Timbre Identification

Classification of Musical Timbre Using Bayesian Networks

Carina Schäffer

carina.schaeffer@rwth-aachen.de

Seminar on Computer Music

June 28, 2017

Outline

Introduction

- Term Definition Timbre
- Problem
- History

Main Body

- Presented Paper
- Algorithms
- Music Example
- Feature Extraction
- Bayesian Network Models
- Experiments
- Results

Conclusion

- Summary

Term Definition Timbre

"That multidimensional attribute of auditory sensation which enables a listener to judge that two non-identical sounds, similarly presented and having the same loudness, pitch, spatial location, and duration, are dissimilar."

"A quality of sound that makes voices or musical instruments sound different from each other." - Cambridge Dictionary

Problem

- Computationally identify different timbres
- Classification by instrument, musician or other preselected timbre feature from the other
- Machine learning problem
- Usage: Genre categorization, automatic score creation, track separation

History

1977	John Grey	Stanford University	Computational musical instrument identification
1999	Marques, Moreno	Cambridge Research Laboratory	SVM (70% accuracy)
2000	Fujinaga, MacMillan	Johns Hopkins University, Baltimore	k-NN system (68% classification)
2005	Kaminskyj, Czaszejko	Monash University, Melbourne	k-NN system (93% classification)
2006	Essid, Richard, David	University of Paris-Saclay	SVM (87% accuracy)

Presented Paper

"Classification of Musical Timbre Using Bayesian Networks" by Patrick J. Donnelly and John W. Sheppard (2013)

- Classification of single, monophonic musical instruments
- Bayesian networks for learning
- Comparison with k-NN systems and SVM

Algorithms I - k-Nearest Neighbor (k-NN)

- a previously unknown example is classified with the most common class amongst its k-nearest neighbors
- Apply some distance metric (Euclidean distance) to determine neighbors
- Method: each sample in the test set is compared to a subset of examples from the training set (using the distance metric) and then assigned with the most common class label among the k nearest neighbors

Algorithms II - Support Vector Machine (SVM)

- discriminant-based method for classification or regression
- constructs a hyperplane in high dimensional space that represents the largest margin separating to classes of data (multiclass problems: "one-versus-all" binary classifiers"
- Linear classifier if kernel function of feature vector is the feature vector itself
- If the kernel is a non-linear function, the features are projected into higher-order space
- Algorithm fits the maximum margin hyperplane in the transformed feature space

Algorithms III - Bayesian Networks

- Probabilistic graph models composed of random variables (represented as nodes) and their conditional dependencies (directed edges)
- Joint probability of represented variables: Product of the individual probabilities of each variable, conditioned on the node's parents
- Bayesian classifier:

$$classify(\boldsymbol{f}) = argmax_{c \in C} P(c) \prod_{f \in \boldsymbol{f}} P(f|parent(f))$$

  $P(c)$ prior probability of class c, $P(f|parent(f))$ conditional probability of feature f given the values of the variable's parents
- Classifier finds the class label with the highest probability of explaining the values of the feature vector

Music Example

Feature Extraction

- audio files: instrument sustains a single note for 1s (each file is 2s long to include attack decay)
- transform audio files to a small vector of relevant numeric features
- Use fast Fourier transform over 20 100ms-slots to get the amplitude as a function of frequencies, then group frequencies into ten exponentially increasing windows (each twice the size of the previous one) on a range from 0 to 22,050Hz
- For each frequency window, extract the peak amplitude as feature
- Choice of features heavily influences the outcome of the chosen learning algorithm

Bayesian Network Models

- Naive Bayes(NB): All evidence nodes are conditionally independent of each other given the class

$$P(c|\boldsymbol{f}) = P(c) \cdot \prod_{f \in \boldsymbol{f}} P(f|c)$$

- Frequency dependencies (BN-F): Each frequency feature is conditionally dependent on the previous frequency feature within a single time window
- Time Dependencies (BN-T): Conditional dependencies in the time domain
- Frequency and Time Dependencies (BN-FT): Both time and frequency dependencies

Experiments

1. Instrument and family identification
2. Instrument Identification within Family
3. Classification Accuracy by Data Set Size
4. Repetition of Experiments 1 and 2 with Iowa Data Set

Results Experiment 1 - Accuracy

Algorithm	*Instrument*	*Family*
NB	81.57	80.94
BN-F	97.53	92.87
BN-T	96.36	94.39
BN-FT	**98.25**	97.09
SVM-L	81.46	85.57
SVM-Q	93.55	95.65
k-NN	92.99	**97.31**

Accuracy of classification (in percent), by instrument ($n = 24$) and by instrument family ($n = 4$), for the EastWest data set. Values in **boldface** indicate best results.

I: BN-FT > BN-F > BN-T > (*k*-NN, SVM-Q) > (SVM-L, NB)
F: (BN-FT, *k*-NN) > SVM-Q > BN-T > BN-F > SVM-L > NB

Results Experiment 1 - Confusion

- Bayesian models: Increased confusion between brass and woodwind instruments, compared to string or percussion instruments
- SVMs, k-NN, NB: Higher confusion between strings and either brass or woodwind

Results Experiment 2

Algorithm	*Strings*	*Woodwinds*	*Brass*	*Percussion*
NB	89.76	84.58	92.43	99.64
BN-F	**99.86**	95.89	**99.70**	99.94
BN-T	99.12	95.56	99.36	99.92
BN-FT	99.60	**97.86**	99.58	**99.96**
SVM-L	98.66	92.01	98.65	98.18
SVM-Q	96.82	94.62	97.35	98.48
k-NN	98.72	92.67	98.63	99.72

Accuracy of classification (in percent), by instrument family ($n = 4$), for the EastWest data set. Values in **boldface** indicate best results.

Results Experiment 3

- Evaluation with data set size from 100 to 1000 samples for each instrument
- Bayesian models: Optimal accuracy at 500 - 800 data samples per instrument
- SVMs and k-NN: Improve with increasing number of samples
- Bayesian models achieved much higher accuracy with far fewer examples than either SVMs or k-NN

Results Experiment 4

Algorithm	*Instrument*	*Family*
NB	46.34	73.30
BN-F	**80.76**	81.82
BN-T	75.25	81.24
BN-FT	**80.31**	87.33
SVM-L	65.36	75.03
SVM-Q	65.89	83.19
k-NN	72.78	**89.67**

Accuracy of classification (in percent), by instrument ($n = 25$) and by instrument family ($n = 3$), for the Iowa data set. Values in boldface indicate best results.

- Significantly smaller data set
- Results consistent with previous results considering the same data size

Summary

- Introduction to Timbre Identification
- Presentation of most important algorithms
- Comparison of Bayesian networks

References I

Acoustic Society of America.
timbre - Welcome to ASA Standards.
http://asastandards.org/Terms/timbre.
Accessed on May 31st, 2017.

Cambridge University Press.
timbre Bedeutung im Cambridge Englisch Wörterbuch.
http://dictionary.cambridge.org/de/worterbuch/englisch/timbre.
Accessed on May 31st, 2017.

Patrick J. Donnelly and John W. Sheppard.
Classification of musical timbre using bayesian networks.
Comput. Music J., 37(4):70–86, December 2014.

References II

Numerical Intelligent Systems Laboratory.
Index of /instruments.
http://nisl.cs.montana.edu/instruments.
Accessed on May 31st, 2017.

University of Iowa Electronic Music Studios.
Musical Instrument Samples.
http://theremin.music.uiowa.edu/MIS.html.
Accessed on May 31st, 2017.