Predicting Genre Preferences from Cultural and Socio-economic Factors for Music Retrieval

Marcin Skowron[1], Florian Lemmerich[2], Bruce Ferwerda[3], Markus Schedl[3]

[1]Austrian Research Institute for Artificial Intelligence, Vienna, Austria
marcin.skowron@ofai.at

[2]GESIS - Leibniz Institute for the Social Sciences & University of Koblenz-Landau, Germany
florian.lemmerich@gesis.org

[3]Johannes Kepler University, Linz, Austria
{bruce.ferwerda}{markus.schedl}@jku.at

Abstract. In absence of individual user information, knowledge about larger user groups (e.g., country characteristics) can be exploited for deriving user preferences in order to provide recommendations to users. In this short paper, we study how to mitigate the cold-start problem on a country level for music retrieval. Specifically, we investigate a large-scale dataset on user listening behavior and show that we can reduce the error for predicting the popularity of genres in a country by about 16.4% over a baseline model using cultural and socio-economics indicators.

1 Introduction

While research that considers individual, user-specific aspects to improve music retrieval and recommendation algorithms has received substantial attention in the past few years, cf. [1, 8, 10], studies on cultural differences between perception of music have not been conducted in the context of retrieval until quite recently [2, 3, 6, 11]. The few existing works almost exclusively analyze the cultural differences in emotion or mood perceived when listening to music, with the aim to integrate such knowledge into music retrieval approaches [7, 13].

Gaining a more fundamental knowledge about the differences in music taste in different countries and about how these differences relate to cultural and socio-economic dimensions can help building culture-aware and cross-cultural music retrieval systems, mitigating the cold-start problem, and improving search or recommendation results by considering the cultural background of users. In this short paper, we approach the cold-start problem in which we do *not* know anything about a new user or the overall music preferences in his country, but assume that country information can be easily inferred from basic user profile information. Given cultural and socio-economic factors that are publicly available, we aim at predicting the music taste for the user's country and by doing so infer an approximation of his music taste using his country's taste as a proxy. Therefore, the specific research question we address is to which extent we can predict the overall music taste in a country given cultural and socio-economic factors.

2 Related Work

Cross-cultural research in the field of music retrieval is very limited. The studies that investigated cultural differences on users' music perception and consumption often limit themselves to a handful of cultures. For example, Hu and Lee [6] showed that there are differences between Americans and Chinese on mood perception in music, whereas Singhi and Brown [11] investigated the influence of lyrics between Canadians and Chinese. Although these findings confirm that cultural differences exist, they cannot easily be generalized to other cultures. More comprehensive studies were conducted by Ferwerda et al. [2, 3] on cultural differences in the need for music diversity. By analyzing the music consumption of users in 97 countries, they identified distinct behavior that could be related to Hofstede's cultural dimensions. In this work, we explore to which extent listeners' music preferences can be predicted across countries using cultural and socio-economical aspects and state-of-the-art machine learning techniques.

3 Datasets

In the following sections we describe how we infer music preferences on the country level and how we model cultural and socio-economic aspects.

3.1 Modeling music preferences

We model music preference on the country level by utilizing the recently published LFM-1b dataset [9],[1] which offers demographic information and detailed listening histories for tens of thousands Last.fm users. We consider in our analysis only countries with at least 100 users in the LFM-1b dataset and for which *all* the Hofstede's cultural dimensions (cf. Section 3.2) are available. The 44 countries meeting both conditions are analyzed in the remaining of this work.

We define country-specific genre profiles that are used as a proxy for music taste as follows. First, the top tags assigned to each artist in the LFM-1b dataset are fetched via the respective Last.fm API endpoint.[2] These tags provide different pieces of information, including instruments ("guitar"), epochs ("80s"), places ("Chicago"), languages ("Swedish"), and personal opinions ("seen live" or "my favorite"). We then filter for tags that encode genre and style information. For this purpose, we use as index terms a dictionary of 20 genre names retrieved from Allmusic.[3] The genre profiles are eventually created as feature vectors describing the share of each genre among all listening events of the respective country's population, according to the LFM-1b dataset. More formally, the weight of genre g in country c is given as $w_{c,g} = \frac{\sum_{a \in A_g} le_{c,a}}{\sum_{a \in A} le_{c,a}}$, where A_g is the set of artists tagged with genre g, A is the entire set of artists, and $le_{c,a}$ is the number of listening events to artist a in country c.

[1] `http://www.cp.jku.at/datasets/LFM-1b`
[2] `http://www.last.fm/api/show/artist.getTopTags`
[3] `http://www.allmusic.com`

3.2 Modeling cultural dimensions

For our study, we rely on Hofstede et al.'s cultural dimensions.[4] It is considered to be the most comprehensive and up to date framework for national cultures. They defined six dimensions to identify cultures [5]:
Power distance is defined as the extent to which power is distributed unequally by less powerful members of institutions (e.g., family). High power distance indicates that a hierarchy is clearly established and executed. Low power distance indicates that authority is questioned and attempted to distribute power equally.
Individualism measures the degree of integration of people into societal groups. High individualism is defined by loose social ties. The main emphasis is on the "I" instead of the "we," while opposite for low individualistic cultures.
Masculinity describes a society's preference for achievement, heroism, assertiveness and material rewards for success Low masculinity represents a preference for cooperation, modesty, caring for the weak and quality of life.
Uncertainty avoidance defines a society's tolerance for ambiguity. High scoring countries are more inclined to opt for stiff codes of behavior, guidelines, laws.
Long-term orientation is associated with the connection of the past with the current and future actions. Lower scoring countries tend to believe that traditions are honored and kept, and value steadfastness. High scoring countries believe more that adaptation and pragmatic problem-solving are necessary.
Indulgence denotes in general the happiness of a country. High indulgence is related to a society that allows relatively free gratification of basic and natural human desires related to enjoying life and having fun (e.g., be in control of their own life and emotions). Whereas low scoring countries show more controlled gratification of needs and regulate it by means of strict social norms.

3.3 Modeling socio-economic dimensions

In addition, we investigate a range of socio-economic indicators to predict music taste. These indicators originate from the *Quality of Government* (QoG) dataset,[5] which collects approximately 2500 variables on country-level information from more than 100 data sources. From this dataset, we extract a subset of 181 variables for which all the scores are available for the set of analyzed countries (cf. Section 3.1). To give some examples, these attributes include information on *GDP*, *income inequality*, *agriculture's share of economy*, *unemployment rate* or *life expectancy*. Details on such variables are provided in [12].

4 Experiments and Results

4.1 Approach and Methods

We want to predict the popularity of each genre in a new country based on cultural and socio-economic data. For that purpose, we employ two ensemble regression methods: gradient boosting and random forests. Gradient boosting is an

[4] `https://geert-hofstede.com/countries.html`
[5] `http://qog.pol.gu.se/data/datadownloads/qogbasicdata`

effective procedure applicable in regression problems offering a natural handling of heterogeneous features and robustness to outliers in output space. Random forests are known to show reasonable performance even with high amounts of noise visible in the features and can be used when the number of features is much larger than the number of observations. Additionally we tested Epsilon-Support Vector Regression with the linear and rbf kernels; their performance on the presented data-set was lower compared with the applied ensamble regression methods. For preprocessing, we tested a variety of techniques including univariate linear regression tests and kernel principle component analysis, but report only the best results here due to limited space. To train and evaluate the regressor, we use scores and features of the 44 countries. As comparative baseline, we consider also the average prevalence of that genre in the training set.

4.2 Results

Table 1 presents genre regressors performance over different sets of features: (i) Hofstede's dimensions, (ii) QoG dimensions, and (iii) the combined Hofstede's and QoG dimensions. We report the root–mean square error (RMSE) calculated over 5 independent, 10–fold cross–validation runs, one for each genre.

The regressors trained on the features inferred from Hofstede's and QoG dimensions outperformed the baseline approach in all the considered music genres. For 9 genres (alternative, pop, folk, rap, rnb, jazz, heavy metal, reggae, easy listening) the lowest RMSE was obtained using the QoG dimensions, in 3 cases (rock, punk, spoken word) the best performing regressors were trained only on Hofstede's dimensions, and for 6 genres (electronic, blues, country, classical, new age, world) the features were obtained from the both resources. The overall best performing regressor and resource type are the random forest regressor trained on the QoG dimensions. Here, the sum of RMSE for all the genres is at 0.1173, which constitutes a 12.2% improvement over the baseline approach. By selecting the best regressors for each genre we obtain an aggregated RMSE of 0.1117, a 16.4% reduction compared to the baseline. For variations that involve the QoG data, the improvements over the baseline are statistically significant according to Bonferroni-adjusted Wilcoxon signed rank tests with the null hypothesis that the error over each genre is equally distributed for the baseline and the respective classifier. This is not the case, if only Hofstede's dimensions are utilized.

In an additional analysis, we investigated which features influence the regressors most. While interpretation of socio-economical dimensions is often difficult, i.e., the best performing regressor uses a large number of relatively weak features of similar informative value, the features obtained from Hofstede's cultural dimensions offer more consistent and interpretable results. Specifically, *Long Term Orientation* is the most informative feature for the largest number of genres, i.e.: rock, alternative, new age, rap, rnb, electronic and jazz; *Power Distance* is the most important feature for classical, blues and reggae genres; *Indulgence* for country, pop and folk; *Masculinity* for heavy metal; *Individualism* for punk and *Uncertainty Avoidance* for the spoken word genre.

Table 1. Results: Music genre preferences regression accuracy from: Hofstede's dimensions, QoG dimensions and the combination of Hofstede's and QoG dimensions; cell values show information on the root mean squared error (RMSE) between the true and predicted genre popularities, for: (Baseline) - global average value for a genre, Gradient Boosting (G. Boost.) and Random Forest (R. Forest) regressors. Asterisk denotes the best performing method for each genre. The last line shows the Bonferroni-adjusted p-value of a Wilcoxon signed rank test compared to the baseline.

		Hofstede's dimensions		QoG dimensions		Hofstede's and QoG dimensions	
Genre	Baseline	G. Boost	R. Forest	G. Boost	R. Forest	G. Boost	R. Forest
rock	0.02592	0.02131	*0.02042	0.02258	0.02255	0.02315	0.02288
punk	0.00728	0.00695	*0.00648	0.00744	0.00724	0.00717	0.00725
spoken word	0.00051	*0.00048	0.00049	0.00049	0.00049	0.00048	0.00048
pop	0.01843	0.01783	0.01749	*0.01488	0.01551	0.01568	0.01626
alternative	0.01823	0.01619	0.01674	0.01541	*0.01483	0.01562	0.01583
folk	0.00930	0.00972	0.01029	0.00869	*0.00830	0.00833	0.00842
rap	0.00573	0.00588	0.00585	*0.00510	0.00520	0.00527	0.00537
rnb	0.00325	0.00316	0.00310	0.00309	*0.00295	0.00305	0.00308
jazz	0.00301	0.00274	0.00290	0.00279	*0.00254	0.00281	0.00272
heavy metal	0.00256	0.00248	0.00246	0.00235	*0.00228	0.00237	0.00230
reggae	0.00169	0.00186	0.00196	0.00150	*0.00139	0.00150	0.00150
easy listening	0.00067	0.00062	0.00064	0.00056	*0.00053	0.00058	0.00058
electronic	0.02557	0.02198	0.02162	0.02353	0.02321	0.02174	*0.02159
blues	0.00508	0.00457	0.00469	0.00449	0.00452	*0.00431	0.00460
country	0.00234	0.00249	0.00260	0.00219	0.00221	0.00222	*0.00218
classical	0.00211	0.00228	0.00232	0.00193	0.00191	*0.00185	0.00188
new age	0.00111	0.00095	0.00094	0.00095	0.00093	0.00094	*0.00091
world	0.00085	0.00071	0.00069	0.00076	0.00072	0.00070	*0.00069
all genres	0.13364	0.12220	0.12168	0.11873	*0.11731	0.11777	0.11852
p-value		0.466	0.733	0.004	0.001	0.001	0.001

5 Conclusions and Future Work

We presented an investigation of the predictive power of cultural and socio-economic dimensions to infer music genre preferences at the country level. We demonstrated that the application of cultural and socio-economics indicators lead to a significant reduction of the error for predicting the popularity of genres in a country by about 16.4% compared to the baseline approach, i.e., predicting the global, country-independent genre preferences. In this study we used a large-scale dataset on user listening behavior obtained from Last.fm user and analyzed how the cultural and socio-economical differences impact the users' music preferences. The study extends the scope of analysis compared to the previous works. In future work we will seek additional data sources, for instance, GPS-tagged microblogs [4], to obtain more fine grained results (e.g., at a regional level). Exploiting such precise data also enables the exploration of differences between rural and urban regions. Further, we will integrate the regressor proposed here into state-of-the-art recommendation algorithms and investigate its performance in comparison to other techniques to alleviate the cold-start problem.

6 Acknowledgments

This research is partially funded by the Austrian Science Fund (FWF) under grant no. P 27530.

References

1. Z. Cheng and J. Shen. Just-for-Me: An Adaptive Personalization System for Location-Aware Social Music Recommendation. In *Proceedings of the 2014 ACM International Conference on Multimedia Retrieval*, Glasgow, UK, April 2014.
2. B. Ferwerda and M. Schedl. Investigating the Relationship Between Diversity in Music Consumption Behavior and Cultural Dimensions: A Cross-country Analysis. In *Workshop on S*, Halifax, Canada, July 2016.
3. B. Ferwerda, A. Vall, M. Tkalčič, and M. Schedl. Exploring Music Diversity Needs Across Countries. In *User Modeling, Adaptation and Personalization*, Halifax, Canada, July 2016.
4. D. Hauger, M. Schedl, A. Košir, and M. Tkalčič. The Million Musical Tweets Dataset: What Can We Learn From Microblogs. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, Brazil, Nov. 2013.
5. G. Hofstede, G. J. Hofstede, and M. Minkov. *Cultures and Organizations: Software of the Mind*. McGraw-Hill, USA, 3rd edition, 2010.
6. X. Hu and J. H. Lee. A Cross-cultural Study of Music Mood Perception Between American and Chinese Listeners. In *Proceedings of the 13th International Society for Music Information Retrieval Conference*, Porto, Portugal, October 2012.
7. X. Hu and Y. H. Yang. Cross-dataset and Cross-cultural Music Mood Prediction: A Case on Western and Chinese Pop Songs. *IEEE Transactions on Affective Computing*, (99), 2016.
8. Y. Hu and M. Ogihara. NextOne Player: A Music Recommendation System Based on User Behavior. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, Miami, FL, USA, October 2011.
9. M. Schedl. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings of the International Conference on Multimedia Retrieval*, USA, 2016.
10. M. Schedl, S. Stober, E. Gómez, N. Orio, and C. C. Liem. User-Aware Music Retrieval. In M. Müller, M. Goto, and M. Schedl, editors, *Multimodal Music Processing*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Germany, 2012.
11. A. Singhi and D. G. Brown. On Cultural, Textual and Experiential Aspects of Music Mood. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan, October 2014.
12. J. Teorell, S. Dahlberg, S. Holmberg, B. Rothstein, A. Khomenko, and R. Svensson. *The Quality of Government Standard Dataset, version Jan16*. University of Gothenburg: The Quality of Government Institute, 2016.
13. J.-C. Wang, Y.-H. Yang, and H.-M. Wang. Affective Music Information Retrieval. In *Emotions and Personality in Personalized Services*, page 227. Springer, 2016.