Proceedings Chapter

Content-based music recommendation using underlying music preference structure

SOLEYMANI, Mohammad, *et al*.

Abstract

The cold start problem for new users or items is a great challenge for recommender systems. New items can be positioned within the existing items using a similarity metric to estimate their ratings. However, the calculation of similarity varies by domain and available resources. In this paper, we propose a content-based music recommender system which is based on a set of attributes derived from psychological studies of music preference. These five attributes, namely, Mellow, Unpretentious, Sophisticated, Intense and Contemporary (emph{MUSIC}), better describe the underlying factors of music preference compared to music genre. Using 249 songs and hundreds of ratings and attribute scores, we first develop an acoustic content-based attribute detection using auditory modulation features and a regression by sparse representation. We then use the estimated attributes in a cold start recommendation scenario. The proposed content-based recommendation significantly outperforms genre-based and user-based recommendation based on the root-mean-square error. The results demonstrate the effectiveness of these attributes in music [...]

Reference

SOLEYMANI, Mohammad, *et al*. Content-based music recommendation using underlying music preference structure. In: *Multimedia and Expo (ICME), 2015 IEEE International Conference on*. IEEE, 2015. p. 1-6

DOI : 10.1109/ICME.2015.7177504

CONTENT-BASED MUSIC RECOMMENDATION USING UNDERLYING MUSIC PREFERENCE STRUCTURE

Mohammad Soleymani[1], *Anna Aljanaki*[2], *Frans Wiering*[2], *Remco C. Veltkamp*[2]

[1] University of Geneva, Switzerland [2] Utrecht University, the Netherlands
mohammad.soleymani@unige.ch,{a.aljanaki, f.wiering, r.c.veltkamp}@uu.nl

ABSTRACT

The cold start problem for new users or items is a great challenge for recommender systems. New items can be positioned within the existing items using a similarity metric to estimate their ratings. However, the calculation of similarity varies by domain and available resources. In this paper, we propose a content-based music recommender system which is based on a set of attributes derived from psychological studies of music preference. These five attributes, namely, Mellow, Unpretentious, Sophisticated, Intense and Contemporary (*MUSIC*), better describe the underlying factors of music preference compared to music genre. Using 249 songs and hundreds of ratings and attribute scores, we first develop an acoustic content-based attribute detection using auditory modulation features and a regression by sparse representation. We then use the estimated attributes in a cold start recommendation scenario. The proposed content-based recommendation significantly outperforms genre-based and user-based recommendation based on the root-mean-square error. The results demonstrate the effectiveness of these attributes in music preference estimation. Such methods will increase the chance of less popular but interesting songs in the long tail to be listened to.

Index Terms— music preferences, music recommendation, music audio analysis

1. INTRODUCTION

Collaborative filtering based recommender systems have several well-known problems, such as the cold start problem (which occurs both with a new user and a new item introduction to the system) and the bias towards more popular items [1]. In the case of music recommendation, other data sources may be used to enhance the recommendation efficacy, such as metadata (genre, country of origin, date and any other available descriptors) and content (i.e., music audio itself).

The work of Soleymani is supported by the Swiss National Science Foundation's Ambizione grant. The work of Aljanaki, Veltkamp and Wiering is supported by the FES project COMMIT/. The authors thank P.J. Rentfrow and B.L. Sturm for kindly providing data and code, respectively.

Perhaps, one of the most useful types of metadata in the case of music is genre. Genre information becomes even more useful for music recommendation when enhanced by genre-similarity information, or even better, genre preference dimensions. Recently, genre preferences and their relationship to personality traits became an active research area [2]. There is no consensus on how many genre preference dimensions there are, but this number seems to be rather small: 4, 5 and 9 dimensions were suggested by various researchers [2].

Though genres are a popular way of describing music, using them in research of music tastes leads to methodological problems. Firstly, there is a large number of genres. For instance, a taxonomy of The Echonest[1] comprises 1301 genres. A lot of them are highly similar, only possessing minor cultural and musical distinctions, and there is no consensus on which genres should be used in preference studies. Secondly, despite their diversity, genres are broad and ill-defined categories. In a genre preference questionnaire, a researcher relies on respondents to have a universal and adequate understanding of genre definitions, which do not exist. Another problem lies in social connotations that are associated with certain genres. Research shows that people tend to associate musical preferences with social stereotypes. And though the music might not appeal to a user, the stereotype does, which might influence his choices, especially in adolescence [3].

It is possible to avoid these pitfalls through assessing music preferences using examples of music. In such a way, music preference dimensions transcend genre, and the musical properties underlying them can be learned directly. Content-based approaches also help to solve the cold start problem when new music is introduced.

A study on music preferences using pieces of music in 26 genres was conducted in 2011 by Rentfrow et al. [4]. Music preference questionnaires were factor analyzed and five factors of music preferences were discovered. The model was named *MUSIC*, after Mellow, Unpretentious, Sophisticated, Intense and Contemporary music preference factors. In 2012, the study was replicated with different subjects, and two new additional studies were conducted in an attempt to uncover the same structure of music preferences within one genre [5].

[1] www.echonest.com

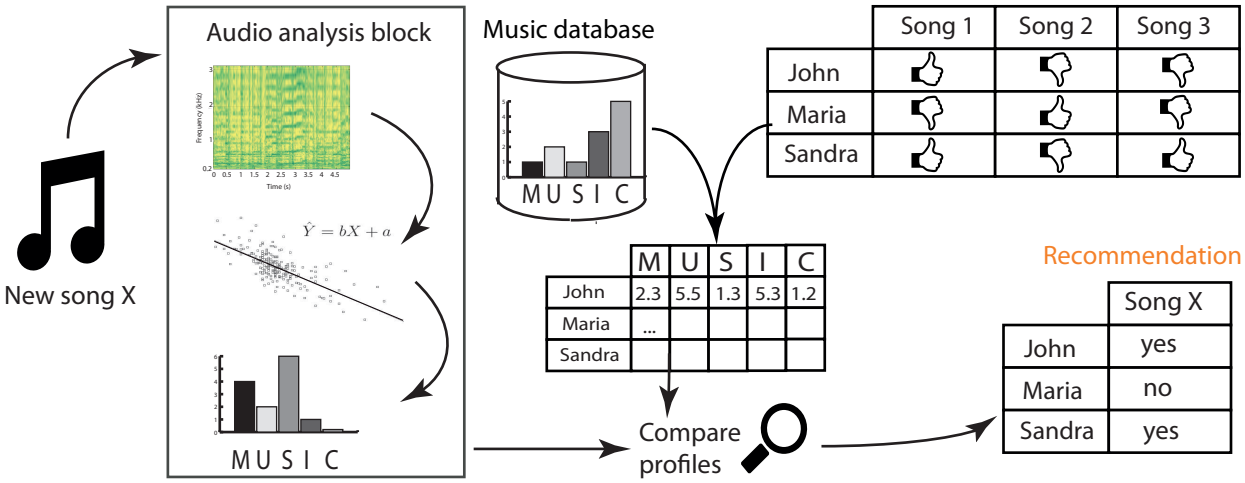

Fig. 1. Scheme of proposed recommender system. When a new song is entered into the system, its *MUSIC* five-factor preference vector is estimated from its acoustic content. Based on the users' previous ratings, the system also knows the five factors of personal preferences of its users. Given the estimated factors of the music and users' factors, the system predicts the potential ratings of songs for each user and recommends the songs.

All the three studies (regardless of whether multiple genres or only one genre were used) reflected the same underlying five factor structure of music preferences. Thus, this structure was shown to transcend genre.

The main contributions of this work are as follows. First, we developed a method to estimate psychologically validated music attributes from audio using sparse representation and auditory modulation features. Second, we validated the effectiveness of these attributes in detecting user preference in a cold start scenario. With this content-based method, we achieve significantly better performance than with different baseline methods relying on genre, artist and crude acoustic similarity. We propose to use these methods to enhance performance of music recommender systems. Fig. 1 shows an overview of the recommendation scenario.

2. RELATED WORK

Recommender systems can rely on listening history, metadata and content. In this section we will only review content-based ones. For a more detailed review on recommender systems we refer the reader to [6].

Content-based music recommender systems can be divided into two categories. The systems from the first category solve a query-by-example problem using some music similarity measure, which can also be personalized [7, 8]. In [8], Sotiropoulos et al. present a system that uses relevance feedback to train neural networks that select features relevant to a particular user from a set of timbral, temporal and tonal features. Systems belonging to the second category learn music preferences from the user profile. In [9], user preferences are learned from user's listening history, and k-Nearest Neighbor (kNN) and feature sub-space ensemble classifiers are applied to select music recommended to a particular user. In [10], user-supplied examples are used to learn high-level music properties, such as genre, instruments, mood, culture, and timbre. Based on learned preferences, recommendations are selected using semantic distance measures or a Gaussian Mixture Model.

To the best of our knowledge, no system has embarked on using the five-factor music attributes [4] in order to make informed decisions about the user's taste. Such a pre-learned model helps to put music in context and reduces the amount of training data necessary for a system to make useful non-trivial recommendations.

3. DATASET

For our experiments we reused the data set and user preference data (user-item matrices) collected and described in the work of Rentfrow et al. [4, 5]. In total, our dataset comprises 249 audio files from 5 substudies. All the audio files are 15 seconds long and extracted from a random point in the song.

In studies [4, 5], the structure of music preferences was analyzed using a series of experiments and their replications. In each experiment, a small set of musical pieces was rated by a considerable number of volunteers either in the laboratory or through the Internet. In each study, the collected preference ratings were factor analyzed using PCA with varimax rotation. Varimax is an orthogonal rotation which enforces independence of factors. The number of factors was determined using multiple criteria, including parallel analysis of Monte Carlo simulations, replicability across factor extraction methods, and interpretability.

The music pieces were selected by experts as being characteristic of a broad spectrum of music genres. The list of genres was compiled through a separate study, in which 5600 people were asked to list their favorite music genres. A list of

Table 1. Dataset compilation and the statistics of the original data collection experiments.

Reference	Number of songs	Number of subjects	Took place on	Genres
[4], Study 1	53	706	the Internet	26 genres
[4], Study 2	46	354	the Internet	26 genres
[5], Study 1	50	1057	Facebook	26 genres
[5], Study 2	50	1321	Facebook/Lab	Jazz
[5], Study 3	50	2160	Facebook/Lab	Rock

23 most commonly occurring genres (jazz, rock, heavy metal, etc.) was compiled, to which, for the sake of completeness, polka, marching band and avant-garde classical were added. The same list of genres was used in studies 1 and 2 of [4], and in study 1 of [5], which were meant to replicate previous experiments using the same genres, but different music and different set of subjects. In studies 2 and 3 of [5], jazz and rock pieces were used, respectively. The set of pieces was selected by experts to reflect the diversity of these two genres; i.e., pieces in swing, bebop, free jazz, big band and other subgenres of jazz were included into the 50 jazz pieces.

Therefore, the 249 pieces cover a very broad range of music genres. The data are summarized in Table 1.

For each musical piece, we have: (i) 15 seconds of audio; (ii) users' ratings for songs on a nine-point scale; (iii) metadata (title, artist, genre); and (iv) factor loadings on five factors of the music preference model (*MUSIC*).

4. METHODS

4.1. Acoustic features

In this paper, we use modulation analysis to extract timbral features from audio. We use the method and code from [11]. Auditory temporal modulation features have been shown to work well for genre recognition [12, 11], which is the reason we employ them.

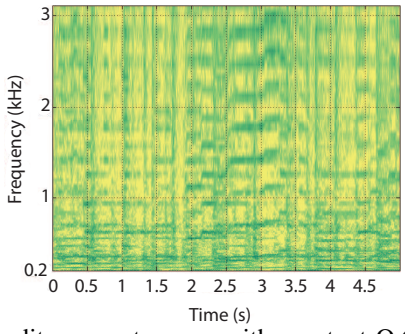

Fig. 2. Auditory spectrogram with constant-Q transform.

Modulation representations of acoustic signals describe the variation of spectral power in scale, rate, time and frequency, which is motivated by the human auditory and visual systems. Perceptual properties of acoustic stimulus, such as speech and music, are encoded by its slow temporal modulations [13]. We extract an auditory representation that maps a piece of music recording to a two-dimensional representation of its slow temporal modulations. These representations form an overcomplete dictionary of basis signals, which are then used for sparse representation-based classification.

To extract features, we downsample each audio file to 16 kHz and split it into overlapping segments of 5 seconds. We then compute auditory spectrograms following the model of the auditory system [14], using constant-Q transform of 96 bandpass filters covering a 4-octave range. Fig. 2 shows the resulting spectrogram for one of the excerpts.

We pass each channel of the auditory spectrogram through a Gabor filterbank sensitive to particular modulation rates, and form the Auditory Temporal Modulations (ATM) by integrating the energy output at each filter. Fig. 3 shows the modulation spectrogram that results from these manipulations.

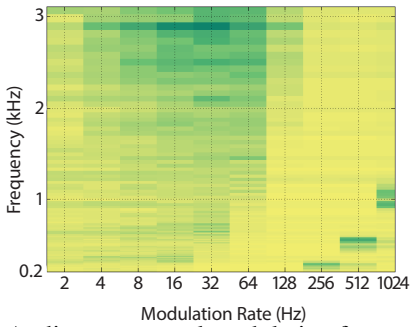

Fig. 3. Auditory temporal modulation feature matrix.

For more details on implementation of auditory temporal modulation features consult [11].

4.2. Attribute learning

Rentfrow et al. introduced a five factor model that describes a set of sufficient attributes in music preferences [4, 5]. The five factors were *M*ellow, *U*npretentious, *S*ophisticated, *I*ntense and *C*ontemporary (*MUSIC*). We aim at automatically estimating these attributes from music content. Given the similarity of this problem to genre recognition, we opted for the well performing models in that domain [15]. Sturm and Noorzad [11] analyzed different approaches for music genre detection. They found that different audio modulation features and dimensionality reduction techniques, such as principal component analysis (PCA) or non-negative matrix factorization (NMF), utilized in [15], do not significantly change the genre recognition rates. Inspired by their findings, we

opted for NSL modulation features [16] with 10 octaves and a filter bank with 128 filters (in total 1280 features), similar to [12]. We also created a baseline feature-set using a set of standard off-the-shelf features from MIRToolbox [17]. The baseline feature-set included the average and standard deviation of the following features using the default settings in MIRtoolbox: short time energy, pitch, 13 Mel Frequency Cepstral Coefficients (MFCC), zero crossing rate, spectrum flux, spectral Rolloff, Harmonic Change Detection Function (HCDF) and roughness (in total 20 features).

Sparse representation has proven its high performance in detecting music genres [15]. Given the nature of the attributes and the similarity between this problem and genre recognition, we used a regression with sparse approximation of data [18], which is defined as follows.

Given a dataset or dictionary of N observations $\mathcal{D} := \{(\mathbf{x}_i, y_i)\}_{i\in\Omega}$ where the input $\mathbf{x}_i = [\mathbf{x}_{1i}, \ldots, \mathbf{x}_{Mi}] \in \mathbb{R}^M$; here $\mathbf{x}_i$ is the feature vector, y_i is the output and $\Omega = 1, \ldots, N$ is the index of the dictionary, a non-parametric regression is defined as:

$$y_i = f(\mathbf{x}_i) + \epsilon_i \tag{1}$$

where $f(\mathbf{x})$ is the function we want to estimate and ϵ_i is an independent error term. The main idea behind the sparse representation approximation regression is to get a local estimate of the function $f(\mathbf{z})$ at a point $\mathbf{z}$ using a linear combination of the outputs of the dictionary samples $(y_i, i \in \Omega)$. In this paper we use the locally constant approximation of the $f(\mathbf{z})$ using the sparse representation as follows:

$$\hat{f}(\mathbf{z}) = \frac{\sum_{i\in\Omega} \alpha_i(\mathbf{z}) y_i}{\sum_{j\in\Omega} \alpha_j(\mathbf{z})} \tag{2}$$

where α is found by sparse approximation using the given sample $\mathbf{z}$ and the dictionary $\mathbf{D}$ defined by:

$$\mathbf{D} := \left[\frac{\mathbf{x_1}}{\|\mathbf{x_1}\|_2}, \frac{\mathbf{x_2}}{\|\mathbf{x_2}\|_2}, \ldots, \frac{\mathbf{x_N}}{\|\mathbf{x_N}\|_2}\right] \tag{3}$$

which in this case is the normalized feature vectors of the training set. The sparse approximation problem can be defined by $\mathbf{z} = \mathbf{D}\mathbf{s} + \epsilon$ where $\mathbf{s}$ is sparse. α is then calculated by the following equation from the sparse approximation weights:

$$\alpha_i(\mathbf{z}) := \left[\frac{\mathrm{S}(\mathbf{z}, \mathbf{x}_i)}{min_{j\in\Omega}\mathrm{S}(\mathbf{z}, \mathbf{x}_j)}\right]^{-1} \frac{s_i}{\|\mathbf{z}\|_2} \tag{4}$$

The sparse representation was calculated using the Spectral Projected Gradient Method for ℓ_1-minimization (SPGL1) [19]. For more details on the regression method we refer the reader to [18].

4.3. Cold-start content-based recommendation

To validate the effectiveness of the automatic attribute detection, we simulated a cold-start scenario where a new song is being introduced to a system. We again estimated every song's rating (for a particular user) from its *MUSIC* estimated attributes using leave one out cross validation. We used the best attribution detection results from NSL features and sparse representation (see Table 2).

Since not every song was present in all the studies (see Table 1 for the list of studies), we first identified the relevant study to extract the corresponding $M \times N$ item-user matrix. If a song was present in more than one study, we picked the larger study. The attributes are originally calculated using factor analysis which states:

$$X - \mu = LF + \epsilon \tag{5}$$

where $X - \mu$ is the $M \times N$ centered song-user (item-user) matrix, $L_{M\times 5}$ is the factor loading matrix containing the attributes and $F_{5\times N}$ is the matrix containing the factors for each user. For each user we estimated the factors $\hat{F}_{5\times N}$ after excluding the ratings of the given song that we use as the test-set using the factor loading of the remaining songs $\check{L}_{(M-1)\times 5}$ and the modified song-matrix $\check{X}_{(M-1)\times N}$ as follows:

$$\hat{F} = \check{L}^{+}(\check{X} - \mu) \tag{6}$$

where $\check{L}^{+}$ denotes the pseudo-inverse of the matrix $\check{L}$. Then the ratings $\hat{\mathbf{x}}_{N\times 1}$ can be estimated from the song attribute vector $\hat{Y}_{1\times 5}$ which is estimated from its acoustic content:

$$\hat{\mathbf{x}} = (\hat{Y}\hat{F})^{\intercal} + \check{\mu} \tag{7}$$

where $\check{\mu}$ is the mean rating for every user in the training set.

We implemented four different baseline methods: user's averaged ratings ($\check{\mu}$), artist-based, genre-based and acoustic similarity based methods. The artist-based and genre-based methods average the scores given by the same user to the songs by the same artist or genre as the test samples. In case the same artist or genre do not exist in the training set, the averaged scores given by the same users from $\check{\mu}$ were used. The acoustic similarity was calculated based on the euclidean distance between the content features of a given song (i) and the rest of the songs from the training set (d_{ji}). To be consistent, we used the NSL modulation features reduced by PCA, similar to the way we detected the attributes. Then the similarity-based estimation $(\hat{x_i})$ is calculated using Equation 8.

$$\hat{x_i} = \frac{1}{M-1}\sum_{j=1}^{M-1} w_{ji}x_j, \text{where } w_{ji} = \frac{e^{-d_{ji}}}{\sum_{k=1}^{M-1} e^{-d_{ki}}} \tag{8}$$

5. EXPERIMENTAL RESULTS

5.1. Automatic attribute detection

Since we only had 249 clips, we segmented the 15 second clips into 11 overlapping 5 seconds excerpts. We then extracted the features using the NSL toolbox[2] and MIRtoolbox. We used multi-linear regression (MLR), support vector

[2] http://www.isr.umd.edu/Labs/NSL/Software.htm

Table 2. Attribute detection results are listed below; for better readability the attribute scores were scaled between [-0.5 ,+0.5]; for RMSE $\in [0, 1]$ the lower the better; for r^2 the higher the better. The negative values of r^2 indicate that the intercept estimation is worse than random. The best results are shown in boldface font. Acronyms: MIRT: MIRtoolbox, RMSE: root-mean-square error, MLR: multi-linear regression, SVR: Support vector regression, RSS: regression with sparse representation.

Attribute		**M**		**U**		**S**		**I**		**C**	
Model	Features	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2
MLR	NSL	0.12	0.00	0.13	-0.19	0.13	0.09	0.11	0.28	0.11	-0.19
	MIRT	0.11	0.15	0.12	0.00	0.12	0.23	0.10	0.41	0.09	**0.24**
SVR	NSL	0.15	-0.54	0.17	-0.84	0.19	-0.99	0.13	0.01	0.13	-0.67
	MIRT	0.14	-0.44	0.17	-0.96	0.19	-0.98	0.13	-0.01	0.13	-0.57
RSS	NSL	**0.10**	**0.21**	**0.12**	**0.07**	**0.12**	**0.28**	**0.10**	**0.47**	**0.09**	0.17
	MIRT	0.11	0.08	0.13	-0.18	0.13	0.09	0.11	0.35	0.10	0.05

regression (SVR) and regression with sparse representation (RSS) to train the models. For the SVR, we used the linear kernel and the hyper-parameter was chosen based on the lowest mean squared error on the training set. We used PCA for dimensionality reduction with the threshold of retaining 95% of the variance. The PCA was applied on the training set in each iteration of the cross validation. After calculating the number of principal components that carry 95% of the variance the corresponding mapping was used to reduce the dimensionality of the test set. The ground truth factor loadings of songs on the five attributes were derived from the studies [4, 5]. We performed a leave-one-out cross validation to evaluate the performance for attribute detection. The results of the 11 excerpts of the same song were averaged to form a single estimation of the attribute score. The results are summarized in Table 2. We used root-mean-square error (RMSE) and the coefficient of determination or r^2 to evaluate attribute detection performance. The combination of auditory modulation features and regression with sparse representation outperformed all the other methods. It is also the only method that consistently achieved positive r^2 which means the estimations are better than random (or averaged scores) for all the attributes.

Looking at the attributes, Intense and Sophisticated were the easiest to model, whereas Mellow, Unpretentious and Contemporary are more problematic to model from low-level acoustic features. This might be caused by the nature of music genres that are the most characteristic of these factors. Aggressive and loud music genres, such as heavy metal, punk and rock, load high on Intense factor. Genres with complex harmonic structure and mostly acoustic instruments, such as jazz, classical and world music, have high loadings on Sophisticated factor. Unpretentious factor includes country, rock'n'roll and pop music, Mellow factor includes pop, soul and r'n'b, and Contemporary – electronica and rap. As it was reported in [4, 5], there are certain acoustic properties, that are characteristic of each of these factors. It appears, that some of these properties are easier to detect with low-level features, such as distinctive electric instruments of Intense factor, or with non-percussive nature of Sophisticated factor. The differences between country, pop and soul music, on the other hand, are more subtle. For instance, Mellow music was perceived as romantic and simple, and Contemporary music was perceived as party music with synthesized instruments. These properties are more high-level and culturally defined, and hence more difficult to learn from modulation features.

5.2. Content-based recommendation

Simulated recommendation performance was evaluated by the root-mean-square error (RMSE) values from different methods (see Table 3). We performed a one-sided non-parametric rank-sum test between the RMSE scores obtained by the attribute-based recommendation and the baseline methods. The p-values of the statistical tests are shown in the last column of the Table 3. We can see that the proposed method performs consistently better than the baseline methods. We therefore validate the effectiveness of these attributes for estimating users' musical preference.

Table 3. Performance of different recommender systems. The ratings are scaled between $[0, 1]$. The last column indicates the p-value of the non-parametric statistical tests between the RSME results of that row and the proposed method. For RMSE, the lower the better.

Method	**RMSE**	p-value
Attribute-based	**0.251±0.039**	-
User's average ratings	0.265±0.037	0.0000
Genre-based	0.264±0.044	0.0002
Similarity-based	0.263±0.042	0.0000
Artist-based	0.278±0.044	0.0003

6. CONCLUSIONS

Current music recommender systems often recommend popular songs and are unable to deal with songs with no ratings. Thus, a content-based method that can position a song within the existing database and assist users to discover new songs is desirable. So far, music genre has been used as an underlying feature for musical preference. However, music genre

is ambiguous, i.e., a song is often associated with a number of genres and songs from the same genre can be very diverse. The five-factor model that was introduced by psychologists [4, 5] is trying to address this shortcoming by identifying the underlying factors that contribute to our preference. In this paper, we first build a system to detect these attributes or factors from the acoustic content of music. We found that the combination of audio modulation features and sparse representation yields the best performance for attribute detection. Then, we demonstrated how these attributes, albeit automatically estimated and inaccurate, can still assist a cold-start recommendation scenario. In this work we solely relied on the acoustic content of the songs. We believe this is an advantage given that the new and less popular songs often do not have any tags or rich metadata associated with them. Such systems can increase the chance of serendipity or the discovery of interesting yet less known songs by users from the very long tail in music recommendation.

7. REFERENCES

[1] O. Celma, *Music recommendation and discovery in the long tail*, Ph.D. thesis, UPF, Barcelona, Spain, 2008.

[2] A. Laplante, "Improving music recommender systems: What can we learn from research on music tastes?," in *Conference of the International Society for Music Information Retrieval (ISMIR)*, 2014, pp. 451–456.

[3] A. C. North and D. J. Hargreaves, "Music and adolescent identity," *Music Education Research*, vol. 1, no. 1, pp. 13–33, 1999.

[4] P. J. Rentfrow, L. R. Goldberg, and D. J. Levitin, "The structure of musical preferences: a five-factor model," *Journal of Personality and Social Psychology*, vol. 100, no. 6, pp. 1139–1157, 2011.

[5] P. J. Rentfrow, L. R. Goldberg, D. J. Stillwell, M. Kosinski, S. D. Gosling, and D. J. Levitin, "The song remains the same: A replication and extension of the MUSIC model," *Music Perception*, vol. 30, no. 2, pp. 161–185, 2012.

[6] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," *Expert Systems with Applications*, vol. 39, no. 11, pp. 10059–10072, 2012.

[7] C. Lu and V. Tseng, "A novel method for personalized music recommendation," *Expert Systems with Applications*, vol. 36, no. 6, pp. 10035–10044, 2009.

[8] D. Sotiropoulos, A. Lampropoulos, and G. Tsihrintzis, "Musiper: A system for modeling music similarity perception based on objective feature subset selection," *User Modeling and User-Adapted Interaction*, vol. 18, no. 4, pp. 315–348, 2007.

[9] M. Grimaldi and P. Cunningham, "Experimenting with music taste prediction by user profiling," in *ACM SIGMM International Workshop on Multimedia Information Retrieval*. 2004, MIR '04, pp. 173–180, ACM.

[10] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambo, E. Gomez, and P. Herrera, "Semantic audio content-based music recommendation and visualization based on user preference examples," *Information Processing and Management*, vol. 49, no. 1, pp. 13–33, Jan. 2013.

[11] B. L. Sturm and P. Noorzad, "On automatic music genre recognition by sparse representation classification using auditory temporal modulations," in *International Symposium on Computer Music Modeling and Retrieval*, 2012.

[12] I. Panagakis, E. Benetos, and C. Kotropoulos, "Music genre classification: A multilinear approach," in *The International Society for Music Information Retrieval Conference (ISMIR)*, 2008, pp. 583–588.

[13] S. Sukittanon, L. E. Atlas, and J. W. Pitton, "Modulation-scale analysis for content identification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 52, no. 10, pp. 3023–3035, 2004.

[14] X. Yang, K. Wang, and S. Shamma, "Auditory representations of acoustic signals," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 824–839, 1992.

[15] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 3, pp. 576–588, 2010.

[16] E. Grassi, J. Tulsi, and S. Shamma, "Measurement of head-related transfer functions based on the empirical transfer function estimate," in *Proc. ICAD*, 2003, pp. 119–121.

[17] O. Lartillot, P. Toiviainen, and T. Eerola, "A matlab toolbox for music information retrieval," in *Data analysis, machine learning and applications*, pp. 261–268. Springer Berlin Heidelberg, 2008.

[18] P. Noorzad and B. L. Sturm, "Regression with sparse approximations of data," in *European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 674–678.

[19] E. van den Berg and M. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.