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

Pawel Lempa †   Michal Ptaszynski †   Fumito Masui †

† Department of Computer Science, Kitami Institute of Technology

lempa@ialab.cs.kitami-it.ac.jp

{ptaszynski,f-masui}@cs.kitami-it.ac.jp

Abstract

In this paper we present our 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.

1 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.

2 Dialogue Systems

Dialogue systems are computer systems which purpose is to converse with the human in most natural way. Often such systems 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 Nakadai et al. (2008) [4] and (2010) [5]. In both works genetic algorithms are used for parameters selection in audition subsystems.

Genetic algorithms, 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 modules. In this case there were used two versions of genetic algorithms, basic and mutated. The goal was to find the best option for this solution.

3 Language Generation

Language generation is a process which aims to create natural language from a data representation such as a knowledge base. Language generators are often used to create textual version of natural language.

The paper by Montero and Araki (2007) [8] describes a method which is based on a small number of phrases chosen randomly from a databese of phrases. From these phrases genetic algorithm generates and evaluates trivial dialogue phrases.

Language generation can also be used to create very sophisticated texts like poetry. Examples of this kind of systems are described in [9] and in [11], where a genetic algorithm is used to find a solution that satisfies the constraints of grammaticality, meaningfulness, and poeticalness.

Good example of the use genetic algorithm is presented in [10]. This paper describes a system of opportunistic text generation. The system reacts react on user choices and adapts to them. In this case a genetic algorithm is used for finding highly-valued as possible tree of facts.

The last example of a paper describing genetic algorithms applied to NLP, [12] describes the use of a genetic algorithm in story generator. In this case GA is utilized for story planer to find space of possible stories. Main reason for using it here is to greatly reduce the risk of getting stuck in local optima.

4 Machine Learning

Machine learning is intended to create algorithms that can learn on the base of initial data. Often this initial data needs to be very large. In this situation a genetic algorithm is a useful solution which can significantly reduce of required information. A good example is presented by Echizenya et al. in [13], [14] and [15]. Thanks to that the number of translation examples for machine translation was greatly reduced. Also important thing is that, the different solutions can use different versions genetic algorithms adjusted to their needs.

The last example could be [17]. In this paper the authors presented a system YALE. It is free open-source environment for KDD and machine learning. It contains a large number of useful features, but for this survey the most important part is Text mining and tracking drifting concepts. Again as presented in section on Dialogue Systems, genetic algorithms are used for feature selection.

5 Conclusions

In this paper we performed a survey on the use of genetic algorithms in Natural Language Processing. We selected three main areas of NLP, namely, 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. We described shortly how different kinds of genetic algorithm can be used according to the demand of application. We presented that a GA can be used as only 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. We also showed that GA has a widely confirmed appliance for research associated with natural language processing.

References

[1] Kenji Araki, Michitomo Kuroda. 2006. Generality of spoken dialogue system using SeGA-IL for different languages. *Proceedings of the IASTED*, San Franisco, 2006.

[2] Yasutomo Kimura, Kenji Araki, Yoshio Momouchi, and Koji Tochinai. 2001. Spoken dialogue processing method using inductive learning with genetic algorithm. *Systems and Computers in Japan*, Vol. 35, No. 12, 2004.

[3] Donn Morrison, Ruili Wang, Liyanage C. De Silva. 2007. Ensemble methods for spoken emotion recognition in call-centres. *Speech Communication*, Elseviere, vol. 49, pp. 98-112, 2007.

[4] Kazuhiro Nakadai, Shunichi Yamamoto, Hiroshi G. Okuno,Hirofumi Nakajima, Yuji Hasegawa, Hiroshi Tsujino. 2008. A Robot Referee for Rock-Paper-Scissors Sound Games. *EEE International Conference on Robotics and Automation*, Pasadena, pp. 3469-3474, 2008.

[5] Kazuhiro Nakadai, Toru Takahashi, Hiroshi G. Okuno, Hirofumi Nakajima, Yuji Hasegawa, Hiroshi Tsujino. 2010. Design and Implementation of Robot Audition System HARK Open Source Software for Listening to Three Simultaneous Speakers. *Advanced Robotics*, vol. 24, Issue 5-6, pp. 739761, 2010.

[6] V. Fischer, J. Fischer, H. Niemann. 1996. An Algorithm for Any-Time Speech Understanding. *Proceedings of the 3rd German-Slovenian Workshop on Speech and Image Analysis*, Ljubljana, 1996.

[7] Noraini Seman, Zainab Abu Bakar, and Nursuriati Jamil. 2013. Improving Speech Recognizer Using Neuro-genetic Weights Connection Strategy for Spoken Query Information Retrieval. *Information Retrieval Technology*, Lecture Notes in Computer Science, vol. 8281, pp. 528-539, 2013.

[8] Calkin S. Montereo, Kenji Araki. 2007. Unsupervised language independent genetic algorithm approach to trivial dialogue phrase generation and evaluation. *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, Vol.4592, pp.388-394, 2007.

[9] Ruli Manurung, Graeme Ritchie, Henry Thompson. 2012. Using Genetic Algorithms to Create Meaningful Poetic Text. *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 24, issue 1, pp. 43-64, 2012.

[10] Chris Mellish, Mick O'Donnell, Jon Oberlander, Alistair Knott. 1998. An Architecture for Opportunistic Text Generation. *Proceeding of the 9th International Workshop on Natural Language Generation*,1998.

[11] Hisar Maruli Manurung. 2003. An evolutionary algorithm approach to poetry generation. *Doctoral Thesis*, Institute for Communicating and Collaborative Systems,School of Informatics University of Edinburgh, 2003.

[12] Neil McIntyre, Mirella Lapata. 2010. Plot Induction and Evolutionary Search for Story Generation. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, pp. 1562-1572, 2010.

[13] Hiroshi Echizen-ya, Kenji Araki, Yoshio Momouchi, Koji Tochinai. 1996. Machine Translation Method Using Inductive Learning with Genetic Algorithms. *Proceeding of the 16th COLING*, Vol. 2, pp. 1020-1023, 1996.

[14] Hiroshi Echizen-ya, Kenji Araki, Yoshio Momouchi, Koji Tochinai. 2004. Machine Translation Using Recursive Chain-Link-Type. *Systems and Computers in Japan*, Vol. 35, No. 2, pp. 1-15, 2004.

[15] Hiroshi Echizen-ya, Kenji Araki, Yoshio Momouchi, Koji Tochinai. 2002. Study of practical effectiveness for machine translation using recursive chain-link-type learning. *Proceeding of the 19th COLING*, Vo. 1, pp. 1-7, 2002.

[16] Colin R. Reeves. 1995. A Genetic Algorithm for Flowshop Sequencing. *Computers & Operations Research*, vol. 22, issue 1, pp. 5-13, 1995.

[17] Ingo Mierswa, Michael Wurst, Ralf Klinkenberg, Martin Scholz, Timm Euler. 2006. YALE: Rapid Prototyping for Complex Data Mining Tasks. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, pp. 935-940, 2006.