An Approach to Model Self-imposed Filter Bubbles

Annalena Aicher and **Wolfgang Minker**
Institute for Communication Engineering
Ulm University, Germany
annalena.aicher@uni-ulm.de

Stefan Ultes
Mercedes Research & Development
Sindelfingen, Germany

Abstract

. When people are confronted with an overwhelming amount of information, they tend to filter out all the parts of the available information that do not fit their existing beliefs or opinions. Within this paper, we propose the first model to describe this "self-imposed filter bubble" (SFB) during argumentative information seeking. Based upon this model, argumentative dialogue systems (ADS) shall be able to learn and adapt their dialogue strategy to overcome this SFB in cooperation with the user.

1 Introduction

Especially when searching for information online, users tend to select claims that adhere to their beliefs and to ignore dissenting information, which coins the terms self-imposed filter bubbles (SFB) (Ekström, 2021) and echo chambers (Quattrociocchi et al., 2016). These phenomena belong to the generic term *confirmation bias* which is typically used in psychological literature (Nickerson, 1998). Allahverdyan and Galstyan (2014) describe confirmation bias as the tendency to acquire or evaluate new information in a way that is consistent with one's preexisting beliefs.

To resolve the confirmation bias of a user in decision making processes Huang et al. (2012) propose the usage of computer-mediated counterargument. Furthermore, Schwind and Buder (2012) regard preference-inconsistent recommendations as a promising approach to trigger critical thinking. Still, if too many counter-arguments are introduced this could lead to unwanted effects negative emotional consequences (annoyance, confusion) (Huang et al., 2012). Consequently, Huang et al. (2012) stress the need for an intelligent system which is able to adapt the frequency, timing and choice of the counter-arguments. To provide such a system, it is crucial to develop and find a model, which can be adapted to the user. The goal of this paper is to present such an abstract model for a user's individual self-imposed filter bubble. It is based on our previous work (Aicher et al., 2022) and consists of the four dimensions *Reflective User Engagement (RUE), Personal Relevance (PR), True Knowledge (TK)*, and *False Knowledge (FK)* and makes it possible to assess the probability of a user being caught in a self-imposed filter bubble with regard to a certain topic. To the best of our knowledge, our approach is the first existing model of a user's SFB which is furthermore suitable to be implemented in an argumentative dialogue system.

2 Self-imposed Filter Bubble Model

As previously mentioned we focus on four dimensions in our model[1]. Their choice is examined in detail in (Aicher et al., 2022) and builds upon findings in well-established state-of-the-art literature (Petty et al., 2009).

2.1 SFB-Model Dimensions

The **RUE** describes the critical-thinking and openmindedness demonstrated by the user. It takes into account the polarity and number of arguments he/she has heard. This can be mapped onto two actions of the user by asking for more information, either on the pro or con side of the topic of the discussion. Thus, it can be interpreted as a weighting how balanced the user is exploring the topic. Due to the limited scope of this paper we refer to our previous work (Aicher et al., 2021) where its calculation is described in detail.

The **PR** refers to the user's individual assessment of how relevant a subtopic is with regard to the topic of the discussion. We assume that the bigger the PR of a certain subtopic is, the higher is the user's interest and motivation to explore arguments belonging to it.

The **TK** serves as a measure for the information gain and is defined as the new information the user

[1]Please note, that we do not claim the dimensions or our model to be complete but a first approach to model SFBs.

is provided with by talking to the system. It can be determined by comparing the total information provided by the system and the information, which is already known to the user. We aim for the user to explore as much information as possible, as this increases the chance to explore other aspects and viewpoints. Thus, the bigger the TK of the users, the more unlikely they find themselves in an SFB.

The **FK** describes the incorrect information a user has on a certain topic[2]. If the user is misinformed on certain aspects, it increases the probability of being stuck in an SFB and reluctant towards contradicting information and viewpoints.

2.2 SFB-Model

Using the dimensions in Subsection 2.1, we define an SFB-vector $\overrightarrow{SFB}$. It has its origin in the origin of the coordinate system and its end is the position of the user in the four-dimensional space at the current state of the interaction.

$$\overrightarrow{SFB} = (PR, RUE, TK, FK)^T. \qquad (1)$$

The SFB is described by a four-dimensional body

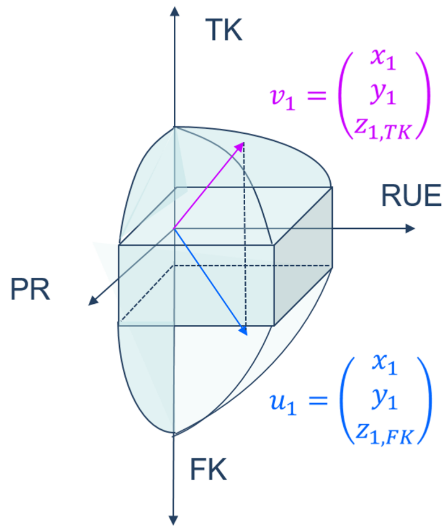

Figure 1: Schematic sketch of an SFB-vector and SFB. For better illustration the four-dimensional SFB-vector is displayed in two split components only differing in their z_1 component. Whereas the blue vector displays TK in the z_1-component, the violet one displays FK. The x_1 component depicts RUE and y_1 PR. The blue filled areas denote the SFB.

describing the probability with which users find themselves within an SFB. Obviously, it is very difficult to determine an exact limit up to which point users are still in their SFB and from which point on-wards they no longer are. The smaller the SFB-vector, the higher the probability that the user is inside the SFB. The longer the SFB vector and the more it extends beyond the SFB, the lower the probability that the user is within the SFB. In Figure 1 an exemplary sketch of this vector and the respective SFB are shown. As a four-dimensional vector cannot be displayed, for better illustration, it was split in two different z_1-components TK and FK. Please note that this sketch is for illustrative purposes only and it is very difficult to determine the “real” shape of the SFB. Therefore the light blue coloured areas indicate a high probability of being inside the bubble, while the non-coloured areas indicate a low probability, without defining the exact boundary of the bubble. To detect and “break” the user’s SFB in an ongoing interaction, the model can be adapted dynamically during the interaction. To estimate the success of breaking the SFB the position of the initial (before the interaction) and final (after the interaction) SFB-vector with respect to the SFB are considered.

3 Conclusion and Future Work

In this work, we introduced a novel model for a user’s self-imposed filter bubble, consisting of four dimensions: *Reflective User Engagement, Personal Relevance, True Knowledge* and *False Knowledge* (but not limited thereto). To the best of our knowledge this model represents the first approach to estimate the probability that users find themselves within an SFB. To break the user’s SFB it is important not to force new information onto the user but to find a more subtle way to weave in information that is not requested (Huang et al., 2012). Our SFB model shall help to identify suitable points of reference (e.g. the most decisive dimensions strengthening the bubble) which can be used as starting point to break the user’s SFB in an engaging cooperative argumentative dialogue. In future work, our model will be implemented in a suitable (cooperative) ADS and evaluated in a user study. Therefore, we will investigate how the change and behaviour of each dimension can be tracked in detail during an ongoing interaction using explicit and implicit methods. Furthermore, other potential dimensions shall be explored, such as user trust, communication styles and a virtual agent interface.

[2]Without loss of generality, the information in the system’s database is assumed to be correct and consequently, information contradicting the former to be incorrect.

References

Annalena Aicher, Wolfgang Minker, and Stefan Ultes. 2021. Determination of reflective user engagement in argumentative dialogue systems.

Annalena Aicher, Wolfgang Minker, and Stefan Ultes. 2022. Towards modelling self-imposed filter bubbles in argumentative dialogue systems. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pages 4126–4134.

Armen E Allahverdyan and Aram Galstyan. 2014. Opinion dynamics with confirmation bias. *PloS one*, 9(7):e99557.

Axel Ekström. 2021. The self-imposed filter bubble hypothesis. Student Paper.

Hsieh-Hong Huang, Jack Shih-Chieh Hsu, and Cheng-Yuan Ku. 2012. Understanding the role of computer-mediated counter-argument in countering confirmation bias. *Decision Support Systems*, 53(3):438–447.

Raymond S Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220.

Richard E Petty, Pablo Briñol, and Joseph R Priester. 2009. Mass media attitude change: Implications of the elaboration likelihood model of persuasion. In *Media effects*, pages 141–180. Routledge.

Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. 2016. Echo chambers on facebook. *Available at SSRN 2795110*.

Christina Schwind and Jürgen Buder. 2012. Reducing confirmation bias and evaluation bias: When are preference-inconsistent recommendations effective–and when not? *Computers in Human Behavior*, 28(6):2280–2290.