Talk

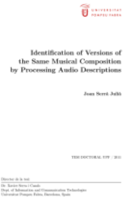
Machine Learning for Music Discovery

Joan Serrà

Artificial Intelligence Research Institute (IIIA-CSIC)
Spanish National Research Council
jserra@iiia.csic.es
http://joanserra.weebly.com

Joan Serrà

- 2007- 2011 Ph.D.
  Music Technology Group
  Universitat Pompeu Fabra
  Barcelona, Spain
- Postdoctoral Resarcher
  Artificial Intelligence Research Institute
  Spanish National Research Council
- Collaborator in the CompMusic project
- Music information retrieval, machine learning, time series analysis, complex networks, ...



Machine Learning for Music Discovery

Focus:

- Learning labels (categories) from music audio data (music signal): Automatic annotation
- Music retrieval: Query-by-example paradigm

Learning labels + Query by example = Category-based retrieval .

- Issues when applying machine learning to music-related tasks

Overview

I. Category-based retrieval (basic concepts)
II. Categorizing and annotating (supervised approaches)
III. Typical approach to supervised music annotation
  I. Feature extraction
  II. Data processing & feature selection
  III. Classification & classifiers
IV. Basic but important issues in classification methodology

Category-based Retrieval

Goal: Find music recordings within a collection that share a specific, predefined characteristic or trait.

Humans naturally group objects with similar characteristics or traits into categories (or classes).

Category: Any word or concept that can be used to describe a set of recordings sharing some *property*. Also referred to as *class*, *type* or *kind*.

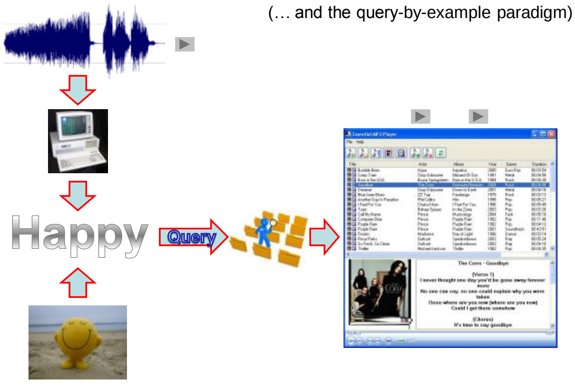
Category-based Retrieval

- Categories are one of the most basic ways to represent and structure human knowledge
- They encapsulate characteristic, intrinsic features of a group of items
- Prototypes can be built and potential category members can be compared to these
- They can overlap and have some kind of “organization” (e.g. fuzzy and hierarchical)

[Zbikowski, *Oxford Univ. Press*, 2007]

Category-based Retrieval

(… and the query-by-example paradigm)



Category-based Retrieval

Motivation

- Search and organization of music collections
- Music discovery; discovery of *related* music documents
- Playlist generation
- Music recommendation
- Music similarity
- …

[Tversky, *Psych. Revs.*, 1977; Bogdanov et al., *IEEE-TMM*, 2011]

Category-based Retrieval

Category-based query examples:

- Genres (e.g. “bossa-nova”)
- Moods (e.g. “sad”)
- Specific instruments (e.g. “extremely distorted electric guitar”)
- Instrument ensembles (e.g. “string quartet”)
- Key/mode (e.g. “G minor pieces”)
- Dynamics (e.g. “lots of loudness changes”)
- Type of recording (e.g. “live recording”)
- Epoch (e.g. “music from the 80s”)
- …

Category-based Retrieval: Annotations

Objective: Label/annotate music documents (ideally with some confidence and probability measurements) in a way that makes sense, e.g. to a user (or to a user community).

Category-based retrieval → Annotating music excerpts → Using/creating a taxonomy

Categories usually imply a taxonomy:

- Based on the opinion of experts, on the “wisdom of the crowd” (folksonomy), on a single user (personomy), …
- Taxonomies can be organized, their members can overlap, different relation types can be applied (ontologies)…

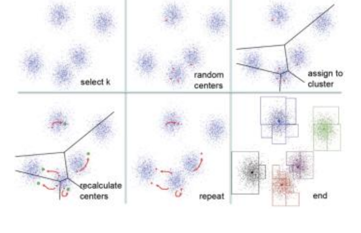
Categorizing/Annontating

Unsupervised vs. Supervised

Categorizing/Annontating

Unsupervised:

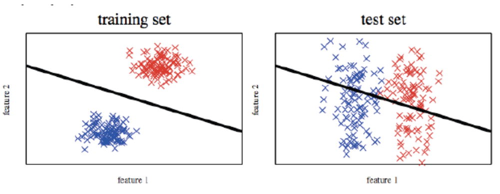
- No taxonomy. Categories emerge from the data organization under a predefined similarity measure
- Typical clustering algorithms: K-means, hierarchical clustering, SOM, growing hierarchical SOM, …



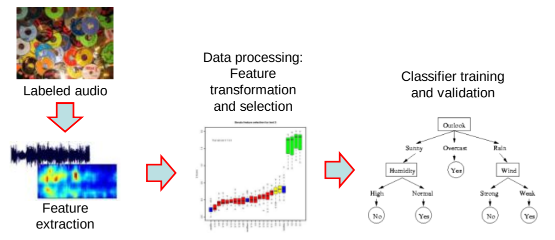
Categorizing/Annontating

Supervised:

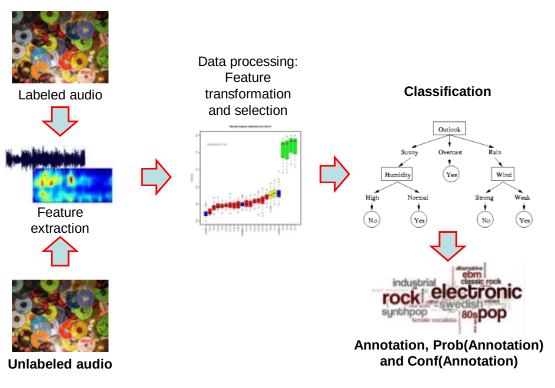
- Use a taxonomy
- Map features to categories (usually without describing the rules of this mapping explicitly)
- Typical supervised learning algorithms: KNN, neural networks, LDA, SVM, trees, …



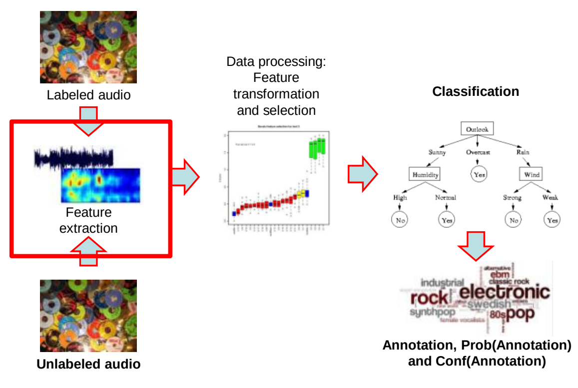
Typical Approach: Automatic Annotation



Typical Approach: Automatic Annotation



Typical Approach: Automatic Annotation



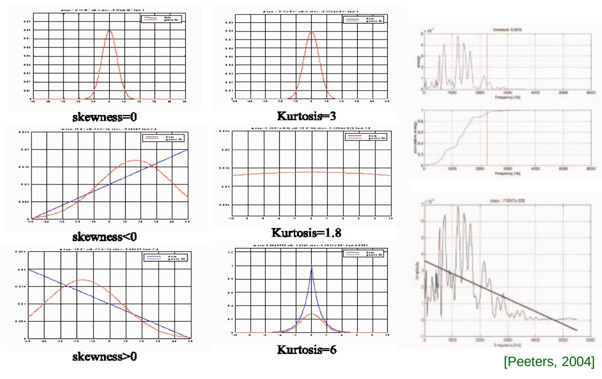
Typical Approach: Automatic Annotation

Feature extraction

- Temporal: ZCR, energy, loudness, rise time, onset asyncrony, amplitude modulation, …
- Spectral: MFCC, LPC, spectral centroid, spectral bandwidth, bark-band energies, …
- Tonal: Chroma/PCP, melody, chords, key, …
- Rhythmic: Beat histogram, IOIs, rhythm transform, FP, …

Features extracted on a short-time frame-by-frame basis

Typical Approach: Automatic Annotation

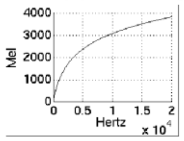


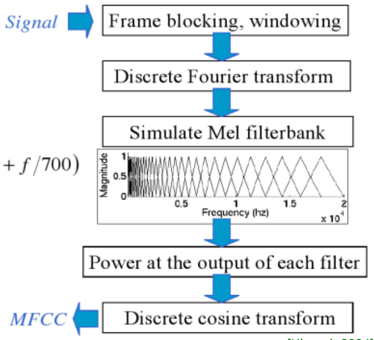
[Peeters, 2004]

Typical Approach: Automatic Annotation
Mel-frequency cepstral coefficients

- Improves over the standard cepstral coefficients
- **Figure**: calculations in practice
- Mel frequency scale:

$Mel(f) = 2595 \log_{10}(1 + f/700)$

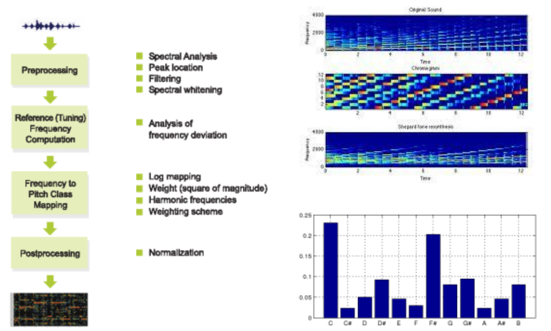




[Klapuri, 2004]

Typical Approach: Automatic Annotation

Feature extraction (PCP/Chroma)



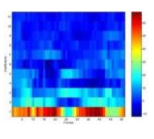
Typical Approach: Automatic Annotation

Feature extraction

But usually summarized to represent global characteristics of the whole audio excerpt:

- Component-wise means
- Component-wise variances
- Covariances
- Higher-order statistical moments (skewness, kurtosis, …)
- Derivatives (1st, 2nd, …)
- Max, min, percentiles, histograms, ...

(+ more *exotic* summarizations)



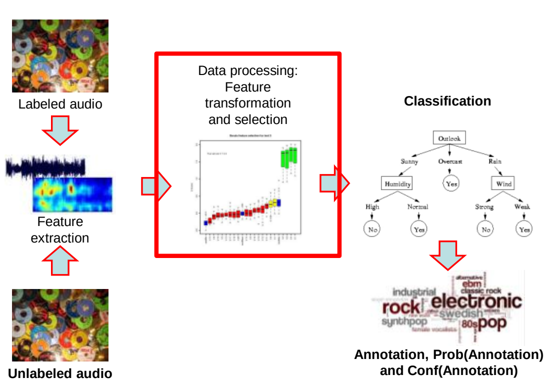
Typical Approach: Automatic Annotation

Feature extraction

Features and their aggregation/summarization depend on the ***task***!

(e.g. MFCCs for raaga recognition??)

Typical Approach: Automatic Annotation



Typical Approach: Automatic Annotation

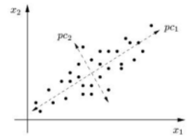
Data processing

Feature transformation and feature selection

- Rescale each feature so that all of them have the same “influence” (normalize, standardize, … Gaussianize?)
- Detect highly correlated features (try to decorrelate them?)
- Remove uninformative features
- Reduce the dimensionality
- Learn about the features and the task!

Typical Approach: Automatic Annotation

Data processing: Feature transformation



Principal Component Analysis (PCA)

- Linear transform
- Eliminates redundancy
- Decorrelates variables
- It can be used to reduce the number of features

- Assumes some "homogeneity" of the distributions
- We lose the information of which specific feature is relevant

Typical Approach: Automatic Annotation

Data processing: Feature selection

$$r_{jii'} = \frac{(\mu_{ji}-\mu_{ji'})^2}{\sigma_{ji}^2+\sigma_{ji'}^2}$$

Fisher's ratio

- Ratio between inter- and intra-class scatter
- Measures discriminative power of feature *j* between classes *i* and *i'*
- Ratio low when classes are well separated

- No standard way of performing this feature selection iteratively. Two of the most basic ones:
  - Sequential backward selection (removing bad features)
  - Sequential forward selection (selecting good features)

Typical Approach: Automatic Annotation

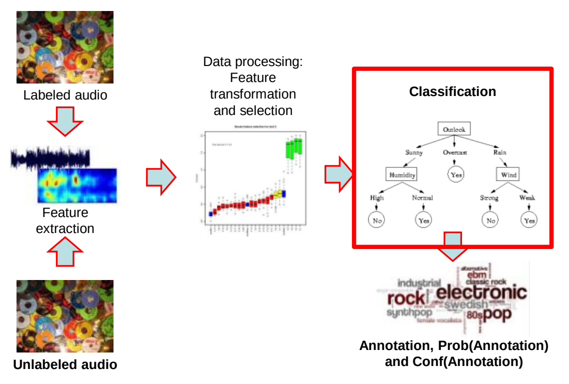
Data processing: Feature selection

Many other feature selection algorithms:

- Correlation-based
- Chi-squared
- Consistency
- Gain ratio
- Information gain
- Bootstrapping
- Classifier-affine feature selection
- …

[Witten & Frank, *Elsevier*, 2005]

Typical Approach: Automatic Annotation



Typical Approach: Automatic Annotation

Classifiers

Bayesian/probabilistic Classifiers

$$p(C|F_1,\dots,F_n) = \frac{p(C)\ p(F_1,\dots,F_n|C)}{p(F_1,\dots,F_n)}.$$

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}.$$

Naïve conditional independence

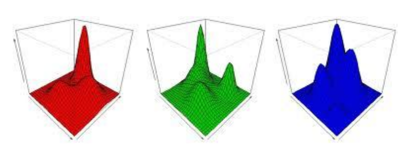
$$p(C|F_1,\dots,F_n) = \frac{1}{Z}p(C)\prod_{i=1}^{n} p(F_i|C)$$

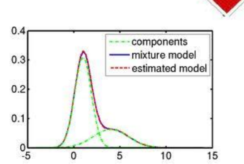
[Witten & Frank, *Elsevier*, 2005]

Typical Approach: Automatic Annotation

Classifiers

Gaussian Mixtures





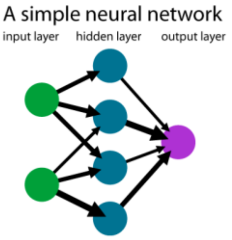
$$P(x=v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}}e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

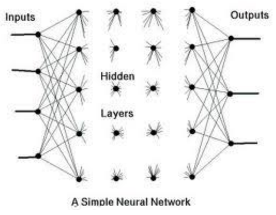
[Witten & Frank, *Elsevier*, 2005]

Typical Approach: Automatic Annotation

Classifiers

Neural Networks

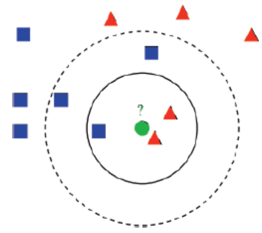




Typical Approach: Automatic Annotation

Classifiers

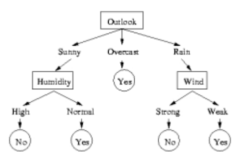
K Nearest Neighbors Classifier

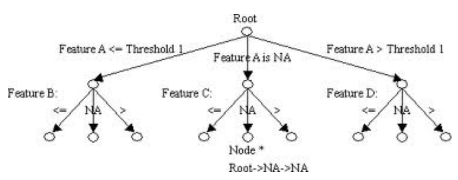


Typical Approach: Automatic Annotation

Classifiers

Trees

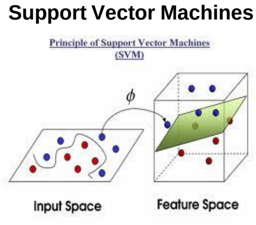


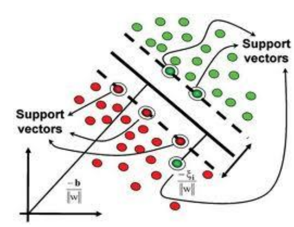


Typical Approach: Automatic Annotation

Classifiers

Support Vector Machines





Typical Approach: Automatic Annotation

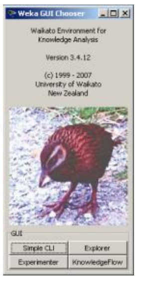
Classifiers

Many other classifiers exist

- Logistic regression
- Perceptrons
- Random forest
- Hidden Markov models
- …
- Boosting
- Ensembles of classifiers
- …

Automatic Annotation

Machine learning / data mining software packages



RAPIDMINER

python

PRTools4
A Matlab Toolbox for Pattern Recognition

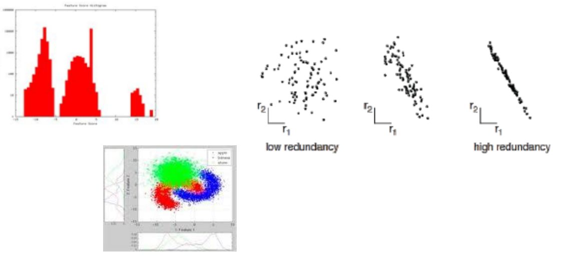
Category-based Retrieval & Annotation

Classification tasks:
General Issues

Category-based Retrieval & Annotation

Classifiers: Issues

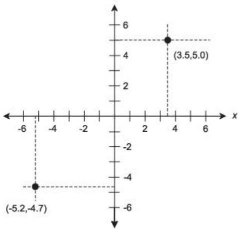
Always perform a visual inspection of your features
(Check that there are no things such as *inf*, or *nan*)



Category-based Retrieval & Annotation

Classifiers: Issues

Do not classify *M* items with *D* features when *D is larger or comparable to M*
As a rule of thumb (personal): *M>20*D*



Category-based Retrieval & Annotation

Classifiers: Issues

Be careful when comparing approaches that use a different number of features

Category-based Retrieval & Annotation

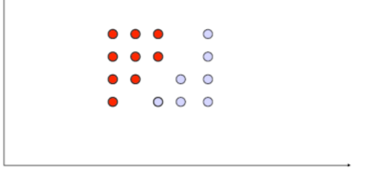
Classifiers: Issues

If possible, use the same amount of instances per class
- "Normalize" by random sampling most populated classes
- Run several experiments (Monte Carlo)

Category-based Retrieval & Annotation

Classifiers: Issues

Avoid overfitting

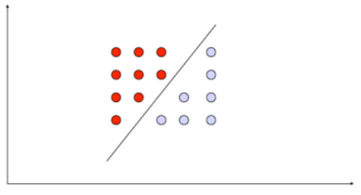


Category-based Retrieval & Annotation

Classifiers: Issues

Avoid overfitting

Diagonal line = ideal solution for the problem

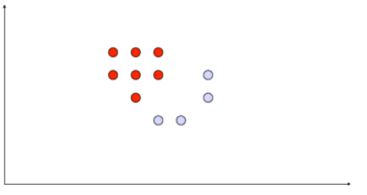


Category-based Retrieval & Annotation

Classifiers: Issues

Avoid overfitting

Let's solve the same problem for a smaller dataset

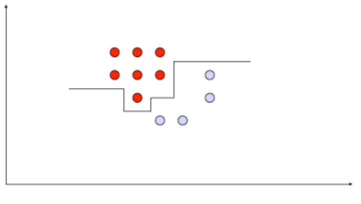


Category-based Retrieval & Annotation

Classifiers: Issues

Avoid overfitting

Let's solve the same problem for a smaller dataset

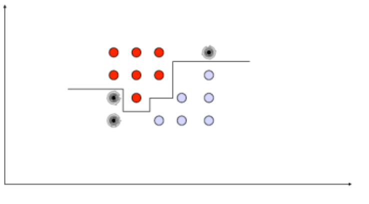


Category-based Retrieval & Annotation

Classifiers: Issues

Avoid overfitting

Classification of new data (3 errors)

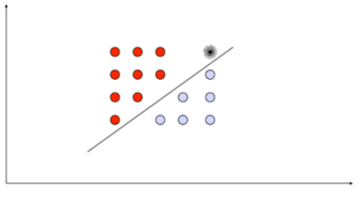


Category-based Retrieval & Annotation

Classifiers: Issues

Avoid overfitting

Solution using a simple line (1 error)



Category-based Retrieval & Annotation

Classifiers: Issues

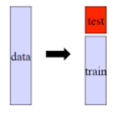
How to **avoid overfitting:**

- Use learning algorithms that intrinsically, by design, generalize well
- Pursue simple classifiers
- Evaluate with the most suitable measure for your task (unbiased and low-variance statistical estimators)
- Estimate the performance of a model with data you did not use to construct the model

Category-based Retrieval & Annotation

Classifiers: Issues

- Hold-out cross-validation



- Cross-fold validation

- LOOCV, Stratified cross-validation, nested cross-validation, …
- With all of them: ensure a sufficient number of repetitions

Category-based Retrieval & Annotation

Classifiers: Issues

Check the default parameters of your classifier

Examples:

- Minimum number of instances per tree branch
- Number of NN in a KNN
- Complexity parameter in an SVM
- Distance-based classifier
  - Which distance
  - Effect of normalization/re-scaling
- …

Category-based Retrieval & Annotation

Classifiers: Issues

If you evaluate your features, demonstrate their usefulness/increment in accuracy

- Use several classifiers (include a "one-rule classifier"?)
- Use several datasets (music collections)
- Always compute/report random baselines (average accuracy, but also their std, min, and max values)
- Include also a baseline feature set in the evaluation

Category-based Retrieval & Annotation

Classifiers: Issues

If you evaluate your features, demonstrate their usefulness/increment in accuracy

Test	Data set A			Data set B		
	Trees	SVM	LLR	Trees	SVM	LLR
MFCC-A	41.3	60.0	**61.3**	56.6	77.6	**78.6**
MFCC-B	41.9	60.6	**63.6**	59.7	76.8	**78.6**
MFCC-C	48.2	**71.6**	71.1	63.1	**81.4**	80.1
OBSC-A	47.4	61.6	**62.4**	58.7	82.8	**83.8**
OBSC-B	46.4	61.4	**64.4**	64.2	**81.4**	81.0
OBSC-C	49.0	67.3	**69.0**	61.4	**82.2**	80.7
SBSC-A	45.5	67.3	**68.1**	63.1	**85.7**	85.5
SBSC-B	48.5	67.0	**68.9**	62.3	**86.8**	85.5
SBSC-C	49.9	72.5	**72.7**	65.1	**86.2**	**86.2**

Category-based Retrieval & Annotation

Classifiers: Issues

Test for statistical significance!

Category-based Retrieval: Conclusions

Scenario: Retrieve music based on automatically annotated labels from audio

Assign a label/class, a probability of that label, and a confidence value

Automatic annotation

Three main steps for **supervised learning**:

- Feature extraction
- Feature transformation and selection
- Classification

Issues with classification tasks

Overfitting, statistical significance, number of features, test several methods/datasets, randomizations, …

Bibliography

- C. M. Bishop.
  Pattern recognition and machine learning.
  Springer, 2007.
- D. Bogdanov, J. Serrà, N. Wack, P. Herrera, and X. Serra.
  Unifying low-level and high-level music similarity measures.
  IEEE Trans. on Multimedia, 13(4): 687-701, 2011.
- P. Cano and M. Koppenberger.
  Automatic sound annotation.
  IEEE Workshop on Machine Learning for Signal Processing, 2004.
- R. O. Duda, P. E. Hart, and D. G. Stork.
  Pattern classification, 2nd ed.
  John Wiley & Sons, 2000.
- Z. Fu, G. Lu, K. M. Ting, D. Zhang.
  A survey of audio-based music classification and annotation.
  IEEE Trans. on Multimedia, 13(2): 303-319, 2011.
- A. K. Jain, R. P. W. Duin, and J. Mao.
  Statistical pattern recognition: a review.
  IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(1): 4-37, 2000.

Bibliography

- A. Klapuri.
  Audio signal classification.
  ISMIR Graduate School, Barcelona, 2004.
- C. Laurier, O. Meyers, J. Serrà, M. Blech, P. Herrera, and X. Serra.
  Indexing music by mood: design and implementation of an automatic content-based annotator.
  Multimedia Tools and Applications, 48(1): 161-184, 2010.
- M. Mandel and D. P. W. Ellis.
  Song-level features and support vector machines for music classification.
  In *Proc. of ISMIR, pp. 594- 599,* 2005
- G. Peeters
  A large set of audio features for sound description (similarity and classification) in the CUIDADO project
  Technical Report, IRCAM, 2004
- N. Scaringella, G. Zoia, and D. Mlynek.
  Automatic genre classification of music content: a survey.
  IEEE Signal Processing Magazine, 23(2):133–141, 2006.

Bibliography

- J. Shen, M. Wang, S. Yan, H. Pang, and X. Hua.
  Effective music tagging through advanced statistical modeling.
  In *Proc. of SIGIR,* 2010.
- E. Tsunoo, T. Akase, N. Ono, and S. Sagayama.
  Musical mood classification by rhythm and bass-line unit pattern analysis.
  In *Proc. ICASSP, pp. 265-268,* 2010.
- D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet.
  Semantic annotation and retrieval of music sound effects.
  IEEE Trans. on Audio, Speech, and Language Processing, 16(2): 467-476, 2008.
- A. Tversky.
  Features of similarity.
  Psychological Reviews, 84(4): 327-352, 1977.
- G. Tzanetakis and P. Cook.
  Musical genre classification of audio signals.
  IEEE Trans. on Speech and Audio Processing, 5(10):293– 302, 2002.
- K. West.
  Novel techniques for audio music classification and search.
  PhD Thesis, University of East Anglia, 2008.

Bibliography

- I. A. Witten and E. Frank.
  Data mining: practical machine learning tools and techniques, 2nd ed.
  Elsevier, 2nd ed., 2005.
- L. M. Zbikowski.
  Conceptualizing music: cognitive structure, theory and analysis.
  Oxford University Press, 2002.

FRSM - 2012

Talk

Machine Learning for Music Discovery

Joan Serrà

Artificial Intelligence Research Institute (IIIA-CSIC)
Spanish National Research Council
jserra@iiia.csic.es
http://joanserra.weebly.com

compmusic

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE - IIIA

CSIC