Annals of the University of Craiova, Mathematics and Computer Science Series
Volume 40(1), 2013, Pages 100–105
ISSN: 1223-6934

A survey on feature ranking by means of evolutionary computation

Ruxandra Stoean and Florin Gorunescu

Abstract. The paper presents a review of current evolutionary algorithms for feature ranking in data mining tasks involving automated learning. This issue is highly important as real-world problems commonly suffer from the curse of dimensionality. By weighting the significance of each attribute from a data set, the less influential indicators can be disposed of before learning actually takes place, making the task become easier and less noisy. Moreover, for several practical domains of application, such as medicine for instance, a ranking of the most indicative attributes for a diagnosis are as vital as the computational learning support for the final decision taking. Evolutionary algorithms are one of the most frequently used heuristics for a diverse range of tasks, due to their simple and flexible nature. Therefore, the current study investigates the numerous recent trends in employing the evolutionary computation field for the subject of feature ranking.

2010 Mathematics Subject Classification. Primary 68T20; Secondary 68T10.
Key words and phrases. Feature ranking, feature selection, artificial learning, evolutionary algorithms, data mining.

1. Introduction

One of the broadly explored research directions in data mining embraces the concern of reducing the high dimensionality of real-world data that are available for an automated learning assignment. The irrelevant, unnecessary, noisy or just less informative features harden the discovery of the inner relationships within the existing information that lead to a reliable computational support for decision making. Feature selection is the more known paradigm for reaching a shorter attribute collection to extract comprehensible knowledge from, while feature ranking is a more direct way to get a fast insight into the relevance of each decision indicator with respect to the output.

Literature entries for feature ranking are numerous and growing. However, the aim of this paper is to review only those involving a natural heuristic search for the important attributes, i.e. by means of evolutionary algorithms. Evolutionary computation is known to be the Swiss army knife of metaheuristics, as its algorithms are able to offer efficient solutions for almost any class of problems, with the additional advantage of an easy and adaptable framework.

The survey thus starts from a definition of feature ranking in section 2.1. It then divides such techniques into two possible categories (learning method reliant or data properties dependent) in section 2.2 and describes the most used criteria to measure attribute weighting of the latter in section 2.3. Several current evolutionary approaches to the task are eventually presented in section 3.

Received January 7, 2013.

2. What is Feature Ranking?

First of all, it is important to draw the distinction between feature ranking and the more familiar term of feature selection.

2.1. Problem Statement.

Feature selection: implies the search for a subset of the problem attributes to be selected as being together optimally decisive for the learning task [13], [1]. The search is thus performed on the basis of a score assigned to the entire potential collection.

Feature ranking: is a special class of feature selection where a weight or score is calculated for each feature according to some criterion. A certain number of attributes from the obtained ranking is eventually taken as the result for the given task.

The advantages and drawbacks of feature ranking can be enumerated in the next scheme.
Therefore, the pros are:

- simplicity in the maneuvering and interpretability of the selected features,
- individualization of the role of each feature in the decision process.

On the other hand, the cons are known to be:

- the relevance of the final feature set as a whole is not measured [21],
- correlations between features are not analyzed [21]; as a consequence, some highly ranking features may bring no additional information and lead to redundancy [11].

In opposition, the disadvantages of feature selection are as follows:

- the selection of different subsets may not be that simple,
- the relevance of each feature is not investigated [21].

There are also several hybrid studies [16], [7], [20] which combine the advantages of both methodologies, while countering their weaknesses. In such an approach, a first method ranks the most informative attributes, while a second one selects an optimal subset from the top ranking features according to the previous step.

2.2. Classification of Feature Ranking Algorithms. Approaches to feature ranking can generally fall into one of the two following categories [8]:

Wrapper methods: use a machine learning approach to compute accuracy based on the considered features and weights,

Filter methods: use a measure on the feature space, independently of a given learning algorithm, to assess the scores of attributes

The pluses of a wrapper technique can be seen as:

- increased dimensionality reduction,
- enhanced prediction accuracy,

while the minuses are:

- results dependency on the learner,
- overtraining,
- high cost.

As for the advantages of a filter approach, these address:

- generalization ability,
- less costs,

while:

- lower reduction of dimensionality,
- lower accuracy

express the disadvantages.

In the next sections, we will address only filter approaches, as being more independent and flexible for a metaheuristic treatment.

2.3. Ranking Criteria. In order to assess a ranking of features by using only data properties, several criteria can be employed. They can be used independently or be incorporated into metaheuristical search, as shown in section 3.

Some effective possible measures (as regarded by literature) can be hence cataloged as those subsequent (in a random order, not that of importance). These are however designed for a classification task, being the most common formulation of real-world data to automatically learn from.

Intraclass distance [21], [18]**:** denotes the distance between samples within a class and must be minimized in order to achieve high data separability.

Interclass distance [21], [18]**:** reversely expresses the distance between samples of distinct classes and must be maximized. A feature ranks higher than another if the obtained ratio between intraclass and interclass distances after its removal is greater (less separability) than that of the latter.

Attribute-class correlation [21], [18]**:** weights the variations of an attribute between every two patterns of the same class (negatively) or distinct classes (positively).

Signal-to-noise ratio [9], [5]**:** computes the difference between the mean values of a feature for each of the two classes and the sum of the corresponding standard deviations; it outputs the absolute value of this difference/sum ratio as the rank of that attribute.

t-test[9]**:** is used in the form that defines the score of an attribute as the ratio of the difference between its mean values for each of the two classes and the standard deviation; the latter takes into account the standard deviation values of the feature for every class and the cardinality of each. The weight of each feature is thus given by its computed absolute score.

Mann-Whitney U test [12]**:** or the equivalent area under curve (AUC) [15] from receiver operating characteristic (ROC) analysis [11]: the weight of a feature is the fraction by which the attribute ranks the positive samples higher than the negative ones (assuming patterns in one class are denoted as positive and those in the other as negative).

In the end, a natural question thus arises: among the multitude of possible weighting means, what criterion is the best? Each of them can be more informative for a given data set or they can be combined into a single measure [18] or multiobjectively [21] to encompass more domain knowledge.

3. Evolutionary Techniques for Feature Ranking

The use of evolutionary algorithms [4] for weighting the importance of features within data mining tasks is a recent and productive direction of research, with many competitive results and insights into the phenomenon. In the following, several interesting existing research ideas of evolutionary filter techniques are thus presented in detail.

3.1. The ROGER approach. The first known entry in the field is represented by the ROGER (ROC-based Genetic LearneR) algorithm [17]. The technique employs evolution strategies to optimize the AUC criterion in ROC analysis [15].

A ROC curve symbolizes the trade-off between the true positive rate (to be maximized) and the false positive rate (to be minimized) of a classifier. The true positive rate is the fraction between true positive samples (whose outcome is predicted to be positive and the real outcome is also positive) and all positive samples. The false positive rate is the ratio between false positive examples (where the predicted target is positive and the real one is in fact negative) and all the negatives.

The AUC is then a reduced single measure and is defined as the probability that the learner will rank a randomly chosen positive sample higher than a randomly chosen negative one [15].

The ROGER method searches for hypotheses that optimize the AUC by $(\mu+\lambda)$-evolution strategies with self-adaptive mutation [17]. In this scheme [4], μ parents generate λ offspring and the survivors of each generation are the best individuals from the pool of both parents and descendants. Self-adaptive mutation assumes that the mutation step size (one or several) is also encoded into the individuals. It suffers a separate modification through specific mechanisms [4] before being used for the mutation of the main variables.

Each hypothesis refers the weighted 1-norm distance [3] of a training example to a generated point in the sample space [11]. Each individual thus comprises the weights and the point of reference. The fitness function then addresses the AUC criterion [11], [17].

Although ROGER shows promising results, it is however limited to linear functions.

3.2. Multiobjective feature ranking. Evolution of individuals encoding weights is also performed in [21]. The final ranking will be obtained by decreasingly arranging the features according to the obtained weights.

Multiobjective combinations of the possible ranking criteria proposed by [18] and described in section 2.3 are taken into account in the creation of the fitness function, where weights are included in the distance measure:

- maximize intraclass similarity and minimize interclass similarity,
- keep the two objectives above and additionally minimize attribute-class correlation.

Instead of a single objective combination of these measures as discussed by [18], their multiple reference is performed through the evolutionary multiobjective scheme called NSGA-II [2]. In this approach, the Pareto-optimal set of solutions is obtained through what is defined as a fast nondominated sorting approach, with induction of elitism and diversity preservation.

The advantage and drawback of the multiobjective feature ranking technique is that the outcome of the NSGA-II application on the possible weight vectors is a set of nondominated solutions. While the end-user is able to look at the reached Pareto front and choose the ranking for his/her preferred evaluation measure, these diverse found individuals have to be eventually aggregated into a single final result [21].

3.3. Feature weighting with genetic programming. Another recent evolutionary possibility for ranking attributes of a data mining problem is denoted by the genetic programming approach in [14]. Individuals of this type of evolutionary algorithms are in fact programs which evolve also according to some given task [4].

In the cited approach to feature ranking [14], a predefined number of genetic programming runs generate (in parallel) a set of optimal programs (classifiers), each program corresponding to a class of the problem. Every classifier obtains an accuracy for the given data mining task and, based on its fitness, features appearing in its scheme are scored accordingly. This score, together with the frequency with which they appear in well performing classifiers, gives each feature a rank of popularity. In the end of the algorithm, a number of the higher ranking attributes are chosen as the solution.

3.4. Neural networks and evolutionary algorithms to rank features. In [6], ensemble learning based on models of neural networks and evolutionary algorithms, together with specific statistical procedures for attribute ordering, achieved the aggregated ranking of features of a classification task. The sensitivity analysis provided by neural networks gives a rate to each feature in agreement with the accuracy deterioration that happens if that attribute is discarded from the model [6], [19].

The employed neural network types included in the collective system are:

- linear network,
- probabilistic neural network,
- radial basis function neural network,
- three and four-layer perceptrons.

Standard evolutionary algorithms are used to search for the optimal synaptic weights involved within the multi-layer perceptron algorithms.

The feature hierarchies obtained following the classical or hybrid models are finally aggregated into a single ranking [6].

3.5. How to handle several solutions? The last problem that has to be considered is that of establishing the final ranking, in situations when several partial ones appear from the following reasons:

- multiple methods applied to the same data [6],
- one technique applied to several remote subsets of the same data [21],
- a multimodal or multiobjective approach leading to several optimal solutions [21].

The solution would be to assign to each feature a score obtained in either one of the several ways presented below:

- the sum of its ranks in each obtained hierarchy [6],
- the most frequent rank in the partial orderings [10],
- the aggregation of several rankings into a generic one (an ordered partition of the set of features) through a measure of similarity between them [21].

4. Concluding Remarks

Feature ranking is an essential step when performing automated learning on high dimensional real-world data and several evolutionary algorithms have been proven to be efficient for this task. Incorporating only data properties and using a flexible framework, these techniques are able to output the decisive factors and their order in influencing the decision making process. Additionally, the irrelevant features can be further discarded and learning can be conducted only with respect to the key indicators.

Starting now from the surveyed algorithms and taking into account the various possibilities of assessing rankings of features, many novel ideas to further explore this area of research can prospectively spring.

References

[1] M. Breaban, H. Luchian, A Unifying Criterion for Unsupervised Clustering and Feature Selection. *Pattern Recognition* **44** (2011), no. 4, 854–865.

[2] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation* **6** (2002), no. 2, 182–197.
[3] `http://en.wikipedia.org/wiki/Distance` - accessed January 2013.
[4] A.E. Eiben and J.E. Smith, *Introduction to Evolutionary Computing*, Springer, Berlin, Heidelberg, New York, 2003.
[5] T. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer and D. Haussler, S upport Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data, *Bioinformatics* **16** (2000), no. 10, 906–914.
[6] F. Gorunescu, S. Belciug, M. Gorunescu and R. Badea, Intelligent Decision-Making for Liver Fibrosis Stadialization Based on Tandem Feature Selection and Evolutionary-Driven Neural Network, *Expert Systems with Applications* **39** (2012), 12824–12832.
[7] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, Gene Selection for Cancer Classification using Support Vector Machines, *Machine Learning* **46** (2002), no. 1–3, 389–422.
[8] J. Handl and J. Knowles, Feature Subset Selection in Unsupervised Learning via Multiobjective Optimization, *International Journal of Computational Intelligence Research* **2** (2006), no. 3, 217–238.
[9] H. Iba, Y. Hasegawa and T. K. Paul, *Applied Genetic Programming and Machine Learning*, CRC Press, 2010.
[10] K. Jong, E. Marchiori and M. Sebag, Ensemble Learning with Evolutionary Computation: Application to Feature Ranking, *Proceedings of Parallel Problem Solving from Nature*, Birmingham, UK, 2004 (Yao, X. et al., Eds.) Springer LNCS, Berlin, Heidelberg, New York (2004) 1133–1142.
[11] K. Jong , J. Mary , A. Cornuejols, E. Marchiori and M. Sebag, Ensemble Feature Ranking, *Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Pisa, Italy, 2004 (Boulicaut, J.-F. et al., Eds.) Springer, New York (2004), 267–278.
[12] `http://en.wikipedia.org/wiki/$Mann-Whitney_U$` - accessed January 2013.
[13] L.C. Molina, L. Belanche, A. Nebot, Feature Selection Algorithms: a Survey and Experimental Evaluation, *Proceedings of the IEEE International Conference on Data Mining* (2002), 306–313.
[14] K. Neshatian, M. Zhang and Peter Andreae, Genetic Programming for Feature Ranking in Classification Problems, *Proceedings of the Seventh International Conference on Simulated Evolution And Learning*, Melbourne, Australia, December 7 - 10, 2008 (Li, X. et al., Eds.) Springer, Berlin, Heidelberg (2008) 544–554.
[15] `http://en.wikipedia.org/wiki/Receiver_operating_characteristic` - accessed January 2013.
[16] R.Ruiz, J. C. Riquelme and J. S. Aguilar-Ruiz, Heuristic Search over a Ranking for Feature Selection, *Proceedings of the International Work-Conference on Artificial Neural Networks*, Barcelona, Spain, 2005 (Cabestany, J. et al., Eds.), Springer, Berlin, Heidelberg (2005), 742-749.
[17] M. Sebag, J. Aze and N. Lucas, ROC-based Evolutionary Learning: Application to Medical Data Mining. In *P. Liardet et al. (ed) Artificial Evolution*, Springer LNCS, Berlin, Heidelberg, New York (2004), 384–396.
[18] L. Wang and X. Fu, *Data Mining with Computational Intelligence*, Springer, 2005.
[19] D. S. Yeung, I. Cloete, D. Shi and W. W. Y. Ng, Sensitivity analysis for neural networks, Natural Computing Series Springer, Berlin, Heidelber, New York, 2010.
[20] L. Yu and H. Liu, Efficient Feature Selection via Analysis of Relevance and Redundancy, *Journal of Machine Learning Research* **5** (2004), 1205–1224.
[21] D. Zaharie, D. Lungeanu and S. Holban, Feature Ranking Based on Weights Estimated by Multiobjective Optimization, *Proceedings of IADIS First European Conference on Data Mining*, Lisabona, Portugal, July 5-7, 2007 (Roth, J. et al., Eds.) (2007), 124–128.

(Ruxandra Stoean) Department of Computer Science, Faculty of Exact Sciences, University of Craiova, 13 A.I. Cuza Street, Craiova, 200585, Romania
E-mail address: `rstoean@inf.ucv.ro`

(Florin Gorunescu) Department of Fundamental Sciences, Faculty of Pharmacy, University of Medicine and Pharmacy of Craiova, 2 Petru Rares Street, Craiova, 200349, Romania
E-mail address: `fgorun@rdslink.ro`