EVALUATION OF A RECOMMENDATION SYSTEM FOR MUSICAL CONTENTS

Luigi Lancieri, Mai Manguin, Stéphane Mangon

Orange Labs, {luigi.lancieri, mai.manguin, stef.mangon}@orange-ftgroup.com

ABSTRACT

This paper considers the evaluation of music recommendation systems. These systems aim at providing those pieces of music most likely to fit users' preferences. The main limits of existing systems are their lack of precision and the high level of resources necessary for musical characterisation and comparison. In this study, we propose an approach which has a good performance-resources ratio. The evaluation process shows that the recommendation system is able to predict the users' preferences with reasonable accuracy. By automatically ranking a list of 5985 unknown songs in order of the user's preference, our method is able to retrieve a known user's preferred song in the first 3.5 % positions of the ranked list.

***Index Terms*—** Music, signal, recommendation,

1. INTRODUCTION

Recommendation systems are often a cheap and effective way to differentiate relevant contents from a mass of information. Unfortunately, recommendation of multimedia content is still weighty to use. Indeed, it implies manual descriptions, or greedy resources in the case of automatic processing. Yet, the prospects are very interesting, given the increasing quantity of multimedia content available and market potential. The IFPI (International Federation of the Phonographic Industry) has announced that the total turnover of online digital music retail services was 10 times greater in 2004 than in 2003. Analysts are very confident and announce that this type of service is expected to generate revenues of the order of $330 million for 2006 [1].

The aim of this study is to assess the performance of a music recommendation system based on rhythm characteristics. The proposed pieces of music are selected on the basis of their similarity to a much-loved set of pieces reported by users. The similarity is estimated on the basis of rhythmic features extracted directly from the content by signal processing. The advantage of the rhythm characteristic that we describe in section 2 is that it can be calculated using less than 10% of the length of each piece of music.

Basically, the principle underlying any recommendation system is that there is a consistency in users' tastes and, more broadly, in users' activity. This implies that statistically, even in cases where a user has varying tastes, his fondness will never be distributed randomly. Without this assumption, it is impossible to anticipate a user preference other than by random selection. To test this hypothesis and validate the interest of our method, we have tried to show that the knowledge of the first half of a user's playlist of preferred songs (i.e. learning set) can help to identify the second half.

The objective of this study is twofold. Firstly, we want to evaluate the relevance of our method of musical characterization in a recommendation system. Secondly, we want to assess the impact of the quantity of music items in the learning set on the system performance. To carry out this study, we used a set of 200 users' opinions in the form of their playlist, each playlist consisting of 30 pieces of music indicated as favourites by the users. We split each playlist into a learning subset and a test subset, each set containing 15 pieces. For each user, the learning set allows us to calculate the music user profile that will be compared to all 5985 (200 x 30 – 15) pierces of music in order to rank them. The results show that for every user the test subset appears in the first 3.5 % of this classification.

This article is structured as follows. We start by giving a general description of our characterization method. Section 3 then gives the principles of our recommendation system. Sections 4 and 5 provide the details and results of the validation process, Section 6 going on to present the state of the art. Finally, the last section concludes and gives some details on future work.

2. MUSIC CHARACTERIZATION METHOD

The following figure shows the base of the signature process through the analysis of samples taken from a sound file. The idea is to capture the image of the swinging of the sound spectrum, as we perceive it while observing the bar graph of an audio reader. The samples to be analyzed were collected in triplets (E0, E1, etc.) of contiguous specimens of duration K. The space of time between each triplet was unspecified and the sum of all triplets covered a limited part of the file.

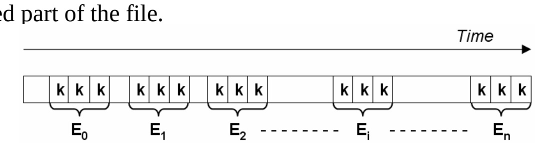

Figure 1 – Collection of samples in triplets *Ei* in a sound file

On each sample k of each triplet, we computed the distribution of frequencies using Fourier's method, and the directing coefficient p of the straight regression line binding the level (y) to each class of spectrum frequency (y). This straight regression line is expressed as: $y = px + b$.

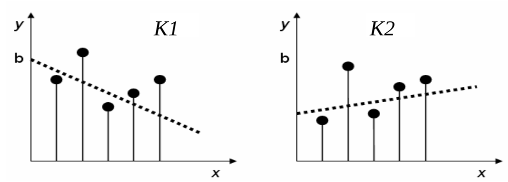

Figure 2 - The slope *p* of the frequency spectrum of 2 consecutive elements *K1* and *K2* of a triplet *Ei*.

The analysis of the behaviour of p (slope of the dotted line, in figure 2) contributes to the evaluation of rhythmic behaviour by measuring the swinging of the spectrum over one period Ei, and average value of the various periods (E1,..,En). By reference to mechanics, this swinging, its speed and its acceleration were calculated as follows.

The first stage consisted of identifying the number of triplets as well as their position in the signal. From a fraction of the sound file, we extracted the first triplet, for which we calculated the 3 spectra then the directing coefficients *Pi* of the straight regression lines. Thus, 3 slope values ($p1_1$, $p1_2$, $p1_3$) were obtained. The speed of swinging was obtained by calculating the difference between 2 consecutive slopes. We obtained 2 values of speed ($v1_1$, $v1_2$) for each triplet. Then, the acceleration *a1*, being a single value by triplet, was evaluated by computing the variation from the 2 speed values (v1, v2). We recomputed these data for the next triplet, and so on until the end of the file. At the end of the operation we had a set of values of coefficients [($p1_1$, $p1_2$, $p1_3$), ($p2_1$, $p2_2$, $p2_3$)… (pn_1, pn_2, pn_3)], speeds [($v1_1$, $v1_2$), ($v2_1$, $v2_1$)… (vn_1, vn_2)], and accelerations (a1, a2… an) for n triplets, representative of the piece of music. The behaviour of swinging (position, speed and acceleration) was obtained by a combination of the average values and standard deviation of all these data (pi, vi, ai). In our previous preliminary studies, we observed that the most representative of these values are both the vi and ai standard deviation [6]. We then used these 2 parameters identified as [Ev] and [Ea] in the next stage of this study. It should also be noted that the entire piece of music did not have to be covered by triplets. An important feature of our approach is that, even with a very low coverage, the characterisation remains good. Actually, we take less than 10 % (i.e. from 10 to 20 sec of music length) scattered over the whole piece of music.

3. RECOMMENDATION SYSTEM

Before presenting the validation process let us describe the main principles of the recommendation system as perceived by the user. The following picture shows the main modules of the system. We started with a set of music from which we computed a signal characterisation Ev, Ea, as described in the previous section. The user is first provided with randomly selected music from this database. Whilst listening to the music, the user rates it (appreciated or not). The appreciated music characterisation is then used to update his playlist and compute his profile. The latter is compared with the characterisation of the raw database set, in order to rank it according to best similarity to the user profile. After a certain time, the music provided to the user will be by order of best - ranked ratio.

At the grass-roots level, the characterization algorithm produces a 2-value vector for each piece of music (standard deviation for speed and acceleration [Ev, Ea]). The user profiles are composed from the characterization of all the contents appreciated.

Similarity can be measured between 2 pieces of music, 2 users or between a user and a piece of music. In the study that follows, we will compare the performance of various kinds of distances (Euclidean, Manhattan, Chi2.). To compare the similarity between a piece of music and the user profile, we will average the differences between the content to be compared and each component of the user profile. For example, if the user profile P1 is composed of 3 items [Ev1, Ea1]. [Ev2, Ea2]. [Ev3, Ea3], and if the content profile C4 = [Ev4, Ea4], the distance between the user and the content is equal to:

$$D\ (P1, C4) = \frac{D(1,4) + D(2,4) + D(3,4)}{3}$$

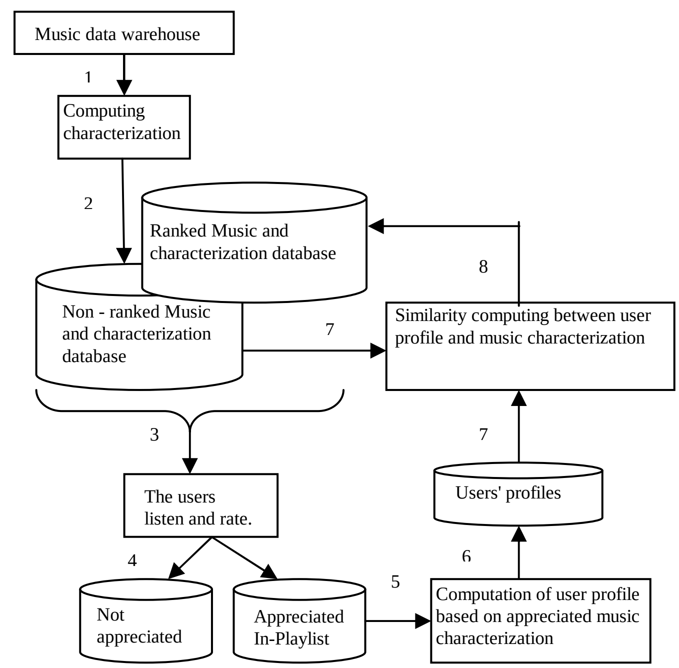

Figure 3 - Synopsis of recommendation system

From the user point of view, this system provides 3 listening functionalities: Random songs, Best of selected and recommended songs.

4. VALIDATION PROCESS

The method of validation aims at making sure that we find the pieces known to be preferred by the users in the best proposals of the recommendation system. For this purpose, we calculated the user profile with a part (learning set) of the contents of his playlist, and then we verified the effectiveness of the recommendation on the other part (test set) of his playlist. The learning set and the test set are then completely different. We wished also to evaluate the influence of the user profile size. Indeed, this profile was computed with 1, 2, 3…, 15 pieces of music.

Hence, if *Eu* is the set of 30 favourite songs (contents of the playlist) from user *u*, we started by building the user profile with a single favourite song (step 1), then 2 favourite songs (step 2), and so on up to 15 (half of the playlist).

At each step we compared the user profile with the entire set of music minus the user learning set ([200 x 30]-15=5985). This allowed us to rank the 5985 items by similarity level with the user profile. The smaller the distance, the better ranked the items are. For each user we identified the best rank. This means that we identified the rank of the item of his test subset playlist that was the best ranked. Finally, we calculated the average and the standard deviation of this best rank for all 200 users. All these operations were carried out at each of the 15 steps of the process.

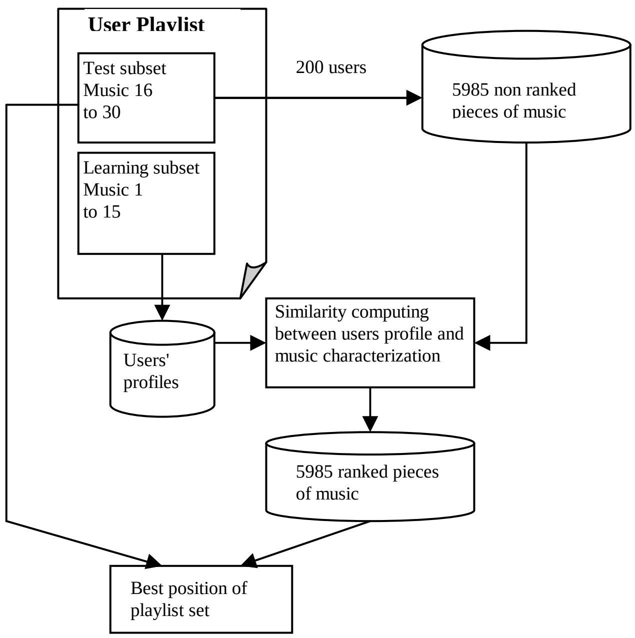

Figure 4 - Synopsis of evaluation process

Step 1: We calculated the user profile with one of the favourite songs from the 15 pieces of his learning playlist. We calculated the distance between this favourite song and the 5985 pieces in order to classify them by level of similarity. In this ranked list, we then determined the best classified music also contained in the 15 pieces of the user test set. We calculated the average of these 200 users' best ranks as well as the standard deviation.

Step 2: we calculated the user profile with two of the favourite songs from the 15 pieces of the learning playlist. We calculated the distance between these songs and the 5985 pieces in order to classify them. In this ranked list, we then determined the best classified music also contained in the 15 pieces of the user test set. As previously, we computed the average and the standard deviation.

We carried out these operations up to step 15, in order to evaluate the influence of the number of music pieces taken into account in the user profile.

5. RESULTS

We have a range of 200 users who each placed in their playlist 30 pieces of music matching their preferences. The number of unique pieces in the set of 6000 (30 x 200) was 5213, which corresponds to an average overlay rate of 13% between the 200 playlists.

The following results are derived from the calculation of the Euclidean distance. As we can see in the following figures, the performance of the recommendation system improves steadily as the number of pieces of music used in the user profile is increased from 1 to 8 (i.e. step 8). The average of the best ranked item evolves from 6.5 % (rank 390 over 5985) to 3.5% (rank 213). We can see that even with 1 song in the profile, the results already give an efficient selection. However, as we can see in the standard deviation curves (figure 6), a one-item profile dimension offers only a very low stability. That means that the results can be very bad for certain users even if the average for all users is reasonable. With 8 items in the user profile, both average rank and stability level are better. An interesting result is that more than 8 items improved neither rank nor stability. Since adding items to the profile does not improve performance though it consumes resources, the best solution is to limit to 8 the number of items used to compute the user profile.

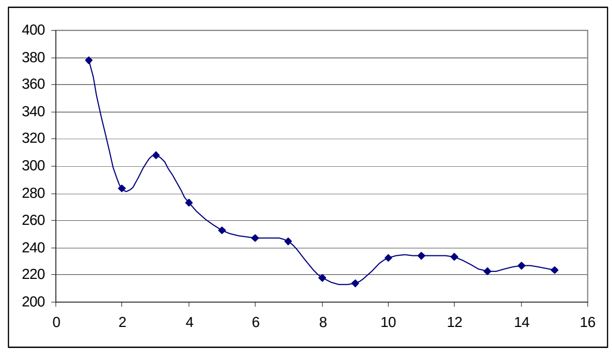

Figure 5 - Recommendation average best rank according to the size of the user profile.

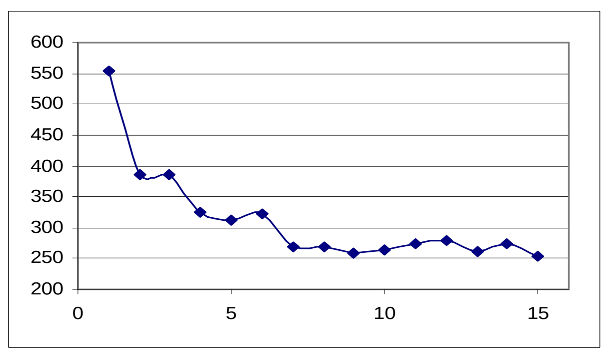

Figure 6 - Recommendation standard deviation according to the size of the user profile.

In this experiment, the Euclidean distance provides the better results. The best average rank using this distance is of 213 (v.s. 222 Manhattan and 282 Chi2)

Let us show the interest of these results by making a comparison with random recommendation. In this very basic process, at each step we selected a random piece of music from the list of 5985 (instead of measuring the similarity), then we selected the best ranked music also contained in the 15 pieces of the user test set, as in the previous method. We also computed the average and the standard deviation for the 200 users. .

As expected, the best "random" results are situated around the midpoint of the list (best average rank 2919, std 557). It is quite obvious that the recommendation based on the learning process gives better results than the random one. This comparison is only intended to give a contrast order of magnitude (213 vs 2919).

6. STATE OF THE ART

First of all, a reminder that the process described in this document is distinguishable from former work by a better descriptive capacity compared to the resources necessary for calculation and storage. The descriptive capacity is related to the rhythmic evaluation by the analysis of the swinging structure. These elements do not need to be obtained on the whole sound file; a limited statistical sampling is enough. We know that the signature requires the storage of only a very limited quantity of numerical data (2 real values). In addition, the signature will be almost independent of the format or the sound quality of the piece, even if the piece is incomplete.

There are various existing techniques for the characterization of musical files and research of similarities (MIR - Music Information Retrieval). These are based on three main approaches: signal processing [4, 5, 6], collaborative filtering [8, 9, 10], and data mining [2,3]. The approaches based on signal processing consist in analyzing directly the content of the audio file (signal and spectrum). In general, these characteristics are modelled by learning systems, and comparisons are carried out to find similarities. For example, in his work, George Tzanetakis extracted a list using characteristics obtained from the signal and spectral data envelope, in particular the centroïd (measure of spectral brightness), the roll-off (measure of the shape of the spectrum), ZeroCrossings (the number of times where the curve of the signal crosses zero) and sometimes even the MFCC (Mel-frequency cepstral coefficients), characteristics usually used in voice recognition. These characteristics are calculated in successive fixed-size analysis windows and on only the 30 first seconds of the audio piece. Another example of technology as regards acoustic prints is the TRM (This Recognizes Music). This technology was developed by the American company Relatable. Basically, this system allows the recognition of pieces of music by acoustic analogy exploiting an “audio code bar” type of print, which generates a single signature. As soon as the numerical print is created, it is sent to the TRM server, which compares the print with that of an existing song in the customer database. The latest commercial version of the TRM server can manage more than 5000 prints (already extracted) per second, or up to several billion requests per day.

The metadata used in collaborative filtering-based recommendation systems can be obtained thanks to the help of experts. For example, in Pandora Music radio [7] (from Music Genome Project), a trained music analyst analyses each song using up to 400 distinct musical characteristics (50 expert musicians). Each analyst spends about 30 minutes per piece of music to identify the pertinent attribute. These attributes capture not only the musical identity of a song, but also the many significant qualities that are relevant to understanding the musical preferences of listeners. This allows Pandora to provide a personalized audio stream consistent with user's preferences. This strategy differs from that of sites like Last FM that is based on collaborative filtering (others users' listening instead of music characteristics). Pandora radio is also available through such devices as mobile phones.

7. CONCLUSION

We propose a music recommendation system based on a light content description process. The evaluation of this system showed that, despite the low resources, the quality of the description was good. Furthermore, this study shows that rhythm is a very fundamental aspect of musical taste. Therefore, there are still improvements to add to our system. For example, we considered that user tastes are monolithic and this needs to be investigated. Even with this assumption, the results are not bad, but we wish to evaluate multi-polar user profiles, where users can have clusters of tastes, which seem more realistic.

8. REFERENCES

[1] On line music market: http://www.clubic.com/actualite-18188-le-marche-de-la-musique-en-ligne-multiplie-par-10-.htmle

[2] N. Durand, L. Lancieri and B. Cremilleux; Recommendation System Based on the Discovery of Meaningful Categorical Cluster, In proceedings of seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems 2003, Oxford, UK, (KES 2003),

[3] L.Lancieri, L Boubchir; Using multiple uncertain examples and adaptative fuzzy reasoning to optimize image characterization.; International journal KNOSYS (Knowledge Based Systems) Elsevier 2006.

[4] George Tzanetakis, George Essl, Perry Cook. Automatic musical genre classification of audio signals. ISMIR, 2001

[5] Scheirer, E. Tempo and beat analysis of acoustic musical signals. Journal of the Acoustical Society of America, 103(1):588-601

[6] L. Tanquerel, L.Lancieri; Fast sound file characterisation method; International Conference on Signal Processing and Multimedia Applications (Sigmap 2007) Barcelona, Spain

[7] Web sites in relation with Pandora music Radio: http://pandora.com; http://content.hamptonroads.com/story.cfm?story=95225&ran=117297; http://www.fastcompany.com/magazine/101/pandora.html

[8] Cory McKay, Ichiro Fujinaga. Automatic genre classification using large high-level musical. ISMIR, 2004

[9] Baumann, S., and O. Hummel. Using cultural metadata for artist recommendations. Proceedings of the International Conference on Web Delivering of Music, 138-41, 2003

[10] Hill, W., L. stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 194-201, 1995