Interpretability of Music Classification as a Criterion for Evolutionary Multi-Objective Feature Selection

technische universität dortmund

Igor Vatolkin[1], **Günter Rudolph**[1], **Claus Weihs**[2]

[1]Chair of Algorithm Engineering, Department of Computer Science, TU Dortmund, Germany
[2]Chair of Computational Statistics, Faculty of Statistics, TU Dortmund, Germany
{igor.vatolkin;guenter.rudolph;claus.weihs}@tu-dortmund.de

ALGORITHM ENGINEERING

Overview

- Categorisation tasks: automatic recognition of music genres and styles.
- Huge number of available features for training of classification models.
- Mostly low-level audio signal characteristics, hard to interpret for music listeners and musicologists.
- Proposal: maximisation of share of interpretable features and minimisation of classification error.
- Optimisation algorithm: evolutionary multi-objective feature selection.
- Hypervolume-based analysis of trade-off between objectives (share of interpretable features and classification error).

1. Introduction and Background

1.1 Music Classification Chain

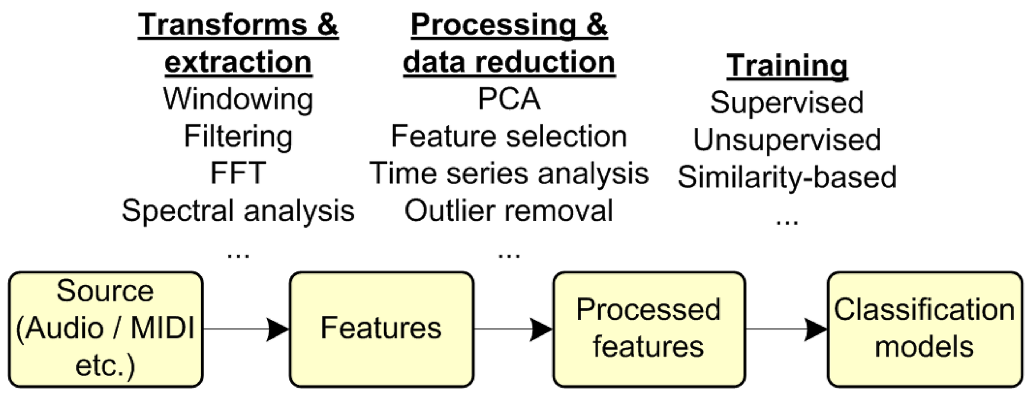

Figure 1: *Basic algorithm steps in music categorisation*

1.2 Advantages of Audio Features

- Compared to MIDI features: extractable for any digital music piece independently of score availability.
- Compared to listener tags and meta data: stable values, less influenced by missing or noisy descriptors (impact of popularity).
- Time consuming extraction can be done offline.

1.3 Advantages of Interpretable Models

- Comprehensible organisation and management of large music collections.
- Theoretical interpretation of relevant properties of genres, styles, etc.
- New recommendation scenarios based on listener preferences across personal categories.
- Automatic composition of music based on high-level characteristics.

1.4 Categorisation of Features

Table 1: *Groups of audio features for music classification*

Group	Examples	No.
LOW-LEVEL AUDIO FEATURES		
Cepstral domain	Mel frequency cepstral coefficients	202
Chroma and harmony	Fundamental frequency, chroma vector	202
ERB and Bark domains	Bark scale magnitudes	106
Phase domain	Angles and distances	4
Rhythm	Characteristics of fluctuation patterns	24
Spectral domain	Spectral centroid, tristimulus	58
Tempo and correlation	Periodicity peak	6
Time domain	Linear prediction coefficients, zerocrossing rate	34
HIGH-LEVEL AUDIO FEATURES		
Chord statistics	Number of recognised chords in 10s	5
Chroma and harmony	Key, consonance, strengths of pitch intervals	258
Instruments	Share of guitar, piano, strings, wind instruments	32
Moods	Aggressive, earnest, energetic, sentimental	64
Structural complexity	Complexity of chords, harmony, instruments	70
Tempo, rhythm, structure	Beats per minute, rhythmic clarity	9
Various features	Activation level, vocal descriptors	128

1.5 Extraction Procedure

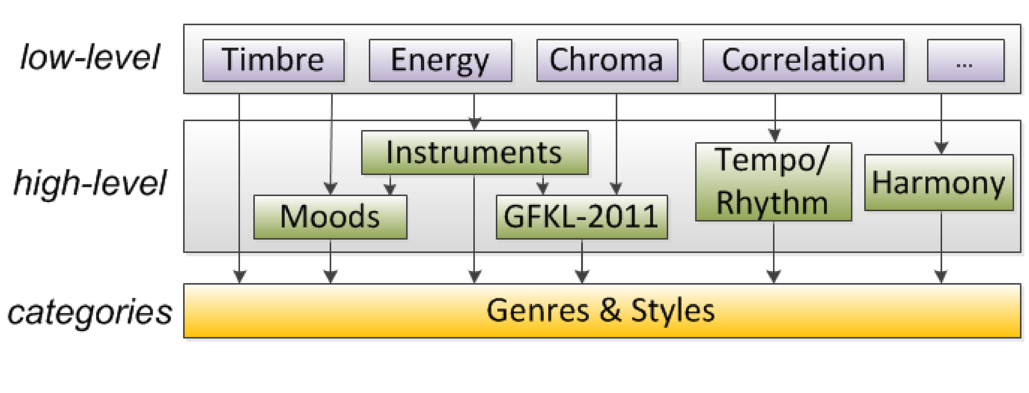

Figure 2: *Extraction of low-level and high-level features after [2]*

2. Optimisation Problem

Given

- TP: true positives,
- TN: true negatives,
- FP: false positives,
- FN: false negatives,
- F: number of features,
- F_{HL}: number of high-level features,

the objectives to optimise are:

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

$$m_{HL} = \frac{F_{HL}}{F}. \quad (2)$$

For

- $\boldsymbol{q}$: binary vector which indicates selected features,
- $\boldsymbol{q}^*$: optimal index vector,
- $\mathcal{F}(\boldsymbol{q})$: set of features indicated in $\boldsymbol{q}$,
- $\boldsymbol{y}$: true labels,
- $\hat{\boldsymbol{y}}$: predicted labels,

the optimisation task is defined as follows:

$$\boldsymbol{q}^* = \arg\min_{\boldsymbol{q}}[m_{BRE}(\boldsymbol{y}, \hat{\boldsymbol{y}}, \mathcal{F}(\boldsymbol{q})), 1 - m_{HL}(\mathcal{F}(\boldsymbol{q}))], \quad (3)$$

3. Experiments

3.1 Setup

- CATEGORISATION TASKS: 6 genres, 8 styles.
- DATA SETS FOR EACH TASK: 20 training tracks, 120 optimisation tracks, 120 holdout (validation) tracks.
- FEATURES: 1,202 (636 low-level, 566 high-level), see Sect. 1.4.
- CLASSIFICATION INSTANCES: 4 s time windows with 50% overlap.
- CLASSIFICATION METHODS: decision tree C4.5, random forest, naive Bayes, linear SVM.
- OPTIMISATION ALGORITHM: (50+1) SMS-EMOA with asymmetric mutation [1].
- INITIAL SOLUTIONS: with equal probability (1) only low-level features allowed, (2) only high-level, (3) both kinds. Selection probability for each feature: 20%.
- EVALUATIONS AND REPETITIONS: 3,000 EA generations, 10 statistical repetitions.

3.2 Results

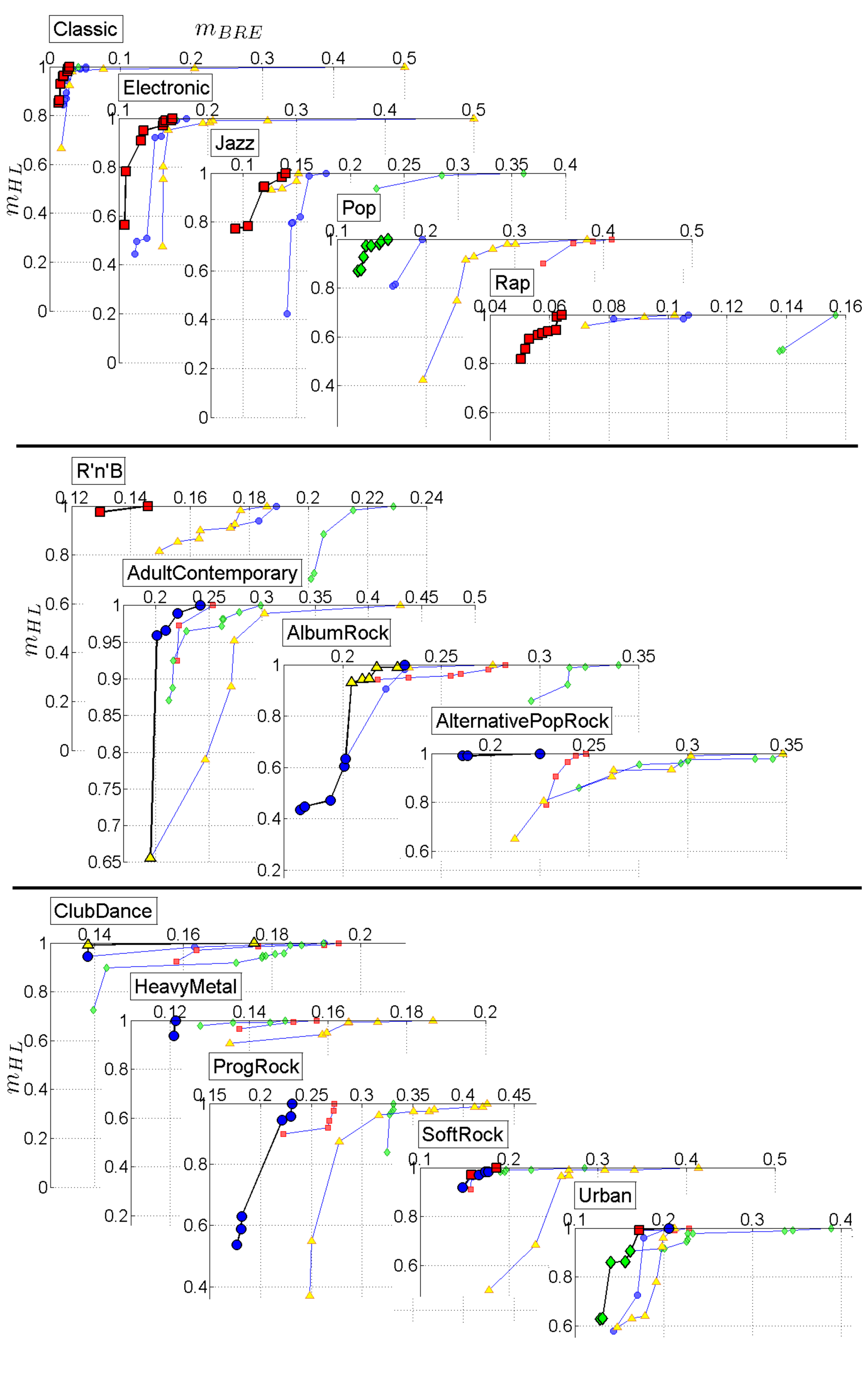

Figure 3: *Non-dominated fronts after optimisation*

4. Analysis of Results

4.1 Trade-Off between Objectives

Given

- N: number of solutions (feature subsets) in a front,
- $\mathbf{r}$: reference point (worst possible solution with $m_{BRE} = 1$ and $m_{HL} = 0$),
- $\boldsymbol{q}_{ID}$: ideal solution at individual best values of m_{BRE} and m_{HL},

the hypervolume is defined as:

$$\mathcal{S}(\boldsymbol{q}_1, ..., \boldsymbol{q}_N) = vol\left(\bigcup_{i=1}^{N}[\boldsymbol{q}_i, \mathbf{r}]\right), \quad (4)$$

and the hypervolume exclusively dominated by the ideal solution is:

$$\mathcal{S}_\alpha = \mathcal{S}(\boldsymbol{q}_{ID}) - \mathcal{S}(\boldsymbol{q}_1, ..., \boldsymbol{q}_N). \quad (5)$$

Table 2: *Characteristics of non-dominated fronts*

Category (6 genres/8 styles)	$\mathcal{S}_\alpha$	m_{BRE} $(\boldsymbol{q}_L)$	m_{HL} $(\boldsymbol{q}_L)$	$\lvert\boldsymbol{q}_L\rvert$	m_{BRE} $(\boldsymbol{q}_R)$	$\lvert\boldsymbol{q}_R\rvert$
Classic	0.00089	0.0128	0.8515	101	0.0276	105
Electronic	0.00566	0.1070	0.5636	110	0.1610	124
Jazz	0.00695	0.0929	0.7714	105	0.1400	109
Pop	0.00149	0.1240	0.8707	116	0.1575	99
Rap	0.00123	0.0504	0.8182	99	0.0642	104
R'n'B	0.00040	0.1297	0.9750	120	0.1458	99
AdultContemporary	0.00299	0.1952	0.6555	119	0.2417	89
AlbumRock	0.01452	0.1785	0.4359	117	0.2316	89
AlternativePopRock	0.00038	0.1859	0.9900	100	0.2251	96
ClubDance	0.00035	0.1384	0.9444	108	0.1760	94
HeavyMetal	0.00002	0.1210	0.9358	109	0.1213	86
ProgRock	0.01757	0.1763	0.5367	177	0.2309	117
SoftRock	0.00148	0.1480	0.9196	112	0.1862	108
Urban	0.00861	0.1290	0.6283	113	0.2061	111

4.2 Most Important Low-Level Features

- Identification of low-level features which strongly contribute to reduction of m_{BRE}.
- Baseline assumption: each low-level feature has equal probability to appear in non-dominated subsets for a given category.
- Feature occurrence rank describes the proportion of times a feature appears in non-dominated subsets related to expected number of occurrences (for estimation procedure see the paper).
- $\mathcal{R}_i$ is the average occurrence rank across categories for which feature i appears in non-dominated subsets.

Table 3: *Most relevant low-level features. LPC: linear prediction coefficient; MFCC: mel frequency cepstral coefficient*

Name	Categories	$\mathcal{R}_i$
Mean(4th delta MFCC)	Elec, Jazz, Rap, Soft	71.167
Stddev(7th LPC)	Adul, Prog, Soft, Urba	37.626
Mean(2nd tristimulus)	Clas, Elec, Albu, Prog	6.568
Stddev(12th delta MFCC)	Elec, Pop, Albu, Prog	5.911

5. Conclusions and Outlook

5.1 Summary of Results

- Large initial set of base features for various music categorisation tasks (1,202 descriptors).
- Strong reduction of number of features compared to complete set without decrease of classification error after feature selection.
- Increase of interpretability without restriction to interpretable features only.
- Analysis of trade-off between interpretable and low-level characteristics.
- Identification of most relevant low-level features.

5.2 Future Research

- Systematic tuning of EA parameters.
- Design of further interpretable features.
- More interpretable classification models (small decision trees, fuzzy rules).

References

[1] Bischl, B., Vatolkin, I., Preuß, M.: Selecting Small Audio Feature Sets in Music Classification by Means of Asymmetric Mutation. *Proc. 11th Int'l Conf. on Parallel Problem Solving From Nature (PPSN)*, pp. 314–323, 2010.

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

EvoMusArt-2015 4th International Conference on Evolutionary and Biologically Inspired Music, Sound, Art and Design, Kopenhagen, Denmark, April 8-10, 2015