

Topical Video-On-Demand Recommendations based on Event Detection

Tobias Dörsch, Andreas Lommatzsch, and Christian Rakow

DAI-Labor, TU Berlin, Ernst-Reuter-Platz 7, D-10587 Berlin, Germany
{Tobias.Doersch,Andreas.Lommatzsch,Christian.Rakow}@dai-labor.de

Abstract. Recommender systems help users to discover relevant items. Traditionally, recommender systems rely on both detailed knowledge of the domain and an extensive user profile. However, small numbers of users, privacy concerns, or a very specific domain limit access or availability to this information. In this work, we present an approach for recommending items based on events relevant to the target group of our system. We exemplify the approach with the aid of a Video-On-Demand platform specialized in independent and art-house movies. Our recommender analyzes domain-specific blogs and news. It extracts current events that can be used for triggering topical recommendations. We show that our approach successfully identifies relevant events and provides highly relevant results without requiring detailed user profiles.

Keywords: recommender, event detection, privacy preserving recommender, Linked Open Data, video on demand

1 Introduction

The rapidly growing amount of items in online shops and entertainment services make it very hard for users to find relevant items. Recommender systems have been developed for supporting users to discover items potentially unknown and matching the user preferences. A widely used approach is user-based collaborative filtering computing the similarity between users and suggesting items that users with similar interests liked. A weakness of collaborative filtering is that current trends and temporal aspects are not taken into account. In many scenarios the context and seasonality have a high influence on the user preferences. Traditionally, experts (e.g. “curators”) compile topical recommendations in video shops or libraries taking into account new releases, trends as well as current events. This motivates us to develop a recommender that scans several different news streams, detects relevant events, and uses this information for computing recommendations.

Video-on-Demand (VoD) systems allow users to watch almost any movie at any time. The challenge for a VoD recommender system is not only identifying items

matching the user preferences but also computing when to recommend an item. In the past, curators knowing the typical seasonal user preferences and the relevant events (awards ceremonies, holidays, etc.) created a schedule when to broadcast a movie. We bring this principle to the VoD recommender. The recommender determines events and trends relevant to a specific target group. Based on these events we compute topical recommendations, which can be weighted by individual preferences. The event-based recommendations are often helpful for escaping the filter bubble and for suggesting items related to current trends.

We develop a recommender system for a VoD service focused on independent and art-house movies. In contrast to main stream VoD services, our portal does not offer blockbuster movies but a carefully selected catalog of films tailored to the needs of a niche market. A remarkable fraction of the offered films are documentaries and films related to current political topics. The requirements in the scenario are providing new relevant recommendations every day without relying on user profiles. We build our system on the idea that recognizing events relevant to our target group is a valuable basis for recommending relevant movies.

The identification of events suitable for recommending items leads to several challenges. To extract events, suitable sources must be identified and appropriate ways of processing and storing the contained information must be developed. This task requires learning algorithms able to identify events in streams of news data suitable for recommending items (films). Dependent from the different events types, adequate methods are needed for the event recognition and for linking events and films. In addition, explanations for the suggested items should be provided for improving the trust in the suggestions since recommendations based on news events are still unfamiliar for most users.

The remaining paper is structured as follows. Sec. 2 summarizes related work and discusses the connection to relevant research domains. Our approach is presented in Sec. 3. In Sec. 4 we evaluate our approach and discuss the strengths and weaknesses of our approach. Finally, a conclusion and an outlook to future work are given in Sec. 5.

2 Related Work

The task of recommending films on a daily basis is related to different domains.

CF-based Recommender Most movie recommender system focus on collaborative filtering (CF) [4]. CF-based approaches analyze the ratings users assign to items. The predictions are calculated by computing the similarity between either users (“user-based CF”) or the similarity between items (“item-based CF”). A requirement for getting high-quality recommendations is that a sufficient number of ratings for every user and every item are available. Well-known problems of CF-based approaches are the popularity bias [6] and the cold start problem [1]. CF-based algorithms tend to suggest popular, often already known items.

Context and Event Detection Beside the individual user preferences several different aspects influence the perceived relevance of movies, e.g. seasonality or the relation

to events. Studies analyzing the messages in social networks show that holidays and recent events have a high impact on the discussed topics [2]. The detection of events and the aggregation of messages related to the events are research topics in the analysis of social networks and news streams. Hennig et al. [3] applied clustering algorithms to news streams for identifying events in the news. The focus of the work lies on extracting and tracking topics but not on recommending items. Macedo et al. [5] developed a system that recommends social events. Based on the analysis of the user's past behavior, the proposed system recommends events based on social distance, and both location and time preferences.

Discussion Contexts and events have a high impact on the interest of users. Hence, building recommender systems computing recommendations based on relevant events is a promising approach helping users to escape the filter bubble and to find items related to the current topics of interest.

3 Approach

We develop a recommender system implementing a 4-layered architecture. The first layer collects news from heterogeneous sources. The second layer aggregates the collected data and extracts potentially relevant events. In addition, semantic data collections are integrated in order to consider expert-defined events (such as birthdays or memorial days). The third layer computes recommendations based on the events relevant to the target group. In the 4th layer, the recommendations are enriched and optimized for presentation. Explanations are generated for improving the trust in the relevance of the recommendations. The architecture of the system is visualized in Fig. 1. In the next paragraphs, we explain the implemented components in detail.

3.1 Collecting Data for Detecting Events

The crawlers continuously collect data being the basis for the identification of events. In order to focus on the events relevant to the target group, we carefully select the sources. In our scenario, we are especially interested in the domains art house, festivals, and documentaries. We analyze the RSS feeds of portals reporting on the domains. In addition, we crawl the TWITTER messages of an expert-defined set of accounts (using the TWITTER streaming API). In addition, we collect tweets from the major news portals for tracking the most relevant topics in the domain of politics. The selection of sources grants us access to up-to-date knowledge from domain experts. These experts typically write about the most relevant events and current trends. In our system we monitor ≈ 800 TWITTER accounts and ≈ 15 RSS feeds.

3.2 Recognition of Relevant Events

We consider two types of events. "Static" events such as birthdays, anniversaries, and memorial days are imported from semantic data collections. "Dynamic" events such as won awards, politic events, or the death of a director are detected in the news streams.

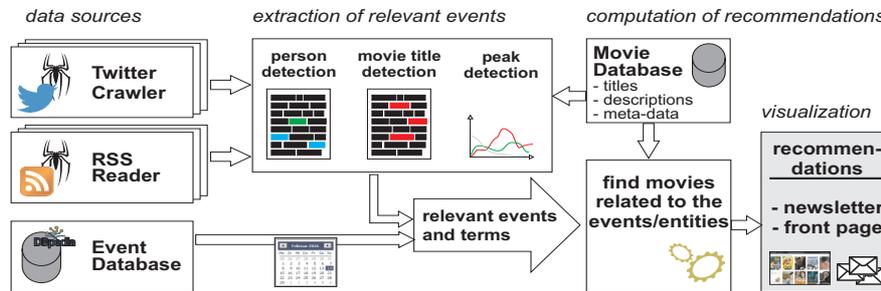


Fig. 1. The main components of our recommender system.

Knowledge Source for Events The “static” events are separated into two groups. The first group is formed by person- and movie-related events. The second group is built by holidays and memorial days typically related to specific keywords and genres.

Relevant Persons: Based on the movie catalog we know the persons related to the potentially relevant movies. We link the persons with DBPEDIA in order to collect all available birthdays of persons related to the movies. The same procedure is done for movie release days and awards won by the movie. The challenge in the task is the ambiguity of names and titles. We address this issue by computing a matching score taking into account context data. We only connect persons with DBPEDIA if the confidence score is above a threshold in order to prevent false positive matches. The score is calculated using the attributes of the entities (occupation, age, synopsis). User feedback is incorporated in order to correct and extend the automatically created links.

Relevant Holidays: In contrast to persons directly listed in the meta-data describing movies, the relations between holidays and movies are computed based on the textual description of the holidays. For this purpose we search the name of the holiday in the movie description and compute the textual similarity between the descriptions of the holidays (retrieved from DBPEDIA) and the synopsis of the movie. If the relatedness is above a threshold (optimized on a training dataset) the movie is linked to the holiday.

Discussion: Aggregating the different types of events, we find on average of about 20 events for each day of the year. This number of potentially relevant events allows us to filter out the most relevant events taking into account feedback from users and experts. In addition, the number of potentially relevant events allows us to ensure the diversity of events (e.g. with respect to actors, directors, composers as well as birthday, anniversaries). Static events are related for a specific day and typically recalled on the date of occurrence. However, some users may still be interested in the event a few days earlier or later (e.g. if they do not use the portal during the week). In order to make these recommendations available to those users, the relevance of static events degrades slowly over the course of 5 days.

Identifying Events in Tweets and RSS Feeds For recognizing events in news streams, we analyze how often a relevant person (listed in the movie catalog) is mentioned in

the news or tweets on a daily basis. An event is detected if a person is much more frequently mentioned than during an “average” day. Due to the large differences in popularity of movies and people, we implemented a 3-dimensional model: The popularity of a topic is identified by a long- and a short-term change in mentions as well as by the number of sources in that the topic is recognized. Since we do not compare topics against each other, each topic must fulfill criteria that are specific to its own time series. This leads to higher diversity in the recommended movies as well as to a broad spectrum of movies. In general, trend-detection for popular persons works more reliably than for unpopular persons. This is due to the fact that an increase of mentions of a popular topic is larger and thus easier to separate from noise.

Discussion The detection of the events and the linking of events with entities is the central component of the recommender. The recognition of static events is computed in advance when new movies are added to the catalog. The linking of dynamic events is done on a daily basis. The process is based on several text mining and similarity computations. A regular desktop computer suffices to complete the computations within minutes as the catalog is of limited size.

3.3 Computing Recommendations

Even though our database contains a large number of events, each type of event should optimally trigger its own set of recommended movies. On the other hand, some topics and events are very well connected in the movie database. In some cases, this leads to a number of recommended movies too large for a recommendation set. In other cases, too few movies are available to fill a type-specific set (i.e. recommendations based on birthdays). In those cases, our system mixes sets together based on the types’ similarity (e.g. birthdays and days of death) to achieve a suitable size for a single set. If we have too many candidates for a set (e.g. 20 birthdays on the same day), we lower the number of allowed movies per event or select the set of trigger events randomly for each user.

3.4 Presentation of results

In our VoD scenario, we present the recommendations computed based on the recognized topics on the front page of the VoD portal and in daily newsletters to registered users.

The topical recommendation of the front page: The landing page of the VoD service presents the sets of recommended movies. A header shows the type of event used for this set and creates the topical connection between each movie. To initially spike the users interest in a recommended movie, the trigger event is presented together with a short description of the event and, if available, context information on the related topic.

Theme-focused Newsletters: The VoD service already sends out a daily newsletter to registered users. This newsletter contains a set of 5 movies that have a deep topical connection. This connection is represented by a motto, for example, “Directors Inspired by Quentin Tarantino” or “Dream of a Better Life”. In order to build newsletter

automatically, we compute the most relevant event of the day and compute related movies. In the next step we try to fill templates created for the newsletters, such as “*Today is the birthday of <X>. His best movies here on <name of the portal>*”.

Discussion The implemented system is based on components and can be easily extended by integrating additional sources or by integrating components tailored to new types of events. The service interface allows the integration in existing recommender system.

4 Evaluation

We analyze the recommendations computed by our system. First, we evaluate the recommendations depending on the type of event used to trigger them. Secondly, we analyze the relevance of the recommendations and the acceptance of the suggested movies. The relevance of recommendations is analyzed based on the web server log of the VoD platform (currently only using content-based item-to-item recommendations) as well as feedback from experts (curators working for the VoD portal).

Recommendations based on News The system looks for trending topics by analyzing search log of the VoD system. Tab. 1 shows an overview of the number of events the recommender extracted based on trending topics for 9 days in January 2015. Analyzing detected events reveals that several trends in news feeds are correlated with “static” events. After the death of a popular actor, the time-series of mentions of that actor oftentimes spikes to an all-time high indicating that people are generally interested in this type of event. For other events, for example festivals and awards ceremonies, the number of mentions of the event increases during certain points in time. Awards ceremonies are often mentioned shortly after nominees for prizes are announced, during the award ceremony itself, and for a short period of time after the event, then oftentimes together winners of awards. Overall 20% of detected trends can be linked to “static” events. However, due to the large number of stored events, only 5% of events are covered by trends. Presenting trends together with a recommended movie proves to be difficult, if no knowledge is available on what event triggered the trend. Test users of our web application reacted positively to the most-popular approach, e.g. presenting the tweet with the highest “favorite” count.

Recommendations based on Static Events **Holidays and memorial days** days such as Veterans Day or Mother’s Day are linked to movies based on the similarity between the description of the holiday and the list of assigned term (retrieved from DBPEDIA) and the movie description. In our scenario, we considered 706 holidays related to at least one of the movies in our catalog. Analyzing the impact of these days on the user behavior, we observed a high variance for country-specific holidays: Displaying the trigger event together with recommended movie caused users to click 8 times more often. Confessional holidays increased clicks only by a factor of 1.7, while international holidays and memorial days increased clicks by a factor of 2.4.

Our **birthday** database contains 2,570 entries listing on average 7.02 birthdays per day. These events are related to 3,291 distinct movies. Tab. 1 shows the statistic of

relevant birthdays for the first days in January 2015. In order to evaluate the relevance of computed recommendations we check the recommendations against the spikes in the web server log. We found that 23% of recommendations derived from birthdays could be recognized by an increased movie related activity in the log file.

Table 1. The statistic of events recognized in news and semantic data.

date	number of relevant trends	number of related films	example person related to one of the trending topics	number of films related to the example person	number of relevant birthdays	number of related films	example person whose birthday this is	number of films related to this person
Jan 1, 2015	0	0	-	0	9	9	Snitz Edwards	1
Jan 2, 2015	2	2	Christian Bale	1	4	6	Lloyd Whitlock	2
Jan 3, 2015	3	4	Van Johnson	2	8	8	Thomas Morris	1
Jan 4, 2015	1	1	Forest Whitaker	1	3	6	August Diehl	4
Jan 5, 2015	6	8	Jessica Chastain	2	7	7	Shea Whigham	1
Jan 6, 2015	5	11	Ethan Hawke	4	13	14	Eddie Redmayne	2
Jan 7, 2015	4	4	Charlie Parker	1	5	6	Nicolas Cage	2
Jan 8, 2015	5	5	Marcus Vetter	1	6	7	Sarah Polley	2
Jan 9, 2015	8	13	Jodie Whittaker	3	11	43	Harun Farocki	32

Death-Days: Compare to birthdays, our database provides a significant smaller number of dates of death. The dataset contains 581 entries covering 286 days of the year. These events are related to 729 distinct movies. Similar to the birthday recommender, most death-days are not recognizable as peaks in the web server log. The death of a person (detected in the news stream) results in an increased user interest. The dates of death retrieved from DBPEDIA relate to dates several years in the past. This explains the different impact of death dates retrieved from the semantic database from death dates detected in the news.

Discussion We showed that our approach allows us to provide useful recommendations without having access to user profiles. The impact of the recommendations depends on the type of the identified event. In general, the relevance of dynamically recognized events (death of an actor, an award won by an actor) is more relevant than “static” events retrieved from a knowledge data base. “Big” birthdays of popular persons are more relevant than “usual” birthdays. Nevertheless “static” events are valuable since these events ensure that we reliably provide a fixed number of recommendations and diversify the result set.

Recommending movies based on events is often unexpected to users. In our discussions with users and the experts from the VoD portal we got positive feedback for the approach. In order to accept the recommendations it is important that users know or at least are interested in the events because the relevance of the events is crucial for the acceptance of the movie recommendations. On the other hand, our approach helps users to discover new content by recommending items based on events users usually would not be aware of.

5 Conclusions and Future Work

In this paper we present our system providing topical film recommendations based on different types of events. We discussed how to detect potentially relevant events from news and social media streams as well as the integration of semantic knowledge sources. In our analysis we found that birthdays of artists have only a very small influence on the user behavior. Events detected in news streams are better suited for recommending movies.

In contrast to traditional CF-based approaches, the developed approach helps users to discover new films. The relevance of suggestions is based on the similarity with current events instead of the similarity with entries in the user profile. We currently work on two personalization approaches. We combine the relevance scores computed using collaborative filtering ensuring that the identified events are matching the individual user preferences. In addition, we plan to allow users to add own sources (RSS feeds). This ensures that the news streams providing the basis for recognizing the relevant event meet the user needs.

Furthermore, we work on combining Named Entity Recognition algorithms and similarity computations based on semantic graphs for detecting movies related to current news. The weighted aggregation of several different relevance measures ensures a higher significance of recommended movies and provides the basis for more detailed explanations. The presented concept for recommending movies can be easily adapted for many additional scenarios, such as online shops. The use of the recent news (or weather data) is a promising new paradigm providing relevant recommendations without requiring detailed (sensitive) user profiles. A careful selection of sources analyzed for detecting events ensures that the recommendations are relevant for specific target groups. Based on the feedback we received for the implemented prototype there is a high potential in this approach.

Acknowledgments The work has been partially done in the EeGoF project supported by the German Federal Ministry for Economic Affairs and Energy. The research leading to these results was performed in the CrowdRec project, which has received funding from the EU 7th Framework Programme FP7/2007-2013 under grant agreement No. 610594.

References

1. J. Bobadilla, F. Ortega, A. Hernando, and J. Bernal. A collaborative filtering approach to mitigate the new user cold start problem. *Know.-Based Syst.*, 26:225–238, Feb. 2012.
2. W. Gao and F. Sebastiani. From classification to quantification in tweet sentiment analysis. *Social Network Analysis and Mining*, 6(1):1–22, 2016.
3. L. Hennig, D. Ploch, D. Prawdzik, B. Armbruster, H. Düwiger, E. W. De Luca, and S. Albayrak. SPIGA - multilingual news aggregator. *Procs. of GSCL 2011*, 2011.
4. J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst. (TOIS)*, 22(1):5–53, 2004.
5. A. Q. Macedo, L. B. Marinho, and R. L. Santos. Context-aware event recommendation in event-based social networks. In *Procs. of the 9th ACM RecSys Conf.*, NY, USA, 2015. ACM.
6. H. Steck. Item popularity and recommendation accuracy. In *Procs. of the 5th ACM Conf. on Recommender Systems*, pages 125–132, New York, NY, USA, 2011. ACM.