Song-level features and SVMs for music classification

Michael Mandel and Dan Ellis {mim,dpwe}@ee.columbia.edu
LabROSA – Columbia University, New York, NY

Summary

- 18-way artist identification task
- Compare all pairs of {artist-, song-level features} × {no SVM, SVM}
- Best performance with song-level features and SVM
- Compare 3 different features / distance measures
- KL divergence on single Gaussians placed 1st in audio artist ID and 2nd in audio genre ID competitions at MIREX '05 with 72.5% and 78.8% accuracies, respectively

Algorithm

- Calculate MFCCs for whole song
- Three different features / metrics
  - Mean of MFCCs and covariance unwrapped and combined into single vector, compared using Mahalanobis distance:

$$D_M(\mathbf{u},\mathbf{v}) = (\mathbf{u}-\mathbf{v})^T\Sigma^{-1}(\mathbf{u}-\mathbf{v}) \quad (1)$$

  - ML Gaussian with mean and covariance of MFCCs, compared using KL divergence:

$$2KL_{\mathcal{N}}(p\,||\,q) = \log\frac{|\Sigma_q|}{|\Sigma_p|} + Tr(\Sigma_q^{-1}\Sigma_p) + (\mu_p-\mu_q)^T\Sigma_q^{-1}(\mu_p-\mu_q) - d \quad (2)$$

  - 20-Gaussian mixture model of MFCCs, compare using KL divergence, estimated with 500 Monte Carlo samples:

$$KL_{GMM}(p\,||\,q) \approx \frac{1}{n}\sum_{i=1}^{n}\log\frac{p(X_i)}{q(X_i)} \quad (3)$$

- Divergences do not fulfill the Mercer conditions on kernels.
  - Symmetrize:

$$D_{KL}(p,q) = KL(p\,||\,q) + KL(q\,||\,p) \quad (4)$$

  - Turn distance measure into similarity (i.e. make positive semidefinite):

$$K(X_i,X_j) = e^{-\gamma D(X_i,X_j)} \quad (5)$$

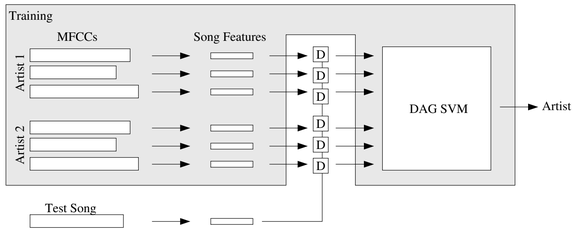
Dataset: *uspop2002*

- To avoid "album effect" each album either training, testing, or validation
- 18 artists (out of 400) had enough albums, 90 albums total
- 1210 songs: 656 training, 451 testing, 103 validation
- Also used 3-fold cross-validation, songs randomly assigned to a group, album effect improves accuracy.

Experiments

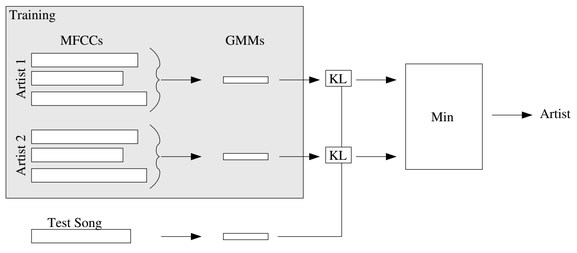
Our solution: classification of song-level features with a DAG-SVM.

- Flexible features and distances, KL divergence between Gaussian song models performed best
- Once features extracted, fast retraining (10s-100s of example songs)
- Useful for relevance feedback, training many classifiers quickly



Brand X: classification of artist-level features without an SVM.

- Gaussian mixture model trained on MFCC frames pooled by artist
- Slow feature extraction and training (10,000s of example frames)
- Too slow to do cross-validation experiments

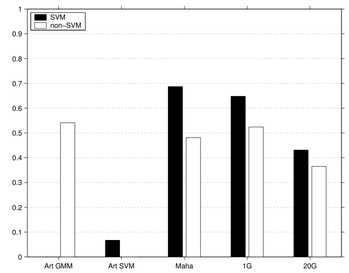


Two other classifiers:

- SVM, no song-level features: Train an 18-way DAG-SVM on many frames from each artist. Classify test song by classifying frames and taking artist with most votes.
- No SVM, song-level features: k-nearest neighbors classifier. Classify test song based on labels of closest training songs using distance metrics mentioned above.

Results

Classification accuracy on separate training and testing albums.
Columns: Artist GMMs, artist SVMs, and kNN and SVMs using Mahalanobis distance, KL divergence between single Gaussians, and KL divergence between 20-component GMMs



Accuracy training and testing on separate albums (Sep) and within albums (Same).

Classifier	Song-Level?	SVM?	Sep	Same
Artist GMM	No	No	.541	—
Artist SVM	No	Yes	.067	—
Song KNN	Yes	No	.524	.722
Song SVM	Yes	Yes	.687	.839

Accuracy for different song-level distance measures.

Classifier	Distance	Sep	Same
KNN	Mahalanobis	.481	.594
KNN	KL-Div 1G	.524	.722
KNN	KL-Div 20G	.365	.515
SVM	Mahalanobis	.687	.792
SVM	KL-Div 1G	.648	.839
SVM	KL-Div 20G	.431	.365

References

Jean-Julien Aucouturier and Francois Pachet. Improving timbre similarity : How high's the sky? *Journal of Negative Results in Speech and Audio Sciences*, 1(1), 2004.

Dan Ellis, Adam Berenzweig, and Brian Whitman. The "uspop2002" pop music data set, 2005. http://labrosa.ee.columbia.edu/projects/musicsim/ uspop2002.html.

Michael I. Mandel, Graham E. Poliner, and Daniel P. W. Ellis. Support vector machine active learning for music retrieval. *ACM Multimedia Systems Journal*, 2005. Submitted for review.

Pedro J. Moreno, Purdy P. Ho, and Nuno Vasconcelos. A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

William D. Penny. Kullback-Liebler divergences of normal, gamma, Dirichlet and Wishart densities. Technical report, Wellcome Department of Cognitive Neurology, 2001.

John C. Platt, Nello Cristianini, and John Shawe-Taylor. Large margin DAGs for multiclass classification. In S.A. Solla, T.K. Leen, and K.-R. Mueller, editors, *Advances in Neural Information Processing Systems 12*, pages 547–553, 2000.