The 2019 Multimedia for Recommender System Task: MovieREC and NewsREEL at MediaEval

Yashar Deldjoo,[1] Benny Kille,[2] Markus Schedl,[3] Andreas Lommatzsch,[2] Jialie Shen[4]
[1]Polytechnic University of Bari, Italy, [2]Technical University of Berlin, Germany,
[3]Johannes Kepler University Linz, Austria, [4]Queen's University Belfast, UK
deldjooy@acm.org,benjamin.kille@dai-labor.de,markus.schedl@jku.at,
andreas.lommatzsch@dai-labor.de,j.Shen@qub.ac.uk

ABSTRACT

The MediaEval 2019 Task "Multimedia for Recommender Systems" investigates the potential of leveraging multimedia content to enhance recommender systems. In this task, participants use a wealth of information from text, images, and audio to predict the success of items. Thereby, we advance the state-of-the-art of content-based recommender systems by leveraging multimedia content.

KEYWORDS

movies, news, recommender systems, multimedia, content-based filtering, feature engineering, text, video, audio, images

1 INTRODUCTION

Recommender systems support users in their decision making by focusing them on a small selection of items out of a large catalogue [11]. To date, most recommendation models use collaborative filtering (CF), content-based filtering on metadata (CBF-metadata), or a combination thereof at their core. CF-based models exploit the collaborative power of interactions encoded in users' implicit or explicit feedbacks to compute recommendations, thereby entirely disregarding the role of content. CBF-metadata models, on the other hand, solely resort to metadata (editorial or user-generated) to generate recommendations, disregarding the perception of media content [5]. We argue that human interpretation of media items is by nature content-oriented in which multimedia content including the audio and visual content play a key role on driving users' preferences [7, 8]. Thus, recommendation systems should offer users the chance to learn more about their multimedia taste (e.g., their visual or musical taste) and their semantic interests [4–6, 12].

The MovieREC and NewsREEL tasks aim to facilitate using multimedia content in recommender systems [7]. Participants can engage with two subtasks covering different domains. The *movie recommendation* task asks participants to predict the average rating for movies, their rating variance, together with popularity scores. The *news recommendation* task challenges participants to predict the number of reads of news articles. In this overview paper, we present the goal of each task, discuss the features provided by the organizers, and provide a description of the ground truth and evaluation methods as well as the required runs.

MediaEval'19, 27-29 October 2019, Sophia Antipolis, France

2 MOVIE RECOMMENDATION

The entertainment industry is a several-hundred-billion-dollar industry. Producing a new movie means that the company is betting on the movie's success. With the goal to make this endeavor successful, producers and investors must utilize various professional promotional methodologies including movie trailers, to publicize the film already a long time before its release. Machine learning techniques can be used to predict the success of a movie, allowing producers and investors to decide whether or not to support similar movies [1]. While previous works have mostly focused on exploiting pre-release factors such as metadata including actors' names, writers, producers, genre, production company, etc. to predict the success of a movie, the goal of the current task is to use content-based features extracted from different modalities (audio, visual, and textual) of the movie trailers to predict how a movie will be received by its viewers. Using movie trailers instead of the entire movies to extract features makes the system more versatile and effective as trailers are more easily available than the full movies.

2.1 Task Description

The input to the system is a set of audio, visual, and text features derived from selected movie trailers. Task participants must create an automatic system that can predict the *average ratings* that users will assign to movies (representing the overall appreciation of the movie by the audience), the *rating variance* (representing the agreement of user ratings)[1] as well as the *popularity score* (characterized by number of ratings given to each movie by all users).

2.2 Data

Participants are supplied with audio and visual features extracted from movie trailer as well as associated metadata (genre and tag labels). The development set (devset) and test set provide features for 10 898 and 2725 trailers. It should be noted that content descriptor types in the current task are similar to MediaEval 2018 Movie Recommendation task [3] with the difference that in MediaEval 2018, *video clips* were used to extract features from audio and visual modalities while in MediaEval 2019 we use movie trailers. A movie can have several associated video clips, but (we assume) it has only one corresponding movie trailer. The content descriptors are organized in three categories.

Metadata descriptors (found in folder `Metadata`) are provided as two CSV files containing *genre* and *user-generated tags* associated with each movie. The metadata features come in pre-computed numerical format instead of the original textual format [3].

[1]Note that in fact it is required to predict standard deviation of ratings, cf. Section 2.3 but due to intelligibility we use the term "variance" instead of standard deviation.

Audio features (found in folder `Audio`) include block level features (BLF) and i-vector features [3]. The BLF data includes the raw features of the 6 sub-components (sub-features) that describe various audio aspects: *spectral aspects*, *harmonic aspects*, *rhythmic aspects*, and *tonal aspects*. The i-vector features, describing *timbre* are computed based on numbers of Gaussian mixtures (GMMs) and the total variability dimension (tvDim). BLF feature vectors are provided in 6 separate CSV files, containing the raw feature vectors of the sub-components. The i-vector features are provided for GMMs = (256, 512) and tvDim = (100, 200).

Visual features (found in folder `Visual`) are contained in two sub-folders: Aesthetic visual features (AVF) and Deep AlexNet Fc7 features [3]. AVF captures three fundamental properties in image composition, *color*, *texture* and *objects*. This is while, the Deep features uncover *syntactic* and *semantic* information about visual content. We provide the AVF and Deep AlexNet features using 4 temporal feature aggregation techniques, based on average value across all frames (denoted `Avg`) and median (denoted `Med`).

Although different hyper parameters are provided, for simplicity, participants can rely on one choice of hyperparameters to build their system. Then, if interested, different hyperparameters could be tested to improve the accuracy of the developed system.

2.3 Run Description and Evaluation Metrics

Every team will be asked to provide 1 submission file, containing 3 predicted scores for the items given in the test set. The scores should be in comma-separated format in the form i.e., $(id,\ s_1,\ s_2,\ s_3)$, where id is the item id, s_1 is the predicted score for rating average, s_2 is the predicted score for rating standard deviation and s_3 is the predicted score for the popularity score. The evaluation of participants' runs is realized by using the standard error metric root-mean-square-error (RMSE) between the predicted scores and the actual scores according to the ground truth , $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - \hat{s}_i)^2}$ where N is the number of scores in the test set on which the system is validated, s_i is the actual score of users given to item i and $\hat{s}_i$ is the predicted score. Note, that during test data release, participants are provided only with the identifiers of test movie trailers and the corresponding features and they are expected to predict three types of the scores.

3 NEWS RECOMMENDATION

The continuing expansion of the world-wide web has lead publishers to distribute news more rapidly and to a larger readership. Readers, on the other hand, struggle to find exactly those stories they want to read. Consequently, publishers have introduced news recommender systems to assist their readers [2, 9].

3.1 Task Description

Participants take on the role of a news recommender system. They must predict which articles will attract most readers for a selection of weeks. This follows the intuition that publishers suggesting these items will maximize their readers' engagement. Conversely, publishers suggesting articles with few reads will experience less engagement. Participants obtain a data set to conduct experiments.

3.2 Data

The data describe the activity on a large German news publisher in the time between 1 January and 31 March, 2019. The data set contains 14 049 articles and 14 638 images published during the thirteen week period. For each article, the data set provides the link to the image, the first 256 characters of text, a stemmed version thereof, the URL to the article, and the URL to the image. In addition, the data show how frequently users read articles for some of the thirteen weeks. More specifically, the data provides the reading statistics for the weeks 1 to 3 and 7 to 9. The data set excludes statistics for the remaining weeks for testing. Furthermore, the data set includes automatically generated labels for the images. We have processed each image with automatic image annotators. We used both TensorFlow and Keras with pre-trained models `VGG16`, `VGG19`, and `Inception`. In total, the data set contains 762 137 labels. Labels carry confidence scores reflecting the degree of certainty with which the model has assigned the label. Finally, the data set includes an activation layer of ImageNet for each image. The data set lacks the images. Participants can collect them using the URLs.

3.3 Run Description and Evaluation Metrics

The task considers four target weeks: 5, 11, 12, and 13. Participants must predict the number of reads for each article in each of these weeks. Let $a \in \mathcal{A}$ refer to the articles and $w \in \mathcal{W}$ represents the weeks. Then, we challenge you to predict the number of reads for article a in week w, $\nu(a, w) \in \mathbb{R}^{+}$. Still, accurately predicting the number of reads is only part of a successful contribution. A publisher needs information about which articles to push to readers thus maximizing their engagement. Consequently, we evaluate submission in terms of precision. Precision measures to what degree a ranked list includes known positive entries. We derive the positive entries from the hold-out evaluation set. Having obtained the predictions, we sort the articles accordingly. Then, we check the top of the list for two settings. We compute the precision, which refers to the fraction of hits in the top of the list [10]. Formally, $p = {}^{|L \cup G|}/_{|L|}$, where p refers to the precision, L to the top of the list, and G to the ground truth of target articles. We consider two settings. First, we the $|L|$ to ten. This reflects how well the algorithm detects the top articles. Second, we set $|L|$ to ten percent of the set of articles. This covers a larger portion of the article collection ans signals whether the algorithm finds interesting articles further down the list. Finally, we determine the best submission in terms of these two figures: p@10 and p@10%. Each participant can submit at most five lists of predictions.

4 CONCLUSIONS

The 2019 Multimedia Recommendation Task provides an unified framework for evaluating participants' recommendation approaches for news and movies. Both task provide multi-media content and meta-data features.Details regarding the methods and results of each individual run can be found in the working note papers of the MediaEval 2019 workshop proceedings.

REFERENCES

[1] Javaria Ahmad, Prakash Duraisamy, Amr Yousef, and Bill Buckles. 2017. Movie success prediction using data mining. In *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 1–4.

[2] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google News Personalization: Scalable Online Collaborative Filtering. In *Proceedings of the 16th International Conference on World Wide Web*. ACM, 271–280.

[3] Yashar Deldjoo, Mihai Gabriel Constantin, Athanasios Dritsas, Bogdan Ionescu, and Markus Schedl. 2018. The MediaEval 2018 Movie Recommendation Task: Recommending Movies Using Content. In *Working Notes Proceedings of the MediaEval 2018 Workshop, Sophia Antipolis, France, 29-31 October 2018.* http://ceur-ws.org/Vol-2283/MediaEval_18_paper_9.pdf

[4] Yashar Deldjoo, Mihai Gabriel Constantin, Hamid Eghbal-Zadeh, Markus Schedl, Bogdan Ionescu, and Paolo Cremonesi. 2018. Audio-Visual Encoding of Multimedia Content to Enhance Movie Recommendations. In *Proceedings of the Twelfth ACM Conference on Recommender Systems*. ACM. https://doi.org/10.1145/3240323.3240407

[5] Yashar Deldjoo, Maurizio Ferrari Dacrema, Mihai Gabriel Constantin, Hamid Eghbal-zadeh, Stefano Cereda, Markus Schedl, Bogdan Ionescu, and Paolo Cremonesi. 2019. Movie genome: alleviating new item cold start in movie recommendation. *User Model. User-Adapt. Interact.* 29, 2 (2019), 291–343. https://doi.org/10.1007/s11257-019-09221-y

[6] Yashar Deldjoo, Mehdi Elahi, Massimo Quadrana, and Paolo Cremonesi. 2018. Using Visual Features based on MPEG-7 and Deep Learning for Movie Recommendation. *International Journal of Multimedia Information Retrieval* (2018), 1–13.

[7] Yashar Deldjoo, Markus Schedl, Paolo Cremonesi, and Gabriella Pasi. 2018. Content-Based Multimedia Recommendation Systems: Definition and Application Domains. In *Proceedings of the 9th Italian Information Retrieval Workshop (IIR 2018)*. Rome, Italy.

[8] Yashar Deldjoo, Markus Schedl, Balázs Hidasi, and Peter Knees. 2018. Multimedia recommender systems. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM, 537–538.

[9] Florent Garcin, Boi Faltings, Olivier Donatsch, Ayar Alazzawi, Christophe Bruttin, and Amr Huber. 2014. Offline and Online Evaluation of News Recommender Systems at swissinfo.ch. In *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 169–176.

[10] Asela Gunawardana and Guy Shani. 2015. Evaluating Recommender Systems. In *Recommender Systems Handbook*. Springer, 265–308.

[11] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, 1–34.

[12] Kirsten Swearingen and Rashmi Sinha. 2002. Interaction Design for Recommender Systems. In *Designing Interactive Systems*, Vol. 6. 312–334.