Can a Rock song have a Jazz audience? Relationship between folksonomy and collaborative filtering in music recommender systems.

Eugenio Tacchini
Università degli Studi di Milano
Via Bramante, 65
26013, Crema (CR), Italy
eugenio.tacchini@unimi.it

Ernesto Damiani
Università degli Studi di Milano
Via Bramante, 65
26013, Crema (CR), Italy
ernesto.damiani@unimi.it

ABSTRACT
In this paper we investigate the relationship between a folksonomy-based music classification and a music classification based on collaborative filtering, i.e. on the users' listening behavior. We found a correlation between folksonomy-based songs clustering and clustering computed using methods based on the audience listening behaviour and, using a combination of the two approaches, we also computed the eclecticism level of a sample set of users, finding that eclecticism seems to be a characteristic which changes according to the genre of music most loved by a user.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – Clustering, Information filtering.

General Terms

Keywords
Collaborative filtering, recommender systems, music, eclecticism, clustering, folksonomy, music genres.

1. INTRODUCTION
Music folksonomies has become a very important part of the social music services ecosystem. Thanks to folksonomy, final users can describe artists and songs exactly as they perceive them (regardless of a priori classifications made by experts). Folksonomy tags have been used, alone or together with other data, as content metadata for recommender systems [1, 2].

In particular, music recommendation services often rely on folksonomy tags as a description of an artist/song as seen through the eyes of a community and such a description can be exploited to provide better recommendations.

In this paper we investigate the relationship between a folksonomy-based music classification and a music classification based on collaborative filtering, i.e. on the users' listening behavior. Our first research question can be stated as follows: "Does folksonomy-based songs clustering compute the same clusters as methods based on the audience listening behaviour?". Getting a clear idea of the relationship between folksonomy-based and collaborative filtering music classification is important not only to understand the dynamics behind social music communities but also to help in the process of combining the two classification components in the context of hybrid recommender systems.

We then used a combination of the two classification methods analyzed in order to compute the eclecticism level of a sample set of users (i.e. the extent to which users listen to songs apparently far from their taste), coming to a second research question: “Is the eclecticism level of a user dependent on the music mostly loved by that user?”, to which we tried to answer in the last part of this paper.

2. RELATED WORK
Music folksonomies has already been studied in the past: an example can be found in [3], where the agreement between expert-based music genres vocabularies and folksonomies was analyzed, coming to the conclusion that, while for some genres experts and wisdom of crowds agree, for other genres clearly disagree.

If we consider our first research question, the closest work is [4], where the authors investigated whether the tagging of artists is coherent with the artist similarities found with collaborative filtering techniques; however, in order to measure the artists similarities, they relied on the “similar artists” feature of last.fm: while the feature algorithm is not public, it is likely that it, at least nowadays (years ago probably the algorithm was different), it is not a pure collaborative filtering approach and the folksonomy itself is exploited in order to provide artists similarities. We instead calculate the collaborative-filtering similarities at the song level (not at the artist level) and we use the well-known k-means clustering algorithm.

To the best of our knowledge, there hasn't been any previous computer science work that addresses our second research question.

3. METHODOLOGY
3.1 Dataset used
The dataset used for the experiments contains the whole listening history of 992 unique users of the social radio last.fm1. The

1 http://www.last.fm
dataset has been built by Oscar Celma and is available for download for non commercial use².

In particular, the dataset contains <user, timestamp, artist, song> tuples (nearly 19 millions) collected using the user.getRecentTracks() method of the last.fm APIs³ and provides data until May 5th, 2009. Furthermore, for most of the tracks and the artists, some heuristics has been applied by Celma to artist names and track titles in order to assign the corresponding MBID, a unique identifier provided by the MusicBrainz project⁴.

3.2 Data pre-processing, music genre assignment

For the purposes of this paper, we chose to keep only the tuples having both an artist MBID and a track MBID; this provide, while analyzing the data, an higher level of confidence respect to relying just on the artist names and track titles, which can contain mistakes (some of them are directly taken from the audio files played locally by the users) and consequently refer to the same item while having a different name or title. After the cleaning procedure, from the original dataset of nearly 19 millions tuples, nearly 17 millions remained, for a total of 961,416 unique songs and 83,982 unique artists.

In order to classify the songs by genres, we decided to use the folksonomy provided by the last.fm users. For each song, we downloaded the tag cloud associated to the corresponding artist using the method artist.gettoptags of the last.fm APIs⁵. While this approach is not perfect, because artist.gettoptags can't distinguish between homonyms, in our opinion is still preferable for this work respect to relying on a classification made by experts because we can get a the description of the artist directly through the eyes of the community of users who belong to the audience of that artist.

Since the tags linked to an artist can be freely chosen by the users, some of the tags don't semantically describe a music genre: “The Beatles”, for example, have both “classic rock” (clearly a music genre) and “the best” (an opinion about the artist) in their tag clouds. Since we can obtain, for each tag belonging to a tag cloud, the number of users who used that tag for an artist, we can easily extract the most used tag from the tag cloud and use that one as the main music genre; we don't have precise figures about this, but we are strongly confident that, in most of the cases, this really represent a music genre.

While we know that an artist can actually play different music genres, we think that setting the genre of a song as the genre of the relative artist can be a good compromise between accuracy and the additional complexity of retrieving the tag cloud for each song through the track.getTopTags method of the last.fm APIs⁶ and evaluate using some heuristics if the tag cloud of a song is rich enough in order to determine the corresponding music genre: the tag cloud of a track is in fact usually poorer respect to the tag cloud of an artist.

After the tag retrieving procedure we managed to determine a genre for 78,336 artists. We computed a popularity score for each genre in the dataset as the sum of the number of unique song

² http://www.dic.upf.edu/~ocelma/MusicRecommendationDataset/lastfm-1K.html
³ http://www.last.fm/api/show?service=278
⁴ http://musicbrainz.org/
⁵ http://www.last.fm/api/show?service=288
⁶ http://www.last.fm.it/api/show?service=289

belonging to a genre which has been played by each user; in figure 1 a generated tag cloud for the top 35 genres is displayed.

3.3 Clustering technique

A selected small number of representative, well known and semantically expressive genres have been selected: classic rock, pop, hip-hop, jazz, metal. The five genres have also been selected for having a comparable popularity score, respectively: 125,870, 99,288, 110,154, 91,829 and 94,369. All and only the tuples relative to those five genres have been selected to fill a matrix M whose rows represent the songs, whose columns represent all the 992 users and whose elements represent the number of time that a song has been played by a user. Note that, for the “metal” genre, the three genres "trash metal", “heavy metal” and “black metal” have been aggregated in order to get a weight comparable to the one of the other four genres.

Each song can therefore be described as a vector of property values, where each user is a property and each playcount value the corresponding property value. When taken together, these vectors compose the usual users/items matrix typically used by collaborative filtering recommender systems. In our case, each playcount value represents an implicit vote that a user gives to an item (song).

Since the resulting matrix was very sparse, we decided to keep just the rows having a minimum of 5 elements > 0, i.e. the songs which have been played at least by 5 users.

The final matrix had 19,760 rows and 992 columns and has been loaded as a text file using the software Cluster 3.0⁷, in order to clusterize the songs. The well known k-means algorithm has been chosen and a preliminary normalization step has been completed by Cluster 3.0 before starting the actual clusterization process: all values in each row of data has been multiplied by a scale factor S in order to get 1.0 as the sum of the squares of the values in each row. The number of chosen clusters (k) was five, in order to verify the correspondence with the five genres selected, and the similarity measure chosen was the Pearson correlation.

4. RESULTS

4.1 Relationship between clustering and music genres

After having clustered the songs according to how much the users listened to them, we analyzed the composition of the resulting

⁷http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm
clusters in order to understand if there were relationships with the music genres. The data are reported in table 1.

Table 1. Composition of the resulting clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Songs count</th>
<th>Genres composition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dominant genre</td>
</tr>
<tr>
<td>1</td>
<td>4,487</td>
<td>metal 88%</td>
</tr>
<tr>
<td>2</td>
<td>4,748</td>
<td>classic rock 69%</td>
</tr>
<tr>
<td>3</td>
<td>4,529</td>
<td>hip-hop 86%</td>
</tr>
<tr>
<td>4</td>
<td>2,773</td>
<td>jazz 80%</td>
</tr>
<tr>
<td>5</td>
<td>3,223</td>
<td>pop 86%</td>
</tr>
</tbody>
</table>

The resulting clusters are quite balanced in terms of the total number of songs, and the data clearly shows that, for each cluster, there is a dominant genre: in four cases over five, at least the 80% of the songs belong to the same genre. Since our clustering is based on the the distance among songs calculated according to the users listening behavior, the approach can be considered very similar to the first step usually processed by a classical item-based collaborative filtering algorithm in the recommender systems domain; this suggests that there should be a strong relationship between the results of an item-based collaborative filtering approach and the results of a content-based (folksonomy) approach.

Analyzing the genre composition, we can also observe, for each genre, which are the “closest” genres. While most of the results are expected, e.g. we already suspected that heavy metal songs have something in common with classic rock songs and not much with pop songs, some others are quite surprising: hip-hop and jazz seem to be not so far from each others.

4.2 Users' eclecticism level

Our second research questions was about the eclecticism level of users: do most of the users play songs belonging to the same cluster or to different clusters? Which are the most eclectic category of users? To answer this question, we computed, for each user, the cluster most played, determining five different category of users (classic rock lovers, jazz lovers, pop lovers and so on); then, for each user category, we calculated how the belonging users are distributed in terms of clusters played among all the category of users.

In order to considered a cluster as “played” by a user, we decided to set a threshold (k) on the percentage of songs played by that user and belonging to that cluster respect to the total number of songs played by the same user. The reason is that having listened just a few songs from a cluster doesn't mean that a user likes that particular kind of songs: the plays could be “casual”, especially if we consider that the listening history provided by last.fm includes both the songs directly played by the users (e.g. from iTunes) and the songs played by the last.fm recommendation radio, which are not necessarily liked by the user. We computed the eclecticism level setting k as 5% and 10%; the results are presented in tables 2 and 3.

Table 2. Users' eclecticism level (k=5)

<table>
<thead>
<tr>
<th>Number of clusters played</th>
<th>Users' percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.28%</td>
</tr>
<tr>
<td>2</td>
<td>22.46%</td>
</tr>
<tr>
<td>3</td>
<td>28.41%</td>
</tr>
<tr>
<td>4</td>
<td>27.69%</td>
</tr>
<tr>
<td>5</td>
<td>6.15%</td>
</tr>
</tbody>
</table>

Table 3. Users' eclecticism level (k=10)

<table>
<thead>
<tr>
<th>Number of clusters played</th>
<th>Users' percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26.56%</td>
</tr>
<tr>
<td>2</td>
<td>34.25%</td>
</tr>
<tr>
<td>3</td>
<td>27.08%</td>
</tr>
<tr>
<td>4</td>
<td>11.38%</td>
</tr>
<tr>
<td>5</td>
<td>0.72%</td>
</tr>
</tbody>
</table>

The results show that the 15.28% (26.56% with k=10) of the users seem to express a very low level of eclecticism; we can also notice that the percentage of very eclectic users (number of clusters played = 5) is very low: 6.15% (0.72% with k=10). It is important to highlight that this result can be considered as a eclecticism measurement relative to the genres and popularity threshold chosen: the users could actually have played songs belonging to other genres (respect to the five considered) and/or songs having a low level of popularity and consequently not selected (see 2.3 for further details). We also found that, as we are going to explain in details, the eclecticism level is not uniformly distributed among all the category of users.

Who are the less and the most eclectic users? To answer this question, we computed, for each user, the cluster most played, determining five different category of users (classic rock lovers, jazz lovers, pop lovers and so on); then, for each user category, we calculated how the belonging users are distributed in terms of eclecticism level. The results are presented in table 4, for k=5.

Table 4. Users' eclecticism level according to genres loved

<table>
<thead>
<tr>
<th>Genre loved</th>
<th>Percentage of users per eclecticism level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Metal</td>
<td>32.98%</td>
</tr>
<tr>
<td>Classic rock</td>
<td>6.82%</td>
</tr>
<tr>
<td>Pop</td>
<td>9.09%</td>
</tr>
<tr>
<td>Jazz</td>
<td>4.65%</td>
</tr>
</tbody>
</table>
The results show that, at least for the data considered, the Heavy Metal lovers are the less eclectic category of users (32.98% of them has eclecticism level 1, more than twice respect to the global percentage, 15.28%, showed in table 2), while the Jazz lovers seems to be the most eclectic ones, which percentage, for levels 3, 4 and 5, always over the ones showed by the other categories of users.

5. CONCLUSIONS AND FUTURE WORK

Using a the listening history of a sample set of last.fm users and the relative music content description provided by the last.fm folksonomy features, we analyzed the relationship between a folksonomy-based music classification and a collaborative filtering based music classification. We found that, for five music genres selected, there is an high correlation between the song clusters computed using the former approach and the ones computed using the latter.

Clustering the songs also allowed us to compute the eclecticism level of users - in term of diverse collaborative-filtering clusters frequently visited by users - and to find that some categories of users seems to be more prone to eclecticism respect to others e.g. Jazz lovers seems to be much more eclectic than Heavy Metal lovers.

We know that the results found could be dependent on the music genres selected; for this reason we have planned to further investigate this phenomenon, taking into account all the most popular genres represented in last.fm. This future extension of the study can also work as a base to develop a global collaborative filtering map of the music.

6. REFERENCES


