Vorabfassung des Artikels
Hirschmeier, Stefan, and Detlef Schoder. "An Approach to Explanations for Public Radio Recommendations." In Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP) Adj. ACM, 2020. https://doi.org/10.1145/3386392.3399306.

An Approach to Explanations for Public Radio Recommendations

ABSTRACT

Explanations are a key concept of public-service remits. For their linear program, public broadcasters regularly report on how they reflect diversity and balance in their program. For personalized public media, however, explanations must be different, as not one, but thousands of individual playouts have to be explained. We propose an approach how to design explanations for content-based radio recommendations based on word embedding approaches and evaluate the acceptance of our explanations with interviews of test listeners. As a result, we reveal an apparent paradox—test listeners appreciate to have explanations, but at the same time do not intend to use the explanation feature often.

1 INTRODUCTION

In return for public funding, public broadcasters have to fulfil a public-service remit, which is written down in broadcasting agreements or broadcasting acts. Broadcasting acts represent a democratically legitimated formulation of requirements that society expects from its public broadcasters, with the overall goal to support opinion diversity and the forming of a mature and individual opinion. As a control mechanism, the broadcaster must periodically (e.g., bi-annually in Germany according to the German Broadcasting Act ("Staatsvertrag für Rundfunk und Telemedien" §11e) write a report on how each program complies with the public-service remit. The report serves as an explanation for the program and is handed to the public and the nation's broadcasting council for review. Thus, explanation of playouts is an integral part of broadcasting acts and the main control mechanism for public-media playouts.

Broadcasting agreements have their origins in a time where personalization was unimaginable or just a minor subject, so that exact requirements for personalized media have not been formulated. Still, broadcasting agreements clearly emphasize explanations of the playout as a key concept to comply with the public-service remit. As a consequence, with the change of the media landscape and its technical possibilities, also the requirements for explanation change and have to be reinterpreted within the new boundaries.

Nowadays, personalized[*] media consumption has become prevalent. In the case of public media, one has to assure a balanced playout, reflecting the diversity of opinions and topics and avoiding filter bubbles for every single playout to fulfil the requirements of the public-service remit. Next to recommendations that match the interests of the user, also so-called anti-recommendations [1] should be included in playouts that may widen the horizon of the recipient. Therefore, for personalized media, explanations are more complex and demanding than for a unified (linear) program for all. Not one program, but thousands of individual playouts have to be explained. As a consequence, we need other explanation mechanisms than before [2], and presumably explanations on two different levels—a general level (explanations that describe the mechanisms of the algorithm) and an individual level (explanations that describe the personal playout for each recipient).

First, on a general level, it is challenging to explain the underlying recommender algorithms and the mechanisms to prevent filter bubbles in a transparent way, as difficult to describe mathematics build the core of recommender systems. Such explanations for the mechanisms may be comprehensible for experts, but not necessarily for the public. Second, on an individual playout level, recommendations have ever since been accompanied by a black box feeling. Users sometimes wonder how and why certain recommendations came to be, as recommendations might be unexpected and leave the user with questions. This was already true for recommender systems that try to match the interests of the user, but it is certainly even more true for public value recommender systems that serve the interests of the society through balanced playouts [3]. Balanced playouts and the idea to prevent filter bubbles might directly contradict with personal interests of the user.

Focusing on individual explanations for public radio, we ask the question: *How to design explanations for recommender systems in public radio?*

[*] We use the terms personalized and individualized interchangeably throughout the paper

The remainder of the paper is structured as follows: First, we present a definition of radio in this paper and related work. In Section 3, we describe our approach to explanations for content-based recommendations within public radio and in Section 4, we present our evaluation results in four findings. Section 5 concludes with a discussion and limitations.

2 BACKGROUND AND RELATED WORK

2.1 Definition of Radio in This Paper

Radio has different meanings. Radio could denote pure music services, news services, journalistic radio, or a mixture of spoken-word content and music. Also, specific variants like talk radio exist. In some countries, radio stations may also have a special political purpose. In this paper, we consider radio as journalistic radio with a mixture of music and spoken-word content, including news.

2.2 Explanatory Approaches for Recommender Systems

Research in the field of explanations for recommender systems is mainly characterized by the types of explanations and by the objectives that should be pursued with explanations.

Tintarev and Masthoff [4] describe seven different objectives that an explanation for a recommender system can serve: transparency, scrutability, trust, effectiveness, persuasiveness, efficiency, and satisfaction. They provide an extensive overview of explanatory approaches in recommender systems and cover explanation styles for collaborative filtering, content-based recommendation, and knowledge-based and demographic explanations.

According to the research of Papadimitriou et al. [5], the explanation styles can be divided into three different categories: human explanations, item explanations, and feature explanations. Human explanations rely on similar users, while item explanations depend on which items a user has previously liked. Feature explanations are derived from individual features of items previously liked by a user. Thus, for instance, explanations such as "customers who bought this item also bought…" are human explanations. Not all explanation styles apply to all types of recommender systems. Because content-based recommendation systems utilize the similarity of items with user ratings to generate recommendations, most explanations for the recommendations of a content-based recommender system depend on the content of elements.

In the following, we present explanations from research and practice that may be suitable for explanations of recommendations in the public radio domain with different explanatory styles.

Feature-Style Explanations. An example of a feature-style explanation is the following: "this item is recommended because you like the features which are a part of the recommended item". The way the recommendations are computed ranges from graph structures [6] to naive Bayes algorithms [7] to incremental nearest neighbor algorithms [8]. Cramer et al. [9] tested the influence of transparency on other explanation objectives, such as user confidence and acceptance of the system. The evaluation of the results confirmed that the users of the feature-style condition found the system most transparent. Participants in the feature-style condition also tend to have a better understanding of how the system works. Mooney & Roy [7] built a content-based recommender with feature-style explanations in the field of book sales. An evaluation showed that feature-style explanations lead to better effectiveness, as human explanations influence users to overestimate the quality of an item from the explanation, which can lead to reduced user confidence in the system [10].

Item-Style Explanations. The item-style explanation often has the following structure: "The activity is recommended because you were interested in this other activity" [5, p. 566]. In literature, one way of using the item-style is to combine it with feature-style explanations to create hybrid explanations [7,11], [12]. In practice, this type of explanation of recommendations often occurs.

Hybrid-Style Explanations. A hybrid style explanation is a way of combining various styles. Papadimitriou et al. [5] conclude from various other sources [13,14] that hybrid-style explanations, as a hybrid between feature-style and item-style explanations, provide better user-perceived transparency, as well as the best user-perceived effectiveness and better user satisfaction than the item-style and the feature-style alone. Hence, the authors conclude that hybrid-style explanations combine individual strengths of feature-style and item-style as well as reducing their weaknesses. Hybrid-style explanations have the following style: "I suggest you Memento since you sometimes like movies as The Prestige and Sherlock Holmes, whose director was an Edgar Award Winner. Moreover, I recommend it because you like Films about Psychiatry as Psycho IV - The Beginning and 2000s films as Psycho IV the Beginning and Sherlock Holmes" [12, p. 98]. Musto et al. [12] tested explanation types for transparency, persuasiveness, commitment, trust, and effectiveness. The results show that the hybrid-style explanations provided both the best transparency and the best persuasiveness.

3 AN APPROACH TO EXPLANATIONS FOR CONTENT-BASED RECOMMENDATIONS

In cooperation with a major German public radio broadcaster, we developed an approach for content-based explanations in personalized public radio. The approach described below meets the requirements of the public-service remit (see [2] for a more detailed description of the requirements) and the individual requirements of the public broadcaster. To fulfil the diversity requirements of the public-service remit, two types of recommendations were needed: a) content-based recommendations that match the interests of the listener and b) content-based anti- recommendations that not necessarily meet the interests of the user but the requirements of the public-service remit.

We had a short teaser text available for all radio content of the last six months. To generate content-based recommendations, we used word embeddings to transform the teaser texts to dense vector representations that are processable in linear algebra. More specifically, we used the word2vec approach of Mikolov et al. [15] to train word vectors of all words in the text corpus. Based on these word vectors, we used the Smooth Inverse Frequency Sentence Embeddings (SIF) approach of Arora et al. [16] to calculate vector representations for the teaser texts.

We calculated the personalized recommendations as follows: First, we identified a set of radio items that are similar to the items that a user liked before. We calculated cosine similarities between the SIF embedding vectors of all liked items and all other available radio items. The items with the highest cosine similarity (calculated via dot products) form a candidate set for recommendations (it is possible to use a minimum threshold for cosine similarity— however, the size of the candidate set should be between 15 and 30). Second, to provide diversity within the recommendations, we used Determinantal Point Processes (DPP) [17,18] to figure out a fairly diverse subset of recommendations within the candidate set (we do not present the DPP process here in detail as the focus of this paper is explanations, not diversity). This diverse subset forms the personalized recommendations given to the user.

Analogously, we calculated anti-recommendations: First, we identified a set of radio items that are dissimilar to the items that a user liked before, again by cosine similarities of the SIF embedding vectors. However, dissimilarities are of a slightly different nature than similarities: While we can identify the similarity between two texts, e.g., politics, it is for example not possible to identify the opposite to politics. So, while similarities between two texts can be multifaceted, the dissimilarity between two texts can be even more multifaceted. For our purpose it was satisficing (i.e., satisfying and sufficient [19]) to identify dissimilar radio items, no matter in which way they are dissimilar. Texts with the lowest cosine similarity between all liked items and all other available items form the candidate set (we used a fixed set of 30). Second, just like for the recommendations, we used DPP to calculate a diverse subset of anti-recommendations within the candidate set.

In the end, we mixed recommendations and anti-recommendations to an individual radio stream for the listener (with x% recommendations and 1-x% anti-recommendations, whereas x can be determined by the broadcaster, e.g., 70%).

Considering explanations, public-service remits do not contain specific requirements. However, the public broadcaster formulated the following requirements for explanations: a) explanations should exist for both recommendations and anti-recommendations, b) explanations for recommendations should be content-based, and c) there should be a variety of explanations so that they do not appear repetitive to the listener.

To meet the requirements, we provided hybrid explanations that consist of two parts 1) item-style explanations and 2) feature-style explanations.

Considering 1), we created item-style explanations, so a user can connect the current recommended item i_{rec} to the most similar item in the set of liked items of the user profile $i_{sim.}$ In the following we present three examples:

- "This item is recommended to you because you liked the item i_{sim}."
- "Within your liked items, the most similar item to this recommendation is i_{sim}"
- "You might like this item because it resembles the item i_{sim} you like."

Considering 2), the second part of the explanation, we provided features-style explanations. For every pair of i_{rec} and i_{sim}, we identified the three most similar terms by calculating the cosine similarities of all word vectors (only taking nouns into account). We added to the explanation: "Topic similarity: *first topic, second topic, third topic*".

Joining item-style explanations and feature-style explanations, we were able to provide hybrid-style explanations for the recommendations (see Figures x).

Considering explanations for the anti-recommendations, we chose formulations as follows (three examples):

- "This item is something completely different to what you usually hear. We hope you like it anyway."
- "This item has been randomly chosen from all other items. Please rate items, so we can offer good recommendations to you."
- "We picked this item randomly. To generate recommendations, you have to indicate which items you like."

In the initial concept, we had longer and more concise explanations, e.g., we differentiated whether a recommendation fits very well, well or medium well to the user profile, and further item-style explanations like "we recommended this item because it fits your profile in general". However, we learned during pretests that they did not provide value to listeners, so we eliminated them from our concept.

4 EVALUATION

For evaluation purposes, we developed a web app (See Fig. 1) for smartphones that consist of a simple player (play, pause, skip, rewind 15 seconds) and recommender interactions (like, explain). The prototype was connected to the media library of a public broadcaster. The possibility to get explanations was designed on-demand, that is, listeners has to tap on a button to receive explanations.

Considering the early stage of developing explanations for public radio, we chose a qualitative evaluation design with personal interviews. We developed an interview guideline for descriptive face to face interviews to get a rich description of the perception and acceptance of explanations. After a small introduction of our research to the test listeners, we presented the prototype to the listeners and asked them to express everything they think and see following the think-aloud method.

During the prototype development, we pre-tested our prototype with ten test listeners to see if we need to adjust the prototype or our interview guideline. After that, for the subsequent evaluation, we conducted 34 interviews of 10 to 15 minutes between November 2019 and January 2020. We decided to gather feedback from both radio listeners and non-listeners, so we interviewed 50% listeners and 50% non-listeners. The age of our test listeners ranged from 21 to 72 years. In the following, we present the main findings. Findings 1 and 2 were already made during the pretest phase and led to changes on the prototype.

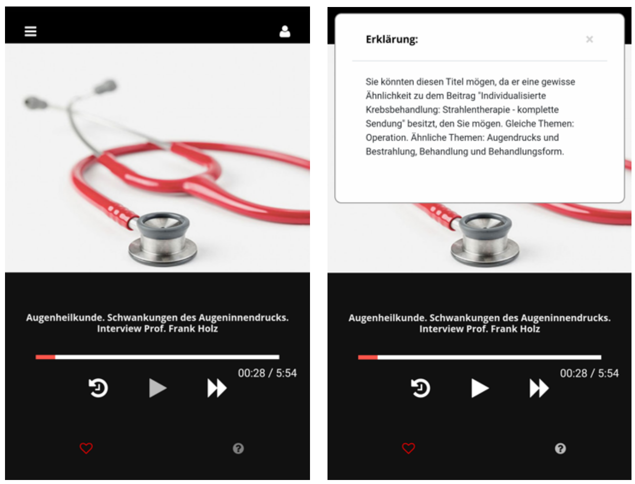

Figure 1: Prototype for content-based personalized radio – standard view (left) and explanation popover (right)

Finding 1. In the pretest we learned that test listeners did not notice differentiations in the explanations how well a recommendation fits to their user profile. More concise explanations were often perceived as too wordy. Test listeners favored shorter explanations instead of longer, more concise explanations.

Finding 2. Less concise explanations like "we recommend this item because fits your profile in general", however, were perceived as too general or meaningless by test users.

In general, test listeners liked the explanations. A majority of 79% rated the explanations as useful and trust-building. "I like to read the explanations when something is suggested to me that I cannot classify". This confirms our assumption that explanations are particularly useful for unexpected, non-recommender items [3].

However, we noticed that test listeners did not extensively use the explanation function. More than half of the users (56%) stated that they would not use the explanations regularly: "I generally like the explanations, because it makes the whole application transparent. But I don't know how much I would really use it. I don't think I'd really use it at all." Interestingly, more than half of the test listeners would appreciate such an explanation function but not use it. In the same vein, 65% of the test listeners agreed that explanations should be displayed on demand only.

Finding 3. Most testers stated that they appreciate the possibility to get an explanation but would not use it regularly. It seems sufficient to know that there is an explanation.

During pretests and evaluation, test users struggled with the explanation button (question mark in a circle). Some test listeners expected a help button behind the '?', others expected further description of the item currently playing. We discussed alternative symbols such as '!', 'i' or stethoscopes, but none of them seemed intuitive.

Finding 4. It is hard to find an intuitive symbol for explanations.

5 DISCUSSION AND CONCLUSION

Interestingly, our approach to explanations for content-based recommendations in public radio revealed an apparent paradox. Listeners seem to like explanations but do not intend to use them frequently. Thus, the added value of explanations seems to be primarily passive. However, if we refer back to the role of explanations in the public-service remit, we find an interesting parallel: Similarly, in the world of linear mass media, most listeners would not want to read the report of the broadcaster to get an explanation of the broadcaster's program. It is sufficient and calming to know that there is an explanation available that could be read if desired. The possibility to consult the explanations seemed to convince the user of the transparency and trustworthiness of the prototype.

As we found out in a conversation with another public broadcaster in Germany, their tests on explanation led to similar results: Users appreciate the possibility to get explanations, but at the same time did not use the feature often.

As a consequence, for practical implications, we have to reconsider how to integrate explanations into the radio experience. It does not seem necessary to place explanations very prominent in the listener's journey. It seems rather appropriate to hide explanations and make them available on demand. The question has to be critically raised how much effort a company should invest into explanations if the feature is rarely used. One could take the position that resources should better be invested in other features that are more present to the user. However, especially from a public-service perspective, one could argue that explanations are, even if rarely used, of high social value and therefore should be designed with care.

For the formulation of explanations, we have indication that explanations should not be too wordy (Finding 1), but also not too general (Finding 2). So far, with our content-based explanations, we were able to explain the recommendation over topical similarities. However, broadcasting agreements contain more requirements for public media playouts, such as a mix of culture, information, entertainment and education, and a mix or regional, national and international matters (see [3]).

Nevertheless, our explanations could help to create both transparency and trust, by exposing the reasoning and data behind a recommendation and thereby increasing the trustworthiness of the recommender system. If these explanations are accompanied by a possibility of intervention to inform the system that it was wrong with a recommendation, such explanations could even strengthen user control [20].

Our work has some limitations. First, we only interviewed test listeners for about 15 minutes and did not observe their click behavior in a long-term test to confirm our findings. Second, our selection of test listeners might have been biased towards age and gender. And finally. we provided item- and feature-based explanations from a similarity perspective, but not form a composition perspective. Future research might therefore find more sophisticated ways to explain individual playouts also from an editorial point of view.

ACKNOWLEDGEMENTS

This work was supported by Deutschlandradio in a joint research project with the University of Cologne. Many thanks to Markus Waldhauser and Dr. Nicola Balkenhol for their support.

REFERENCES

[1] N. Pöchhacker, M. Burkhardt, A. Geipel, and J.-H. Passoth, "Interventionen in die Produktion algorithmischer Öffentlichkeiten: Recommender Systeme als Herausforderung für öffentlich-rechtliche Sendeanstalten.," kommunikation @ gesellschaft, vol. 18, 2017.
[2] S. Hirschmeier and V. Beule, "Compliance of Personalized Radio with Public-Service Remits," in Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP) Adjunct, Singapore, 2018, [Online]. Available: https://doi.org/10.1145/3213586.3226204.
[3] S. Hirschmeier, "Towards Explanations of Anti-Recommender Content in Public Radio," in Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP) Adjunct, Larnaca, Cyprus, 2019, pp. 229–233, doi: 10.1145/3314183.3323454.
[4] F. Ricci, L. Rokach, and B. Shapira, Introduction to Recommender Systems Handbook. Springer, 2015.
[5] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "A generalized taxonomy of explanations styles for traditional and social recommender systems," Data Min Knowl Disc, vol. 24, no. 3, pp. 555–583, May 2012, doi: 10.1007/s10618-011-0215-0.
[6] A. Passant, "dbrec — Music Recommendations Using DBpedia," in The Semantic Web – ISWC 2010, Berlin, Heidelberg, 2010, pp. 209–224, doi: 10.1007/978-3-642-17749-1_14.
[7] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in Proceedings of the fifth ACM conference on Digital libraries, San Antonio, Texas, USA, Jun. 2000, pp. 195–204, doi: 10.1145/336597.336662.
[8] D. Billsus and M. J. Pazzani, "A personal news agent that talks, learns and explains," in Proceedings of the third annual conference on Autonomous Agents, Seattle, Washington, USA, Apr. 1999, pp. 268–275, doi: 10.1145/301136.301208.
[9] H. Cramer et al., "The effects of transparency on trust in and acceptance of a content-based art recommender," User Model User-Adap Inter, vol. 18, no. 5, p. 455, Aug. 2008, doi: 10.1007/s11257-008-9051-3.
[10] M. Bilgic and R. J. Mooney, "Explaining Recommendations: Satisfaction vs. Promotion," in Beyond Personalization 2005: A Workshop on the Next Stage of Recommender Systems Research at the 2005 International Conference on Intelligent User Interfaces, 2005, Accessed: Apr. 15, 2020. [Online]. Available: http://www.cs.utexas.edu/~ai-lab/pub-view.php?PubID=51494.
[11] C. Musto, F. Narducci, P. Lops, M. De Gemmis, and G. Semeraro, "ExpLOD: A Framework for Explaining Recommendations based on the Linked Open Data Cloud," in Proceedings of the 10th ACM Conference on Recommender Systems, Boston, Massachusetts, USA, Sep. 2016, pp. 151–154, doi: 10.1145/2959100.2959173.
[12] C. Musto, F. Narducci, P. Lops, M. de Gemmis, and G. Semeraro, "Linked open data-based explanations for transparent recommender systems," International Journal of Human-Computer Studies, vol. 121, pp. 93–107, Jan. 2019, doi: 10.1016/j.ijhcs.2018.03.003.
[13] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "Justified Recommendations based on Content and Rating Data," in WebKDD Workshop on Web Mining and Web Usage Analysis, 2008.
[14] J. Vig, S. Sen, and J. Riedl, "Tagsplanations: explaining recommendations using tags," in IUI '09, 2009, doi: 10.1145/1502650.1502661.
[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," International Conference on Learning Representations (ICLR) Workshop, 2013, Accessed: Mar. 08, 2018. [Online]. Available: http://arxiv.org/abs/1301.3781.
[16] S. Arora, Y. Liang, and T. Ma, "A Simple but Tough-to-Beat Baseline for Sentence Embeddings," in Proceedings of the International Conference on Learning Representations (ICLR), 2017.

[17] A. Kulesza and B. Taskar, “k-DPPs: fixed-size determinantal point processes,” in Proceedings of the International Conference on International Conference on Machine Learning, 2011, pp. 1193–1200, Accessed: Apr. 14, 2020. [Online].
[18] O. Macchi, “The coincidence approach to stochastic point processes,” Advances in Applied Probability, vol. 7, no. 1, pp. 83–122, Mar. 1975, doi: 10.2307/1425855.
[19] H. A. Simon, “Theories of Decision-Making in Economics and Behavioral Science,” The American Economic Review, vol. 49, no. 3, pp. 253–283, 1959.
[20] N. Tintarev and J. Masthoff, “Designing and Evaluating Explanations for Recommender Systems,” in Recommender Systems Handbook, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 479–510.