Rushed or Relaxed? – How the Situation on the Road Influences the Driver's Preferences for Music Tracks

Linas Baltrunas
Telefonica Research,
Plaza de E. Lluchi Martin 5,
Barcelona, Spain
Linas@tid.es

Bernd Ludwig
University of Regensburg,
Universitätsstraße 31,
Regensburg, Germany
bernd.ludwig@ur.de

Francesco Ricci
Free University of Bolzano,
Piazza Domenicani 3,
Bolzano, Italy
fricci@unibz.it

ABSTRACT

In context-aware recommender systems, the dependency of the user's ratings on factors that describe important aspects of the recommendation context is used to provide more relevant recommendations.

Individual users may be influenced differently by the same set of contextual factors. By understanding this kind of dependency between the user's ratings (evaluations) and context, it is possible to identify user profiles and use them to predict precisely the user ratings for items to be recommended. In this paper, we present our methodology to identify user profiles in a corpus of ratings for music tracks. These ratings were collected in a user study, which simulated typical situations that occur while driving a car. We present the findings derived from the data, and argue that it is feasible to distinguish different typologies of users from the ratings they give to music tracks in specific contexts.

Categories and Subject Descriptors

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

Keywords

Recommender Systems, Context-based Reasoning, Collaborative Filtering

1. INTRODUCTION

Recommender systems predict user ratings for items on the basis of previous ratings for similar items or similar users [5]. As users may rate the same item differently depending on the situation in which they will experience or use the item, context-aware recommender systems [4, 6, 3, 1] have become a popular research focus. The main idea is to model context as a set of variables (contextual factors) each of which can take one of a finite set of discrete values (contextual value). The user ratings are stochastically dependent on the contextual values.

For a recommender system, there is a major implication from this observation. If we can assess such an influence for individual users we are able to better personalize recommendations. Beyond this, it may even be possible to group users influenced in a similar way by certain contextual conditions. This knowledge could lead to an improved prediction of ratings for items not previously rated by the user.

With this in mind, it seems worth understanding the influence of context on user ratings. In previous work [2], we reported on a collection of ratings data for music tracks while users experienced different stereotypical situations while driving a car. In this report, we focus on the analysis of this data with respect to the aims discussed above. Whether or not a particular aspect of context is important for predicting user ratings, is dependent on the user to whom the recommendations are targeted. Our data suggest that different users have different perceptions of their surroundings and that these perceptions may influence musical preferences. Our data reveal that people assign different ratings to the same music track in different contexts and in many cases these differences are statistically significant.

Our paper is structured as follows: In the next section we briefly present our data. Next, we introduce the mathematical tools we use to analyze the influence of context on user ratings. In sections to follow, we present evidence that context can provoke a change the music genres preferences of the user. In the final section, we discuss whether or not the influence of the context on ratings can even be observed for individual users, and conclude the paper with a discussion of the results and outline our plans for future work.

2. DATA CORPUS AND CONTEXT MODEL

As described in [2], we collected two independent data samples. In these experiments, driving situations were simulated with descriptions on a website. In the first experiment, we intended to capture the influence of context on the active and conscious decision of a user to listen a tracks of a certain genre if at the same time he was exposed to a certain contextual factor. For this purpose, users were asked to focus on one context factor at a time and rate the *influence* of this context factor on their *decision to listen* to a track of a randomly proposed genre on a three-level scale (`POSITIVE`, `NEGATIVE`, or `NONE`). In this way, the decision making process in this experiment was modeled as an active modification of the user's attitude towards a genre. Over a period of three weeks, we acquired 2436 ratings from 59 users (Users were recruited via email-lists and social networks). This study was considered a pilot, and in order to avoid the sparse data

Context Factor	$MI_Y(X, Y)$
`sleepiness`	0.169766732
`traffic conditions`	0.034971332
`weather`	0.027759496
`driving style`	0.025347564
`road type`	0.022788139
`natural phenomena`	0.015574021
`mood`	0.013993043
`landscape`	0.010431354

Figure 1: Mutual Information between Influence of Context on Ratings and Context Factors

problem a small number of tracks for each genre were proposed. 95 ratings were collected per contextual factor.

For our model of context, we relied on cognitive task analyses of car driving and considered three different kinds of a driver's perceptions and actions as potentially relevant:

Context Factor	**Possible Values**
driving style	relaxed driving, sport driving
road type	city, highway, serpentine
landscape	coast line, country side, mountains/hills, urban
sleepiness	awake, sleepy
traffic conditions	free road, many cars, traffic jam
mood	active, happy, lazy, sad
weather	cloudy, snowing, sunny, rainy
natural phenomena	day time, morning, night, afternoon

Situations where more than one passenger was present were beyond the scope of our research.

For the second sample, we collected tracks with ratings on a five star scale. The sample consists of 955 ratings ignoring any context factor and 2865 ratings taking one contextual condition into account. The ratings were given by 66 different users (including many who had participated in the first study). 69 to 167 ratings were collected per contextual factor depending on the assumed relevance for the experiment (see Figure 1 and the discussion in Sect. 3).

3. RELEVANCE OF CONTEXT FACTORS

When analyzing the dependency between contextual factors and ratings we could not make any modeling assumptions regarding the nature of the dependency. The same holds for inter-factor dependencies. Therefore, parametric models for the dependency such as linear regression are not appropriate. Instead, we had to find a non-parametric model. In information theory, the concept of mutual information of two random variables is known exactly for this purpose: it provides means to quantify the mutual dependence of two random variables.

In our case, we can apply mutual information to quantitatively assess the difference in the average ratings for music ignoring any influence of context compared to the average rating taking single contextual factors into account. More formally, we define a random variable X for the event that users assign one of the ratings 1, 2, 3, 4, or 5 to a genre (in the first sample) or to a track (in the second sample).

Secondly, we define another random variable Y for the event that one of the context factors holds in the current situation. Mutual information (MI) between X and Y is then defined as:

$$MI(X, Y) = \sum_{y \in Y} \sum_{x \in X} P(x, y) \cdot \log \frac{P(x, y)}{P(x) \cdot P(y)}$$

MI can be normalized to the interval $[-1; 1]$ by computing its value relative to the entropy of Y:

$$MI_Y(X, Y) = \frac{MI(X, Y)}{-\sum_{y \in Y} P(y) \cdot \log P(y)}$$

For X we have 2436 ratings (see Section 2 above). For each of the context factors, we collected 95 ratings. Figure 1 gives a numeric overview of the average ratings in the second data set and the impact of the single context factors on the average rating.

The results indicate that users are influenced heavily by variable driving conditions such as their own physical condition (sleepiness) and external factors such as traffic and weather. Personal factors, such as their mood, and factor not directly related to the car driving task, such as the landscape in which users are traveling, are of minor impact.

In the next step of our analysis, we wanted to understand whether the influence of context depends on the user preference for a music track. We hypothesized that if the user more strongly likes or dislike a track then his rating can be significantly influenced by contextual factors. In order to analyze this hypothesis we grouped the data into 5 partitions for each of the 5 possible ratings a user could assign to a track. I.e. the partition 1 ("the tracks disliked without considering context") contains all tracks rated with 1 (while different context factors were activated), and partition 5 ("the highly preferred tracks") contains the tracks rated with 5 in any context. Again, the influence of the context factors can be computed by measuring the mutual information and therefore the dependence between the random variable "a track is rated r without considering context" ($r \in \{1, 2, 3, 4, 5\}$) and the random variable "context factor c is active while a track is rated r". Figure 2 shows the results of this experiment. A first look at the numbers gives the impression that the mutual information is generally higher than in the experiment documented in Figure 1. To test this in a statistically sound way, we compared the mutual information values for each partition to those shown in Figure 1 using a t-test. The results are given in the last column. With the exception of partition 3 which groups the tracks that users did rate neutrally, for each partition the difference is statistically significant (the dot stands for $\alpha = 0.5$, $**$ for $\alpha = 0.01$, $***$ for $\alpha = 0.001$). These findings suggest that when users have strong positive or negative opinions for certain tracks, the conditions they experience while driving a car can influence more their ratings for these tracks.

We also analyzed the influence of context on the preferences for certain music genres. For this purpose, we analyzed the data coming from the first study (see above). We formalized the user responses (`POSITIVE`, `NEGATIVE`, or `NONE`) as a random variable I. Given this variable, the genre G and the activated context factor C given, we can estimate the probability distribution $P(I|G, C)$ from the first data set and compare it to the distribution $P(I|G)$ which does not take any context into account. For our purposes, it is again interesting to compute the mutual information for the above random variables $(C|G)$ and $(I|G)$. The following table presents the top-3 results for all combinations of genres and context factors:

Context Factor	Partition 1	2	3	4	5
`driving style`	0.145373959	0.048822968	0.18469473	0.035874718	0.028085475
`landscape`	0.039462852	0.025682432	0.05470132	0.042950347	0.038938108
`mood`	0.017266963	0.029724906	0.052830753	0.046422692	0.093026607
`natural phenomena`	0.022655695	0.053228548	0.084777547	0.024086852	0.082907254
`road type`	0.062203817	0.027293531	0.040344565	0.073388508	0.143056622
`sleepiness`	0.136737517	0.17566705	0.053153867	0.396715694	0.31060986
`traffic conditions`	0.036059416	0.121036344	0.124320839	0.032237073	0.139863842
`weather`	0.089973183	0.064745768	0.03265592	0.019943082	0.053972648
Level of Significance	.	$**$		.	$**$

Figure 2: Mutual Information between Influence of Context on Ratings (`POSITIVE`, `NEGATIVE`, or `NONE`) and Context Factors Given a Certain Rating *(key: '.': $\alpha = 0.5$. $*, *$: $\alpha = 0.01$)*

Blues	driving style	0.324193188
	road type	0.216609802
	sleepiness	0.144555483
Classics	driving style	0.77439747
	sleepiness	0.209061123
	weather	0.090901095
Country	sleepiness	0.469360938
	driving style	0.363527911
	weather	0.185619311
Disco	mood	0.177643232
	weather	0.17086365
	sleepiness	0.147782999
Hip Hop	traffic conditions	0.192705142
	mood	0.151120854
	sleepiness	0.105843345
Jazz	sleepiness	0.168519565
	road type	0.127974728
	weather	0.106333439
Metal	driving style	0.462220717
	weather	0.264904662
	sleepiness	0.196577939
Pop	sleepiness	0.418648658
	driving style	0.344360938
	road type	0.268688459
Reggae	sleepiness	0.549730059
	driving style	0.382254696
	traffic conditions	0.321430505
Rock	traffic conditions	0.238140493
	sleepiness	0.224814184
	driving style	0.132856064

From these results, we can learn two lessons. First, within a given genre, the mutual information is very high only for some factors. Evidently, these have a strong influence on the user ratings. This outcome was not obvious before the experiment as the user preferences could have been stronger than the influence of the driving situation. However, some of these factors influence the ratings for (almost) all genres. We may conclude that they are strongly related to the cognitive and emotional state of a driver and therefore constitute important features of recommending music in car.

Second, as the influence of context is evident, we may conclude that even users with strong preferences for certain tracks may change their opinion if they experience their driving situation intensively enough.

4. INDIVIDUAL USER TYPES

We now investigate the influence of context on individual users. We analyze the user ratings of the four users who gave most of the ratings in our second data collection phase (see above). We show that different contextual factors can influence different users in different ways. In the following tables, *Mean with context* (MCY) is the average rating of a user for all items rated under the assumption that the given contextual factor holds. *Mean without context* (MCN) is the average (of all users) rating for the same items without considering context. Differences in these averages are compared using a t-test in order to assess whether a contextual factor actually influences the user's ratings in a significant way. We indicate the statistical significance of the difference between MCY and MCN with the p-value of the t-test.

We note that a recommender system can exploit the results of our data analysis when building a prediction model that integrates the average rating of many users for an item, a personalized component for a particular user, and a component for the context (see [2] for details).

User 1: Preferences above Average.

As can be seen in column MCN in Table 3b, this user, on average, rated the tracks in the data base higher than the others. The comparison with MCN of all users (see Table 3a) suggests that for this user many of the tracks were perceived very positively in driving situations demanding the driver's attention. In fact, driving on a highway, on a serpentine or mountain road leads to an increase of the average rating (compared to MCN for all users). On the other hand, situations that can be perceived as negative (e.g. traffic jam) provoke a decrease of the user ratings. This observation similarly holds for some other factors: *lots of cars*, a situation quite similar to *traffic jam*, or driving in *morning* time. Interestingly, *sport driving* – which stands for a consciously sportive style of driving – has negative influence on the average ratings of this user. Hence we hypothesize that the user is affected negatively by the tracks (mainly pop music) in situations that are likely to produce stress.

User 2: Preferences around Average with Positive Tendency towards Tracks.

In this example the user has a personal average rating similar to the other users. This phenomenon is not an ef-

Factor	MCN	MCY	Tendency	α
highway	2.498429	3.521739	↑	$***$
traffic jam	2.498429	1.647059	↓	$*,*$
city	2.498429	3.800000	↑	$**$
serpentine	2.498429	3.529412	↑	$**$
sport driving	2.498429	1.705882	↓	$**$
lots of cars	2.498429	1.894737	↓	$**$
coast line	2.498429	3.500000	↑	$*$
mountains/hills	2.498429	3.307692	↑	.
active	2.498429	1.866667	↓	.
country side	2.498429	3.272727	↑	.

(a) MCN of all Users versus MCY for User 1

Factor	MCN	MCY	Tendency	α
traffic jam	3.077586	1.647059	↓	$***$
lots of cars	3.077586	1.894737	↓	$***$
sport driving	3.077586	1.705882	↓	$***$
active	3.077586	1.866667	↓	$**$
morning	3.077586	2.000000	↓	$**$
city	3.077586	3.800000	↑	$*$

(b) MCN versus MCY of User 1

Figure 3: Profile of User 1. Only those factors with statistical significance are shown.

Factor	MCN	MCY	Tendency	α
happy	2.498429	1.444444	↓	$**$
serpentine	2.498429	1.709677	↓	$**$
urban	2.498429	1.760000	↓	$*$
awake	2.498429	3.642857	↑	$*$
country side	2.498429	1.807692	↓	$*$
sad	2.498429	1.846154	↓	$*$
afternoon	2.498429	2.000000	↓	.
relaxed driving	2.498429	2.025641	↓	.

(a) MCN of all Users versus MCY of User 2

Factor	MCN	MCY	Tendency	α
happy	2.432692	1.444444	↓	$**$
serpentine	2.432692	1.709677	↓	$*$
awake	2.432692	3.642857	↑	$*$
urban	2.432692	1.760000	↓	$*$
country side	2.432692	1.807692	↓	.
sad	2.432692	1.846154	↓	.

(b) MCN versus MCY of User 2

Figure 4: Profile of User 2. Only those factors with statistical significance are shown.

fect of any context. The sign of the significant differences between MCN and MCY in Table 4a indicate that this user likes the tracks in the corpus when he feels *awake*. Being *sad*, he would never like to listen to the tracks. In general, for this user the traffic situation (differently from user 1) seems to play a minor role. Many significant differences in his ratings can be found comparing his MCY with his non-contextualized ratings (own MCN) as well as with the rating of all the users (MCN), for personal factors such as the mood and the perception of the surrounding landscape.

User 3: Preferences slightly below or on Average with Negative Tendency towards the Tracks.

In this user profile, the factors provoking significant differences between MCN and MCY (see Table 5a) are mostly personal ones or factors that indirectly influence personal attitudes or the cognitive load of the driver (i.e. road type).

As many of the tracks used for our data collection were pop songs, and on average the user assigns low ratings, we can conclude that he has a strong dislike for this kind of music. This impression is strengthened by the observation that negative emotions (such as sad) lead to even worse ratings for tracks than on average for this user.

User 4: Preferences below Average.

In this user profile, there are several highly significant differences between the MCN of all users and MCY (see Table 6a). In every case, the tendency is negative indicating that there are almost no situations in which tracks from the data set should be recommended to such a user. Probably this user does not like the tracks in the corpus, or he even does not like to listen to music at all while driving. The significance level of the difference between the personal MCN and MCY (see Table 6b), here is slightly smaller than in the previous comparison. Moreover, there is one personal factor (awake) under which the user rated significantly higher. But, as there are many factors with almost identical ratings to the already low non-contextualized ratings, in most situations the items should not be recommended to this user. From this observation, we can assume that as this user dislikes tracks very strongly, it is hard to find context factors that may change his attitude.

5. CONCLUSIONS AND FUTURE WORK

We have presented a non-parametric approach to assess the impact of a set of contextual factors on the user ratings. Our findings from the analysis of two data collections suggest that the perceptions and experiences during the execution of a task influence user preferences even for non-crucial items such as music tracks to be played in a car.

5.1 Influence of Context

We found empirical evidence that the driving situation indeed influences the driver's preferences for music. The influence of context may even be strong enough to modify the preference of a user for his favorite tracks.

The findings also suggest that the cognitive load of the driver, his emotional, mental, and physical state, and current traffic conditions influence his preferences.

These findings are surely affected by the set of tracks used in the study. We used this set as the reported experiments were developed within an industrial project, and the tracks were provided by the media platform of the industrial partner. It is an interesting task to collect data for other set of tracks – in a wider set of types of tracks or with a different specialization – and repeat the analysis.

Factor	MCN	MCY	Tendency	α
sad	2.498429	1.333333	$\downarrow$	$**$
day time	2.498429	1.666667	$\downarrow$	$**$
active	2.498429	1.769231	$\downarrow$	$*$
serpentine	2.498429	1.714286	$\downarrow$	$*$
coast line	2.498429	2.000000	$\downarrow$	.

(a) MCN of all Users versus MCY of User 3

Factor	MCN	MCY	Tendency	α
sad	2.329787	1.333333	$\downarrow$	$**$
day time	2.329787	1.666667	$\downarrow$	$*$
active	2.329787	1.769231	$\downarrow$	.

(b) MCN versus MCY of User 3

Figure 5: Profile of User 3. Only those factors with statistical significance are shown.

Factor	MCN	MCY	Tendency	α
day time	2.498429	1.166667	$\downarrow$	$***$
afternoon	2.498429	1.666667	$\downarrow$	$**$
highway	2.498429	1.700000	$\downarrow$	$*$
urban	2.498429	1.769231	$\downarrow$	$*$
morning	2.498429	1.714286	$\downarrow$	.
mountains/hills	2.498429	1.714286	$\downarrow$	.
country side	2.498429	1.700000	$\downarrow$	.

(a) MCN of all Users versus MCY of User 4

Factor	MCN	MCY	Tendency	α
day time	2.175676	1.166667	$\downarrow$	$***$
awake	2.175676	3.222222	$\uparrow$	.
afternoon	2.175676	1.666667	$\downarrow$	.

(b) MCN versus MCY of User 4

Figure 6: Profile of User 4. Only those factors with statistical significance are shown.

5.2 Critical Discussion of the Study Design

It is important to note the constraints and conditions of our study design. First of all, in the web survey, we created fictive situations that the subject should imagine. Hence, the test persons may have overestimated the relevance of the contextual factors on their music preferences. Hence, a different study where users are actually facing certain contextual conditions is in order. But before performing that evaluation, our study clearly indicates that users perceive context as important and influential, and different users, with different music preferences, have completely different perceptions. To assess this result quantitatively, the web survey and the described methods represent a simple way to collect and analyze data. In fact, we exploited our results in the implementation of a real music recommender system and player [2]. Besides, it is also important to note that during our study users rated the music tracks just after listening to them. This is not always the case in many recommender systems (e.g. MovieLens or Netflix), where often the ratings are provided long after the user experienced the items.

5.3 Consequences for Future Work

Currently, we are preparing a new study with an improved experimental setup: we are merging our prototype with another application that allows to log onboard data in a car. We will equip cars of test persons with this tool and collect data in real driving situations. The logged data will allow us to detect the values of certain contextual factors from onboard information about the car and its navigation system. Furthermore, we will be able to combine this data with feedback from the users (e.g., which of the recommended tracks are played or skipped). From such a new collection of data, gained in a naturalistic setting, we will validate the findings of our simulation study.

6. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 217 – 250. Springer, 2011.

[2] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, A. Aydin, K.-H. Luke, , and R. Schwaiger. Incarmusic: Context-aware music recommendations in a car. In *(to appear) Proceedings of the 12th International Conference on Electronic Commerce and Web Technologies*, 2011.

[3] L. Baltrunas, M. Kaminskas, F. Ricci, L. Rokach, B. Shapira, and K.-H. Luke. Best usage context prediction for music tracks. In *2nd Workshop on Context-Aware Recommender Systems*, 2010.

[4] A. Chen. Context-aware collaborative filtering system: Predicting the user's preference in the ubiquitous computing environment. In T. Strang and C. Linnhoff-Popien, editors, *Location- and Context-Awareness*, volume 3479 of *Lecture Notes in Computer Science*, pages 244–253. Springer Berlin / Heidelberg, 2005.

[5] Y. Koren and R. Bell. Advances in collaborative filtering. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*. Springer, 2011.

[6] G.-E. Yap, A.-H. Tan, and H.-H. Pang. Discovering causal dependencies in mobile context-aware recommenders. In *MDM 06: Proceedings of the 7th International Conference on Mobile Data Management*, page 4, Washington, DC, USA, 2006. IEEE Computer Society.