

Location-Adapted Music Recommendation Using Tags

Marius Kaminskas and Francesco Ricci

Free University of Bolzano,
Piazza Domenicani 3, 39100 Bolzano, Italy
{mkaminskas,fricci}@unibz.it

Abstract. Context-aware music recommender systems are capable to suggest music items taking into consideration contextual conditions, such as the user mood or location, that may influence the user preferences at a particular moment. In this paper we consider a particular kind of context aware recommendation task — selecting music content that fits a place of interest (POI). To address this problem we have used emotional tags attached by a users' population to both music and POIs. Moreover, we have considered a set of similarity metrics for tagged resources to establish a match between music tracks and POIs. In order to test our hypothesis, i.e., that the users will reckon that a music track suits a POI when this track is selected by our approach, we have designed a live user experiment where subjects are repeatedly presented with POIs and a selection of music tracks, some of them matching the presented POI and some not. The results of the experiment show that there is a strong overlap between the users' selections and the best matching music that is recommended by the system for a POI.

Keywords: recommender systems, location-aware, context, music, social tagging, emotions.

1 Introduction

Music recommender systems are decision support tools that reduce the information overload by retrieving relevant music items based on a user's profile. The recommendation process can be content-based, i.e., using features of the music liked by the user to predict what the target user may like [4], or collaborative-based, which finds users with similar music preferences and recommend to the target user items liked by these users [7]. However, most of the available music recommender systems suggest music regardless of the contextual conditions which can be important to predict the user's preferences at a particular moment. In fact, a study on users' musical information needs [6] has shown that people often seek music for a contextual situation like an occasion, an event or an emotional state rather than by artist or song information. To cope with these needs, recently there has been an emerging interest in contextual, or situational music selection [8,2]. The idea of such music selection is to recommend music

depending on the user's actual situation, emotional state, or any other contextual condition that might be relevant to increase the user's satisfaction for the selected music.

In this line of research, we are considering the problem of retrieving music that fits the surroundings of the user, and more specifically, we are looking for effective ways to match musical content with places of interest (POIs). For instance, a Vivaldi's concerto seems a better choice for a "calle" (a narrow street) in Venice than a Schubert's lied, that may better suit an old street in Vienna. Being able to select music for a place can be used for creating new engaging location-aware music delivery services. In particular, we are considering a scenario where a tourist is sightseeing a city using a mobile city guide (Android). It recommends a walking itinerary and, while the user is visiting the suggested POIs, it plays music that, according to the technique described in this paper, matches the visited POIs. The goal is to enhance the user's experience and create a more engaging travel guide tool. This application will be described in a forthcoming paper; here we want to focus on the enabling music to POI adaptation technologies.

It is clearly challenging to match music to a place so that the user can recognize this adaptation or, even without explicitly recognizing it, the user can appreciate such a selection and prefer it to other music not explicitly matching the place. The core technical issue to be solved is related to the fact that music and POIs are different objects and there is no obvious way to match one with the other. In recommender systems literature, it is normal to establish the similarity of two items, either using their feature based descriptions [11], or using their ratings given by a set of users [14]. The first approach requires that the two items, whose similarity is sought, share a common set of features, while the second one requires that a large number of users co-rated the two items. The first approach is therefore difficult to apply when the two items are not of the same type, while the second would only predict that a user that likes (dislikes) the first item will also like (dislike) the second. But this is not really a sign that the two items match together and they can, for instance, be recommended together. The problem of matching POI with a music track is more closely related to that found in cross-selling, e.g., recommending a type of boots that suit a kind of ski. This is a rather unpopular recommendation problem, that have only be considered by applications that recommend a good bundling of items, e.g., a travel planning [13] or music compilation [1].

In the research described here we decided to use the first approach mentioned above, and we looked for a common set of tags as shared representation for music and POIs. With such a representation the matching can then be performed by comparing the tag profiles of the two items. Since emotionally related tags can be used both for POIs and music we decided to explore this path. Music and places can both raise emotions and we conjectured that the commonality of the raised emotions could provide the base for establishing a degree of match between a place and a music track. Moreover, using tags to describe both music and POIs

is a promising and viable approach since there is a rapid growth of the amount of user-generated tagging data (a phenomena also known as social, or collaborative tagging) [3,15].

In a previous work [5] we conducted a preliminary user study aiming at verifying that matching music to places can enhance the user satisfaction for recommended itineraries. In that study the tags used for describing music and POIs were not only describing emotions and included a standard set of English descriptive adjectives. Furthermore, we used a rather simple similarity metric without comparing it to other competing metrics. This paper extends substantially that early work in three directions:

- A well founded set of emotional tags, suggested by recent studies on music cognition [16] is used for representing music and POIs.
- A set of similarity metrics, which were previously tested in a different setting [10], is evaluated first offline, comparing them in term of the generated ranked music recommendations for a POI, and then with a live user study where users expressed their subjective evaluations for the proposed match between music and POI.
- The matching of music to POIs is carefully evaluated for each considered POI as opposed to the previous study where the user feedback was collected for itineraries, i.e., a collection of three POIs.

In conclusion, in this paper we show that matching POIs with music is feasible. We have found that the proposed approach can recommend music tracks that users consider more suitable for a POI, compared to other apparently similar tracks but not matching the POI. This is an interesting result as it makes it possible to develop context-aware music retrieval systems that can enhance the user experience in many practical applications, such as travel guides, music recommender systems, automatic soundtrack generation systems.

The rest of this paper is structured as follows. We first describe the process of collecting the tagging data for music and POIs as well as present the statistics of collected data. Then we describe the used similarity metrics and the process of evaluating our matching strategies with real users. We then review some of the related work, and finally we draw conclusions and discuss some future work.

2 Tagging Music and POIs

As we mentioned above, in this research we have investigated if emotional tags attached to music and POIs can be used as a common set of descriptive features for establishing a match between these two types of items. Since emotionally tagged music tracks or POIs are not directly available, we have designed a web interface to collect this type of tagging data (Figure 1). We have considered 75 classical music tracks and movie soundtracks, and 50 POIs in the city of Bolzano and surrounding areas. The tagging was performed by volunteer users; mostly students and researchers at the Free University of Bolzano. The users were asked to tag the items using a restricted tag vocabulary consisting of adjectives from

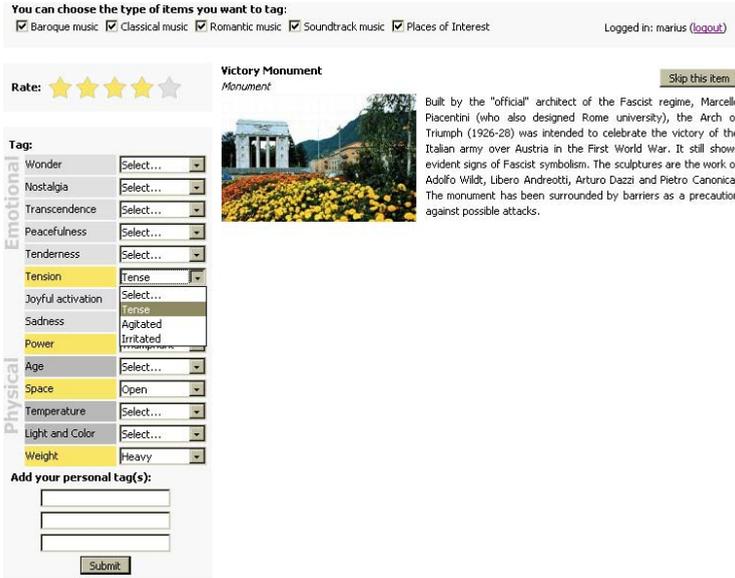


Fig. 1. Tagging interface

Table 1. Emotional tags from the GEMS model

Category	Tags
<i>Wonder</i>	Allured, Amazed, Moved, Admiring
<i>Transcendence</i>	Fascinated, Overwhelmed, Thrills, Transcendence
<i>Tenderness</i>	Mellowed, Tender, Affectionate, In love
<i>Nostalgia</i>	Sentimental, Dreamy, Melancholic, Nostalgic
<i>Peacefulness</i>	Calm, Serene, Soothed, Meditative
<i>Power</i>	Triumphant, Energetic, Strong, Fiery
<i>Joyful Activation</i>	Joyful, Animated, Bouncy, Amused
<i>Tension</i>	Tense, Agitated, Irritated
<i>Sadness</i>	Sad, Tearful

the Geneva Emotional Music Scale (GEMS) model described in [16]. The GEMS model consists of nine groups of emotions, each group having 2-4 emotional tags (Table 1). In addition to the emotional tags from GEMS model, we used tags describing physical characteristics of items that proved to be useful in a previous user study [5]. These are: Ancient, Modern — Colorful, Bright, Dark, Dull — Open, Closed — Light, Heavy — Cold, Mild, Warm.

The tagging was performed by 32 volunteer users. In total, 817 tags were collected for the POIs (16.34 tags per POI on average) and 1025 tags for the music tracks (13.67 tags per track on average). We call the tag-based description of an item its tag-profile. Figure 2 shows tag usage statistics grouped by tag categories. These results show that certain types of emotional tags have been applied to both music tracks and POIs with similar probabilities. This particularly applies to the categories “Peacefulness” and “Power”. Whereas some

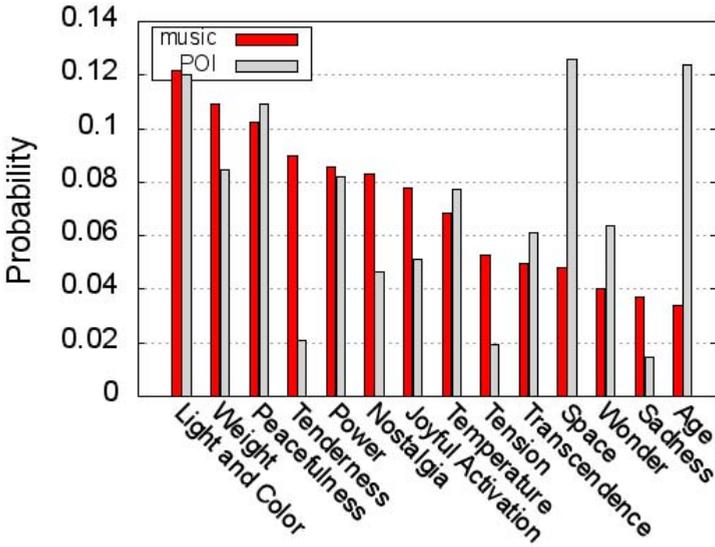


Fig. 2. Tag usage probability across tag categories

other categories were clearly used more often for music than POIs. These include “Tenderness”, “Nostalgia”, “Tension”, and “Sadness”. Tags in these categories are not likely to be useful for direct matching of music and POI, at least for our selection of POIs. From the physical tag categories, “Light and Color”, “Weight” and “Temperature” are the most likely to be useful due to the similar usage of such tags in both music and POI profiles. Contrastingly, tags describing age and space of items were applied mostly to POIs. These findings are interesting per se, but also indicate that an effective similarity metric for this task should be robust against the differences in the overall tags distributions observed for the two types of items. We observe that in previous research works that used tag-based similarity metrics this was not an issue as the items to be matched were confined to be in a single domain. Observing that the tags in the same emotion category are synonyms, we have also considered a simpler model where the original tags are replaced with their emotion category. For instance, the tags *allured*, *amazed*, *moved* and *admiring* appearing in any item tag-profile were substituted with *wonder*. Such merging of tags improved the tag coverage and reduced the dimensionality of item profiles from 46 (the initial number of individual tags) to 22 (9 emotion categories + 13 physical tags). In the next section we will present some similarity metrics that are either using the original tag-profile or this more compact tag-profile based on the adjectives’ categories.

3 Similarity Metrics

In order to match POIs with music, we decided to consider a well established set of similarity metrics that are applicable to tagged resources. Markines et al. [10]

evaluated the performance of some similarity metrics using classical IR evaluation measures, when computing the similarity between tagged resources. However, this study was conducted on a single folksonomy data set (BibSonomy.org — a social bookmarking system), with the task being to predict URL-to-URL similarity. The ground truth for resources' similarity was the graph-based similarity of URLs. Since our task was to match tagged objects from different domains (music and POIs) where the ground-truth similarity could only be assessed by subjective users' evaluations, we could not directly rely on the outcome of that study; we had to evaluate these metrics for our specific task.

In the following definitions, u and v represent items (either a music track or a POI), y represents a tag, X_u — the set of tags with not null frequency in the tag-profile of the item u , w_{uy} — the frequency of tag y in the item's u tag-profile and $p(y)$ — the fraction of items (both music tracks and POIs) annotated with y .

- The Matching similarity between two items is defined as:

$$\text{MatchingSim}(u, v) = \frac{\sum_{y \in X_u \cap X_v} \log p(y)}{\sum_{y \in T} \log p(y)}$$

where T is the set of all tags in the database.

- The Overlap similarity:

$$\text{OverlapSim}(u, v) = \frac{\sum_{y \in X_u \cap X_v} \log p(y)}{\max(\sum_{y \in X_u} \log p(y), \sum_{y \in X_v} \log p(y))}$$

- The Jaccard similarity:

$$\text{JaccardSim}(u, v) = \frac{\sum_{y \in X_u \cap X_v} \log p(y)}{\sum_{y \in X_u \cup X_v} \log p(y)}$$

- The Dice similarity:

$$\text{DiceSim}(u, v) = \frac{2 \sum_{y \in X_u \cap X_v} \log p(y)}{\sum_{y \in X_u} \log p(y) + \sum_{y \in X_v} \log p(y)}$$

- The Cosine similarity:

$$\text{CosineSim}(u, v) = \frac{\sum_y w_{uy} w_{vy}}{\sqrt{\sum_y w_{uy}^2} \sqrt{\sum_y w_{vy}^2}}$$

- The TF-IDF similarity: in addition to the above metrics, we have introduced a variation of Cosine similarity metric, where the tag frequencies w_{uy} in an item profile X_u are replaced with the TF-IDF weight of tag y . In order to compute the TF-IDF weights, for each item u all the tags assigned to the item (with repetitions) are considered as a document representing the item.

The usage of logarithms in the first four metrics is related to Shannon information theory. Intuitively, a very common tag will have a high probability and

therefore a very small log probability. Thus, it will bring a small contribution to the similarity score. We observe that all six metrics range in $[0, 1]$, making their comparison easier. Moreover, we note that they can be applied both to the original tag-profiles and to the merged tag-profiles mentioned previously. As a result, we have 12 different ways to compute the similarity between a music track and a POI.

4 System Evaluation

Since there is no accessible ground-truth telling if a music track suits a POI, we designed a live user study where the user is requested to assess that relationship. However, since a user cannot express many of these subjective judgments in a single session we had to narrow down the number of similarity metrics compared in the live user study. For that reason, the full evaluation of the effectiveness of the matching computed by the considered similarity metrics was carried out in two steps. First, we computed the correlation of the ranked lists produced by the considered similarity metrics when matching (scoring) the available music tracks to a given POI. The more correlated two rankings are the more similar will be the recommended music tracks for a given POI. This initial analysis allowed us to study the general properties of the metrics, and to discard some redundant metrics. Second, the selected metrics were used in the live user experiment to collect the users' subjective evaluations, i.e., if a music track suits a POI.

4.1 Offline Similarity Metrics Comparison

In order to compare the considered similarity metrics, without relying on subjective user-based evaluations, we sorted the music tracks recommended for a given POI using the different similarity metrics and we computed, pairwise, the Spearman's correlation of these ranked lists produced by the different metrics. Averaging, for each pair of metrics, the correlation of the ranked lists of music tracks generated by two metrics for all the POIs in our database we produced an average correlation score between every pair of metrics.

When comparing the similarity metrics, using the original tag-profiles, i.e., without merging the count for the tags in the same emotional category, we observed two clusters of metrics: [Matching, Overlap, Jaccard, Dice] all have an average correlation higher than 0.8 between each other. Likewise, Cosine and TF-IDF similarities have a correlation higher than 0.7, but were less correlated with the metrics in the first cluster (e.g., correlation of TF-IDF with Overlap was 0.64, Cosine with Jaccard — 0.68). The same clusters of metrics were observed when we compared the similarity metrics computed on the merged tag-profiles. Hence, relying on these results we have selected one representative metric for each cluster; i.e., Jaccard for the first and Cosine for the second. We observe that there is a major difference between these two metrics: cosine considers tag frequency in items' profiles while Jaccard only considers each co-occurring tag once. In the online evaluation study with real users, which is described later, we used these two metrics.

Table 2. Average Spearman’s correlation of different similarity metrics. The asterisk marks similarity computed when merging the emotional tags. Grey cells indicate an average correlation lower than 0.7.

	Matching	Overlap	Jaccard	Dice	Cosine	TF-IDF	Matching*	Overlap*	Jaccard*	Dice*	Cosine*	TF-IDF*
Jaccard	0.94	0.92	1.00	1.00	0.82	0.70	0.70	0.60	0.72	0.72	0.64	0.70
Cosine	0.78	0.78	0.82	0.82	1.00	0.73	0.64	0.58	0.67	0.67	0.76	0.73
Jaccard*	0.75	0.65	0.72	0.72	0.67	0.88	0.93	0.83	1.00	1.00	0.68	0.85
Cosine*	0.63	0.60	0.64	0.64	0.76	0.83	0.62	0.63	0.68	0.68	1.00	0.83

The effect of merging the tag profiles is evident by looking at Table 2, where the average correlation of the four selected metrics with all the initially considered similarity metrics is shown. The average correlation of the ranking produced by using the merged tag-profiles and the original tag-profiles is much lower; this can be seen by comparing the figures shown in the first (second) and third (fourth) rows of this table (e.g., Jaccard-Matching is 0.94, vs. Jaccard*-Matching is 0.75). Whereas the differences in the correlation between the same pair of metrics, when applied to a merged or not merged pair of tag-profiles, is not large. This can be seen comparing the first six entries of the first (second) row with the corresponding last six entries of the third (fourth) row. In other words, if two metrics are correlated when using the not merged tag-profiles they are still correlated when using the merged profiles.

4.2 Online User Study

In the final and more important evaluation experiment we measured whether the users consider the music tracks with largest similarity to a POI as those really better suited for that POI. In order to perform this evaluation we have designed a second web interface (Figure 3). Here, the users were repeatedly asked to consider a POI (taken from a POI collection), and while looking at this POI, to listen to some selected music tracks. The user was asked to check all the tracks that in her opinion suit that POI.

During each evaluation step the music recommendations for a POI were selected using two out of the four considered similarity metrics. The selected tracks included the two best matching tracks for each metric (highest similarity). In addition, we introduced in the recommendation lists two tracks that were mostly different from the matching tracks, i.e., having low similarity to the given POI. Introducing the low similarity tracks allowed us to directly compare tracks that were supposed to fit the POI with those not. In total a maximum of six tracks were suggested for each POI, but usually less tracks were shown as the tracks selected by the similarity metrics often overlap.

The goal of this analysis was to see whether the users actually agree with the music-to-POI matching computed using our approach. We note that the outcome

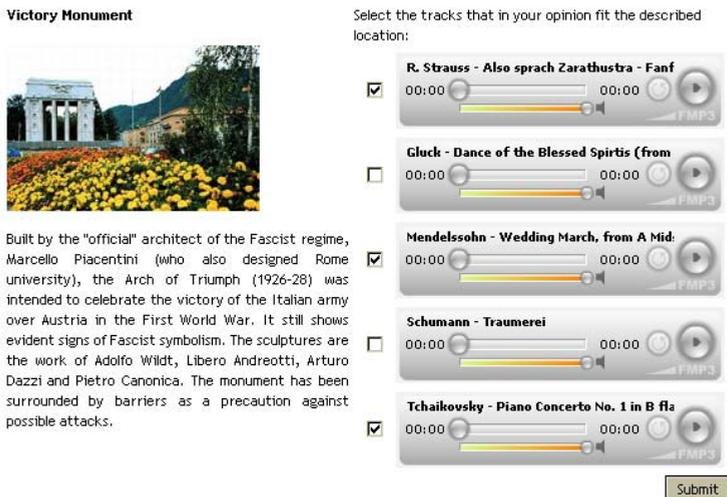


Fig. 3. The interface for evaluating music matching to POI

of this evaluation was not evident at all, since with a superficial evaluation, even the less similar tracks could be considered as suited and there are not large differences among the considered tracks (all of them are popular classical music tracks).

For example, consider the evaluation step shown in Figure 3. The POI Victory Monument was tagged as *bright, heavy, open, strong, triumphant, tense*, etc. In this case, the two metrics used to select the tracks are: Jaccard (suggesting tracks 1 and 5), and Jaccard with merged tag profiles (suggesting tracks 3 and 5). The low similarity tracks are tracks 2 and 4. Track 1 has been tagged as *open, heavy, triumphant, amazed* etc.; track 3 — *open, bright, agitated, bouncy, in love, triumphant*, etc.; track 5 — *open, heavy, triumphant, strong, cold*, etc. Contrastingly, tracks 2 and 4 have been tagged as *serene, light, colorful* etc. Looking at the tag profiles, it is easy to understand why the similarity metrics suggest tracks 1,3 and 5. However, the user is neither aware of the items' tag profiles, nor of the different ways the tracks were selected. It was therefore crucial to see if a person, just by listening to the selected music tracks, would agree with the match produced by our approach.

The online evaluation was carried out by 10 users in total performing 154 evaluation steps, that is, each user considered on average 15.6 POIs and the music suggested for these POIs. In order to compare the effectiveness of different metrics in selecting the best tracks, we have computed the probability that a metric produces a music track that is considered suited for a POI by the users. The probability was computed as the ratio of the number of times any track produced by a metric was selected over the total number of evaluation steps where this metric was used, i.e., tracks produced by this metric were presented.

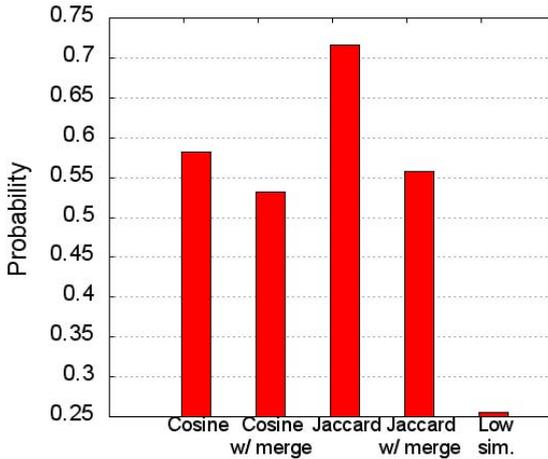


Fig. 4. The selection probability for the five evaluated matching approaches

Note that each time a music track, which was suggested by multiple metrics, was selected as appropriate for a POI, the probabilities for all these metrics were increased.

Looking at the results of this experiment (see Figure 4) it is clear that all the four tested similarity metrics performed significantly better than the low similarity matching (99% confidence level of the two-proportion z-test). Among the 4 metrics, Jaccard performs significantly better than the others (95% confidence level of the two-proportion z-test). A possible reason for the better performance of Jaccard compared to Cosine is that Jaccard metric in contrast to Cosine uses the probability that a tag can be found in a corpus; thus a frequent tag contributes to the similarity score less than a rare tag. In conclusion, we can affirm that the users consider the music tracks suggested by our approach as more suited for POIs than other not matching tracks. Furthermore, Jaccard selects the tracks that the users most frequently choose as suited for the illustrated POIs.

5 Related Work

Matching music with the user’s location has not been investigated so far. The most closely related works are in the area of context-aware music recommendation. For instance, in [12,8] music is adapted to environment parameters, such as the location of the user, the time or the weather conditions. In other research works emotions have been used as a link between music and other types of content, e.g., websites [2] and images [9]. [2] mined emotions directly from textual representation of websites and music tracks, while [9] used a machine learning approach trained on hand labeled training data. We note that these works used emotion models that are different from the one used in our work.

Other research activities have been conducted on tagging behavior and tag distribution in social tagging systems [3,15]. However, to the best of our knowledge the usage of social tags for matching heterogeneous objects has not been investigated so far. We have used the similarity measures presented in [10] as a reference point, but we could not rely on their evaluation results since our study deals with different types of resources and a different ground truth. However, in [10], Matching, Overlap, Dice and Jaccard metrics performed slightly better than Cosine metric — a result that was also observed in our experiment.

6 Conclusions and Future Work

In this paper we have analyzed a new problem in music recommendation, i.e., recommending music tracks that suit a place of interest (POI). We have developed an approach that exploits user assigned emotional tags to both music tracks and POIs. We have collected and analyzed the tagging data obtained from real users through a custom developed interface that enables users to uniformly tag both music tracks and POIs. Then, we have performed an online experiment where the users were required to evaluate the appropriateness of the music selected by the system for POIs. The results show that users tend to agree with the matching produced using our proposed approach. Moreover, Jaccard similarity measure was shown to produce the matching preferred by most of the users.

We are currently testing the music-to-POI matching approach on a mobile device. We have developed a mobile tourist guide for the city of Bolzano. The guide offers an itinerary and plays matching music tracks as the user approaches a POI. We intend to collect the user feedback and further explore the effects of such matching on users' experience. An important next step is understanding if certain tags (emotions) contribute more to the perceived match between a POI and a music track (currently all tags have equal importance). Another important future step is to move from manually labeled data towards automatic tag acquisition for both music and POIs. We intend to investigate public folksonomies (e.g., Flickr, Last.fm) and blogs as possible data sources. Additionally, we want to study if personal preferences should be taken into account in this task as currently the same match is provided for all the users. Furthermore, we intend to use ontologies as an additional way to match music with POIs.

References

1. Baccigalupo, C., Plaza, E.: Case-based sequential ordering of songs for playlist recommendation. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds.) ECCBR 2006. LNCS (LNAI), vol. 4106, pp. 286–300. Springer, Heidelberg (2006)
2. Cai, R., Zhang, C., Wang, C., Zhang, L., Ma, W.-Y.: MusicSense: contextual music recommendation using emotional allocation modeling. In: Proceedings of the 15th International Conference on Multimedia, pp. 553–556. ACM, New York (2007)
3. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. *Journal of Information Science* 32(2), 198 (2006)

4. Hoashi, K., Matsumoto, K., Inoue, N.: Personalization of user profiles for content-based music retrieval based on relevance feedback. In: Proceedings of the Eleventh ACM International Conference on Multimedia, pp. 110–119. ACM, New York (2003)
5. Kaminskas, M., Ricci, F.: Matching places of interest with music. In: Workshop on Exploring Musical Information Spaces, WEMIS 2009, pp. 68–73. University of Alicante (2009)
6. Kim, J.Y., Belkin, N.J.: Categories of music description and search terms and phrases used by non-music experts. In: Proceedings of the 3rd International Conference on Music Information Retrieval, Paris, France (2002)
7. Konstas, I., Stathopoulos, V., Jose, J.M.: On social networks and collaborative recommendation. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 195–202. ACM, New York (2009)
8. Lee, J.S., Lee, J.C.: Context awareness by case-based reasoning in a music recommendation system. In: Ichikawa, H., Cho, W.-D., Chen, Y., Youn, H.Y. (eds.) UCS 2007. LNCS, vol. 4836, pp. 45–58. Springer, Heidelberg (2007), http://dx.doi.org/10.1007/978-3-540-76772-5_4
9. Li, C.-T., Shan, M.-K.: Emotion-based impressionism slideshow with automatic music accompaniment. In: Proceedings of the 15th International Conference on Multimedia, pp. 839–842. ACM, New York (2007)
10. Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., Stumme, G.: Evaluating similarity measures for emergent semantics of social tagging. In: Proceedings of the 18th International Conference on World Wide Web, pp. 641–650. ACM, New York (2009)
11. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. *The Adaptive Web*, 325–341 (2007)
12. Reddy, S., Mascia, J.: Lifetrak: music in tune with your life. In: HCM 2006: Proceedings of the 1st ACM International Workshop on Human-centered Multimedia, pp. 25–34. ACM, New York (2006)
13. Ricci, F., Cavada, D., Mirzadeh, N., Venturini, A.: Case-based travel recommendations. In: Fesenmaier, D.R., Woeber, K.W., Werthner, H. (eds.) *Destination Recommendation Systems: Behavioural Foundations and Applications*, pp. 67–93. CABI (2006)
14. Schafer, J.B., Frankowski, D., Herlocker, J.L., Sen, S.: Collaborative filtering recommender systems. *The Adaptive Web*, 291–324 (2007)
15. Sen, S., Lam, S., Rashid, A., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F., Riedl, J.: Tagging, communities, vocabulary, evolution. In: Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 181–190. ACM, New York (2006)
16. Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion* 8(4), 494–521 (2008)