COMPARATIVE MUSIC SIMILARITY MODELLING USING TRANSFER LEARNING ACROSS USER GROUPS

Daniel Wolff, Andrew MacFarlane and Tillman Weyde
Music Informatics Research Group – Department of Computer Science
City University London
{daniel.wolff.1, a.macfarlane-1, t.e.weyde}@city.ac.uk

ABSTRACT

We introduce a new application of transfer learning for training and comparing music similarity models based on relative user data: The proposed Relative Information-Theoretic Metric Learning (RITML) algorithm adapts a Mahalanobis distance using an iterative application of the ITML algorithm, thereby extending it to relative similarity data. RITML supports transfer learning by training models with respect to a given template model that can provide prior information for regularisation. With this feature we use information from larger datasets to build better models for more specific datasets, such as user groups from different cultures or of different age. We then evaluate what model parameters, in this case acoustic features, are relevant for the specific models when compared to the general user data.

We to this end introduce the new CASimIR dataset, the first openly available relative similarity dataset with user attributes. With two age-related subsets, we show that transfer learning with RITML leads to better age-specific models. RITML here improves learning on small datasets. Using the larger MagnaTagATune dataset, we show that RITML performs as well as state-of-the-art algorithms in terms of general similarity estimation.

1. INTRODUCTION

Music similarity models are a central part of many applications in music research, particularly Music Information Retrieval (MIR). When training similarity models, it turns out that learnt models vary considerably for different data sets and application scenarios. Recently, context-sensitive models have been introduced, e.g. for the task of music recommendation (Stober [9] provides an overview). The main problem with context-sensitive similarity models is currently to obtain enough data to train the models for each context. Transfer learning promises to enable effective training of models for specific contexts by including information from related datasets. We here present an approach of transfer learning in music similarity that improves results of specialised models, using our W_0-RITML extension of Information-Theoretic Metric Learning (ITML). The template-based optimisation in W_0-RITML allows for a comparison of the general and specialised models – it derives the latter from the former – which we suggest as a tool for comparative analysis of similarity data by (e.g. cultural) provenance.

We are particularly interested in modelling relative similarity ratings collected from participants during Games With a Purpose (GWAPs). Using similarity data from user groups promises to provide tailored model performance and the opportunity to compare such groups via the trained similarity models. The new CASimIR dataset presented in Section 3 contains such similarity ratings and information about the contributing subjects. We use this extra data to group users and here exemplarily train age-specific music similarity models based on age-bounded subsets. However, the relatively small size of the CASimIR dataset requires a different approach to training the group-specific models as existing algorithms are not sufficiently effective for this purpose.

We contribute a solution to this problem with a novel generic algorithm for transfer learning with similarity models: The RITML algorithm (see Section 5.2) extends on ITML to allow for learning a Mahalanobis metric from relative similarity data like in CASimIR. With W_0-RITML, information learnt from remaining data can be successfully transferred to an age-bounded dataset via a Mahalanobis matrix. This transfer-learning increases performance on small datasets and provides interpretable values in the Mahalanobis matrix. The Mahalanobis matrix provides a compact representation of similarity information in a dataset. This is useful in scenarios where the music data is difficult to access due to its data volume or copyright restrictions. The CASimIR dataset and code used in this paper are available online [1] .

2. RESEARCH BACKGROUND

Transfer learning relates to many areas and approaches in machine learning. A general overview of transfer learning is given in Pan and Yang [6]. In their categorisation, our task is an inductive knowledge transfer from one similarity modelling task to another via model parameters. Note that

[1] http://mirg.city.ac.uk/datasets/ismir2015dw

in our example the tasks differ only in the dataset, but our method can also be used for more divergent tasks.

In MIR, transfer learning is a relatively new method. In 2013, [2] described multi-task learning using a shared latent representation for auto-tagging, genre classification and genre-based music similarity. This representation includes both the features and the labels for the different tasks. In experiments on several datasets they showed improvement of classification accuracy and modelling similarity according to genre.

We here work with relative similarity ratings from humans in our new CASimIR dataset for group-specific modelling. Furthermore, we use the MagnaTagATune dataset [3] for comparison on non-specific similarity learning. Here, the Support Vector Machine (SVM) approach developed by Schultz and Joachims [7] and applied in [10, 11] is used as state-of-the art baseline.

Another state-of-the-art algorithm for learning from relative similarity data is Metric Learning To Rank (MLR). McFee et al. [4] introduce MLR for parametrising a linear combination of content-based features using collaborative filtering data. Their post-training analysis of feature weights revealed that tags relating to genre or radio stations were assigned greater weights than those related to music theoretical terms.

3. A DATASET FOR USER-AWARE SIMILARITY

In order to perform a related analysis and comparisons of models between different user groups, we have collected the CASimIR datasets using Spot the Odd Song Out [13], an online [2] multi-player Game With a Purpose (GWAP). The similarity module of the Spot the Odd Song Out game collects relative similarity data using an odd-one-out survey: From a set of three music clips, participants are asked to choose the clip most dissimilar to the remaining clips, i.e. the *odd song out*. The game motivates players by rewarding blind agreement. For various reasons, including personal data protection, little music annotation data is publicly available with information about the provider of the data and their context.

Although the game can collect anonymised personal information including gender, nationality, spoken languages and musical experience, the amount and type information available varies between participants, as data provision is voluntary. Our overarching goal is to study the relation between similarity and culture and we thus link annotations to cultural profiles rather than indexing specific participants. With this paper we publish the first set of similarity data with anonymised profiles.

3.1 Constraints from Relative Similarity Ratings

The MagnaTagATune and CASimIR datasets both contain relative similarity ratings. A participant's rating of C_k as the odd one out (of the triplet C_i, C_j, C_k) results in 2 relative similarity constraints: clips C_i and C_j are more similar than C_i and C_k, and clips C_j and C_i are more similar than C_j and C_k. These constraints are denoted as (i, j, k) and (j, i, k), respectively which are contained in the constraint set $\hat{Q}$.

Human ratings regularly produce inconsistent constraints. We use the graph representation of the similarity data as suggested by [5] to analyse and filter inconsistencies: Each constraint (i, j, k) is represented by an edge connecting two vertices $(i, j) \overset{\alpha_{ijk}}{\rightarrow} (i, k)$ corresponding to two clip pairs, with the edge weight $\alpha_{ijk} = 1$. When combining all constraints in a graph, the weights α_{ijk} are accumulated. Inconsistencies then appear as cycles in the graph, which in their most common form are of length 2:

$$(i, j) \underset{\alpha_{ikj}}{\overset{\alpha_{ijk}}{\rightleftarrows}} (i, k).$$

We remedy such cycles by removing the edge with the smaller weight and assigning the weight $|\alpha_{ijk} - \alpha_{ikj}|$ to the remaining edge. For both the MagnaTagATune and CASimIR datasets this already creates a cycle-free graph Q as no larger cycles remain. The cycle-free sets Q are used in this study for training and evaluation.

Compared to the MagnaTagATune dataset, the CASimIR dataset features more frequent recurrences of clips between the triplets presented to the users. Recurring clips relate the corresponding similarity data, and result in large connected components in the CASimIR similarity graph: While the maximal number of clips directly or transitively related to each other through similarity data in the MagnaTagATune dataset was 3 (see [11]), most clips in the CASimIR similarity data are related to at least 5 other clips. The repetition of clips across triplets results in fewer unique referenced clips: the current CASimIR similarity dataset contains only 180 clips referenced by 2102 ratings, while MagnaTagATune references 2000 ratings with about 500 clips, and has 1019 clips with 7650 ratings in total.

3.2 Analysis of Age-bounded Similarity Ratings

The additional participant attributes allow us to select subsets of similarity data according to specific profiles of the participants. This enables the training of more specific models that support better similarity predictions for the relevant group of users, and allows for comparison of different models.

As an example of group-based similarity modelling we choose age as a separating criterion on the CASimIR similarity data from over 256 participants: We divide the complete set of similarity ratings R into two *age-bounded* subsets $R^{\leq 25}$ of data provided by participants not older than 25 years and $R^{>25}$ containing data of older participants. The boundary of 25 years was chosen as the best approximation to equal sizes of the subsets (data input is only in 5 year bands). As shown in Table 1, the number of ratings is higher for the $R^{\leq 25}$ dataset.

[2] http://mirg.city.ac.uk/camir/game/

	R	$R^{\leq 25}$	$R^{>25}$	$R^{\complement(\leq 25)}$	$R^{\complement(>25)}$
ratings	2102	919	644	1183	1458
constr.	914	723	576	732	809
clips	180	171	163	175	176

Table 1. Number of votes, unique constraints and referenced clips, after filtering inconsistencies, per dataset.

539 similarity ratings are not associated to a valid age and stored separately in $R^{\emptyset}$. For the two age-bounded datasets, we furthermore define complementary datasets $R^{\complement(\leq 25)}$ and $R^{\complement(>25)}$ combining the remaining similarity data, e.g. $R^{\complement(\leq 25)}$ = $R^{>25} \cup R^{\emptyset}$. These complementary sets will be used for training of template models for transfer learning.

After splitting, the above (sub)sets of ratings are transferred into constraints (see Section 3.1) and separately filtered for inconsistencies. We now use the corresponding sets of unique constraints Q, $Q^{\leq 25}$, $Q^{>25}$, $Q^{\complement(\leq 25)}$ and $Q^{\complement(>25)}$ for training and testing of models. The number of constraints are also noted in Table 1, together with the total number of clips referenced by the constraint sets. Due to multiple ratings referring to the same constraint and filtering the constraint count is lower than the number of ratings.

4. SIMILARITY MODELLING

The computational representations of music through features, related to physical, musical, and cultural attributes determine the basis of similarity models. Both the MagnaTagATune and CASimIR datasets contain pre-computed features created by The Echo Nest API. For our experiments with CASimIR we derive acoustic features from this data which are aggregated to the clip-level. The 41-dimensional features contain 12 chroma and 12 timbre features, both aggregated via averaging, 2 weight vectors and further features after [8, 11]:

chroma	timbre
segmentDurationMean	tempo
segmentDurationVariance	beatVariance
timeLoudnessMaxMean	tatum
loudness	tatumConfidence
loudnessMaxMean	numTatumsPerBeat
loudnessMaxVariance	timeSignature
loudnessBeginMean	timeSignatureStability
loudnessBeginVariance	–

Table 2. Features used in our experiments.

For experiments with the MagnaTagATune dataset we will use the similar features provided in [12] which contain preprocessed tag information in addition to the acoustic features described above. For the CASimIR dataset, using unprocessed tags from Last.fm did not increase performance in earlier experiments due to very sparse tag assignments. Therefore, our experiments on CASimIR use acoustic features only. For a clip C_i, we refer to its feature vector as $x_i \in \mathbb{R}^N$.

4.1 Mahalanobis Distances

We use the inverse of the distance of two feature vectors as the similarity of the two corresponding clips. The mathematical form of the Mahalanobis distance is used to specify a parametrised distance measure. Given two feature vectors $x_i, x_j \in \mathbb{R}^N$, the distance can be expressed as

$$d_W(x_i, x_j) = \sqrt{(x_i - x_j)^\intercal W (x_i - x_j)},$$

where $W \in \mathbb{R}^{N \times N}$ is a square matrix parametrising the distance function: the *Mahalanobis matrix*. d_W qualifies as a metric if W is positive definite and symmetric.

5. MODEL TRAINING WITH RITML

We now discuss our algorithm which can adapt Mahalanobis distances in order to fit relative similarity data. It is based on the ITML algorithm as described below, which cannot be used directly with relative similarity data. Instead, ITML requires upper or lower bounds on the similarity of two clips, e.g. $d_W(x_i, x_j) < m_{i,j}$ for similar clips. In Section 5.2 we will iteratively derive such constraints during the RITML optimisation process.

5.1 Information-Theoretic Metric Learning

Davis et al. [1] describe Information-Theoretic Metric Learning (ITML) for learning a Mahalanobis distance from absolute distance constraints (e.g. requiring $d_W(x_i, x_j) < 0.5$). A particularly interesting feature of ITML is that a template Mahalanobis matrix $W_0 \in \mathbb{R}^{n \times n}$ can be provided for regularisation. This W_0 can be from a metric that is predefined or learnt on a different dataset. If W_0 is not specified, the identity transform is used. The regularisation of ITML exploits an interpretation of Mahalanobis matrices as multivariate Gaussian distributions: The distance between two Mahalanobis distance functions parametrised by W and W_0 is measured by the relative entropy of the corresponding distributions, which in [1] uses the LogDet divergence D_{ld}:

$$\begin{aligned} D_{ld}(W, W_0) &= \mathrm{tr}(W W_0^{-1}) - \log\det(W W_0^{-1}) - n \\ &= 2 * \mathrm{KL}\left(P(x_i; W_0) \parallel P(x_i; W)\right). \end{aligned}$$

KL refers to the Kullback-Leibler divergence. For details of the transformation see [1]. Given the constraints in form of similar (R_s) and dissimilar (R_d) clip indices as well as upper and lower bounds u_{ij}, l_{ij}, the optimisation problem is then posed as follows:

$$\begin{aligned} &\mathrm{ITML}(W, \xi, c, R_s, R_d) = \\ &\qquad \underset{W \succeq 0, \xi}{\operatorname{argmin}} D_{ld}(W, W_0) + c \cdot D_{ld}(\mathrm{diag}(\xi), \mathrm{diag}(\xi_0)) \\ &\text{s.t.} \quad \mathrm{tr}(W d^L_{i,j} (d^L_{i,j})^\intercal) \leq \xi_{ij} \quad \forall (i,j) \in R_s \\ &\qquad \mathrm{tr}(W d^L_{i,j} (d^L_{i,j})^\intercal) \geq \xi_{ij} \quad \forall (i,j) \in R_d \\ &\text{with} \quad d^L_{i,j} = (x_i - x_j). \end{aligned}$$

Here, ξ_{ij} are slack variables enabling and controlling the violation of individual constraints. The ξ_{ij} are initialised to given upper bounds u_{ij}, if $(i,j) \in R_s$ or lower bounds l_{ij}, if $(i,j) \in R_d$. During optimisation, they are regularised by comparison to the template slack ξ_0 using triangular matrices diag(ξ) and diag(ξ_0).

5.2 Relative Learning with RITML

In order to allow for training with relative similarity constraints, we present Relative Information-Theoretic Metric Learning (RITML) based on ITML. Motivated by [14], we embed ITML into an iterative adaptation of the upper and lower bounds.

We start with a training set of relative constraints $(i,j,k) \in Q_t$. We require standard ITML parameters such as c, as well as the relative learning parameters including shrinkage factor η, margin τ and number of cycles k at the beginning. We use the identity matrix for the template W_0. During iteration m, the active training set of violated constraints Q^m is calculated as

$$Q^m = \{(i,j,k) \in Q_t \mid d_{W^m}(x_i,x_j) > d_{W^m}(x_i,x_k)\}\,.$$

Q^m is then further divided into the sets of similar and dissimilar constraints R_s^m and R_d^m:

$$R_s^m = \{(i,j) \mid (i,j,k) \in Q^m\}$$
$$R_d^m = \{(i,k) \mid (i,j,k) \in Q^m\},$$

Afterwards, absolute distance constraints ξ_{ij} for the following ITML instance are acquired by adding a margin τ to the average distance values $\mu = \frac{d_{W^m}(x_i,x_j)+d_{W^m}(x_i,x_k)}{2}$ of the clip pairs:

$$\xi_{ij}^m = \begin{cases} \mu - \tau & (i,j) \in R_s^m \\ \mu + \tau & (i,j) \in R_d^m \end{cases} \quad \forall (i,j,k) \in Q^m$$

Now, with ξ^m containing the upper and lower bounds, ΔW can be calculated using

$$\Delta W = \text{ITML}(W^m, \xi^m, \gamma, R_s^m, R_d^m) \qquad (1)$$

and the final Mahalanobis matrix is accumulated over iterations using the model update function

$$W^{m+1} = \frac{m * W^m + \eta * \Delta W}{m+1}.$$

In order for the algorithm to converge, the cardinality of the active training set $|Q^m|$ needs to decrease. In our experiments, $k = 200$ training iterations are usually sufficient. Otherwise an early stopping of the algorithm takes place if $|Q^m|$ does not decrease for 50 iterations. In this case the W^m for the smallest $|Q^m|$ within the last 50 iterations is returned. RITML does not guarantee d_W to be a metric.

Algorithm 1: Relative Training with RITML

Data: Constraints Q_t, features x_i, template matrix W_0, regularisation factor c, shrinkage factor η, margin τ, number of cycles k

m = 0 ;
while $m \leq k \wedge Q^* \neq \emptyset$ **do**
  Update training sets Q^m, R_s^m and R_d^m ;
  Update absolute constraints ξ^m ;
  Calculate parameter change ΔW ;
  Calculate W^{m+1} ;
  m = m+1 ;
end
return Mahalanobis matrix W^k

5.3 Transfer Learning with W_0-RITML

The property that motivates our usage of RITML is that it enables *transfer learning*: If a specific starting value or template of W_0 other than the identity matrix is provided, the optimisation tends to produce results close to the provided W_0. In order to sustain this effect for large numbers of iterations we modify Equation (1) such that regularisation is fixed towards W_0 instead of the Euclidean distance:

$$\Delta W = \text{ITML}(\mathbf{W_0}, \xi^m, \gamma, R_s^m, R_d^m)$$

This constitutes the W_0-RITML algorithm for transfer learning with Mahalanobis matrices.

6. EXPERIMENTS

For all our experiments we use the 10-fold cross-validation with *inductive sampling* as described in [11]: Instead of dividing the similarity constraints themselves into test/training sets, the data are divided on the basis of connected clusters in the similarity data. This approach prevents the recurrence of clips from a training-set in the corresponding test set. It also leads to a greater variance in test-set sizes for CASimIR where the clusters of connected similarity data are larger.

We evaluate the algorithms' performance based on the percentage of training and test constraints fulfilled by the trained model. Our main focus is on the test-set results as we are interested how well the learnt models generalise to unseen data. As a baseline we use the Euclidean distance on the features. We have tested results for statistical significance using the Wilcoxon signed rank test on cross-validation folds' results with a threshold of $p < 5\%$.

Both SVM as implemented in *svmlight*[7] and RITML have hyper-parameters affecting the performance on different datasets. The results reported here were selected on the basis of best test-set performances after a grid-search over a range of value combinations identified as reasonable in preliminary experiments: The regularisation trade-off c is a parameter common to SVM, RITML and W_0-RITML with a similar effective range: we explored a $c \in [0.001, 10]$ using an approximately logarithmic scale. For RITML and

W_0-RITML we additionally used
$\tau \in \{10^{-4}, 10^{-3} \ldots, 10^{-1}, 0.5, 1 \ldots 10\}$ and
$\eta \in \{0.1, 0.15 \ldots 0.95\}$.

6.1 Comparing the Performance of RITML

For a comparable evaluation of RITML we chose the MagnaTagATune-based dataset and constraint sampling published in [12]. Their evaluation compares various algorithms for learning a Mahalanobis metric using two different samplings. The inductive sampling used here corresponds to the *sampling B* in their text. Table 3 shows the results on MagnaTagATune and on the complete CASimIR dataset (Q).

Algorithm	MagnaTagATune	CASimIR
Euclidean	59.80 / 59.77	59.75 / 59.82
RITML	71.12 / 73.41	**64.23** / **93.36**
SVM	**71.20** / 85.75	63.22 / 69.11
MLR	68.90 / **100.0**	62.79 / 73.37

Table 3. Comparison of Test / Training set performance on the MagnaTagATune and CASimIR datasets for baseline, RITML and SVM. Reported are the number of constraints fulfilled by the learnt distance measures.

For MagnaTagATune, RITML achieves similar generalisation results as SVM (with parameters SVM: $c = 0.7$ and RITML: $c = 1, \eta = 0.85, \tau = 0.5$), while MLR overfits to the training data. For both the MagnaTagATune and CASimIR datasets all methods perform significantly better than the baseline. The RITML results are therefore comparable to the state-of-the-art. The training results on MagnaTagATune with SVM and MLR are far better than the test results, indicating overfitting, which does not occur for RITML. Interestingly, on the CASimIR dataset, the situation between RITML and SVM is reversed. Results published by [11] for acoustic-only features on MagnaTagATune show a performance of 66% on MagnaTagATune, but the lower performance on CASimIR can be explained by the smaller number of training examples.

6.2 Transfer Learning

A core motivation for transfer learning is the training on highly specialised but small datasets. To evaluate the W_0-RITML method for transfer learning, we firstly compared the SVM and RITML algorithms with the baseline on the age-bounded datasets $Q^{>25}$ and $Q^{\leq 25}$ in Table 4. The rightmost column shows the average performance across both age-bounded datasets. Expectedly, on these smaller datasets generalisation results for RITML as well as the reference SVM and MLR are lower than on the whole CASimIR. Only for RITML an increase of 4.37% from the baseline is notable for the slightly larger $Q^{>25}$ which improves the average score for RITML.

We now apply transfer learning to improve generalisation results on the age-bounded sets. The overall process is depicted in Figure 1. First, an similarity modelling experiment is performed on both of the complementary subsets

Algorithm	$Q^{\leq 25}$	$Q^{>25}$	Average
Euclidean	59.32 / 59.95	59.15 / 59.63	59.23 / 59.79
RITML	63.69 / **75.87**	61.02 / 67.95	62.35 / 71.91
SVM	61.56 / 72.78	61.34 / 71.43	61.45 / 72.10
MLR	62.06 / 75.79	62.58 / **78.47**	62.32 / **77.13**
W_0-Direct	63.96 / 66.17	64.82 / 69.57	64.39 / 67.87
W_0-RITML	**65.53** / 70.82	**67.07** / 73.22	**66.30** / 72.02

Table 4. Comparison of Test / Training set performance on the age-bounded datasets. Training on single datasets (top 3 rows) and transfer learning with W_0-RITML and W_0-Direct.

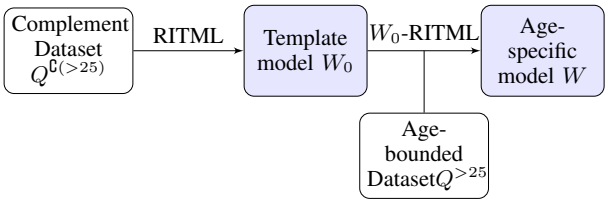


Figure 1. Flow diagram for transfer learning, exemplified for the $Q^{>25}$ dataset.

$Q^{\complement(>25)}$ and $Q^{\complement(\leq 25)}$ using cross-validation with training and test data from only these sets. Comparing the individual results for validation folds we choose the Mahalanobis matrices with the greatest test-set performance as template matrix W_0. The template matrix W_0 learnt on $Q^{\complement(\leq 25)}$ is then used for transfer learning on $Q^{\leq 25}$, using W_0-RITML. For comparison of the effectiveness of the fine-tuning with W_0-RITML, we report the performance achieved with the unmodified W_0 on $Q^{\leq 25}$ as W_0-Direct. This process is repeated analogously for $Q^{\leq 25}$ by applying the template matrix W_0 from $Q^{\complement(\leq 25)}$ on $Q^{\leq 25}$.

The highlighted lower columns of Table 4 show the results for transfer learning: Row W_0-Direct reports the direct performances of the template Mahalanobis matrices W_0. The results of fine-tuning these models with W_0-RITML are reported in the last row. We here find that using the matrices trained on the larger datasets, and thus transfer learning, generally improves results. Only the results for W_0-RITML provide gains $> 6.21\%$ that are statistically significant when compared to the baseline. As the average result of W_0-RITML also significantly outperforms the average SVM performance, W_0-RITML works best for adapting models to specialised datasets.

A drawback of RITML is that it is computationally demanding: For the Q dataset, RITML uses 50 seconds where SVM converges in 5 seconds. On the other hand, SVM learns a diagonal W which reduces the number of parameters and model flexibility.

6.3 Model Comparison

In order to identify specificities of the $Q^{>25}$ dataset in comparison to the remaining $Q^{\complement(>25)}$, we now analyse changes made to the template matrix W_0 in the fine-tuning process. Instead of starting from the Euclidean metric, models learnt from the W_0-RITML method have a model already

adapted to similarity data as basis.

Figure 2 shows the relative difference $\hat{W} - \hat{W}_0$ of the Mahalanobis matrix before (W_0) and after (W) fine tuning. As the fine tuning process rescales the similarity measure and thereby W, the matrices have been normalised to the interval of $[0, 1]$ via [3]

$$\hat{W} = \frac{W - \min_{i,j}(w_{ij})}{\max_{i,j}(W - \min_{i,j} w_{ij})_{ij}}. \quad (2)$$

The axes of the figure correspond to feature types, which for better overview have been grouped into chroma, timbre and ranges of the features in Table 2. The template matrix W_0 in Figure 3a has large values only in the diagonal and homogeneous small values off the diagonal. In comparison to this, Figure 2 shows that specific combinations of timbre features (in the bottom centre) with (B)eat and tempo statistics were raised in importance by W_0-RITML, resulting in the final matrix W as shown in Figure 3b. Also, the centre of the matrix shows increased values for combinations of different timbre coefficients. The strongest increases (20-24%) in weights are reported for the off-diagonal fields of $C_{11}C_1, T_6T_5, B_4T_4$ and B_4T_5, where C, T relate to chroma and timbre coefficients and B_4 refers to the tatumConfidence feature. Weights are increased mainly at the cost of diagonal elements, and suggest at a specialisation of the model to the specificities of the $Q^{>25}$ similarity subset. For this data collected from users aged over 25, the analysed W_0-RITML model with stronger influence of the timbre and beat-statistics features performs best in our evaluation.

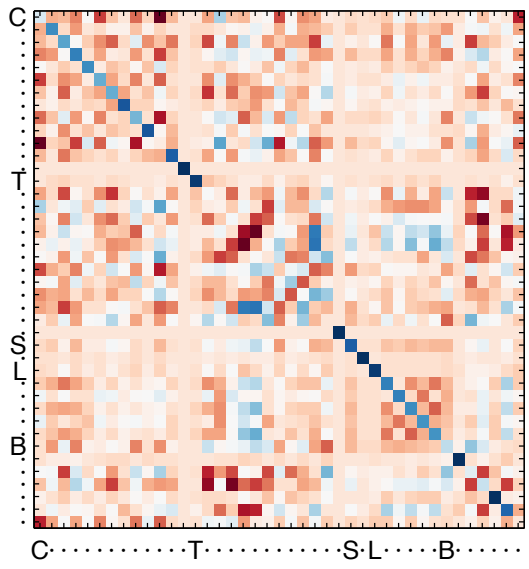
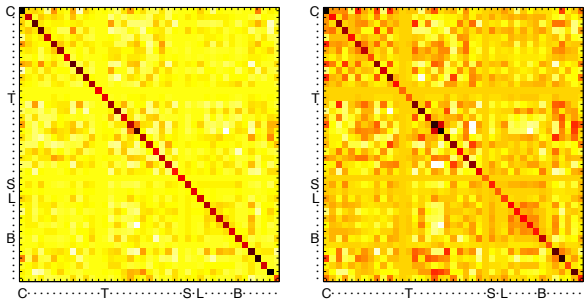


Figure 2. Learnt model difference for W_0-RITML on $Q^{>25}$. Axis labels represent ranges of feature types: (C)hroma, (T)imbre, as well as (S)egment, (L)oudness and (B)eat+Tempo statistics. Dark red / blue colours correspond to strong weight increase / decrease.

[3] Subtraction and division are applied to W in a point-wise manner.



(a) (b)

Figure 3. **(a)** Template matrix W_0 before and **(b)** final matrix W after fine-tuning with W_0-RITML on $Q^{>25}$. The latter shows higher variance in off-diagonal entries for the specialised model. Axis labels represent ranges of feature types: (C)hroma, (T)imbre, as well as (S)egment, (L)oudness and (B)eat+Tempo statistics. Dark red colours correspond to strong weight increase, light yellow to decrease.

7. CONCLUSION & FUTURE WORK

We presented a method for analysing music similarity data of different user groups via models trained with transfer learning. To this end, the new RITML algorithm was developed extending ITML to relative similarity data. A key feature of RITML is that it enables transfer learning with template Mahalanobis matrices via W_0-RITML. Our evaluation of the algorithm was performed on two datasets: The evaluation on the commonly used MagnaTagATune dataset showed that RITML performs comparably to state-of-the-art algorithms for metric learning.

For evaluation of transfer learning with W_0-RITML we provide the CASimIR similarity dataset, the first open dataset containing user attributes associated to relative similarity data. Tests on the whole CASimIR dataset corroborated our finding that RITML competes with current similarity learning methods. Our analysis of W_0-RITML was performed on age-bounded subsets of the dataset. Results showed that transfer learning with W_0-RITML outperforms the standard SVM algorithm on small datasets.

Our comparison of models allowed us to point out specific features and combinations that determine similarity in user data. For this first evaluation we chose age to group users. We hope this will motivate further research in comparison of similarity models and adaptation to data with regard to cultural and user context.

For future work we are interested in collecting larger similarity datasets, and applying the methods introduced here for improved validation of results and the analysis of more specific user groups. The set-up used for our experiments motivates transfer learning across the MagnaTagATune and CASimIR datasets with W_0-RITML for further analysis of the transferability of similarity information via Mahalanobis matrices.

8. REFERENCES

[1] Jason V. Davis, B. Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. Information-theoretic metric learning. In *Proc. of ICML '07*, pages 209–216, New York, NY, USA, 2007. ACM.

[2] Philippe Hamel, Matthew E. P. Davies, Kazuyoshi Yoshii, and Masataka Goto. Transfer learning in mir: Sharing learned latent representations for music audio classification and similarity. In Alceu de Souza Britto Jr., Fabien Gouyon, and Simon Dixon, editors, *ISMIR*, pages 9–14, 2013.

[3] Edith Law and Luis Von Ahn. Input-agreement: A new mechanism for collecting data using human computation games. In *Proc. of CHI*. ACM Press, 2009.

[4] B. McFee, L. Barrington, and G. Lanckriet. Learning similarity from collaborative filters. In *Proc. of ISMIR 2010*, pages 345–350, 2010.

[5] Brian McFee and Gert R. G. Lanckriet. Partial order embedding with multiple kernels. In *Proc. of the 26th International Conference on Machine Learning (ICML'09)*, pages 721–728, June 2009.

[6] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010.

[7] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. In *Advances in Neural Information Processing Systems (NIPS)*. MIT Press, 2003.

[8] Malcolm Slaney, Kilian Q. Weinberger, and William White. Learning a metric for music similarity. In Juan Pablo Bello, Elaine Chew, and Douglas Turnbull, editors, *Proc. of ISMIR 2008*, pages 313–318, 2008.

[9] Sebastian Stober. *Adaptive Methods for User-Centered Organization of Music Collections*. PhD thesis, Otto-von-Guericke-University, Magdeburg, Germany, Nov 2011. published by Dr. Hut Verlag, ISBN 978-3-8439-0229-8.

[10] Sebastian Stober and Andreas Nürnberger. Similarity adaptation in an exploratory retrieval scenario. In *Proc. of AMR 2010*, Linz, Austria, Aug 2010.

[11] Daniel Wolff and Tillman Weyde. Learning music similarity from relative user ratings. *Information Retrieval*, pages 1–28, 2013.

[12] Daniel Wolff, Sebastian Stober, Andreas Nürnberger, and Tillman Weyde. A systematic comparison of music similarity adaptation approaches. In *Proc. of ISMIR 2012*, pages 103–108, 2012.

[13] Daniel Wolff, Guillaume Bellec, Anders Friberg, Andrew MacFarlane, and Tillman Weyde. Creating audio based experiments as social web games with the casimir framework. In *Proc. of AES 53rd International Conference: Semantic Audio*, Jan 2014.

[14] Zhaohui Zheng, Keke Chen, Gordon Sun, and Hongyuan Zha. A regression framework for learning ranking functions using relative relevance judgments. In *Proc. of SIGIR '07*, pages 287–294, New York, NY, USA, 2007. ACM.