Performance of Specific vs. Generic Feature Sets in Polyphonic Music Instrument Recognition

Igor Vatolkin[1], Anil Nagathil[2], Wolfgang Theimer[3], Rainer Martin[2]

[1]Chair of Algorithm Engineering, TU Dortmund

[2]Institute of Communication Acoustics, Ruhr-Universität Bochum

[2]Research in Motion, Bochum

22.03.2013

EMO 2013, Sheffield, UK

Outline

- Introduction
- Algorithms
- Experimental study
- Results and discussion
- Conclusions

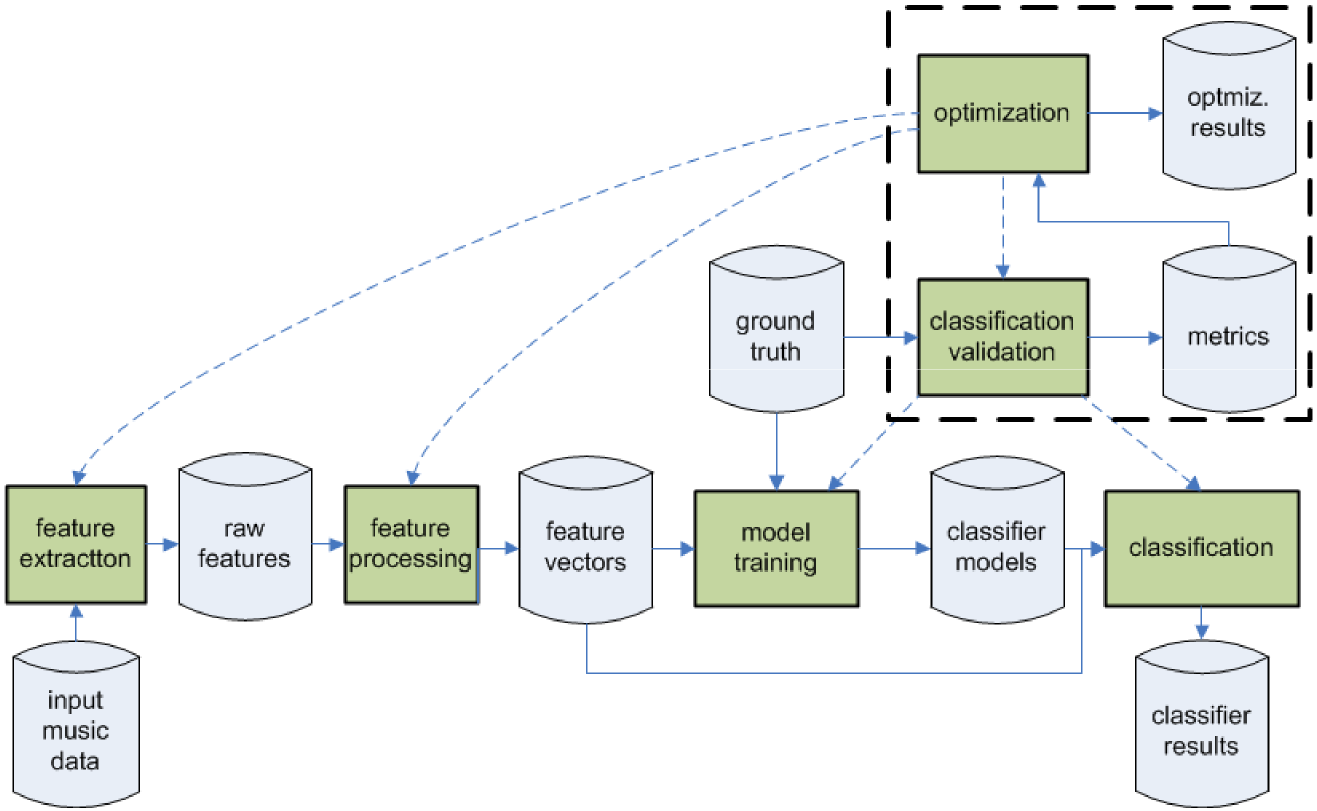
Outline

- **Introduction**
- Algorithms
- Experimental study
- Results and discussion
- Conclusions

Purpose of Instrument Recognition

- Management of large music collections
- Music recommendation
- Analysis of genre and style characteristics
- Note / vocals correction
- Music transcription

Supervised Classification Chain



Outline

- Introduction
- **Algorithms**
- Experimental study
- Results and discussion
- Conclusions

Reasons for Feature Selection

- Classifier performance suffers from too many irrelevant features
- The same base feature set for different tasks
- Optimization of feature extraction/processing/classification
  - Runtime demands decrease
  - Storage space demands decrease
- Reduced risk of model overfitting
- Understanding of classification category properties

Multi-Objective Feature Selection

- Feature selection:

$$\theta^* = \arg\min_{\theta}[I(Y;\Phi(X,\theta))], X \in \mathbb{R}^d$$

  - X : original feature set
  - Φ : selected feature set
  - θ : indices of selected features
  - Y : target variable (category)
  - I : relevance function

- Multi-objective feature selection:

$$\theta^* = \arg\min_{\theta}[I_1(Y;\Phi(X,\theta)),\ldots,I_O(Y;\Phi(X,\theta))]$$

Optimization Algorithm

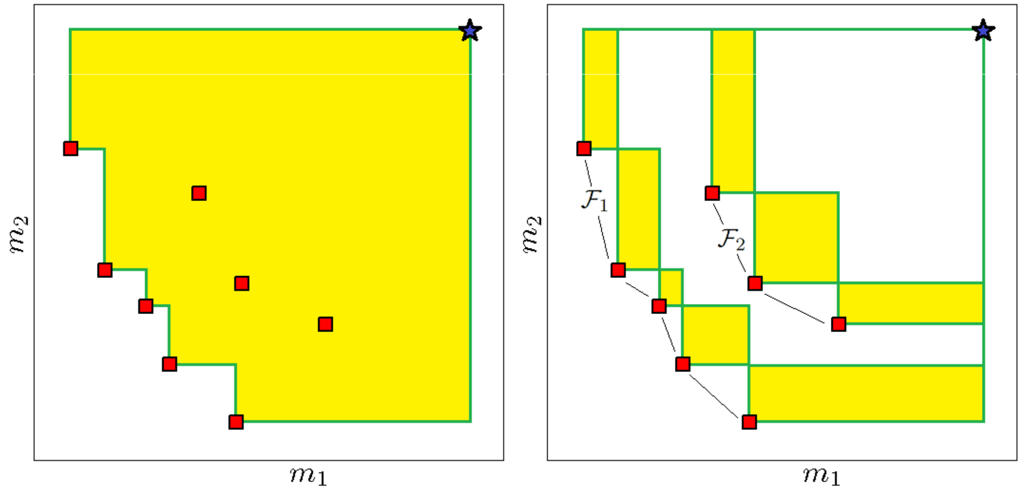
- SMS-EMOA [N. Beume, B. Naujoks, M. Emmerich: SMS-EMOA: Multiobjective Selection Based on Dominated Hypervolume. European Journal of Operational Research 181(3):1653–1669, 2007]

$$\mathcal{S}(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \text{vol}\left(\bigcup_{i=1}^{N}[\mathbf{x}_i,\mathbf{r}]\right) \quad \Delta\mathcal{S}(\mathbf{x}_i) = \mathcal{S}(\mathbf{x}_1,\ldots,\mathbf{x}_N) - \mathcal{S}(\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\ldots,\mathbf{x}_N)$$

- Asymmetric mutation

$$p_m(i) = \frac{\gamma}{|X|}|\mathbf{v}_i - p_{01}|$$

  - $p_m(i)$: bit flip prob.
  - γ : step size
  - $\mathbf{v}_i$: bit vector value
  - p_{01}: prob. of switching 0 to 1



Outline

- Introduction
- Algorithms
- **Experimental study**
- Results and discussion
- Conclusions

technische universität dortmund

Experiment Data Sets 1/2

- Binary categorization tasks
  - Piano
  - Guitars (acoustic / electric)
  - Wind (flute / trumpet)
  - Strings (viola / violine / cello)
- Instrument samples from McGill, RWC and Iowa databases
  - Different playing styles
  - Approximately the same loudness
  - 3000 chords mixed randomly (3 or 4 tones)

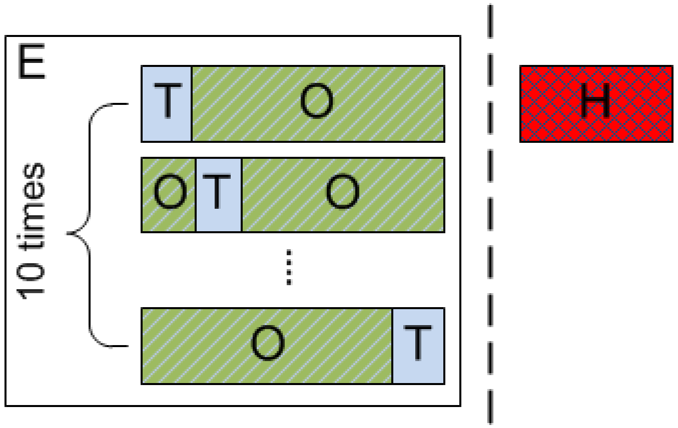
piano
piano
piano

violin
viola
cello

acoustic guitar
piano
trumpet
electric guitar

Experiment Data Sets 2/2

- Experiment set E: 2000 chords
- Holdout set H: 1000 chords
- 10-fold cross-validation
  - Training set T: 200 chords
  - Optimization set O: 1800 chords



Experiments - Features

- Features
  - 353 block characteristics from

    [M. Eichhoff, C. Weihs: Musical Instrument Recognition by High-Level Features. Proc. GFKL 2010, 373–381, 2012]
  - 795 short-term signal characteristics extracted by AMUSE

    [I. Vatolkin, W. Theimer, M. Botteck: AMUSE (Advanced MUSic Explorer) – A Multitool Framework for Music Data Analysis, Proc. ISMIR, 33–38, 2012]
  - 102 constant-Q features

    [J. C. Brown: Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features. J. Acoust. Soc. Am., 105(3), 1933–1941, 1999]

Experiments – Classification Methods

- Decision tree C4.5
- Random forest
- Naive Bayes
- Support vector machine with linear kernel

Experiments – Optimization Metrics

- Specific feature selection

$$E^2 = \frac{1}{L}\sum_{i=1}^{L}(\hat{s}_i - s_i)^2 \qquad f_r = \frac{|\Phi(X,\theta)|}{|X|}$$

- Generic feature selection

$$\widehat{E^2} = \frac{1}{C}\sum_{k=1}^{C}\left(\frac{1}{F}\sum_{j=1}^{F}\left(\frac{1}{L}\sum_{i=1}^{L}(\hat{s}_i - s_i)^2\right)\right) \qquad f_r = \frac{|\Phi(X,\theta)|}{|X|}$$

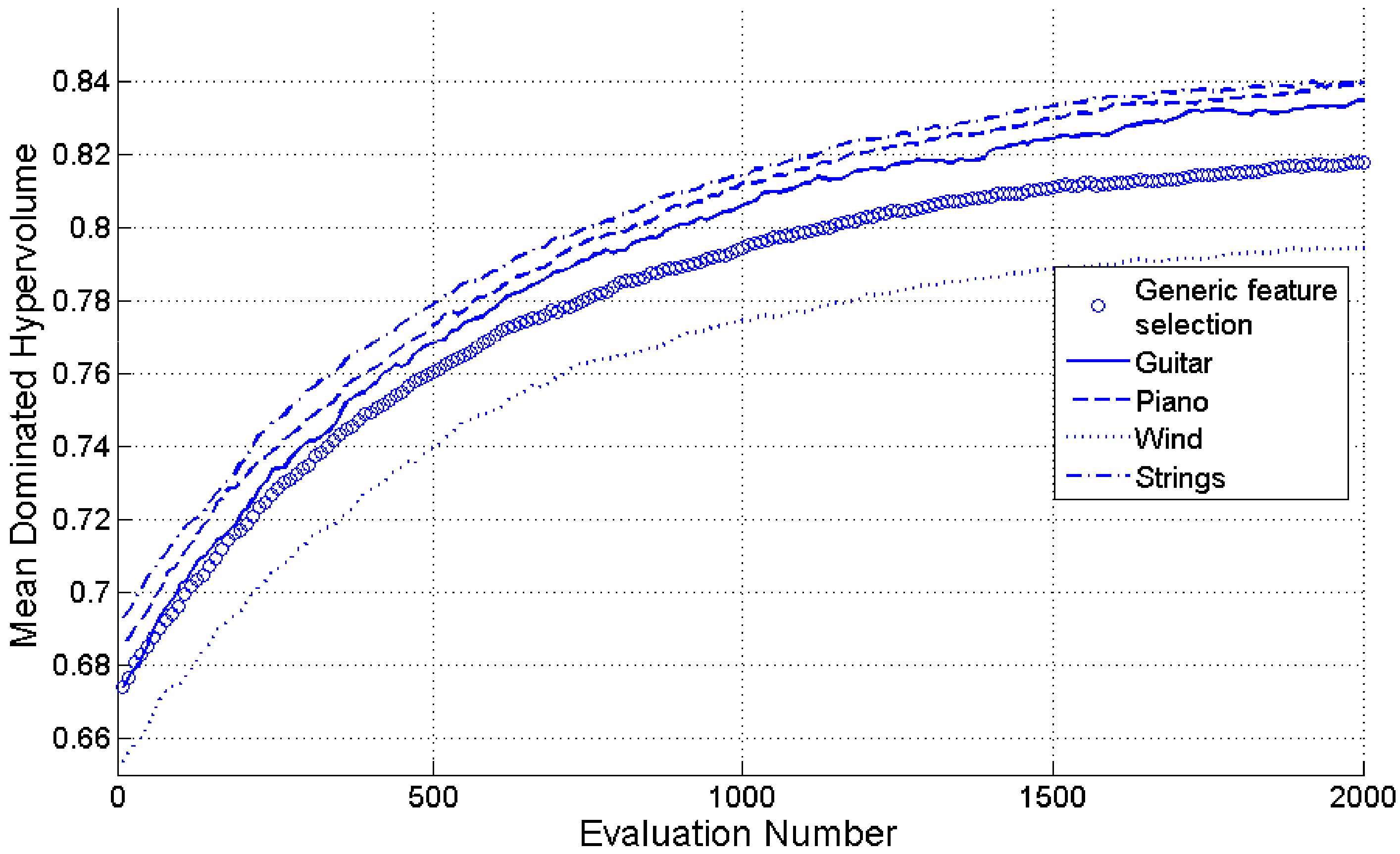
  - $s_i, \hat{s}_i \in \{0;1\}$: labeled and predicted relationships
  - L: number of chords
  - F: number of folds
  - C: number of classification tasks

Outline

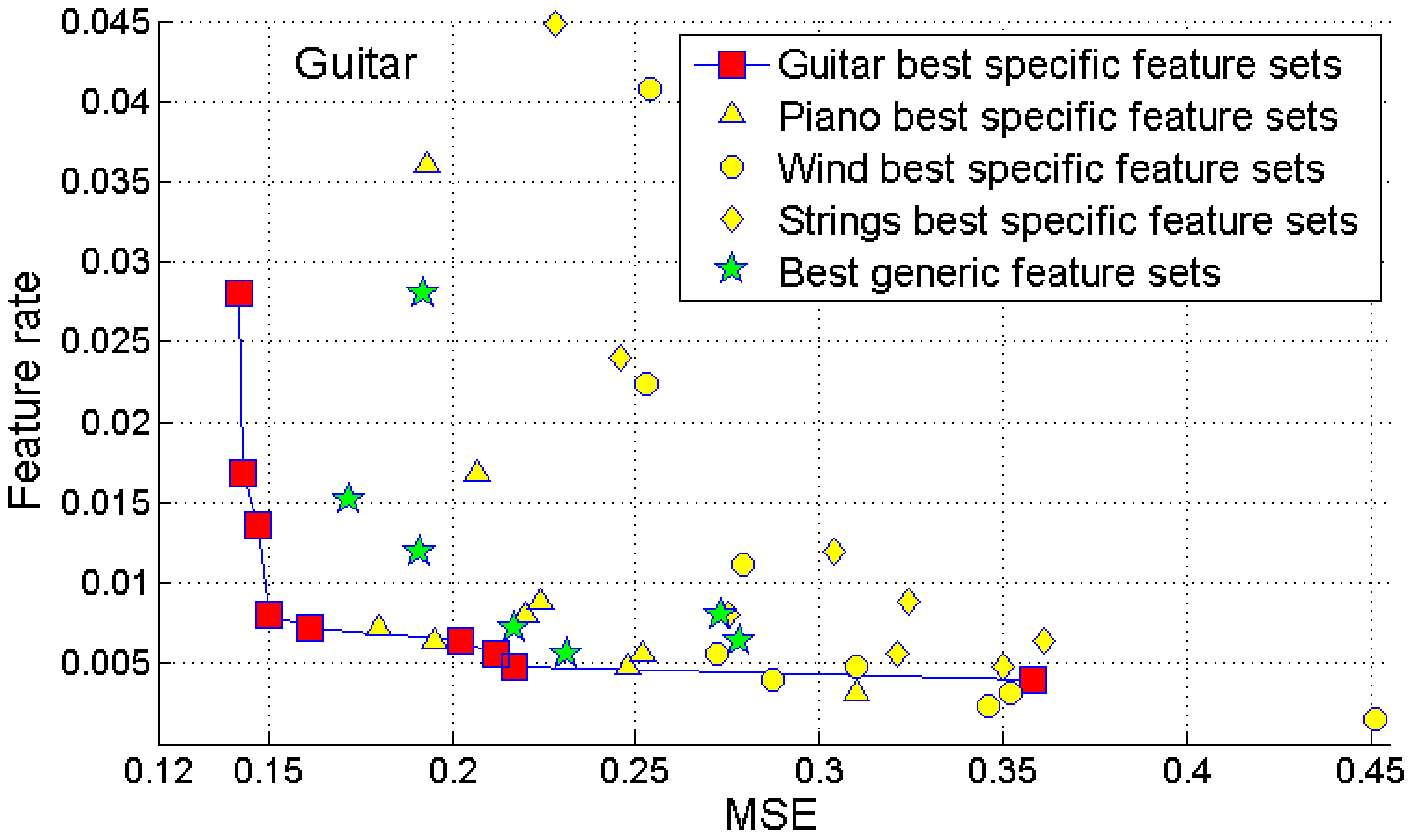
- Introduction
- Algorithms
- Experimental study
- **Results and discussion**
- Conclusions

technische universität
dortmund

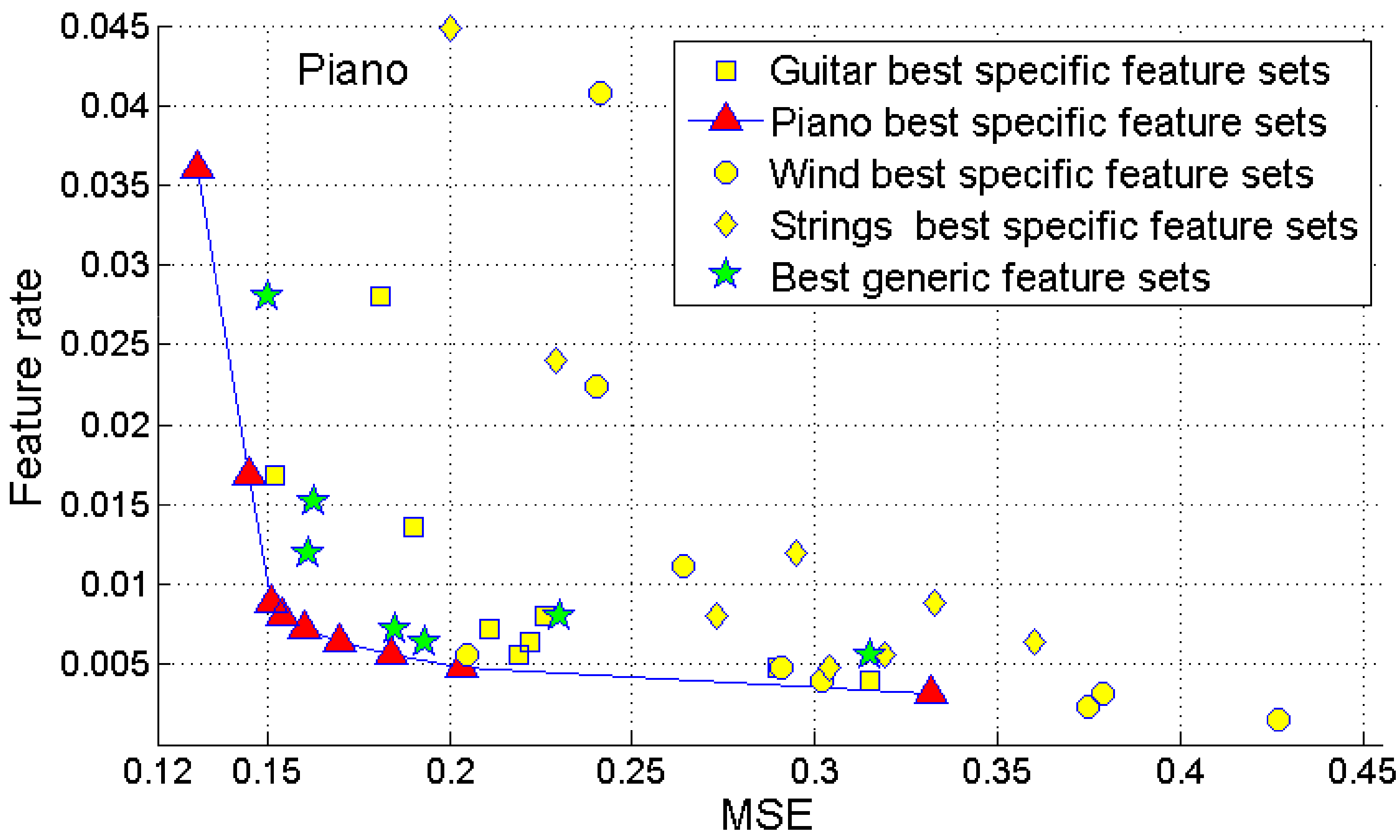
Optimization Performance on Holdout Set



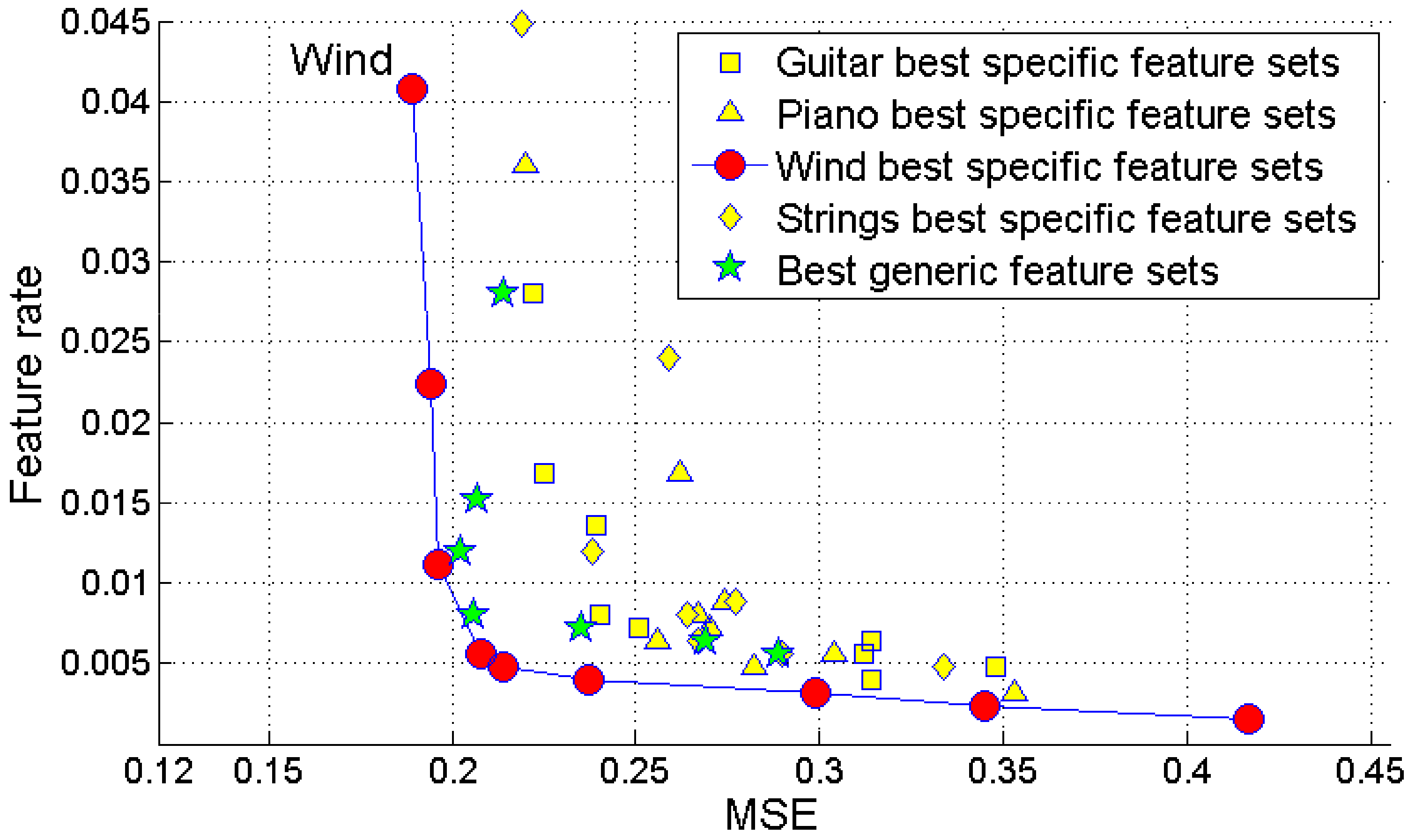
Specific vs. Generic Features 1/4



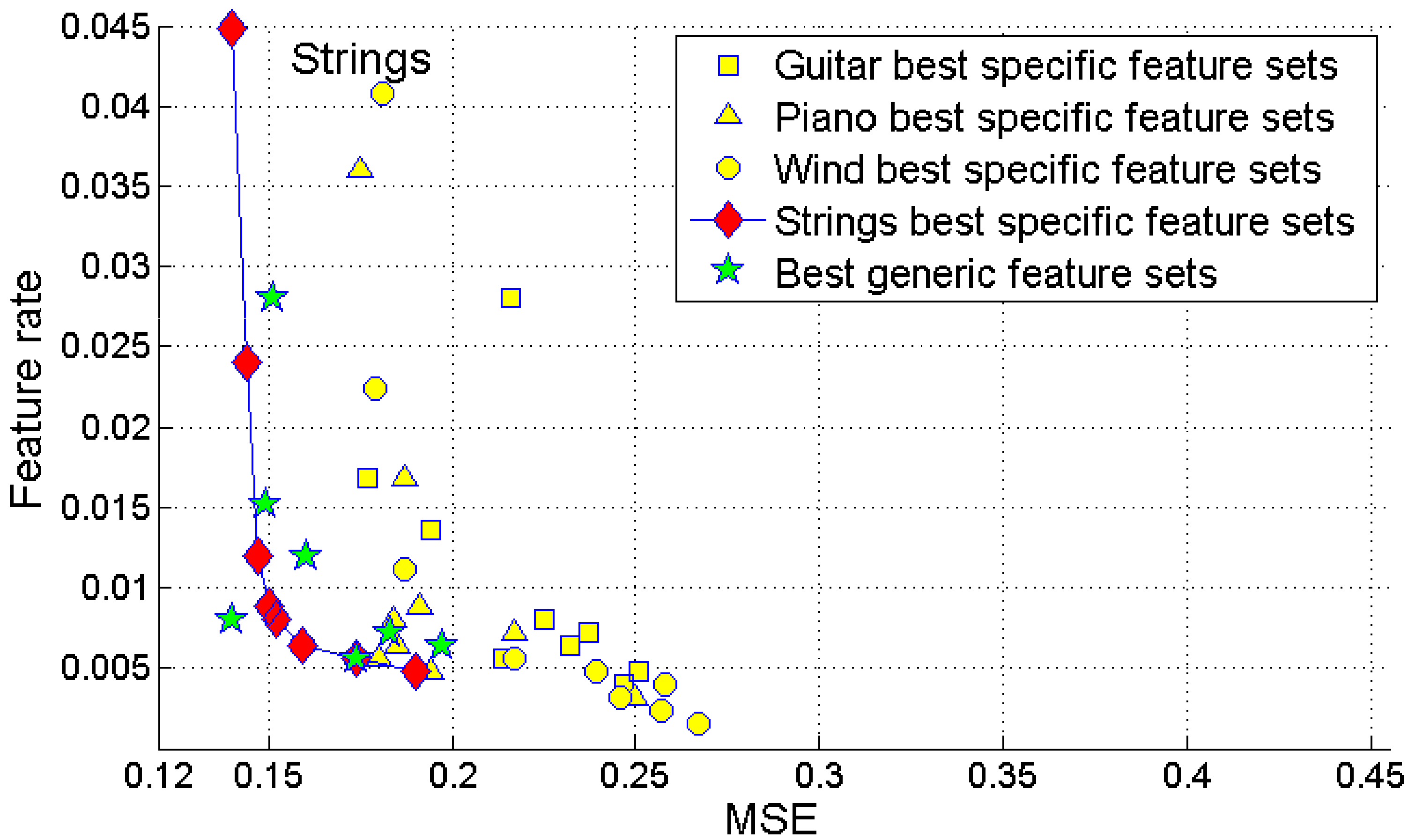
Specific vs. Generic Features 2/4



Specific vs. Generic Features 3/4



Specific vs. Generic Features 4/4



Outline

- Introduction
- Algorithms
- Experimental study
- Results and discussion
- **Conclusions**

Summary

- Feature selection increases hypervolume on holdout set
- Task specific features often the best
- Generic features perform quite well
  - For Guitar at worst
  - For Strings at best
- Specific features from other tasks are less suited

Current and Further Work

- Instruments as high-level features for genre and style prediction [PhD, to appear]
- More instruments
- Playing styles analysis (e.g. open vs. fretted strings)
- Generic feature selection for further task groups
- Other multi-objective scenarios [I. Vatolkin, M. Preuß, G. Rudolph: Multi-Objective Feature Selection in Music Genre and Style Recognition Tasks. Proc. GECCO, 411–418, 2011]
  - Training set size vs. classification quality
  - Model stability vs. classification quality
- Nested resampling with early stopping [J. Loughrey, P. Cunningham. Overfitting in Wrapper-Based Feature Subset Selection: The Harder You Try the Worse It Gets., Proc. SGAI, 33–43,2004]