Exploration of Two-Objective Scenarios on Supervised Evolutionary Feature Selection: a Survey and a Case Study (Application to Music Categorisation)

technische universität dortmund

Igor Vatolkin[1]

[1]Chair of Algorithm Engineering, Department of Computer Science, TU Dortmund, Germany

igor.vatolkin@tu-dortmund.de

Overview

- Almost all studies on feature selection for supervised classification are limited to single-objective optimisation.
- Typically classification performance measures are optimised (accuracy, classification error, precision, recall).
- Literature survey: past and recent studies on evolutionary multi-objective feature selection with the focus on the combinations of objectives (see the paper).
- Case study: exploration of 28 pairs of objectives for supervised music classification.
- Measurement of suitability for multi-objective optimisation with the help of two hypervolume-based statistics.

1. Introduction and Background

1.1 Definition of Feature Selection Problem

Given

- $\boldsymbol{q}$: binary vector to indicate selected features,
- $\boldsymbol{q}^*$: optimal index vector,
- $\mathcal{F}$: set of all features,
- $\Phi(\mathcal{F},\boldsymbol{q})$: set of features indicated in $\boldsymbol{q}$,
- $\boldsymbol{y}$: true labels,
- $\hat{\boldsymbol{y}}$: predicted labels,
- m: relevance measure (objective function to optimise),

the SINGLE-OBJECTIVE FEATURE SELECTION is defined as:

$$\boldsymbol{q}^* = \arg\min_{\boldsymbol{q}} \left[m\left(\boldsymbol{y}, \hat{\boldsymbol{y}}, \Phi(\mathcal{F},\boldsymbol{q})\right)\right], \tag{1}$$

and for

- K relevance measures (objective functions) $m_1, ...m_K$,

the MULTI-OBJECTIVE FEATURE SELECTION is defined as:

$$\boldsymbol{q}^* = \arg\min_{\boldsymbol{q}} [m_1\left(\boldsymbol{y}, \hat{\boldsymbol{y}}, \Phi(\mathcal{F},\boldsymbol{q})\right), ..., m_K\left(\boldsymbol{y}, \hat{\boldsymbol{y}}, \Phi(\mathcal{F},\boldsymbol{q})\right)]. \tag{2}$$

1.2 Music Classification Chain

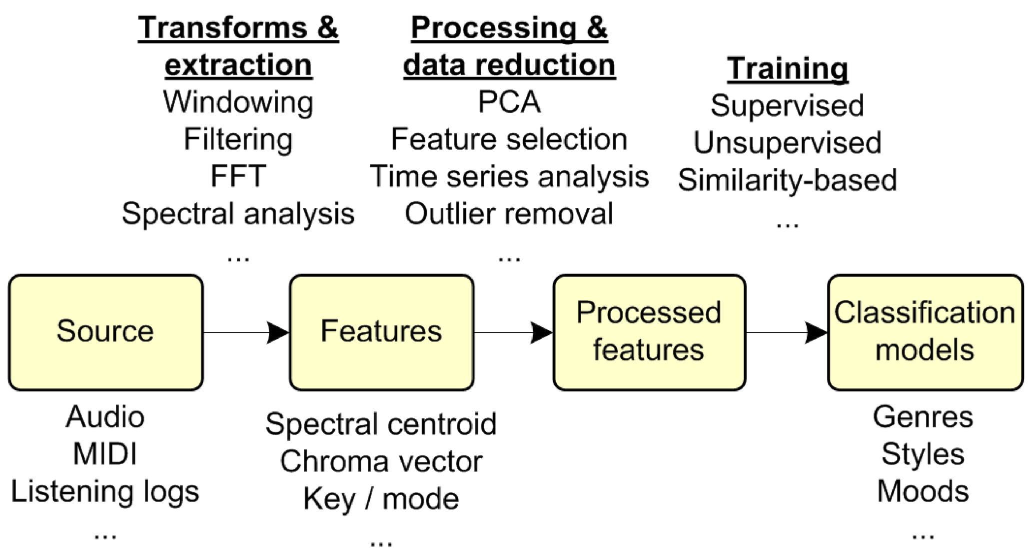

Figure 1: *Basic algorithm steps in music categorisation*

1.3 Categorisation of Relevance Measures

Groups of measures proposed for the evaluation of music classification in [1]:

- **Classification performance**: commonly applied methods based on the confusion matrix: accuracy, precision, recall, etc., also constructed for imbalanced data sets.
- **Resources**: demands on runtime and storage space for various steps in algorithm chain, see Sect. 1.2.
- **Model complexity**: measures for the identification of simple and fast models which are more robust against overfitting.
- **User related**: personal satisfaction with classification results and reduction of any personal efforts necessary to train classification models.
- **Specific performance**: evaluation of a particular task (e.g., music segmentation, tempo recognition).

2. Experiments

2.1 Setup

- CATEGORISATION TASKS: 6 genres (Classic, Electronic, Jazz, Pop, Rap, R&B), 8 styles (AdultContemporary, AlbumRock, AlternativePopRock, ClubDance, etc.).
- DATA SETS FOR EACH TASK: 20 training tracks, 120 optimisation tracks.
- FEATURES: 636 audio signal characteristics.
- CLASSIFICATION INSTANCES: 4 s time windows with 50% overlap.
- CLASSIFICATION METHODS: random forest, naive Bayes, linear SVM.
- OPTIMISATION ALGORITHM: (50+1) SMS-EMOA (for details see [2]), 3,000 generations, 28 pairs of 8 evaluation measures (see Sect. 2.2).
- OVERALL NUMBER OF EXPERIMENTS: 28 evaluation scenarios · 14 categorisation tasks · 3 classifiers · 5 statistical repetitions = 5,880.

2.2 Evaluation Measures

Given

- TP: true positives, • TN: true negatives,
- FP: false positives, • FN: false negatives,
- T: number of classification instances,
- $R(\cdot)$: the rank after the sorting of instances,

the following measures to optimise

- m_{BRE}: balanced relative error, • m_{FR}: feature rate,
- m_{PREC}: precision, • m_{REC}: recall,
- m_{SPEC}: specificity, • m_{F1}: F1-measure,
- m_{GEO}: geometric mean,
- m_{SPEAR}: Spearman's correlation coefficient between true and predicted labels

are defined as:

$$m_{BRE} = \frac{1}{2}\left(\frac{FN}{TP+FN} + \frac{FP}{TN+FP}\right), \tag{3}$$

$$m_{FR} = \frac{|\Phi(\mathcal{F},\boldsymbol{q})|}{|\mathcal{F}|}, \tag{4}$$

$$m_{PREC} = \frac{TP}{TP+FP}, \tag{5}$$

$$m_{REC} = \frac{TP}{TP+FN}, \tag{6}$$

$$m_{SPEC} = \frac{TN}{FP+TN}, \tag{7}$$

$$m_{F1} = \frac{2\cdot m_{PREC}\cdot m_{REC}}{m_{PREC}+m_{REC}}, \tag{8}$$

$$m_{GEO} = \sqrt{m_{REC}\cdot m_{SPEC}}, \tag{9}$$

$$m_{SPEAR} = \frac{\sum_{i=1}^{T}\left(R(\hat{y}(i))\cdot R(y(i))\right) - T\left(\frac{T+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^{T}\left(R^2(\hat{y}(i))\right) - T\left(\frac{T+1}{2}\right)^2\right)\cdot\left(\sum_{i=1}^{T}\left(R^2(y(i))\right) - T\left(\frac{T+1}{2}\right)^2\right)}}. \tag{10}$$

2.3 Evaluation of Multi-Objectiveness

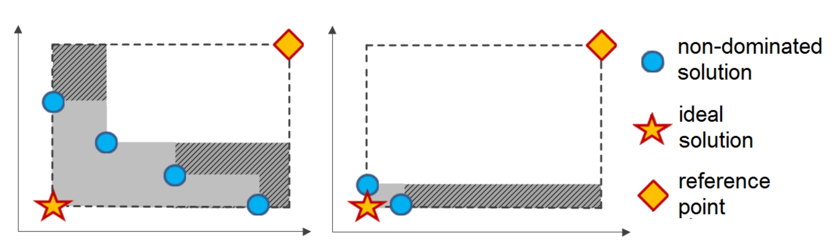

Figure 2: *Stronger (left) and weaker (right) advantage of multi-objective against single-objective approach*

Given

- N: number of solutions $\boldsymbol{q}_1, ..., \boldsymbol{q}_N$ in a front,
- $\mathbf{r}$: reference point, • $\boldsymbol{q}_{ID}$: ideal solution,

the HYPERVOLUME is defined as:

$$\mathcal{S}(\boldsymbol{q}_1, ..., \boldsymbol{q}_N) = vol\left(\bigcup_{i=1}^{N}[\boldsymbol{q}_i, \mathbf{r}]\right), \tag{11}$$

and share of the hypervolume exclusively dominated by the ideal solution is:

$$\epsilon_{ID} = \frac{\mathcal{S}(\boldsymbol{q}_{ID}) - \mathcal{S}(\boldsymbol{q}_1, ..., \boldsymbol{q}_N)}{\mathcal{S}(\boldsymbol{q}_{ID})}\cdot 100\%. \tag{12}$$

The share of the hypervolume of the front without the solution with maximum contribution to hypervolume is:

$$\epsilon_{MAX} = \frac{\mathcal{S}(\boldsymbol{q}_1, ..., \boldsymbol{q}_N) - \max\limits_{i\in\{1,...,N\}}\mathcal{S}(\boldsymbol{q}_i)}{\mathcal{S}(\boldsymbol{q}_1, ..., \boldsymbol{q}_N)}\cdot 100\%. \tag{13}$$

3. Analysis of Results

3.1 Trade-offs between Objectives

- (a): Electronic, $\epsilon_{ID} = 33.90\%$, $\epsilon_{MAX} = 39.52\%$.
- (b): Classic, $\epsilon_{ID} = 4.20\%$, $\epsilon_{MAX} = 12.99\%$.
- (c): R&B, $\epsilon_{ID} = 21.48\%$, $\epsilon_{MAX} = 33.66\%$.
- (d): AdultContemporary, $\epsilon_{ID} = 24.14\%$, $\epsilon_{MAX} = 27.40\%$.
- (e): Rap, $\epsilon_{ID} = 0.02\%$, $\epsilon_{MAX} = 0.71\%$.
- (f): Rap, $\epsilon_{ID} = 0\%$, $\epsilon_{MAX} = 0.14\%$.

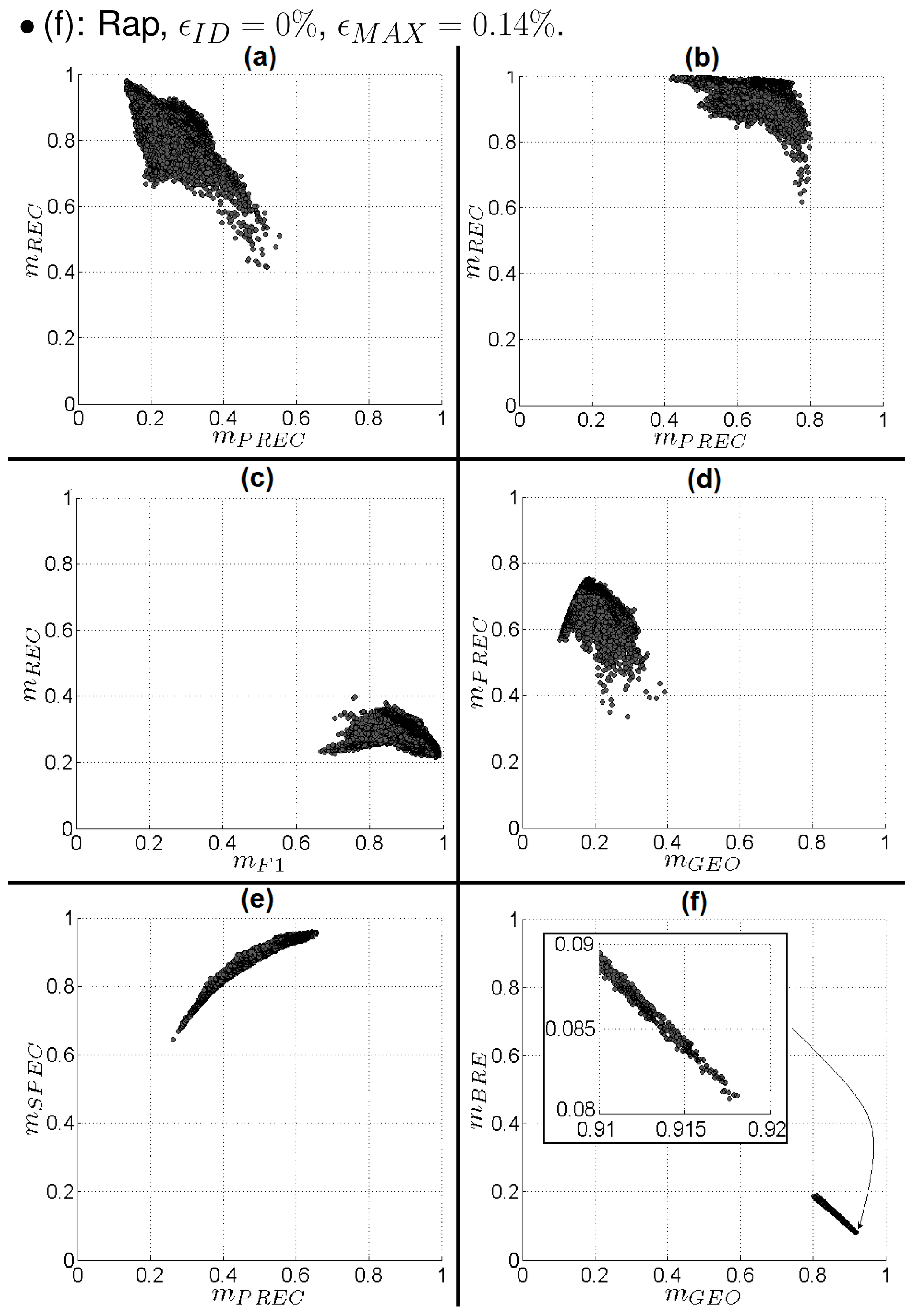

Figure 3: *Combinations of categories and objectives*

3.2 Comparison of Objective Pairs

- (black): pair in the row has a significantly higher ϵ.
- (white): pair in the row has a significantly lower ϵ.
- (grey): no significant difference.

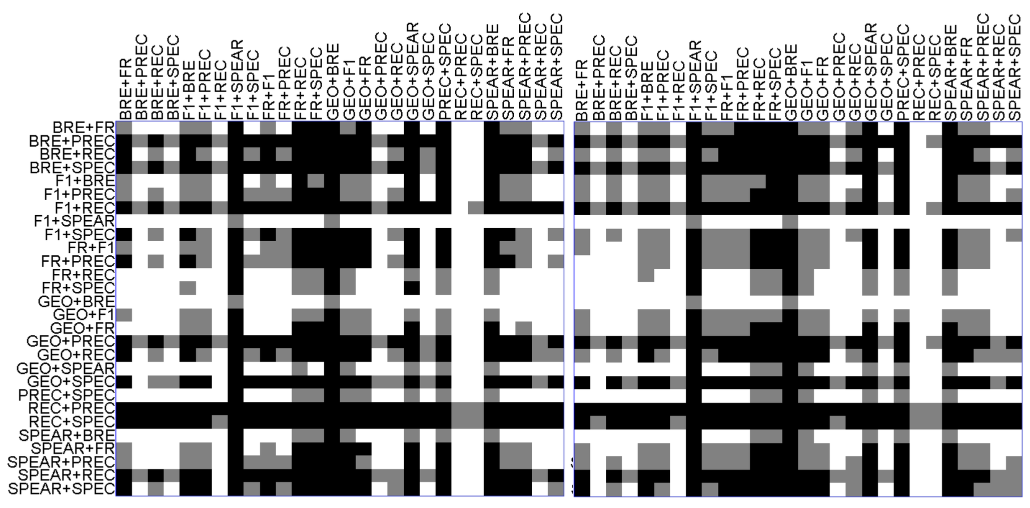

Figure 4: *Comparison of objective pairs based on ϵ_{ID} (left) and ϵ_{MAX} (right)*

3.3 Correlation between ϵ_{ID} and ϵ_{MAX}

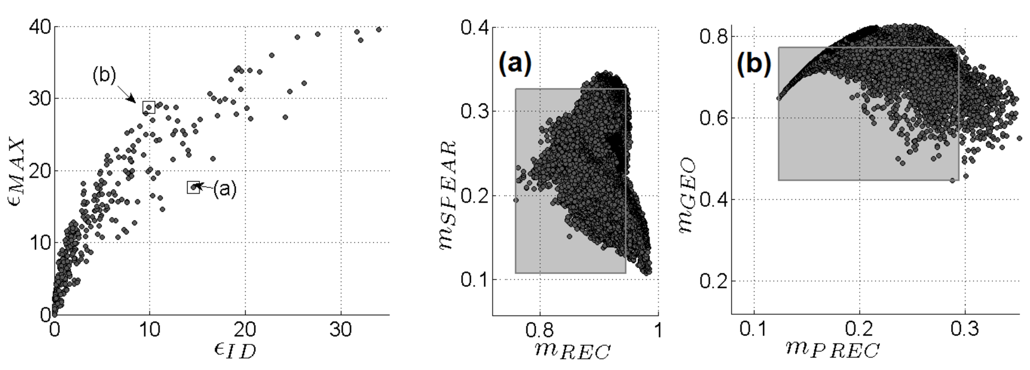

Figure 5: *Example of difference between ϵ_{ID} and ϵ_{MAX}*

4. Future Research

- Other objectives from different groups (cf. Sect. 1.3).
- Three and more objectives at the same time.
- Further classification tasks and scenarios.
- Impact of optimisation parameters.

References

[1] Vatolkin, I., Preuß, M., Rudolph, G.: Multi-Objective Feature Selection in Music Genre and Style Recognition Tasks. *Proceedings of the 13th Annual Genetic and Evolutionary Computation Conference (GECCO)*, pp. 411–418, 2011.

[2] Vatolkin, I.: Improving Supervised Music Classification by Means of Multi-Objective Evolutionary Feature Selection. PhD thesis, TU Dortmund, 2013.

EMO 2015 8th International Conference on Evolutionary Multi-Criterion Optimization, Guimarães, Portugal, 29 March - 1 April, 2015