BBC Recommender System

Mid Fidelity Prototype

Techincal Report

Michal Kubina

1 Introduction

Nowadays video streaming services are even more used than TV services (Kelly and Sørensen, 2021). And they usually are quite different compared to television broadcasting. The user can choose whatever they like to watch no matter what time it is. This introduces a new way how to interact with the user. This is done by recommendation systems that recommend content to the consumers. In other words, the recommendation system evaluates if an item might be useful to the user based on the user's data and content of the recommended items (Ricci et al., 2011).

However, the user itself is not a single entity influenced by the recommender system. There are direct and indirect stakeholders who could and sometimes should have a say in the recommendations themselves (Voinov and Bousquet, 2010). Moreover, public recommender systems are even harder to create as the public recommender will deviate from short-term preferences, for common good (Smets et al., 2020).

But how do the recommendations work? Recommendation systems can be a content-based, user-based, hybrid, or other specific types (Fayyaz et al., 2020). In my case, I am using scraped data by myself - the content which is available on BBC video-on-demand service. These data points are shows and movies available there. My public recommender system is content-based since we have no user history.

But how the values should be defined and the stakeholders themselves? How the recommender should be defined to be sensitive to all possible stakeholders and which values it should comprehend, implement and evaluate?

A value sensitive design is a common method that consists of three parts: conceptual, empirical, and technical investigation (Friedman et al., 2002). These methods can be seen as a building rock that can be used iteratively and without a specific order. Considering literature, there is a vast variability of usage of this design (Winkler and Spiekermann, 2021). In the conceptual part, it is common to define stakeholders and values. In the empirical part, we try to get insights from stakeholders about values and the technical part considers the implementation itself (Simon et al., 2020).

In my case, I am trying to follow a value sensitive design but simultaneously I try to use it in a new and innovative way by following this procedure:

1. Definition of specific stakeholders
2. General semi-structured survey
3. Value specific semi-structured survey
4. Interface design and implementation
5. User testing (survey)

As I mentioned earlier, there can be multiple direct and indirect stakeholders including endpoint users, government, advertising agencies, and other entities. For my specific recommender system, I am focusing on a specific group of direct stakeholders.

In other words, my persona can be defined as a young university student who understands that there is a technical side to recommender systems. In other words, the persona understands that the raw recommendations are based on some type of algorithm. Moreover, the persona is capable of understanding technical parameters that might be present in the recommender system's interface.

For my research, I interviewed 13 international students from Utrecht University with their ages ranging from 20 to 28. After conducting a semi-structured interview asking about recommender systems generally, it was noticeable that the participants of the interviews lacked the possibility of influencing the recommendations. Moreover, they expressed curiosity about seeing content that is not their usual taste. Finally, nearly every participant implied that they would acknowledge understanding why the things they see in recommendations systems are recommended.

2 Literature Review

However, how these findings can be expressed and named? We can start by defining a value first.

From an economical point of view, value of an object can be described in terms of money. From a human and more general perspective, value can be described as something important to a person or a group of people in their life (Friedman et al., 2002).

After conducting the first review, I concluded that three values are projecting and intersecting in the minds of users. Thus, there is a need for autonomy, transparency, and exposure diversity.

Autonomy can be described as the ability of the user to control the interface, have a say in or influence the recommendations, and act upon own decisions (Friedman, 1998).

Transparency as a value evolved simultaneously with the upcoming era of law advances (Ball, 2009). Recommended systems usually work as black-box algorithms and this issue needs to be challenged. This can be resolved by implementing a transparent recommender system. Transparency, in this case, can be described as the ability of the user to understand recommendations, thus understanding the context between input and output (Sinha and Medhurst, 2002).

Sometimes we as users want to see content that is not usual to our taste. In other words, we want to be exposed to content we would not normally choose. This value can be described as exposure diversity. Though there is no specific and exact definition of this abstract value, it can be understood as a way how to correct mistakes of not recommending items the users would like and simultaneously getting users out of their usual homogenous virtual environment (Helberger et al., 2018).

I believe that these three values constitute a mix of values that needs to be implemented in a more significant way in today's recommender systems.

After the definition of these values, I conducted another semi-structured interview with the 13 students to explore more the specific values and I asked questions to get a general idea of how to operationalize them in the method and interface itself.

3 Method

From the answers, I noticed that the stakeholders do not generally care about the type of algorithm and method which is used for producing recommendations. They just want to be informed about the method and be able to operate with the parameters. Thus, I will explain the algorithm in the following sentences and I will elaborate on how it operationalizes autonomy, transparency, and exposure diversity from the perspective of the BBC data set.

My single row of data corresponds to specific content. Each show or movie has a specific name, image, category, date, and description. These are the five features that are used in my recommender system. The original data frame is filtered by category and date. This is a way how to implement user autonomy which is consequently handled by interface.

However, the inner working takes into account just a description of the shows and movies. The punctuations were filtered from descriptions together with stop words. Consequently, tokenization of descriptions was conducted. Firstly, the term frequency-inverse document frequency matrix is computed from the corpus of descriptions. This method looks at how common (or uncommon) a word is amongst the corpus of documents (Bafna et al., 2016). K-Means is then used to cluster documents according to the content of their respective tf-idf vector based on the chosen K (Sinaga and Yang, 2020).

Similar and dissimilar content is chosen based on the output of the K-means clustering algorithm and thus this is a way how exposure diversity is implemented. Moreover, the user autonomy is operationalized by the user's ability

to choose the number of clusters of the algorithm.

The last value, transparency is implemented by using a user interface which is described in the upcoming section.

4 Interface Design

My interface can be seen as a very simple application. The main dashboard fits just one page or screen. The reason behind this is that the users expressed the need to keep the recommender system easy to use.

The dashboard can be structured into three rows. The first row displays the parameters that the user can alter, buttons, and the chosen content. The second row contains very similar content to the chosen main content and the third row contains content that is not similar to the chosen content.

From the value point of view, autonomy is supported by the ability to choose categories of movies and shows, their date, and the number of clusters present in the main algorithm. These ideas were also taken from the semi-structured interview. Another support for this value is a button that can delete content from the data if the user does not like it all.

Transparency as a value is implemented by a message describing the inner working in the first row on the right side of the dashboard. Moreover, the fact that the user can choose the number of clusters in the main algorithm is also a step for greater transparency of this recommender system. Moreover, the resulting cluster can be seen on a button to view the content.

Exposure diversity is implemented by two main ideas. First idea is a button "shuffle me" in the first row. This button chooses random content in terms of chosen parameters and thus changes the dynamics of a whole dashboard. The reason behind this approach is to test the stakeholders' satisfaction even though they expressed just a small need to get out of their usual bubble. This might help the users to get off their usual content and if necessary, they can tweak the main algorithm by deleting unwanted content by using the button "delete me". Moreover, the second row contains content that is clustered the same as the main chosen content and the third row contains all the data that are clustered differently. Each displayed show or movie can be naturally chosen as the main content by clicking on the button open.

After completing the interface, user testing was conducted on the 13 participants. All participants were asked to try the interface for a few minutes and then to ask the following questions which were not specifically mentioned values on purpose

- Do you feel that you are more in control compared to the usual recommender system? (autonomy)
- Would you say that you are more exposed to your unusual content (exposure diversity)
- Do you feel that the results are presented clearly? (transparency)

All participants agreed that they feel more in control. Eight participants agreed that they feel more exposed to the unusual content. However, just two participants of the user testing were satisfied with the level of transparency.

5 Conclusion

To start with the technical side point of view, there is a limitation of the prototype in terms of inner working. Sometimes there can be no data for the chosen set of groups and date range. In future work, this could be solved with a pop-up window informing the user of a wrong and not ideal selection. Moreover, filtering of the data set could be a big help since there is sometimes duplicate content.

Another limitation might be a small group of users who were interviewed and tested the system. All personas were international master's students of the Utrecht University from multiple different study areas. Still, this might be a very specific subgroup in terms of chosen persona and the results might be biased towards more technical-oriented people.

Considering the result of the final user testing, it seems that my recommender system is capable of satisfying values such as autonomy and exposure diversity. However, the users, after conducting user testing, were not satisfied with transparency. Thus, in future work, this value should be highlighted and the interface should be changed. For example, implementing the information about specific tf-idf scores of shows and movies could be added to the interface. This might make the transparency as a value more present.

References

Bafna, P., Pramod, D., & Vaidya, A. (2016). Document clustering: Tf-idf approach. *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 61–66. https://doi.org/10.1109/ICEEOT.2016.7754750

Ball, C. (2009). What is transparency? *Public Integrity*, *11*, 293–308. https://doi.org/10.2753/PIN1099-9922110400

Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Applied Sciences*, *10*, 7748. https://doi.org/10.3390/app10217748

Friedman, B. (1998). User autonomy: Who should control what and when? a chi 96 workshop. *SIGCHI Bull.*, *30*(1), 26–29. https://doi.org/10.1145/280571.280583

Friedman, B., Kahn, P. H., & Borning, A. (2002). Value sensitive design: Theory and methods.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207. https://doi.org/10.1080/1369118X.2016.1271900

Kelly, J., & Sørensen, J. (2021). "what's on the interface tonight?": A longitudinal analysis of the publishing strategies of public service video-on-demand platforms in the uk and denmark. *MedieKultur: Journal of media and communication research*, *37*, 066–090. https://doi.org/10.7146/mediekultur.v37i70.122386

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. *Recommender Systems Handbook*.

Simon, J., Wong, P.-H., & Rieder, G. (2020). Algorithmic bias and the value sensitive design approach. *Internet Policy Review*, *9*. https://doi.org/10.14763/2020.4.1534

Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised k-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Sinha, R., & Medhurst, K. (2002). The role of transparency in recommender systems. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/506443.506619

Smets, A., Walravens, N., & Ballon, P. (2020). Designing recommender systems for the common good. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 276–278. https://doi.org/10.1145/3386392.3399570

Voinov, A., & Bousquet, F. (2010). Modelling with stakeholders [Thematic Issue - Modelling with Stakeholders]. *Environmental Modelling Software*, *25*(11), 1268–1281. https://doi.org/https://doi.org/10.1016/j.envsoft.2010.03.007

Winkler, T., & Spiekermann, S. (2021). Twenty years of value sensitive design: A review of methodological practices in vsd projects. *Ethics and Information Technology*, *23*. https://doi.org/10.1007/s10676-018-9476-2