kitami Institute of Technology

ABSTRACT

In this paper we present a survey on application of genetic algorithms in Natural Language Processing (NLP). We focus on main issues of NLP and describe research presented in papers related with this topics.

Introduction

Natural Language Processing (NLP) is an interdisciplinary field connecting artificial intelligence and linguistics. NLP focuses on a number of topics like language generation, machine translation, dialogue systems etc. Researches around the world have proposed different approaches to solve problems within previously mentioned topics. One of them is to use genetic algorithms. A genetic algorithm is an algorithm looking for the best solution using heuristic searching based on natural selection known in genetics. This approach has proven its utility in many cases, sometimes replacing other tools used in artificial intelligence, but also often used as a part of them to resolve a given research problem. Therefore genetic algorithms are often used with e.g neural networks. In the following sections we will describe the application of genetic algorithms in selected papers related with this topics.

Language Generation

Language generation is a process which aims to create natural language as a data representation such as a knowledge base. Language generators are often used to create textual version of natural language. The paper by Monaco and Araki (2007) [8] describes a method that is in contrast with the human-generated natural way. Often such algorithms are used for communication based on text, speech or symbols. Dialogue systems tend to be a part of more general solutions. For example, full functional robot for whom dialogue system is used for communication with external environment. This kind of robots are described in papers by Nakada et al. (2006) [4] and (2010) [5]. In both works genetic algorithms are used for parameters selection in audition subsystems. Operations research, are also frequently used for their most practical function, namely, for optimization. In the context of dialogue systems genetic algorithms can help optimize the settings of time and accuracy of the used methods [3] or cost function [6].

More advanced versions of genetic algorithms are also used. Thanks to large diversity and ease in adoption of GA, they were used in more specified situations. Genetic Algorithm with Sexual selection (GAwSS) was used by Araki and Kuroda (2006) [1]. Thanks to that, a system based on the method using GAwSS, in its initial state can be trained to sufficient level without any prior language information (vocabulary or grammar).

Another example of system which does not need a predefined training data is described Kimura et al. (2001) [2]. In this case conversation rules for the dialog agent are created on base of statements and the system responses.

The last case of GA used in dialogue system is described in [7]. The system proposed in the paper uses a dynamic weight connection of Artificial Intelligence learning algorithm with neural network and genetic algorithm. At first they are created as separate systems. In this paper there were used two versions of genetic algorithms, basic and mutated. The goal was to find the best option for this solution.

CONCLUSIONS

A survey on the use of genetic algorithms in Natural Language Processing was performed. Three main areas of NLP were selected: Dialogue Systems, Language Generation and Machine learning. In each part we presented most representative examples of research applying genetic algorithms to Natural Language Processing tasks. There were described shortly how different kinds of genetic algorithm can be used according to the demand of application. It was presented that a GA can be used to solve a small part of system, that purpose is only to select the best settings for other parts of the system, but also that GA can be in the main focus of whole system. Also was showed that GA has a widely confirmed application for research associated with natural language processing.

REFERENCES
