People, places and playlists: modeling soundscapes in a mobile context

Nima Zandi, Rasmus Handler, Jakob Eg Larsen, and Michael Kai Petersen

Technical University of Denmark,
DTU Informatics, Cognitive Systems Section
Richard Petersens Plads, Building 321,
DK-2800 Kgs. Lyngby, Denmark,
{jel|mkp}@imm.dtu.dk

Abstract. In this paper we present an initial study of music listening patterns on mobile devices combined with contextual information. The study included N=7 participants that carried a smart phone for a duration of two weeks. The participants used the main features of the phone along the music player capabilities. All phone activities and data from embedded sensors were recorded along the music being played on the device. We report initial indications that listening patterns in terms of music genre preferences are influenced by whether the user is in a static environment or on the move. Applying a simple decision tree algorithm to identify what contexts determine the preferences indicate that our listening patterns change over time, suggesting that music applications utilizing context information must be designed to adapt to our shifting preferences as they continuously evolve.

Keywords: Mobile, context-awareness, music, genre, listening pattern

1 Introduction

The aspects of context-awareness and context-aware applications have gotten much attention for more than a decade. In this paper our focus is on the consumption of music content in mobile scenarios involving multiple contexts. It is our assumption that the way people access and use content is highly dependant on the particular context. The context include the time, place, tasks, motivation, and the history of the interaction.

When people carry their mobile phones throughout the day it provides a unique opportunity to capture contextual information from the wide range of embedded sensors. Together these sensors provide an interesting source of information about activities, people, places and other entities. The advancements of mobile phones has increased the potential for novel context-aware mobile applications. Present off-the-shelf smart phones have several embedded sensors, such as, GPS, accelerometer, light sensor, proximity sensor, microphone, camera, as well as multiple network connectivity options, such as, GSM, WLAN, and Bluetooth. As such, the mobile phone can potentially serve as a proxy in terms of providing information about the context of the human user [1].

In a recent study Song et al. [2] discuss the predictability of human behavioral patterns based on detected movement patterns using GSM cellular information for location approximation. The interesting result is a predictability as high as 93%, which provides an indication of the potential for future context aware mobile applications. Mobile applications involving music has also gained interest recently. Especially novel ways of navigating music collections is available in the commercial space, examples include Moodagent for the iPhone and Playlist DJ for Symbian, which allows the user to navigate a music collection in terms of mood, rhythm and style of the music. We hypothesize that contextual information obtained from a mobile device can offer useful information in terms of providing additional input for music recommendation for the individual and in a social context.

2 Context

In order to capture contextual information on mobile phones in our study, we utilize an existing solution – Mobile Context Toolbox – available for the Symbian S60 mobile operating system [1]. The system is built in multiple layers (as depicted in Fig. 1) on top of the Nokia S60 platform using Python for S60 (PyS60) with a set of extensions for accessing low-level sensors and application data.

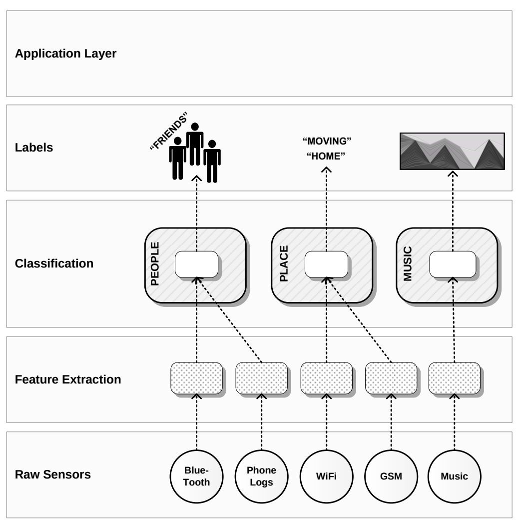


Fig. 1. Mobile Context Toolbox architecture

The underlying toolbox provides interfaces for accessing multiple low-level sensors, encapsulated in higher-level adapters. The inspiration for the layered approach is the framework described by Salber et al. [3] where the emphasis is on a clean cut between system resources, inferring contextual information and applications using it. Thus the details of the invidual sensors are abstracted to infer higher-level contextual information and focus on the feature extraction and classification of the obtained sensor data. Taking this approach simplifies the process of aggregating sensor data into higher-level contexts, as-well as making the framework extensible and adaptable to changes in the underlying platform.

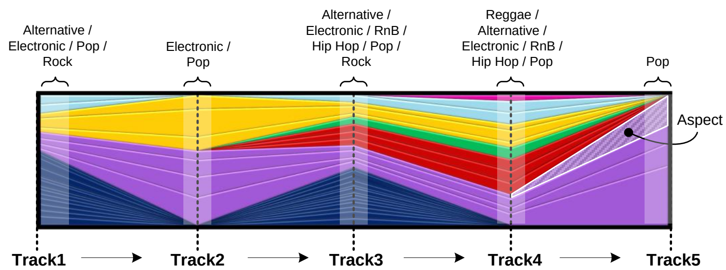
In the present study we have focused on obtaining contextual information about people, places and music. As shown in Fig. 1, information is obtained from the Bluetooth sensor and phone log in order to extract features related to people. Wi-Fi and GSM cellular information is used to extract features about the present place. Finally, the existing Mobile Context Toolbox has been extended with a virtual sensor obtaining information from the embedded music player application on the phone. Thus it acquires information about tracks being played, that is, extract music features including song title and artist information obtained from the embedded track metadata. Each of these features are translated into meaningful labels, and finally the application layer can utilize the contextual information inferred from the system by means of these contextual labels.

3 Music

The track information from the music player is incorporated as yet another sensor. The aim is to model not only the people and places as our mobile context, but also the constantly changing frame of mind reflected in the music we listen to. Although the artist and genre descriptions associated with the tracks as ID3 tags might be limited to terms such as 'pop', we are able to extend this based on available data from music social networks like *last.fm*. Tag-clouds generated by hundred thousands of users, might be interpreted as high dimensional representations of artist information, musical genres as well as the perceived emotional context of songs. Studies have shown that people often tend to agree not only on what emotional tags to use but also what tracks they are applied to [4]. The genres which form tag-clouds are far from crisp categories, but have large overlaps between what might vaguely be termed 'rock' versus 'pop', or could be labeled 'indie' in contrast to 'alternative'. When applying probabilistic latent semantic analysis to extract the underlying topics behind the most frequently co-occurring words, aspects like 'pop' will appear multiple times in different contexts coupled with tags such as 'soft, rock' or 'love, romantic' while emotions such as 'chillout' would be fused together with other tags like 'electronica, ambient, downtempo' [5]. This allows us to build track and session signatures based on the ID3 tags that are enhanced to capture the underlying semantic aspects of the music. Initially defining track signatures for the songs that are being played, each $Track_i$, $i = 1, \cdots, m$, consists of aspects $\{A_1, ..., A_{91}\}$ where A_j is based on tags $\{T_{j1}, ..., T_{jn}\}$. For every j'th aspect, A_j, in $Track_i$, the ratio of tags is calculated and combined to obtain a track signature for the i'th track. Subsequently we collect tracks into sessions if the user has been listening to at least 3 songs in a row, by combining the above track information into a session signature

$$Session_l^{Signature} = \{\frac{1}{M_l}\sum_{i=1}^{M_l} AspectHitRatio_{1i}, ..., \frac{1}{M_l}\sum_{i=1}^{M_l} AspectHitRatio_{Ni}\}$$

constituting an average of the ratios of tag co-occurrences defined in the track signature, where $M_l = \#\{\text{tracks in session nr. l}\}$. Generalizing the tag co-occurences into broader categories of musical style, allow us to define the changing genre characteristics over time, within the sequences of tracks that constitute a playlist (Fig.2). For visual clarity the different colors are assigned to the different genres used. Based on Pearson correlation we are able to define the similarities between different sessions, across multiple users within varying mobile contexts.



(a) Track sequence signature visualization

(b) Genre colors

Fig. 2. An example generalized track sequence signature (a) derived by enhancing the track ID3 metadata with social network tags, that capture the changing genre characteristics over time. Each music genre has a unique color (b) as shown in (a).

4 Experiment

Data was acquired from N=7 participants over a duration of two weeks. The participants were asked to use a Nokia N95 8GB smart phone with our Mobile Context Toolbox preinstalled along a collection of MP3 music files. Participants were asked to use the phone as their standard phone (with personal SIM card), and as their MP3 player for the duration of the experiment. Table 1 provides an overview of the contextual data obtained related to places and people for all 7 participants, where * marks the number of unique data points. In addition the total user interaction with the embedded music player application is included. 'Played' refers to the number of tracks that has started playing on the music player, and 'listened' includes the number of tracks where more than half of the track has been played, similar to the criteria used in *last.fm*.

P	BT	BT*	WiFi	WiFi*	GSM	GSM*	Ph	Ph*	Played	Listened	Unique
1	12620	1387	147333	2892	2043	242	343	35	337	160	85
2	10422	475	69973	4013	2792	344	305	35	474	153	100
3	7406	203	63495	777	402	105	92	20	375	190	48
4	3095	546	25375	1087	3516	296	286	28	524	292	68
5	5032	798	89917	2329	1573	175	164	35	173	110	58
6	15127	2675	75948	3915	1985	472	398	41	742	167	124
7	7654	1849	88222	2306	1132	169	113	17	198	94	65
Total	61356		560263		13443		1701		2823	1166	

Table 1. Overview of obtained contextual data for all 7 participants. BT is Bluetooth data, Ph is phone call/sms log, and Played, Listened and Unique refer to music tracks.

Contextual categorization of music playlists were generated over 2 weeks, where each sequence signature has been linked to labels such as 'at home', 'in transition' or 'weekend', inferred from low level location and motion data collected, as shown in Table 2. In essence capturing the contextual listening habits of a user combined with the genre preferences derived from the sequences of tracks that have been played. The upper and lower rows in the table indicate sessions from the first and second week respectively based on data generated by participant P_1.

Nr.	Transition	Home	O.K.P	U.Place	T.O.D	WeekDay	S.Category
1	True	False	False	False	Day	Work	C
2	True	True	False	False	Day	Work	A
3	True	False	False	False	Evening	Work	B
4	True	False	False	False	Day	Work	A
5	True	False	False	False	Day	Work	C
6	False	True	False	False	Evening	Work	A
7	True	False	False	False	Day	Work	D
8	True	False	False	False	Day	Work	E
9	False	False	True	False	Day	Work	A
10	True	True	False	False	Day	Work	A
11	False	True	False	False	Day	Weekend	B
12	True	False	False	False	Evening	Weekend	D
13	True	False	False	False	Day	Work	B
14	False	False	True	False	Day	Work	F
15	False	False	True	False	Day	Work	G
16	False	False	True	False	Day	Work	F
17	True	False	False	False	Day	Work	H
18	True	False	False	False	Day	Work	D
19	False	False	True	False	Day	Work	G
20	False	False	True	False	Day	Work	C

Table 2. Contextual categorization of music playlists generated over 2 weeks for participant P_1. O.K.P is other known place, such as work, U. place is unknown place, T.O.D is time of the day, and S.Category is session category.

5 Results and Discussion

Learning how music genre preferences are associated with labels such as 'at home', 'in transition' or 'other known place', inferred from low level location and motion data continuously collected by the Mobile Context Toolbox, makes it possible to contextually categorize music based on the constantly changing usage scenarios. Even when only considering a limited test based with seven participants forming a small scale network, patterns emerge that indicate how it might be possible to share music by traversing the social graph and find users in a similar context or listening to playlists resembling our own, as illustrated in Fig. 3.

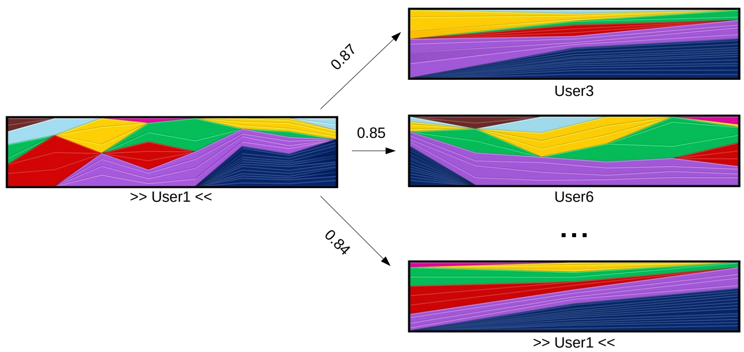
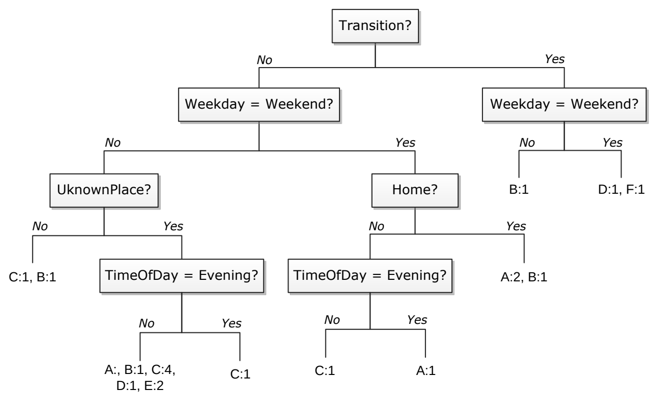
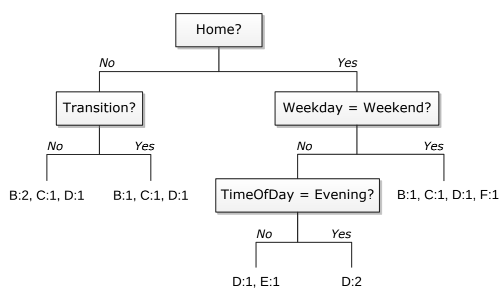


Fig. 3. Traversing the social graph similar music playlists are identified and ranked using Pearson correlation as distance measure based on participant P_1's session signature. The correlated sessions are retrieved from different levels in the social graph, e.g. not only friends have similar sessions but also friends of friends. Here the highest-ranked session is from a friend, whereas the following session is from a friend of a friend.

When applying a simple machine learning algorithm to untangle the user choices that define the contextual genre preferences, it becomes apparent that these structures are extremely dynamic. Not only do the contextual labels that influence what music we listen to differ from one user to another, but also change from one week to another when viewed from the perspective of a single user. This means training a classifier on contextual listening patterns over a short period and use it for future prediction is not likely to work, unless it is continuously adjusted to shifting preferences. Nevertheless clear tendencies in listening patterns appear to emerge for habits associated with contexts corresponding to static scenarios like 'at work' or 'gym' versus what genres are being exploring when on the move. The study by Song et al. [2] use entropy as a measure to assess the degree of predictability in patterns of movement based on extrapolated GSM cellular information for location approximation, which indicates that human behavior may be determinable up to 93%, of the time. However, when minimizing entropy in our approach, based on a decision tree algorithm that finds the contextual labels which as top nodes best define the genre preferences for each usage scenario, it seems that size matters when selecting a temporal frame. That is, a dynamic approach to prediction appears to be essential once we move beyond determining likely nodes of location and enter the unchartered territory of how our media preferences are influenced by the people and places constituting our constantly shifting mobile context.



(a)



(b)

Fig. 4. Decision trees classifying the user choices that determine the associations between contextual labels and playlist genre characteristics, trained based on data generated in the first (a) and second (b) week respectively for participant P_4. Splitting the data on a weekly basis highlights the dynamic character of the contextual listening patterns. The top nodes corresponding to the conditions that are most significant for defining the correlations between mobile contexts and genre preferences differ not only between users, but also change on an individual basis within the two week period.

Example decision trees for participant P_4 is provided in Fig. 4. In the decision tree the nodes are the spatiotemporal context labels generated by the Mobile Context Toolbox, whereas the leafs at the end of branches define the preferred genre categories corresponding to each usage scenario. The algorithm finds the variables that best divide the data, meaning that the nodes which are pushed towards the top represent the conditions which in terms of information gain contribute the most to explaining the underlying structure.

6 Conclusions

Although our study is based on a very limited number of users constituting a small scale social network, our initial findings suggest that it is possible to define contextual categories linking our music genre preferences to labels continuously inferred from low level location and motion data generated by the Mobile Context Toolbox running on the smart phones. The study has supported our expectations that listening patterns in terms of preferred music genres are influenced by conditions defining whether we are in a static environment or on the move. Applying a simple decision tree algorithm to identify what contextual labels determine music preferences, our results indicate that our listening patterns are continuously transformed over time. This indicates that even though we may observe distinct tendencies in habits related to the underlying context, future recommender systems must allow the application to adapt to our changing music listening patterns as they are influenced by context but also appear to continuously evolve over time.

Acknowledgments The authors would like to thank the 7 participants which took part in the experiments. Also thanks to Nokia Denmark and to Forum Nokia for the equipment used in the experiments.

References

1. Larsen, J., Jensen, K.: Mobile Context Toolbox-an extensible context framework for S60 mobile phones. In: Smart Sensing and Context: 4th European Conference, EuroSSC 2009, Guildford, UK, September 16-18, 2009. Proceedings, Springer-Verlag New York Inc (2009) 193
2. Song, C., Qu, Z., Blumm, N., Barabasi, A.: Limits of predictability in human mobility. Science **327**(5968) (2010) 1018
3. Salber, D., Dey, A.K., Abowd, G.D.: The context toolkit: aiding the development of context-enabled applications. In: CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (1999) 434–441
4. Hu, X., X, M.B., Downie, S.: Creating a simplified music mood classification ground-truth set. In: Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR). (2007) 309–310
5. Levy, M., Sandler, M.: Learning latent semantic models for music from social tags. Journal of New Music Reseach **37**(2) (2008) 137–150