AudioLens: Audio-Aware Video Recommendation for Mitigating New Item Problem

Mohammad H. Rimaz[1][0000−0002−7777−9556], Reza Hosseini[2], Mehdi Elahi[3][0000−0003−2203−9195], and Farshad Bakhshandegan Moghaddam[4]

[1] Technical University of Kaiserslautern, Erwin-Schrödinger-Str 52, 67663 Kaiserslautern, Germany
mrimaz@rhrk.uni-kl.de
[2] Vaillant Group Business Services, Berghauser Str. 63, 42859 Remscheid, Germany
seyed-reza.hosseini@vaillant-group.com
[3] University of Bergen, Fosswinckelsgt. 6, 39100 Bergen, Norway
mehdi.elahi@uib.no
[4] University of Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Germany
farshad.bakhshandegan@uni-bonn.de

Abstract. From the early years, the research on recommender systems has been largely focused on developing advanced recommender algorithms. These sophisticated algorithms are capable of exploiting a wide range of data, associated with video items, and build quality recommendations for users. It is true that the excellency of recommender systems can be very much boosted with the performance of their recommender algorithms. However, the most advanced algorithms may still fail to recommend video items that the system has no form of representative data associated to them (e.g., tags and ratings). This is a situation called *New Item* problem and it is part of a major challenge called *Cold Start*. This problem happens when a new item is added to the catalog of the system and no data is available for that item. This can be a serious issue in video-sharing applications where hundreds of hours of videos are uploaded in every minute, and considerable number of these videos may have no or very limited amount of associated data.

In this paper, we address this problem by proposing recommendation based on novel features that do not require human-annotation, as they can be extracted completely automatic. This enables these features to be used in the cold start situation where any other source of data could be missing. Our proposed features describe audio aspects of video items (e.g., energy, tempo, and danceability, and speechiness) which can capture a different (still important) picture of user preferences. While recommendation based on such preferences could be important, very limited attention has been paid to this type of approaches.

We have collected a large dataset of unique audio features (from Spotify) extracted from more than 9000 movies. We have conducted a set of experiments using this dataset and evaluated our proposed recommendation technique in terms of different metrics, i.e., Precision@K, Recall@K, RMSE, and Coverage. The results have shown the superior performance

of recommendations based on audio features, used individually or combined, in the cold start evaluation scenario.

Keywords: Recommender Systems · Audio Visual · Multimedia · Cold Start.

1 Introduction

It is known that *YouTube*, an example of popular video-sharing applications, has about 1.5 billion active users who consume incredible number of 5 billion videos per day [5]. Hence, it is not uncommon to observe confused video consumers with problem in deciding what to watch from a missive volume and variety of videos [3].

Recommender Systems can cope with this problem by supporting the users when making decision on what to watch [24,31,27]. Recommender systems can build personalized video suggestions based on the *particular* interests of users for videos and find what can better match users' needs and constraints [30,32]. Over the many years, wide range of video recommendation algorithms have been proposed and evaluated presenting excellency in performance. These algorithms can receive a variety of data sources, e.g., content-associated data (tags), and generate personalized recommendations on top of this data [17,27,9,36,2].

While the performance of these recommender algorithms can impact the quality of the generated recommendations, however, any type of algorithms may fail to generate relevant recommendations of video items which have no or very limited amount of associated data [14,34,39,26]. This is a situation known as *New Item Cold Start* problem, which typically occurs when a new item is added to the catalog of the system and no input data is available for that item [22]. This is a major problem in video-sharing applications, such as YouTube where hundreds of hours of videos are uploaded in every minute, by millions of active video makers [6].

Furthermore, collecting the traditional types of content-associated data, that are typically represented by semantic attributes (e.g., tags), requires either a group of experts or a network of users [12,29,7,6,40]. This indeed is an expensive process and needs human efforts. Then recommendations based on these costly semantic attributes still may not properly capture the true users' preferences, e.g., the user tastes associated with audio characteristics of videos.

In addressing this problem, this article investigates the potential behind different types of audio features representative of video content in building quality recommendations for users. We have exploited two different audio features that can be extracted completely *automatic* without any need for costly *manual* human annotation. Hence they can be exploited by any content-based recommender algorithm capable of incorporating them in the recommendation process.

[5] `https://www.omnicoreagency.com/youtube-statistics`
[6] `http://tubularinsights.com/hours-minute-uploaded-youtube/`

We have compared quality of recommendation based on the (automatic) audio features against other types of (automatic) features and (manual) tags. The comparisons have been conducted with respect to various evaluation metrics (i.e., Precision@N, Recall@N, RMSE, and Coverage) using a large dataset of more than ≈ 18M ratings obtained from a large network of ≈ 162K users who provided the ratings for ≈ 9K movies.

The overall results of the evaluation have shown the consistent superiority of the recommendations based on our novel audio features over the traditional tags.

- we propose a novel technique for video recommendation based on audio features (e.g., energy, tempo, and danceability, and speechiness) that can be extracted automatically, without any need for costly human-annotation;
- we will publish a large dataset [7], which is the most important contribution of this paper, that contains a wide range of audio features (collected from Spotify) for 9,104 movies, linked directly with the user ratings and tags (+ other descriptors such as visual features);
- we have evaluated the recommendations based on novel audio features in cold start scenarios, when features are used *individually* or when used in *combination* with other features; we tested the recommendation quality exploiting using millions of ratings given by hundreds of thousands of users;

2 Related Works

This work is related to two research fields, i.e., the *Cold Start* problem and *Audio-aware* Recommendation Systems.

One of the major problems of recommender systems in general is the cold start problem, i.e., when a new user or a new item is added to the catalog and the system does not have sufficient data associated with these users/items [4]. In such a case, the system cannot properly recommend existing items to a new user (new user problem) or recommend a new item to the existing users (new item problem) [1]. In video domain, one of the effective approaches that can tackle the cold start problem exploit different forms of video content for generating recommendation [2]. Such video content can be manually added, e.g. tags [25,15], or automatically extracted, e.g., visual descriptors [11,23,5,20,33].

Another form of content data that can be used for video recommendation is based on audio descriptors [35]. Very limited works have focused on investigating such type of descriptors and their potential in representing user preferences. As an example, in [28] the correlation between user music taste and his/her personality has been discussed. Several medium and weak correlations between music audio features and personality traits have been shown, and their results have provided useful insights into the relationship between the personality and the music preference. Moreover, authors in [18] have collected a dataset of movies and television shows matched with subtitles and soundtracks and analyzed the

[7] https://github.com/mhrimaz/audio-lens

relationship between story, song, and the user taste. However, they have taken a non-personalized approach and used IMDb ratings. [41] has investigated the effect of the movie soundtrack search volume on the movie revenue in different time periods. It has shown that the online search volume of a movie soundtrack has an effect on the movie revenue. [16] has investigated the relationship between the musical and visual art preferences, and the role of personality traits in predicting preferences for different musical styles and visual art motives. Beside this, [10] recommender system has integrated the some forms of deep learning features as well as block-level and i-vector audio features of more than 4,000 movie trailers.

This work differs from the prior works in different aspects. In terms of dataset, prior works extracted the audio features from movie trailers or short clips (e.g., in [10]), while in our dataset, the audio features have been extracted from original score soundtracks for full-length movies. Even though in [18] the authors take a similar approach, however, their focus is not really personalization. Moreover, we cover almost double in number of items. Second, in our experiments, we use the recently released MovieLens25M dataset, with much larger number of ratings. Finally, unlike previous datasets, e.g., introduced by [18,10], our data went through extensive manual checks, and errors have been corrected with careful expert checks.

3 Proposed Method

3.1 Data Collection Process

We did the data collection process in two phases. In the first phase, we queried albums in Spotify with a specific pattern *"{movie_name} (Original Motion Picture Soundtrack) {year}"*. This naming pattern is quite prevalent within the music industry, and many publisher's releases follow this naming convention. This phase was completely automated using the Spotify Search API [8] to find a Spotify identifier (Spotify ID) for each movie. Each Spotify ID could represent an album or a playlist. However, there are several shortcomings to this approach. First, many albums do not follow this naming convention (e.g., *"Toy Story (Soundtrack)"*). Second, some movies do not have any related published album, whereas their soundtrack is a playlist in Spotify. To alleviate this problem, and enhance the quality of our dataset, in the second phase, we carefully checked each individual entry, manually. A team of 7 trained person taught to check the matching manually. Several criteria and identifier factors have been used to check the correctness of matching. First and foremost, the album's poster and the movie's poster should usually look identical or share some common elements. Moreover, composer information and track names checked against various online resources, including IMDb's soundtrack section and Wikipedia. In some few cases, the decision is inconclusive, which in such cases, we simply removed the entry from the dataset. We manually checked the corresponding movie or playlist Spotify

[8] `https://developer.spotify.com/documentation/web-api/`

identifier for missing popular movies with the highest number of ratings in the IMDb platform. We decided to use IMDb since many new releases may have very low number of ratings in MovieLens25M [19] [9] dataset released on January 2019. With the advent of sophisticated signal processing techniques, automatically extracting musical and vocal features from a full-length movie would be possible for real-world recommender systems. Since this is out of the scope of this research, we used already existed Spotify API. By having a manual checking procedure, we are ensuring to have a high quality and error-prone dataset for further researches.

3.2 Dataset Description

Our dataset provides a link between every movie and its corresponding soundtrack in Spotify (using Spotify ID). We found the Spotify ID for 9,104 movies. These movies received 18,745,630 ratings from 16,254 users. For each Spotify ID, we could find a corresponding album or playlist, which contains the number of included music tracks. Using the unique ID, we could collect the representing audio features provided by Spotify Audio Feature API[10]. The following list, briefly explains our collected audio features:

- ***f1*: Acousticness** is a confidence measure from 0.0 to 1.0 (high confidence) of whether the track is acoustic.
- ***f2*: Danceability** describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. The value is in the range of [0,1].
- ***f3*: Energy** is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.
- ***f4*: Instrumentalness** predicts whether a track contains no vocals. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
- ***f5*: Liveness** shows the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
- ***f6*: Loudness** is the overall loudness of the entire track in decibels (dB) ranging typically between -60 and 0 db. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude).
- ***f7*: Popularity** of a track is a value between 0 and 100, and is based on the total number of plays the track has had and how recent those plays are.

[9] `https://grouplens.org/datasets/movielens/25m/`

[10] `https://developer.spotify.com/web-api/get-audio-features`

- ***f8*: Speechiness** detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value.
- ***f9*: Tempo** is the speed or pace of a given piece and is the overall estimated tempo of a track in beats per minute.
- ***f10*: Track Duration** is the duration of the track in milliseconds.
- ***f11*: Valence** is a measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. More positive tracks (e.g. happy, cheerful, euphoric) have higher valence sound, while tracks with low valence sound more negative (e.g. sad, depressed, angry).
- ***f12*: Key** is the estimated overall key of the track. The values ranging from 0 to 11 mapping to pitches using standard Pitch Class notation [11] (E.g. 0 = C, 1 = C-sharp/D-flat, 2 = D, and -1 if no key was detected).
- ***f13*: Mode** indicates the modality (major is 1 and minor is 0) of a track, the type of scale from which its melodic content is derived. Note that the major key (e.g. C major) could more likely be confused with the minor key at 3 semitones lower (e.g. A minor) as both keys carry the same pitches.
- ***f14*: Time Signature** specifies how many beats are in each bar (or measure). It ranges from 3 to 7 indicating time signatures of "3/4", to "7/4".

3.3 Recommendation algorithm

We adopted a classical "K-Nearest Neighbor" content-based algorithm. Given a set of users $u \in U$ and a catalogue of items $i \in I$, a set of preference scores r_{ui} provided by user u to item i has been collected. Moreover, each item $i \in I$ is associated to its feature vector f_i. For each couple of items i and j, the similarity score s_{ij} is computed using *cosine similarity*. For each item i the set of its nearest neighbors closer that a specified threshold NN_i is built, Then, for each user $u \in U$, the predicted preference score $\hat{r}_{ui}$ for an unseen item i is computed as follows

$$s_{ij} = \frac{{f_i}^T f_j}{f_i f_j} \qquad \textit{and} \qquad \hat{r}_{ui} = \frac{\sum_{j \in NN_i, r_{uj}>0} r_{uj} s_{ij}}{\sum_{j \in NN_i, r_{uj}>0} s_{ij}} \tag{1}$$

3.4 Baselines

We have compared our proposed recommendation technique (*AudioLens*) against recommendation based on a range of automatic and manual features. For automatic features, that can be used in cold start situation, we considered recommendation based on **Musical Keys** and **Visual features**. Musical keys can be also collected from Spotify and be a informative descriptor of the musics composed for movies. Visual features is a novel form of content descriptors that has been shown to be effective in cold start situation. In our experiment, we

[11] `https://en.wikipedia.org/wiki/Pitch_class`

used a recent dataset *MA14KD* [12] that have shown promising results in recommender systems [13]. In addition, we combined both audio and visual features and formed **Hybrid features** in order to compare the recommendation based on these features used individually or in combination. All of these features can be extracted automatically and adopted for recommendation in cold start situation. For the sake of comparison, we consider recommendation based on manual **Tags** which certainly need human-annotation and may be missing in cold start situation. However, this form of recommendation can still be included as a traditional baseline in our experiment.

4 Experimental Result

4.1 Evaluation Methodology

For evaluation, we followed a methodology similar to the one proposed by [8]. We used a large rating dataset, i.e., MovieLense25M with 25M ratings, and filtered out users who have rated at least 10 relevant items (i.e., items with ratings equal or higher than 4). This ensured us that each user has a minimum number of favorite items. Then we randomly selected 4000 users for our experiment. For each selected user, we choose 2 items with rating equal or higher than 4 (forming a favorite set of items). Then we randomly add 500 items not rated by the user to this set. After that we predict the ratings for all the 502 movies using the recommender system and order them according to the predicted ratings. For each $1 \leq N \leq 502$, number of hits will be the number of favorite movies appear in top N movies (e.g. 0, 1 or 2). Assume T is the total number of favorite items in the test set for all selected users ($T = 8000$ in our case), then:

$$recall@N = \frac{\#hits}{T} \qquad and \qquad precision@N = \frac{\#hits}{N \cdot T} = \frac{recall@N}{N} \tag{2}$$

In addition to these metrics, we also computed *Root Mean Squared Error (RMSE)*, i.e., the rating prediction error, and *Coverage* [35], i.e., the proportion of items over which the system is capable of generating recommendations [21].

4.2 Experiment A: Exploratory Analysis

In experiment A, we performed a set of exploratory analysis. Due to the space limit, we focus on reporting some interesting results we observed by analyzing the time evolution of the audio features over the history of (sound) cinema. For that, we computed the yearly average of every audio features for the period of 1940 to 2020. Figures 1 and 2 illustrate the obtained results. Interestingly, there are two opposite trends in the evolution of the audio features over time, i.e., a positive trend (for audio features such as **Energy**, **Danceability**, and **Tempo** shown in Figure 1), and a negative trend (for audio features such as

[12] `https://zenodo.org/record/3266236#.Xx7hLPgzako`

Liveness, **Acousticness**, and **Instrumentalness** shown in Figure 2). These trends indicate that while, over the history of cinema, the musics of the movies have become more energetic with higher tempo (and perhaps more danceable), at the same time, the musics are also losing their liveness, acousticness, and instrumentalness.

Another interesting observation is that, according to our collected audio features, the musics of the newer movies (produced after 2000) have different characteristics compared to the older movies (produced before 2000). In earlier years of cinema, the musics of the movies illustrate more diversity in terms of our audio features. This could mean that composers have been making a more similar type of music for newer movies. In addition, the observed trend for newer movies goes slightly into the opposite direction compared to the older movies (e.g., see the u-turn in figure 2-middle, around 2000s). This might be due to the fact that the music production has encountered a big shift in 2000s with the introduction of *digital* composition techniques [13]. We could not present all figures for the other audio features, due to space limit. However similar trends have been observed for them.

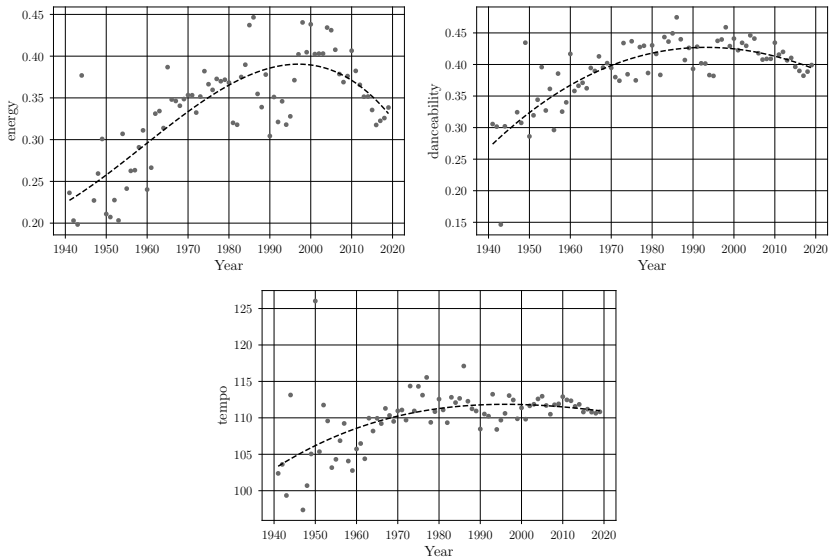

Fig. 1: Time evolution of **Energy** (left), **Danceability** (left), and **Tempo** (bottom) audio features over history of cinema.

[13] https://www.filmindependent.org/blog/know-score-brief-history-film-music/

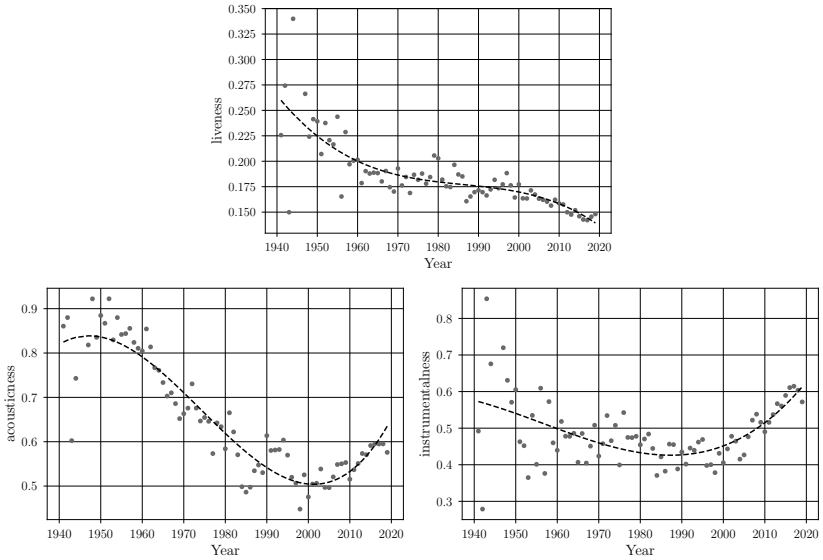

Fig. 2: Time evolution of **Liveness** (top), **Acousticness** (left), and **Instrumentalness** (right) features over history of cinema.

4.3 Experiment B: Recommendation Quality

In experiment B, we evaluated our audio-aware recommendation technique (*AudioLens*) and compared it against different baselines, e.g, recommendation based on other automatic features (i.e., musical key, visual features, and hybrid features) as well as recommendation based on manual tags (see Section 3.1 for more details). Figure 3 and Figure 4 illustrate the results.

In terms of the precision@N, as shown in Figure 3 (left), the best results have been consistently achieved by AudioLens, i.e., our proposed recommendation approach based (automatic) audio features. The precision value of AudioLens is 0.0023, 0.0023, 0.0022, 0.0022 for recommendation sizes (N) of 5, 10, 15, and 20, respectively. The second best approach is recommendations based on visual features which achieves precision of 0.0017, 0.0018, 0.0019, and 0.0019 for growing recommendation sizes of 5, 10, 15, and 20. The worst results have been achieved for recommendation based on (manual) tags with values of 0.0008, 0.0010, 0.0011, and 0.0012 for different recommendation sizes.

In terms of Recall@N, similar results have been observed, as depicted in Figure 3 (right). Again, recommendation based on (automatic) audio features (AudioLens) obtains the best results, visual features are the second best, and again, the worst results achieved by recommendation based on tags.

In terms of RMSE, presented in Figure 4 (left), our proposed recommendation technique based on the audio features (AudioLens) achieves superior results compared to the other features, with RMSE values of 0.83. Recommendation based on visual features has also obtained relatively good results with RMSE values of 0.86. The results of the other features were not substantially different

from each other, and indeed, despite the differences in the feature types, they perform similarly in terms of rating prediction accuracy.

Finally, in terms of Coverage, illustrated in Figure 4 (right), all (automatic) audio and visual features achieves the best coverage of 100%. This means that these features can be used to cover the entire item catalog of a recommender system. This is while recommendation based on tags achieves the worst results, i.e., coverage of 93%.

An important observation we made is that, recommendation based on (automatic) hybrid features has not achieved a superior performance compared to the recommendation based on (automatic) audio features. This means that a combining the audio and visual features will not necessarily result in improvement on recommendation quality. This could be related to the hybridization method, as we used a simple combination of audio and visual features, while a more advanced feature fusion method can be expected to enhance these outcomes.

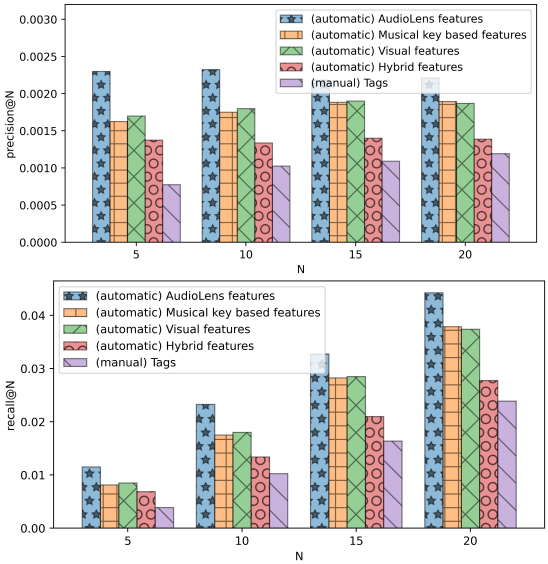

Fig. 3: Quality of movie recommendation, based on different content features, w.r.t, **Precision (top)** and **Recall (bottom)**

5 Conclusion

This paper addresses the cold start problem by proposing a recommendation technique based on *audio* features that can be automatically extracted with no need for human involvement. These novel features can represent video items when neither any rating nor any tag is available for a new video item. We have conducted a preliminary experiments to better investigate the potential power

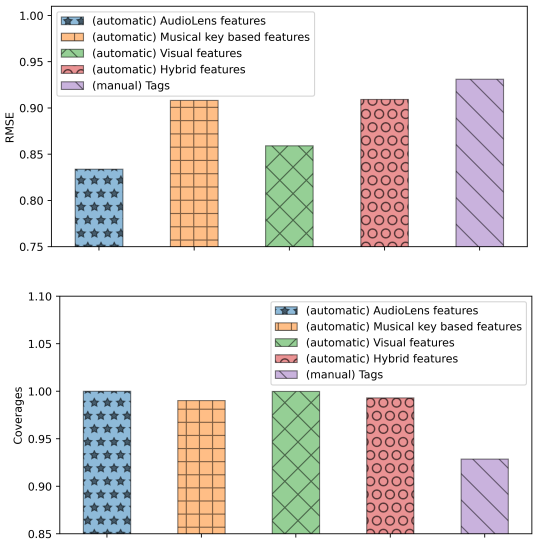

Fig. 4: Quality of movie recommendation, based on different content features, w.r.t, **RMSE (top)** and **Coverage (bottom)**

of these audio features in generating video recommendation and compared the results against user tags labeled manually. The experiment has been conducted using our new dataset with novel audio features extracted from more than 9000 movies. The results of the experiment have shown consistent superiority of these audio features in generating relevant recommendation and hence effectively dealing with the the cold start problem.

Our plans for future work includes eliciting user-generated video content from other video sharing social networks (e.g., Instagram). We also plan to obtain the implicit preferences of music listeners through their facial appearance using recent findings [37] that have shown correlation between peoples musical preferences and their facial expressions [38].

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005). https://doi.org/10.1109/TKDE.2005.99
2. Aggarwal, C.C.: Content-based recommender systems. In: Recommender Systems, pp. 139–166. Springer (2016)
3. Anderson, C.: The Long Tail. Random House Business (2006)
4. Bakhshandegan Moghaddam, F., Elahi, M.: Cold start solutionsfor recommendation systems. Big Data Recommender Systems, Recent Trends and Advances IET (2019)

5. Brezeale, D., Cook, D.J.: Automatic video classification: A survey of the literature. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on **38**(3), 416–430 (2008)
6. Cantador, I., Bellogín, A., Vallet, D.: Content-based recommendation in social tagging systems. In: Proceedings of the fourth ACM conference on Recommender systems. pp. 237–240. ACM (2010)
7. Cantador, I., Konstas, I., Jose, J.M.: Categorising social tags to improve folksonomy-based recommendations. Web semantics: science, services and agents on the World Wide Web **9**(1), 1–15 (2011)
8. Cremonesi, P., Koren, Y., Turrin, R.: Performance of recommender algorithms on top-n recommendation tasks. In: Proceedings of the fourth ACM conference on Recommender systems. pp. 39–46 (2010)
9. De Gemmis, M., Lops, P., Semeraro, G., Basile, P.: Integrating tags in a semantic content-based recommender. In: Proceedings of the 2008 ACM conference on Recommender systems. pp. 163–170. ACM (2008)
10. Deldjoo, Y., Constantin, M.G., Eghbal-Zadeh, H., Ionescu, B., Schedl, M., Cremonesi, P.: Audio-visual encoding of multimedia content for enhancing movie recommendations. In: Proceedings of the 12th ACM Conference on Recommender Systems. p. 455–459. RecSys '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3240323.3240407
11. Deldjoo, Y., Elahi, M., Cremonesi, P., Garzotto, F., Piazzolla, P., Quadrana, M.: Content-based video recommendation system based on stylistic visual features. Journal on Data Semantics pp. 1–15 (2016)
12. Di Noia, T., Mirizzi, R., Ostuni, V.C., Romito, D., Zanker, M.: Linked open data to support content-based recommender systems. In: Proceedings of the 8th International Conference on Semantic Systems. pp. 1–8. ACM (2012)
13. Elahi, M., Hosseini, R., Rimaz, M.H., Moghaddam, F.B., Trattner, C.: Visually-aware video recommendation in the cold start. In: Proceedings of the 31st ACM Conference on Hypertext and Social Media. pp. 225–229 (2020)
14. Elahi, M., Ricci, F., Rubens, N.: A survey of active learning in collaborative filtering recommender systems. Computer Science Review (2016)
15. Enrich, M., Braunhofer, M., Ricci, F.: Cold-start management with cross-domain collaborative filtering and tags. In: Huemer, C., Lops, P. (eds.) E-Commerce and Web Technologies - 14th International Conference, EC-Web 2013, Prague, Czech Republic, August 27-28, 2013. Proceedings. Lecture Notes in Business Information Processing, vol. 152, pp. 101–112. Springer (2013). https://doi.org/10.1007/978-3-642-39878-0_10, `https://doi.org/10.1007/978-3-642-39878-0_10`
16. Ercegovac, I.R., Dobrota, S., Kuščević, D.: Relationship between music and visual art preferences and some personality traits. Empirical Studies of the Arts **33**(2), 207–227 (2015). https://doi.org/10.1177/0276237415597390
17. Gedikli, F., Jannach, D.: Improving recommendation accuracy based on item-specific tag preferences. ACM Transactions on Intelligent Systems and Technology (TIST) **4**(1), 11 (2013)
18. Gillick, J., Bamman, D.: Telling stories with soundtracks: An empirical analysis of music in film. In: Proceedings of the First Workshop on Storytelling. pp. 33–42. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). https://doi.org/10.18653/v1/W18-1504, `https://www.aclweb.org/anthology/W18-1504`
19. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. ACM Trans. Interact. Intell. Syst. **5**(4) (Dec 2015). https://doi.org/10.1145/2827872

20. Hazrati, N., Elahi, M.: Addressing the new item problem in video recommender systems by incorporation of visual features with restricted boltzmann machines. Expert Systems p. e12645 (2020)
21. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. **22**(1), 5–53 (2004). https://doi.org/10.1145/963770.963772
22. Hornick, M.F., Tamayo, P.: Extending recommender systems for disjoint user/item sets: The conference recommendation problem. IEEE Transactions on Knowledge and Data Engineering (8), 1478–1490 (2012)
23. Hu, W., Xie, N., Li, Zeng, X., Maybank, S.: A survey on visual content-based video indexing and retrieval. Trans. Sys. Man Cyber Part C **41**(6), 797–819 (Nov 2011). https://doi.org/10.1109/TSMCC.2011.2109710
24. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems: An Introduction. Cambridge University Press (2010)
25. Liang, H., Xu, Y., Li, Y., Nayak, R.: Tag based collaborative filtering for recommender systems. In: Wen, P., Li, Y., Polkowski, L., Yao, Y., Tsumoto, S., Wang, G. (eds.) Rough Sets and Knowledge Technology, 4th International Conference, RSKT 2009, Gold Coast, Australia, July 14-16, 2009. Proceedings. Lecture Notes in Computer Science, vol. 5589, pp. 666–673. Springer (2009). https://doi.org/10.1007/978-3-642-02962-2_84
26. Lika, B., Kolomvatsos, K., Hadjiefthymiades, S.: Facing the cold start problem in recommender systems. Expert Systems with Applications **41**(4), 2065–2073 (2014)
27. Lops, P., De Gemmis, M., Semeraro, G.: Content-based recommender systems: State of the art and trends. In: Recommender systems handbook, pp. 73–105. Springer (2011)
28. Melchiorre, A.B., Schedl, M.: Personality correlates of music audio preferences for modelling music listeners. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. p. 313–317. UMAP '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3340631.3394874
29. Milicevic, A.K., Nanopoulos, A., Ivanovic, M.: Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions. Artificial Intelligence Review **33**(3), 187–209 (2010)
30. Resnick, P., Varian, H.R.: Recommender systems. Commun. ACM **40**(3), 56–58 (1997). https://doi.org/10.1145/245108.245121
31. Ricci, F., Rokach, L., Shapira, B.: Recommender systems: Introduction and challenges. In: Recommender Systems Handbook, pp. 1–34. Springer US (2015)
32. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B.: Recommender Systems Handbook. Springer (2011)
33. Rimaz, M.H., Elahi, M., Bakhshandegan Moghadam, F., Trattner, C., Hosseini, R., Tkalčič, M.: Exploring the power of visual features for the recommendation of movies. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization. pp. 303–308 (2019)
34. Rubens, N., Elahi, M., Sugiyama, M., Kaplan, D.: Active learning in recommender systems. In: Recommender systems handbook, pp. 809–846. Springer (2015)
35. Schedl, M., Zamani, H., Chen, C., Deldjoo, Y., Elahi, M.: Current challenges and visions in music recommender systems research. Int. J. Multim. Inf. Retr. **7**(2), 95–116 (2018). https://doi.org/10.1007/s13735-018-0154-2
36. Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R.: Personalized recommendation in social tagging systems using hierarchical clustering. In: Proceedings of the 2008 ACM conference on Recommender systems. pp. 259–266. ACM (2008)

37. Tkalčič, M., Maleki, N., Pesek, M., Elahi, M., Ricci, F., Marolt, M.: A research tool for user preferences elicitation with facial expressions. In: Proceedings of the Eleventh ACM Conference on Recommender Systems. pp. 353–354. ACM (2017)
38. Tkalčič, M., Maleki, N., Pesek, M., Elahi, M., Ricci, F., Marolt, M.: Prediction of music pairwise preferences from facial expressions. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. p. 150–159. IUI '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3301275.3302266
39. Vlachos, M., Duenner, C., Heckel, R., Vassiliadis, V.G., Parnell, T., Atasu, K.: Addressing interpretability and cold-start in matrix factorization for recommender systems. IEEE Transactions on Knowledge and Data Engineering (2018)
40. Wang, L., Zeng, X., Koehl, L., Chen, Y.: Intelligent fashion recommender system: Fuzzy logic in personalized garment design. IEEE Trans. Human-Machine Systems **45**(1), 95–109 (2015)
41. Xu, H., Goonawardene, N.: Does movie soundtrack matter? the role of soundtrack in predicting movie revenue. In: Siau, K., Li, Q., Guo, X. (eds.) 18th Pacific Asia Conference on Information Systems, PACIS 2014, Chengdu, China, June 24-28, 2014. p. 350 (2014), `http://aisel.aisnet.org/pacis2014/350`