ENSEMBLE LEARNING FOR HYBRID MUSIC RECOMMENDATION

Marco Tiemann, Steffen Pauws, Fabio Vignoli
Philips Research Europe
High Tech Campus 34
5656 AE Eindhoven, The Netherlands
{marco.tiemann, steffen.pauws, fabio.vignoli}@philips.com

ABSTRACT

We investigate ensemble learning methods for hybrid music recommenders, combining a social and a content-based recommender algorithm in an initial experiment by applying a simple combination rule to merge recommender results. A first experiment suggests that such a combination can reduce the mean absolute prediction error compared to the used recommenders' individual errors.

1 INTRODUCTION

Music recommender systems that are publicly available today apply one of two recommender paradigms: *social recommenders* predict preferences based purely on user preference data. *Content-based recommenders* represent songs as feature vectors, where the features can be assigned manually or extracted automatically directly from audio data. Users are characterized by profiles, and songs that are similar to a user's profile are recommended. *Hybrid recommenders* combine recommender paradigms to improve the overall accuracy of predictions and reduce problems of specific recommender algorithms. To this end, hybrid recommenders usually combine social and content -based recommendation using one of a number of possible combination methods [1]. They can be categorized into recommenders that 1) integrate social and content-based recommendation in a single algorithm or 2) combine the outputs of independent recommender algorithms. We follow the second approach. Our goal is to integrate many distinct independent recommenders (referred to as *base recommenders* in this paper) flexibly in a generalized hybrid recommender, using ensemble learning methods to create diverse base recommenders and to combine their output with established and novel combination methods.

2 APPLYING ENSEMBLE LEARNING

In *ensemble learning* [3], several weak learners are created and used in regression or classification tasks. Their individual output is combined into a unified single output by applying a combination rule. Ensemble learning methods can often outperform single strong learners in practical applications. However, ensemble methods can only be applied successfully when the used weak learners are sufficiently *diverse* – weak learners must be able to correct errors made by other weak learners. Hence many ensemble learning methods actively select or create suitably diverse weak learners by techniques such as training data resampling. Some ensemble learning methods can also iteratively adapt to newly presented evidence by modifying either combination rules or weak learners as new evidence is successively presented. Ensemble learning is a promising concept for hybrid music recommendation. It can be used to flexibly integrate heterogeneous data sources and different algorithms. Weak learner diversification methods and combination rules with well explored properties are readily available in existing work on ensemble learning, and iterative algorithms can be used to optimize recommender accuracy as additional observations of user preferences becomes available.

3 A HYBRID MUSIC RECOMMENDER

In this paper we focus on combining recommender output. We investigate to what extend commonly used social and content-based recommenders provide sufficiently diverse output to improve recommendations by combining them without applying diversification methods. A frequently used social recommender method is item-based collaborative filtering [4]. Using this method, first the similarity between two items i and j is computed using preference data of all users $u \in U$ considered. The Pearson correlation coefficient is frequently used to determine this similarity:

$$s(i,j) = \frac{\sum_u \left(R_{u,i} - \bar{R}_i\right)\left(R_{u,j} - \bar{R}_j\right)}{\sqrt{\sum_u \left(R_{u,i} - \bar{R}_i\right)^2 \sum_u \left(R_{u,j} - \bar{R}_j\right)^2}} \quad (1)$$

for all users $u \in U$, where $R_{u,i}$, $R_{u,j}$ are preferences of user u for items i and j respectively, and $\bar{R}_i$, $\bar{R}_j$ are mean preferences for these items. Then, a predicted preference $P_{u,i}$ for a target item i is computed as the weighted sum of preference values of user u for all items $j \in J$ that are correlated to item i and for which the user i's preferences are known, scaled by the sum of similarity terms:

$$P_{u,i} = \frac{\sum_j s(i,j) R_{u,j}}{\sum_j s(i,j)} \quad (2)$$

A content-based recommender can be implemented similarly to this social recommender algorithm. We first construct a similarity matrix of songs using a song similarity measure introduced in [5]. Analogous to the social recommendation algorithm described above, recommendations are computed as given in Formula 2 using the referenced content-based similarity function $s(i, j)$ instead of the correlation of user preference data.

A combination rule using *decision templates* [2] is applied to combine the base recommender output. A decision template is a record of weak learner estimates for an observed value. It can be expressed as a decision rule in the form $\{P_{r_1}, ..., P_{r_n}\} \rightarrow R$ for a set of base recommenders $\{r_1, ..., r_n\}$, where P_{r_n} is the prediction of recommender n and R is the observed preference for a particular recommendation. For classification, decision templates are created as averaged class probabilities over all instances classified into a particular class. This method has been shown to perform well for different classification tasks [2]. For the regression task of predicting preference values, we apply a variant of this method. Decision template rules are retained without averaging them for each user. To combine base recommenders for a new instance, their individual predictions are computed. Then the Euclidean distance of the vector of computed results to each stored decision template rule head is computed, and the nearest neighbor rule's tail (the observed preference) is the resulting preference prediction.

4 EXPERIMENT

We conducted a preliminary experiment to evaluate the performance of the described hybrid recommender. For this a music dataset containing a collection of 63,949 popular music pieces was used. For each song manually assigned metadata are provided and perceptual audio features were extracted for the content-based recommender described in Section 3. The dataset contains 1,139,979 *play counts*, the number of times that a song has been played by a user, for 6,939 participants. Participants that played less than 100 distinct songs, played less than 50 songs more than once, or played no song at least 10 times were discarded for the experiment to ensure that sufficient test and training data for cross-validation were available. After this procedure 735 participants remained. In order to avoid bias introduced by large play counts, these were capped at a value of 10.

The evaluation task was to predict play counts for songs in the range [1,10] for each of the 735 users. For each user, 10-fold cross validation was used to split the available play counts into training and validation sets. The Mean Absolute Error (MAE) and the standard deviation σ of errors, averaged over all users $u \in U$ and scaled by the total count of evaluated instances n were computed for each recommender r. The MAE indicates the mean deviation of predicted from observed values and is the most frequently used recommender algorithm evaluation measure [4].

$$MAE_r = \frac{1}{n} \sum_u \sum_i \|R_{u,i} - P_{r,u,i}\| \qquad (3)$$

In this initial experiment, the item-based social recommender reached a MAE of 2.608 (2.736) and σ 3.374, the content-based recommender reached a MAE of 2.884 (2.993) and σ 3.651, and the hybrid recommender a MAE of 2.349 [1] and σ 2.817. We attribute the higher measured MAE for the social recommender compared to MAE values for other, ratings-based datasets (such as [4]) to the usage of implicit listening data and the high sparsity of the dataset. The presented hybrid recommender reduces the MAE by 0.259 compared to the best performing base recommender.

5 CONCLUSION AND FUTURE WORK

We have introduced a hybrid recommender that uses a combination rule adapted from ensemble learning methods to combine base recommenders. In a preliminary experiment performed on observed listening data the technique leads to a MAE that is about 10% smaller than that of the best base recommender. This implies that a degree of diversity between social and content-based recommenders exists that can be exploited. Future work will focus on integrating more hetereogeneous data sources, applying diversification techniques on them, developing specifically suited combination rules for hybrid music recommenders, and collecting explicit user ratings for a comparative evaluation of such approaches.

6 REFERENCES

[1] R. Burke. "Hybrid Recommender Systems: Survey and Experiments." *User Modeling and User-Adapted Interaction, Volume 12, Issue 4, 2002, 331-370.*

[2] L. Kuncheva, J. Bedzek and R. Duin. "Decision Templates for Multiple Classifier Fusion: an Experimental Comparison" *Pattern Recognition, Volume 34, Issue 2, 2001, 299-314.*

[3] R. Polikar. "Ensemble Based Systems in Decision Making." *IEEE Systems Magazine, Issue 3, 2006, 21-45.*

[4] B. Sarwar, G. Karypis, J. Konstan and J. Riedl. "Item-based Collaborative Filtering Recommendation Algorithms." *Proc. 10th International World Wide Web Conference, 2001.*

[5] F. Vignoli and S. Pauws. "A Music Retrieval System Based on User-Driven Similarity and its Evaluation." *Proc. Sixth International Symposium on Music Information Retrieval, 2005.*

[1] MAE values for the base recommenders are rounded to the nearest possible play count to compensate for the inherent rounding effect of the decision template combination rule; original values are given in brackets.