Exploring the Power of Visual Features for Recommendation of Movies

Mohammad Hossein Rimaz
K. N. Toosi University of Technology
Tehran, Iran
mhrimaz@email.kntu.ac.ir

Mehdi Elahi
Free University of Bozen-Bolzano
Bolzano, Bozen, Italy
meelahi@unibz.it

Farshad Bakhshandegan Moghadam
Karlsruhe Institute of Technology
Karlsruhe, Germany
farshad.moghaddam@student.kit.edu

Christoph Trattner
University of Bergen
Bergen, Norway
christoph.trattner@uib.no

Reza Hosseini
PmOne Analytics GmbH
Paderborn, Germany
reza.hosseini@pmone.com

Marko Tkalčič
Free University of Bozen-Bolzano
Bolzano, Bozen, Italy
Marko.Tkalcic@unibz.it

ABSTRACT

In this paper, we explore the potential of using visual features in movie Recommender Systems. This type of content features can be extracted automatically without any human involvement and have been shown to be very effective in representing the visual content of movies. We have performed the following experiments, using a large dataset of movie trailers: (i) Experiment A: an exploratory analysis as an initial investigation on the data, and (ii) Experiment B: building a movie recommender based on the visual features and evaluating the performance. The observed results have shown promising potential of visual features in representing the movies and the excellency of recommendation based on these features.

ACM Reference Format:
Mohammad Hossein Rimaz, Mehdi Elahi, Farshad Bakhshandegan Moghadam, Christoph Trattner, Reza Hosseini, and Marko Tkalčič. 2019. Exploring the Power of Visual Features for Recommendation of Movies. In *27th Conference on User Modeling, Adaptation and Personalization (UMAP '19), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3320435.3320470

1 INTRODUCTION

Classic movie Recommender Systems, typically, rely on movie content features (such as genre and tags) and build recommendation models on top of this data [1, 28, 30, 44]. While the performance of these systems may depend largely on the performance of their core algorithms, the quality of the generated recommendations can still very much depend on the quantity and quality of the available content data [20, 29, 42, 50].

Furthermore, since the early works on content-based recommender systems, the focus of the research community has been mainly on the usage of semantic attributes. While semantic attributes are very informative of the movie content, they may still not necessarily represent entirely what a movie item visually represents. In particular, the actual users' interests could be related to several factors, that are more *visual* than *semantic* [19]. Indeed, it is known that a user's visual perception plays an important role in forming her taste on movies and this is largely dependent on the visual style of the movies. Semantic attributes are not well-capable of entirely capturing such visual characteristics, encoded by movie makers.

In this paper, we investigate the power of different visual features in representing movies' content in the context of the recommender systems. These visual features, in contrary to the semantic features, are extracted *automatically* with no need for any expensive human-annotation. We provide a comprehensive analysis, which includes an exploratory analysis of the visual features (Experiment A), as well as an evaluation of the quality of the recommendations based on visual features (Experiment B). In particular, we compare the quality of recommendations based on visual features when used *individually* or *combined* with the semantic attributes.

The experiments have been conducted by using the Movielens 1M dataset. The visual features were extracted from more than 1800 movie trailers. Prior work has reported the visual similarity between full-length movies and their respective trailers [12]. Our results show that recommendations based on visual features have a higher quality in comparison to the recommendations based on semantic attributes.

2 RELATED WORK

Multimedia recommender systems typically model content of the items based on two types of item features, referred to as *high-Level* features (attributes) and *low-Level* features [2, 3, 7, 33, 35]. Both types of features can be extracted from multimedia content automatically [13, 14, 16] or manually with the help of an expert or users of the system [37]. Recommender systems adopt implicit or explicit preferences of users on these high-level or low-level features in order to generate personalized recommendations for users [1, 9, 10]. As an example, in the music domain, the low-level features are acoustic properties, such as rhythm or timbre and they can be extracted and used to compute the similarity among music tracks [4, 5, 27] or classify music genre [31, 43].

UMAP '19, June 9–12, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6021-0/19/06…$15.00
https://doi.org/10.1145/3320435.3320470

Plenty of approaches have been already proposed in order to exploit both high-level and low-level content features (as side-information) in recommender systems. Majority of such approaches address the big challenge of recommender systems, i.e., *Cold Start* problem [15, 17, 18, 26, 34]. However, these approaches typically rely to the assumption that user preferences are mainly influenced by high-level features of movies (e.g., plot, genre, director, actors) and, to a lesser extent, by low-level features (as stylistic properties) [11].

Recent studies on recommender systems suggest that users' preferences are influenced more by the low-level features (representing the visual aspects of movies) and less by their high-level features (representing the semantic or syntactic properties) [12, 23, 32, 41]. Low-level features, such as color energy, shot duration, and lighting key [51] have a proved effect on user emotions [40]. Moreover, motions of objects, camera movements, and shot length duration are delicately planned by directors and cinematographers in order to influence the audience perception [24]. Many algorithms were introduced to solve the problem of extracting such features from image sequences [21, 47].

Overall, the usage of the low-level features has drawn minor attention in recommender systems' research field (e.g., in [32]), while it has been extensively investigated in the related fields such as computer vision and video retrieval [36, 45]. The works presented in [6, 25] provide comprehensive surveys on the relevant state-of-the-art techniques related to video content analysis and classification, and discuss a large body of low-level features (visual, auditory or textual) that can be considered for these purposes. In [39] Rasheed et al. proposes a framework for movie genre classification based only on commutable visual cues. In [38] Rasheed et al discuss a similar approach by considering also audio features. In [46] Svanera et al. propose a deep learning approach to automatically recognize the director of a movie based on low-level visual features. Finally, in [52] Zhou et al. propose a framework for the automatic movie genre classification, using temporally-structured features from movie trailers.

3 METHOD DESCRIPTION

We have built a *pure* Content-Based recommender system (CB), which relies solely on semantic item features attributes, i.e genre, tags, or visual features, as well as a similarity metric. The similarity metric is used to measure the similarity among items. Then a model is built, based on the user preferences, exploited to learn the taste of a target user and to recommend to her items that are similar to those that she liked in the past. We have used one of the most common similarity metrics, the *Cosine similarity*. As baselines, we used the genre and tag attributes.

Each model, built by using either of these (baseline) attributes or proposed visual features, uses its own similarity matrix. By computing the similarity matrix, we can predict the rating of a movie by looking at the *K-Nearest Neighbors* of an item. Hence, the recommender algorithm computes a score for all the items in the catalogue. In order to calculate the score for an item, the weighted average of the k-nearest neighbors of the item is used:

$$predict(u,i) = \frac{\sum_{j=x_1}^{x_k} similarity(i,x_j) * x_j}{\sum_{j=x_1}^{x_k} similarity(i,x_j)}$$

All results have been achieved by extending the *Surprise* Python Library[1]. However, we also attempted to double-checked the results with the *Hi-Rec* Java framework [2], as it is well-compatible with the mise-en-scène dataset, described in the next section.

4 EXPERIMENTAL RESULTS

4.1 Dataset

We have used the mise-en-scène dataset [19], which contains visual features for 13373 movies. The visual features are the followings:

- *f1:* Average shot length
- *f2:* Mean of color variance across the key Frames
- *f3:* Standard deviation of color variance across the key Frames
- *f4:* Mean of motion average across all the frames
- *f5:* Mean of motion standard deviation across all the frames
- *f6:* Mean of lighting key across the key frames
- *f7:* Number of shots

All of these features have been normalized by passing them through a "Natural Logarithm" function and then a "Quantile Normalization" scheme [3]. We combined these dataset with the MovieLens 1M [22] dataset. The final dataset contains 10920 unique tags, 19 unique genres from the MovieLens dataset, and low-level visual features from the mise-en-scène dataset. For the evaluation purpose we filtered out the ratings of the users that have less than 10 ratings with the values 4 or 5. An item that has received rating 4 or 5, from a user, is assumed to be a relevant item for that user. The final dataset contains 666713 ratings provided by 5690 users to 1828 movies.

4.2 Experiment A: Exploratory Analysis

4.2.1 **Evolution of Visual Features Over Time.** We have analyzed the time evolution of the visual features over a long period of cinema's history (from 1918 to 2000). Figure 1 illustrates the results for f_1 (average shot length). The figure shows that this visual feature has been constantly decreased over the time. Hence, the recent movies may have more camera changes in their scenes. We note again that the visual features are all normalized with the *log* function, and hence, even small changes could actually mean a lot.

Figure 2 shows the evolution of f_2 (mean of color variance) over the time. As it can be seen there is a slight positive trend in this figure, which indicates that the recent movies contain larger variation in their colors. We observed similar results for f_3 (Standard deviation of color variance) and hence did not include the figure.

Figure 3 represents the evolution of f_4 (mean of motion average). This feature represents the average motion within a movie. We have observed again slightly positive trend for this feature, which may represent that the objects and the cameras move faster in the recent movies. We have observed a similar result for f_5 (standard deviation of the motion average).

Figure 4 shows the time evolution of f_6 (mean of lighting key). For this feature, we have seen a slight drop over the considered

[1] http://surpriselib.com/
[2] https://fmoghaddam.github.io/Hi-Rec/
[3] https://www.researchgate.net/publication/305682388_Mise-en-Scene_Dataset_Stylistic_Visual_Features_of_Movie_Trailers_description

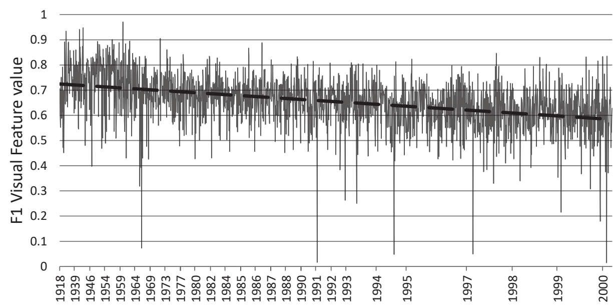

Figure 1: Time evolution of the *average shot length* (f_1) in movies, over the history of cinema. The values are *log* normalized.

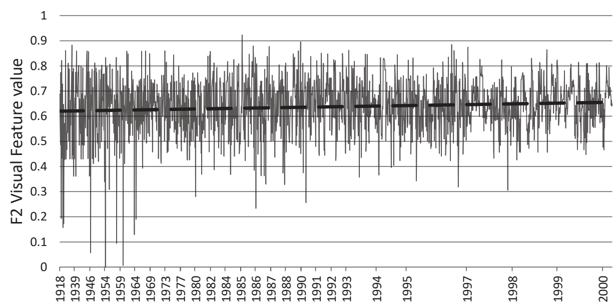

Figure 2: Time evolution of the *mean of color variance* (f_2), over history of cinema. The values are *log* normalized.

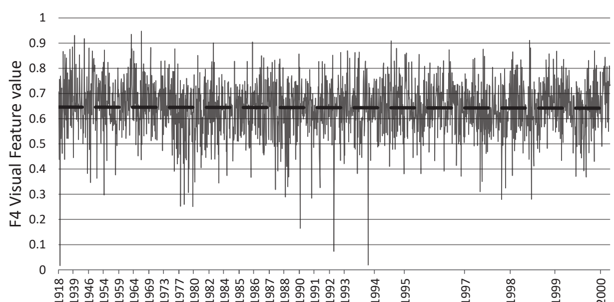

Figure 3: Time evolution of the *mean of motion average* (f_4) over the history of cinema. The values are *log* normalized.

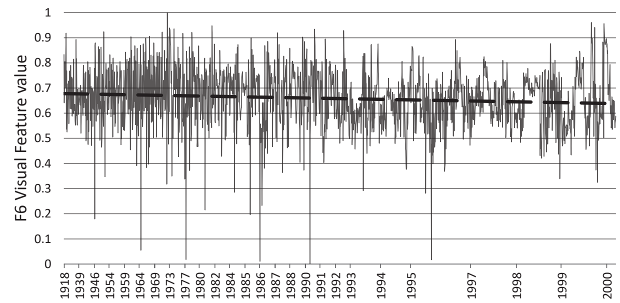

Figure 4: Time evolution of the *lighting key* (f_6) over the history of cinema. The values are *log* normalized.

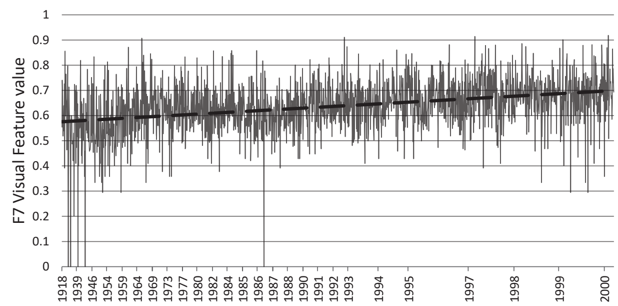

Figure 5: Time evolution of the *number of shots* (f_7) over the history of cinema. The values are *log* normalized.

period. This could mean that, in average, older movies were brighter and had less contrast than newer movies. This could be due to the quality of earlier cameras which were not capable of capturing very sharp movies or good quality in the dark conditions.

Finally, Figure 5 shows the evolution of feature f_7 (number of shots). As the figure shows, the number of shots in movies has been increasing over time. This means that recent movies contain larger numbers of shots compared to the older movies.

We believe that these insights can be exploited in recommender systems, but requires further investigation. As an example, in order to recommend more novel movies, the recommender can be fine-tuned to suggest more movies with a larger number of shots or shorter shot length.

4.2.2 **Clustering the Movies based on Visual Features.** We have adopted *K-Means* clustering in order to investigate the (visually similar) clusters that could exist among movies. In order to identify the number of clusters, we have performed the *Elbow Method*, which varies the number of cluster in a range of 1 to 70. The results are plotted in Figure 6. This figure shows that the right number of clusters could be around 30.

We have then checked the most popular movies within each cluster. Table 1 shows the most popular movies within 4 of these clusters (as some examples). We have noticed that some of the movies have certain kind of *semantic correlation*, in addition to their visual similarity. For instance, Mission Impossible (1996) and Saving Private Ryan (1998) are both movies about "a mission", full of action scenes. On the other hand, sometimes the movies within a cluster are more visually similar rather than semantically.

In addition, we checked the extreme cases where the movies has the *minimum* or *maximum* values of the visual features. In Table 2, we list these movies for some important visual features. For instance, the movie Psycho (1960) has the minimum variation of colors, Hold Back the Dawn (1941) has the maximum average shot length.

Table 1: Most popular movies within the example of 4 movie clusters based on visual features.

Example Cluster	#Ratings	Most Popular Movies	Director	IMDB Genre	IMDB Plot Keywords
Cluster 0	47048	Independence Day (1996)	Roland Emmerich	Action | Adventure | Sci-Fi	saving the world mission | 1990s
	44208	Dances with Wolves (1990)	Kevin Costner	Adventure | Drama | Western	friendship | 19th century
Cluster 1	37127	Mission: Impossible (1996)	Brian De Palma	Action | Adventure | Thriller	ethan hunt character | train
	37110	Saving Private Ryan (1998)	Steven Spielberg	Drama | War	rescue mission | world war two
Cluster 2	43295	Raiders of the Lost Ark (1981)	Steven Spielberg	Action | Adventure	indiana jones | egypt
	41426	Back to the Future (1985)	Robert Zemeckis	Adventure | Comedy | Sci-Fi	time machine | future
Cluster 3	54502	Star Wars: A New Hope (1977)	George Lucas	Action | Adventure | Fantasy	rebellion | empire
	52244	Terminator 2 (1991)	James Cameron	Action | Sci-Fi	time travel | liquid metal

Table 2: Movies with the *minimum* or *maximum* values for each of the visual features.

Feature Type	Ext	Value	Movie	Director	IMDB Genre	IMDB Plot Keywords
Average shot length	max	0.9183	Hold Back the Dawn (1941)	Mitchell Leisen	Drama | Romance	deception | mexico
	min	0.2904	Equilibrium (2002)	Kurt Wimmer	Action | Drama | Sci-Fi	dystopia | suppress of emotion
Color variation	max	0.8625	In the Line of Fire (1993)	Wolfgang Petersen	Action | Crime | Drama	interrupted sex | assassin
	min	0.0062	Psycho (1960)	Alfred Hitchcock	Horror | Mystery | Thriller	motel | shower
Motion Average	max	0.9476	The Graduate (1967)	Mike Nichols	Comedy | Drama | Romance	college graduate | love triangle
	min	0.1645	Miller's Crossing (1990)	Joel & Ethan Coen	Crime | Drama | Thriller	irish mob | organized crime
Lighting Key	max	0.8540	The Graduate (1967)	Mike Nichols	Comedy | Drama | Romance	college graduate | love triangle
	min	0.3976	The Dirty Dozen (1967)	Robert Aldrich	Action | Adventure | War	commando mission | criminal

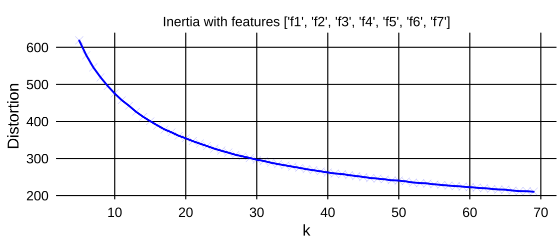

Figure 6: Elbow Method: Identifying the right number of clusters within the movies.

4.3 Experiment B: Recommendation

Table 3 presents the results of the experiment B where we evaluate the quality of content-based recommendations based on visual features. All the algorithms have been trained and tested on a 70%-30% data split.

In terms of the precision@10 metric, the best results have been achieved by the recommendation approach based on visual features. The precision value of visual features is 0.832 while this is 0.684 for genre, and 0.651 for tags. Using all features together results in a precision score of 0.666. This indicates that a simple merge of the features may not be the best technique for feature fusion and more advanced method should be exploited.

In terms of RMSE and MAE, the visual features have still obtained relatively good results. However, the results were not substantially better than the other baseline attributes. The RMSE and MAE of the visual features were 1.091 and 0.846. This values were 1.093 and 0.846 for genre, and were 1.137 and 0.870 for tags. Combining all features resulted in RMSE and MAE values of 1.153 and 0.876.

We have used a simple combination of visual features, but a more sophisticated feature fusion method could improve these results.

Table 3: Quality of movie recommendation, based on different content features, w.r.t, RMSE, MAE and Precision@10.

Recommender	RMSE	MAE	Precision@10
ContentBased:Visual	**1.091**	**0.846**	**0.832**
ContentBased:Genre	1.093	0.846	0.684
ContentBased:Tag	1.137	0.870	0.651
ContentBased:All	1.153	0.876	0.666
Distribution-based	1.491	1.195	0.611

The worst results for all metrics have been obtained by a non-personalized baseline (distribution based recommendation). These results show that by exploiting visual features, the content-based recommender algorithm outperforms all other baselines, in terms of both considered evaluation metrics.

5 CONCLUSION AND FEATURE WORK

In this paper, we have investigated the power of visual features for the movie recommendation task. We have performed a preliminary exploratory analysis and an experiment for evaluating the quality of recommendations based on visual features. The results were promising and indicative of the potential power of such features.

In the future, we will design a novel web application for real user studies and will evaluate the quality of recommendations based on visual features in an online scenario. We are interested in investigating the importance of user interface design, in developing a visually-aware movie recommender system [8, 16]. Finally, we plan to exploit a new tool [48, 49] that we have developed recently, which is capable of learning user preferences from their facial expressions. We will use this tool in order to study the potential correlation between users' facial expressions and the visual features within the movies.

REFERENCES

[1] Charu C Aggarwal. 2016. *Content-Based Recommender Systems*. Springer International Publishing, Chapter 4, 139–166. https://doi.org/10.1007/978-3-319-29659-3_4

[2] Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. 2007. Open User Profiles for Adaptive News Systems: Help or Harm?. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 11–20. https://doi.org/10.1145/1242572.1242575

[3] Daniel Billsus and Michael J. Pazzani. 2000. User Modeling for Adaptive News Access. *User Modeling and User-Adapted Interaction* 10, 2-3 (Feb. 2000), 147–180. https://doi.org/10.1023/A:1026501525781

[4] Dmitry Bogdanov and Perfecto Herrera. 2011. How Much Metadata Do We Need in Music Recommendation? A Subjective Evaluation Using Preference Sets.. In *International Society for Music Information Retrieval Conference (ISMIR)*. University of Miami, Miami, Florida, USA, 97–102.

[5] Dmitry Bogdanov, Joan Serra, Nicolas Wack, Perfecto Herrera, and Xavier Serra. 2011. Unifying Low-Level and High-Level Music Similarity Measures. *IEEE Transactions on Multimedia* 13, 4 (aug 2011), 687–701. https://doi.org/10.1109/tmm.2011.2125784

[6] D. Brezeale and D. J. Cook. 2008. Automatic Video Classification: A Survey of the Literature. *Trans. Sys. Man Cyber Part C* 38, 3 (May 2008), 416–430. https://doi.org/10.1109/TSMCC.2008.919173

[7] Iván Cantador, Martin Szomszor, Harith Alani, Miriam Fernandez, and Pablo Castells. 2008. Enriching ontological user profiles with tagging history for multi-domain recommendations. In *1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb 2008)*. http://oro.open.ac.uk/20015/

[8] Paolo Cremonesi, Mehdi Elahi, and Franca Garzotto. 2017. User interface patterns in recommendation-empowered content intensive multimedia applications. *Multimedia Tools and Applications* 76, 4 (01 Feb 2017), 5275–5309. https://doi.org/10.1007/s11042-016-3946-5

[9] Paolo Cremonesi, Franca Garzotto, and Maurizio Ferrari Dacrema. 2018. *User Preference Sources: Explicit vs. Implicit Feedback*. World Scientific.

[10] Marco de Gemmis, Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Giovanni Semeraro. 2015. *Semantics-Aware Content-Based Recommender Systems*. Springer US, Chapter 4, 119–159. https://doi.org/10.1007/978-1-4899-7637-6_4

[11] Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi, Franca Garzotto, and Pietro Piazzolla. 2016. Recommending Movies Based on Mise-en-Scène Design. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1540–1547. https://doi.org/10.1145/2851581.2892551

[12] Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi, Franca Garzotto, Pietro Piazzolla, and Massimo Quadrana. 2016. Content-Based Video Recommendation System Based on Stylistic Visual Features. *Journal on Data Semantics* 5, 2 (Jun 2016), 99–113. https://doi.org/10.1007/s13740-016-0060-9

[13] Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi, Farshad Bakhshandegan Moghaddam, and Andrea Luigi Edoardo Caielli. 2017. How to Combine Visual Features with Tags to Improve Movie Recommendation Accuracy?. In *E-Commerce and Web Technologies*, Derek Bridge and Heiner Stuckenschmidt (Eds.). Springer International Publishing, Cham, 34–45. https://doi.org/10.1007/978-3-319-53676-7_3

[14] Yashar Deldjoo, Mehdi Elahi, Massimo Quadrana, and Paolo Cremonesi. 2015. Toward Building a Content-Based Video Recommendation System Based on Low-Level Features. In *Lecture Notes in Business Information Processing*, Dietmar Stuckenschmidt, Heinerand Jannach (Ed.). Springer International Publishing, 45–56. https://doi.org/10.1007/978-3-319-27729-5_4

[15] 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* 7, 4 (01 Nov 2018), 207–219. https://doi.org/10.1007/s13735-018-0155-1

[16] Yashar Deldjoo, Mehdi Elahi, Massimo Quadrana, Paolo Cremonesi, and Franca Garzotto. 2015. Toward Effective Movie Recommendations Based on Mise-en-Scène Film Styles. In *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter (CHItaly 2015)*. ACM, New York, NY, USA, 162–165. https://doi.org/10.1145/2808435.2808460

[17] Mehdi Elahi. 2014. *Empirical Evaluation of Active Learning Strategies in Collaborative Filtering*. Ph.D. Dissertation. Free University of Bozen - Bolzano.

[18] Mehdi Elahi, Matthias Braunhofer, Tural Gurbanov, and Francesco Ricci. 2018. *User Preference Elicitation, Rating Sparsity and Cold Start*. Chapter Chapter 8, 253–294. https://doi.org/10.1142/9789813275355_0008 arXiv:https://www.worldscientific.com/doi/pdf/10.1142/9789813275355_0008

[19] Mehdi Elahi, Yashar Deldjoo, Farshad Bakhshandegan Moghaddam, Leonardo Cella, Stefano Cereda, and Paolo Cremonesi. 2017. Exploring the Semantic Gap for Movie Recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. ACM, New York, NY, USA, 326–330. https://doi.org/10.1145/3109859.3109908

[20] Mehdi Elahi, Francesco Ricci, and Neil Rubens. 2016. A Survey of Active Learning in Collaborative Filtering Recommender Systems. *Comput. Sci. Rev.* 20, C (May 2016), 29–50. https://doi.org/10.1016/j.cosrev.2016.05.002

[21] Ralph Ewerth, Martin Schwalb, Paul Tessmann, and Bernd Freisleben. 2004. Estimation of arbitrary camera motion in MPEG videos. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. IEEE. https://doi.org/10.1109/icpr.2004.1334181

[22] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* 5, 4, Article 19 (Dec. 2015), 19 pages. https://doi.org/10.1145/2827872

[23] Ruining He, Chen Fang, Zhaowen Wang, and Julian McAuley. 2016. Vista: A Visually, Socially, and Temporally-aware Model for Artistic Recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 309–316. https://doi.org/10.1145/2959100.2959152

[24] Timothy Heiderich. 2018. Cinematography techniques: The different types of shots in film. *Ontario Mining Assosiation. Recuperado de https://www. oma. on. ca/en/contestpages/resources/free-report-cinematography. pdf* (2018).

[25] Weiming Hu, Nianhua Xie, Li, Xianglin Zeng, and Stephen Maybank. 2011. A Survey on Visual Content-Based Video Indexing and Retrieval. *Trans. Sys. Man Cyber Part C* 41, 6 (Nov. 2011), 797–819. https://doi.org/10.1109/TSMCC.2011.2109710

[26] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. 2017. Visually-Aware Fashion Recommendation and Design with Generative Image Models. *CoRR* abs/1711.02231 (2017). arXiv:1711.02231 http://arxiv.org/abs/1711.02231

[27] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer. 2007. A music search engine built upon audio-based and web-based similarity measures. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07*. ACM Press. https://doi.org/10.1145/1277741.1277818

[28] Michael S Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain. 2006. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 2, 1 (2006), 1–19.

[29] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the Cold Start Problem in Recommender Systems. *Expert Syst. Appl.* 41, 4 (March 2014), 2065–2073. https://doi.org/10.1016/j.eswa.2013.09.005

[30] Pasquale Lops, Marco De Gemmis, Giovanni Semeraro, Cataldo Musto, and Fedelucio Narducci. 2013. Content-based and Collaborative Techniques for Tag Recommendation: An Empirical Evaluation. *J. Intell. Inf. Syst.* 40, 1 (Feb. 2013), 41–61. https://doi.org/10.1007/s10844-012-0215-6

[31] Martin F. McKinney and Jeroen Breebaart. 2003. Features for audio and music classification. In *ISMIR 2003, 4th International Conference on Music Information Retrieval, Baltimore, Maryland, USA, October 27-30, 2003, Proceedings*. Johns Hopkins University. http://ismir2003.ismir.net/papers/McKinney.PDF

[32] Pablo Messina, Vicente Dominquez, Denis Parra, Christoph Trattner, and Alvaro Soto. 2018. Exploring Content-based Artwork Recommendation with Metadata and Visual Features. *UMUAI* (2018).

[33] Stuart E. Middleton, Nigel R. Shadbolt, and David C. De Roure. 2004. Ontological User Profiling in Recommender Systems. *ACM Trans. Inf. Syst.* 22, 1 (Jan. 2004), 54–88. https://doi.org/10.1145/963770.963773

[34] Farshad Bakhshandegan Moghaddam and Mehdi Elahi. [n. d.]. *Cold Start Solutions For Recommendation Systems*. Chapter 2, book title: Big Data Recommender System.

[35] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Giovanni Semeraro, Marco de Gemmis, Mauro Barbieri, Jan Korst, Verus Pronk, and Ramon Clout. 2012. Enhanced Semantic TV-show Representation for Personalized Electronic Program Guides. In *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization (UMAP'12)*. Springer-Verlag, Berlin, Heidelberg, 188–199. https://doi.org/10.1007/978-3-642-31454-4_16

[36] H. R. Naphide and Thomas Huang. 2001. A probabilistic framework for semantic video indexing, filtering, and retrieval. *IEEE Transactions on Multimedia* 3, 1 (March 2001), 141–151. https://doi.org/10.1109/6046.909601

[37] Pandora. 2019. Music Genome Project. Retrieved January 11, 2019 from https://www.pandora.com/about/mgp

[38] Zeeshan Rasheed. 2003. *Video Categorization Using Semantics and Semiotics*. Ph.D. Dissertation. Orlando, FL, USA. AAI3110078.

[39] Zeeshan Rasheed, Yaser Sheikh, and Mubarak Shah. 2005. On the Use of Computable Features for Film Classification. *IEEE Trans. Cir. and Sys. for Video Technol.* 15, 1 (Jan. 2005), 52–64. https://doi.org/10.1109/TCSVT.2004.839993

[40] Ralph J. Roberts, Lisa D. Hager, and Christine Heron. 1994. Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General* 123, 4 (1994), 374–393. https://doi.org/10.1037/0096-3445.123.4.374

[41] Sujoy Roy and Sharath Chandra Guntuku. 2016. Latent Factor Representations for Cold-Start Video Recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 99–106. https://doi.org/10.1145/2959100.2959172

[42] Neil Rubens, Mehdi Elahi, Masashi Sugiyama, and Dain Kaplan. 2015. Active learning in recommender systems. In *Recommender systems handbook*. Springer, 809–846. https://doi.org/10.1007/978-1-4899-7637-6_24

[43] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* 7, 2 (01 Jun 2018), 95–116. https://doi.org/10.1007/s13735-018-0154-2

[44] Andriy Shepitsen, Jonathan Gemmell, Bamshad Mobasher, and Robin Burke. 2008. Personalized recommendation in social tagging systems using hierarchical clustering. In *Proceedings of the 2008 ACM conference on Recommender systems*. ACM, 259–266. https://doi.org/10.1145/1454008.1454048

[45] Cees G.M. Snoek and Marcel Worring. 2005. Multimodal Video Indexing: A Review of the State-of-the-art. *Multimedia Tools and Applications* 25, 1 (01 Jan 2005), 5–35. https://doi.org/10.1023/B:MTAP.0000046380.27575.a5

[46] Michele Svanera, Mattia Savardi, Alberto Signoroni, András Bálint Kovács, and Sergio Benini. 2018. Who is the director of this movie? Automatic style recognition based on shot features. *CoRR* abs/1807.09560 (2018). arXiv:1807.09560 http://arxiv.org/abs/1807.09560

[47] Yap-Peng Tan, D. D. Saur, S. R. Kulkami, and P. J. Ramadge. 2000. Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation. *IEEE Trans. Cir. and Sys. for Video Technol.* 10, 1 (Feb. 2000), 133–146. https://doi.org/10.1109/76.825867

[48] Marko Tkalčič, Nima Maleki, Matevž Pesek, Mehdi Elahi, Francesco Ricci, and Matija Marolt. 2017. A Research Tool for User Preferences Elicitation with Facial Expressions. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. ACM, New York, NY, USA, 353–354. https://doi.org/10.1145/3109859.3109978

[49] Marko Tkalčič, Nima Maleki, Matevž Pesek, Mehdi Elahi, Francesco Ricci, and Matija Marolt. 2019. Prediction of Music Pairwise Preferences from Facial Expressions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. ACM, New York, NY, USA, 150–159. https://doi.org/10.1145/3301275.3302266

[50] Michail Vlachos, Celestine Duenner, Reinhard Heckel, Vassilios G Vassiliadis, Thomas Parnell, and Kubilay Atasu. 2018. Addressing Interpretability and Cold-Start in Matrix Factorization for Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering* (2018). https://doi.org/10.1109/TKDE.2018.2829521

[51] Hee Lin Wang and Loong-Fah Cheong. 2006. Affective understanding in film. *IEEE Transactions on Circuits and Systems for Video Technology* 16, 6 (jun 2006), 689–704. https://doi.org/10.1109/tcsvt.2006.873781

[52] Howard Zhou, Tucker Hermans, Asmita V. Karandikar, and James M. Rehg. 2010. Movie Genre Classification via Scene Categorization. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 747–750. https://doi.org/10.1145/1873951.1874068