AI to SOAP optimization

3.1   Introduction

In previous chapter it was discussed about the previous attempts to optimize web service message. Then it has highlighted the advantages of using context based approach for optimizing SOAP message communication. The most challenging part of this approach is the process of identifying the static / dynamic content of the message. To identify the common properties of a set of strings first of all they should be arranged in some groupings of those texts based on their common properties. After the grouping is done it is possible to identify the common patterns of the group and then classify the common pattern as the static content. This chapter discuss about some previous researches carried out to classify text using Genetic Algorithms and Multi Agent System. It also discusses of some Multi Agent Models that can be used in data clustering.

3.2   Text Classification using GA

Genetic algorithms were formally introduced in the United States in the 1970s by John Holland at University of Michigan. Originally it had been used on the optimization problem of continuing price - performance of computational systems. In particular, genetic algorithms work very well on mixed (continuous and discrete), combinatorial problems. They are less susceptible to getting stuck at local optima than gradient search methods. But genetic algorithms tend to be computationally expensive.

To use a genetic algorithm, a solution to the problem must be represented as a genome chromosome. The genetic algorithm then creates a population of solutions and applies genetic operators such as cloning, mutation and crossover to evolve the solutions in order to find an optimum solution.

A text based document classification method has been presented by S. M. Kamruzzaman and Farhana Haider on their paper [24]. They have considered the

existing supervised learning algorithms that are used to automatically classify text documents. Their argument is that ordinary supervised learning methods need more sample documents to train the classification algorithm but their algorithm does not need such large number of sample documents. The classification method proposed on this paper is based on the association rules of the words in a particular document but not the words on that. Set of features has to be derived from those associations and finally classification is done using Genetic Algorithms. They have used a concept of positive and negative set of words in which when the set of associated words do not match the considered class then it is considered as negative set of associations. Otherwise it is considered as a positive. They have tested their solution using abstracts of 103 research papers on different areas and it has been proven that the proposed solution requires less number of training documents than conventional methods. Even though this research has been conducted on the area of classification of text documents still it supports the idea of classifying text based messages using Genetic Algorithm based solutions.

The research done by Te-Sheng Li has proposed a method of genetic algorithm based on neural network for feature selection [29]. The method proposed on this paper combines genetic algorithms with artificial neural network classifiers. Genetic algorithm has been used to optimize the feature vector by removing both irrelevant and redundant features and find optimal feature set. It has been studied about the linear feature selection methods on this research and Li has stated that the disadvantage of the linear feature selection is that still it does not reduce the observed variables into a smaller number of projections. Even though linear feature selection transform observed data from one form to other still it does not reduce the number of original features and therefore final feature vector also contains redundant features. In the proposed method of Li has used the KNN-GA [28] method as an example from the literature. Li has introduced the GA to that algorithm to optimize the search method for searching optimal solution. It has been proposed as two phase approach during the first phase it trains a set of neural network using the training data. During second phase it uses the Genetic Algorithms for optimization.

3.3   Text classification using Swarm-Intelligence

In order to find benefits of Multi Agent System to text classification, the basic concepts and definitions related to multi-agent systems need to be understood.

The computer science community has produced various definitions for an agent. However all the definitions share a basic set of concepts. Those are the behavior of an agent, its environment, and autonomy. According to Wooldridge definition [16], an agent is a software (or hardware) entity that is situated in some environment and enjoys the properties of autonomy, social ability, reactivity, pro-activeness. The environment is everything external to the agent. An agent can alter the environment by taking some action. Autonomy means the agent operates without the direct intervention of system users and has some kind of agent's action and their internal state. Social ability of an agent is that they interact with other agents on the environment to complete the job. Agents also react to the environment changes and try to preserve the environment.

A multi-agent system is a system comprising two or more intelligent agents. Each separate agent has local goals. Agents in a multi-agent system do communicate with each other using some kind of agent communication language to complete its own tasks. Because of the emergent properties of the Multi Agent System the collaborative output will be a rational solution to a complex problem.

Andre L. Vizine and his colleges have presented a method of automatically grouping of PDF documents [30] based on the ant-clustering algorithm [6]. Two modifications have been introduced to the ordinary ant-algorithm on their proposal. First is to use a metric to evaluate the similarity degree of text data. Second is to modify the cooling schedule with user defined parameters to improve the convergence of the algorithm. They have identified two main areas of swarm intelligence. That is researches done based on social insects and the research work done based on the ability of human societies to process knowledge. On their method of classifying PDF documents first the PDF document will be converted into text document. Then the document has to be transformed into collections of words that will represent an object on the grid. This transformation has been based on the relative frequency of the words on the document. Then the frequencies has to be normalized in order overcome the influence

of number of words on the document. Then the grid has to be constructed and then ant-algorithm with those modifications has to be applied on the grid. It appears that they have not considered about the performance of the algorithm, may be because the classification of PDF documents is not much time critical.

Gorge Rzevski and his colleges have presented a data clustering method using Agents on their paper [9]. They have discussed three basic negotiation models. The club model, the shareholder model and the tax model are those. The club model is built based on the membership fee of the club. The elements of the cluster have to pay a fee while they are member of the cluster. The membership fee is fixed. Once a data element is entered into a cluster it is reluctant to change the membership. The shareholder model is based on the energy level of the cluster and the individual data elements of the cluster. Data elements can increase the energy level by entering the cluster or by leaving the cluster. They can lose energy if they make a wrong clustering decision. Tax model is based on the charging of taxes from the members of the cluster while they are within the cluster. The tax amount can be varying from cluster to cluster. A cluster can reduce the tax amount to attract members into it. If a member cannot afford to pay the taxes then it has to leave the cluster. This paper has mentioned that the user has to make a hypothesis of data patterns before it starts the clustering using current data mining algorithms. They have mentioned that as a disadvantage of current algorithms.

3.4   Summary

In this chapter it has been discussed about genetic algorithms and multi agent negotiation techniques as the technologies to be adapted to SOAP optimization process. From literature it has been proven that it is possible to use methodologies such as Genetic Algorithms and Multi Agent Negotiation as a solution for data clustering problem. There are many advantages like no need of pre-identified hypothesis of data and will work under the dynamic environments by using these methods. In next chapter it has discussed about the proposed approach to optimize the SOAP message communication using multi agent negotiation.