A Cartesian Ensemble of Feature Subspace Classifiers for Music Categorization

Thomas Lidy Rudolf Mayer Andreas Rauber[1]
Pedro J. Ponce de León Antonio Pertusa Jose M. Iñesta[2]

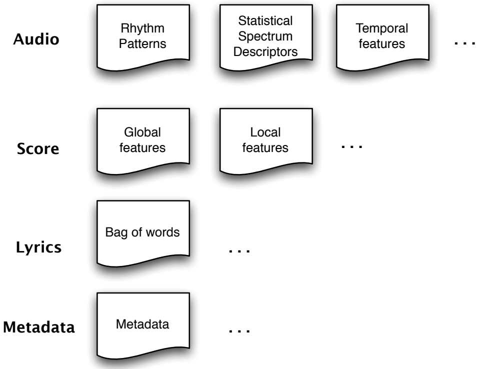
[1] *Information & Software Engineering Group (IFS)*
Department of Software Technology and Interactive Systems
Vienna University of Technology, Austria
http://www.ifs.tuwien.ac.at/mir

[2] Universitat d'Alacant
Universidad de Alicante

Pattern Recognition and Artificial Intelligence Group
Department of Software and Computing Systems
University of Alicante, Spain
http://grfia.dlsi.ua.es/cm

ISMIR Conference, 2010

Motivation



- Given a tagged corpus, several feature sets from different 'modalities' are available (e.g., audio, symbolic, lyrics,...)
- Improve classification through combination of feature sets/classification schemes
- Release the user from explicitly choosing 'the best' single feature set/classifier combination.

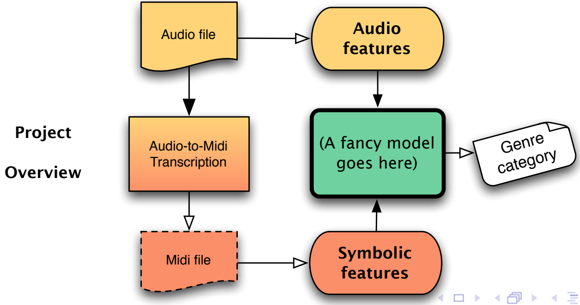
Motivation

Funding: Bilateral (Spain-Austria) R&D programm

Project:

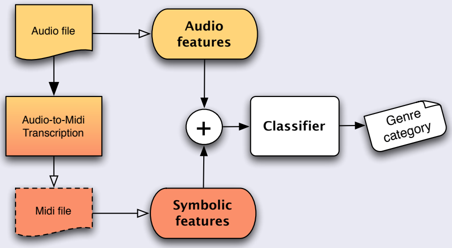
- Music genre classification by combining audio and symbolic descriptors through an automatic transcription system.

Period: January 2008 - July 2010



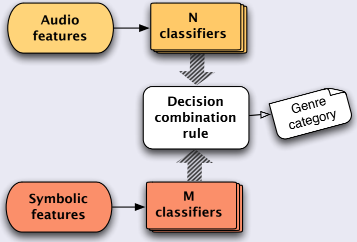
Music categorization by combined approaches

Early fusion: Audio and symbolic feature subspace concatenation



ISMIR 2007
MIREX 2007
MIREX 2008

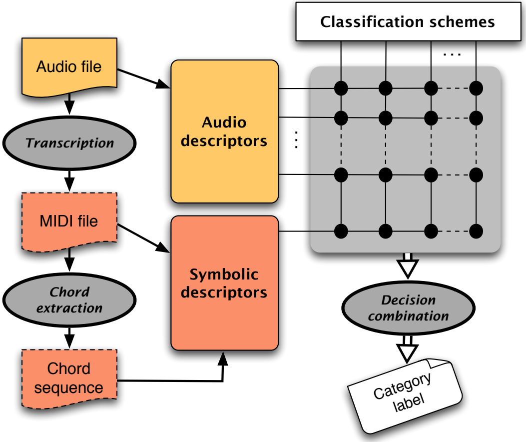
Late fusion: model outcomes combination



ISMIR 2010

- Base models can come from different machine learning paradigms.
- Key factor: The more *diverse and accurate* the ensemble of classifiers, the more improvement is expected.
- Ensemble diversity: How varied model opinions are.
- A wide range of decision combination rules exists.

Late fusion: the *Cartesian Ensemble*



- **D** feature subspaces,
- **C** classification schemes, then
- **DxC** models to combine
- Build on top of the Weka[a] data mining toolkit.

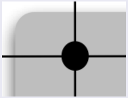
[a] M. Hall, et al.(2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1. www.cs.waikato.ac.nz/ml/weka/

Input section

- Feature sets in Weka and SomLIB format currently supported.
- Feature subspaces aligned through a common ID attribute.
- Labeled samples are mandatory only in first subspace.

Model training

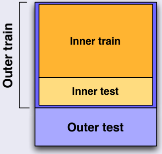
Model training



(single model)

- Each model built using a given classification scheme and feature subspace
- All possible feature subspace/scheme models are built

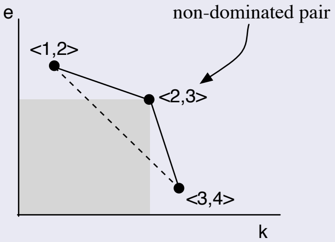
Model accuracy estimation



- Model accuracy estimated through *inner crossvalidation*.
- Needed for model selection and weighted decision combination rules.

Model selection

Pareto-optimal classifier selection



[Remember:] *The more* <u>*diverse*</u> *and* <u>*accurate*</u> *the ensemble, the more improvement is expected.*

- Selects pairs of models based on <u>accuracy</u> and <u>diversity</u> metrics.
- All *non-dominated by all criteria* pairs are selected.

Given <i,j>, κ_{ij} is the *inter-rater agreement*, e_{ij} is *pair average error rate*.

$$\kappa_{ij} = \frac{\sum_k m_{kk} - \mathrm{ABC}}{1 - \mathrm{ABC}}$$

$$\mathrm{ABC} = \sum_r (\sum_s m_{r,s})(\sum_s m_{s,r})$$

$$e_{ij} = 1 - \frac{\alpha_i + \alpha_j}{2}$$

Late fusion strategies: combining model outcomes

Unweighted combination

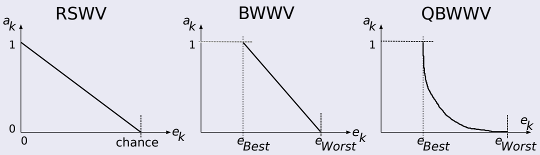
MAJ	Majority vote rule
AVG	Average of p.p.
MAX	Maximum of p.p.
MED	Median of p.p.

(p.p.: posterior probability)

Weighted majority vote rules

SWV	Simple Weighted
RSWV	Rescaled Simple Weighted
BWWV	Best-Worst Weighted
QBWWV	Quadratic Best-Worst Weighted
WMV	Weighted Majority

Model weight: based on model estimated accuracy



Evaluation: Music categorization

Corpora

Dataset	Files	Genres	File length
9GDB	856	9	full
GTZAN	1000	10	30 sec
ISMIRgenre	1458	6	full
ISMIRrhythm	698	8	30 sec
LatinMusic	3225	10	full
Africa-function	1024	27	full
Africa-instrument	1024	11	full
Africa-country	1024	11	full
Africa-ethnic	1024	40	full

Feature subspaces

Audio features

Feature subspace	no. feats.
Rhythm Pattern (RP)	1440
Rhythm Histogram (RH)	60
Statistical Spectrum Descriptor (SSD)	168
Modulation Variance Descriptor (MVD)	420
Temporal RH (TRH)	420
Temporal SSD (TSSD)	1176

Symbolic features

Feature subspace	no. feats.
Global features	52
Chord Relative Frequency	9

(Chord extraction algorithm: [Pardo & Birmingham, 2002])

Evaluation

Outer c.v. 10 folds

Inner c.v. 3 folds

Classification schemes (10)

Scheme	Paradigm
Naïve Bayes (NB)	Bayes rule
Nearest Neighbor (1-NN)	lazy learner
3-NN, Manhattan dist.	lazy learner
RIPPER	rule learner
C4.5	decision tree
REPTree	decision tree
Random Forest (RF)	decision tree ensemble
SVM, linear kernel (SVM-lin)	statistical learning theory
SVM, quadratic kernel (SVM-quad)	"
SVM, Puk kernel (SVM-Puk)	"

8 feature subspaces $\times$ 10 schemes $=$ 80 models

Ensemble vs. single best model results

Ensemble vs. single best model accuracy (in %)

Corpus	*Single best*	*Ensemble*	*Comb. rule*
9GDB	78.15 (2.25)	**81.66** (3.96)	AVG
GTZAN	72.60 (3.92)	**77.50** (4.30)	QBWWV
ISMIRgenre	81.28 (3.13)	**84.02** (1.50)	QBWWV
ISMIRrhythm	87.97 (4.28)	**89.11** (4.62)	BWWV
LatinMusic	89.46 (1.62)	**92.71** (0.99)	QBWWV
Africa-country	86.29 (2.30)	**89.03** (1.63)	QBWWV
Africa-ethnic	81.10 (2.41)	**82.97** (3.30)	WMV
Africa-function	51.06 (6.63)	**54.84** (6.29)	QBWWV
Africa-instrument	69.90 (4.69)	**73.00** (4.25)	WMV

Extending feature subspaces: segmenting the input

- Segment each audio file into 3 equal-sized segments.
- $6 \times 3 = 18$ audio subspaces
- Symbolic features were not segmented.
- Results inferior than using full song features.

Ensemble cross-validation execution times

Corpus	files	train (sec.)	test (sec.)
9GDB	856	6645	140
GTZAN	1000	10702	345
ISMIRgenre	1458	12510	275
ISMIRrhythm	698	5466	185

Test times are averaged over decision combination methods.
Roughly, 10 sec. per sample on a Quad machine (e.g., 3 hours for GTZAN)

Conclusions

- A generic ensemble framework based on feature subspaces was devised.
- The ensemble improves classification accuracy over best single model.
- The user is released from having to choose a particular feature subspace/classifier.
- Relying on the QBWWV decision combination rule seems feasible.

Further work

- Reduce training times by feature selection. Preliminary results presented at MML 2010.
- Add other input modalities: Lyric features, metadata, symbolic features by statistical language modeling techniques...

Thanks!

A Cartesian Ensemble of Feature Subspace Classifiers for Music Categorization

Thomas Lidy, Rudolf Mayer, Andreas Rauber
Pedro J. Ponce de León, Antonio Pertusa, Jose M. Iñesta

Information & Software Engineering Group (IFS)
Department of Software Technology and Interactive Systems
Vienna University of Technology, Austria
http://www.ifs.tuwien.ac.at/mir

Pattern Recognition and Artificial Intelligence Group
Department of Software and Computing Systems
University of Alicante, Spain
http://grfia.dlsi.ua.es/cm