Combining Feature Kernels for Semantic Music Retrieval

Luke Barrington, Mehrdad Yazdani, Doug Turnbull & Gert Lanckriet
Computer Audition Laboratory, University of California, San Diego
lukeinusa@gmail.com

Semantic Music Rerieval

Given a tag, we want relevant songs.

Train a Support Vector Machine (SVM) for each tag, by building **kernels** from features that describe the songs.

The distance of a new song from a tag's SVM boundary determines if that song is relevant to the tag.

We use **four different feature types** to describe CAL-500[1] songs and retrieve using 61 tags.

Content Features

Chroma ~ harmonic content
12-D representation of chromatic energy
Chroma ROC = 0.6

MFCC ~ spectral / timbral content
39-D summary of the signal's spectrum
MFCC ROC = 0.71

Context Features

last.fm ~ social tag context
User-generated tags of songs and artists
last.fm ROC = 0.68

Web Docs ~ online music context
Relevance Scoring[2] of music web pages
Web Docs ROC = 0.67

Best Single Feature

3 of the 4 features are most predictive of some of the 61 tags:

Chroma	**0**
MFCC	**37**
last.fm	**16**
Web Docs	**8**

Best Single Feature ROC = 0.73*

* this is, of course, cheating as it is not possible to know in advance which feature will be best at predicting a given tag. Researchers with access to such an oracle are encouraged to contact the Computer Audition Laboratory with predictions about Apple's share price

Optimum Feature Combination

The best solution would be to combine the information represented by each feature.

We achieve this by learning the optimum weighted combination of feature kernels:

$$K_{opt} = \sum_{i=1}^{m} \mu_i K_i$$

Lanckriet et al. [2004][3] propose a convex solution to this problem that learns both this optimal weighted kernel and the resulting SVM.

By combining the information from all kernels, we can achieve performance that is greater than even the best individual kernel.

Caveat: the combined kernel SVM tends to require more training data than a single kernel SVM

1 D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. *Semantic annotation and retrieval of music and sound effects*. IEEE TASLP, 16(2):467–476, February 2008.
2 P. Knees, T. Pohle, M. Schedl, D. Schnitzer, and K. Seyerlehner. *A Document-centered Approach to a Natural Language Music Search Engine*. European Conference on Information Retrieval, 2008.
3 G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M.I. Jordan. *Learning the kernel matrix with semi-definite programming*. Journal of Machine Learning Research, 5:27–72, 2004.