

Reading Interesting News While Forming New Interests

We live in the information age. Since the advent of the digital revolution information is increasingly easier to produce, copy and share. This has caused an unprecedented availability of information on any topic. This increased availability however comes at a price. The more information that is available, the more one has to wade through irrelevant or uninteresting content to get to the relevant information. This is called the 'Information Overload' problem. One of the domains where the effects of information overload are most visible is news consumption [1]. With myriads of topics constantly weaving in and out of importance and many different media sources covering different issues, it is easy to become overwhelmed or occasionally miss highly relevant news stories.

My research is focused on designing a news filtering system that will alleviate information overload. My working prototype is called pakrat and can be found online: <http://www.pakr.at>. Pakrat is an adaptive news reader. It continually learns what topics a user is interested in and updates its topical filters accordingly. This results in personalized news feeds that are customized to individual users' interests.

So how does Pakrat learn a user's interests? The first step is to map each news article to a set of topics that it is about. As users read the news, they will interact with articles they find interesting, for example by clicking on the title-link to go to the full story (Pakrat provides a summary in its feed) or by marking a story 'to-read-later' (A pakrat feature). These types of interactions with a news story indicate user interest and therefore signal pakrat to add the corresponding topics to the user's interest profile. The interest profile is a set of topics that the user has shown interest in, weighted by the extent of that interest.

An important detail was glossed over in the above paragraph. How is a news article mapped to the set of topics that it is about? Pakrat uses a wikifier to do this. Wikifiers are algorithms that given a piece of text, output a set of Wikipedia topics that capture what the text is about. They do this by first detecting all the topics that occur in the text and then pruning the ones that are least related to the other detected topics. The wikifier I use in my current prototype is part of the Wikipedia Miner package developed by Milne and Witten [2]. Using wikifiers in this context is a novel approach. According to a recent literature review [3] wikifiers have not yet been utilized for news recommendation and filtering systems and might have great advantages.

With any information filtering tool there is always the risk of over-customization. This is where, as a result of the tool's adaptation to the user's tastes, the personalized content suffers from lack of diversity and covers only narrow spectrums centered around topics that the user has already shown interest in. Although this helps the user keep track of existing interests, it hinders the discovery of new interesting topics and lessens the diversity of user interests. This phenomenon is known as the filter bubble. To break the bubble and strike a balance between keeping track of existing interests and discovering new ones pakrat employs two simple techniques:

1- Ranking the articles in the user's personalized feed based on recency as opposed to similarity to existing user interests. The latter approach is taken by many news filtering systems but it gives old, enduring user interests an advantage over newly formed ones and is therefore conducive to filter bubbles. A user study is currently in the planning to measure the effects of these different ranking strategies on user interest diversity over time.

2 - Pakrat's user interface employs a two feed structure. One is the personalized feed that is customized to the user's interests. Another is the general feed. This feed is populated with news articles on topics that the user has not shown any previous interest in. The idea here is that while the user is given plenty of interesting news to read via the personalized feed, the presence of the general feed might spark new interests in completely new domains.

The final aim is to create a more efficient and engaging news reading experience that will result in better informed and more engaged citizens.

References:

- [1] Avery E. Holton and Hsiang Iris Chyi. Cyberpsychology, Behavior, and Social Networking. November 2012, 15(11): 619-624. doi:10.1089/cyber.2011.0610.
- [2] Milne, D., & Witten, I. H. (2012). An open-source toolkit for mining Wikipedia. Artificial Intelligence.
- [3] Content-based Recommender Systems: State of the Art and Trends In Recommender Systems Handbook (2011), pp. 73-105, doi:10.1007/978-0-387-85820-3_3
- [4] Pariser, Eli. The filter bubble: What the Internet is hiding from you. Penguin Press HC, 2011