EXPLORING THE RELATION BETWEEN NOVELTY ASPECTS AND PREFERENCES IN MUSIC LISTENING

Andryw Marques
Universidade Federal
de Campina Grande
andrywmr@lsd.ufcg.edu.br

Nazareno Andrade
Universidade Federal
de Campina Grande
nazareno@computacao.ufcg.edu.br

Leandro Balby
Universidade Federal
de Campina Grande
lbmarinho@computacao.ufcg.edu.br

ABSTRACT

The discovery of new music, e.g. song tracks and artists, is a central aspect of music consumption. In order to assist users in this task, several mechanisms have been proposed to incorporate novelty awareness into music recommender systems. In this paper, we complement these efforts by investigating how the music preferences of users are affected by two different aspects of novel artists, namely familiarity and mainstreamness. We collected historical data from Last.fm users, a popular online music discovery service, to investigate how these aspects of novel artists relate to the preferences of music listeners for novel artists. The results of this analysis suggests that the users tend to cluster according to their novelty related preferences. We then conducted a comprehensive study on these groups, from where we derive implications and useful insights for developers of music retrieval services.

1. INTRODUCTION

The discovery of new songs and artists that one likes is a central aspect of music consumption. To a higher or lesser degree, all music listeners look for novelty over time. Due to the huge music collections presently available to listeners, it becomes increasingly difficult to sift for novel and relevant music. There is thus a potential for efficient and novelty-aware retrieval services that assist listeners.

Both commercial systems and research efforts have made progress in incorporating novelty in music retrieval systems (e.g.. [7,10,11]). In general, novelty concerns an item previously unknown to a consumer. For example, a movie still not watched or a band still not listened to. However, it is also possible to extend this unifaceted view of novelty. For example, an item that is both previously unknown and has a different style from all other previously known items brings to a consumer an extra dimension of novelty. The distinction of extra dimensions in novelty is relevant as consumers might have preferences for novel items along multiple of such dimensions: some users may prefer novel items that are similar (or familiar) to their current preferences, but not mainstream (or popular), and vice-versa. Although sometimes mentioned in the literature, this multifaceted view of novelty remains largely unexplored in the Music Information Retrieval field.

In this work we conduct an exploratory analysis of the impact of different dimensions of novelty in the preferences of music listeners. Our ultimate goal is to elucidate what kinds of novelties would be preferred by music listeners, thus paving the way for more efficient and informative music retrieval services. More concretely, we examine how two important aspects of music novelty – familiarity and mainstreamness – affect the preferences of music listeners to the new artists they discover.

Our analysis relies on historical music listening data from 17,000 Last.fm users (Section 3), and on a model of listening behavior and novelty aspects we propose (Section 4). On the one hand, our results suggests that there is no global correlation between familiarity or mainstreamness and novelty relevance for the listeners in our sample. On the other hand, when considered individually, most listeners do have a significant correlation between either familiarity or mainstreamness of novel artists heard and the relevance they see in these artists (Section 5). This suggests that novelty-based personalized music retrieval systems likely have more chances of success than non-personalized ones.

Another consequence of this result is that listeners form groups concerning their novelty-related preferences. This observation motivates us to perform a cluster analysis on the listeners. This analysis unveils seven archetypical profiles that explain how different groups have preferences for the different aspects of novelty we consider (Section 6).

2. RELATED WORK

Several studies point to the relevance of considering novelty in music recommendation systems. In general, different efforts agree that it is necessary to evaluate such systems not only by their accuracy and recall, but also by how many new and relevant items are introduced to a user [4,5]. In materializing this view, several studies in the literature discuss how to build recommender systems that introduce novelty adequately [7, 10, 11].

Specifically about novelties, Vargas et al. [8] put forward a formal framework to analyze the relation between users, items, and novelty in a recommendation system. The authors make a distinction between choice (a user picks an item), discovery (a user is introduced to an novel item) and relevance (a user likes an item).

Related to the discovery, novel items can be defined in a twofold manner [1]. On the one hand, novel items have (usually content-based) characteristics not shared by items previously declared as relevant by the user [7, 11]. On the other hand, novel items are defined in terms of popularity among users, mainly non popular items. This attribute of items is influential in their discovery potential [2, 4], thus interplaying with their novelty.

With respect to the reaction to novelty in music, a related field is that of the reaction to new opinions (in general). Munson and Resnick [6] ran experimental studies that suggest that online users can be clustered into three distinct subgroups with respect to their preferences for new opinions: diversity-seeking, challenge-averse and support-seeking.

In general, although past work has utilized the concept of novelty for recommending music and other types of items, and has sometimes dealt with the relation between novelty and diversity or popularity, the literature currently lacks studies specifically formalizing novelty aspects and relating them to the relevance of novel items. This study contributes to fill this gap, formalizing the concepts of familiarity and mainstreamness of novel music artists in relation to a listener. According to Vargas et al.'s framework, we relate these two aspects with the relevance of items chosen by listeners so as to inform the design of discovery mechanisms. Finally, our results with respect to different types of listeners with respect to their reaction to novelty bears similarities to Munson and Resnick's results.

3. DATA COLLECTED

Our dataset is comprised of two parts: the first contains the subjects whose behavior will be analyzed, together with their historical listening habits, and the second concerns metadata about the artists listened by the chosen subjects. All data was collected from Last.fm through its publicly available data access API.

For the remainder of this study, we differentiate between the *experiment period*, set as the six months between March to September 2012, the *observation period*, which includes the experiment period and also the following six months, and the *prior listening history*, which is the year preceding the experiment period. The distinction between experiment and observation period is used to control the bias in relevance measuring, and is further discussed in Section 4.4. Figure 1 illustrates the periods on a timeline.

3.1 Subject data

Our goal is to identify and collect data about a sample of Last.fm users which has been exposed to novel artists during a period of interest. Starting with the profile of the first author, a snowball sampling procedure was conducted on the Last.fm friendship network until 100,000

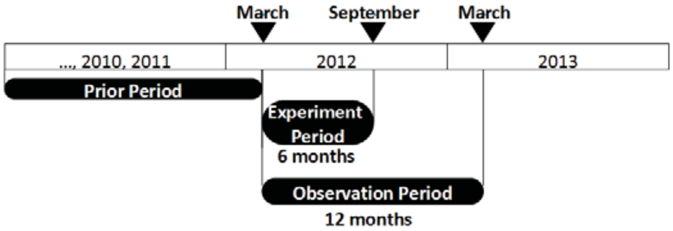


Figure 1: Time periods used in the experiment.

users were identified. Next, data about the artists listened to by the users prior and during the experiment period were collected. The former is collected as the set of the 200 artists most listened by the user since he or she joined Last.fm, plus the union of the sets of 100 most listened artists by the user in each week of his or her prior listening history. The data collected about the listening habits during the observation period is the union of the sets of the 100 most listened artists by the user in each week in this period.

After this data collection, a filtering process was conducted, with the goal of selecting the users who (i) have highly active listening habits, (ii) are likely to have informed a large portion of these habits to the system we collected the data from, (iii) have experienced a number of novelties in their listening that enable us to investigate the relation between novelty characteristics and user preferences, and (iv) likely gear most of their music listening, instead of using a Last.fm's radio created by a recommendation algorithm [1] .

For (i) and (ii) we kept only users who had at least 100 artists in their prior listening history, and which were active in at least three quarters of the weeks in the observation period, having at least 100 song executions in a week to be considered active. For (iii), we selected the users with at least 10 novel artists in the experiment period that were listened to at least 10 times each. For (iv), we infer that a user does not chiefly use the Last.fm recommendations if that user listened 15 or more songs of a same artist in a week. Our assumption is that this denounces a level of gearing incompatible with radio licensing. Finally, we filtered out also a small number of registered users with unrealistically high music listening frequency (more than 16 hours a day during the observation period). This process produced a sample with 17,183 Last.fm users.

3.2 Artist data

For each artist present in the subjects' listening data, artist popularity and tag data was collected from Last.fm. Popularity data was collected as the number of users who listened the artist as informed by Last.fm on March 10th 2013. A tag is a label given by a user to the artist, ranging, in Last.fm, from music genres (eg. rock, samba) to mood (eg. sad, lively) and other contextual metadata (eg. summer, lovesong). Each tag has a popularity, reflecting how often users have assigned the tag to a given artist. After

[1] http://www.lastfm.com/listen

collecting tag data from all artists in the subjects listening data, we filter out unpopular tags, defined as having popularity lower than 0.15 of the most popular tag for the artist. Furthermore, tags denoting personal circumstances, such as *seen live* and *favorite* were manually removed.

4. CHARACTERIZING LISTENER PROFILES AND NOVELTIES

To examine how different characteristics of novel items affect their relevance for listeners, we resort on three constructs we now describe in turn: a model of a listener profile, a set of dimensions on which we consider novel items may vary, and a definition of relevance.

4.1 Novel Items

For our experiment, the items considered are artists, and an artist is novel for a subject if two conditions are met: (a) this artist is absent from the subject's listening history prior to the experiment period and (b) the artist was listened to at least five times in a week by this subject during the experiment period. The experiment period is used for identifying novel artists listened by the subject, while the whole observation experiment period is used for an unbiased evaluation of how much attention the subject payed to these novel items (a process detailed in Section 4.4).

With this definition, we identify 652,511 novel items in our data, with a subject discovering on average 38.3 (std. dev. 31.2) novel artists during our measurement.

4.2 Listener Profiles

We model each listener profile as a set of clusters of artists in his or her listening history prior to the experiment period. These clusters are obtained applying the DBScan clustering algorithm [3] to the set of artists known to the listener are not considered to be novel to the user in our experiment.

Artists are represented as vectors, where each vector component represents the number of times a given tag was assigned to the artist being considered. More formally, let A be the set of artists, and T the set of tags. Let $f : A \times T \rightarrow \mathbb{R}$ denote the frequency of which a given tag $t \in T$ was assigned to a given artist $a \in A$.[2] Now, the vector representing an artist $a \in A$ is defined as $\vec{a} := (f(a, t_1), f(a, t_2), \ldots, f(a, t_{|T|}))$ The cosine similarity between two artists is defined, as usual, as:

$$\cos(\vec{a}, \vec{a}') := \frac{\langle \vec{a}, \vec{a}' \rangle}{\|\vec{a}\| \|\vec{a}'\|} \tag{1}$$

The clusters are then computed based on the cosine similarity between the vector representations of artists as mentioned above. The two parameters of the DBScan algorithm – the minimum number of points for a cluster to be considered and the minimum similarity for two points in a cluster – were empirically defined as 3 and 0.875 (resp.) after consulting an online community of music aficionados [3] about which of several clustering solutions better described members' taste profiles. Profiles obtained through this process for our subjects have an average of 5.5 (std. dev. 2.9) clusters found among the artists present in the subject's listening history.

[2] http://www.lastfm.com/api

4.3 Characteristics of novelties

We consider novel items/artists have two other aspects of novelty besides being unknown to a listener: familiarity and mainstreamness. The former captures the notion that characteristics of the music from an artist may or may not be familiar to a listener. The latter models the intuition that the artist, although not listened, may be known to the listener through other media and the general public acclaim, ie. mainstream.

4.3.1 Familiarity

Familiarity is defined based on the similarity between an artist and those that compose a listener profile. This models the intuition that for a jazz fan, novel heavy metal artists are likely to sound less familiar than a novel artists from a subgenre of jazz. Formally, let $P := \{C_1, ..., C_n\}$ denote a listener's profile, formed by the clusters of artists C_i. Also, let $\vec{c_i}$ be the centroid of the cluster C_i, and p_i be the proportion of all song executions by the listener in his or her prior listening history that were from artists in C_i. Then, the familiarity between an artist a and a listener profile P is the weighted arithmetic mean of the similarity between the artist and the centroids $\vec{c_i}$, with p_i as weights:

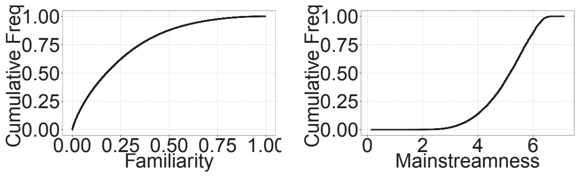
$$\text{fam}(a, P) = \frac{\sum_{i=1}^{n} cos(\vec{a}, \vec{c_i}) \times p_i}{\sum_{i=1}^{n} p_i} \tag{2}$$

Figure 2(a) displays the cumulative distribution of the familiarity of all novel artists in our experiment to their listener profiles. Notice that there is an overall skew towards more familiar artists.

4.3.2 Mainstreamness

For our experiment, mainstreamness is defined as the log of the overall popularity of the artist in Last.fm. This definition aims to capture how likely it is that the artist and an opinion about the artist are known to a subject before the subject has carefully listened to the artist. This knowledge likely comes from other medias, such as mentions in newspapers or television, and from the subject's social network. Our hypothesis is that some users may have a preference for mainstream novel items known to be liked by the general public. Because the artists popularity distribution observed is highly skewed, our analysis uses the log of this popularity in all results reported. Figure 2(b) displays the cumulative distribution of the mainstreamness of all novel artists in our experiment. This distribution has a skew towards highly mainstream artists.

[3] A facebook group of music releases named 'Revista Billboard'



(a) Familiarity. (b) Mainstreamness.

Figure 2: Cumulative distribution of Familiarity and Mainstreamness for all novel artists in the experiment period.

4.4 Preferences for novel items

We measure subjects' preferences for different types of novel artists gauging how much or how often each novel artist was listened to during the observation period. An artist/item is thus said more relevant than another if the former was listened to more times or more frequently than the latter during a time span.

To put forward this approach we consider two metrics for item relevance in our analysis. Let δ be the period between the first time a subject listened to a novel artist and the end of our observation period, then (a) the *total attention* a subject gives to an artist is the total number of executions of songs of this artist during δ divided by the number of weeks in δ; and (b) the *period of attention* of a subject to a novel artist is the fraction of the number of weeks in δ in which the subject listened at least once to the artist. Because novel artists are found over a period of time, some of these artist have a smaller time window in our experiment period in which they can be listened to. We tackle this issue with two measures. First, taking δ as the denominator of attention metrics limits the potential bias in the counting of song executions or weeks of attention. Second, as mentioned in Section 4.1, only novelties discovered in the experiment period are considered in the analysis, but the whole observation period is used to count song executions and weeks of attention. This gives each novel item a minimum of six months of observation in our traces, which limits possible biases caused by too short observations.

Figure 3 shows the cumulative distribution of the total attention log and period of attention for each novelty found by the subjects in our experiment. In both cases, attention is concentrated on a small proportion of the novel items discovered.

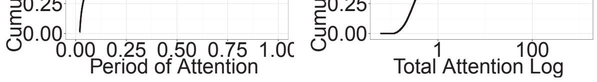


Figure 3: Cumulative distribution of the Total Attention Log and Period of Attention for each novelty

Table 1: Correlation (Spearman's coefficient) between novelty characteristics and relevance, analyzing all novel items together.

Dimensions	PoA	F	M
Total Attention	0.74	0.04	0.04
Period of Attention (PoA)	-	0.04	0.05
Familiarity (F)	-	-	0.01
Mainstreamness (M)	-	-	-

5. PREFERENCES FOR NOVELTY CHARACTERISTICS

Our central research question is to understand how considering different aspects of novelty can enhance our understanding about listener preferences. To address this question, we first evaluate whether there is a correlation between novelty characteristics – familiarity and mainstreamness – and relevance – total attention and attention period.

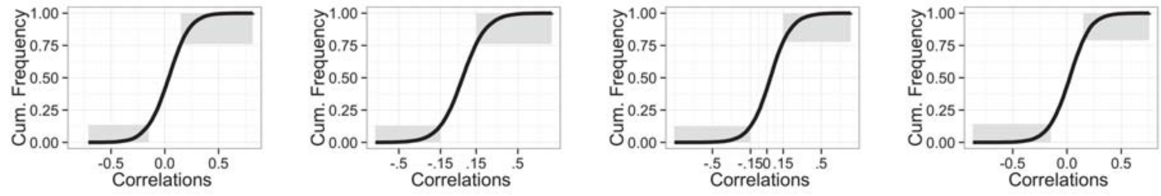
Table 1 shows that analyzing all novel items together, there seems to be no relevant correlation between novelty characteristics and relevance that is valid for all subjects. On the other hand, analyzing the correlation between novelty characteristics and relevance for the novel items of each subject individually, there is a different pattern. For each pair of variables we examine, there is a significant correlation (above 0.15 or below -0.15) between the variables for approximately one third of all subjects (Figure 4). In aggregate, 70% of all subjects have in their data a significant correlation between at least one of the novelty characteristics and a measure of relevance.

Taken together, our results point that *while there is no overall common behavior regarding subjects' preferences for different types of novelty, most users have some preference for a type of novelty in their listening behavior.*

6. LISTENER GROUPS WITH RESPECT TO NOVELTY

The analysis of the correlation between novelty types and preferences indicates that there are different types of listener in our data. In this section we apply clustering analysis to delve further in the identification and analysis of such groups. For that, we use only the set of subjects for which at least one of the correlations analyzed in the previous section is not in the [-0.15; 0.15] interval (total=12,114).

Our analysis uses the Ward hierarchical clustering method [9] considering normalized versions of two dimensions that describe each subject's preferences for novelty aspects and two dimensions that measure the subject's usual musical taste. The dimensions related to preferences are (a) the correlation between total attention devoted to a novel artist and the artist's familiarity for the listener and (b) the correlation between total attention to a novel artist and the artist's mainstreamness. To represent usual tastes, we use (c) the average mainstreamness of the artists in the listener's profile, and (d) the number of clusters present in the subject's listening profile. Upon examination, the cor-



(a) Period of Attention and Mainstreamness (b) Period of Attention and Familiarity (c) Total Attention and Familiarity (d) Total Attention and Mainstreamness

Figure 4: Distribution of the Spearman correlations coefficient between novelty characteristics and relevance in each listener's data. Shaded areas highlight the portion of subjects with correlation higher than 0.15 or lower than -0.15.

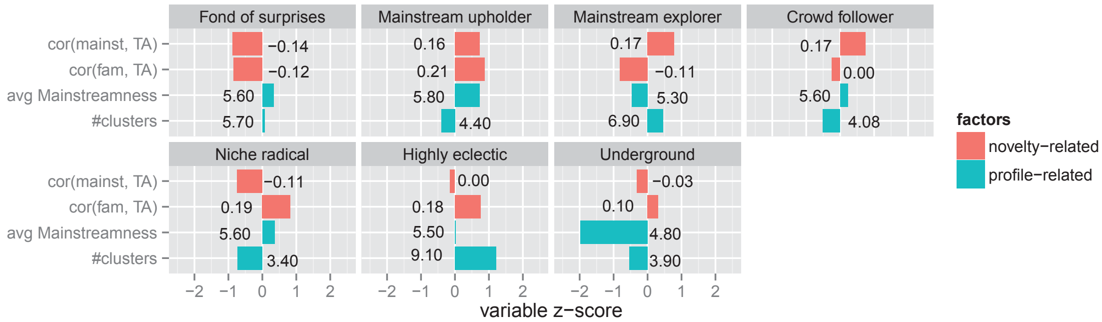


Figure 5: Centroids of the seven clusters found in the analysis. Variables as normalized as the z-score (horizontal axis), where zero represents the average across all listeners, and the unit of variation is a standard deviation in the considered metric. In the vertical axis, fam stands for familiarity, mainst for maisntreamness and TA for total attention. The numbers on the chart are the denormalized values.

relations between novelty aspects and period of attention for novel artists were discarded as redundant dimensions for the clustering analysis.

Applying this approach points to satisfying clustering solutions with respect to intra and intergroup heterogeneity that have between 5 and 8 clusters. Further examination considering the descriptive power of the solutions leads to the choice of the 7-cluster solution, whose centroids are shown in Figure 5. Analyzing these centroids, we labelled the groups as follows:

1. *Fond of surprises*: The largest cluster (n=2,738), contains subjects with a distinct preference for novelty that is both unfamiliar and of lower mainstreamness.

2. *Mainstream upholder*: Subjects with a listening history marked by mainstream artists, and which clearly prefer novelty of higher mainstreamness and familiarity (n=1,552).

3. *Mainstream explorer*: Listeners with noted preference for novel artists from the mainstream, but who value new artists that are not familiar. Seem to be exploring and discovering relevant artists among the mainstream that were previously outside their profile, which is itself formed by a high number of clusters (n=1,535).

4. *Crowd follower*: Subjects with a limited number of clusters in their listening profile, and with a clear preference for mainstream novelty (n=2130).

5. *Niche radical*: Listeners who habitually focus on a small number of clusters of artists, and which have a strong preference for novelty familiar and of below-average mainstreamness, likely in one or more niches (n=1,172).

6. *Highly eclectic*: Somewhat the opposite of niche radicals, these are subjects whose listening profiles had high number of artist clusters, and who value novelty already familiar (n=1,874).

7. *Underground*: Subjects who usually listen to artists far from the mainstream, have a relatively homogeneous listening profile, and prefer novelty close to this profile (n=1,143).

Interestingly, the largest single cluster in our results — and simultaneously one of the most distinct of them — is that of subjects fond of surprises. These are listeners who exhibit a clear preference for diversity and novelty among lesser known artists. Our results point that subjects in this group are interested in discovering new artists in the long tail of popularity, and at the same time exploring unfamiliar genres, even though their profile is, on average formed by mainstream artists. As an example of fond of surprises, a listener has a profile composed by *Indie* artists, but he preferred as novel artists some *Country* unpopular artists.

Diametrically opposite to subjects in this cluster, those in the niche follower group seem eager to discover more music similar to that they have listened in the past. Moreover, what was listened in the past is focused on a small area of the artists space, and novelty is more appreciated if it has low mainstreamness.

Not surprisingly, taken together, subject groups related to a marked preference for mainstream novel artists (mainstream upholders, mainstream explorers and crowd followers) form a large portion (43%) of our sample. Nevertheless, we see a distinction between listeners who have a markedly mainstream listening history prior to the experiment (mainstream upholders), and those who don't. As an example of mainstream upholder, a listener has a profile composed by popular *Indie* artists and he liked others popular *Indie* artists as novel items.

Mainstream explorers seem to be subjects who discovered a set of new mainstream artists during the experiment period that are different from their previous profile. As an instance of this group, a listener has a profile composed of *Sludge Metal* and *Hardcore* artists, but he preferred some popular *Hip-Hop* artists as novel items. Crowd followers, on the other hand, are subjects with a simpler profile previous to the experiment, and which enjoined novelty according to mainstreamness, irrespective of similarity to their previous profiles.

The underground cluster seems to capture listeners who devote their attention to a set of artists extremely different from the mainstream. Finally, the highly eclectic cluster models the profile of listeners who transit among many artist clusters, but which have marked preferences for novelty based on this diverse profile.

7. DISCUSSION AND IMPLICATIONS

The main objective of this study is to investigate the use of multiple dimensions of novelty to further understand listeners' preferences regarding novel artists. This understanding, in turn, is aimed at improving the design of music information retrieval systems.

Our main findings are threefold. First, there is no overall correlation between familiarity of mainstreamness and relevance in the novel artists discovered by our subjects. Second, when considered individually, most subjects had a clear correlation between either familiarity or mainstreamness and the relevance seen in novel artists. Third, it is possible to cluster listeners with some correlation between novelty aspects and preferences in seven groups that reveal how different audiences approach novelty.

Considered together, these results point to the need of a personalized approach in assisting a listener to discover relevant novel artists. Moreover, this personalization should be related not only to which other artists have been enjoyed before, but should be also aware that some listeners have clear preferences regarding how mainstream or familiar novel artists are. The information that these dimensions are relevant to model listeners' behavior, and that it is highly variable among a listener population should be considered in the design of future mechanisms. On another perspective, the groups we find to describe the archetypical behaviors among our subjects can be used to develop interfaces or mechanisms that target different types of users.

There are a number of opportunities on our study and directions in which future work could expand it. Notedly, considering data from a population outside Last.fm is necessary to evaluate the generalizability of our results. Also, our clustering is aimed at a descriptive analysis. Finding the listener clustering that provides maximum efficacy gain to a predictive model is likely a fruitful avenue of work. Finally, the validation of the use of familiarity and mainstreamness as dimensions of novelty in an experiment of musical recommendation is necessary to further validate the use of multiple dimensions to address novelty.

8. REFERENCES

[1] A. Bellogin, I. Cantador and P. Castells "A Study of Heterogeneity in Recommendations for a Social Music Service," *Proc. of the 1st HetRec*, pp. 1–8, 2010.

[2] Ò. Celma, P, Herrera: "A new approach to evaluating novel recommendations," *Proc. of the ACM RecSys 2008*, pp. 179–186, 2008.

[3] M. Ester, H. Kriegel, J. Sander and X. Xu: "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proc.of the ACM KDD, pp. 226-231*, pp. 226–231, 1996.

[4] J. L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, Vol. 22, No. 1, pp. 5–53, 2004.

[5] S. M. McNee, J. Riedl and J. A. Konstan "Being accurate is not enough: how accuracy metrics have hurt recommender systems," *Proc. of CHI '06*, pp. 1097–1101, 2006.

[6] S. Munson and P. Resnick "Presenting diverse political opinions: how and how much," *Proc. of CHI*, pp. 1457–1466, 2010.

[7] M. Nakatsuji, Y. Fujiwara, A. Tanaka, T. Uchiyama, K. Fujimura, T. Ishida: "Classical music for rock fans?: novel recommendations for expanding user interests," *Proc. of the 19th ACM CIKM*, pp. 949–958, 2010.

[8] S. Vargas and P. Castells: "Rank and relevance in novelty and diversity metrics for recommender systems," *Proc. of the ACM RecSys 2011*, pp. 109–116, 2011.

[9] J. H. Ward Jr. "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, Vol. 58, No. 301, pp. 236-244, 1963.

[10] Y. C. Zhang, D. O. Séaghdha, D. Quercia and T. Jambor: "Auralist: introducing serendipity into music recommendation," *Proc. of the WSDM*, pp. 13–22, 2012.

[11] C. Ziegle, S. M. McNee, J. A. Konstan, G. Lausen: "Improving recommendation lists through topic diversification," *Proc. of the 14th WWW*, pp. 22–32, 2005.