Tell Me What You Want:
Embedding Narratives for Movie Recommendations

Lukas Eberhard
Graz University of Technology
Graz, Austria
lukas.eberhard@tugraz.at

Simon Walk
Detego GmbH
Graz, Austria
s.walk@detego.com

Denis Helic
Graz University of Technology
Graz, Austria
dhelic@tugraz.at

ABSTRACT

Recommender systems are efficient *exploration* tools providing their users with valuable suggestions about items, such as products or movies. However, in scenarios where users have more specific ideas about what they are looking for (e.g., they provide describing narratives, such as *"Movies with minimal story, but incredible atmosphere, such as No Country for Old Men"*), traditional recommender systems struggle to provide relevant suggestions. In this paper, we study this problem by investigating a large collection of such narratives from the movie domain. We start by empirically analyzing a dataset containing free-text narratives representing movie suggestion requests from reddit users as well as community suggestions to those requests. We find that community suggestions are frequently more diverse than requests, making a recommendation task a challenging one. In a prediction experiment, we use embedding algorithms to assess the importance of request features including movie descriptions, genres, and plot keywords, by computing recommendations. Our findings suggest that, in our dataset, positive movies and keywords have the strongest, whereas negative movie features have the weakest predictive power. We strongly believe that our new insights into narratives for recommender systems represent an important stepping stone towards novel applications, such as interactive recommender applications.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; *Users and interactive retrieval*; *Personalization*.

KEYWORDS

Narrative-driven recommendations; Recommender systems; Empirical analysis

---

SUBMISSION
"**[Request] Movies about writing/writers.** Two of my favourites are *Secret Window* and *Stranger Than Fiction*. I also liked *The Ghost Writer*. [...] I'm not a fan of horror. I know there are probably a lot of 'inspirational' movies about writing out there (I vaguely recall one with Sean Connery?). [...]"

COMMENTS
"Adaptation."
"Sean Connery movie was Finding Forrester."

⋮

Box 1: *Request & Suggestions* Example. In the request of this reddit submission[1] crowdworkers annotated three positive movies (i.e., *Secret Window, Stranger Than Fiction, The Ghost Writer*), a negative genre (i.e., *horror*), several positive keywords (i.e., *writing, writers, inspirational*), and a positive actor (i.e., *Sean Connery*). As suggestions from the reddit community, the crowdworkers extracted the movies *Adaptation* and *Finding Forrester* from the comments section.

1 INTRODUCTION

Search engines are omni-present tools designed to help users retrieve information when they specifically *know what they are looking for* (i.e., they can articulate what they want with a few simple keywords). On the other hand, users rely on recommender systems when they *are unable to specifically state what they seek* (i.e., they vaguely know what they want but can not articulate it). In that case, recommender systems allow users to *explore* large collections of items and find interesting items by, for example, browsing [10]. **Problem.** In contrast to those two information seeking situations, in narrative scenarios users have a more specific idea what they are looking for, but the information need is often too complex to be articulated in the form of a few simple keywords. For example, in online boards they describe what they are looking for in the form of a narrative request while other users provide relevant recommendations (cf. Box 1).

Results of our recent work on such recommendation scenarios [5] indicate that the problem of narrative-driven recommendations is hard and difficult to address with traditional recommender approaches. However, our research community still lacks a deeper understanding of the potential causes for the hardness of this problem. For instances, we miss insights into user preferences in narratives, and, particularly, whether users tend to illustrate their needs

---

[1]https://www.reddit.com/r/MovieSuggestions/comments/ssuhu

through (i) examples, (ii) by describing the characteristics of desired items, or (iii) by a specific combination of both examples and characteristics. Also, the question whether positively associated aspects (e.g., *writing*, *writers*, *Secret Window* and *Stranger Than Fiction* in Box 1) are more important for calculating recommendations than negatively associated aspects (e.g., *horror* in Box 1) is still unanswered in the recent research on this topic.

This Work. In this paper, we set out to learn more about narratives by analyzing a movie suggestion board from reddit (r/Movie-Suggestions[2]). We start our study by empirically analyzing a dataset that consists of narrative requests and corresponding movie suggestions from the reddit community. To quantify the difficulty of this problem we analyze the diversity of requests and their corresponding suggestions. Further, we evaluate the effects of positive vs. negative features of users on reddit. Next, we utilize document and graph embedding techniques to (i) compute algorithmic recommendations and to (ii) evaluate the importance of features extracted from requests by comparing algorithmic recommendations with movie suggestions from the reddit community. The results of our empirical analysis reveal that community suggestions are often-times more diverse than requests, meaning that similar requests are frequently answered with highly diverse movie suggestions by the reddit community. The results of our recommendation experiment show that positive movies and keywords have the strongest, whereas negative movie features the weakest predictive power for narrative-driven movie recommendations on reddit.

Related Work. A special context-aware recommendation scenario called narrative-driven recommendation was proposed by Bogers and Koolen [4]. In the recommendation calculation process, besides the user history a narrative explanation of the current recommendation needs of the respective user is utilized.

Although Glenski and Weninger [6] showed that simple models are able to predict user interactions, such as likes, votes, clicks, and views, on reddit, recommending movies based on narrative requests on reddit is hard [5]. In our previous work [5], we determined the suitability of well-established recommender algorithms for calculating narrative-driven movie recommendations. For evaluation, we built a crowdsourced dataset from reddit submissions providing narrative movie recommendation requests and comments including corresponding movie suggestions by the reddit community. The obtained results reveal that the problem of predicting narrative-driven recommendations is a hard problem and needs further investigation to get a better understanding of narrative aspects.

Contrarily, in this work we use this crowdsourced dataset and follow up by introducing the first in-depth empirical analysis of movie recommendation requests on reddit. Moreover, we evaluate the embedding of narratives through the document and the graph embedding techniques doc2vec and node2vec.

In the large and well-investigated research field of recommender systems and algorithms [1, 4, 7] there exists a vast variety of studies based on word2vec [12] and its extensions doc2vec [11] and node2vec [8] partly exhibiting outstanding performances [2, 5, 13, 14]. Kallumadi and Hsu [9] evaluated the effectiveness of query-based interactive movie recommendations on IMDb data using graph-level embeddings. They created meta paths with different

entities (e.g., users, movies, genres) to build movie networks as basis for their embeddings and obtained good results with node2vec.

In contrast to their work, we use multiple networks, each of which consists of nodes from the same type. Further, we combine node2vec with doc2vec embedding vectors based on textual movie information.

2 EMPIRICAL ANALYSIS OF NARRATIVES

We empirically analyze the publicly available crowdsourced dataset[3] that we extracted from reddit [3] in our previous work [5]. The dataset contains about 1,500 narrative requests that all received at least ten suggestions and about 21,000 suggestion lists including more than 43,000 individual suggestions. We list further details of our dataset in Table 1. Each request in our dataset includes one or more positive movies, which are examples of movies that the user liked before. Moreover, requests frequently include additional descriptions, such as negative movies (movies that the user did not like before), positive and negative keywords describing further aspects of the movies, positive and negative genres, and finally positive and negative examples of movie actors. In Box 1 we show a typical example of such a request from our dataset, in which crowdworkers annotated three positive movies, one negative genre, three positive keywords, and one positive actor. We also show two examples of the suggestion lists, each of them having a single suggestion.
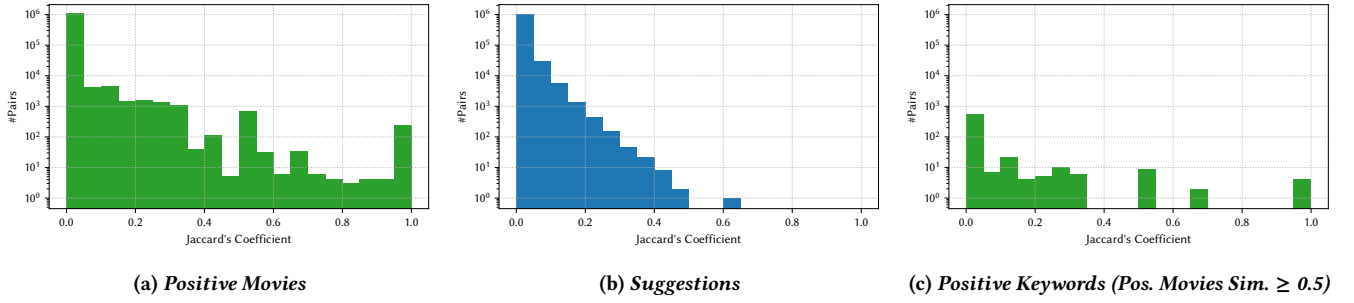
2.1 Dataset Characterization

Popularity Bias. We observe a heterogeneous distributions of positive movies and suggestions. Particularly, while the majority of the movies is mentioned only a few times, there also exist a few highly popular movies indicating a strong popularity bias. To check whether movies used as examples and suggestions are correlated we

---

[3]https://www.rbz.io/datasets

Table 1: *Reddit Dataset Characteristics*. This table shows the statistics of the crowdsourced reddit dataset [5].

#Requests	1,480
#Request Authors	1,244
#Movies in Requests	5,521
#Requests With Positive Movies	1,480
#Requests With Negative Movies	77
#Keywords in Requests (Without Common Words)	4,492 (3,947)
#Requests With Positive Keywords	1,202
#Requests With Negative Keywords	152
#Genres in Requests	762
#Requests With Positive Genres	459
#Requests With Negative Genres	55
#Actors in Requests	100
#Requests With Positive Actors	73
#Requests With Negative Actors	7
#Suggestions	43,402
#Suggestion Authors	7,431
Average #Suggestions per Request	29.33
Average Duration Between Request and Suggestion	31 h 41 min

---

[2]https://www.reddit.com/r/MovieSuggestions

(a) *Positive Movies*

(b) *Suggestions*

(c) *Positive Keywords (Pos. Movies Sim. ≥ 0.5)*

Figure 1: *Overlap of Movies & Keywords in Requests & Suggestions.* **These figures (y axes on the log scale) show the Jaccard's coefficient of positive movies (a), suggestions (b), and positive keywords (c), with positive movies similarity ≥ 0.5 for all request pairs. We observe positive movies and positive keywords in requests having a longer tail with a higher probability for high similarity values as compared to suggestions. For example, we find 1043 pairs consisting of 622 (42.03%) requests with positive movies, 708 pairs consisting of 410 (34.11%) requests with positive keywords, and only a single suggestion pair with Jaccard's coefficients ≥ 0.5. These findings indicate a strong diversity of suggestions in our dataset.**

compute Spearman's coefficient. We find that popular movies, such as *Fight Club* or *Memento*, are frequently used as positive examples as well as suggestions (Spearman's coefficient ρ=0.638, p<.001). In contrast to positive movies and suggestions, we do not observe any conclusive patterns in the distribution of negative movies due to their infrequent occurrences.

Hence, our initial findings indicate that automatic recommender algorithms trained on this or similar data may be strongly influenced by the popularity bias and should follow advanced strategies to account for this bias in order to, for example, increase beyond-accuracy metrics, such as diversity or novelty.

Users Prefer Dark Over Hollywood Movies. Looking at relative occurrence frequency among positive and negative keywords we find that users on reddit prefer to watch *dark* movies in *psychological* settings rather than *hollywood* or *gory zombie* movies.

Summary. Our initial dataset characterization suggests a strong popularity bias towards movies popular in this particular reddit community. The fact that there is a significant correlation between positive movies and suggestions also indicates that users frequently describe similar requests with differing movie examples. This highlights the importance of context in such narrative-driven recommendation requests as we observe no distinguishing use of positive and negative keywords. Thus, users signal their recommendation needs through specific and distinctive combinations of various features including example movies, genres, or keywords.

2.2 Suggestions Diversity

We continue our empirical analysis by investigating how diverse community suggestions are. Particularly, as reddit submissions resemble a typical discussion board structure, we expect that each following user suggesting movies recommends movies different to the already suggested ones. Moreover, as users can browse the submission lists and corresponding suggestions from the past, we also expect a high suggestion diversity across different requests. Taken all together, we hypothesize that highly diverse community responses will render narrative-driven recommendations as a challenging task for automatic recommender algorithms.

Overlapping Examples Result in Differing Suggestions. In the first step, we investigate how overlap in positive movie examples across requests relates to overlap (or lack thereof) in suggestions. Thus, we compute a standard overlap measure, the Jaccard's coefficient, of positive movies from all pairs of requests and corresponding suggestions. We plot all overlap distributions in Figure 1. We find that both requests and suggestions are typically dissimilar to the majority of other requests with average of 0.003 (sample standard deviation s=0.028) and suggestions with average of 0.008 (s=0.019). However, while there is still a substantial number of requests highly similar to at least one other request (e.g., with similarities ≥ 0.5). In contrast to requests, we do not observe such similarities in pairs of suggestions.

Varying Tastes Lead to Suggestion Diversity. One possible explanation for the diversity in suggestions may be that suggestions are provided by different users with diverging movie preferences. As we expect that users will follow their own tastes and preferences while suggesting movies, the varying user preferences will result in different suggestions even to identical requests.

To test this intuition, we start by removing all suggestion lists from deleted users (3% of suggestion lists), who are denoted as "[deleted]" in our dataset. Next, we analyze the overlap of suggesting users for requests with identical positive movies. In 230 of such request pairs we find an average overlap (as measured with the Jaccard's coefficient) between suggesting users to be 0.010 (s=0.024). Similarly, the average overlap of suggesting users for the request pairs with similarities ≥ 0.5 is as low as 0.008 (s=0.020). Thus, this observation corroborates the previous finding suggesting that typical users respond sporadically by answering only to a small number of posted requests or by providing only a small number of suggestions. Finally, to assess the effect of user personal preferences on the suggestions that they make we first eliminate the effects of confounding factors, such as additional positive keywords or genres. Therefore, we control for these features and extract a subset of requests with identical positive movies and no additional information. We find 14 such pairs comprising 17 requests (1% of all requests). Among those request pairs we find no overlap between