International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 05 Issue: 12 | Dec -2018 www.irjet.net p-ISSN: 2395-0072

SURVEY OF FEATURE SELECTION BASED ON ANT COLONY

Kritika[1], Ritika Mehra[2]

[1]*Research Scholar, Dept. of Computer Science and Engineering RPIIT Technical Campus Karnal Haryana, India*
[2]*Assistant Professor, Dept. of Computer Science and Engineering RPIIT Technical Campus Karnal Haryana, India*

Abstract *The Feature Selection approaches can generally be divided into three groups: filter, wrapper, and hybrid approaches. The filter approach operates independently of any learning algorithm. Due to its computational efficiency, the filter methods are very popular to high-dimension data. Traditional search and optimization methods such as gradient-based methods are difficult to extend to the multi objective case because their basic design precludes the consideration of multiple solutions. In contrast, population-based methods such as evolutionary algorithms are well-suited for handling such situation*

Key Words ANN, TACO, SVM, FS, INN

I. INTRODUCTION

Feature selection problem deals with selection of an optimum relevant set of features or attributes that are necessary for the recognition process (classification or clustering). It helps reduce the dimensionality of the measurement space. The goal of feature selection is mainly threefold. First, it is practically and computationally difficult to work with all the features if the number of features is too large. Second, many of the given features may be noisy, redundant, and irrelevant to the classification or clustering task at hand. Finally, it is a problem when the number of features becomes much larger than the number of input data points. For such cases, reduction in dimensionality is required to permit meaningful data analysis. Feature selection facilitates the use of easily computable algorithms for efficient classification or clustering.

In general, the feature selection problem (Ω, P) can formally be defined as an optimization problem: determine the feature set F^* for which

$P(F^*) = \min P(F,X)$

$F \in \Omega$

Where Ω is the set of possible feature subsets, F refers to a feature subset, and $P: \Omega \times \psi \rightarrow (R)$ denotes a criterion to measure the quality of a feature subset with respect to its utility in classifying/clustering the set of points $X \in \psi$. The elements of X, which are vectors in d-dimensional space, are projected into the subspace of dimension $dF=|F|=d$ defined by F. P is used to judge the quality of this subspace.

Feature selection can be either supervised or unsupervised. For the supervised case, the actual class labels of the data points are known. In filter approaches for supervised feature selection, features are selected based on their discriminatory power with regard to the target classes. In wrapper approaches for supervised feature selection, the utility of F is usually measured in terms of the performance of a classifier by comparing the class labels predicted by the classifier for feature space F with the actual class labels. For the unsupervised case, actual class labels are not available. Hence, in filter approaches, features are selected based on the distribution of their values across the set of point vectors available. In wrapper-based unsupervised feature selection, the utility of a feature subset F is generally computed in terms of the performance of a clustering algorithm when applied to the input dataset in the feature space F.

1.1 Dimensionality Reduction through Feature Selection

Feature selection is the process of detecting the relevant features and discarding the irrelevant ones. A correct selection of the features can lead to an improvement of the inductive learner, either in terms of learning speed, generalization capacity or simplicity of the induced model. Moreover, there are some other benefits associated with a smaller number of features: a reduced measurement cost and hopefully a better understanding of the domain.

There are several situations that can hinder the process of feature selection, such as the presence of irrelevant and redundant features, noise in the data or interaction between attributes. Feature selection (FS), since it is an important activity in data preprocessing, has been widely studied in the past years by the machine learning researchers. This technique has found success in many different real-world applications like DNA microarray analysis, intrusion detection, text categorization or information retrieval, including image retrieval or music information retrieval.

There exist numerous papers and books proving the benefits of the feature selection process. However, most researchers agree that there is not a so-called "best method" and their efforts are focused on finding a good method for a specific problem setting. Therefore, new feature selection methods are constantly appearing using different strategies:

A. combining several feature selection methods, which could be done by using algorithms from the same approach, such as two filters, or coordinating

algorithms from two different approaches, usually filters and wrappers;

B. combining FS approaches with other techniques, such as feature extraction or tree ensembles;
C. reinterpreting existing algorithms, sometimes to adapt them to specific problems;
D. Creating new methods to deal with still unresolved situations; and
E. Using an ensemble of feature selection techniques to ensure a better behavior.

Bearing in mind the large amount of FS methods available, it is easy to note that carrying out a comparative study is an arduous task. Another problem is to test the effectiveness of these FS methods when real datasets are employed, usually without knowing the relevant features.

In these cases the performance of the FS methods clearly relies on the performance of the learning method used afterward and it can vary notably from one method to another. Moreover, performance can be measured using many different metrics such as computer resources (memory and time), accuracy, ratio of features selected, etc. Besides, datasets may include a great number of challenges: multiple class output, noisy data, huge number of irrelevant features, redundant or repeated features, ratio number of samples/number of features very close to zero and so on. It can be noticed that a comparative study tackling all these considerations could be unapproachable, and therefore, most of the interesting comparative studies are focused on the problem to be solved. The majority of current real datasets (microarray, text retrieval, etc.) also present noisy data; however, no specific FS comparative studies dealing with this complex problem were found in the literature, although some interesting works have been proposed. From a theoretical perspective, in, a survey of feature selection methods was presented, providing some guidelines in selecting feature selection algorithms, paving the way to build an integrated system for intelligent feature selection.

More experimental work on feature selection algorithms for comparative purposes can be found in, some of which were performed over artificially generated data, like the widely used Parity, LED or Monks problems. Several authors choose to use artificial data since the desired output is known, therefore a feature selection algorithm can be evaluated with independence of the classifier used. Although the final goal of a feature selection method is to test its effectiveness over a real dataset, the first step should be on synthetic data. The reason for this is twofold:

I. Controlled experiments can be developed by systematically varying chosen experimental conditions, like adding more irrelevant features or noise in the input. This fact facilitates to draw more useful conclusions and to test the strengths and weaknesses of the existing algorithms.

II. The main advantage of artificial scenarios is the knowledge of the set of optimal features that must be selected; thus, the degree of closeness to any of these solutions can be assessed in a confident way.

Although works studying some of these problems can be found, up to the authors' knowledge a complete study, such as the one described in here, has not been carried out. Besides, a very interesting problem, since it is very probable in very datasets, such as the alteration of the input variables, has not been addressed elsewhere.

II. RELATED WORKS

Lale, Özbakir, Adil Baykasoglu, [1] In this paper, extracting classification rules from data is an important task of data mining and gaining considerable more attention in recent years. In this paper, a new meta-heuristic algorithm which is called as TACO-miner is proposed for rule extraction from artificial neural networks (ANN). The proposed rule extraction algorithm actually works on the trained ANNs in order to discover the hidden knowledge which is available in the form of connection weights within ANN structure. The proposed algorithm is mainly based on a meta-heuristic which is known as touring ant colony optimization (TACO) and consists of two-step hierarchical structure. The proposed algorithm is experimentally evaluated on six binary and n-ary classification benchmark data sets. Results of the comparative study show that TACO-miner is able to discover accurate and concise classification rules.

Sheng Ding[2] In this paper, It proposes a new strategy combining with the SVM(support vector machine) classifier for features selection that retains sufficient information for classification purpose. Our proposed approach uses F-score models to optimize feature space by removing both irrelevant and redundant features. To improve classification accuracy, the parameters optimization of the penalty constant C and the bandwidth of the radial basis function (RBF) kernel Â¿ is an important step in establishing an efficient and high-performance support vector machine (SVM) model. Aiming at optimizing the parameters of SVM, this paper also presents a grid based ant colony optimization (ACO) algorithm to choose parameters C and Â¿ automatically for SVM instead of selecting parameters randomly by human's experience and traditional grid searching algorithm, so that the classification feature numbers can be reduced and the classification performance can be improved simultaneously. Some experimental results confirm the feasibility and efficiency of the approach. Susana M. **Vieira, João MC Sousa [3]** In this paper, the available set of potential features in real-world databases is sometimes very large, and it can be necessary to find a small subset for classification purposes. One of the most important techniques in data pre-processing for classification is feature selection. Less relevant or highly correlated features decrease, in general, the classification accuracy and enlarge the complexity of the classifier. The goal is to find a reduced

set of features that reveals the best classification accuracy for a classifier. Rule-based fuzzy models can be acquired from numerical data, and be used as classifiers. As rule based structures revealed to be a useful qualitative description for classification systems, this work uses fuzzy models as classifiers. This paper proposes an algorithm for feature selection based on two cooperative ant colonies, which minimizes two objectives: the number of features and the classification error. Two pheromone matrices and two different heuristics are used for these objectives. The performance of the method is compared with other features selection methods, achieving equal or better performance. Mahesh, Pal, and

Giles M. Foody [4] In this paper, support vector machines (SVM) are attractive for the classification of remotely sensed data with some claims that the method is insensitive to the dimensionality of the data and, therefore, does not require a dimensionality-reduction analysis in preprocessing. Here, a series of classification analyses with two hyperspectral sensor data sets reveals that the accuracy of a classification by an SVM does vary as a function of the number of features used. Critically, it is shown that the accuracy of a classification may decline significantly (at 0.05 level of statistical significance) with the addition of features, particularly if a small training sample is used. This highlights a dependence of the accuracy of classification by an SVM on the dimensionality of the data and, therefore, the potential value of undertaking a feature-selection analysis prior to classification. Additionally, it is demonstrated that, even when a large training sample is available, feature selection may still be useful. For example, the accuracy derived from the use of a small number of features may be non inferior (at 0.05 level of significance) to that derived from the use of a larger feature set providing potential advantages in relation to issues such as data storage and computational processing costs. Feature selection may, therefore, be a valuable analysis to include in preprocessing operations for classification by an SVM.

Md Monirul Kabir [5] In this paper presents a new feature selection (FS) algorithm based on the wrapper approach using neural networks (NNs). The vital aspect of this algorithm is the automatic determination of NN architectures during the FS process. Our algorithm uses a constructive approach involving correlation information in selecting features and determining NN architectures. We call this algorithm as constructive approach for FS (CAFS). The aim of using correlation information in CAFS is to encourage the search strategy for selecting less correlated (distinct) features if they enhance accuracy of NNs. Such an encouragement will reduce redundancy of information resulting in compact NN architectures. We evaluate the performance of CAFS on eight benchmark classification problems. The experimental results show the essence of CAFS in selecting features with compact NN architectures.

Chao-TonSu, and Hung-Chun Lin [6] In this paper, methods of feature selection have been increasingly emphasized as venues for reducing cost and shortening the length of time required for computation in data mining. This study utilizes electromagnetism-like mechanism as a wrapper approach to feature selection. Birbil and Fang proposed EM in 2003. EM uses the attraction–repulsion mechanism of the electromagnetism theory to ascertain the optimal solution. Although EM has been applied to the topic of optimization in continuous space and a small number of studies on discrete problems, it has not been applied to the subject of feature selection. In this study, EM combined with 1-nearest-neighbor (1NN) was applied for feature selection and classification. This study utilized the total force exerted on a particle and evaluated this force to determine which features are to be selected. The most crucial features were selected according to the proposed method based on the minimum miss-classification rate, which was attained through 1NN. An unknown datum was classified by 1NN based on the chosen reduced model. To estimate the effectiveness of the proposed method, a numerical experiment was conducted using several data sets with diverse sizes, features, separability, and classes. Experimental results indicated that the proposed method outperformed other well-known algorithms in not only balanced classification accuracy but also efficiency of feature selection. Lastly, this study used an actual case concerning gestational diabetes mellitus to demonstrate the workability of the proposed method.

Kuan-Cheng Lin [7], Feature selection, which is a type of optimization problem, is generally achieved by combining an optimization algorithmwith a classifier. Genetic algorithms and particle swarm optimization (PSO) are two commonly used optimal algorithms. Recently, cat swarm optimization (CSO) has been proposed and demonstrated to outperform PSO. However, CSO is limited by long computation times. In this paper, we modify CSO to present an improved algorithm, ICSO. We then apply the ICSO algorithm to select features in a text classification experiment for big data. Results show that the proposed ICSO outperforms traditional CSO. For big data classification, the results show that using term frequency-inverse document frequency (TF-IDF) with ICSO for feature selection is more accurate than using TF-IDF alone.

Majdi Mafarja [8], Searching for the optimal subset of features is known as a challenging problem in feature selection process. To deal with the difficulties involved in this problem, a robust and reliable optimization algorithm is required. In this paper, Grasshopper Optimization Algorithm (GOA) is employed as a search strategy to design a wrapper-based feature selection method. The GOA is a recent population-based metaheuristic that mimics the swarming behaviors of grasshoppers. In this work, an efficient optimizer based on the simultaneous use of the GOA, selection operators, and Evolutionary Population Dynamics

(EPD) is proposed in the form of four different strategies to mitigate the immature convergence and stagnation drawbacks of the conventional GOA. In the first two approaches, one of the top three agents and a randomly generated one are selected to reposition a solution from the worst half of the population. In the third and fourth approaches, to give a chance to the low fitness solutions in reforming the population, Roulette Wheel Selection (RWS) and Tournament Selection (TS) are utilized to select the guiding agent from the first half. The proposed GOA_EPD approaches are employed to tackle various feature selection tasks. The proposed approaches are benchmarked on 22 UCI datasets. The comprehensive results and various comparisons reveal that the EPD has a remarkable impact on the efficacy of the GOA and using the selection mechanism enhanced the capability of the proposed approach to outperform other optimizers and find the best solutions with improved convergence trends. Furthermore, the comparative experiments demonstrate the superiority of the proposed approaches when compared to other similar methods in the literature.

III. MULTI OBJECTIVE OPTIMIZATION

In this section, some basic concepts of MOO are first introduced. Then, an overview of available MOEAs is provided.

3.1 Concepts of Multi objective Optimization

In many real-world situations, there may be several objectives that must be optimized simultaneously in order to solve a certain problem. This is in contrast to the problems tackled by conventional EAs, which involve optimization of just a single criterion. The main difficulty in considering multi objective optimization is that there is no accepted definition of optimum in this case, and therefore it is difficult to compare one solution with another one. In general, these problems admit multiple solutions, each of which is considered acceptable and equivalent when the relative importance of the objectives is unknown. The best solution is subjective and depends on the need of the designer or decision maker.

We are interested in the multi objective optimization problem (MOP), which can be stated as follows

Minimize $\vec{F}(\vec{x}) := [f1(\vec{x}), f_2(\vec{x}), \ldots, fk(\vec{x})$

subject to

$gi(\vec{x}) \leq 0$ i=1,2,... ,m

$h_i(\vec{x}) = 0$ i=1,2,... ,p

where $\vec{x} = [x_1, x_2, \ldots, xn]^T$ is the vector of decision variables, $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, i =1, ..., k are the objective functions and $g_i, h_j : \mathbb{R}^n \rightarrow \mathbb{R}$, i =1, ...,m, j=1, ..., p are the constraintfunctions of the problem

3.2 Multi objective Evolutionary Algorithms

Traditional search and optimization methods such as gradient-based methods are difficult to extend to the multi objective case because their basic design precludes the consideration of multiple solutions. In contrast, population-based methods such as evolutionary algorithms are well-suited for handling such situations. There are different approaches for solving multi objective optimization problems.

MOEAs have evolved over several years, starting from traditional aggregating approaches to the elitist Pareto-based approaches and, more recently, to the indicator-based algorithms. In the aggregating approaches, multiple objective functions are combined into a single scalar value using weights, and the resulting single-objective function is then optimized using conventional evolutionary algorithms. In population based non-Pareto approaches such as the vector evaluated genetic algorithm, a special selection operator is used and a number of subpopulations are generated by applying proportional selection based on each objective function in turn. Among the Pareto-based approaches, multiple objectives GA, niched Pareto GA (NPGA), and no dominated sorting GA (NSGA) are the most representative nonelitist MOEAs. Although these techniques take into account the concept of Pareto optimality in their selection mechanism, they do not incorporate elitism and, therefore, they cannot guarantee that the nondominated solutions obtained during the search are preserved. In the late 1990s, a number of elitist models of Pareto-based multi objective evolutionary algorithms were proposed. The most representative elitist MOEAs include strength Pareto evolutionary algorithm (SPEA) and SPEA2, Pareto archived evolutionary strategy (PAES), Pareto envelope-based selection algorithm (PESA) and PESA-II , and nondominated sorting genetic algorithm-II (NSGA-II). Most of the recent applications of MOEAs for data mining problems have used one of these Pareto-based elitist approaches as their underlying optimization strategy. A more recent trend regarding the design of MOEAs is to adopt a selection mechanism based on some performance measure. For example, the indicator-based evolutionary algorithmis intended to be adapted to the user's preferences by formalizing such preferences in terms of continuous generalizations of the dominance relation. Since then, other indicator-based approaches, such as the S metric selection evolutionary multi objective optimization algorithm (SMS-EMOA) (which is based on the hyper volume) have also been proposed. The main advantage of indicator based MOEAs

such as SMS-EMOA is that they seem to scale better in the presence of many objectives (four or more). However, approaches based on the hyper volume are very computationally expensive. Since we do not review any application of an indicator-based MOEA in data mining, these approaches are not discussed further in this paper, and they are mentioned only for the sake of completeness.

The FS approaches can generally be divided into three groups: filter, wrapper, and hybrid approaches. The filter approach operates independently of any learning algorithm. These methods rank the features by some criteria and omit all features that do not achieve a sufficient score. Due to its computational efficiency, the filter methods are very popular to high-dimension data. Some popular filter methods are F-score criterion, mutual information, information gain and correlation. The wrapper approach involves with the predetermined learning model, selects features on measuring the learning performance of the particular learning model. Although wrappers may produce better results, they are expensive to run and can break down with very large numbers of features. This is due to the use of learning algorithms in the evaluation of feature subsets every time. Filter and wrapper are two complementary approaches, then the hybrid approach attempts to take advantage of the filter and wrapper approaches by exploiting their complementary strengths.

For a large number of features, evaluating all states is computationally non-feasible requiring heuristic search methods. More recently, nature inspired metaheuristic algorithms have been used to select features, namely: particle swarm optimization (PSO), genetic algorithm (GA)-based attribute reduction, gravitational search algorithm (GSA). These methods attempt to achieve better solutions by application of knowledge from previous iterations.

Ant colony optimization (ACO)is another promising approach to solve the combinational optimization problems and has been widely employed in feature selection. It was initially used for solving Traveling Salesman Problem (TSP) and then has been successfully applied to a large number of NP-hard problems such as Quadratic Assignment Problem (QAP), vehicle routing, system fault detecting, scheduling, etc. In recent years, some ACO-based methods for feature selecting are reported. The hybrid of ACO and mutual information has been used for feature selection in the forecaster.

VI. CONCLUSION

As feature selection is one of the important activities in various fields such as computer vision and pattern recognition. Traditional search and optimization methods such as gradient-based methods are difficult to extend to the multi objective case because their basic design precludes the consideration of multiple solutions. In this paper we survey feature selection and various Feature selection related to multi-objective optimization using meta heuristics. In contrast, population-based methods such as evolutionary algorithms are well-suited for handling such situations. Ant colony optimization (ACO) is another promising approach to solve the combinational optimization problems and has been widely employed in feature selection. It was initially used for solving Traveling Salesman Problem (TSP) and then has been successfully applied to a large number of NP-hard problems such as Quadratic Assignment Problem (QAP), vehicle routing, system fault detecting, scheduling, etc. In recent years, some ACO-based methods for feature selecting are reported. In future we will work on the hybrid of ACO and correlation information for feature selection.

REFERENCES

[1] Lale, Özbakir, Adil Baykasoglu, Sinem Kulluk, and Hüseyin Yapici. "**TACO-miner: an ant colony based algorithm for rule extraction from trained neural networks**." Expert Systems with Applications 36, no. 10 (2009): 12295-12305.

[2] Sheng Ding. "**Feature selection based F-score and ACO algorithm in support vector machine."** In Knowledge Acquisition and Modeling, 2009. KAM'09. Second International Symposium on, vol. 1, pp. 19-23. IEEE, 2009.

[3] Susana M. Vieira, João MC Sousa, and Thomas A. Runkler. "**Two cooperative ant colonies for feature selection using fuzzy models."** Expert Systems with Applications 37, no. 4 (2010): 2714-2723.

[4] Mahesh, Pal, and Giles M. Foody. "**Feature selection for classification of hyperspectral data by SVM**." Geoscience and Remote Sensing, IEEE Transactions on 48, no. 5 (2010): 2297-2307.

[5] Md Monirul Kabir, Md Monirul Islam, and Kazuyuki Murase. "**A new wrapper feature selection approach using neural network."** Neurocomputing 73, no. 16 (2010): 3273-3283.

[6] Chao-TonSu, and Hung-Chun Lin. "**Applying electromagnetism-like mechanism for feature selection**." Information Sciences 181, no. 5 (2011): 972-986.

[7] Lin, Kuan-Cheng, et al. "**Feature selection based on an improved cat swarm optimization algorithm for big data classification**." The Journal of Supercomputing 72.8 (2016): 3210-3221.

[8] Mafarja, M., Aljarah, I., Heidari, A. A., Hammouri, A. I., Faris, H., Ala'M, A. Z., & Mirjalili, S. (2018). **Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems. Knowledge-Based Systems**, 145, 25-45.