Multi-label Learning Approaches for Music Instrument Recognition

Eleftherios Spyromitros - Xioufis, Grigorios Tsoumakas and Ioannis Vlahavas

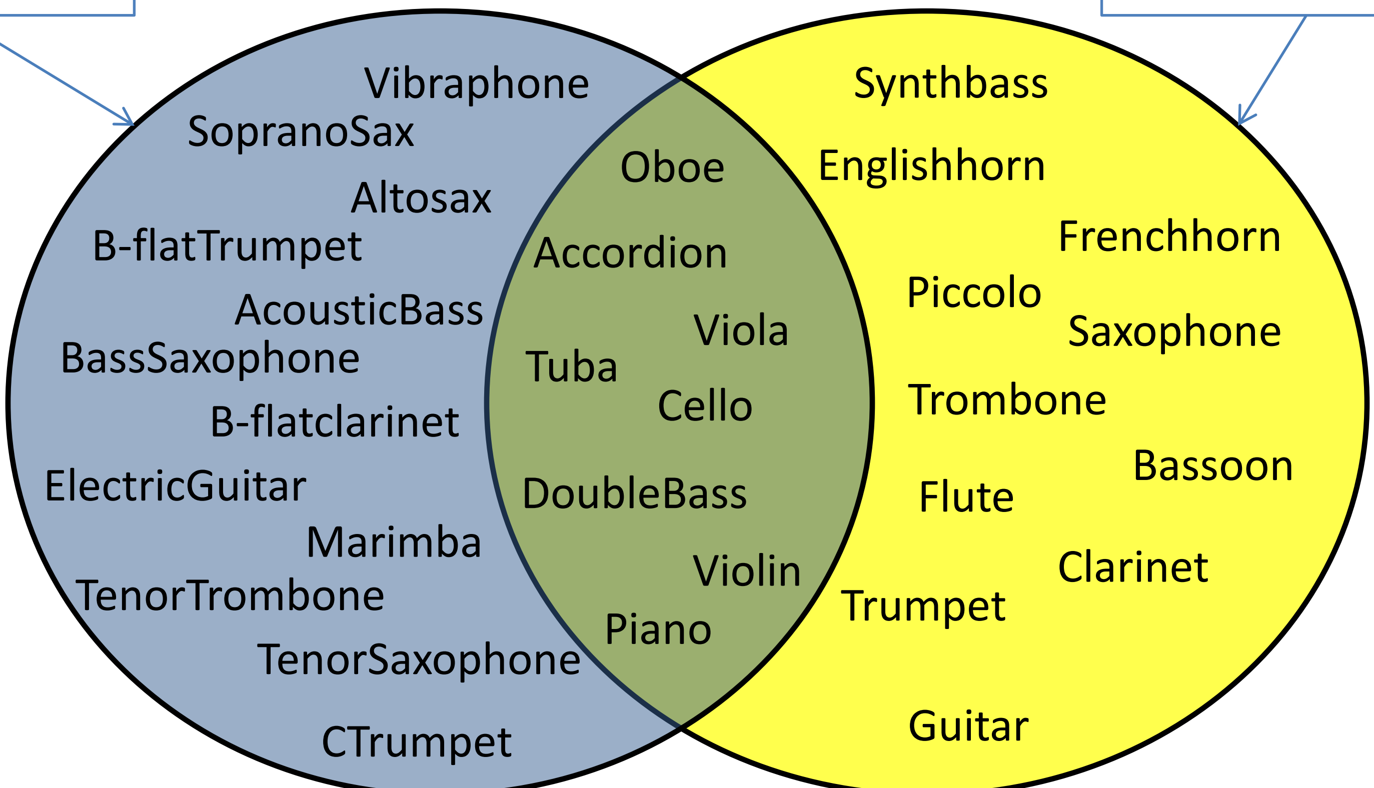
Machine Learning & Knowledge Discovery Group
Department of Informatics
Aristotle University of Thessaloniki
Greece

The training sets

Pairs
5422 recordings
21 instruments

Only 8 in common

Single Instruments
114914 recordings
19 Instruments



Additional complexity

- Relations between instruments of the two datasets --> complexity:
  - Examples of the specialized class could be considered as examples of the general class
    - C-Trumpet and B-FlatTrumpet are kinds of Trumpet
    - TenorTrombone is a kind of Trombone
  - Difficult to distinguish different kinds of the same instrument
    - soprano or alto saxophone?
- The following statements brought additional complexity:
  - The pairs of the training set **do not** occur in the test set
  - **Not all** 32 instruments of the training data ~~must~~ appear in the test data
  - Some instruments of the test set ~~may~~ appear **only** in single instruments data

The trick

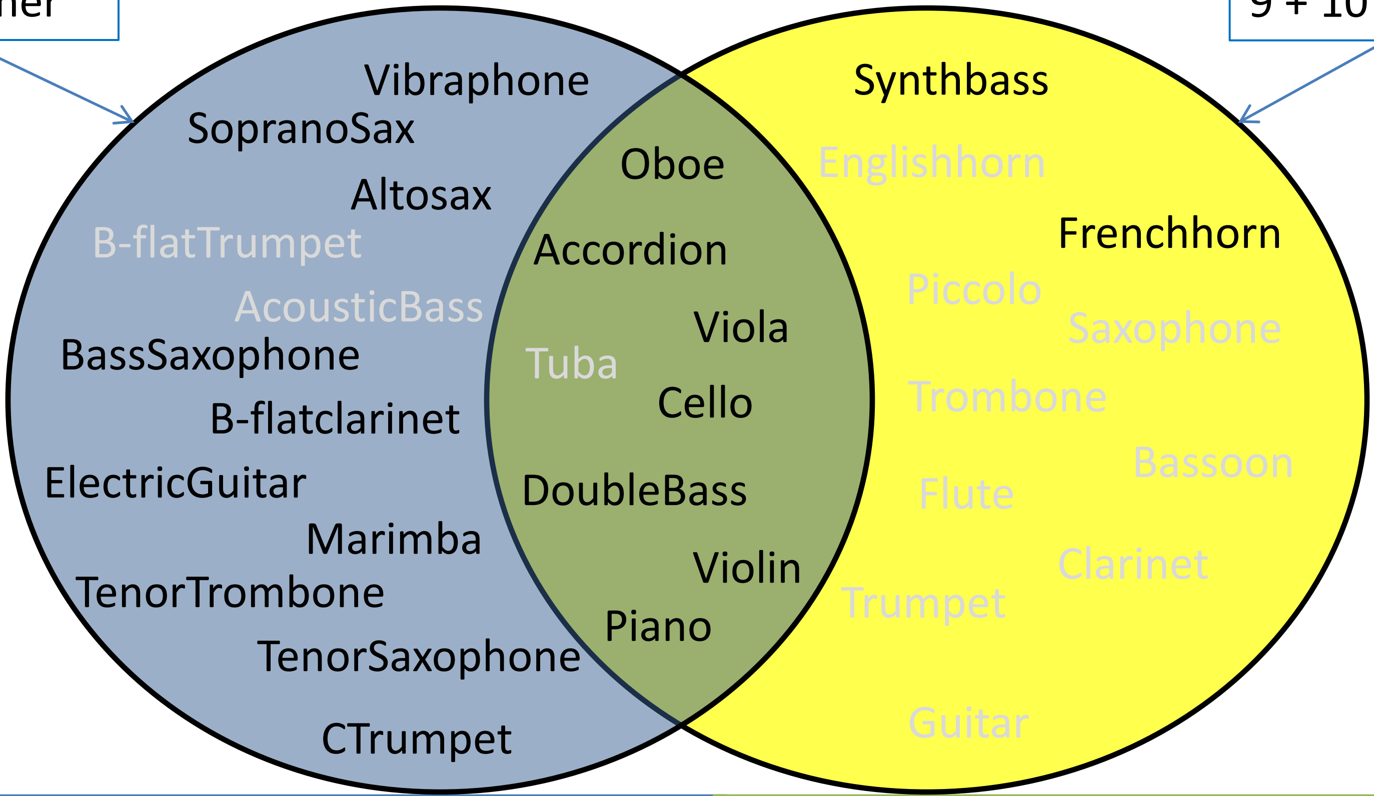
- Lets make things more clear!
- The evaluation system allowed a trick:
  - 32 ‘dummy’ predictions containing the same instrument for every test instance were sent
  - The resulting accuracy represented the percentage of each instrument in the validation set (35% of the test set)
  - This allowed a very close approximation of the label distribution in the full test set
- Findings

Findings about the test set

20 out of the 32 instruments appear in the test set

Pairs set
18 + 3 other

Pairs set
9 + 10 other



Trying different inputs

- Which is the best input?
  - Using only the pairs dataset
  - Using only the single-instruments dataset
  - The union of the datasets (pairs + single-instrument examples)
- The results (of a comparison using various learning methods)
  - **Only pairs** better than **only single instruments** (expected)
    - many instruments of the test set do not appear in the single-instruments set
  - **Only pairs** better than **the union** (**unexpected**)
    - examples for all instruments are there

The final training set

- Further experiments revealed that:
  - Using only examples of pairs is better than combining them with single instrument examples.
  - Using single instrument examples is beneficial only if pair examples are not available.
- The final set used to train the winning method:
  - All the 5422 example pairs
  - The 340 single-instrument examples of SynthBass and Frenchhorn
  - All the given feature attributes (except for the 5 additional attributes of the single-instruments set)

A multi-label classification problem

- Single-label classification:
  - One categorical target variable
- Multi-label classification:
  - Multiple target variables (with possible associations between them)
- Recognition of instrument pairs:
  - A special multi-label case
  - Each example is associated with exactly 2 labels
- Two families of multi-label methods:
  - **Problem transformation**
  - Algorithm adaptation

Preliminary experiments

- Various multi-label methods of the problem transformation family
  - state-of-the-art: ECC [Read et al., ECML 09], RAKEL [Tsoumakas et al., TKDE 11]
  - baseline: Binary Relevance (BR), Label Powerset (LP)
- Coupled with various base classifiers
  - SVMs, Decision Trees, etc.
- BR was found competitive
  - especially when coupled with strong base classifiers

Binary Relevance (BR)

- How it works
  - Learns one binary classifier for each label
- Trained on transformed training sets
  - The examples having λ are positive
  - All the rest are negative
- Limitations
  1. Does not consider label correlations
  2. Leads to class imbalance
- In our case
  - Limitation 1 is not important (different correlations appear in the test set)
  - Focus on limitation 2

Ex#	Label set
1	{λ1,λ4}
2	{λ3,λ4}
3	{λ2}
4	{λ2,λ1}

Ex#	λ1
1	+
2	-
3	-
4	+

Ex#	λ2
1	-
2	-
3	+
4	+

Tuning the base classifier

- Random Forest (RF) was used as a base classifier
- How to deal with class imbalance?
  - Combine RF with Asymmetric Bagging [Tao et al, TPAMI06]
- Asymmetric Bagging Random Forest (ABRF):
  1. Take a bootstrap sample **only from the negative examples**
  2. Use the negative sample **+ all the positive examples** and train a RF
  3. Repeat the above steps n times and aggregate the decisions of all the generated random trees
- The best performance
  - 10 forests (of 10 random trees each) trained on 10 balanced training sets

Typical ranking approach

- Output of an ABRF classifier for each label:
  - A confidence score for the label being true
  - Equal to: $\frac{\#\ trees\ voting\ yes}{\#\ total\ trees}$
  - e.g.

Viola	Piano	Cello	Violin
0.34	0.67	0.22	0.56

- Focus
  - Produce an accurate ranking
  - Pick the 2 top-ranked instruments
- Typical approach
  - Use the confidence scores to produce a ranking
  - e.g.

1^{st}	2^{nd}	3^{rd}	4^{th}
Piano	Violin	Viola	Cello

An alternative approach

- Use the trained classifiers to generate confidence scores for all test instances
- For each test instance:
  - Find how the confidence score assigned to each label is ranked in the list of confidence scores given for that label
  - Output the 2 labels with the lowest ranks
  - Instance 1 would take the labels {Cello, Piano} or {Cello, Viola}

Test set
Inst. #1
Inst. #2
Inst. #3

ABRS classifiers

# inst	Viola	Piano	Cello
1	0.45	0.33	0.77
2	0.97	0.50	0.21
3	0.44	0.11	0.62

# inst	Viola	Piano	Cello
1	2nd	2nd	1st
2	1st	1st	3rd
3	3rd	3rd	2nd

Taking the priors into account

- However
  - Label priors was used to approximate the # examples per label in test set
  - Being 2^{nd} out of 3 is better than 2^{nd} out of 1
  - Output the 2 labels with the lowest “normalized” ranks

Labels	Priors	#
Viola	0.33	1
Cello	1.00	3
Piano	0.66	2

# inst	Viola	Piano	Cello
1	2^{nd}	2^{nd}	1^{st}
2	1^{st}	1^{st}	3^{rd}
3	3^{rd}	3^{rd}	2^{nd}

# inst	Viola	Piano	Cello
1	2/1	**2/2**	**1/3**
2	**1/1**	**1/2**	3/3
3	3/1	**3/2**	**2/3**

Post-processing filter

- Avoid outputting instrument pairs of the training set:
  - substitute the second-ranked instrument
- Assumption:
  - the classifier is more confident for the first-ranked instrument

Conclusions

- Motivation
  - Explore the potential of multi-label learning methods
- Conclusions
  - Baseline is sometimes better than state-of-the-art
  - Pairs of instruments are better recognized using pair examples
- Future
  - Generalization to an arbitrary number of instruments playing together
- Software
  - **Mulan** http://mulan.sourceforge.net
  - **Weka** http://www.cs.waikato.ac.nz/ml/weka/

Acknowledgements

- Acknowledgements
  - To my teacher and friend Grigorios Tsoumakas for the fair play!

I was 1st only for a while!

I was 2nd only for a while!

Rank	Team	Time of Submission	Preliminary Result	Final Result
1	– Eleftherios Spyromitros Xioufis (lefman), *Aristotle University of Thessaloniki*	Mar 21, 22:33:11	0.7239	0.72273
2	– MOZ Grigorios Tsoumakas (tsoumakas)	Mar 12, 16:42:28	0.7106	0.71133

THANK YOU!

QUESTIONS?