Towards Ensemble Learning for Hybrid Music Recommendation

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

ABSTRACT

We investigate ensemble learning methods for hybrid music recommender algorithms, combining a social and a content-based recommender algorithm as weak learners by applying a combination rule to unify the weak learners' output. A first experiment suggests that such a combination can already reduce the mean absolute prediction error compared to the weak learners' individual errors.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Correlation and regression analysis; H.3.3 [**Information Search and Retrieval**]: Information filtering

General Terms

Algorithms, Experimentation

1. INTRODUCTION

Music recommender systems that are commercially used today apply one of two recommendation approaches: *social* and *content-based recommenders*. *Social recommenders* predict preferences based purely on user preference data. *Content-based recommenders* predict preferences based on extracted audio features or using catalogue metadata as descriptive features. *Hybrid recommenders* combine recommendation approaches to improve the overall accuracy of predictions and reduce problems of specific approaches. To this end, in many applications hybrid recommenders combine social and content-based recommendation in one of a number of possible configurations [2]. Hybrid recommenders can be categorized into ones that 1) integrate social and content-based recommendation in a single algorithm or 2) combine output of independent algorithms. We follow the second approach, since our goal is to flexibly integrate many distinct independent recommenders (referred to as *weak learners* in this paper) in a generalized hybrid recommender to which new recommender algorithms can be added with no or minimal changes. To this end, ensemble learning methods can be applied to create diverse weak learners and to unify their output using established or novel combination methods.

RecSys'07, October 19–20, 2007, Minneapolis, Minnesota, USA.

2. APPLYING ENSEMBLE LEARNING

In *ensemble learning* [4], several weak learners are created and used in regression or classification tasks. The weak learners' output is combined into a unified output by applying a combination rule. Ensemble learning methods can outperform single strong learners in many practical applications. However, ensemble learning methods can only be successfully applied when the used weak learners are sufficiently *diverse* – weak learners must be able to correct errors made by other weak learners in an ensemble. Hence many ensemble learning methods actively select or create suitably diverse weak learners with techniques such as training data resampling. Incremental ensemble learning methods have also been studied, and incremental variants of established ensemble learning methods can often be derived easily. Ensemble learning is a promising concept for hybrid music recommendation. It can be applied as a *data fusion* technique to flexibly integrate heterogeneous data sources as well as different classification or regression algorithms. Weak learner diversification methods and combination rules with well-explored properties are readily available in existing work, and incremental algorithms can be applied to successively increase recommender accuracy as additional observations of user preferences become available.

3. A HYBRID MUSIC RECOMMENDER

In this paper we focus on combining recommender output. We investigate whether and, if so, to what extent unmodified social and content-based recommenders that are used in practical applications provide output that is sufficiently diverse to lead to a reduction in recommendation error when their output is combined using a suitable combination rule. If this is the case, these recommenders can be used as weak learners in learner ensembles.

A frequently used social recommender technique is *item-based collaborative filtering* [1], where first the similarities of all items in an item set are computed using the preference data of all users available. The *Pearson correlation coefficient* can be used to determine the similarity of two items i and j:

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

for all users $u \in U$, where $R_{u,i}$, $R_{u,j}$ denote preferences of user u for items i and j respectively, and $\bar{R}_i$, $\bar{R}_j$ denote mean preferences for these items. Then, the 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 that are correlated to item i and for which the user u's preferences are known, scaled by the sum of similarity terms:

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

We implemented a content-based music recommender that performs the same two steps outlined for the item-based collaborative filtering algorithm. For the content-based recommender, however, item similarities are determined using a song similarity measure described in [5]. This measure combines catalogue metadata and features extracted directly from audio data to determine song similarities. With the similarity matrix obtained in this way, recommendations are then computed analogous as described in Formula 2.

A *decision template* based combination rule [3] is applied to unify the weak learners' output. A decision template is a record of weak learner estimates for an observed value and of the observed value itself. It can be expressed as a decision rule in the form $\{P_{r_1}, ..., P_{r_n}\} \rightarrow R$ for a set of weak learners $\{r_1, ..., r_n\}$, where P_{r_n} is the estimate of the nth learner and R denotes the actually observed value. For classification tasks, a decision template for a class consists of a single decision rule, which is created by averaging over all available rules for that class. For regression tasks, all observed rules can be stored and used directly without aggregating them. In our implementation, all observed decision rules of a user are stored in a set of decision rules for this user. To predict a value for a new instance, the weak learners' individual estimates for the instance are computed and stored in a vector $\{P_{r_1}, ..., P_{r_n}\}$. Then, the Euclidean distance of this vector to each stored decision rule head is computed, and the nearest neighbor rule's tail (the observed value) is used as unified estimate. Decision templates have been shown to generally perform well for a variety of classification tasks [3].

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, metadata and perceptual audio features for the content-based recommender described in Section 3 were made available. The dataset contains 1,139,979 *play counts* denoting the number of times that a song has been played by a user for a set of 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 task performed was to predict play counts for songs for each of these 735 users in the range $\{1, 10\}$. Training and validation sets were created with 10-fold cross validation. The Mean Absolute Error (MAE), averaged over 10 validation folds per user, all users and scaled by the total count of evaluated instances n was computed for each recommender r. The MAE indicates the averaged deviation of predicted values from observed values and is the most frequently used evaluation measure for recommender algorithms [1].

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

In this initial experiment, the item-based social recommender achieved a MAE of 2.608 (2.736), the content-based recommender reached a MAE of 2.884 (2.993), and the hybrid recommender a MAE of 2.349[1]. The implicit listening data used for this experiment must be considered to be noisier input data than explicit ratings, which may explain why the MAE of the item-based collaborative filtering algorithm is higher than in other experiments on ratings-based datasets [1]. The presented hybrid recommender reduces the MAE by 0.259 - about 10% - compared to the best performing weak learner. This indicates that a combination of weak learners using this combination rule can indeed reduce recommendation errors.

5. CONCLUSION AND FUTURE WORK

We have introduced a hybrid recommender that uses a combination rule adapted from ensemble learning methods to combine different recommender algorithms as weak learners. A preliminary experiment performed on observed listening data indicates that this technique may be suitable to reduce the MAE compared to the weaker learners. 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 heterogeneous 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 in the music domain. Additionally, we are reproducing experiments with well-known reference datasets from the movie domain for comparability of results.

6. REFERENCES

[1] J. K. B. Sarwar, G. Karypis and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. 10th International World Wide Web Conference*, 2001.

[2] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.

[3] J. B. L. Kuncheva and R. Duin. Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, 34(2):299–314, 2001.

[4] R. Polikar. Ensemble based systems in decision making. *IEEE Systems Magazine*, (3):21–45, 2006.

[5] F. Vignoli and S. Pauws. A music retrieval system based on user-driven similarity and its evaluation. In *Proc. Sixth International Symposium on Music Information Retrieval*, 2005.

[1] MAE values for weak learners are rounded to the nearest integer value to compensate for inherent rounding effects of the decision template combination rule; unmodified MAE values are given in brackets.