# Location-Aware Music Recommendation Using Auto-Tagging and Hybrid Matching

Marius Kaminskas Free University of Bozen-Bolzano Faculty of Computer Science Bozen-Bolzano, Italy mkaminskas@unibz.it Francesco Ricci Free University of Bozen-Bolzano Faculty of Computer Science Bozen-Bolzano, Italy fricci@unibz.it Markus Schedl Johannes Kepler University Department of Computational Perception Linz, Austria markus.schedl@jku.at

# ABSTRACT

We propose a novel approach to context-aware music recommendation – recommending music suited for places of interest (POIs). The suggested hybrid approach combines two techniques – one based on representing both POIs and music with tags, and the other based on the knowledge of the semantic relations between the two types of items.

We show that our approach can be scaled up using a novel music auto-tagging technique and we compare it in a live user study to: two non-hybrid solutions, either based on tags or on semantic relations; and to a context-free but personalized recommendation approach. In the considered scenario, i.e., a situation defined by a context (the POI), we show that personalization (via music preference) is not sufficient and it is important to implement effective adaptation techniques to the user's context. In fact, we show that the users are more satisfied with the recommendations generated by combining the tag-based and knowledge-based context adaptation techniques, which exploit orthogonal types of relations between places and music tracks.

# **Categories and Subject Descriptors**

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Information Filtering

### **General Terms**

Algorithms, Design, Experimentation, Human Factors

# Keywords

Music recommendation; context-aware recommendation; hybrid recommendation; auto-tagging; tag prediction

### 1. INTRODUCTION

Music recommender systems are decision support tools that reduce information overload by retrieving relevant music items based on a user's profile [8]. However, most of the

RecSys '13 Hong Kong, China ACM 978-1-4503-2409-0/13/10. http://dx.doi.org/10.1145/2507157.2507180. available music recommenders suggest music without taking into consideration the user's context, e.g., her mood, current location, activity, or any other contextual condition that might influence the user's perception or evaluation of music [1]. In response to these observations, in recent years a new research topic of contextual music retrieval and recommendation has emerged [12]. In this work we address a particular type of context-aware music recommendation - recommending music suited for places of interest (POIs). Finding music items that suit POIs can be exploited in a number of engaging information services. In particular, we have considered a scenario in which a mobile city guide provides an enhanced presentation of a place for a tourist by playing music that is related to the place, i.e., music that is culturally or emotionally associated with the place (e.g., a Mozart sonata in Salzburg or a Bach fugue in a Gothic Cathedral) [6].

The main challenge to face when addressing the above mentioned goal is related to the fact that POIs and music are two rather different domains, and there is no obvious way to match such heterogeneous items. In previous works [6, 11], we have developed two alternative techniques to solve the problem of matching music and POIs – one based on a common representation of both types of items, and the other exploiting semantic relations between the items.

The first technique represents items in the two domains with a common set of features – user-assigned tags describing the emotional properties of music and POIs [6]. We decided to use tags that describe emotional properties of the items, since music and places can both raise emotions, and the commonality of these emotions could provide the base for establishing a degree of match between a place and a music track. The tag-based representation allows matching music tracks and POIs by comparing the tag profiles of the items.

The second technique is based on extracting explicit knowledge about items in the two domains [11]. Obtaining such knowledge automatically has become possible with the advent of the Semantic Web and specifically with its reference implementation in the Linked Data initiative [5]. We have exploited the DBpedia Linked Data repository to build a framework in which semantic networks linking items from music and POI domains are automatically created. These networks are represented as graphs, on which a graph-based algorithm is run to rank the items in the music domain with respect to their relatedness to an item in the POI domain.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

These two techniques - tag-based and knowledge-based represent two complementary ways of establishing a match between a place and a music track: a track may "feel right" for a POI or it can be linked to a POI by factual relations (e.g., belonging to the same cultural era or composed by someone whose life is related to the POI). In previous works, we have evaluated the two techniques independently on different datasets and against a simple baseline (random recommendations). In this paper, we take some next steps toward location-aware music recommendation by (i) proposing a hybrid recommendation approach which combines the tag-based and knowledge-based techniques using rank aggregation, (ii) proposing an extension to the *tag-based* technique to scale up the tagging process for the music tracks employing a state-of-the-art auto-tagger, and (iii) evaluating the performance of these techniques on a dataset of POIs and music tracks against personalized recommendations. We have validated the research hypothesis that the combination of knowledge-based and tag-based techniques is preferred by the users over the recommendations produced by each technique separately and over the classical personalization approach.

The remainder of the paper is structured as follows. In Section 2 we overview works related to location-aware music recommendation and music auto-tagging. Section 3 describes the used dataset of music tracks and POIs. In Section 4 we describe the automatic tag prediction technique that we have used for tagging music tracks. Section 5 describes the user study to evaluate the proposed recommendation approach. Finally, we summarize the findings of the evaluation and define future work directions in Section 6.

### 2. RELATED WORK

Finding music that suits a POI can be viewed as a *context*aware recommendation problem [1], the place being the context for consuming the recommendations (music). There are several works on context-aware music recommendation that exploit location-related context information. Reddy and Mascia [18] present a mobile music recommender system Lifetrak that generates a playlist using the user's music library based on the current context of the user (location, time, weather, and activity information). Similarly to our tag-based approach [6], Lifetrak requires the user to label songs with tags from a controlled vocabulary. However, in contrast to our approach, the tags in the vocabulary directly represent the values of the previously mentioned context parameters. So for instance, songs have to be labeled with a ZIP code of a certain area to be recommended for that location.

More recently, Ankolekar and Sandholm [2] presented a mobile audio application that plays audio content associated with a particular location with the goal of enhancing the sense of being in a place by creating its emotional atmosphere. Instead of establishing relations between music and location automatically, the presented approach relies on crowd-sourcing – users are allowed to assign audio pieces (either a music track or a sound clip) to a specific location (represented by the geographic coordinates).

Our research is also related to *music auto-tagging*, the process of automatically assigning labels to music pieces. This task is typically performed employing a supervised learning approach; based on a training set of music content features and semantic labels, a classifier is trained and subsequently used to predict tags for unseen pieces. Rhythm or timbre descriptors are frequently used as content features [16] and sometimes high-level features are included [23].

Recent work in music auto-tagging includes [23], where Sordo proposed a simple though efficient algorithm based on a weighted vote k-Nearest Neighbour (kNN) classifier to propagate tags from training data to unseen music items. Moreover, in order to increase the computational efficiency, he used *feature selection* and *dimensionality reduction* of the input feature vectors before training the classifier. A similar approach is suggested in [13] for auto-tagging artists. Kim et al. analyzed artist similarity functions based on different data sources: artist co-occurrences in Last.fm playlists, Last.fm tags, web pages about the artists, and music content features. It is found that the similarity measure based on Last.fm co-occurrences performs best in terms of precision and recall. Mandel et al. [16] used conditional Restricted Boltzmann Machines [22] to learn tag language models over three sets of vocabularies: annotations by users of Amazon's Mechanical Turk, of the tagging game MajorMiner [15], and of Last.fm. Theses models are learned on the level of song segments. Optionally, track level and user level annotations are considered.

A recent trend in music auto-tagging is adopting two-stage algorithms. In the first stage, such algorithms infer higherlevel information from content features, such as term weight vector representations. These representations are then fed into a machine learning algorithm to learn semantic labels [9, 17]. For instance, Miotto et al. [17] first model semantic multinomials over tags based on music content features. To account for co-occurrence relations between tags, they subsequently learn a Dirichlet mixture model of the semantic space, which eventually yields a contextual multinomial used for tag prediction.

In this work, we employ a variant of the auto-tagger presented in [19], where Seyerlehner et al. proposed a combination of different audio features described within their block-level framework [21]. This tagger showed superior performance in the "Audio Tag Classification" and the "Audio Tag Affinity Estimation" tasks, run at the 2012 Music Information Retrieval Evaluation eXchange (MIREX)<sup>1</sup>. We extended the set proposed in [21] with two additional audio features.

# 3. THE DATASET

In order to evaluate the proposed hybrid recommendation approach, we required a dataset of POIs and music tracks on which both *tag-based* and *knowledge-based* techniques could be applied. Since the ground truth relatedness of music and POIs can only be assessed by subjective users' evaluations, and the users cannot perform a large number of judgments during an evaluation session, we have limited the size of the dataset to 25 well-known POIs from 17 major city tourism destinations in Europe (Madrid, Berlin, Florence, Paris, etc.).

Acquiring music tracks related to the selected POIs was performed by retrieving the top-5 musicians for each POI using the *knowledge-based* technique [11], and aggregating them into a single set. This resulted in a collection of 123 musicians (there were two repetitions in the initial list of 125 musicians). Subsequently, we retrieved 3 music tracks

<sup>&</sup>lt;sup>1</sup>http://www.music-ir.org/mirex

 
 Table 1: The controlled vocabulary used for tagbased matching of music and POIs

| Tags  | Type      |  |  |  |
|---|-----------|--|--|--|
| Affectionate, Agitated, Animated, Bouncy,         |           |  |  |  |
| Calm, Energetic, Melancholic, Sad, Sentimen-      | GEMS      |  |  |  |
| tal, Serene, Spiritual, Strong, Tender, Thrilling |           |  |  |  |
| Ancient, Modern, Bright, Dark, Colorful,          | Dhysical  |  |  |  |
| Heavy, Lightweight, Open, Cold, Warm              | 1 Hysical |  |  |  |

for each musician by taking the top-ranked results returned by the YouTube search interface. Doing so ensured the collected tracks to be representative of the musicians in our dataset. Eventually, a set of 369 music tracks belonging to 9 music genres was obtained: Classical, Medieval, Opera, Folk, Electronic, Hip Hop, Jazz, Pop, and Rock.

We note that using the *knowledge-based* technique, which retrieves musicians from the DBpedia knowledge base [11], allowed us to avoid the *popularity bias*, i.e. selecting only well-known musicians. In fact, the collected dataset covers a wide variety of less-known musicians from the so-called "long tail" of music content [7].

In order to evaluate the *tag-based* approach on this dataset, the POIs and music tracks had to be tagged using the controlled tag vocabulary as described in [6]. While tagging the 25 POIs was easily performed through a tagging survey, tagging 369 music tracks required substantial user effort and was not practical. Therefore, to address this issue we have selected a third of the music tracks (123 tracks – one random track per musician) to be tagged through a survey. The collected annotations were used as training data for automatically tagging the remaining tracks in the dataset (see next section).

The tagging survey was performed through a web interface where the users were asked to annotate POIs and music tracks using a controlled vocabulary consisting of adjectives from the GEMS emotion model [26] and adjectives describing physical characteristics of items (Table 1). In [6] we have shown that this tag vocabulary is suitable for annotating both music and POIs and performs well for establishing relations between the two types of items. Compared to our previous research, in this work we have used a revised version of the vocabulary. In fact, by analyzing the data collected during the previous tagging surveys [6], we observed that certain tags were rarely used and therefore were marginally providing useful information when recommending music for POIs. We therefore discarded the tags that in the previous surveys were applied to less than 1% of the items. The revision allowed us to reduce the size of the vocabulary from 46 to 24 tags, which was beneficial for the music auto-tagging procedure, in which a separate classifier is trained for each tag (see next section).

The tagging procedure was performed by 10 volunteers recruited via email – students and researchers from the Free University of Bolzano and other European universities. We note that the tagging of music and POIs is a subjective task and users may disagree whether certain tags apply to an item [24]. Consequently, aggregating tags provided by different users into items' tag profiles may result in contradicting tags. Therefore, to ensure the quality of the acquired tags, we considered the agreement between users, which is a standard measure of quality for user-generated tags [14]. We cleaned the data by keeping for each item only the tags on which at least 2 taggers agreed. As a result, we obtained an average of 5.1 distinct tags for the POIs and 2.5 for the music tracks. The lower number of tags for music tracks can be explained by the large number of items and the small number of taggers – on average a music track was tagged by 2 users. However, the quality of the acquired data was more important than the quantity, since it was crucial to train the music auto-tagger on reliable data. In the next section we will show how the evaluation of the auto-tagging algorithm allowed us to collect more tagging data for the music tracks.

# 4. MUSIC AUTO-TAGGING

We used a variant of the high-performance, state-of-theart auto-tagger proposed in [19]. It was applied in the data acquisition phase, to reduce the human effort required in the annotation process, and to scale up tagging. We hence used the tags manually obtained for a third of the music dataset to train the auto-tagger, and automatically label the remaining 246 music tracks.

Our auto-tagger is based on a set of audio features defined within the *block-level framework* (BLF) [21]. This framework describes a music piece by first modeling it as overlapping blocks of the Cent spectrum representation of the audio signal. More precisely, a window size of 2048 samples and a hop size of 512 samples are used. The Short Time Fourier Transform (STFT) is then computed on the Hanning-windowed samples of the audio signal. To account for the musical nature of the audio under consideration, the resulting magnitude spectrum with linear frequency resolution is mapped onto the logarithmic Cent scale.



Figure 1: Overview of the block-level framework.

Based on these Cent spectrum representations, we defined several features that are computed on blocks of frames (Figure 1). The features which describe a single block were eventually aggregated over all blocks of the piece under consideration, using a summarization function. This yielded the following music descriptors used in the auto-tagger: *Spectral Pattern* (SP) characterizes the frequency content, *Delta Spectral Pattern* (DSP) emphasizes note onsets, *Variance Delta Spectral Pattern* (VDSP) aims at capturing variations of onsets over time, *Logarithmic Fluctuation Pattern* (LFP) describes the periodicity of beats, *Correlation Pattern* (CP) models the correlation between different frequency bands, and Spectral Contrast Pattern (SCP) uses the difference between spectral peaks and valleys to identify tonal and percussive components. Figure 2 illustrates the different feature values for a Jazz piece. The y-axis represents the frequency bands and the x-axis the sorted temporal components of the blocks. In addition to these block-level descriptors, we further included two variants of Mel Frequency Cepstral Coefficient (MFCC) features, which are commonly used to capture music timbre [25, 3], although they originate from speech processing. Finally, the song level feature vectors resulting from these individual music descriptors were concatenated and associations between songs and tags learned using a Random Forest classifier. We used a RF classifier because in previous experiments its performance was among the best and comparable to that of a Support Vector Machine classifier using Principal Component Analysis. Moreover, RF has a much better computational performance [20].



Figure 2: BLF features for a Jazz piece.

#### **Music Auto-Tagger Evaluation** 4.1

As described in Section 3, the set of 123 music tracks was initially tagged by a small number of human taggers to acquire data for training the music auto-tagger. Following this initial data acquisition survey, which resulted in 2.5 tags per track on average, we evaluated the auto-tagger on the 123 tracks using leave-one-out train/test data split and crossvalidation. The results showed that the auto-tagger can accurately predict the small number of tags for the tracks. However, more rich tagging data had to be obtained before the auto-tagger could be trained and applied to the untagged portion of the dataset.

Therefore, we conducted a second tagging survey on the same set of 123 music tracks, aimed to measure how much the users agree with predictions generated by the auto-tagger and to acquire more tagging data. Figure 3 shows the interface of the web application used in this user study. The participants of the study were asked to tag the music tracks using the controlled tag vocabulary, with 6 of the usable tags pre-selected by the system: 3 tags predicted for the music track by the auto-tagger and 3 tags randomly selected from the remaining 21 tags in the vocabulary.

31 users participated in the study. There was no overlap with the users involved in the initial tagging of the tracks. On average, each music track was tagged by 2.6 users. The participants of the study were informed about the nature of pre-selected tags: "Note that some tags are pre-selected by a recommender system and some at random. If you disagree with any of the suggestions, please de-select them". This

| Tag: |              |   | Fritz Kreisler - Liebesfreud Skip this item |   |
|------|--------------|---|---|---|
|      | Melancholic  |   | Bright                                      | http://en.wikipedia.org/wiki/mtz_kreisien   |
|      | Heavy        |   | Animated                                    | 00:08 00:31 0   |
| ~    | Tender       |   | Energetic                                   | Triedrich Tritz' Kreisler (February 2, 1875 – January 29, 1962) was a<br>Austrian-born violinist and composer. One of the most famous violin master<br>of his or any other day, he was known for his sweet tone and expressiv<br>phrasing. Like many great violinists of his generation, he produced<br>characteristic sound which was immediately recognizable as his own. Althoug<br>he derived in many respects from the Franco-Belgian school, his style i<br>nonetheless reminiscent of the gemütlich (cozy) lifestyle of pre-war Vienna." |
|      | Cold         |   | Spiritual                                   |   |
| ~    | Modern       | ~ | Serene                                      |   |
|      | Ancient      |   | Calm  |   |
|      | Affectionate |   | Sad   |   |
| ~    | Dark         |   | Strong                                      |   |
| ~    | Lightweight  |   | Colorful                                    |   |
| ~    | Open         |   | Thrilling                                   |   |
|      | Warm         |   | Agitated                                    |   |
|      | Sentimental  |   | Bouncy                                      |   |

Submit

Figure 3: Interface of the experiment on automatic tag prediction.

information was provided to encourage the users' critical thinking when assessing the pre-selected tags.

In total the users performed 318 evaluation sessions (i.e., tagged a music track) assigning on average 5.29 tags to a track per session. Out of 1683 acquired music annotations, 652 were those predicted by the auto-tagger, 377 randomly pre-selected, and 654 newly assigned by the users. This shows that the users were active in assigning new tags, instead of merely revising the pre-selected tags.

We have calculated the probability for a predicted/random tag to be accepted by a user as the ratio of the number of times such tags were selected by users over the total number of times they were displayed by the system. The results (Table 2) show a higher selection probability for the predicted tags compared to random tags to be accepted. Note that the probability for a random tag to be selected by a user is still relatively large, which can be explained by two facts. Firstly, due to the large number of tags assigned to a music track per session, the baseline probability for any tag in the vocabulary to be selected by a user is large: 5.29/24 = 0.22. Secondly, by suggesting certain tags we created a presentation bias – the users were paying more attention to the pre-selected tags, even those at random. Nevertheless, the difference between the two obtained probabilities is statistically significant with p < 0.001 in a two-proportion z-test. Therefore, this result confirms that the users agree with the tag predictions produced by the auto-tagger.

Table 2: Selection probabilities of predicted and randomly pre-selected tags.

| Predicted tags | Random tags |
|----------------|-------------|
| 0.683          | 0.395       |

In addition to validating the effectiveness of the autotagger, the described experiment also allowed us to collect more tagging data for music tracks, since the study participants not only reviewed the predicted tags, but also added new ones. Upon the completion of the user study we had the 123 tracks tagged with 5.8 distinct tags on average.

Finally, having acquired sufficient tagging data for the 123 tracks and validated the effectiveness of the auto-tagging algorithm, we ran the auto-tagger on the full set of 369 tracks, training it on the 123 tagged tracks considered so far.

The auto-tagger outputs a probability for each tag in the vocabulary to be relevant for a track. However, the Jaccard similarity metric between sets of tags, which we use for the *taq-based* music-to-POI similarity computation, requires to compute the intersection and union of the items' tag profiles [6]. Therefore, we decided to generate binary tag assignments based on the probabilistic output of the auto-tagger by applying a threshold to the tag prediction probabilities. The standard value for the binarization threshold is 0.5, although it can be set higher or lower to trade precision with recall [20]. Typically, an optimal threshold value is determined using cross-validation (e.g., in [20] the authors report an optimal value of 0.25). However, in our research, the performance of the auto-tagger is defined by the quality of music-to-POI recommendations which have to be assessed by users. As it was not feasible to conduct multiple user studies with different binarization threshold values, we determined the threshold value based on the average size of items' tag profiles it produced. Empirical analysis showed that a threshold of 0.4 produced an average tag profile of 5.2 tags which is in accordance with the average profile size of manually tagged items (5.1 tags for POIs and 5.8 for music tracks).

# 5. EVALUATION AND RESULTS

As described in the previous sections, we have collected a dataset of 25 POIs and 369 music tracks from 9 music genres. Both *tag-based* and *knowledge-based* techniques can be applied to this dataset, since the music tracks and POIs were tagged using the controlled tag vocabulary, and the tracks belong to musicians semantically related to the POIs.

We designed an evaluation study to compare the performance of the two techniques and evaluate their combination. Additionally, we wanted to compare the performance of *tag-based* music recommendations obtained with manually vs. automatically tagged music tracks. Therefore, we considered two variants of the *tag-based* technique: one using only the manually annotated 123 tracks, and the other using the full set of 369 auto-tagged tracks. Moreover, we implemented a baseline approach which exploits only the users' music genre preferences. Hence, in total we considered five approaches to recommend music for POIs, which we describe in the next section.

### 5.1 Music Recommendation Approaches

#### 5.1.1 Genre-based

The genre-based music recommendation approach implements a basic personalization technique: the music tracks are recommended based on the users' genre preferences. We aimed to compare a personalized music recommendation approach with the *knowledge-based* and *tag-based* approaches, which are not personalized, but rather directly match music with the users' context (i.e., POIs). We did not implement a more sophisticated personalization approach since the users were requested to evaluate if the proposed music tracks suit the presented POI and therefore it was not necessary and even deleterious to totally match the detailed user's music preferences in this particular contextual situation.

In order to obtain the users' preferences, we asked the study participants to select their preferred music genres prior to performing the evaluation. The genre taxonomy was based on the music tracks in our dataset, and included: Classical, Medieval, Opera, Folk, Electronic, Hip Hop, Jazz, Pop, and Rock. For each displayed POI, the *genre-based* track is randomly selected from the whole set of music tracks belonging to the user's preferred genres.

#### 5.1.2 Knowledge-based

The knowledge-based music recommendation approach employs the technique presented in [11]. Given a POI, this approach ranks musicians by their relatedness to the POI. The relatedness is computed from the semantic relations between the POI and musicians extracted from the DBpedia knowledge base<sup>2</sup>.

To extract relations between POIs and musicians, we considered only a subspace of DBpedia by identifying classes and relations belonging to the two domains of interest places and music. We identified three types of relations that are useful for matching POIs with musicians: location relations (e.g., POI located in City, City birthplace of Musician), time relations (e.g., POI construction Date, Musician active Date), and relations through historical/cultural eras (e.g., POI has category Architecture Style, Musician has category Music Genre). Subsequently, for each POI, we built a network consisting of a directed acyclic graph whose nodes represent the identified classes, and edges represent the selected relations. We assigned relevance weights to nodes and edges of the graph, and performed a weight spreading algorithm over the graph to compute relatedness scores for musician nodes in the graph.

As explained in Section 3, we have downloaded three representative music tracks for each musician in the dataset. Using this approach, we implicitly assume that there are no major differences between these tracks of the same musician. Therefore, for each POI, the *knowledge-based* track is randomly selected from the three music tracks by the topranked musician.

#### 5.1.3 Tag-based

The *tag-based* music recommendation approach uses the technique presented in [6] and is applied to the previously described set of 123 manually tagged tracks. This approach computes the weighted Jaccard similarity between a POI u and a music track v as:

$$\operatorname{JaccardSim}(u,v) = \frac{\sum_{t \in X_u \cap X_v} \log f(t)}{\sum_{t \in X_u \cup X_v} \log f(t)}$$
(1)

where  $X_u$  and  $X_v$  are the items' tag profiles and f(t) is the fraction of items in our dataset (both POIs and music tracks) annotated with the tag t. For each POI, the *tagbased* approach recommends the top-ranked music track.

#### 5.1.4 Auto-tag-based

The *auto-tag-based* approach is applied to the full dataset of 369 music tracks which are auto-tagged using the subset of 123 manually tagged tracks for training (see Section 4).

Given the tag vocabulary of size K (in our case K = 24), for each music track v the auto-tagger produces a tag affinity vector:

$$a_v = (p(t_1), ..., p(t_K))$$

```
<sup>2</sup>http://dbpedia.org/
```

where  $p(t_i)$  denotes the probability for a tag  $t_i$  to be relevant for the track. Then, the track's tag profile  $X_v$  is defined as:

$$X_v = \{t_i \mid p(t_i) \ge 0.4\}, \ i = \{1, \dots, K\}.$$

The similarity between a POI u and a music track v is computed using Equation 1. For each POI, the *auto-tagbased* approach recommends the top-ranked music track.

#### 5.1.5 Combined

This music recommendation approach is a hybrid combination of the *knowledge-based* and *auto-tag-based* techniques, employing a rank aggregation technique [10]. Since the music-to-POI similarities produced by the two techniques have different value ranges, we used the normalized Borda count rank aggregation method to give equal importance to the two. This method is also applied in other recommender systems [4] and works as follows: Given a POI u, the *knowledge-based* and *auto-tag-based* approaches produce the rankings of music tracks  $\sigma_u^{kb}$  and  $\sigma_u^{tb}$ . We denote the position of a track v in these rankings as  $\sigma_u^{kb}(v)$  and  $\sigma_u^{tb}(v)$ respectively. Then we compute the combined score of the track v for the POI u as:

CombinedScore
$$(u, v) = \frac{N_{kb} - \sigma_u^{kb}(v) + 1}{N_{kb}} + \frac{N_{tb} - \sigma_u^{tb}(v) + 1}{N_{tb}}$$

where  $N_{kb}$  and  $N_{tb}$  are the total number of tracks in the corresponding rankings. Finally, for each POI the *combined* approach recommends the top-scored music track.

### 5.2 Web-based Evaluation

To determine which approach produces better music recommendations we designed a web-based interface (Figure 4) for collecting the users' subjective assessments of whether a music track suits a POI. The participants of the experiment were repeatedly asked to consider a POI, and while looking at its images and description, to listen to the recommended music tracks. They were asked to check all the tracks that in their opinion suit that POI. During each evaluation step, the music recommendations for a POI were selected using the proposed five approaches described above - the personalized baseline approach and the four contextaware approaches. The order of the recommendations was randomized, and the user was not aware of the algorithms that were used to generate the recommendations. In total, a maximum of five tracks corresponding to the top-ranked tracks given by each approach were suggested for each POI, but sometimes less tracks were shown as the tracks selected by the different approaches may overlap.

#### 5.3 Results

A total of 58 users participated in the evaluation study, performing 564 evaluation sessions: viewing a POI, listening to the suggested music tracks, and providing feedback. 764 music tracks were selected by the users as well-suited for POIs. Figure 5 shows the performance of the recommendation approaches, computed as the ratio of the number of times a track produced by each approach was considered as well-suited over the total number of evaluation sessions. All context-aware approaches perform significantly better than the personalized *genre-based* track selection (p < 0.001 in a two-proportion z-test). This result shows that in a situation

defined by a context (i.e., a POI), it is not sufficient to recommend a music track liked by the user, but it is important to adapt music recommendations to the user's context.

Among the four context-aware approaches, the *combined* approach produced the best results, outperforming the others with statistical significance at p < 0.01. These results confirm our hypothesis that the users are more satisfied with the recommendations when combining the *tag-based* and *knowledge-based* techniques, which represent orthogonal types of relations between a place and a music track.



Figure 5: Selection probabilities of the different music recommendation approaches.

While our research hypothesis is supported by the users? feedback averaged across the full set of POIs, it is also interesting to look at the users' judgments of music appropriateness for the individual POIs. Figure 6 shows two interesting cases. For the POI Casa Batlló, the knowledge-based approach outperforms both *tag-based* approaches, while for the Royal Palace of Brussels, the *tag-based* approaches perform better. These results can be explained by analyzing the tag profiles of the items. We observe the differences of tag distributions in the profiles of the two POIs. Casa Batlló is tagged as [Modern, Spiritual, Energetic, Warm, Colorful, Animated, Lightweight], with over half of the tags in the profile being very frequent in our dataset (among both music tracks and POIs). Instead, the Royal Palace is tagged as [Calm, Heavy, Strong, Ancient, Open], with most of the tags being quite rare in the dataset. Note that the Jaccard similarity metric, used in the tag-based approach, weights the tags with respect to their frequency in the dataset (by assigning more weight to rare tags, see Equation 1). Therefore, the approach is better at identifying a good match for a POI tagged with a smaller set of tags unique in the set of music tracks compared to a POI tagged with a large number of common tags.

Furthermore, the different performance of the *knowledge-based* approach for the two POIs suggests that for certain POIs the technique fails to provide a relevant recommendation. In fact, we observe that for the Royal Palace of Brussels, the approach recommends a modern music piece by a Brussels-born composer, however, not fitting the nature of this particular POI. Notwithstanding the shortcomings of each individual technique discussed above, the obtained results suggest that the proposed hybrid *combined* approach can effectively balance the effects of the two alternative techniques and therefore produce better recommendations.



Figure 4: Screenshot of the web application used to evaluate the different music recommendation approaches.



Figure 6: Selection probabilities of the different music recommendation approaches for two individual POIs.

Finally, the evaluation results suggest that the *tag-based* music recommendation approach can be successfully scaled up using automatic tag prediction techniques. The *auto-tag-based* approach even outperformed the *tag-based* approach with marginal significance (p = 0.078). This can be explained by the larger variety of music in the auto-tagged music dataset – using the *auto-tag-based* approach the recommendations were selected from the full set of 369 music tracks, while the *tag-based* approach used only the subset of 123 manually annotated tracks. Scaling up the process of tag generation without harming performance is hence the vital advantage of using the auto-tagger.

# 6. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a novel hybrid approach to recommend music related to POIs. This approach is based on combining two music recommendation techniques: a *tag-based* one and a *knowledge-based* one [6, 11].

Since the *tag-based* technique was previously assessed on a manually tagged dataset of limited size [6], in this work we have shown that it can scale up employing a music autotagging technique. In fact, in a user study we have demonstrated that music auto-tagging can be successfully applied in tag-based music recommendation for POIs. Subsequently, we have presented a user study where the proposed approach was evaluated on a dataset of POIs and music tracks and was compared with a personalized recommendation approach that considers users' music genre preferences and does not adapt music to the POIs. The different recommendation techniques were evaluated in a web-based user study where the users were required to evaluate the appropriateness of the music selected by the system for a set of 25 POIs. The results of the evaluation study confirm our hypothesis that the combination of the tag-based and knowledge-based techniques produces better recommendations. We believe that these findings provide a solid base for the implementation of real-life location-aware music recommenders.

We stress that the ultimate goal of this research is a location-aware music delivery service that combines the *tagbased* and the *knowledge-based* music recommendation techniques. While in this work we demonstrate the effectiveness of such combination through a web-based evaluation, it is important to evaluate the approach in real-life settings and with a larger music/POI dataset to confirm our findings. Moreover, additional evaluation would help us understand which type of associations between music and POIs – emotion-based or knowledge-based – the users prefer in different recommendation scenarios (e.g., sightseeing, choosing a holiday destination, or acquiring knowledge about a destination).

Other important future work directions include automatically acquiring tags describing POIs (using web mining techniques and folksonomy datasets like Flickr as sources of tagging information); analyzing the importance of individual audio features when automatically tagging music tracks; using more complex hybrid recommendation strategies to combine our music recommendations with, e.g., those produced by Last.fm. Finally, we believe that the designed contextaware recommendation solutions may be adapted to other types of content, for instance, recommending music content that fits movies, books, or paintings.

## 7. ACKNOWLEDGMENTS

This research is supported by the Austrian Science Fund (FWF): P22856 and P25655; and by the EU FP7/2007-2013 through the PHENICX project under grant agreement no. 601166. The authors would further like to thank Klaus Seyerlehner for providing source code and expertise on music auto-tagging.

# 8. REFERENCES

- G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin. Context-aware recommender systems. AI Magazine, 32(3):67–80, 2011.
- [2] A. Ankolekar and T. Sandholm. Foxtrot: a soundtrack for where you are. In Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications, pages 26–31, 2011.
- [3] J.-J. Aucouturier and F. Pachet. Improving Timbre Similarity: How High is the Sky? Journal of Negative Results in Speech and Audio Sciences, 1(1):1–13, 2004.
- [4] L. Baltrunas, T. Makcinskas, and F. Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the 4th ACM Conference on Recommender Systems (RecSys)*, pages 119–126, 2010.
- [5] C. Bizer, T. Heath, and T. Berners-Lee. Linked data the story so far. International Journal on Semantic Web and Information Systems, 5(3):1-22, 2009.
- [6] M. Braunhofer, M. Kaminskas, and F. Ricci. Location-aware music recommendation. International Journal of Multimedia Information Retrieval, 2(1):31–44, 2013.
- [7] Ò. Celma. Music Recommendation and Discovery in the Long Tail. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2008.
- [8] O. Celma and P. Lamere. If you like radiohead, you might like this article. AI Magazine, 32(3):57–66, 2011.
- [9] E. Coviello, A. B. Chan, and G. Lanckriet. Time Series Models for Semantic Music Annotation. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(5):1343–1359, 2011.
- [10] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings* of the 10th international conference on World Wide Web (WWW), pages 613–622, 2001.
- [11] I. Fernández-Tobías, I. Cantador, M. Kaminskas, and F. Ricci. Knowledge-based music retrieval for places of interest. In Proceedings of the 2nd International Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies (MIRUM), pages 19–24, 2012.
- [12] M. Kaminskas and F. Ricci. Contextual Music Information Retrieval and Recommendation: State of

the Art and Challenges. *Computer Science Review*, 6:89–119, 2012.

- [13] J. H. Kim, B. Tomasik, and D. Turnbull. Using Artist Similarity to Propagate Semantic Information. In Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR), pages 375–380, 2009.
- [14] E. Law, L. Von Ahn, R. Dannenberg, and M. Crawford. Tagatune: A game for music and sound annotation. In *Proceedings of the 8th International Society for Music Information Retrieval Conference* (ISMIR), pages 361–364, 2007.
- [15] M. I. Mandel and D. P. W. Ellis. A Web-Based Game for Collecting Music Metadata. *Journal of New Music Research*, 37(2):151–165, 2008.
- [16] M. I. Mandel, R. Pascanu, D. Eck, Y. Bengio, L. M. Aiello, R. Schifanella, and F. Menczer. Contextual Tag Inference. ACM Transactions on Multimedia Computing, Communications and Applications, 7S(1):1–18, 2011.
- [17] R. Miotto, L. Barrington, and G. Lanckriet. Improving Auto-tagging by Modeling Semantic Co-occurrences. In Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), 2010.
- [18] S. Reddy and J. Mascia. Lifetrak: music in tune with your life. In Proceedings of the 1st ACM International Workshop on Human-centered Multimedia (HCM), pages 25–34, 2006.
- [19] K. Seyerlehner, M. Schedl, P. Knees, and R. Sonnleitner. A Refined Block-Level Feature Set for Classification, Similarity and Tag Prediction. *Extended Abstract to the Music Information Retrieval Evaluation eXchange (MIREX)*, 2011.
- [20] K. Seyerlehner, M. Schedl, T. Pohle, and P. Knees. Using block-level features for genre classification, tag classification and music similarity estimation. *Extended Abstract to the Music Information Retrieval Evaluation eXchange (MIREX)*, 2010.
- [21] K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees. Automatic Music Tag Classification based on Block-Level Features. In *Proceedings of the 7th Sound* and Music Computing Conference (SMC), pages 126–133, 2010.
- [22] P. Smolensky. Information Processing in Dynamical Systems: Foundations of Harmony Theory. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, pages 194–281, 1986.
- [23] M. Sordo. Semantic Annotation of Music Collections: A Computational Approach. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2012.
- [24] D. Turnbull, L. Barrington, and G. Lanckriet. Five approaches to collecting tags for music. In *Proceedings* of the 9th International Society for Music Information Retrieval Conference (ISMIR), pages 225–230, 2008.
- [25] G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on* Speech and Audio Processing, 10(5):293–302, 2002.
- [26] M. Zentner, D. Grandjean, and K. R. Scherer. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4):494–521, 2008.