GA-CFS APPROACH TO INCREASE THE ACCURACY OF ESTIMATES IN ELECTIONS PARTICIPATION

Seyyed Reza Khaze\textsuperscript{1}, Ramin Jafarzadeh\textsuperscript{2}, Laya Ebrahimi\textsuperscript{3}, Isa Maleki\textsuperscript{4}

\textsuperscript{1,4}Department of Computer Engineering, Dehdasht Branch, Islamic Azad University, Dehdasht, Iran
\textsuperscript{2} Department of Computer Engineering, Khoy Branch, Islamic Azad University, Khoy, Iran
\textsuperscript{3} Department of Computer Engineering, Science and Research Branch, Islamic Azad University, West Azerbaijan, Iran

ABSTRACT

Prediction the main indexes of participation in election and its effective factors are serious challenges for political and social planners. By respect to abnormal nature of offered analyzes by political scientists, data searchers tried to solve the problems of other methods by discovering hidden sciences of data. In this paper, we represent combined method of Genetic Algorithm (GA) and Correlation-based feature selection (CFS) for increasing precision of classifying in methods based on data searching for participation in election which identifies and removes noised Feature of total set of them. Results of our paper indicated that our offered method could increase precision of other methods prediction.

KEYWORDS

Stock Market, Stock index, S&P 500, Data Mining, Regression, Dataset

1. INTRODUCTION

Participation is a voluntarily contribution of a group of people, movement and forms in political and social plans [1] which is from people and groups sentiment for solving their problems and can have main role in national development and for increasing the credit of different governments political and social leadership. every person and group like to live in a system which can answer his mental, political, opinionative and even financial needs and without participating decisions he can't find his position and such system. On the other hand, by participation in political and social life we can find background of stability in political systems. Election is a kind of action which defines and determine amount of political participation in social system [2] and can be the best sample of people's political participation. Therefore participation analyze in election can be true analyze of totality of peoples political participation. By respect to its importance, scientists and analyzers in political and social sciences in worldwide are developing models for determination factors of people's political participation by which prepare programs based on the most important causes of participation or not participation for people [3]. Government also tries to meet the shortages and factors which prevent people in addition to reinforce them to participate based on offered plans. For preventing the probable hurts to government without participation of people and for decreasing the public dissatisfying, before election, its better that people participation be predicted and encouraging and preventing factors to participation be determined which leads to less costs for a political system. Political experiences proved that political leadership which believe in democracy and people who live in such society and are satisfied by their politicians have a people force against internal and external threats, make other societies to respect and
suitable action in international organizations and societies with them. So far analyzers have offered various methods for prediction people participation in election and its effective factors which have probably and uncertain nature and they don’t have high certain coefficient [4]. On the other hand, political and social experts offered many factors for people participation which have wide range and respecting to all of them require time and is impossible. Today, most of scientists and analyzers believe that Data mining can discover hidden rules of data [5]. In this paper we tried to solve problems of probably and uncertain nature by methods of searching data and by offering trust method, discover the most important factors which are effective on people participation in election. This model can make minimum the cost of many factors which are taken account by different experts and help them to increase amount of people participation in social decision by respect to the most important discovered factors from knowledge of data by investment of them. In this paper, at first, we investigated people participation prediction and its effective factors by respect to previous works and then, we tried to increase percentage of recognition and precision of prediction in previous methods by common model of searching data.

2. PREVIOUS WORK

Amin Babazadeh Sangar & et al. [6] prepared classification tree, K-Nearest Neighbor (KNN) and naive Bayes for prediction the amount of people participation in election and discovering effective factors by orange tools based on CRISP which is a total schematic of architecting of system for prediction the participation in president election by KNN algorithm and its results are good. The KNN decision rule has been a ubiquitous classification tool with good scalability. Past experience has shown that the optimal choice of K depends upon the data, making it laborious to tune the parameter for different applications [6]. The performance of a KNN classifier is primarily determined by the choice of K as well as the distance metric applied.

For test and assessment the offered model, they used subjective study of 100 person's opinion who can participate in 11th president election in Islamic republic of Iran in kohgiloye town and prediction of their participation in election. In their paper they introduced factors by domain which are observed in Table (1) as effective parameter for determining the amount of people participation in election.

<table>
<thead>
<tr>
<th>No</th>
<th>Important Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age of people</td>
</tr>
<tr>
<td>2</td>
<td>Degree of people</td>
</tr>
<tr>
<td>3</td>
<td>Jobs of people</td>
</tr>
<tr>
<td>4</td>
<td>Political Orientation of people</td>
</tr>
<tr>
<td>5</td>
<td>Attitudes to Government Services</td>
</tr>
<tr>
<td>6</td>
<td>Opinion of people about Participation type in elections</td>
</tr>
<tr>
<td>7</td>
<td>Opinion of people about general policy in international affairs</td>
</tr>
<tr>
<td>8</td>
<td>Opinion of people about the election official</td>
</tr>
<tr>
<td>9</td>
<td>Opinion of people about Candidates</td>
</tr>
</tbody>
</table>

They indicated that for prediction the participation KNN algorithm in compare to 2 other algorithms predict and classify with more precision. In this paper, we tried to increase amount of precision of used algorithm.
3. THE PROPOSED METHOD

In this paper, we represented a combined method of GA and CFS and by recognition the noisy and unrelated Feature from its set by Amin Babazadeh et al [6] helps to nearest neighbor algorithm, means decision tree and simple Bayes that classify by better percentage. The classifier tree is one of the possible approaches to multistage decision making [7]. In fact, decision making tree is a type of proper tool and function to classify data, do estimation and provide anticipation due to the features and characteristics which data have till now. Naïve Bayes models are popular in machine learning applications, due to their simplicity in allowing each attribute to contribute towards the final decision equally and independently from the other attributes [8]. This simplicity equates to computational efficiency, which makes Naïve Bayes techniques attractive and suitable for many domains. Bayesian theory is a basic statistical method to solve problems of pattern recognition and classification.

GA was at first offered by John Holland based on Darwin natural evolution opinion [9]. GA is a repeated accidentally process which doesn’t grant the convergence. But with some strategies this chance is increased and this offered algorithm is implanted. The condition for repeating stop in this algorithm can be acceptable with some constant and predetermined amounts such as number of generations or reaching to level. GA is an effective optimization method which collected positive feature of accidentally methods so much. But always, it can be offered some changes in GA by respect to feature of the problem [10]. The first step in GA is primary population accidentally or discovered. Each member of population is named a chromosome, and shows an answer for problem. Chromosomes are completed during repeated periods and each period is named a generation. In each period population is changed and created a new generation which is effective for optimum answer. For keeping the best answer of each generation and preventing their destroying we can use elite technics. Chromosomes reform and evolution is done in 2 ways. In the first step, number of chromosomes are accidentally selected and are combined to create new elements. Second step is named jumping in which in each repeating one or more chromosomes are accidentally selected and one of its gene is also accidentally selected and is changed based on special mechanism and thus, new chromosomes are created. In final step primary population is selected from extended population among the best members and is considered as new generation. There are different methods for selection step and we can point to roulette selection. GA is different in combination method, jumping, and selection or application steps [11].

Feature subset selection has extended scope in research and statistic and pattern recognition methods [12]. Feature selection can have role in economic cost and human force decreasing for data collecting and implementation and analyzing the levels [13]. Selection the Feature subset is important in searching data. High number of Feature lead to be difficult classifying methods testing and training [14] and less precision in prediction [15]. In selecting the Feature we try to identify unneeded and unrelated data with preprocessing and remove them from Feature domain [16]. The main problem in recognition machine learning is a set of Feature which is more important for making classify model and prediction [17]. One of the methods for Feature selection is CFS [18]. So far CFS has been used in many natural and artificial sets. Experiments indicated that this method in compare to learning algorithm such as decision tree, nearest neighbor and simple Bayes which were used by Amin Babazadeh et al, has better speed for representing noisy and unrelated Feature. This method proved that in most cases, classify and prediction is equal or better than other methods. CFS assessor is value of one subset of adjectives by respect to ability for individual prediction for each property and by respect to length of increasement degree between them. Unity coefficient is used for unity estimating among adjectives subset and class label and for unity relating among considered feature [19]. CFS is used for the best property of the subset and can be with some search methods such as backward, forward selection, GA, best-
first search, bi-directional search, elimination. In this paper, we considered GA by name search method among effective Feature for participation in election and CFS was considered as assessment mechanism for effective subsets for participation in election or fitness function and its total scheme is observed in figure 1.

Figure 1. The Proposed Model Is Based On a GA-CFS Approach

We set effective parameters in GA corresponded to details in table 2.

Table 2. Characteristics of the GA for Proposed Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>20</td>
</tr>
<tr>
<td>Number of generations</td>
<td>100000</td>
</tr>
<tr>
<td>Probability of crossover</td>
<td>0.6</td>
</tr>
<tr>
<td>Probability of mutation</td>
<td>0.033</td>
</tr>
</tbody>
</table>

For calculating fitness function, we used CFS is observed in equation (1).

$$r_{zC} = \frac{kr_{zi}}{\sqrt{k + k(k - 1)r_{zi}}}$$
\( r_{zc} \) is unity between Feature subset and variables class. Parameter \( K \) is number of Feature in subsets and \( r_{zi} \) is average unity between Feature subsets in variables class and \( r_{ii} \) is average unity of feature subsets. Our methods by combination two algorithms indicated that finally in 19728th generation, crossover and mutation don’t effect on alternation the function and finished search and assessment. At the end, three feature from feature sets which prepared by Amin Babazadeh et al are selected as ineffective and noisy feature and we removed them of data sets [6]. These features are age, job and opinion about candidates.

4. EVALUATION OF THE PROPOSED METHOD

This section includes the results obtained by the methods described in the paper. For the experiments of the paper have been used the database [6]. We classify them by removing noisy feature from property sets in Table 1. Its results are based on offered method in figure 2.

![Accuracy Chart](chart.png)

Figure 2. The proposed approach is compared with other methods

It can be observed that offered method could have better performance in decision tree and simple Bayes and increase precision of classifying but in compare to KNN its 1% weaker. Thus, in average precision, offered method could recognize noisy data and increase prediction precision by removing them of data set. GA combined performs anticipation and classification, high accuracy.

5. CONCLUSIONS

In this paper, GA is combined with Classifier Trees, KNN and Naive Bayes algorithms. We, a combined method of GA and CFS is introduced which classifies by better percentage with noisy and unrelated feature recognition for predicting the participation in election. We used GA as search method among effective feature in participation in election and used CFS as assessment mechanism for effective subsets in election participation and after removing unrelated and noisy feature, classified data sets for other investigators and its results indicated that average amount of prediction for algorithms by offered method has better results. Therefore, Obtained results demonstrate the good accuracy of the proposed GA-based combined other algorithms.
REFERENCES


Authors

Seyyed Reza Khaze is a Lecturer and Member of the Research Committee of the Department of Computer Engineering, Dehdasht Branch, Islamic Azad University, Iran. He is a Member of Editorial Board and Review Board in Several International Journals and National Conferences. His interested research areas are in the Software Cost Estimation, Machine learning, Data Mining, Optimization and Artificial Intelligence.
Ramin Jafarzadeh is Faculty member of the Department of Computer Engineering, Khoy Branch, Islamic Azad University, Iran. Currently He Has a PhD Candidate In Department Of Computer Engineering At Science And Research Branch, Islamic Azad University, Iran. His Interested Research Areas Are In The Cloud Computing And Data Mining Artificial Intelligence.

Laya Ebrahimi is a M.Sc. student in Computer Engineering Department, Science and Research Branch, Islamic Azad University, West Azerbaijan, Iran. His interested research areas are Meta Heuristic Algorithms, Data Mining and Machine learning Techniques.

Isa Maleki is a Lecturer and Member of The Research Committee of The Department of Computer Engineering, Dehdasht Branch, Islamic Azad University, Dehdasht, Iran. He Also Has Research Collaboration with Dehdasht Universities Research Association Ngo. He is a Member of Review Board in Several National Conferences. His Interested Research Areas Are in the Software Cost Estimation, Machine learning, Data Mining, Optimization and Artificial Intelligence.