Semi-supervised Learning for Musical Instrument Recognition

Aleksandr Diment, Toni Heittola, Tuomas Virtanen

Department of Signal Processing, Tampere University of Technology, Finland

Introduction

Motivation

- Supervised learning needs annotated data (costly).
- Semi-supervised learning (SSL) enables utilising additional unannotated data (easier to obtain).

Objectives

- To show whether SSL is capable of introducing improvement in the performance of an instrument recogniser.
- SSL is studied on an example of the iterative EM-based algorithm.
- Extensions for a smoother model transition are proposed.

Methodology

Features

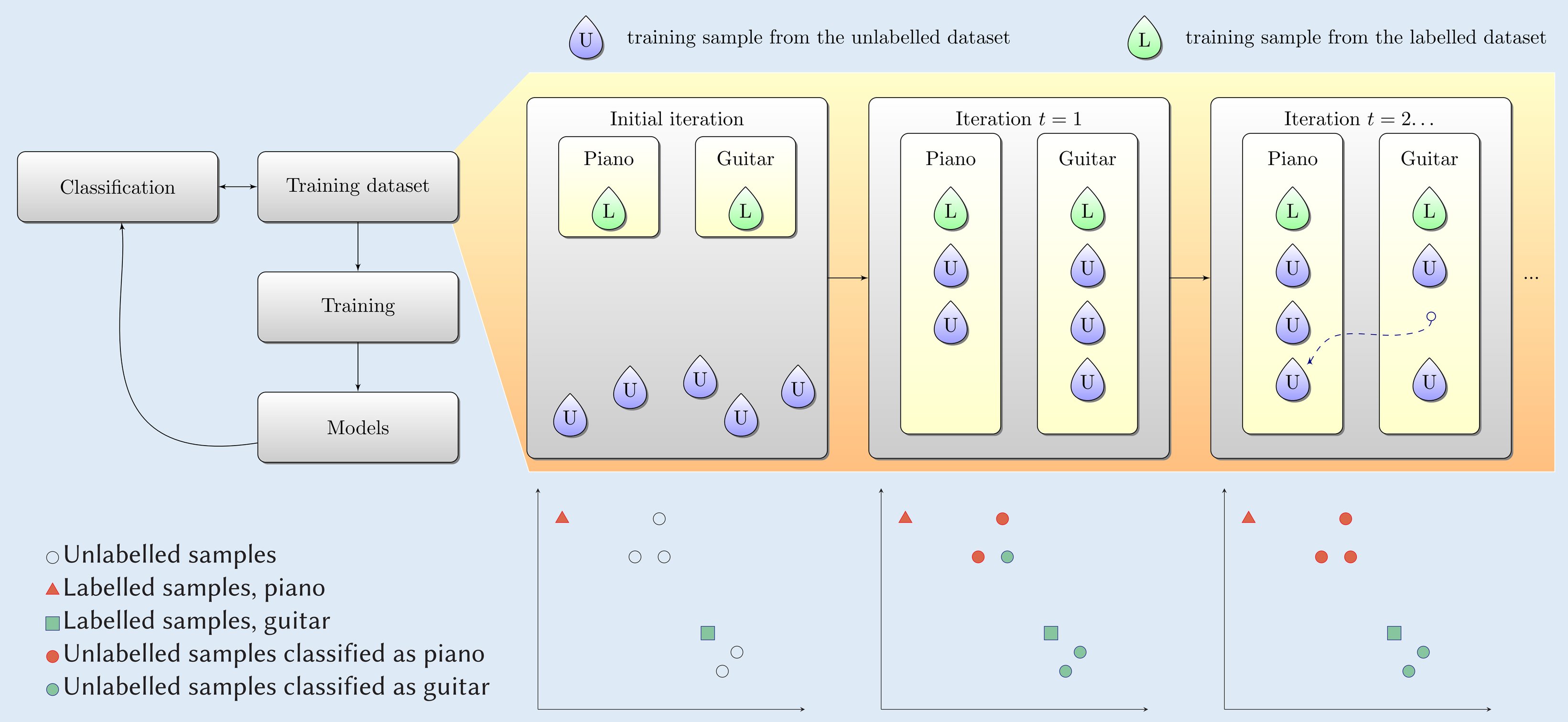
Static and delta MFCCs (mel-frequency cepstral coefficients).

Supervised training

The GMMs are obtained based on labelled data (EM algorithm).

Semi-supervised training

- Iterative EM-based algorithm (Moreno *et al.*, 2003).
- Incorporating unlabelled data: labels are predicted, and together with the labelled data it is used to re-estimate model parameters.
- Prediction and re-estimation are repeated iteratively.



Proposed extensions

Class-wise retraining

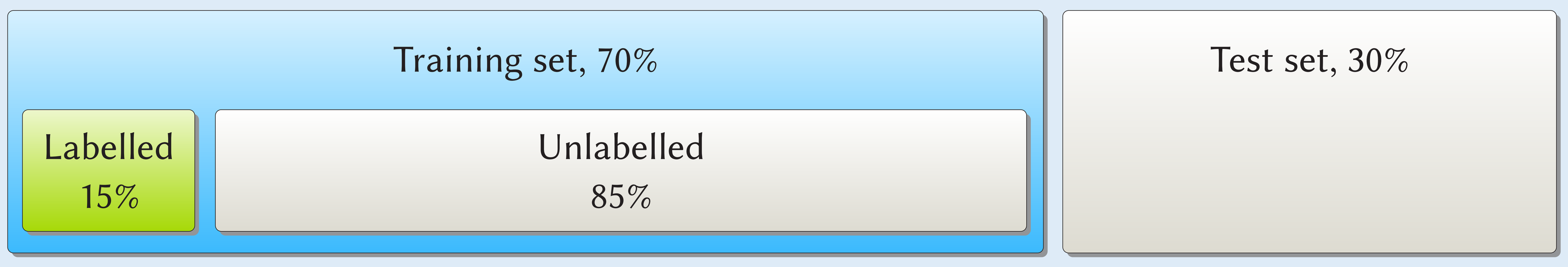
- Coupled model degradation effect: an erroneous re-classification degrades both the true and erroneous classes' models.
- Proposal: retrain models of only one class per iteration.
- Smoother transition between the models.
- Fewer local peaks in the accuracy curve. Easier convergence.

Labelled data weighting

- Insignificant change when increasing amount of unlabelled data.
- Solution: to de-weight the impact of the unlabelled data by replicating the labelled data several times. The replication factor is reduced along the iterations.

Acoustic material

- Separate monophonic note recordings.
- Nine instruments from the RWC Music DB (with # notes):
  - Piano (792), Classic Guitar (702), Electric Guitar (702), Electric Bass (507), Trombone (278), Tuba (270), Bassoon (360), Clarinet (360) and Banjo (941).

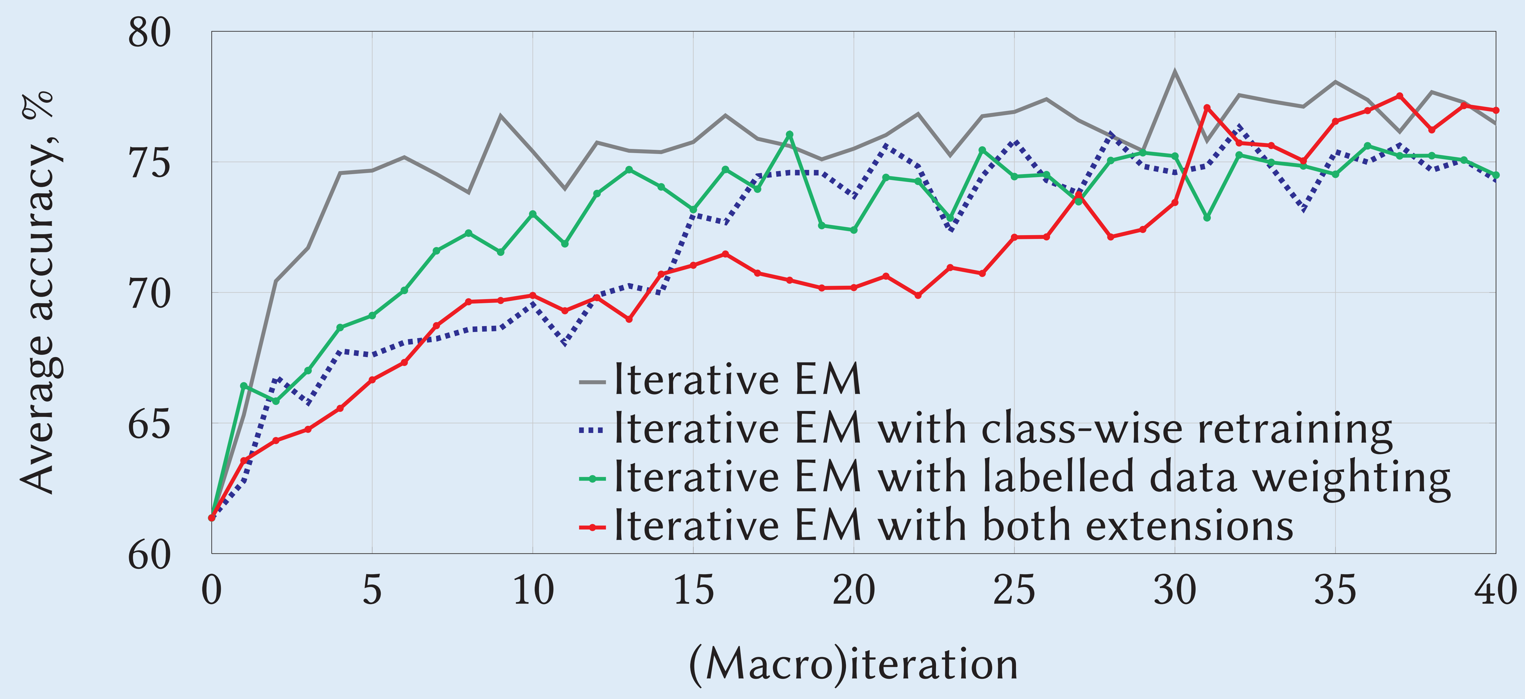


Results

Evaluation scenarios:

- A fully supervised case: all the data is labelled. Upper limit for the SSL performance (as if SSL with all the labels estimated correctly).
- The iterative EM-based algorithm, extensions, combination.

Test case	Recognition accuracy, %	
Supervised, 100% of data labelled	83.8	
Semi-supervised, 15% of data labelled	**initial**	**final**
iterative EM	61.4	76.5
iterative EM with class-wise retraining	61.4	74.3
iterative EM with labelled data weighting	61.4	75.1
iterative EM with both extensions	**61.4**	**77.0**



- Similar improvement of 12-16% with all algorithms.
- The basic algorithm: most oscillating, but reaches max earlier.
- The extensions (especially the class-wise retraining): smoother transition between the models.

Conclusions

- The applicability of SSL for instrument recognition explored.
- The EM-based SSL algorithm + two proposed extensions for a smoother transition between the models implemented.
- Evaluation with only 15% data labelled: up to 16% improvement.
- Future: A more sophisticated feature extraction, more instruments, added noise, reverberation and distortions.

http://www.cs.tut.fi/~diment/

aleksandr.diment@tut.fi