A Systematic Comparison of Music Similarity Adaptation Approaches

Daniel Wolff[2], Sebastian Stober[1], Tillman Weyde[2] & Andreas Nürnberger[1]

Abstract

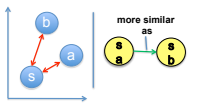
In order to support individual user perspectives and different retrieval tasks, music similarity can no longer be considered as a static element of MIR systems. Various approaches have been proposed recently that allow dynamic adaptation of music similarity measures. This paper provides a systematic comparison of algorithms for metric learning and higher-level facet distance weighting on the MagnaTagATune dataset. A cross-validation variant taking into account clip availability is presented. Applied on user generated similarity data, its effect on adaptation performance is analyzed. Special attention is paid to the amount of training data necessary for making similarity predictions on unknown data, the number of model parameters and the amount of information available about the music itself.

Clip Features & Facets

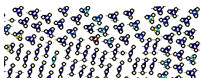
feature	dim.	value description	#facets
key	1	0 to 11 (one of the 12 keys) or −1 (none)	
mode	1	0 (minor), 1 (major), or −1 (none)	
loudness	1	overall value in decibel (dB)	
tempo	1	in beats per minute (bpm)	1 each
time signature	1	3 to 7 ($\frac{3}{4}$ to $\frac{7}{4}$), 1 (complex), or −1 (none)	
danceability	1	between 0 (low) and 1 (high)	
energy	1	between 0 (low) and 1 (high)	
pitch mean	12	dimensions correspond to pitch classes	1 / 12
pitch std. dev.	12	dimensions correspond to pitch classes	1 / 12
timbre mean	12	normalized timbre PCA coefficients	1 / 12
timbre std. dev.	12	normalized timbre PCA coefficients	1 / 12
tags	99	binary vector (very sparse)	14 / 99
genres	44	binary vector (very sparse)	1
			26 / 155

Data Partitioning

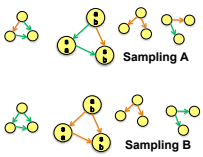
- We train relative constraints $d(s,a) < d(s,b)$ for clips s, a, b.



- In MagnaTagATune, constraints are distributed unevenly over pairs of clips, grouping triplets.



- Random selection of constraints does not guarantee train and test sets to be disjunct.
- Sampling along triplets solves this problem.



Linear Combination of Facet Distances

idea: $d(a,b) = w_{rhythm} \cdot \delta_{rhythm}(a,b) + w_{timbre} \cdot \delta_{timbre}(a,b) + \ldots$

with:
- facet distances: $\delta_f(a,b) \geq 0 \quad \delta_f(a,b) = \delta_f(b,a)$
- facet weights: $w_f \geq 0 \quad \sum_f w_f = const.$

➞ constraints reformulation:

$$d(s,a) < d(s,b) \Leftrightarrow \sum_f w_f(\delta_f(s,b) - \delta_f(s,a)) = \mathbf{w}^\intercal \mathbf{x} > 0$$

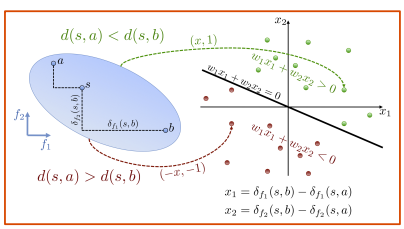
➞ constraint optimization or binary classification problem

Candidate Algorithms:

a) gradient descent (heuristic error reduction)

b) quadratic programming (slack minimization)

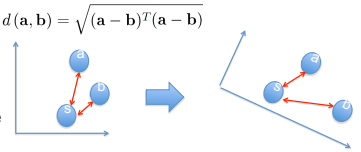
c) linear SVM (LIBLINEAR) (margin maximization)



Metric Learning

- vector space combining all facets
- Mahalanobis metrics generalize the Euclidean metric.
- Mahal. matrix $\mathbf{W}$ determines characteristics of distance measure

$$d(\mathbf{a},\mathbf{b}) = \sqrt{(\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b})}$$



$$d_{\mathbf{W}}(\mathbf{a},\mathbf{b}) = \sqrt{(\mathbf{a}-\mathbf{b})^T\mathbf{W}(\mathbf{a}-\mathbf{b})}$$

SVMLIGHT

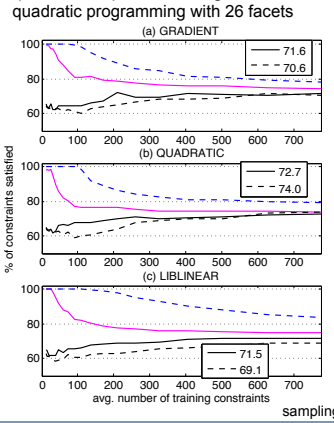
- optimises diagonal matrix $\mathbf{W}$
- #params = dim($\mathbf{W}$)
- optimization performed on vector space
- allows slack penalty

MLR (Metric Learning to Rank)

- different versions for learning full (MLR) and diagonal matrix $\mathbf{W}$ (DMLR)
- #params = dim($\mathbf{W}$)² or dim($\mathbf{W}$) (DMLR)
- based on Sturctural SVM for optimising to Information retrieval quality measures: **AUC**
- allows slack penalty

26 (—) vs. 155 (- -) Facets

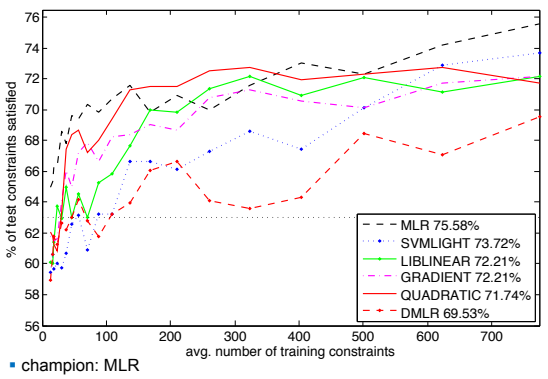
- overfitting: simpler model generalizes quicker
- quickest adaptation to a good trade-off by quadratic programming with 26 facets



sampling: B

Evaluation: Performance across all methods

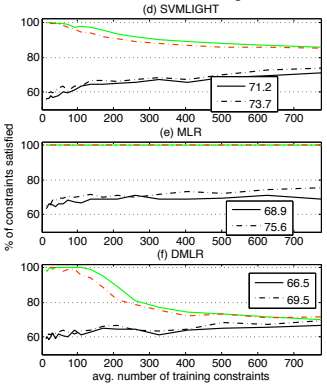
- 10-fold cross-validation
- training sets grow in size (expanding subsets)
- sampling A used for overall comparison



- champion: MLR

Sampling A (- -) vs. B (—)

- significant only for metric learning
- sampling B: all approaches degrade
- effects of transductive learning



[1] DKE Data & Knowledge Engineering Group, Otto von Guericke Universität Magdeburg, INF Fakultät für Informatik
- e-mail: sebastian.stober@ovgu.de, andreas.nuernberger@ovgu.de
- website: http://www.dke-research.de

[2] MIRG Music Informatics Research Group, City University London
- e-mail: daniel.wolff.1@city.ac.uk, t.e.weyde@city.ac.uk
- website: http://mi.soi.city.ac.uk