Sparsity-based Evolutionary Multi-objective Feature Selection for Multi-label Classification

Kaan Demir
Victoria University of Wellington
Kaan.Demir@ecs.vuw.ac.nz

Bach Hoai Nguyen
Victoria University of Wellington
Hoai.Bach.Nguyen@ecs.vuw.ac.nz

Bing Xue
Victoria University of Wellington
Bing.Xue@ecs.vuw.ac.nz

Mengjie Zhang
Victoria University of Wellington
Mengjie.Zhang@ecs.vuw.ac.nz

ABSTRACT

Feature selection (FS) is typically an essential pre-processing step for many machine learning tasks. However, most existing FS approaches focus on single-label classification, whereas multi-label classification (MLC) is an emerging topic where each instance can be assigned to more than one class label. Sparsity-based FS is an efficient and effective method that can be applied to MLC. However, existing sparsity-based approaches based on gradient descent can get trapped at local optima, and cannot optimise the selected number of features and classification performance simultaneously that are often in conflict, thus making FS a multi-objective problem. Evolutionary multi-objective optimisation (EMO) can be applied to FS due to its potential global search ability. EMO-based algorithms have not been utilised in sparsity-based approaches for FS in MLC. This paper proposes a novel sparsity-based MLC FS method based on Multi-objective Evolutionary Algorithm with Decomposition (MOEA/D). The experimental results indicate improvement in the MLC performance in comparison to a existing multi-objective FS algorithms.

CCS CONCEPTS

• **Computing methodologies → Feature selection**.

KEYWORDS

Feature Selection, Multi-objective, Multi-label Classification

1 INTRODUCTION

Classification is the task of predicting a class label based on the instance's features. Redundant and irrelevant features are known to negatively impact the classification performance [1], hence feature

selection (FS) is proposed to find small discriminate feature subsets that can improve the classification performance over using all features [1, 5]. Most existing FS methods are focused on single-label classification, where each instance is assigned to a single class label. However, it is increasingly more common for instances of data to possess a set of class labels, which is known as multi-label classification (MLC). Sparsity-based FS [2] can efficiently and effectively be applied to solve FS problems for MLC. However, most existing sparsity-based methods are limited by their tendency to converge at local optima due to gradient-descent. Furthermore, the two main objectives of FS, to maximise the classification performance and to minimise the number of selected features, are often in conflict with one another [5]. Thus, FS is a multi-objective problem.

Evolutionary Multi-objective Optimisation (EMO) can effectively solve FS problems by employing a population of solutions, which can explore the large search space of FS. In particular, Multi-objective evolutionary algorithm based on decomposition (MOEA/D) [8], a state-of-the-art EMO algorithm, is capable of evolving more diverse solutions in comparison to NSGAII and SPEA2 [6].

To the best of our knowledge, sparsity-based EMO has not yet been applied to solve FS for MLC. Therefore, to address the limitations of existing sparsity-based methods, this paper proposes a novel sparsity-based FS algorithm based on MOEA/D. The proposed method is expected to efficiently evolve a diverse set of feature subsets with high classification performance.

2 PROPOSED METHOD

In order to handle multi-label data, a matrix-based representation enabling sparsity-based optimisation is proposed for MOEA/D (MOEA/D-SPAR). A candidate solution is represented by a matrix $\mathbf{Z}^{d \times q}$ as shown in Equation (1), where d is the number of original features and q is the number of class labels. The ith row, $1 \leq i \leq d$, represents the importance of the ith feature to q labels.

$$\mathbf{Z} = \begin{matrix} & l_{.,1} & \cdots & l_{.,q} \\ f_{1,.} \\ \vdots \\ f_{d,.} \end{matrix} \begin{pmatrix} z_{1,1} & \cdots & z_{1,q} \\ \vdots & \ddots & \vdots \\ z_{d,1} & \cdots & z_{d,q} \end{pmatrix} \quad (1)$$

The importance indicates how useful the given feature is in classifying each of the class labels. The selected features of a candidate solution can be obtained by choosing the top ranked feature rows, ordered from highest to lowest of their row norms.

Table 1: Datasets

Dataset	#Instances	#Features	#Labels	Domain	Cardinality	Density
Birds	645	260	19	audio	1.014	0.053
Emotions	593	72	6	music	1.869	0.311
Enron	1702	1001	53	text	3.378	0.064
Genbase	662	1186	27	biology	1.252	0.046
Image	2000	294	5	Image	1.236	0.247
Langlog	1460	1004	75	text	1.180	0.016
Scene	2407	294	6	Image	1.074	0.179
Yeast	2417	103	14	biology	4.237	0.303

In multi-objective sparsity-based FS, the classification performance is measured by a reconstruction error that is unbounded [2]. It has been shown that normalising such unbounded objectives may have negative impact on the performance [4]. Thus, this work aims to propose a novel decomposition method that does not require to normalise the objectives. The proposed mechanism decomposes the problem by using N reference points, where N is the population size. A reference point lies on the fRatio axis, which represents an ideal feature subset (with a classification error of 0) given the corresponding feature ratio. Using this method, each reference point can form a sub-problem which aims to find the best feature subset given a predefined feature ratio, allowing MOEA/D to directly optimise the reconstruction error, hence not requiring objective space normalization. The reconstruction error-based objective functions are given by Equation (2)

$$\underbrace{R_1(\mathbf{Z}) = ||\mathbf{XZ} - \mathbf{Y}||_F,}_{\text{classification loss}} \quad \underbrace{R_2(\mathbf{Z}) = ||\mathbf{XZ} - \mathbf{Y}||_F + ||\mathbf{Z}||_F^2}_{\text{classification loss with regularization}} \quad (2)$$

where $\mathbf{X} \in \mathbb{R}^{n \times d}$ is the training instance matrix, $\mathbf{Y} \in \mathbb{R}^{n \times q}$ is the training label matrix, $\mathbf{Z} \in \mathbb{R}^{d \times q}$ is the solution matrix, n denotes the number of instances, and F denotes the Frobenius norm of the matrix.

Introducing sparsity regularization is expected to reduce the overall magnitude of all values in the weight matrices, thus encouraging the evolution of sparse solutions that can shrink the weights of less important features toward zero, hence improving the overall quality of the population. In this work, each of the two loss functions $R_1(\mathbf{Z})$ and $R_2(\mathbf{Z})$ are used together with the objective of minimising the number of selected features to form two seperated algorithms, called MOEA/D-SPAR-$R_1(\mathbf{Z})$ and MOEA/D-SPAR-$R_2(\mathbf{Z})$, respectively.

Furthermore, due to the sparsity-based representation and differentiable objective functions, a gradient descent-based local search can be used. This can help MOEA/D-SPAR to evolve higher quality solutions to evolve in the final front. After a solution is chosen for crossover or mutation, a copy of it can be enhanced with local search by setting the derivatives of either $R_1(\mathbf{Z})$, or $R_2(\mathbf{Z})$ to zero.

3 EXPERIMENT DESIGN AND RESULTS

The following eight widely used benchmark datasets are chosen. For each of the eight datasets, a 70/30 stratified split is applied. A wrapper based multi-label FS algorithm using MOEA/D (WMOEA/D) is used as a benchmark that is based on previous work [3]. ML-KNN [7] with Hamming-Loss is utilised as the wrapped classification

Table 2: *Test* Set Friedman/Nemenyi Multitest

	Metric	WMOEAD	MOEA/D-SPAR-$R_1(\mathbf{Z})$	MOEA/D-SPAR-$R_2(\mathbf{Z})$
W/D/L	HV	0/0/16	9/5/2	10/5/1
W/D/L	IGD	1/3/12	7/5/4	10/4/2

algorithm. The population size N is set to 200 [6]. Based on initial experiments, the maximum number of iterations is set to 800. In total, each algorithm on each dataset is run over 30 independant trials.

Table 2 shows a Friedman/Nemenyi significance test, with $\alpha = 0.05$, of the hypervolume (HV) and inverted generational distance (IGD) metrics of fronts plotted against their Hamming-Loss and *fRatio* values. An aggregate row is presented which sums the wins (W), draws (D) and losses (L) of each proposed method in comparison to all other methods. Both of the proposed methods $R_1(\mathbf{Z})$ and $R_2(\mathbf{Z})$ outperform WMOEA/D. In particular, $R_2(\mathbf{Z})$ mostly outperforms all other methods on both HV and IGD metrics. WMOEA/D is outperformed by the proposed methods on almost all datasets, indicating that that sparsity-based EMO is capable of solving multi-label FS problems much better than conventional EC-based FS.

4 CONCLUSIONS

The main contributions of this paper included a new matrix representation that allows MOEA/D to perform sparsity-based FS - MOEA/D-SPAR, a new decomposition method that can control the number of selected features with a reference point-based strategy, two new multi-label sparsity-based objective functions for MOEA/D-SPAR: $R_1(\mathbf{Z})$, and $R_2(\mathbf{Z})$, and a new efficient local search strategy to help MOEA/D-SPAR generate high quality solutions. The experimental results show a significant improvement in performance in comparison with a wrapper-based multi-objective FS algorithm WMOEA/D.

REFERENCES

[1] M. Dash and H. Liu. 1997. Feature selection for classification. *Intelligent Data Analysis* 1, 1 (1997), 131 – 156. https://doi.org/10.1016/S1088-467X(97)00008-5
[2] Ling Jian, Jundong Li, Kai Shu, and Huan Liu. 2016. Multi-Label Informed Feature Selection. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 1627–1633.
[3] S. Kashef and H. Nezamabadi-pour. 2017. An effective method of multi-label feature selection employing evolutionary algorithms. In *2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*. 21–25.
[4] Y. Liu, H. Ishibuchi, G. G. Yen, Y. Nojima, N. Masuyama, and Y. Han. 2020. On the Normalization in Evolutionary Multi-Modal Multi-Objective Optimization. In *2020 IEEE Congress on Evolutionary Computation (CEC)*. 1–8. https://doi.org/10.1109/CEC48606.2020.9185899
[5] Bach Hoai Nguyen, Bing Xue, and Mengjie Zhang. 2020. A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation* 54 (2020), 100663.
[6] Hoai Bach Nguyen, Bing Xue, Hisao Ishibuchi, Peter Andreae, and Mengjie Zhang. 2017. Multiple Reference Points MOEA/D for Feature Selection. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (Berlin, Germany) *(GECCO '17)*. Association for Computing Machinery, New York, NY, USA, 157–158.
[7] Min-Ling Zhang and Zhi-Hua Zhou. 2007. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition* 40, 7 (2007), 2038 – 2048. https://doi.org/10.1016/j.patcog.2006.12.019
[8] Q. Zhang and H. Li. 2007. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation* 11, 6 (Dec 2007), 712–731.