Multimedia Recommender Systems

Yashar Deldjoo
Politecnico di Milano
deldjooy@acm.org

Markus Schedl
Johannes Kepler University
markus.schedl@jku.at

Balázs Hidasi
Gravity R&D
balazs.hidasi@gravityrd.com

Peter Knees*
TU Wien
peter.knees@tuwien.ac.at

ABSTRACT

This tutorial introduces *multimedia recommender systems* (MMRS), in particular, recommender systems that leverage multimedia content to recommend different media types. In contrast to the still most frequently adopted collaborative filtering approaches, we focus on content-based MMRS and on hybrids of collaborative filtering and content-based filtering. The target recommendation domains of the tutorial are *movies*, *music* and *images*. We present state-of-the-art approaches for multimedia feature extraction (text, audio, visual), including deep learning methods, and recommendation approaches tailored to the multimedia domain. Furthermore, by introducing common evaluation techniques, pointing to publicly available datasets specific to the multimedia domain, and discussing the grand challenges in MMRS research, this tutorial provides the audience with a profound introduction to MMRS and an inspiration to conduct further research.

KEYWORDS

multimedia recommender systems; video recommendation; music recommendation; image recommender systems; feature extraction; deep learning

ACM Reference Format:
Yashar Deldjoo, Markus Schedl, Balázs Hidasi, and Peter Knees. 2018. Multimedia Recommender Systems. In *Twelfth ACM Conference on Recommender Systems (RecSys '18), October 2–7, 2018, Vancouver, BC, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3240323.3241620

1 MOTIVATION AND BACKGROUND

Data available on the Web and by content providers nowadays encompass several different media types, including text, audio, video, and images. The abundance of this kind of data is made accessible by multimedia recommender systems (MMRS), in which either the input features (item descriptors) or output items (recommendations) are composed of several media types.

The majority of MMRS algorithms effect recommendations using either content-based filtering (CBF) based on textual data such as metadata or collaborative filtering (CF) leveraging the correlations among user interactions. However, the content of a multimedia item can be described in more versatile ways. For a movie, these include its genre, actors, and mise-en-scène reflected in its audio-visual content. For a music piece, style, rhythm, instrumentation, lyrics, but also cultural background of the performer are important descriptors, among others. Still, metadata features are the most commonly used in today's recommender systems

In stark contrast, in the multimedia community, extracting content descriptors from different media types is a well-established research area. So is the automatic inference of semantic descriptors by means of machine and deep learning. This tutorial therefore aims at bridging the gap between the multimedia, machine learning, and recommender systems communities. We believe that recommender systems research can strongly benefit from knowledge in multimedia signal processing established over the past years for solving various multimedia recommendation tasks.

2 TUTORIAL DESCRIPTION

We first introduce the notion of MMRS [8]. In particular, we present the typical viewpoints of the multimedia and the recommender systems communities and discuss how they can be connected for mutual benefit. We further categorize MMRS in terms of the stage in the recommendation process at which multimedia content can be used (e.g., feature representation as input or as items to recommend). Based on this categorization, we discuss which recommendation algorithms can be applied for which scenario (e.g., CF-MMRS, CB-MMRS, MM-driven RS [8]).

In the main part, we focus on the domains of movie and music recommendation and partly as well image recommendation, covering the following topics:

Multimedia feature extraction: We categorize multimedia features into audio/music, image/video, text, and metadata, and present the state of the art in feature extraction from each modality. We particularly discuss i-vectors [4, 9, 25] and block-level features [4, 15] for audio/music, aesthetic features and AlexNet deep features [4, 19, 20] for image/video, and features derived from lyrics and subtitles via vector space models and topic modeling [3, 18, 21] for text.

MMRS approaches: We elaborate on the state-of-the-art approaches that exploit the introduced multimedia features to build MMRS. More precisely, we clarify that multimedia recommendation is not only about recommending a particular media type. Rather, there exists a variety of other tasks in which the analysis of multimedia input can be usefully exploited to provide recommendations of various kinds. In particular, we categorize three main types of

*Peter Knees acknowledges support by the *SmarterJam* project (FFG grant no. 858514).

RecSys '18, October 2–7, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5901-6/18/10.
https://doi.org/10.1145/3240323.3241620

systems: (i) CB-MMRS, (ii) CF-MMRS, and (iii) MM-driven RS and show how these systems differently incorporate MM content in the recommendation process.

Feature extraction via deep learning: We provide examples for automatic feature extraction by deep neural networks, discussing convolutional and recurrent networks as well as architectures for standalone feature extraction using these components. We further discuss how to integrate any type of extracted latent content features into latent feature based CF models to enable hybridization.

End-to-end deep models: One of the advantages of deep learning is modularity, which allows for easy integration of multiple information sources into a single, which can be trained by end-to-end using gradient descent. In theory, these models completely eliminate manual feature engineering, if enough data is available. We examine this statement and also compare end-to-end training and pretraining of features.

Evaluation and datasets: We discuss the particularities when evaluating MMRS (e.g., the need to consider sequential characteristics in playlist recommendation or the strong contextual component for outfit recommendation via fashion images) and point to a few existing datasets that integrate multimedia descriptors and preference information, such as MMTF-14K[1] for movies [4] and the Million Song Dataset[2] (and its extensions) for music [2].

In the last part, we discuss the grand challenges MMRS research is facing, such as (i) the establishment of standardized and public datasets that integrate rating data and multimedia content descriptors [4], (ii) the need for transparent and fair recommendation approaches based on multimedia descriptors, and (iii) sequence-aware MMRS that consider users' context and intent. By providing some practical guidelines, we finally intend to help researchers new to the area of MMRS shaping their ideas for future research directions on this interesting topic.

3 INSTRUCTORS

Dr. Yashar Deldjoo completed his PhD at Politecnico di Milano, Italy. His research interests include *recommender systems and personalization*, *multimedia*, and *machine learning*. Selected publications: [4–7, 10, 24, 26]

Dr. Markus Schedl is an Associate Professor at the Johannes Kepler University Linz, Institute of Computational Perception. His research interests include *music recommender systems*, *data analytics*, and *social media mining*. Selected publications: [4, 5, 8, 15, 22, 23].

Dr. Balázs Hidasi is the Head of Research and Data Mining at Gravity R&D. His main research areas are *deep learning for recommender systems*, *matrix and tensor factorization*, *session-based and context-aware recommendations*. Selected publications: [11–14, 16].

Dr. Peter Knees is an Assistant Professor of the Faculty of Informatics of TU Wien. His research interests include *music information retrieval* and *recommender systems in creative domains*. Selected publications: [1, 17, 18, 22].

REFERENCES

[1] K. Andersen and P. Knees. Conversations with expert users in music retrieval and research challenges for Creative MIR. In *Proc 17th International Society for Music Information Retrieval Conference*, ISMIR, 2016.

[2] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere. The million song dataset. In *Proc 12th International Society for Music Information Retrieval Conference*, ISMIR, 2011.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Machine Learning Research*, 3:993–1022, 2003.

[4] Y. Deldjoo, M. G. Constantin, B. Ionescu, M. Schedl, and P. Cremonesi. MMTF-14K: A Multifaceted Movie Trailer Dataset for Recommendation and Retrieval. In *Proc 9th ACM Multimedia Systems Conference*, MMSys, 2018.

[5] Y. Deldjoo, P. Cremonesi, M. Schedl, and M. Quadrana. The effect of different video summarization models on the quality of video recommendation based on low-level visual features. In *Proc 15th International Workshop on Content-Based Multimedia Indexing*, CBMI, 2017.

[6] Y. Deldjoo, M. Elahi, M. Quadrana, and P. Cremonesi. Using visual features based on MPEG-7 and deep learning for movie recommendation. *International Journal of Multimedia Information Retrieval*,1–13, 2018.

[7] Y. Deldjoo, C. Frà, M. Valla, A. Paladini, D. Anghileri, M. A. Tuncil, F. Garzotta, P. Cremonesi, et al. Enhancing children's experience with recommendation systems. In *Proc Workshop on Children and Recommender Systems - 11th ACM Conference of Recommender Systems*, KidRec, 2017.

[8] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi. Content-Based Multimedia Recommendation Systems: Definition and Application Domains. In *Proc 9th Italian Information Retrieval Workshop*, IIR, 2018.

[9] H. Eghbal-Zadeh, M. Schedl, and G. Widmer. Timbral modeling for music artist recognition using i-vectors. In *Proc 23rd European Signal Processing Conference*, EUSIPCO, 2015.

[10] M. Elahi, Y. Deldjoo, F. Bakhshandegan Moghaddam, L. Cella, S. Cereda, and P. Cremonesi. Exploring the semantic gap for movie recommendations. In *Proc 11th ACM Conference on Recommender Systems*, RecSys, 2017.

[11] B. Hidasi, A. Karatzoglou, O. Sar-Shalom, S. Dieleman, B. Shapira, and D. Tikk. Dlrs 2017: Second workshop on deep learning for recommender systems. In *Proc 11th ACM Conference on Recommender Systems*, RecSys, 2017.

[12] B. Hidasi and D. Tikk. Enhancing matrix factorization through initialization for implicit feedback databases. In *Proc 2nd Workshop on Context-awareness in Retrieval and Recommendation*, CaRR, 2012.

[13] B. Hidasi and D. Tikk. Initializing matrix factorization methods on implicit feedback databases. *Journal of Universal Computer Science*, 19(12):1834–1853, 2013.

[14] B. Hidasi and D. Tikk. General factorization framework for context-aware recommendations. *Data Mining and Knowledge Discovery*, 30(2):342–371, 2016. First online: 07 May 2015.

[15] M. Kaminskas, F. Ricci, and M. Schedl. Location-aware Music Recommendation Using Auto-Tagging and Hybrid Matching. In *Proc 7th ACM Conference on Recommender Systems*, RecSys, 2013.

[16] A. Karatzoglou and B. Hidasi. Deep learning for recommender systems. In *Proc 11th ACM Conference on Recommender Systems*, RecSys, 2017.

[17] P. Knees and K. Andersen. Building physical props for imagining future recommender systems. In *Proc Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces*, HUMANIZE, 2017.

[18] P. Knees and M. Schedl. A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Transactions on Multimedia Computing, Communications, and Applications*, TOMM, 10(1), 2013.

[19] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. In *European Conference on Computer Vision*, Springer, 2016.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 2012.

[21] C. Laurier, J. Grivolla, and P. Herrera. Multimodal Music Mood Classification Using Audio and Lyrics. In *Proc 7th International Conference on Machine Learning and Applications*, ICMLA, 2008.

[22] M. Schedl, P. Knees, B. McFee, D. Bogdanov, and M. Kaminskas. *Recommender Systems Handbook*, ch. Music Recommender Systems. Springer, 2nd ed., 2015.

[23] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, and M. Elahi. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7(2):95–116, 2018.

[24] Y. Deldjoo, M. Elahi, P. Cremonesi, F. Garzotto, and P. Piazzolla. Recommending movies based on mise-en-scene design. *In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1540–1547, 2016

[25] A. Vall, H. Eghbal-zadeh, M. Dorfer, M. Schedl, and G. Widmer. Music playlist continuation by learning from hand-curated examples and song features: Alleviating the cold-start problem for rare and out-of-set songs. In *Proc 2nd Workshop on Deep Learning for Recommender Systems*, 2017.

[26] Y. Deldjoo, M. G. Constantin, H. Eghbal-Zadeh, M. Schedl, B. Ionescu, and P.Cremonesi. Audio-Visual Encoding of Multimedia Content to Enhance Movie Recommendations In *Proc 12th ACM Conference on Recommender Systems*, RecSys, 2018.

[1] https://mmprj.github.io/mtrm_dataset/benchmark.html

[2] https://labrosa.ee.columbia.edu/millionsong