Semi-supervised Musical Instrument Recognition

Master’s Thesis Presentation

Aleksandr Diment

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

Musical instrument recognition
Semi-supervised learning
Objectives and main results

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 ⇒
- 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

1 Introduction
2 System description
   Building blocks
      Feature extraction
      Training algorithms
      Labelled data weighting
      One-class-at-a-time training
3 Evaluation
4 Conclusions

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

Input signal

Pre-emphasis

Frame blocking and windowing

$|\text{FFT}|^2$

Log

DCT

$\frac{d}{dt}$

Static coefficients

Delta coefficients

Figure 2: A block diagram of MFCCs calculation.
System description

Training algorithms

The EM algorithm ⇒ 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 algorithms¹.

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.

⇒ de-weight the contribution of the unlabelled data.

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

⇒ previous models of one class are unaffected by training another one

⇒ fewer peaks

⇒ safer to chose an arbitrary iteration index for termination.
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.

---

## Evaluation

### Datasets

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

<table>
<thead>
<tr>
<th>Instrument</th>
<th># notes</th>
<th>Instrument</th>
<th># notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic Guitar</td>
<td>702</td>
<td>Pianoforte</td>
<td>792</td>
</tr>
<tr>
<td>Electric Guitar</td>
<td>702</td>
<td>Classic Guitar</td>
<td>702</td>
</tr>
<tr>
<td>Tuba</td>
<td>270</td>
<td>Electric Guitar</td>
<td>702</td>
</tr>
<tr>
<td>Bassoon</td>
<td>360</td>
<td>Electric Bass</td>
<td>507</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trombone</td>
<td>278</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tuba</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Horn</td>
<td>288</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bassoon</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clarinet</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Banjo</td>
<td>941</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2 212</strong></td>
<td><strong>Total</strong></td>
<td><strong>5 200</strong></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Instrument set size</th>
<th>Recognition accuracy, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 instruments</td>
<td>92.1</td>
</tr>
<tr>
<td>10 instruments</td>
<td>82.9</td>
</tr>
</tbody>
</table>
Evaluation

Results with the iterative EM-based algorithm

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

<table>
<thead>
<tr>
<th>Instrument set size</th>
<th>Recognition accuracy, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>initial</td>
</tr>
<tr>
<td>4 instruments</td>
<td>89.19</td>
</tr>
<tr>
<td>10 instruments</td>
<td>58.73</td>
</tr>
</tbody>
</table>
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.
1 Introduction

2 System description

3 Evaluation

4 Conclusions
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.