TAMPERE UNIVERSITY OF TECHNOLOGY

Semi-supervised Musical Instrument Recognition

Master's Thesis Presentation

Aleksandr Diment[1]

[1]Tampere University of Technology, Finland

Supervisors:
Adj.Prof. Tuomas Virtanen, MSc Toni Heittola

17 May 2013

Outline

1. Introduction
2. System description
3. Evaluation
4. Conclusions

1 Introduction

2 System description

3 Evaluation

4 Conclusions

Introduction

Musical instrument recognition

Music information retrieval: obtaining information of various kinds from music.

Situationally tailored playlisting, personalised radio etc.

Instrument recognition $\Rightarrow$

- automatic music database annotation;
- automatic music transcription;
- musical genre classification.

Introduction

Semi-supervised learning

Traditional learning paradigms:

- unsupervised: no additional knowledge about the data samples
- supervised: a label is assigned to each data sample

Supervised: large amounts of annotated training data needed.

SSL: only part of training data needs to be annotated.

Not yet applied for instrument recognition.

Introduction

Objectives and main results

Objectives:

- studying techniques for musical instrument recognition;
- studying various SSL schemes;

Main results:

- developed pattern recognition system for instrument recognition with SSL;
- two algorithms are implemented.

System description

Building blocks

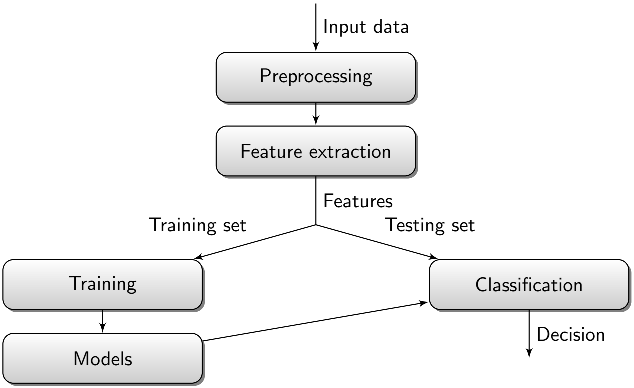


Figure 1: A block diagram of a typical pattern classification system.

System description

Feature extraction

Relevant information is within the timbre, unique to a particular group of instruments.

Set of tonal qualities which characterise a particular musical sound.

Everything except pitch, loudness and duration.

System description

Feature extraction

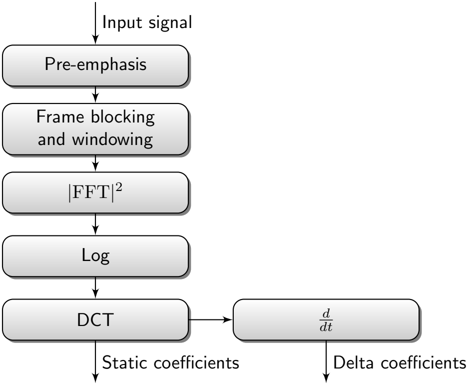


Figure 2: A block diagram of MFCCs calculation.

System description

Training algorithms

The EM algorithm $\Rightarrow$ MLEs of model parameters when treating observations as incomplete data.

Used when obtaining direct equations for the model is impossible.

Extending EM for SSL: the *iterative* and *incremental* EM-based algorithms[1].

[1] P. J. Moreno and S. Agarwal, "An experimental study of EM-based algorithms for semi-supervised learning in audio classification", in *Proc. of the ICML-2003 Workshop on the Continuum from Labeled to Unlabeled Data*, 2003.

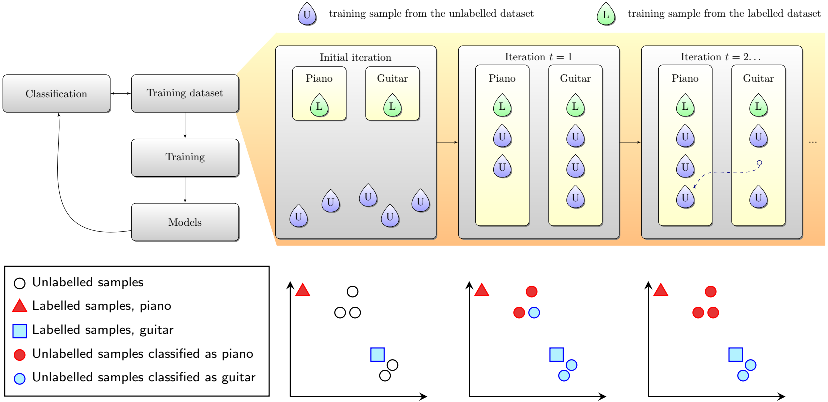


Figure 3: Schematic and feature space representation of the training stage with three iterations of the **iterative** EM-based algorithm.

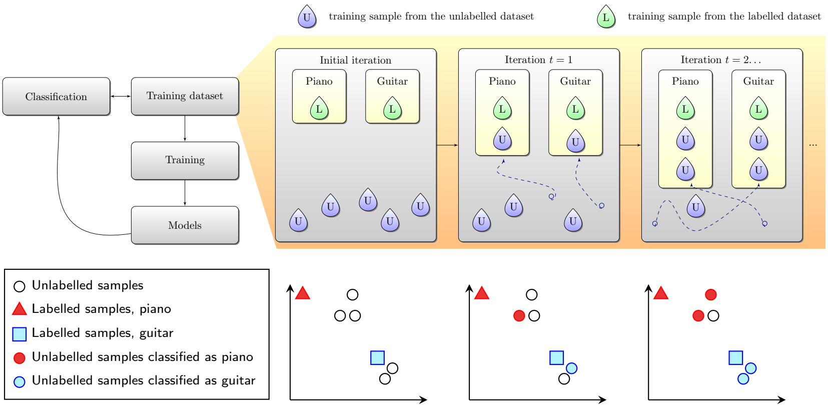


Figure 4: Schematic and feature space representation of the training stage with three iterations of the **incremental** EM-based algorithm.

System description

Labelled data weighting

The algorithms improve the performance in case the initial labelled data size is relatively low.

Not much difference when adding large or small amount of unlabelled data.

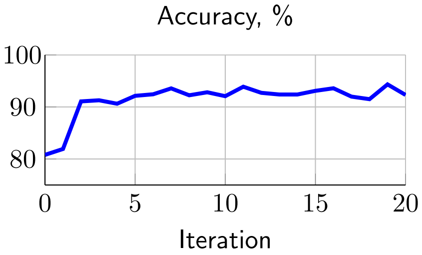
$\Rightarrow$ de-weight the contribution of the unlabelled data.

Simply: $S^l = \omega(t) \diamondsuit S^l$.

System description

One-class-at-a-time training

Another issue of the iterative algorithm: resulting classification accuracy is oscillating along the iteration axis.



System description

One-class-at-a-time training

One-class-at-a time approach: at an iteration only one class is retrained.

$\Rightarrow$ previous models of one class are unaffected by training another one

$\Rightarrow$ fewer peaks

$\Rightarrow$ safer to chose an arbitrary iteration index for termination.

1 Introduction

2 System description

3 Evaluation

4 Conclusions

Evaluation

Datasets

- RWC Music Database[2].
- Three variations per instrument.
- Separate notes within the whole range with a step of a semitone.
- 44.1 kHz, 16 bit.

[2] M. Goto et al., “RWC music database: music genre database and musical instrument sound database”, in *Proc. of the 4th Int. Conf. on Music Information Retrieval (ISMIR)*, 2003, pp. 229–230.

Evaluation

Datasets

Table 1: List of instruments and number of recordings of the notes used in the smaller and larger sets, respectively.

Instrument	# notes
Acoustic Guitar	702
Electric Guitar	702
Tuba	270
Bassoon	360
Total	2 212

Instrument	# notes
Pianoforte	792
Classic Guitar	702
Electric Guitar	702
Electric Bass	507
Trombone	278
Tuba	270
Horn	288
Bassoon	360
Clarinet	360
Banjo	941
Total	5 200

Evaluation

Baseline results

Baseline scenario: treating all available data as labelled. Upper bound for possible SSL performance.

Table 2: Average classification accuracy across all classes in fully-supervised case.

Instrument set size	Recognition accuracy, %
4 instruments	92.1
10 instruments	82.9

Evaluation

Results with the iterative EM-based algorithm

Table 3: Classification results with the iterative EM-based SSL algorithm when incorporating both modifications.

Instrument set size	**Recognition accuracy, %**			
	initial	maximum	absolute gain	relative gain
4 instruments	89.19	95.30	6.11	6.85
10 instruments	58.73	68.43	9.70	14.18

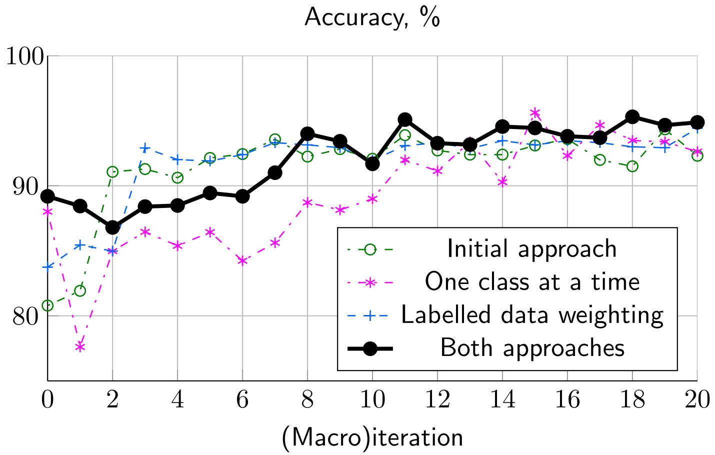


Figure 5: Comparison of the modifications to the *iterative* algorithm with their combination and the initial version, smaller instrument set.

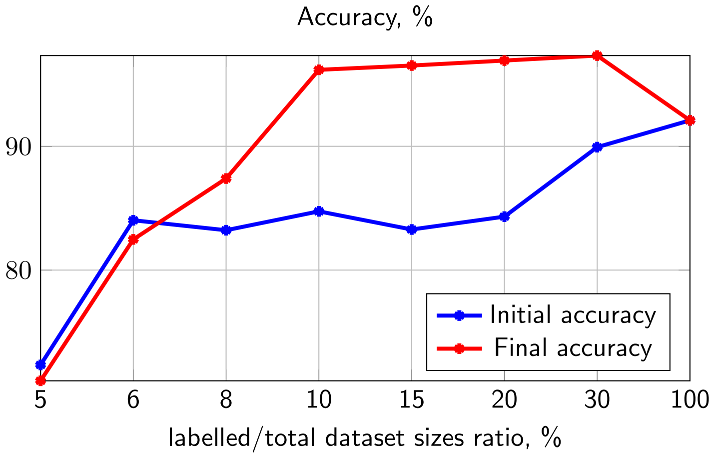


Figure 6: Comparison of the accuracies of the initial models and the final iteration of the *incremental* EM-based SSL algorithm as a function of the relative labelled dataset size with the *smaller* instrument set.

Conclusions

- SSL for instrument recognition works. Gain 9.7%;
- as little as 7% data needs to be annotated;
- proposed extensions simplify termination and increase accuracy.

Future work:

- more complex scenarios (more instruments, noise, reverberation. . .);
- neighbouring problems.