WAVELET-BASED MUSIC RECOMMENDATION

Claudio Biancalana, Fabio Gasparetti, Alessandro Micarelli, Alfonso Miola and Giuseppe Sansonetti
Department of Computer Science and Automation, ROMA TRE University, Via della Vasca Navale 79, Rome, Italy
{*claudio.biancalana, gaspare, micarel, miola, gsansone*}*@dia.uniroma3.it*

Keywords: Wavelet, Music Recommendation, Signal Processing

Abstract: Recommender Systems provide suggestions for items (e.g., movies or songs) to be of use to a user. They must take into account information to deliver more useful (perceived) recommendations. Current music recommender takes an initial input of a song and plays music with similar characteristics, or music that other users have listened to along with the input song. Listening behaviors in terms of temporal information associated to ratings or playbacks are usually ignored. We propose a recommender that predicts the most rated songs that a given user is likely to play in the future analyzing and comparing user listening habits by means of signal processing techniques.

Recommender systems provide suggestions based on user preferences in order to recommend items likely to be of interest to a user. It is obvious that user preferences are influenced by the current context, such as the current time of the day, mood, or current activities. Nevertheless, a few recommender systems explicitly include this information in the preference models.

A special group of recommender systems are the ones based on the collaborative approach (Resnick et al., 1994; Shardanand and Maes, 1995; Breese et al., 1998). The system generates recommendations using only information about rating profiles for different users. Collaborative systems locate peer users with a rating history similar to the current user and generate recommendations using this neighborhood.

Collaborative filtering (CF) systems have been successful in several recommender systems. The availability of large datasets and additional information that is easy collectable from the web, makes this task interesting.

There are several issues that do not allow us to directly apply the traditional CF approach for music recommendation. The space of possible items (i.e., tracks) can be very large and, similarly, the user space can also be enormous. Often user ratings are not available or they cover only a small subset of the user library of songs. Moreover, when new users enter to the system or new songs are added to the global library, it is not possible to provide any recommendation to them due to the lack of any preference information (the so known cold-start problem). There is no chance to use taxonomies or ontologies to represent the new items and facilitate the clustering as happens in different domains (e.g., (Acampora et al., 2010a; Micarelli et al., 2009)) Content-based approaches collect information describing the items and then, based on the user preferences, they predict which tracks the user might enjoy (see for example the Pandora service[1]). The key component of this approach is the similarity function among the songs. Nevertheless, there is a strong limitation of the highlevel descriptors that can be automatically extracted from the tracks (Celma, 2010).

One more relevant issue that traditional CF approaches do not take into consideration is the listening behavior of the user in terms of temporal information. The timestamp of an item (i.e., when the song song is played) is an important factor for the recommendation algorithm. Usually, the prediction function treats the older items as less relevant than the new ones, but any further reasoning about the temporal information is simply ignored.

In this paper, we discuss a recommendation approach based on signal processing. In particular, a traditional CF approach is enhanced considering an improved similarity function between users. The user listening habits are represented by signals. Wavelet theory is used to study the related time-frequency representations of signals and draw similarity between listening behaviors. Signal processing techniques are not employed to extract features from the songs, but for representing and comparing those behaviors in or-

[1] www.pandora.com

der to group similar users together. This is the novelty of the approach in comparison with the current literature.

The rest of this paper is organized as follows. Section 1 briefly introduces some related studies on music recommendation. Section 2 details our proposed approaches. Last, in Section 3 a brief account of the testbed we are developing for the evaluation is given. Conclusions close the discussion.

1 RELATED WORK

Many algorithms have been developed to address the personalized recommendation problem. Content-based approaches aim at including different sources of information (Semeraro et al., 2009; Groh and Ehmig, 2007; Micarelli et al., 2006) or better modelling the user interests (Gasparetti and Micarelli, 2007). User-based collaborative filtering (CF) is widely used, and the main idea is to find the items liked by other people with similar taste. Different from the user-based CF, the item-based CF recommends the items which are similar with the user's collected items (Schafer et al., 2007). Context-aware high-level frameworks (e.g., (Acampora et al., 2010b; Gaeta et al., 2009)) are not easily adaptable to this specific domain because of the peculiar characteristic of the items. For example, in (Biancalana et al., 2011a) the authors devise a neural network context-aware recommender extracting different features from point of interests. In the music scenario, techniques that automatically extract features from the played songs are not easily conceivable.

As for music recommendation, the most compressive survey on the literature is to be found in (Celma, 2010). The author groups the recommendation approaches in four categories: (1) collaborative filtering, based on explicit or implicit feedbacks; (2) content-based filtering, by means of manual or automatic feature extraction; (3) context-based filtering, the take advantage of potential user tags associated to each single song; and (4) hybrid approaches that combine more then one of the above-mentioned ones.

To the best of our knowledge, there are currently no attempts to include temporal behavior in user habits in the music recommendation task. A preliminary attempt has been suggest for the movie domain in (Biancalana et al., 2011b). The proposed approach can be categorized as context-based, where the similarity of different songs is evaluated according to the implicit listening behavior that the user exhibits.

2 WAVELET-BASED RECOMMENDATION

Traditional user-based CF approaches relies on similar users which have similar rating patterns, that is, the prediction of a rating $r_{u,s}$ by user u for the track $track_k$ is evaluated as an aggregate of the rating of some other users for the same item $track_k$. We call these similar users *neighbors*. If a user v is similar to a user u, we say that v is a neighbor of u. User-based algorithms generate a prediction for a track $track_k$ by analyzing ratings for $track_k$ from users in u's neighborhood.

In order to draw the distance (or similarity) between two users, the Pearson correlation coefficient is usually employed (Resnick et al., 1994):

$$sim(u,v) = \frac{\sum_{s \in S_{u,v}} (r_{u,s} - \overline{r}_u)(r_{v,s} - \overline{r}_v)}{\sqrt{\sum_{s \in S_{u,v}} (r_{u,s} - \overline{r}_u)^2 \sum_{s \in S_{u,v}} (r_{v,s} - \overline{r}_v)^2}} \tag{1}$$

where $S_{u,v}$ denotes the set of co-rated items between u and v, $r_{u,s}$ is the rating of the user u for the item s, and $\overline{r}_u$ is the average of the ratings of the user u.

Pearson correlation ranges from 1.0 for users with perfect agreement to -1.0 for users with perfect disagreement. In this way, it is possible to generate a prediction of rating for the user u and the item s as follows:

$$pred(u,s) = \overline{r}_u + \frac{\sum_{v \in NNu} sim(u,v)(r_{v,s} - \overline{r}_v)}{\sum_{v \in NNu} sim(u,v)} \tag{2}$$

where NNu is the set of users in the u's neighborhood.

The proposed recommendation approach is enhanced considering a user similarity function that analyzes contextual factors that are included in the data collected during the normal usage of the recommender system. In particular, the timestamp associated to playbacks.

In our recommender we employ Discrete wavelet transforms (DWT). The basic principles of wavelet theory were put forth in a paper by Gabor in 1945 (Gabor, 1946). In comparison with the Fourier transform, wavelets are localized in both time (or location) and frequency instead of just frequency. A wavelet is a function used to represent a time signal into different scale components. Usually one can assign a frequency range to each scale component. Each scale component can then be studied with a resolution that matches its scale. The DWT is computed by successive lowpass and highpass filtering of the discrete time-domain signal as shown in Fig. 1. This is called the Mallat pyramid algorithm, a computationally efficient method of implementing the wavelet transform.

The input signal is assumed to be a set of discrete-time samples, i.e., a sequence $x[n]$, where n is an integer. Whereas the basis function of the Fourier transform is a sinusoid, the wavelet basis is a set of functions. In our approach we decide to employ the popular Haar wavelets.

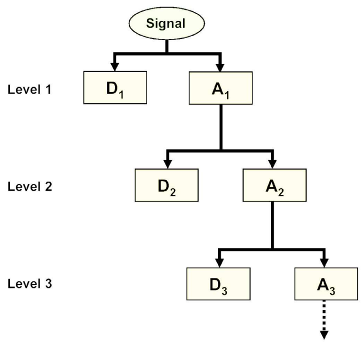

Figure 1: At each decomposition level, the half band filters produce signals spanning only half the frequency band.

For an input represented by a list of $2n$ numbers, the Haar wavelet transform may be considered to simply pair up input values, storing the difference and passing the sum. This process is repeated recursively, pairing up the sums to provide the next scale: finally resulting in $2n-1$ differences and one final sum. In essence, the obtained decomposition can be thought of as representing a frequency decomposition of the input.

The algorithm to compute the similarity between two users is based on the comparison of the Wavelet transforms obtained by the two signals related to the listening behavior of the users. In particular, the input dataset is composed of users, songs and tuples $< u_i, track_k, timestamp >$ that represent tracks of a library L listened by users in a given moment. The signal is built in the following way:

Two users are considered similar if they listen to the same songs in the same time of the day. If the two users listen to the same songs in a given period of time but this two periods do not coincide, e.g., the user u played some songs in January and v played the same songs in March, traditional comparison metrics return low similarity between the two users.

The Euclidean distance between the coefficients of the two wavelets allows us to ignore potential time shifts between listening behaviors. Moreover, it takes into account the frequency of items, i.e., the times a song has been played in a given period. In this way, we are able to recognize similarities between user habits analyzing different scales or approximations of the input signals produced by the wavelet tree decomposition. This is a well-known characteristic of the wavelet theory that is exploit several times in the Content-based image retrieval domain.

Algorithm 1 Similarity between users u and v

for all $track_k \in L$ **do**
  $vv_{u,k}[t_h] \leftarrow$ number of times the song $track_k$ has been played by the user u in the time interval $t_h < t < t_h + \delta T$
  $vv_{u,k}[t_h] \leftarrow$ number of times the song $track_k$ has been played by the user v in the time interval $t_h < t < t_h + \delta T$
end for
for all $track_k \in L$ **do**
  $w_{u,k} \leftarrow$ discrete Haar Wavelet transform of signal $vv_{u,k}$
  $w_{v,k} \leftarrow$ discrete Haar Wavelet transform of signal $vv_{v,k}$
end for
$sim_{u,v} \leftarrow$ Euclidean distance between the two vectors w_u and w_v

3 HOW TO EVALUATE

We are currently devising a testbed that include enough real usage data extracted by a public domain datasets, i.e., Last.fm Dataset - 1k[2], in order to compare the performance with traditional recommender approaches by means of standard evaluation measures. That dataset contains tuples in the following form:

$$< user, timestamp, artist, song > \quad (3)$$

collected from Last.fm website that correspond to one song played by a specific user. There are 992 unique users and more than 19 Million entries, therefore enough data to represent user listening habits. The pair $< artist, song >$ is easily mapped to the $track$ variable used in the wavelet-based recommender. We will looking to standard measures of prediction between the recommended songs and the actual songs played by the users.

4 CONCLUSIONS

In this paper, we have discussed a recommender approach based on signals related to the user listening behavior.

[2] last.fm

Further work will be done along several research directions. Some factors that should be included in the recommendation process are the novelty of songs and the user authority. New songs have a higher potential of being interesting than old songs. Moreover, some collaborative approaches have tried to diversify the ratings from users, identifying more authoritative users that should be taken more into consideration when predictions have to be suggested. Serendipity and negative preferences are further factors that music recommenders should include in their predictive analysis (Iaquinta et al., 2008; Musto et al., 2011).

REFERENCES

Acampora, G., Gaeta, M., and Loia, V. (2010a). Exploring e-learning knowledge through ontological memetic agents. *Comp. Intell. Mag.*, 5:66–77.

Acampora, G., Gaeta, M., Loia, V., and Vasilakos, A. V. (2010b). Interoperable and adaptive fuzzy services for ambient intelligence applications. *ACM Trans. Auton. Adapt. Syst.*, 5:8:1–8:26.

Biancalana, C., Flamini, A., Gasparetti, F., Micarelli, A., Millevolte, S., and Sansonetti, G. (2011a). Enhancing traditional local search recommendations with context-awareness. In Konstan, J. A., Conejo, R., Marzo, J. L., and Oliver, N., editors, *User Modeling, Adaption and Personalization - 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings*, volume 6787 of *Lecture Notes in Computer Science*, pages 335–340. Springer.

Biancalana, C., Gasparetti, F., Micarelli, A., Miola, A., and Sansonetti, G. (2011b). Context-aware movie recommendation based on signal processing and machine learning. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, CAMRa '11, pages 5–10, New York, NY, USA. ACM.

Breese, J. S., Heckerman, D., and Kadie, C. M. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In Cooper, G. F. and Moral, S., editors, *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 43–52.

Celma, . (2010). *Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space.* Springer.

Gabor, D. (1946). Theory of communication. *J. Inst. Elect. Eng.*, 93:429–457.

Gaeta, A., Gaeta, M., and Ritrovato, P. (2009). A grid based software architecture for delivery of adaptive and personalised learning experiences. *Personal Ubiquitous Comput.*, 13:207–217.

Gasparetti, F. and Micarelli, A. (2007). Personalized search based on a memory retrieval theory. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI): Special Issue on Personalization Techniques for Recommender Systems and Intelligent User Interfaces*, 21(2):207–224.

Groh, G. and Ehmig, C. (2007). Recommendations in taste related domains: collaborative filtering vs. social filtering. In *GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work*, pages 127–136, New York, NY, USA. ACM.

Iaquinta, L., de Gemmis, M., Lops, P., Semeraro, G., Filannino, M., and Molino, P. (2008). Introducing serendipity in a content-based recommender system. In Xhafa, F., Herrera, F., Abraham, A., Köppen, M., and Benítez, J. M., editors, *International Conference on Hybrid Intelligent Systems (HIS 2008*, pages 168–173. IEEE Computer Society.

Micarelli, A., Gasparetti, F., and Biancalana, C. (2006). Intelligent search on the internet. In Stock, O. and Schaerf, M., editors, *Reasoning, Action and Interaction in AI Theories and Systems*, volume 4155 of

Lecture Notes in Computer Science, pages 247–264. Springer.

Micarelli, A., Sciarrone, F., and Gasparetti, F. (2009). A case-based approach to adaptive hypermedia navigation. *IJWLTT*, 4(1):35–53.

Musto, C., Semeraro, G., Lops, P., and de Gemmis, M. (2011). Random indexing and negative user preferences for enhancing content-based recommender systems. In Huemer, C. and Setzer, T., editors, *E-Commerce and Web Technologies*, volume 85 of *Lecture Notes in Business Information Processing*, pages 270–281. Springer.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, CSCW '94, pages 175–186, New York, NY, USA. ACM.

Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). Collaborative filtering recommender systems. In Brusilovsky, P., Kobsa, A., and Nejdl, W., editors, *The Adaptive Web: Methods and Strategies of Web Personalization*, volume 4321 of *Lecture Notes in Computer Science*, pages 291–324. Springer Berlin, Heidelberg, Berlin, Heidelberg, and New York.

Semeraro, G., Lops, P., Basile, P., and de Gemmis, M. (2009). Knowledge infusion into content-based recommender systems. In Bergman, L. D., Tuzhilin, A., Burke, R. D., Felfernig, A., and Schmidt-Thieme, L., editors, *Proceedings of the 2009 ACM Conference on Recommender Systems, RecSys 2009*, pages 301–304. ACM.

Shardanand, U. and Maes, P. (1995). Social information filtering: algorithms for automating word of mouth. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '95, pages 210–217, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.