A Memetic-Based Approach for Web-Based Question Answering

Iman Khodadi
Faculty of Electrical and Computer Engineering of Tarbiat Modares University, Tehran, Iran
E-mail: iman.khodadi@modares.ac.ir

Mohammad Saniee Abadeh
Faculty of Electrical and Computer Engineering of Tarbiat Modares University, Tehran, Iran
E-mail: saniee@modares.ac.ir

Abstract—In this paper we proposed an evolutionary approach for answering open-domain factoid questions, which include searching among sentences that are candidate for the final answer with Memetic Algorithm (MA), and using lexical and syntactic features for calculating fitness of the sentences. Our main purpose is making a search engine with accurate answering ability, or a web-based Question Answering (QA) system. The Text Retrieval Conference (TREC) QA Tracks data are used to develop and evaluate the approach. The answering process begins with retrieving related documents from a search engine. Then, MA searches among all the sentences of these documents and finds the best one. Finally, one or more words will be extracted based on our hand-made patterns. The results of different approaches for local search, mutation, and crossover, and also different values for number of reproduction and retrieved documents are investigated in the empirical study section. The results are promising with sufficient retrieved documents and find the best one. Finally, one or more words will be extracted based on our hand-made patterns. The results show that the Mametic-based approach is more efficient.

Index Terms—Question Answering, Memetic Algorithm, Information Extraction (IE), Natural Language Processing (NLP), Local Search, Evolutionary Computing, Dynamic Mutation Ratio

I. INTRODUCTION

Question Answering systems are advanced form of search engines and can provide accurate answer to a query, instead of a list of links to potentially relevant web pages. So, QA systems have additional step that extract exact answer from one of the retrieved sources. One of the main advantages of these systems is providing an easy interaction with huge set of text sources. These systems can resolve the information need indicated in a query, retrieve the related information, and extract an answer from them in a form with respect to the question [1]. An important usage of these systems is being an answering system. The Text Retrieval Conference (TREC) QA Tracks data are used to develop and evaluate the approach. The answering process begins with retrieving related documents from a search engine. Then, MA searches among all the sentences of these documents and finds the best one. Finally, one or more words will be extracted based on our hand-made patterns. The results show that the Memetic-based approach is more efficient.

Early QA systems were designed in order to enable users to ask questions from structured data, like personnel data [2]. These structured-based systems can answer questions from a specific subject and that is why they are called Restricted-domain QA (RDQA) systems. But there are systems that can answer questions, independent of the domain and they are called Open-domain QA (ODQA) systems. The first ODQA system was MURAX [3] and it used Information Retrieval (IR) with NLP to answer questions. The ODQA approach differs significantly from RDQA, where a natural language query is transformed into a Structured Query Language (SQL). Instead, in ODQA, the answer must be extracted rather than executed [2].

Although off-line QA systems existed before the search engines, but the first web-based QA was developed many years after the appearance of the search engines in 2004, called START [4].

The direction of research in ODQA systems has been mainly handled by TREC. The Text Retrieval Conference arranged a competition for ODQA systems in 1999, called QA Track. The early competitions focused on factoid question, that is, questions requiring a simple fact. But two other question types were added later, named list questions, that is, questions requiring a list of items, and “other” questions, that is, questions requiring a fact about a subject that is not mentioned in the factoid and list questions. In addition to these three types, there are other question types based on TREC classification, including: definition, hypothetical, causal, relationship, procedural, and confirmation [5]. The target of this paper is answering the factoid questions.

Another related issue is the corpus that the answers are extracted from. There are famous collections such as AQUAINT, which is used in TREC QA Track, but our approach was designed for web data, so the Internet texts are used.

The remainder of this paper is organized as follows: in section 2, related works will be mentioned. In section 3, the overall structure of our approach will be represented, and sections 4 and 5 are empirical study and conclusions.
because its local search function has a valuable effect to
A Memetic-Based Approach for Web-Based Question Answering
more questions and answers.
our hand-made patterns can be upgraded with evaluating
and dynamic mutation ratio. And a set of patterns
approach is more efficient than examining the sentences
number of reproduction and number of retrieved
documents are also investigated. But we only performed answering of the factoid questions and answering to other types can be
investigated in future works. The features that were used can also become more sophisticated in future works. And our hand-made patterns can be upgraded with evaluating more questions and answers.

V. CONCLUSION
In this paper we proposed an evolutionary approach for Question answering systems. Memetic algorithm is used because its local search function has a valuable effect to reach a high accuracy. The results showed that our approach is more efficient than examining the sentences one by one, with respect to the accuracy and process time. A set of equations are proposed for sentence fitness, local search, and dynamic mutation ratio. And a set of patterns are made for questions and answers. The best values for number of reproduction and number of retrieved documents are also investigated.

In this work we proposed an evolutionary approach for Question answering systems. Memetic algorithm is used because its local search function has a valuable effect to reach a high accuracy. The results showed that our approach is more efficient than examining the sentences one by one, with respect to the accuracy and process time. A set of equations are proposed for sentence fitness, local search, and dynamic mutation ratio. And a set of patterns are made for questions and answers. The best values for number of reproduction and number of retrieved documents are also investigated. But we only performed answering of the factoid questions and answering to other types can be investigated in future works. The features that were used can also become more sophisticated in future works. And our hand-made patterns can be upgraded with evaluating more questions and answers.

REFERENCES

Authors’ Profiles
Iman Khodadi was born in Tehran, Iran. He is currently a M.Sc. student in Software Engineering at Tarbiat Modares University, Tehran, Iran, in 2014. He received his B.Sc. in Software Engineering from Science and Culture University, Tehran, Iran, in 2012. His research interests are natural language processing, evolutionary computing, and machine learning. He is a Lecturer at Science and Culture University, Tehran, Iran.

Mohammad Saniee Abadeh received his B.S. degree in Computer Engineering from Isfahan University of Technology, Isfahan, Iran, in 2001, the M.S. degree in Artificial Intelligence from Iran University of Science and Technology, Tehran, Iran, in 2003 and his Ph.D. degree in Artificial Intelligence at the Department of Computer Engineering in Sharif University of Technology, Tehran, Iran in February 2008. His research has focused on developing advanced meta-heuristic algorithms for data mining and knowledge discovery purposes. His interests include data mining, bio-inspired computing, computational intelligence, evolutionary algorithms, fuzzy genetic systems and Memetic algorithms. He is currently a faculty member at the Faculty of Electrical and Computer Engineering at Tarbiat Modares University.