LEMONS: Listenable Explanations for Music recOmmeNder Systems

Alessandro B. Melchiorre[1,2(✉)], Verena Haunschmid[1], Markus Schedl[1,2], and Gerhard Widmer[1,2]

[1] Johannes Kepler University (JKU), Linz, Austria
{alessandro.melchiorre,verena.haunschmid, markus.schedl,gerhard.widmer}@jku.at
[2] Linz Institute of Technology (LIT), Linz, Austria

Abstract. Although current music recommender systems suggest new tracks to their users, they do not provide listenable explanations of why a user should listen to them. LEMONS (Demonstration video: https://youtu.be/giSPrPnZ7mc) is a new system that addresses this gap by (1) adopting a deep learning approach to generate audio content-based recommendations from the audio tracks and (2) providing listenable explanations based on the time-source segmentation of the recommended tracks using the recently proposed audioLIME.

Keywords: Music recommendation · Explainability · audioLIME · Content-based recommendation

1 Introduction

Motivated by the impact of explainability on transparency, user satisfaction, and scrutability [1,2], different types of explanations in recommender system (RS) research have been proposed [3,4]. The adopted explanation method depends on the type of model input (e.g., user-item interaction data, content features, or contextual information), the RS algorithm (e.g., CF or CBF), and the modality used to give explanations (e.g., textually [5–8], visually [9], or graph-based user preferences [4,10,11]), cf. [4]. In music RS, research on explaining recommendations has considered music data [12–14], user data [14,15], context information [16], or a combination of the above [6,14,17], which are predominantly used to create textual explanations (such as "because you like jazz", "because users with similar taste listen to it", or "because it's Monday morning", respectively). To the best of our knowledge, none of the existing approaches provides explanations in the same modality of music itself, i.e. listenable. We address this shortcoming in the LEMONS demo[1] at hand by (1) adopting an audio-based music recommender system and (2) providing listenable explanations of the recommended tracks. LEMONS is based on the recently proposed audioLIME method [18].

A. B. Melchiorre and V. Haunschmid—These authors contributed equally.
[1] https://github.com/cpjku/lemons.

D. Hiemstra et al. (Eds.): ECIR 2021, LNCS 12657, pp. 531–536, 2021.
https://doi.org/10.1007/978-3-030-72240-1_60

2 System Overview

Music Recommender System. Existing approaches in content-based music RS usually employ metadata or acoustic features extracted from the audio track to make recommendations, which, in turn, can be used to create explanations [13, 14]. However, these approaches lead to non-listenable explanations as the audio information is either lost or compressed. In contrast, we provide explanations a user can listen to with an audio-based recommendation model inspired by state-of-the-art approaches for music tagging [19,20]. Focusing on one user at a time, we train a fully convolutional neural network[2] to predict the relevance of a specific track for the user by using its audio as input. More precisely, we consider the tracks listened to by the user as relevant while randomly selected tracks never interacted with as non-relevant [21]. We split the tracks into train, validation, and test set in an 80-10-10 fashion and select the model that achieves the best results in terms of AUC and MAP on the validation set. The results on the test set averaged across the users are 0.734 ± 0.130 MAP and 0.758 ± 0.113 AUC.

Generating Listenable Explanations. Explanations are computed post-hoc using audioLIME [18], an extension of LIME [22] for audio data. audioLIME extracts interpretable components from audios by using source separation estimates and temporal segmentation [18,23]. These interpretable components are then used as input features to fit a simple linear model that mimics the underlying RS model. The components with a positive weight are interpreted as having a positive contribution to the recommended track relevance, while the opposite is true for negative weights. When computing explanations using audioLIME, we also care how well the linear model approximates the RS model, which is reported by the fidelity score, the coefficient of determination R^2 between the linear explanation model and the RS model.

Data. We use the Million Song Dataset (MSD) and the Taste Profile Dataset [24] for training the recommender systems, as they provide listening data for about 1 million users and 300,000 songs. For this demo, we carefully select 7 users who listened to more than 900 tracks and who differ by their music preferences. The music audio data was originally obtained from 7digital[3] and the snippets' durations range from 30s to 60s. We also include and test our system on the musdb18 dataset [25], which comprises 150 songs ($\sim$10 h) belonging to 9 different genres.

3 Demonstration Overview

The landing page of our demo is shown in Fig. 1. It first introduces the 7 users from the MSD that serve as different personas (e.g., a listener with very specific

[2] Details about training and architecture can be found in our GitHub repository.
[3] https://www.7digital.com/.

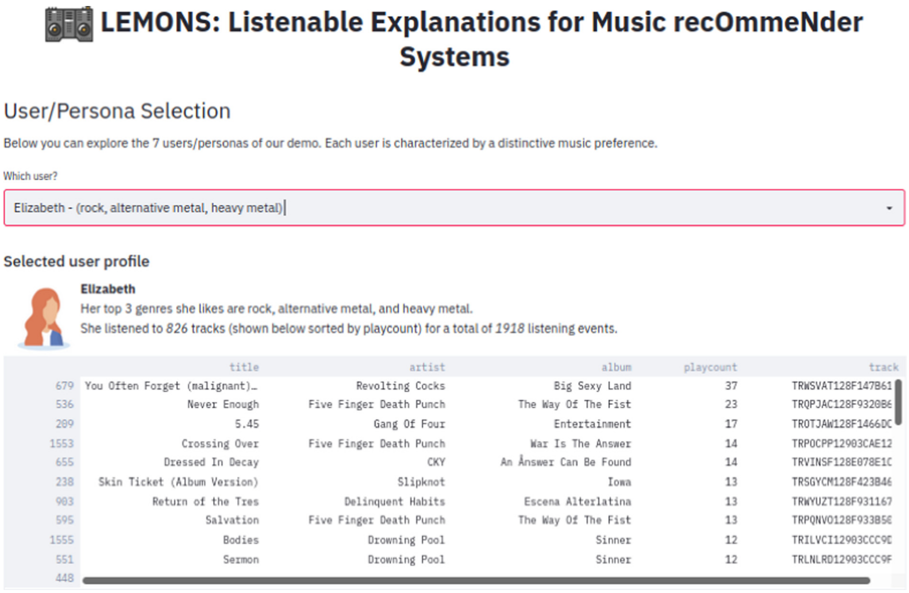

Fig. 1. Introduction of personas' music taste, listening statistics, and listened to tracks.

genre taste, very diverse taste, or a chart music follower), from which one can be selected. The selected user's profile is then shown below along with a short description of their music preferences, some music listening statistics, and the tracks they listened to. On the left (not shown in the figure), a sidebar provides clarification on how the RS and the listenable explanations work. Thereafter, the music dataset from which recommendations are computed (either MSD or musdb18) can be selected. The recommended tracks are presented to the user as a ranked list, in decreasing order of relevance. The demo user can select a song, play it, and seek within a visualization of its waveform.

As shown in Fig. 2, we offer three types of listenable explanations for the selected song depending on the interpretable components used: (1) *time-based* explanations use time segmentation to split the audio into five equally long segments, (2) *source-based* explanations use Spleeter [26] to separate the audio into 5 sources (vocals, drums, bass, piano, and other), (3) *time-and-source-based* explanations combine both, resulting in 25 interpretable components. We also describe the selected type of explanation accompanied by an illustrating image.

When the *Compute Explanation* button is pressed, the system generates the explanation and provides the fidelity score. We present two interfaces for the listenable explanations: "Top Highlight" and "Top-3". Top Highlight allows listening to the single interpretable component that influences the recommendation the most. Top-3, instead, selects the 3 most influential components. A *time-and-source-based* explanation for a track could sound like drums and bass playing in the first segment and drums playing in the third segment.

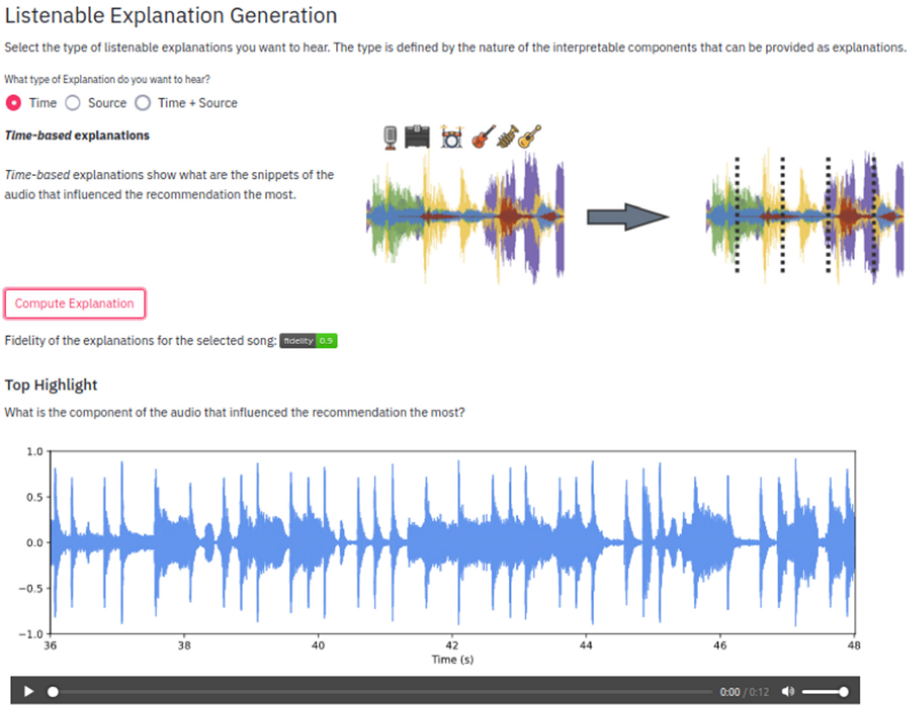

Fig. 2. Listenable Explanations: After having selected the explanation type (e.g. *time-based*), the demo shows the fidelity score and the listenable explanation interfaces. In this example, "Top Highlight" shows that the most influential component is the snippet from seconds 36 to 48.

4 Conclusion and Future Work

We presented a novel approach to generate listenable explanations for music recommender systems (LEMONS). For this purpose, we integrated audioLIME into a content-based recommender system, to uncover the pivotal components in the music audio signal which serve as explanations of why a track has been recommended to the user. As a next step, we plan to conduct a user study to investigate the quality and usefulness of the offered explanations from an end user's perspective. In addition, future work includes integrating a music segmentation technique to provide more meaningful segments for the time-based explanations (e.g., verse, chorus, or motif), and extending the purely content-based approach to a hybrid one by integrating collaborative listening data.

References

1. Tintarev, N., Masthoff, J.: Explaining recommendations: design and evaluation. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 353–382. Springer, Boston (2015). https://doi.org/10.1007/978-1-4899-7637-6_10

2. Balog, K., Radlinski, F.: Measuring recommendation explanation quality: the conflicting goals of explanations. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 329–338. Association for Computing Machinery (2020)
3. Arrieta, A.B., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020)
4. Zhang, Y., Chen, X.: Explainable recommendation: a survey and new perspectives. Found. Trends Inf. Retrieval **14**(1), 1–101 (2020)
5. Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., Ma, S.: Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 83–92. Association for Computing Machinery (2014)
6. Tsukuda, K., Goto, M.: Explainable recommendation for repeat consumption. In: 14th ACM Conference on Recommender Systems, pp. 462–467. Association for Computing Machinery (2020)
7. Li, P., Wang, Z., Ren, Z., Bing, L., Lam, W.: Neural rating regression with abstractive tips generation for recommendation share on. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 345–354. Association for Computing Machinery (2017)
8. Chang, S., Harper, F.M., Terveen, L.G.: Crowd-based personalized natural language explanations for recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 175–182. Association for Computing Machinery (2016)
9. Chen, X., et al.: Personalized fashion recommendation with visual explanations based on multimodal attention network: towards visually explainable recommendation. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 765–774. Association for Computing Machinery (2019)
10. Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., Getoor, L.: User preferences for hybrid explanations. In: Proceedings of the 11th ACM Conference on Recommender Systems, pp. 84–88. Association for Computing Machinery (2017)
11. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pp. 241–250. Association for Computing Machinery (2000)
12. Vig, J., Sen, S., Riedl, J.: Tagsplanations: explaining recommendations using tags. In: Proceedings of the 14th International Conference on Intelligent User Interfaces, pp. 47–56. Association for Computing Machinery (2009)
13. Green, S.J., et al.: Generating transparent, steerable recommendations from textual descriptions of items. In: Proceedings of the 3rd ACM Conference on Recommender Systems, pp. 329–338. Association for Computing Machinery (2009)
14. Millecamp, M., Htun, N.N., Conati, C., Verbert, K.: To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In: Proceedings of the 24th International Conference on Intelligent User Interfaces, pp. 397–407. Association for Computing Machinery (2019)
15. Sharma, A., Cosley, D.: Do social explanations work? Studying and modeling the effects of social explanations in recommender systems. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1133–1144. Association for Computing Machinery (2013)
16. Zhao, G., et al.: Personalized reason generation for explainable song recommendation. ACM Trans. Intell. Syst. Technol. **10**(4), 1–21 (2019)

17. Wang, X., Wang, D., Xu, C., He, X., Cao, Y., Chua, T.S.: Explainable reasoning over knowledge graphs for recommendation. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence, vol. 33, pp. 5329–5336. Association for the Advancement of Artificial Intelligence Press (2019)
18. Haunschmid, V., Manilow, E., Widmer, G.: audioLIME: listenable explanations using source separation. In: 13th International Workshop on Machine Learning and Music, pp. 20–24 (2020)
19. Won, M., Ferraro, A., Bogdanov, D., Serra, X.: Evaluation of CNN-based automatic music tagging models. In: Proceedings of 17th Sound and Music Computing (2020)
20. Choi, K., Fazekas, G., Sandler, M.: Automatic tagging using deep convolutional neural networks. In: Proceedings of the 17th International Conference on Music Information Retrieval (ISMIR 2016), pp. 805–811 (2016)
21. Pan, R., et al.: One-class collaborative filtering. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 502–511. Institute of Electrical and Electronics Engineers (2008)
22. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should i trust you?": explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144. Association for Computing Machinery (2016)
23. Haunschmid, V., Manilow, E., Widmer, G.: Towards Musically Meaningful Explanations Using Source Separation. CoRR abs/2009.02051 (2020). https://arxiv.org/abs/2009.02051
24. Bertin-Mahieux, T., Ellis, D.P., Whitman, B., Lamere, P.: The million song dataset. In: Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011), pp. 591–596. University of Miami (2011)
25. Rafii, Z., Liutkus, A., Stöter, F.R., Mimilakis, S.I., Bittner, R.: MUSDB18 - A Corpus for Music Separation (2017)
26. Hennequin, R., Khlif, A., Voituret, F., Moussallam, M.: Spleeter: a fast and efficient music source separation tool with pre-trained models. J. Open Source Softw. **5**(50), 2154 (2020)