Walk the Talk

Analyzing the relation between implicit and explicit feedback for preference elicitation

Denis Parra and Xavier Amatriain

[1] University of Pittsburgh
[2] Telefonica Research

Abstract. Most of the approaches for understanding user preferences or taste are based on having explicit feedback from users. However, in many real-life situations we need to rely on implicit feedback such as the amount of times a user has bought an item or listened to a song.
To analyze the relation between implicit and explicit feedback, we conduct a user experiment in the music domain. We find that there is a strong relation between implicit feedback and ratings. We analyze the effect of context variables on the ratings and find that *recentness* – *i.e.* time elapsed since the user interacted with the item – has a significant effect. We also analyze several user variables and conclude that the way that the user interacts with the items also affects the ratings. Finally, we propose a simple linear model that relates these variables to the rating we can expect to an item. Such mapping would allow to easily adapt any existing approach that uses explicit feedback to the implicit case.

1 Introduction

The rise of recent web applications such as online social networks and e-commerce has uncovered the unforeseen potential of user mining and modeling. Applications such as Recommender Systems [1] rely on understanding user preferences in order to tailor the response and produce a personalized output. User preferences are modeled by taking into account either *explicit* or *implicit* user feedback.

Implicit feedback [2] is obtained by measuring the interaction of the user with the different items: if a given user buys a lot of strawberries or listens to many tracks by John Lennon, we may assume that she likes both strawberries and music by John Lennon. Implicit feedback is obtained without incurring into any overhead on the user, since it is obtained from direct usage [3]. However, it is not clear that we can trust a simple one-to-one mapping between usage and preference [4].

On the other hand, explicit feedback is obtained by directly querying the user. Users are usually presented with an integer scale where they quantify how much they like the items. In principle, explicit feedback is a more robust way to extract preference, since the user is reporting directly on this variable, removing the need of an indirect inference. However, it is also known that this kind of feedback is affected by user inconsistencies known as *natural noise* [5]. Users might also be pressed to report different preferences because of peer or social pressure. Besides, the fact that we are introducing a user overhead, makes it difficult to have a complete view on the user preferences [6].

Therefore, none of the two existing strategies clearly outperforms the other. Ideally, we would like to use implicit feedback, minimizing the impact on the user, but having a robust and proven way to map this data to the actual user preference. Our target scenario is one in which by sampling a few ratings given by some of the users we can design an appropriate mapping. This would allows us to then use implicit feedback with any method proved valid for explicit ratings.

2 Preliminaries and Related Work

Although implicit feedback is much more readily available in practical applications, most of the research literature focuses on the use of explicit feedback input. The main reason is that this explicit feedback is considered the ground truth on the user preferences and the recommender problem is then assimilated into a predictive model. The current work is motivated by some of our previous work in doing contextual recommendations based on implicit feedback [7]. In that case, we modeled implicit data following the approach by Celma [8] in which playcounts are directly binned into ratings. However, we found results to be unsatisfactory and uncovered the need for more work in this area.

In one of the few papers addressing the implicit feedback recommendation problem [9], Hu *et al.* list their observations regarding implicit feedback: (1) **There is no negative feedback**. In explicit feedback, users may rate items they like or they don't. In implicit feedback, we cannot assume zero feedback means the user did not like the item. (2) **Implicit feedback is noisy**. We would like to directly relate amount of implicit feedback to level of preference. But this might not always true. (3) **Preference vs. Confidence**. The numerical value of explicit feedback indicates preference while the numerical value of implicit feedback indicates confidence on whether the user likes the item. (4) **Evaluation of implicit feedback**. There is a lack of clear metrics for evaluating a recommender system using implicit feedback.

Our approach starts off from different hypothesis, some of which in fact contradict the previous. In particular: (1) While it is true that you cannot interpret no implicit feedback as negative feedback – and this is true also for explicit feedback–, implicit data can include negative feedback. As long as the granularity of the items is comparable, and there is enough variability, you should be able to assume that *low* feedback is negative feedback. For example, if you are comparing TV series, you can assume that the user did not like a series she watched only once. You could not assume this with cinema movies, since most users will only watch movies once and therefore there is not enough variability. However, you could group them into, for instance, genres, and again assume that the user does not like least watched genres. (2) Implicit feedback is noisy but, as we showed in previous work [5], so is explicit feedback. (3) The numerical value of implicit feedback can be directly mapped to preference given the appropriate mapping and this is the main goal of our work. On the other hand, we do agree that there is no appropriate evaluation approaches for implicit feedback and this is in fact one of the motivations of our work: if we find an appropriate way to map implicit to explicit feedback we can ensure an evaluation that is as good as the one we have in the explicit case.

Our hypothesis that there is some observable correlation between implicit and explicit feedback can be tracked in the literature. Already in 1994, Morita and Shinoda [10] proved that there was a correlation between reading time on online news and self-reported preference. Konstan *et al.* [11] did a similar experiment with the larger user base of the Grouplens project and again found this to be true. Oard and Kim [12] performed experiments using not only reading time but also other actions like printing an article to find a positive correlation between implicit feedback and ratings. Koh *et al.* did a thorough study of rating behavior in two popular websites [13]. They hypothesize that the overall popularity or average rating of an item will influence raters. The conclusion on this issue is that, while there is an effect, this depends on the cultural background of the raters.

There are two recent works that are worth mentioning since they approach the issue of implicit feedback in the music domain. Jawasher *et. al* analyze the characteristics of user implicit and explicit feedback in the context of last.fm music service [14]. The authors also report on some experiments using standard Collaborative Filtering techniques on both implicit and explicit data. However, their results are not conclusive due to limitations in the dataset. In particular, it should be noted that they only used explicit feedback available in the last.fm profiles, which is limitted to the *love/ban* binary categories. This data is very sparse and, as the authors report, almost non-existant for some users or artists. On the other hand, Kurdomova *et. al* use a Bayesian approach to learn a classifier on multiple implicit feedback variables [15]. Using these features, the authors are able to classify liked and disliked items with an accuracy of 0.75, uncovering the potential of mapping implicit feedback directly to preferences.

All these previous works, provide a qualitative intuition of the potential of implicit feedback and its relation to explicit ratings. However, they do not measure the significance of the effect of the variables, nor propose a predictive model for ratings. In this context, the main contributions of our work are: (1) A study of the relation between implicit and explicit feedback in the music domain; (2) An analysis of the effect of other context and user variables; (3) A predictive linear model that can be used to infer unknown user ratings given their implicit feedback; (4) A general approach to building such a linear mapping in other domains.

3 Experimental Setup

We conducted an online user study among users of the last.fm music service. The goal of the study was to gather explicit feedback on music albums to compare to the user implicit feedback. We obtained data on the user's listening history to use as implicit feedback by directly crawling the last.fm page related to the user taking the survey.

Explicit feedback was obtained by asking users to rate albums on a 1 to 5 star scale – see Figure 1. The items to rate were obtained from the list of albums in the user's playlist. Each user responded to a personalized survey that was generated from their last.fm profile listening history.

User Demographics In order to be accepted to the study, users had to (a) be 18 years old, and (b) have 5000 songs in their lastfm listening history. The reason for this latter requirement is that we wanted to ensure a meaningful sampling of the listening habits for the users selected in the study. This is not a limitation of

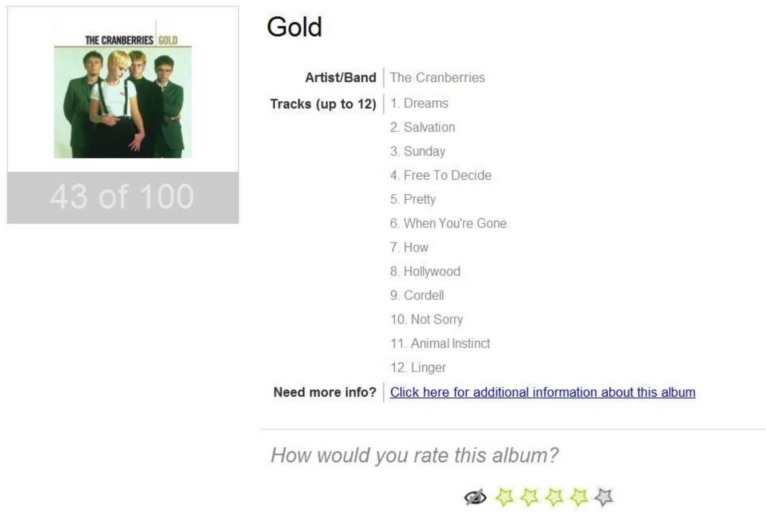

Fig. 1. Rating interface screenshot.

the approach, but rather a way to ensure the model is derived from meaningful data – although once derived it could be applied to any user. 151 users started the user experiment, and 127 completed the process. We filtered out outliers which did not present a meaningful variance in their ratings so our final study is based on 114 users.

Before starting the rating exercise, we queried users about a number of demographic variables. Out of the final users, 82% were male and 18% were female. Although we had representatives from 23 different countries, the sample was biased towards three countries: Spain (25 users), U.S. (15 users), and UK (16 users).

When asked about their internet use, more than 80% admitted to be heavy users with 20 or more hours per week. The percentage of heavy music listeners was lower, but still noticeable, with more than 50% of our users listening to music for over 20 hours per week. We were interested in having more information related to their music listening habits. Out of our subjects, almost 9% responded that they did not attend music concerts. On the other hand, 30% went to 11 or more concerts a year. Most of our subjects (35%) said that they only read music magazines or blogs *sometimes*. However, the most involved music enthusiasts who read them at least every week accounted for over 20% of our sample. Over 50% of our subjects admitted rating music online never or seldom. And only 9% reported doing so consistently (often or every week). 45% of our subjects said they bought 1 to 10 physical records a year. However, a non-negligible 18% said they did not buy any. On the other extreme, only 5% reported buying more than 21 records a year. If we look at online music shopping, more than 35% of our subjects report never doing so, but 8% say they do it once a month or more. Finally, since we are asking users to report on their preferences on “albums”, we wanted to verify whether they usually listened to albums as a whole or to single songs. Only 14% of our subjects preferred to listen to single tracks while over 45% preferred listening to full albums. The other 40% reported listening to music either way.

Item Sampling We were interested in analyzing how a number of variables influence the relation between implicit and explicit feedback. On the other hand, we want our users to face a reasonable number of items to rate.

We decided to control the following variables: user popularity, general popularity, and recentness. Our initial hypothesis is that implicit feedback is directly correlated with user explicit feedback. However, global popularity might also affect since users might feel the *social pressure* to rate higher, items with more popular acceptance. In a similar way, we hypothesize that users might rate higher those items that have been listened to recently. Our main variables are therefore: (1) *Implicit Feedback (IF)*: playcount for a user on a given item; (2) *Global Popularity (GP)*: global playcount for all users on a given item; (3) *Recentness (R)* : time elapsed since user played a given item.

Depending on the user's listening habits, a naive, random sampling strategy might yield only very popular items. Therefore, both the number of control variables and the sampling strategy that we adopt is critical. For each of the three control variables we define three bins – *low*, *medium*, and *high*. This effectively defines 27 buckets, where we place all items for a given user. Bins are not defined by simply dividing the scale for each variable in three. All these variables follow a powerlaw-like distribution. Therefore, our bins are defined logarithmically in order to guarantee that the number of items in a bucket remains reasonably homogeneous. We then follow a random sampling strategy for each bucket. Some particular combinations of variables are more unlikely than others. Therefore, and despite of our goal of having a homogeneous distribution, we obtain buckets that include anywhere from 1 to 8% of the total number of items.

4 General Analysis

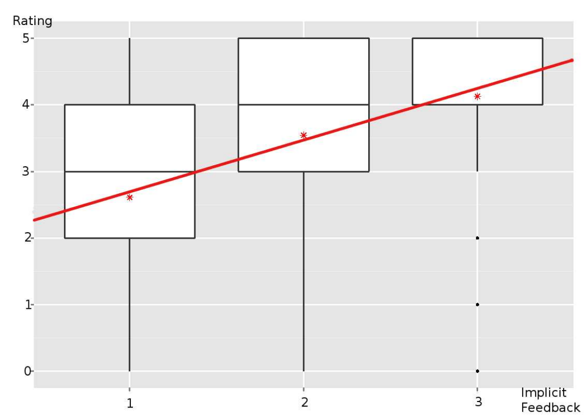

Fig. 2. Relation between implicit feedback and explicit ratings

Are user Ratings related to implicit feedback? Our initial assumption is that we have a dependent variable – the explicit rating given by the user – that depends on the user implicit feedback but also on the two other in dependent variables – overall popularity and recentness. In this initial qualitative analysis we shall first look at how these three variables affect the user ratings. We leave quantitative and significance analysis to the next section where we shall make use of multiple regression.

Figure 2 illustrates the relation between implicit feedback and ratings. Note how there is a clear correlation between the distribution of ratings and the implicit feedback. As we can see, the more implicit feedback, the higher the rating value where the distribution of ratings is centered. Note that ratings are quantized to the closest integer and this forces the median for implicit feedback of 2 and 3 to be located at 4. However, the mean, also depicted as an asterisk, clearly shows an ascending trend.

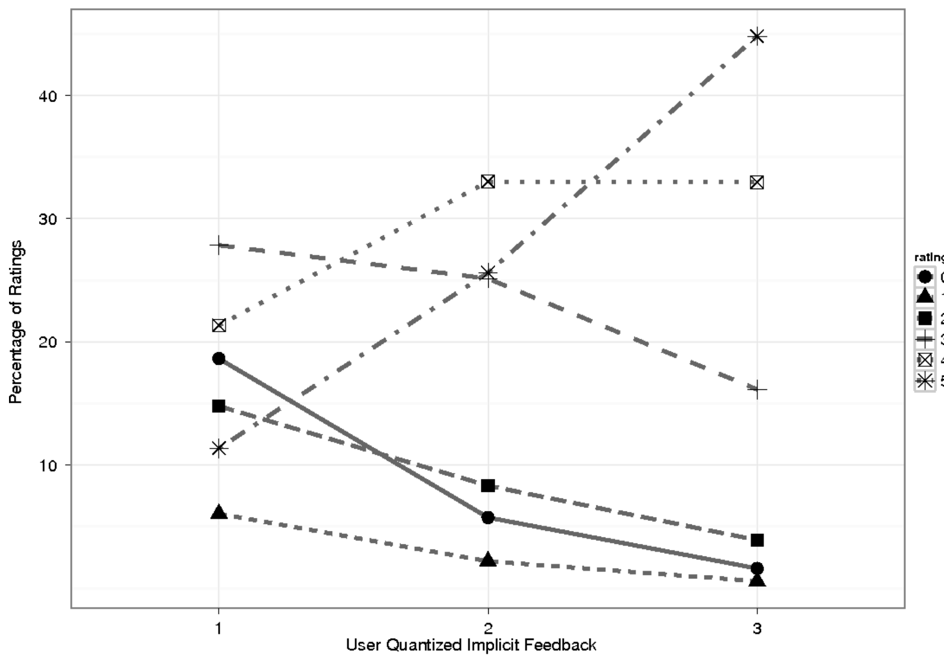

Fig. 3. Distribution of ratings given different values of implicit feedback

Figure 3 depicts the distribution of user ratings given different values for implicit feedback. Implicit feedback – *i.e.* how much they have listened to the album – is quantized as explained in the previous section and increases from 1 to 3 in the horizontal axis. We see that positive ratings that users give to the albums – ratings 4 and 5 – increase proportionally to the implicit feedback while negative ratings – ratings 1 and 2 – decrease. Rating 0 – the especial case where the user decided not to rate an item – also decreases with user feedback. Interestingly, the middle rating – *i.e.* 3 – also decreases with implicit feedback.

If we look into more details, we see that the descending slope for negative ratings 1 and 2 is constant and approximately the same. On the other hand, rating 3 is more or less stable from user feedback 1 to 2 and rapidly decreases for 3. In other words, the probability a user rates an album that she has listened to *a lot* with a 3 is significantly lower than one with an average number of listens. However, there is little difference between albums with medium and low implicit feedback.

For positive ratings we see a clear and almost constant ascending slope for the 5. However, the 4 has a different behavior that is somewhat complementary to the 3. There is a significant difference between the proportions of 4 given to low and medium feedback, but this proportion remains constant between medium and high feedback.

4.1 Effect of Other Independent Variables on Ratings

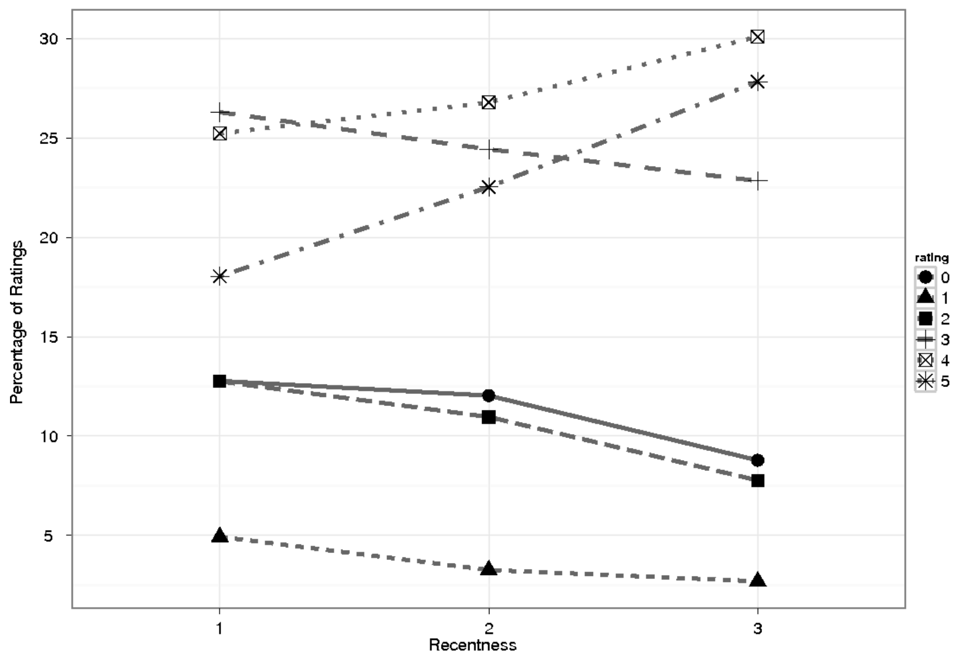

Fig. 4. Distribution of ratings given different values of recentness

Recentness. We look at the effect of the recentness factor in Figure 4. We see that this factor has a noticeable effect on all ratings. For positive ratings 4 and 5, the percentage increases almost linearly with the quantized recentness. All negative ratings and 3 decrease their percentage for more recent ratings. In other words, albums that were listened to more recently tend to receive more positive ratings and less negative ones. And, differently to what we saw with the implicit feedback variable, descending and ascending lines seem to have a similar and approximately constant slope.

Overall Popularity. We also analyze the effect of overall popularity in Figure 5. We don't see a significant effect of the independent variable for any

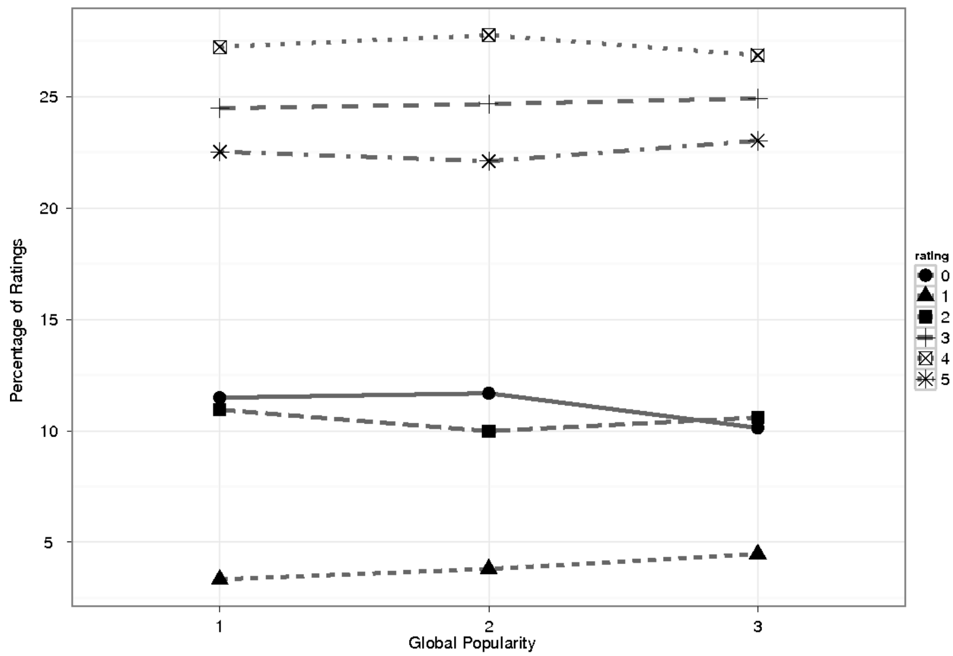

Fig. 5. Distribution of ratings given different values of overall popularity

of the ratings. Therefore, this first rough analysis seems to discard the effect of overall popularity in the user explicit rating.

Interaction Analysis. Next, we analyze the possible interaction between different pairs of variables by analyzing the corresponding interaction plots – we cannot include these figures due to space constraints. The only two variables that showed some coupling were recentness and implicit feedback. In particular, we found that for albums listened to more recently, the user needs a higher number of listens to give them a high rating. When we analyze the detailed effects of this coupling, however, we see that its effect on the average rating are not very significant.

Effect of User Variables. In the previous analysis, we are assuming that users rate items in a similar way, regardless of their demographic or musical background. However, it may well be that these variables influence and are somehow correlated with the user response to some items. In order to analyze these possible effects, we perform a multi-way ANOVA analysis on the different user variables we gathered from our initial survey.

By analyzing these results, we realize that there is only one variable with a significant contribution to explain the variance of our data – *Sig.* value below 0.05. The variable, *listen to tracks or albums*, encodes the way in which users listen to music. In particular, we were asking users whether they tend to listen to full albums, single tracks – e.g. through a radio stream –, or both. There was a significant difference on the average rating among the levels of *listen to tracks* *or albums*, $F(2, 62) = 3.949$, $p = 0.024$[3]. In order to further inspect this possible effect, we look at the relation between this variable and the different ratings in figure 6.

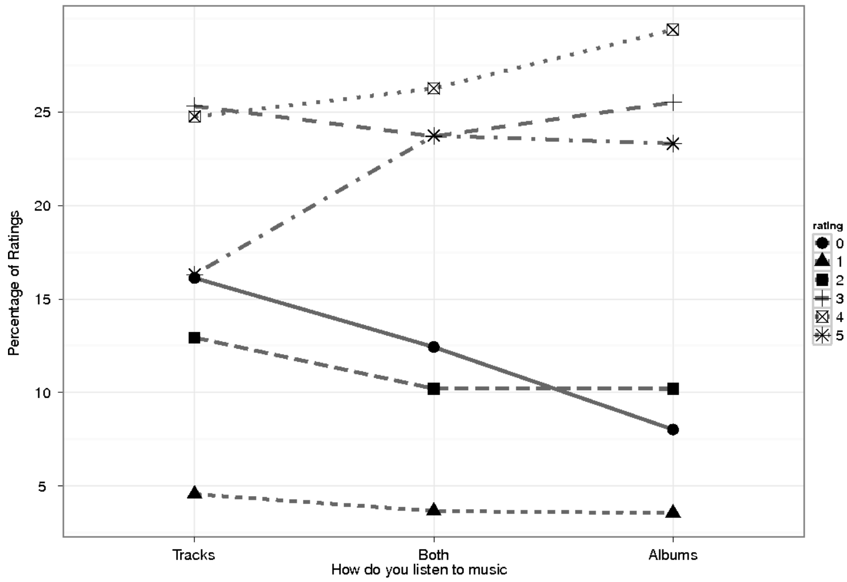

Fig. 6. Effect of listening style on percentage of ratings.

We see a number of clear trends. First, the percentage of zeros – *i.e.* not rated – is much higher for users that listen to tracks. In fact, the percentage of unrated items for users who listen to tracks – 16% – doubles the percentage in users who listen preferably to albums. This might be an expected effect since these users might not have a well-formed opinion on the quality of the album or even of its content. However, we also see significant differences in other ratings. In particular, users who tend to listen to tracks seem to have a much more critical opinion of albums since we observe a clear decrease in the percentage of positive ratings. The highest rating – 5 – captures around 23% of the ratings for the users who listen to albums and even a bit more for those who listen to albums and independent tracks. But it only corresponds to 16% of the ratings for users who listen preferably to tracks.

We conclude that, except for the user listening style, all demographic and usage variables seem to have little or no effect on the rating of items. But the kind and granularity of the interaction of the user with content should be taken into account when interpreting implicit feedback.

5 Predicting ratings

Our ultimate goal is to come up with a general model that can directly map implicit data to explicit ratings. We aim at having some kind of parametric

[3] In our survey, in order to obtain an appropriate level of granularity, we only asked users to rate albums, not tracks.

model that given input data on implicit user feedback is able to predict the rating that the user would give.

We approach our goal by performing a regression analysis in order to capture which independent variable (IV) or combination of IVs better accounts for the variations in the rating, the dependent variable (DV). In order to obtain fully meaningful results from a regression analysis, the model needs to observe some assumptions [16]. Given the way that our model is constructed, we cannot guarantee that it observes all the conditions. However, we conduct the regression analysis with the goal of having a preliminary idea. We shall then evaluate the models directly and verify our preliminary findings using a hold-out method to measure prediction error.

Types of variables In this analysis, the IVs – implicit feedback, global popularity, and recentness –, which take values either 1, 2, or 3, are considered continuous. The DV, rating, although it is rigorously an ordinal variable, is also considered continuous in order to assess the model using a common measure such as RMSE, which will make results comparable with previous research. Note that, initially, we do not consider the special case of the 0 value – *i.e.* unrated – as part of this ordinal scale. However, we will analyze the effect of including or excluding this variable in the predictive power of the derived model at the end of this Section.

Model comparison and selection After removing observations where the albums were unrated, we performed a regression analysis comparing 4 models using the remaining 10122 ratings. In all of the models, the DV is rating, but the IVs are respectively for models 1 through 3: i) implicit feedback (IF), ii) implicit feedback and recentness (RE) iii) implicit feedback, recentness and global popularity (GP). In the last model we check for possible interactions between implicit feedback with recentness. The exact formulation of each model is as follows:

- Model 1: $r_{iu} = \beta_0 + \beta_1 \cdot if_{iu}$
- Model 2: $r_{iu} = \beta_0 + \beta_1 \cdot if_{iu} + \beta_2 \cdot re_{iu}$
- Model 3: $r_{iu} = \beta_0 + \beta_1 \cdot if_{iu} + \beta_2 \cdot re_{iu} + \beta_3 \cdot gp_i$
- Model 4: $r_{iu} = \beta_0 + \beta_1 \cdot if_{iu} + \beta_2 \cdot re_{iu} + \beta_3 \cdot if_{iu} \cdot re_{iu}$

Model	R^2	*F-value*	*p-value*	β_0	β_1	β_2	β_3
1	0.125	$F(1, 10120) = 1146$	< 0.001	2.726	0.499	-	-
2	0.1358	$F(2, 10019) = 794.8$	< 0.001	2.491	0.484	0.133	-
3	0.1362	$F(3, 10018) = 531.8$	< 0.001	2.435	0.486	0.134	0.0285
4	0.1368	$F(3, 10018) = 534.7$	< 0.001	2.677	0.379	0.038	0.053

Table 1. Regression Results. R^2, F-value, and p-value for the 5 models.

Table 1 presents the results of each model. The results show that all models significantly explain the variance in the data. Besides, there are clear trends in the results. By including the variable **recentness** as a predictor, model 2 increases the amount of variability of the DV explained by the model in 10% with respect to model 1, which is reflected in the R^2 value. Although including the variable **global popularity** as a predictor increases the value of R^2, this increment is very small. This result, in addition to the fact that global popularity is not correlated to the other two IVs, supports our assumptions that the variables implicit feedback and recentness are more strongly related to rating, and subsequently, would be more useful to predict it. We see that model 4, which considers the interactions between implicit feedback and recentness, shows an important improvement over model 2. This supports our initial finding of such an interaction reported in Section 4.1. Note also that the β coefficients remain fairly constant throughout the models giving always much more importance to implicit feedback.

Predictive power of the models Finally, to test our findings in the regression analysis, we fit the 4 linear models described in the previous section using 80% of the observations, and then by doing 10-fold cross validation, we compare the predictive power of our models. We do so by measuring the root-mean squared error (RMSE) between our predictions and the actual ratings. The results in Table 2 show that all of our models improve the user average baseline significantly – 7% in the worst case. The improvement in performance when introducing other variables, is less clear. In the best case, introducing recentness, improves our accuracy in 0.5%. And, as we would expect, introducing global popularity does not improve the results significantly.

Model	RMSE1	RMSE2
User average	1.5308	1.1051
1	1.4206	1.0402
2	1.4136	1.034
3	1.4130	1.0338
4	1.4127	1.0332

Table 2. Predictive power including (RMSE1) and excluding (RMSE2) unrated items.

Predicting known ratings The above results refer to a model that predicts both ratings and non-ratings. The zero value in our ratings refers to the user not giving any feedback. Therefore, we are deriving models in which we can not only predict what the rating will be but also if the user decided to rate or not. However, we are also interested in evaluating the predictive capabilities of the models on known ratings. That is, given a pair of user and item for which we know there is the rating, how well can we predict its rating?

By comparing the results in columns 2 and 3 in Table 2, we can see that by excluding non-rated items our models have a significant gain in predictive power (RMSE decreases in more than 25%). However, relative performance of each model remains approximately the same. The improvement over the baseline predictor of user average is 6.5%.

Adding effect of user interaction style In Section 4.1, we analyzed the effect of user variables and concluded that the only one that had a significant effect was the way the user interacted with the items. We are interested in analyzing the effect of this variable on the predictive capabilities of our models. Our hypothesis is that our models should be able to predict better ratings for users who interact with music at the album level. In order to check this, we split our data into three different sets: (a) those who interact at the track level; (b) those who interact at the album level; and (c) those who interact either way.

Model	Tracks	Tracks/Albums	Albums
User average	1.1833	1.1501	1.1306
1	1.0417	1.0579	1.0257
2	1.0383	1.0512	1.0169
3	1.0386	1.0507	1.0159
4	1.0384	1.049	1.0159

Table 3. Predictive power of the regression models depending on the user interaction style. Values represent RMSE of 10-fold cross validation

In Table 3 we see that all of our models perform better when predicting users that listen preferably to albums. The decrease in RMSE is around 10%. This finding is supported when comparing the results to those obtained for the whole dataset and reported in Table 2 and finding a general improvement for users who listen to albums despite the fact that the user average baseline decreases its performance for this population. In fact, the average improvement of our predictive models over the baseline user average predictor is over 10% when segmenting the population into these three groups.

6 Conclusions and Future Work

Our analysis shows a clear relation between the amount of times users listen to an album, and the rating they report. We also find that the time elapsed since the user interacted with the album, have a significant effect but others, but the global item popularity does not influence the rating. We analyze the effect of several demographic and usage variables and find that only the granularity of the interaction style has a significant effect: .

Using the results of our analysis, we create a predictive model in which we can predict a user rating, given information of how the user interacted with an item. We perform a regression analysis to come up with several linear models and evaluate their fit to this purpose. We conclude that we can predict user ratings with an acceptable level of accuracy using a simple model that takes into account implicit feedback and recentness. In the best case, we measure an improvement over the baseline user average predictor of more than 10%.

The same approach to create a linear mapping could be applied to any domain for which we have a sample of ratings, and information about relevant context and user variables. Our results open up many possibilities for using implicit feedback in predictive tasks, especially in the context of recommender systems. Since we have a model that relates this implicit feedback to ratings, we can think of applying any of the methods used for explicit feedback on implicit data. Nevertheless, the particular model should be validated on other domains and datasets. In future work, we also plan on exploring other possible parametric approaches such as hierarchical or Bayesian models.

References

1. F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors. *Recommender Systems Handbook*. Springer, 2011.
2. D. W. Oard and J. Kim. Implicit feedback for recommender systems. In *AAAI Workshop on Recommender Systems*, 1998.
3. G. Potter. Putting the collaborator back into collaborative filtering. In *2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, 2008.

4. D. M. Nichols. Implicit rating and filtering. In *In Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering*, pages 31–36, 1997.
5. X. Amatriain, J.M. Pujol, and N. Oliver. I like it... i like it not: Evaluating user ratings noise in recommender systems. In *Proc. of the 2009 Conference on User Modeling, Adaptation, and Personalization*, 2009.
6. G. Jawaheer, M. Szomszor, and P. Kostkova. Characterisation of explicit feedback in an online music recommendation service. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 317–320, 2010.
7. L. Baltrunas and X. Amatriain. Towards time-dependant recommendation based on implicit feedback. In *Workshop on Context-Aware Recommender Systems (CARS 2009) in ACM Recsys 2009*, 2009.
8. O. Celma. Music recommendation and discovery in the long tail, 2008.
9. Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *Proceedings of ICDM 2008*, 2008.
10. M. Morita and Y. Shinoda. Information filtering based on user behavior analysis and best match text retrieval. In *SIGIR '94: Proceedings of the 17th annual international ACM SIGIR conference*, pages 272–281, New York, NY, USA, 1994. Springer-Verlag New York, Inc.
11. Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. Grouplens: applying collaborative filtering to usenet news. *Commun. ACM*, 40(3):77–87, 1997.
12. D. Oard and J. Kim. Modeling information content using observable behavior. In *Proc. of the ASIST Annual Meeting*, pages p481–88, 2001.
13. N.S. Koh, N. Hu, and E. K. Clemons. Do online reviews reflect a product's true perceived quality? - an investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications*, 2010.
14. Gawesh Jawaheer, Martin Szomszor, and Patty Kostkova. Comparison of implicit and explicit feedback from an online music recommendation service. In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, 2010.
15. S. Kordumova, I. Kostadinovska, M. Barbieri, V. Pronk, and J. Korst. Personalized implicit learning in a music recommender system. In *UMAP 2010*, 2010.
16. M. Kutner, C. Nachtschiem, W. Wasserman, and J. Neter. *Applied Linear Statistical Models*. McGraw-Hill, 4th edition, 1996.