Differential Evolution (DE) for Multi-Objective Feature Selection in Classification

Bing Xue[1], Wenlong Fu[2], and Mengjie Zhang[1]
[1]School of Engineering and Computer Science,
[2]School of Mathematics, Statistics and Operations Research, Victoria University of Wellington
PO Box 600, Wellington 6140, New Zealand
{Bing.Xue, Mengjie.Zhang}@ecs.vuw.ac.nz, Wenlong.Fu@msor.vuw.ac.nz

ABSTRACT

Feature selection has two main conflicting objectives, which are to minimise the number of features and maximise the classification accuracy. Evolutionary computation techniques are particularly suitable for solving mult-objective tasks. Based on differential evolution (DE), this paper develops a multi-objective feature selection algorithm (DEMOFS). DEMOFS is examined and compared with two traditional feature selection algorithms and a DE based single objective feature selection algorithm. DEFS aims to minimise the classification error rate of the selected features. Experiments on nine benchmark datasets show that in almost all cases, DEMOFS outperforms DEFS and the two traditional feature selection methods in terms of both the number of features and the classification performance.

Categories and Subject Descriptors

I.2 [**Computing Methodologies**]: Artificial Intelligence

Keywords

Differential evolution, Feature selection, Multi-objective optimisation, Classification

1. PROPOSED APPROACHES

1.1 Single Objective Algorithms: DEFS

DEFS is a basic DE based feature selection algorithm, where the objective of DEFS is to select a feature subset with the lowest classification error rate or the highest classification accuracy. Therefore, the classification error rate (i.e. a wrapper measure) is used as the fitness function, which is a minimisation function shown by Equation 1).

$$Fit_1 = ErrorRate = \frac{FP + FN}{TP + TN + FP + FN} \tag{1}$$

where FP, FN, TP and TN stand for false positives, false negatives, true positives, and true negatives, respectively.

In DEFS, each individual is encoded as a vector of real numbers, $x_i = (x_{i1}, x_{i2}, .., x_{id}, .., x_{iD})$, where D is the dimensionality of the dataset. $0 \le x_{id} \le 1$ shows the probability of the dth feature being selected. A threshold θ is used to determine whether this feature is selected. If $\theta \le x_{id}$, the dth feature is selected. Otherwise, the dth feature is not selected.

The DE scheme *DE/rand/1/bin* [4] is used following [3]. The evolutionary feature selection process of DEFS is performed on the training set of a classification problem to select a small number of features. After this training process, the selected feature subset will be used to transform the test set and the classification performance of the selected features will be evaluated on the transformed test set.

Algorithm 1: Pseudo-Code of DEMOFS

```
begin
    Initialise the population with randomly
    individuals; while Maximum Generation is not
    met do
        evaluate the two objective values of each
        individual ;    /* number of features and
        the classification error rate on the
        Training set */
        for i=1 to Population Size (P) do
            create candidate C from parent i;
            evaluate the two objective values of C;
            if C dominates parent i then
                use C to replace parent i;
            else if parent i dominates C then
                C is discarded;
            else if parent i and C are non-dominated
            to each other then
                C is added to the population;
        if Population exceeds the maximum size
        (maxP) then
            truncate the population according to
            non-dominated sorting
        randomly enumerate the individuals in the
        population;
    calculate the testing classification error rate of the
    non-dominated solutions (feature subsets);
    return the non-dominated solutions and their
    training and testing classification error rates.
```

1.2 Multi-objective Algorithm: DEMOFS

DE was originally used for single objective optimisation. To extend DE for multi-objective optimisation, Robič and Bogdand [3] developed a multi-objective DE algorithm named DEMO [3]. DEMO is based on the idea of a popular multi-

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.
ACM 978-1-4503-2881-4/14/07.
http://dx.doi.org/10.1145/2598394.2598493.

objective evolutionary algorithm, i.e. non-dominated sorting based genetic algorithm II (NSGAII) [1]. Experiments on multi-objective benchmark function optimisation showed that DEMO achieved better performance than popular evolutionary multi-objective algorithms [3], but DEMO has never been applied to feature selection problems. To test the use of DE for multi-objective feature selection, we investigate a new feature selection algorithm based on DEMO, where DEMO is used to simultaneously minimise the classification error rate and the number of features. This feature selection algorithm is named DEMOFS. Algorithm 1 shows the pseudo-code of DEMOFS.

2. EXPERIMENTS

The performance of the two DE based feature selection algorithms are examined on nine datasets selected from the UCI machine learning repository [2]. For each dataset, the instances are randomly divided into two sets: 70% as the training set and 30% as the test set. The crossover rate is set as 0.3. The DE based algorithms are stochastic methods and each of them has been performed for 30 independent runs on each dataset. Other settings of the experiments and the results presentation follow the literature [6, 5].

2.1 Results of DEFS

In all cases, DEFS selected around half of the available features and achieved similar or better classification performance than using all features.

The results show that DEFS uses DE as the search technique can be successfully used for feature selection in classification to reduce the dimensionality of the data, but reserve useful information and/or remove redundant/irrelevant information to maintain or improve the classification performance.

2.2 Results of DEMOFS

For all the nine datasets, the "average" set contains at least one feature subset, which included a smaller number of features and achieved at least similar classification performance as using all features. Note that the solutions from the same run are non-dominated to each other, but when combining solutions from all the 30 independent runs, some solutions may dominate others. Therefore, some of the solutions in the "average" set dominate others. In all the nine datasets, the "best" set of solutions selected a significantly smaller number of features and achieved better classification performance than using all features.

The results show that DEMOFS uses a multi-objective search mechanism can effectively explore the solution space to obtain a set of trade-off solutions, which can reduce the number of features and improve the classification performance over using all features.

2.3 Comparisons between DEFS and DEMOFS

The classification performance of the "average" set is usually slightly worse or similar to that of DEFS, but the "best" set is always better than DEFS. This is reasonable because DEFS, and DEMOFS share the same (total) number of evaluations, but DEFS focuses on the optimisation of the classification performance and return only one single solution from each run. DEMOFS returns a set of solutions with trade-off between the classification performance and the number of features. Therefore, when averaging the classification error rates of feature subsets from different runs, it may be slightly worse the DEFS.

3. CONCLUSIONS AND FUTURE WORK

The goal of this paper was to develop a DE based multi-objective feature selection approach to selecting a set of non-dominated feature subsets, which include a small number of features and achieve high classification performance than using all features. The goal has been achieved by using a multi-objective DE algorithm to minimise both the classification error rate and the number of selected features to form the algorithm named DEMOFS. DEMOFS was compared with two traditional feature selection algorithms (LFS and GSBS) and a DE based single objective algorithms (DEFS), which aimed to minimise the classification error rates. Experiments on nine commonly used datasets showed that DEFS, or DEMOFS can select a or a set of small feature subset (s) to achieve similar or even better classification performance than using all features. DEMOFS outperformed both DEFS because it aimed to minimise both the classification error rate and the number of features. All the three DE based algorithms outperformed the two traditional algorithms.

This is the initial and also the first work of using multi-objective DE for feature selection, but it discovers that DE can be successfully used for multi-objective feature selection. It also provides motivations for further investigating EC particularly DE methods for multi-objective feature selection. The performance of DEMOFS needs to compare with other EC based multi-objective feature selection algorithms, such as PSO and GAs.

4. REFERENCES

[1] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182 –197, 2002.

[2] A. Frank and A. Asuncion. UCI machine learning repository, 2010.

[3] T. Robič and B. Filipič. Demo: Differential evolution for multiobjective optimization. In *Evolutionary Multi-Criterion Optimization*, volume 3410 of *Lecture Notes in Computer Science*, pages 520–533. 2005.

[4] R. Storn. On the usage of differential evolution for function optimization. In *1996 Biennial Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, pages 519–523, 1996.

[5] B. Xue, L. Cervante, L. Shang, W. N. Browne, and M. Zhang. A multi-objective particle swarm optimisation for filter based feature selection in classification problems. *Connection Science*, 24(2-3):91–116, 2012.

[6] B. Xue, M. Zhang, and W. Browne. Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics*, 43(6):1656–1671, 2013.