A Novel Association Rules Method Based on Genetic Algorithm and Fuzzy Set Strategy for Web Mining

Chunlai Chai
College of Computer Science and Information Engineering
Zhejiang GongShang University, Hangzhou, China
Email:warrior21cn@hotmail.com

Biwei Li
College of Computer Science and Information Engineering
Zhejiang GongShang University, Hangzhou, China
Email:warrior@hotmail.com

***Abstract*— the use of fuzzy techniques has been considered to be one of the key components of data mining systems because of the affinity with human knowledge representation. A hybridization of fuzzy sets with genetic algorithms is described for Web mining in this paper. It is based on a hybrid technique that combines the strengths of rough set theory and genetic algorithm. The algorithm through the introduction of selection operators, crossover operators and mutation operators, improves the global convergence speed, and can effectively avoid prematurity. The role of fuzzy sets in handling the different types of uncertainties/impreciseness is highlighted. Experimental results indicate that this adaptive method significantly improves the performance in Web mining.**

***Index Terms*—Web mining, fuzzy sets, genetic algorithm, data mining**

I. INTRODUCTION

Web mining can be viewed as the use of data mining techniques to automatically retrieve, extract and evaluate information for knowledge discovery from web documents and services. The web is a vast collection of completely uncontrolled heterogeneous documents. Thus, it is huge, diverse, and dynamic, and raises the issues of scalability, heterogeneity, and dynamism, respectively [1]. Due to these characteristics, we are currently drowning in information, but starving for knowledge; thereby making the web a fertile area of data mining research with the huge amount of information available online.

Web mining can be broadly defined as the discovery and analysis of useful information from the WWW. In web mining data can be collected at the server side, client side, proxy servers, or obtained from an organization's database. Depending on the location of the source, the type of collected data differs. It also has extreme variation both in its content (e.g., text, image, audio, symbolic) and meta information, that might be available. This makes the techniques to be used for a particular task in web mining widely varying. Web mining basically deals with mining large and hyper-linked information base having the aforesaid characteristics. Also, being an interactive medium, human interface is a key component of most web applications. Thus, web mining, though considered to be a particular application of data mining, warrants a separate field of research, mainly because of the aforesaid characteristics of the data and human related issues.

Web mining can be broadly categorized as Web Content Mining, Web Structure Mining and Web Usage Mining .Web Content Mining of multimedia documents, involving text, hypertext, images, audio and video information. This deals with the extraction of concept hierarchies/relations from the Web, and their automatic categorization; Web Structure Mining of inter-document links, provided as a graph of links in a site or between sites; Web Usage Mining of the data generated by the users' interactions with the Web, typically represented as Web server access logs, user profiles, user queries and mouse-clicks. This includes trend analysis, and Web access association/sequential pattern analysis.

Web mining refers to the use of data mining techniques to automatically retrieve, extract and evaluate (generalize/analyze) information for knowledge discovery from Web documents and services. Web data is typically unlabelled, distributed, heterogeneous, semi-structured, time varying, and high dimensional. Almost 90% of the data is useless, and often does not represent any relevant information that the user is looking for. Taking into account the huge amount of data storage and manipulation needed for a simple query, the processing essentially requires adequate tools suitable for extracting only the relevant, sometimes hidden, knowledge as the final result of the problem under consideration.

Early research in the field concentrated on Boolean association rules, which are concerned only with whether

Manuscript received August 1, 2009; revised August 15, 2009; accepted October 15, 2009.
Corresponding author: Chunlai Chai
Email: :warrior21cn@hotmail.com

doi:10.4304/jcp.5.9.1448-1455

an item is present in a transaction or not, without considering its quantity [2].However, quantity is a very useful piece of information. Realizing the importance of quantity, people started to concentrate on quantitative attributes. The main reason for quantitative association rules mining is that numerical attributes typically contain many distinct values. The support for any particular value is likely to be low, while the support for intervals is much higher. Although current quantitative association rules mining algorithms solved some of the problems particular to quantitative attributes, they introduced some other problems. In other words, existing quantitative mining algorithms either ignore or over-emphasize elements near the boundary of an interval. The use of sharp boundary intervals is also not intuitive with respect to human perception as illustrated next.

To proceed toward web intelligence, obviating the need for human intervention, we need to incorporate and embed artificial intelligence into web tools. At present, the principal tools include fuzzy sets, artificial neural networks (ANNs), genetic algorithms (GA), and rough set (RS) theory. Fuzzy sets provide a natural framework for the process in dealing with uncertainty. Fuzzy set Theory that provides the framework for dealing with linguistic terms has been more frequently used in intelligent systems, because of its simplicity and similarity to human reasoning. The theory has been successfully applied to many fields such as engineering, manufacturing, economics, and others. Neural networks (NNs) are widely used for modeling complex functions, and provide learning and generalization capabilities. GA is an efficient search and optimization tool. RS help in granular computation and knowledge discovery. Rough set theory, which was introduced by Pawlak in the early 1980s, is a new mathematical tool that can be employed to handle uncertainty and vagueness. It focuses on the discovery of patterns in inconsistent data and can be used as the basis to perform formal reasoning under uncertainty, machine learning and rule discovery. Compared to other approaches in handling uncertainty, rough set theory has its unique advantages. It does not require any preliminary or additional information about the empirical training data such as probability distributions in statistics. Fuzzy logic has been used for analyzing inference based on functional dependencies (FDs), between variables, in database relations. Fuzzy inference generalizes both imprecise (set-valued) and precise inference [3]. Similarly, fuzzy relational databases generalize their classical and imprecise counterparts by supporting fuzzy information storage and retrieval. Inference analysis is performed using a special abstract model which maintains vital links to classical, imprecise and fuzzy relational database models. These links increase the utility of the inference formalism in practical applications involving "catalytic inference analysis," including knowledge discovery and database security. FDs are an interesting notion from a knowledge discovery standpoint since they allow one to express, in a condensed form, some properties of the real world which are valid on a given database. These properties can then be used in various applications such as reverse engineering or query optimization. Bosc et al. use a data mining algorithm to extract/discover extended FDs, represented by gradual rules composed of linguistic variables.

There is a growing indisputable role of fuzzy set technology in the realm of data mining [4]. Various data browsers have been implemented using fuzzy set theory [5]. Analysis of real-world data in data mining often necessitates simultaneous dealing with different types of variables, viz., categorical/symbolic data and numerical data. Nauck [6] has developed a learning algorithm that creates mixed fuzzy rules involving both categorical and numeric attributes. Pedrycz discusses some constructive and fuzzy set-driven computational vehicles of knowledge discovery, and establishes the relationship between data mining and fuzzy modeling. The role of fuzzy sets is categorized below based on the different functions of data mining that are modeled.

The fuzzy set concept has recently been used more frequently in mining quantitative association rules. Unlike classical set theory where membership is binary, the fuzzy set theory introduced by Zadeh [7] provides an excellent means to model the "fuzzy" boundaries of linguistic terms by introducing gradual membership. Some example linguistic terms include "poor", "young", "rich", "excellent", etc. Based on this and instead of using sharp boundary intervals, some work has recently been done on the use of fuzzy sets in discovering association rules for quantitative attributes e.g., [8,9]. However, in existing approaches fuzzy sets are either supplied by an expert or determined by applying an existing known clustering algorithm. The former is not realistic, in general, because it is extremely hard for an expert to specify fuzzy sets in a dynamic environment. On the other hand, approaches that applied classical clustering algorithms to decide on fuzzy sets have not produced satisfactory results.

To handle this problem, in this paper we present a GA-based method to derive the fuzzy sets from a set of given transactions. Data mining tools such as GA are presently used to recognize patterns, anticipate changes, and learn the buying habits and preferences of electronic commerce customers in Internet-based transactions [10, 11].In a similar fashion to that used by physical retailers who use data-mining technologies in the design of their stores, web teams can use GA to assist them in mining for the most effective web-site design for electronic commerce. Our method, integrates Fuzzy Set Theory and Genetic algorithm in order to obtain association rules that can be expressed in linguistic terms, which are more natural and understandable for human beings; this type of knowledge makes the discovered rules more useful. The method finds the optimum centroid points for a given number of clusters such that the membership functions generated using these points will extract the maximum number of large itemsets. The experiment results support the efficiency and effectiveness of the proposed method. Fuzzy genes are used as intelligent agents for Web mining. This incorporates a hybridization of fuzzy sets

with GA in the soft computing framework. User profiles are built from the user preferences, represented by chromosomes made up of a vector of fuzzy genes. Each chromosome is associated with a fitness corresponding to the system's belief in the hypothesis that the chromosome, as a query, represents the user's information needs. Every gene represents, by a fuzzy set, the number of occurrences that characterizes the documents considered relevant by the user. The fitness of the chromosome is adjusted based on the comparison between the user's evaluation of the retrieved documents and the score computed by the system. GA is used to track the user's preferences and adapt the profile by incorporating her/his relevance feedback, while fuzzy sets handle the imprecision in the user's preferences and evaluation of the retrieved documents.

We put forward of A hybridization of fuzzy sets with genetic algorithms, called FSGA is described for Web mining in this paper. The rest of this paper is organized as follows. Related work is discussed in Section 2. The FSGA process to find fuzzy sets and their membership functions is described in Section 3. Experiments are given in Section 4. Section 5 provides the conclusion and scope of future research in the area of web mining.

II. RELATED WORKS

Recently, some research works have been done on the use of Fuzzy Set Theory in discovering association rules for dealing continuous attributes. Fuzzy sets provide a smooth transition between members and non members of a set. Fuzzy association rules are also easily understandable to humans because of the linguistic terms associated with fuzzy sets. In addition to fuzziness, researchers proposed different approaches to overcome the interval sharp boundary problem. Miller and Yang proposed a distance-based association rules mining process, which improves the semantics of the intervals. Hirota and Pedrycz proposed a context sensitive fuzzy clustering method based on fuzzy C-means to construct rule-based models. Au and Chan proposed the F-APACS method in order to solve the qualitative knowledge discovery problem .In [12] illustrated fuzzy versions of confidence and support . Gyenesei presented two different methods for mining fuzzy continuous association rules, namely without normalization and with normalization. The experiments of Gyenesei showed that the numbers of large itemsets and interesting rules found by the fuzzy method are larger than the discrete method defined by Srikant and Agrawal.

In [13] introduced fuzzy linguistic summaries on different attributes. Hirota and Pedrycz [14, 15] proposed a context sensitive fuzzy clustering method based on fuzzy C-means to construct rule-based models. However, the context-sensitive fuzzy C-means method cannot deal with the data consisting ofboth numerical and categorical attributes. To solve the qualitative knowledge discovery problem, in [16] applied fuzzy linguistic terms to relational databases with numerical and categorical attributes. Later, they proposed the F-APACS method to discover fuzzy association rules. They utilized adjacent difference analysis and fuzziness in finding the minimum support and confidence values instead of having them supplied by a user. They determine both positive and negative associations. In [17] proposed an algorithm that integrates fuzzy set concepts and Apriori mining algorithm to find interesting fuzzy association rules from given transactional data. In another paper, Hong et al. proposed definitions for the support and confidence of fuzzy membership grades and designed a data mining approach based on fuzzy sets to find association rules with linguistic terms of human knowledge [18]. In [19] illustrated fuzzy versions of confidence and support that can be used to evaluate each association rule. The authors employed these measures of fuzzy rules for function approximation and pattern classification problems. In [20] presented two different methods for mining fuzzy quantitative association rules, namely without normalization and with normalization. The experiments of Gyenesei showed that the numbers of large itemsets and interesting rules found by the fuzzy method are larger than the discrete method defined by Srikant and Agrawal [21]. The approach developed by Zhang [22] extends the equi-depth partitioning with fuzzy terms. However, it assumes fuzzy terms as predefined. Fu et al. proposed an automated method to find fuzzy sets for the mining of fuzzy association rules. Their method is based on CLARANS clustering algorithm. After obtaining the k medoids for each quantitative attribute, these medoids are used to classify each quantitative attribute into k fuzzy sets. On the other hand, in a previous part of our research, we used less centroids than the other approaches described in the literature, and the membership functions of the determined sets are adjusted accordingly [23]. The fuzzy c-medoids and fuzzy c-trimmed medoids are used to cluster relational data from Web documents and snippets in [24]. The algorithms are applied to a collection of 1042 abstracts from the Cambridge Scientific Abstract Web site, corresponding to 10 topics. A preprocessing stage is used to filter and removes irrelevant words, in order to generate the input feature vector that is computed using an inverted document frequency method. This 500-dimensional feature vector (keywords) is reduced using principal component analysis, resulting in a selection of 10 eigenvector values. The algorithms are also tested on a collection of snippets, corresponding to 200 Web documents collected by a search engine in response to a query "salsa". A fuzzy decision tree that uses a fuzzy inductive learning to acquire relations from examples is presented in [25]. The authors generate a concept relation dictionary and a classification tree from a random set of daily business reports database of text classes concerning retailing. An approach in Ref. [26] uses fuzzy association thesaurus and query expansion for information retrieval. Fuzzy composition operations like max – min, max – product and sum – product are used for constructing the thesaurus. Interactive query expansion shows the user, upon initial query, a ranked list of documents suggested by the system based on the fuzzy relation composition. A system that looks for Web documents using link-based

search and a fuzzy concept network is described in [27]. The user's subjective interests are appropriately represented by a fuzzy concept network based on user profile. Missing information is inferred from a transitive closure of a matrix of knowledge in the network. The degree of relevance in the network is fuzzed as a value between 0 and 1. The importance of the document is computed by the fuzzy retrieval system that personalizes the results given by the search engine. It ranks the retrieved documents and finds the authoritative and hub sources. Five best authoritative sources for query are selected as the most representative documents corresponding to a user's query. Documents are ranked according to the user's interests stored in her/his profile (say, ten concepts as "Book", "Java", etc.) The experimental results present the ranked evaluation by three users and the personalized result in response to their query for "Java". Chiang et al. have used fuzzy linguistic summary for mining time series data. The system provides human interaction, in the form of a graphic display tool, to help users premise a database and determine what knowledge could be discovered.

Evolutionary algorithms such us GA have also been used in the field of data mining since they are powerful search techniques in solving difficult problems. M. Kaya proposed a GA-based clustering method to derive a predefined number of membership functions for getting a maximum profit. Hong et. al. proposed a GA-based fuzzy data-mining method for extracting both association rules and membership functions from quantitative transactions. R. Mendez et. al. proposed a co-evolutionary system for discovering fuzzy classification rules. The system uses two evolutionary algorithms: a GP algorithm evolving a population of fuzzy rule sets and a simple evolutionary algorithm evolving a population of fuzzy membership functions definitions. The two populations co-evolve, so that the final result is a fuzzy rule set and a set of membership functions definitions that are well adapted to each other.

III. PROBLEM DEFINITIONS

Let p be a set of web pages, and indicate with $p \in P$ a page in that set. Now assume that P is the result of a standard query to a database of pages, and thus represents a set of pages that satisfy some conditions expressed by the user (for example, that satisfy a Boolean expression on some search keywords). Each page $p \in P$ is associated with a score, based on the query that generated P, that would determine the order by which the pages are presented to the user who submits the query. The role of this ordering is crucial for the quality of the search: in fact, if the dimension of P is relevant, the probability that the user considers a page p strongly decreases as the position of p in the order increases.

This may lead to two major drawbacks: (1) the pages in the first positions may be very similar (or even equal) to each other; (2) pages that do not have a very high score but are representative of some aspect of set P may appear in a very low position in the ordering, with a negligible chance of being looked at by the user. Our method tries to overcome both drawbacks, focusing on the determination of a small set of pages, selected from the initial set p, that, besides containing pages with a high score, also are di0erent from each other and are chosen from die rent regions of some space where the pages are represented. An association rule describes an interesting association relationship among different attributes. A Boolean association involves binary attributes, a generalized association involves attributes that are hierarchically related, and a quantitative association involves attributes that can take on quantitative or categorical values.

Fuzzy sets can be either provided by an expert or automatically derived from the contents of the existing transactions. The definition "X causally increases Y", implies that an increase/decrease of X causally increases/decreases Y. Let $T = \{t_1, t_2, ..., t_i, ..., t_n\}$ be a database of transactions; each transaction t_i represents the ith tuple in T. We use $I = \{i_1, i_2, ..., i_k, ..., i_m\}$ to represent all attributes (items) that appear in T; each attribute i_k may have a binary, categorical or quantitative underlying domain D_{ik}. Besides, each quantitative attribute i_k is associated with at least two fuzzy sets. Explicitly, it is possible to define some fuzzy sets for attribute i_k with a membership function per fuzzy set such that each value of attribute i_k qualifies to be in one or more of the fuzzy sets specified for i_k. The degree of membership of each value of attribute i_k in any of its fuzzy sets is directly based on the evaluation of the membership function of the particular fuzzy set with the value of i_k as input. So, given a database of transactions T, its set of attributes I, and the fuzzy sets associated with quantitative attributes in I. Note that each transaction t_i contains values of some attributes from I and each quantitative attribute in I has two or more corresponding fuzzy sets. The target is to find out some interesting and potentially useful regularities, i.e., fuzzy association rules with enough support and high confidence. We use the following form for fuzzy association rules.

Definition 1. A fuzzy association rule is expressed as

If $X = \{x_1, x_2, ..., x_m\}$ is $A = \{a_1, a_2, ..., a_n\}$ then $Y = \{y_1, y_2, ..., y_n\}$ is $B = \{b_1, b_2, ..., b_m\}$,

In which, X and Y are disjoint sets of attributes called itemsets, i.e., $X \subset I$, $Y \subset I$ and $X \cap Y = \phi$; A and B contain the fuzzy sets associated with corresponding attributes in X and Y, respectively, i.e.. Finally, "X is A" is called the antecedent of the rule while "Y is B" is called the consequent of the rule.

IV. FSGA-BASED FOR WEB MINING

GA is stochastic and evolutionary search techniques based on the principles of biological evolution, natural selection, and genetic recombination. They simulate the principle of 'survival of the fittest' in a population of potential solutions known as chromosomes. Each chromosome represents one possible solution to the problem or a rule in a classification. The population evolves over time through a process of competition whereby the fitness of each chromosome is evaluated using a fitness function. During each generation, a new population of chromosomes is formed in two steps. First, the chromosomes in the current population are selected to reproduce on the basis of their relative fitness. Second, the selected chromosomes are recombined using idealized genetic operators, namely crossover and mutation, to form a new set of chromosomes that are to be evaluated as the new solution of the problem. GA is conceptually simple but computationally powerful. They are used to solve a wide variety of problems, particularly in the areas of optimization and machine learning. GA begins with a population of chromosomes either generated randomly or gleaned from some known domain knowledge. Subsequently, it proceeds to evaluate the fitness of all the chromosomes, select good chromosomes for reproduction, and produce the next generation of chromosomes. More specifically, each chromosome is evaluated according to a given performance criterion or fitness function, and is assigned a fitness score. Using the fitness value attained by each chromosome, good chromosomes are selected to undergo reproduction. Reproduction involves the creation of offspring using two operators, namely crossover and mutation. By randomly selecting a common crossover site on two parent chromosomes, two new chromosomes are produced. During the process of reproduction, mutation may take place. The process of GA evolution goes on towards the direction of maximizing the value of the fitness function. It includes selection of the fittest, crossover and mutation. By the selection operator, solutions with higher fitness values are selected with a higher probability. Crossover means exchanging substrings from pairs of chromosomes to form new pairs of chromosomes. The single point crossover, which separates chromosomes into two substrings, and the double point crossover, which separates them into three substrings, is the most popular crossover methods. Mutation involves generating mutations of the chromosomes. Mutation prevents the search process from falling into local maxima, but a mutation rate that is too high may cause great fluctuation, so the mutation rate is generally set at a low value.

We cluster the values of quantitative attributes into fuzzy sets with respect to a given fitness evaluation criteria. For this purpose, the GA will be employed to adjust the appropriate centroid values of the clusters. GA is very important for the immunology and attracts much attention from the artificial intelligence researchers. We use real-valued coding, where chromosomes are represented as floating point numbers and their genes are the real parameters.

The fitness function measures the goodness of an individual in a given population. It is one of the key issues to a successful GA, simply because the main task in a GA is to optimize a fitness function. The fitness function accepts a decoded chromosome and produces an objective value as a measure of the performance of the input chromosome. The aim of the FSGA employed in this study is to maximize the number of all the large itemsets extracted by the adjusted membership functions. During each generation, individuals with higher fitness values survive while those with lower fitness values are destroyed. In other words, individuals who are strong according to parent selection policy are candidates to form a new population. Parent selection mimics the survival of the best individuals in the given population.

A number of different selection implementations have been proposed in the literature [28], such as roulette wheel selection, tournament selection, and linear normalization selection. Here linear normalization selection, which has a high selection pressure [28], has been implemented. In linear normalization selection, an individual is ranked according to its fitness, and then it is allowed to generate a number of offspring proportional to its rank position. Using the rank position rather than the actual fitness values avoids problems that occur when fitness values are very close to each other (in which case no individual would be favored) or when an extremely fit individual is present in the population (in such a case it would generate most of the offspring in the next generation). This selection technique pushes the population toward the solution in a reasonably fast manner, avoiding the risk of a single individual dominating the population in the space of one or two generations.

The order crossover operator selects at random a substring in one of the parent tours, and the order of the cities in the selected positions of this parent is imposed on the other parent to produce one child. The other child is generated in an analogous manner for the other parent. After crossover, the mutation operation is performed. We adopt the same mutation operation that is widely used: given a mutation rate, do the mutations on a randomly selected bit to change from 0(1) to 1(0).

The solution process of FSGA as follows:

Step 1: Randomly generate $Size$ populations as initial population $P(k)$, $k = 0$;

Step 2: Encode chromosomes into a string representation;

Step3: Calculate the fitness value of each chromosome in each population

Step4: Execute selection , crossover and mutation for $P(k)$ to generate $P(k)'$;

Step 5: Let $k = k + 1$.if $k < MaxGen$, gather the sets of membership functions, each of which has the highest fitness value in its population; otherwise, go back to step 3.

V. EXPERIMENTS

In this section, we present our simulation results for the comparison of FSGA with GA and fuzzy sets. We set the population size, crossover rate, mutation rate, and stopping condition as the controlling parameters of FSGA search for our experiments. We use 50 organisms in the population and set the crossover at 0.7 and mutation rate at 0.1.As a stopping condition, we use 1000 trials (500 generations).

Fig. 1-3 shows the comparisons of the number of fuzzy association rules of the three mining algorithm. In these experiments, the FSGA scheduling policy yields the

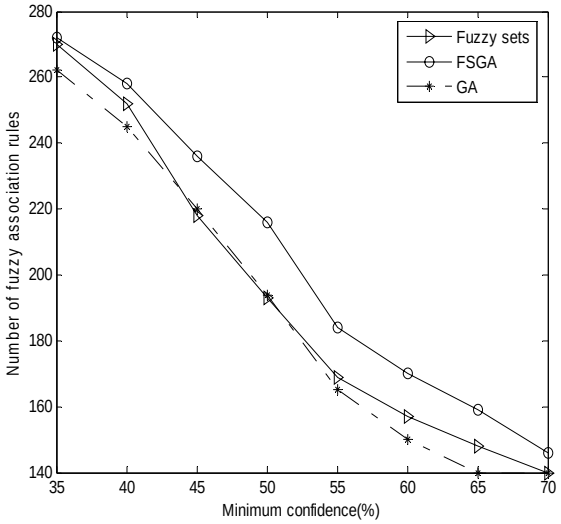

Fig.1. Number of fuzzy association rules for different minimum confidence values

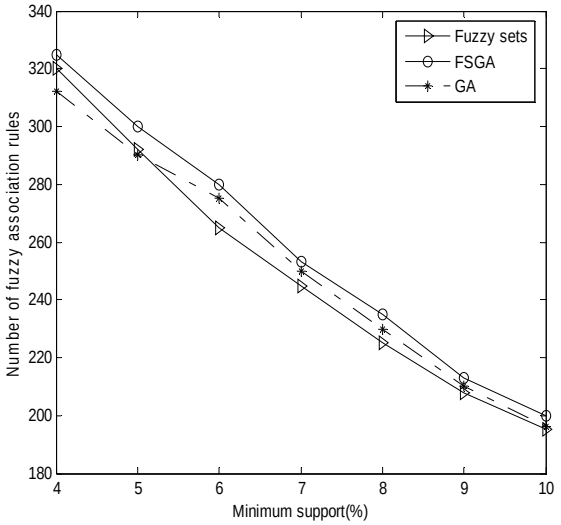

Fig.3. Number of fuzzy association rules for different minimum support values

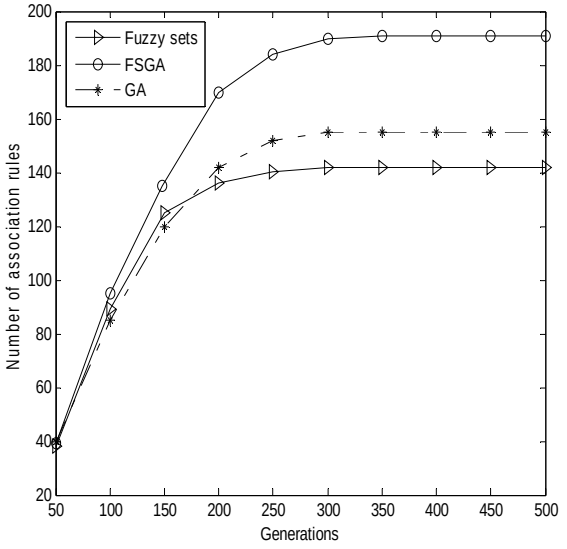

Fig.3. Number of fuzzy association rules for different generations

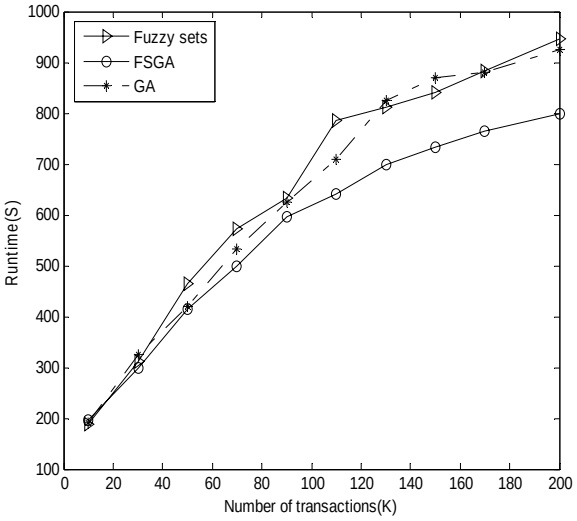

Fig.4. The runtime for three Web mining algorithms

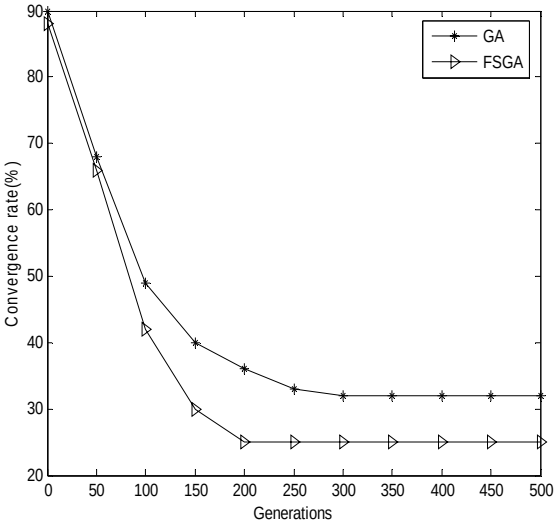

Fig.5. The convergence for GA and FSGA

best results compared to the other examined approaches. Fig. 4 show the runtime required by each of the three algorithms to find the required fuzzy association rules. As can be seen easily from Fig. 4, FSGA outperforms other algorithms, as runtime is concerned. The convergence rate with generations for FSGA and GA is shown in Fig.5. From Fig.5, it is clear that FSGA convergence at a much faster rate than GA. It proves that FSGA is able to obtain the optimal one than GA for Web mining.

VI. CONCLUSIONS AND FUTURE WORKS

The web is a vast collection of completely uncontrolled heterogeneous documents. Fuzzy sets are suitable for handling the issues related to understandability of patterns, incomplete/noisy data, mixed media information and human interaction, and can provide approximate solutions faster. Genetic algorithms provide efficient search algorithms to select a model, from mixed media data, based on some preference criterion/objective function. To prevent the user from being overwhelmed by a large number of uninteresting patterns. A hybridization of fuzzy sets with genetic algorithms is described for Web mining in this paper. It is based on a hybrid technique that combines the strengths of rough set theory and genetic algorithm. In future, we will research on merging simulation annealing and particle swarm optimization in data mining fields.

ACKNOWLEDGMENT

This work is supported by Education project of Zhejiang (20070597), the foundation of science and technology department of Zhejiang province (2008C13074,2008C23025). The authors are grateful for theanonymous reviewers who made constructive comments.

REFERENCES

[1] Pawlak, Z.. Rough sets. International Journal of Computer and Information Sciences, 1982，11(5), 341–356.
[2] P. Bosc, O. Pivert, and L. Ughetto, “Database mining for the discovery of extended functional dependencies,” in Proc. NAFIPS 99, New York, June 1999, pp. 580–584.
[3] D. A. Chiang, L. R. Chow, and Y. F. Wang, “Mining time series data by a fuzzy linguistic summary system,” Fuzzy Sets Syst., 2000, pp. 419–432.
[4] R. R. Yager, “Database discovery using fuzzy sets,” Int. J. Intell. Syst., vol. 11, pp. 691–712, 1996.
[5] J. F. Baldwin, “Knowledge from data using fuzzy methods,” Pattern Recognition Lett., vol. 17, pp. 593–600, 1996.
[6] D. Nauck, “Using symbolic data in neuro-fuzzy classification,” in Proc. NAFIPS 99, New York, June 1999, pp. 536–540.
[7] L.A. Zadeh, Fuzzy sets, Inform. and Control ,1965.pp.338–353.
[8] S.Yue, E. Tsang, D.Yeung, D. Shi, Mining fuzzy association rules with weighted items, Proc. IEEE Internat. Conf. Systems Man Cybernet. (2000) 1906–1911.
[9] C.M. Kuok, A.W. Fu, M.H.Wong, Mining fuzzy association rules in databases, SIGMOD Rec. 17 (1) (1998) 41–46.
[10] J. McCarthy, Phenomenal data mining, association for computing machinery, Communications of the ACM ,2000,43 (8): 75–80.
[11] T.K. Sung, N. Chang, G. Lee.Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction, Journal of Management Information Systems1999, 16 (1): 63–86.
[12] W. Pedrycz, Fuzzy sets technology in knowledge discovery, Fuzzy Sets and Systems (1998) 279–290.
[13] R.R.Yager, Fuzzy summaries in database mining, Proc. Conf. Artif. Intell. Appl. (1995) 265–269.
[14] K. Hirota, W. Pedrycz, Linguistic data mining and fuzzy modelling, Proc. IEEE Internat. Conf. Fuzzy Systems, vol. 2, 1996, pp. 1448–1496.
[15] W. Pedrycz, Fuzzy sets technology in knowledge discovery, Fuzzy Sets and Systems, 1998, pp. 279–290.
[16] W.H. Au, K.C.C. Chan, An effective algorithm for discovering fuzzy rules in relational databases, in: Proc. IEEE Internat. Conf. Fuzzy Systems, 1998, pp. 1314–1319.
[17] T.P. Hong, C.S. Kuo, S.C. Chi,A fuzzy data mining algorithm for quantitative values, in: Proc. Internat. Conf. Knowledge-Based Intelligent Information Engineering Systems, 1999, pp. 480–483.
[18] T.P. Hong, C.S. Kuo, S.C. Chi, Mining association rules from quantitative data, Intell. Data Anal. 3 (1999) 363–376.
[19] H. Ishibuchi, T. Nakashima, T.Yamamoto, Fuzzy association rules for handling continuous attributes, in: Proc. IEEE ISIE, 2001, pp. 118–121.
[20] A. Gyenesei, A fuzzy approach for mining quantitative association rules, TUCS Technical Report No: 336, March 2000.
[21] R. Srikant, R. Agrawal, Mining quantitative association rules in large relational tables, in: Proc. ACM SIGMOD Internat. Conf. Management of Data, 1996, pp. 1–12.
[22] W. Zhang, Mining fuzzy quantitative association rules, Proc. IEEE Internat. Conf. Tools Artif. Intell. (1999) 99–102.
[23] M. Kaya, R. Alhajj, F. Polat, A. Arslan, Efficient automated mining of fuzzy association rules, in: Proc. Internat. Conf. Database and Expert Systems with Applications, 2002.
[24] R. Krishnapuram, A. Joshi, L. Yi, A fuzzy relative of the k-medoids algorithm with application to document and snippet clustering, in: Proceedings ofIEEE International Conference on Fuzzy Systems (FUZZ IEEE’99), Korea, August 1999, pp. 3:1281–3:1286.
[25] S. Sakurai, Y. Ichimura, A. Suyama, R. Orihara, Inductive learning ofa knowledge dictionary for a text mining system, in: L. Monostori, J. Vancza, M. Ali (Eds.), Proc. 14th Internat. Conf. Industrial and Engineering Applications ofArti8cial Intelligence and Expert Systems (IEA/AIE 2001), Lecture Notes in Arti8cial Intelligence, vol. 2070,Springer, Berlin, 2001, pp. 247–252.
[26] H.-M. Lee, S.-K. Lin, C.-W. Huang, Interactive query expansion b ased on fuzzy association thesaurus for web information retrieval, in: Proc. the 10th IEEE Internat. Conf. on Fuzzy Systems, 2001, pp. 2:724–2:727.
[27] K.-J. Kim, S.-B. Cho, A personalized web search engine using fuzzy concept network with link structure, in: Proc. Joint Nineth IFSA World Congress and 20th NAFIPS Internat. Conf. 2001, pp. 1:81–1:86.
[28] Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. In: Proc. national academy of sciences,pp 2907–2912

Chunlai Chai received M.S. degrees in Computer Science from Hohai University, China, in 2003. He is currently an teacher in College of Computer Science and Information Engineering, Zhejiang GongShang University. He has published over 10 papers in international journals and conferences in the areas of wireless communications, E-Commerce and Data Mining. His current research interests are in the areas of wireless communications, protocol design and optimization. The research activities have been supported by the Natural Science Foundation of China,Education project of Zhejiang and Natural Science Foundation of Zhejiang province.

.

Biwei Li ,Ph.D candidate. Majored in Economics and Management. Graduated from Zhejiang University with MBA degree. Main research areas include data mining, corporate governance and financial anomalies. Publised several papers on web mining and E-education