Enhancing the NPO Start Recommendation System with Metadata

submitted in partial fulfillment for the degree of master of science

Eileen Kapel
12369462

master information studies
data science
faculty of science
university of amsterdam

YYYY-MM-DD

	External Supervisor	External Supervisor	Second External Supervisor
Title, Name	Dr Maarten Marx	Arno van Rijswijk	Robbert van Waardhuizen
Affiliation	UvA, FNWI, IvI	NPO	NPO
Email	maartenmarx@uva.nl	arno.van.rijswijk@npo.nl	robbert.van.waardhuizen@npo.nl

Universiteit van Amsterdam

Amsterdam Data Science

nederlandse publieke omroep

ABSTRACT

KEYWORDS

Recommendation systems; Video recommendation systems; Metadata

1 INTRODUCTION

Humans are tasked with making thousands of decisions daily that may range from selecting what outfit to wear to which television series to watch. Recommendations can make the process of these decisions more efficient by sorting through potentially relevant information and making recommendations customised to the individual user [7, 10]. One system that employs recommendations is NPO Start[1]. It is a video-on-demand service from the NPO (Dutch Public Service Media) that gives people the ability to watch series, movies and documentaries as well as watch live television online. Personalised recommendations are available for registered users of the service, which is based on the collaborative filtering method. However, there is a lot of metadata available about the offered content that is unused in this current method. In this thesis, the metadata of broadcasts will be utilised to determine if it can improve the performance of the current video recommendation system.

This is achieved by answering the main research question: *'Can a hybrid recommendation system using metadata perform better than the current recommendation system of NPO Start which uses collaborative filtering?'*. The research is split into three sub-questions:

RQ.1 What is the performance of the current recommendation system?

RQ.2 Can the performance of the current recommendation system be improved by implementing a hybrid recommendation system?

RQ.3 Which metadata features improve the performance of the hybrid recommendation system the most?

The thesis is structured as follows: first, various related works of literature are outlined in section 2 that relate to the goal of the thesis. The methodology employed during the research is discussed in section 3 and is followed by its results in section 4. Subsequently, the conclusions of these results are presented in section 5 after which a discussion of the choices and possible future work follows in section 6.

2 RELATED LITERATURE

Recommendations are based on ratings that are explicitly given by users or ratings that are implicitly inferred from users their actions [10], like a click or a certain watch duration. There are three main approaches for building a system that gives out recommendations. The first approach utilises information about items for a content-based system. The second approach utilises user interaction information with items for a collaborative filtering system. Lastly, there is a hybrid recommendation system, which is a combination of the two previous approaches, that exploits item information and user interaction information to provide recommendations. The first two approaches each have their own shortcomings, like overspecialisation, rating sparsity and cold-start [1, 6], that hybrid systems aim to overcome to provide more accurate recommendations.

In this section, the approach of the current recommendation system of NPO Start is outlined. This is followed by an overview of the current state of hybrid recommendation research. Furthermore, a few personalised services that employ recommendation systems are described and, lastly, the role of metadata in recommendation systems is touched upon.

2.1 The NPO Start Recommendation System

NPO Start is a service that offers users the ability to watch video content on demand. This video content is displayed to users in so-called “ribbons“ or rows that have a certain theme, like ‘Populair‘, ‘Nieuw‘ and ‘Aanbevolen voor jou‘. Each ribbon consists of a ranked list of several items and an item represents a series that can be streamed.

Users of the service that have an account have the ability to receive several personalised ribbons that contain items that are recommended to a specific user. These recommendations are materialised on the front page of the service in the two ribbons ‘Aanbevolen voor jou‘ en ‘Probeer ook eens‘. This thesis focuses on the ‘Aanbevolen voor jou‘ ribbon. The recommendations for this ribbon are produced by a collaborative filtering approach that utilises the history of user interaction information with items. These user interactions are grouped on series level and evaluated by pairs of series that are frequently watched together, or coincide often, with the history of the user. Of these coincidences, the top 100 pairs are extracted after they are ordered based on their frequency. A sliding window technique of the past 21 days is used for the interaction information, which is based on the intuition that recent events are more relevant than older ones [1]. These recommendation lists are updated hourly at peak times and bi-hourly off-peak.

2.2 Hybrid Recommendation Systems

Hybrid recommendation systems are able to deal with scenarios where there is little data on new users and items, also called the cold-start problem, or when there is sparse interaction data available. This is achieved by joining content and collaborative data in the system to produce recommendations.

Different approaches exist for building a hybrid recommendation system, however most employ matrix factorisation for the collaborative filtering part of the system. This approach was popularised by the solution of the Netflix Prize competition that employed matrix factorisation using the alternating least square (ALS) algorithm [2, 4, 5]. Some works that have successfully employed matrix factorisation for their hybrid recommendation systems are Rubtsov et al. [11], Ludewig et al. [8] and Al-Ghossein et al. [1]. Rubtsov et al. made use of the LightFM library, which is a Python implementation of a matrix factorisation model that can deal with user and item information [6], paired with a weighted approximate-rank pairwise loss for their approach in the ACM Recommender

[1] www.npostart.nl

Systems Challenge 2018. The challenge was focused on automatic playlist continuation and Rubtsov et al. their hybrid recommendation system was rewarded first place. Ludewig et al. made use of matrix factorisation in their model by combining it with a k-nearest-neighbour technique and using the ALS algorithm. Content was incorporated into the model by weighing the matrix factorisation results with the IDF (inverse document frequency) score of titles to produce the final list of recommendations. Lastly, the hybrid recommendation system by Al-Ghossein et al. merged matrix factorisation with topics extracted using topic modeling for online recommendation.

Furthermore, neural networks have also been combined into recommendation systems that employ hybrid matrix factorisation, due to its recent popularity. Volkovs et al. [13] produced a two stage model for automatic playlist continuation that first employs weighted regularized matrix factorisation to retrieve a subset of candidates and then uses convolutional neural networks and neighbour-based models for detecting similar patterns. In the second stage features of playlists and songs are combined with the items after which the final ranking of recommendations is produced.

2.3 Personalised services

Recommendation systems are frequently employed by services to provide a personalised experience to their users. This occurs in different domains, e.g. product recommendation by Amazon, music recommendation by Spotify and video recommendation by YouTube and Netflix.

Netflix is a special case where the whole experience of its service is defined by recommendation systems [3]. This is primarily showcased on its homepage which consists of rows of recommended videos with a similar theme. This is similar to the ribbons of NPO Start however NPO Start only employs recommended videos in a few ribbons. A few of the row themes are ‘Top Picks‘, ‘Trending Now‘ and genre rows, like suspenseful movies or action & adventures. Each row theme uses different data, signals and models to produce recommended videos that are ranked using a personalised video ranker. On top of that, pages are also personalised by selecting and ordering rows that are relevant and diverse for each user. Aside from the homepage, personalisation is also offered while using search for which play data, search data and metadata is used. Netflix uses topic modeling, matrix factorisation and probabilistic graphical models in its recommendation system. A/B testing is used to improve algorithms to enhance member retention and offline experiments are used for faster innovation.

The streaming service Spotify also employs recommendations but for automatic playlist continuation, which is a task that adds one or more tracks to a playlist that fits the original playlist [14]. Spotify organised the ACM Recommender Systems Challenge 2018 with the purpose of improving this task. The most frequent approaches by teams that participated in the challenge consist of two-stage architectures, matrix factorisation, neural network models and learning to rank models.

2.4 Metadata

Multimedia content, like songs, films and series, are often described using a set of metadata elements. Metadata is information that describes the main attributes of an item [7]. Multimedia content datasets, like the MovieLens and Netflix dataset, generally contain metadata in the form of a title, plot, genre, actors, directors and a release year.

Metadata is often used in content-based recommendation systems and thus also in hybrid recommendation systems. However, good performance of such systems is predicated on the availability of high-quality metadata [6, 11]. This is evident in the work of Soares & Viana [12] where the version of their recommendation system that used more granular metadata, e.g. genres and sub-genres, resulted in recommendations of a higher quality. If high-quality metadata is not available then good quality metadata can be obtained from item descriptions, like actor lists and synopses [6]. However, metadata of a lower quality, e.g. by being sparse, may result in overfitting and causes models to not make use of content in an effective way [14].

Feature selection is frequently used to improve the quality of metadata. It is a method where rather than using all the features only a subset of the features is used [9]. By carefully selecting this subset of features a better effectiveness of the system can be achieved. For example, the hybrid recommendation system by Soar et al. that only employed the director as feature, instead of all the features, resulted in recommendations that were more precise [12]. A single feature can also be made more precise by taking advantage of mutual information [9], e.g. using frequent words or information retrieval methods like TF (term frequency), DF (document frequency), and TF-IDF (term frequency-inverse document frequency). This was employed by the hybrid recommendation system of Rubtsov et al. which used the top-2000 most frequent words in titles as a feature in their recommendation system opposed to all the words [11].

3 METHODOLOGY

3.1 Description of the Data

3.1.1 User Interaction Information.

3.1.2 Content Features.

3.1.3 Feature Encoding.

3.2 Methods

3.2.1 The LightFM Model.

3.2.2 The Experimental Setup.

3.3 Evaluation

3.3.1 Mean Precision@k.

4 RESULTS

4.1 RQ1

4.2 RQ2

4.3 RQ3

5 CONCLUSIONS

6 DISCUSSION

ACKNOWLEDGMENTS

Acknowledgements

REFERENCES

[1] Marie Al-Ghossein, Pierre-Alexandre Murena, Talel Abdessalem, Anthony Barré, and Antoine Cornuéjols. 2018. Adaptive collaborative topic modeling for online recommendation. *Proceedings of the 12th ACM Conference on Recommender Systems* (2018), 338–346.

[2] Robert Bell, Yehuda Koren, and Chris Volinsky. 2007. Modeling relationships at multiple scales to improve accuracy of large recommender systems. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 95–104.

[3] Carlos A Gomez-Uribe and Neil Hunt. 2016. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6, 4 (2016), 13.

[4] Yehuda Koren. 2009. The bellkor solution to the netflix grand prize. *Netflix prize documentation* 81, 2009 (2009), 1–10.

[5] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.

[6] Maciej Kula. 2015. Metadata embeddings for user and item cold-start recommendations. *arXiv preprint arXiv:1507.08439* (2015).

[7] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2011. Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*. Springer, 73–105.

[8] Malte Ludewig, Iman Kamehkhosh, Nick Landia, and Dietmar Jannach. 2018. Effective Nearest-Neighbor Music Recommendations. In *Proceedings of the ACM Recommender Systems Challenge 2018*. ACM, 3.

[9] Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. 2010. Introduction to information retrieval. *Natural Language Engineering* 16, 1 (2010), 100–103.

[10] Michael J Pazzani. 1999. A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review* 13, 5-6 (1999), 393–408.

[11] Vasiliy Rubtsov, Mikhail Kamenshchikov, Ilya Valyaev, Vasiliy Leksin, and Dmitry I Ignatov. 2018. A hybrid two-stage recommender system for automatic playlist continuation. In *Proceedings of the ACM Recommender Systems Challenge 2018*. ACM, 16.

[12] Márcio Soares and Paula Viana. 2015. Tuning metadata for better movie content-based recommendation systems. *Multimedia Tools and Applications* 74, 17 (2015), 7015–7036.

[13] Maksims Volkovs, Himanshu Rai, Zhaoyue Cheng, Ga Wu, Yichao Lu, and Scott Sanner. 2018. Two-stage model for automatic playlist continuation at scale. In *Proceedings of the ACM Recommender Systems Challenge 2018*. ACM, 9.

[14] Hamed Zamani, Markus Schedl, Paul Lamere, and Ching-Wei Chen. 2018. An Analysis of Approaches Taken in the ACM RecSys Challenge 2018 for Automatic Music Playlist Continuation. *Proceedings of the 12th ACM Conference on Recommender Systems* (2018), 527–528.

A CONTENT FEATURES

Table 1: Overview of the Content Features

Feature	Type	Description
broadcaster	categorical	Broadcaster of the broadcast, e.g. NOS.
credits	list	The people accredited in the broadcast, such as presenters or guests.
description	string	Description of the broadcast. This is either the main description, otherwise the short description or the kicker.
genres	list	Genres of the broadcast denoted by a genre id and name, e.g. (3.0.1.6, [Amusement]).
mid	string	Unique media identifier.
series reference	string	A reference to the series of which the item is part of.
subtitles	string	The subtitles of the broadcast, which were extracted using the POMS subtitles API.
title	string	The main title of the broadcast.