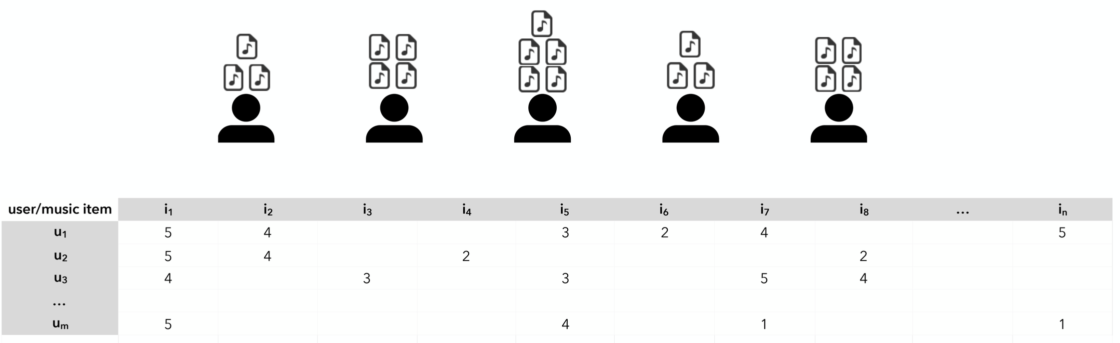
AUTOMATIC **MUSIC RECOMMENDATION** SYSTEMS: DO **DEMOGRAPHIC, PROFILING, AND CONTEXTUAL** FEATURES IMPROVE THEIR **PERFORMANCE**?

Gabriel Vigliensoni
gabriel@music.mcgill.ca

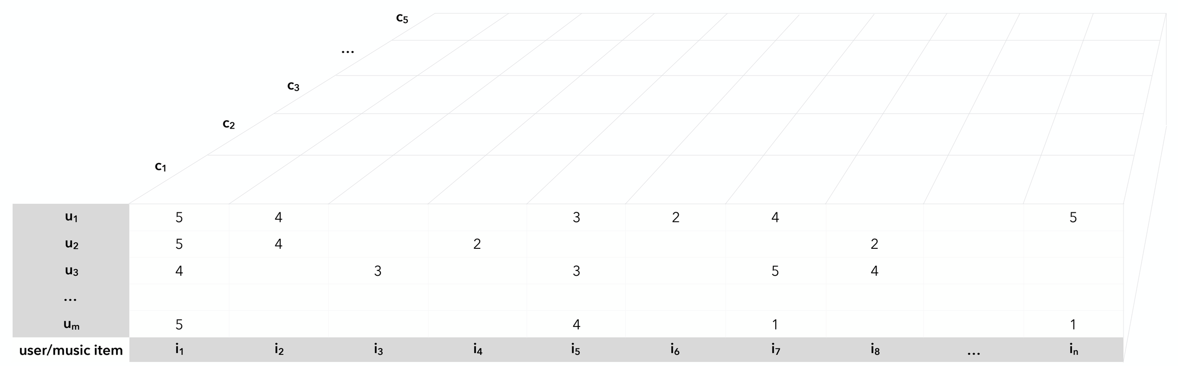
Ichiro Fujinaga
ich@music.mcgill.ca

Introduction

- Traditional music recommendation systems' performance rely on the accuracy of statis-tical models learned from past preferences of users on music items.



- Additional sources of data (e.g., demographic, profiling, and contextual features) may be used to improve the accuracy of music recommendation models.



- Our goal is to evaluate if the use of these additional sources of data improve the accuracy of a recommendation model.

Dataset and features

- We evaluate the accuracy of a music artist recommendation model learned from past preferences of listeners on music items, and their interaction with several combina-tions of people's demographic, profiling, and contextual features.

Dataset and demographic features

- We collected a large dataset of music listening histories with listener's demographic information.

Dataset general characteristics
More than half-million listeners' listening histories
LastFM API
Listening history time span > 2 years
Average number of logs per day > 10

Items	No.	Demographic	%	Age groups	%
Listeners	594K	Age	70.5	15–24	57.5
Artists	555K	Country	81.8	25–34	35.8
Albums	900K	Gender	81.6	35–44	5.5
Tracks	7M			45–54	1.2
Logs	27MM				

Table 1. Dataset and listeners' demographic characteristics.

Listener profiling features

- We hypothesized that by better understanding listeners and their listening behaviour, we will be able to more accurately model the user needs.
- We characterized people by their self-declared demographic features: *age*, *country*, and *gender*.
- We profiled listeners with a set of custom-designed music listening behavioural features:
  - **Exploratoryness**: how much a listener explores different music.
  - **Mainstreamness**: how similar a listener's listening history is to what everyone else listened to.
  - **Genderness**: how close a listener's listening history is to what people self-declared as females or males are listening to.

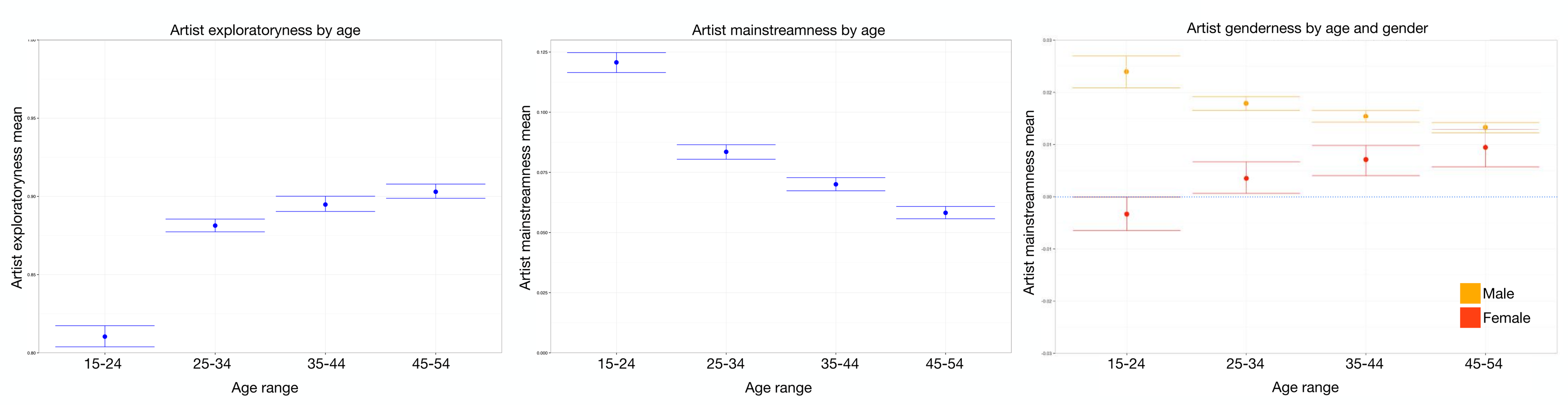


Figure 1. Feature means and 95% CI bars for a subset of the dataset listeners

Experimental procedure

- We randomly sampled 10% per-user music listening histories, split the data into training (.9) and testing (.1) datasets, and aggregated the number of playcounts per user into a 1–5 Likert preference scale.
- We used the Factorization Machines method (Rendell 2010) to estimate models based on latent factors, the additional features, and their interactions within a single framework.
- Models were learned from the training dataset for all 32 possible combinations of user-side features, and evaluated the accuracy of those models in the testing dataset.

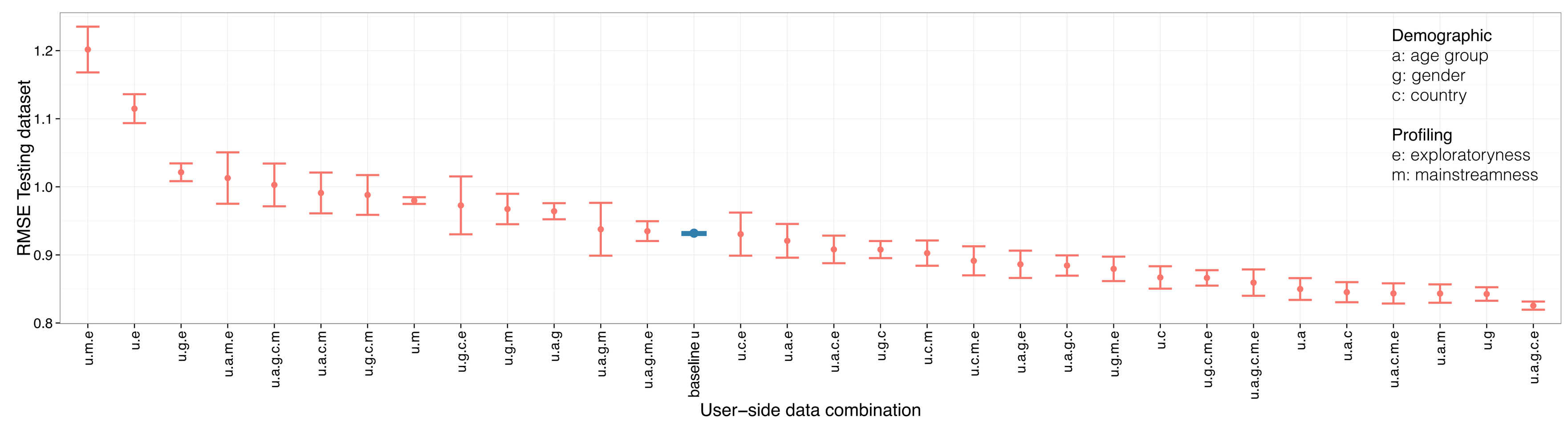


Figure 2. RMSE means and 95% CI bars for learned models evaluated in the testing dataset, with 32 combinations of the user-side features. Baseline for comparison is combination u: user's preferences only, without any additional features. Lower RMSE values mean a more accurate model.

Listening preference and context

- We hypothesized that if people listen to different music during the weekdays than on weekends, we could create more accurate models by using data from only the week-days or weekends, respectively.
- We created two additional preference matrices of listeners for artists using only listen-ing data from weekdays and weekends and performed the same experiment.

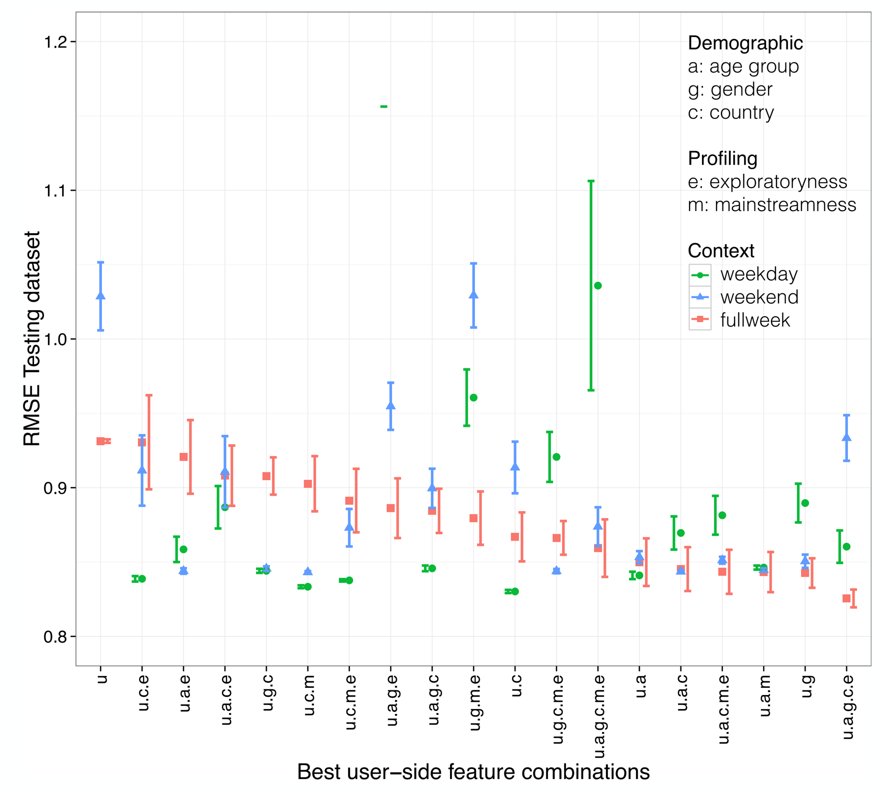


Figure 3. RMSE means and 95% CI bars for learned models with *weekday*, *weekend*, and *full-week* data.
Only those feature combinations with a better RMSE value than the baseline for *full-week* data are shown.

Results

- Custom-developed features effectively profiled some characteristics of listeners' music listening behaviour (Fig. 1).
- Estimating an artist recommendation model using listeners' self-declared age, coun-try, and gender improve the recommendation accuracy by 8 percent (Fig. 2: (u.a.g.c)).
- The model accuracy was increased to 12 percent when the profiling feature explorato-ryness was added (Fig. 2: (u.a.g.c.e)).
- Although for some feature combinations the use of split listening data improved the estimated model accuracy, the best combination of features did benefit from having full-week data (Fig. 3).

Acknowledgements

This research was supported by Comisión Nacional de Ciencia y Tecnología, Gobierno de Chile, and the Social Sciences and Humanities Research Council of Canada. Important parts of this work used ComputeCanada's High Performance Computing resources. The authors would like to thank Emily Hopkins for her thorough proofreading of this paper.

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB McGill CIRMMT Centre for Interdisciplinary Research in Music Media and Technology Social Sciences and Humanities Research Council of Canada Conseil de recherches en sciences humaines du Canada Canada

Presented at the 17th International Society for Music Information Retrieval Conference, New York City, NY, USA. 7–11 August 2016.