Filter Bubble: How to burst your filter bubble

Sneha Prakash

Asst.Professor
Department of computer science, DePaul Institute of science and Technology,
Angamaly, Ernakulum, Kerala, 683573,
Snehaprakash119@gmail.com

Abstract: *The number of peoples who rely on internet for information seeking is doubling in every seconds.as the information produced by the internet is huge it have been a problem to provide the users relevant information.so the developers developed algorithms for identifying users nature from their clickstream data and there by identifying what type of information they like to receive as an answer for their search. Later they named this process as Web Personalization. As a result of web personalization the users are looped Eli pariser an Internet Activist named this loop as "Filter Bubble". This preliminary work explores filter bubble, its pros and cons and techniques to burst filter bubble.*

Keywords: Filter Bubble, Web Personalization, cookies, clickstream

1. Introduction

Searching web has become an integral part of our life. Now a days the web we use are personalized, This is done by suggesting the user some links, sites ,text ,products or messages.so the user can easily access the information he needs which will provide the user a feel that he is using his personal web. According to [1]Web personalization can be described as any action that makes the Web experience of a user customized to the user's taste or preferences. We can also say that Personalization is a technique where different users receive different results for the same search query. The increasing personalization is leading to concerns about Filter Bubble effects, where certain users are simply unable to access information that the search engines algorithm decides as irrelevant.

2. Web personalization

The web is a huge repository of millions of data .the data includes web pages, links, products etc. Therefore the number of users driven to the web increases day by day. The users like to have a customized/personalized web experience so that they can easily access the web they need. The most used search engines Google and Bing are now using personalized searches. Not only search engines social networking sites-commerce sites etc. are now applying personalization techniques. According to [3] Web personalization can be seen as an interdisciplinary field that includes several search domains from user modeling, social networks, web data mining, human machine interactions to web usage mining. Web personalization is the process of customizing a Web site to the needs of each specific user or set of users, taking advantage of the knowledge acquired through the analysis of the user's navigational behavior [4]. Integrating usage data with content, structure or user profile data enhances the results of the personalization process [5].ultimately we can say that web personalization is done to provide each user their personal web. Web personalization has been recently gaining great momentum in research and in various commercial web applications. One of the interesting applications of

personalization on Web is the recommender systems [6] .the recommender system analyze the users usage patterns using techniques association rules, clustering. These recommender systems actually puts the user in a self-loop so that he can only see information's he likes or of the same view. This self-loops are called as **FILTER BUBBLES**.so we can say that Filter Bubble is a Result of Personalization.

3. Filter bubble

Filter bubble is the result of web personalization and as a result the user get isolated from the views which he did not agree. The problem which lies here is that the user's likes and dislikes are guessed by algorithms which apply personalized search and sometimes relevant information for the user may be hided.
Best examples for personalized search is googles personalized search and Facebooks personalized search.

It was Facebook and Google were the Pariser first noticed the effects of personalization. About Facebook Pariser said that "I noticed one day that the conservatives had disappeared from my Facebook feed", he tells us. "And what it turned out was going on was that Facebook was looking at which links I clicked on, and it was noticing that, actually, I was clicking more on my liberal friends' links that on my conservative friends' links. And without consulting me about it, it had edited them out." [7].

The same kind of editing, Pariser found on Google also: He asked two friends to search for "Egypt" on Google. The results were drastically different: "Daniel didn't get anything about the protests in Egypt at all in his first page of Google results. Scott's results were full of them. And this was the big story of the day at that time. That's how different these results are becoming." [7]

The personalization was not only done by these websites .it was also done by other sites like news portals. Online shopping sites etc.pariser worried about this search because he says there is so many problems which user have to face in personalization.

Among them one problem is the "friendly world syndrome": "Some important public problems will disappear. Few people seek out information about homelessness, or share it, for that matter. In general, dry,

complex, slow moving problems – a lot of the truly significant issues – won't make the cut." [8] .This problem can be related to issue: "Instead of a balanced information diet, you can end up surrounded by information junk food." [7]

However, at the center of all these tendencies there is one effect that Pariser calls "the you loop:" "The bubble tends to highlight the information that conforms to our ideas of the world which is easy and pleasurable; consuming information that challenges us to think in new ways or question our assumptions is frustrating and difficult." [7] Personalized filtering directs us towards doing the former. The filter bubble isn't tuned for a diversity of ideas or of people. It's not designed to introduce us to new cultures. As a result, living inside it, we may miss some of the mental flexibility and openness that contact with difference creates." [8].

We can say that filtering is a threat to freedom of users.it will be hard for the users to come out of the filter bubble (the you loop).

4. Detecting filter bubble

The users from all over the world use search engines to search the web. According to the users the result produced by search engines are unbiased and neutral. The filter bubbles present will be producing results which will create separations in society. In The preliminary work [9] the authors studied about whether the filter bubble can be measured and described and is an initial investigation towards the larger goal of identifying how non-search experts might understand how the filter bubble impacts their search results.in this work I will be summarizing the work [9] and its results.

People uses search engines to access information, services, entertainment and other resources on the web. Built on revenue largely derived from advertisements, web search is a multi-billion dollar industry. Many search engines, including the top 2 most popular search engines Google and Bing [10] customize user search results based on user's search history, location, and past click behavior. This is how a filter bubble is generated as termed by Eli Pariser [7]. In the study [9], they investigated whether the filter bubble can be measured and described in a way that might be understood by non-search experts.

4.1 Methodology [9]

To measure the filter bubble, they recruited 20 users from Amazon's Mechanical Turk to conduct five unique search queries. Users were instructed to: 1) copy a search link to their clipboard, 2) open a private browsing window, 3) paste a link in the address bar, and 4) press enter. Instructions were provided on how to open a private browsing window for the web browser the participant was using. As a final step, they asked Participants to copy the HTML results from the "view source" option in their browser, and paste the text into a text box in the main study window. Users repeated this process ten times.

4.2 Data analysis[9]

Search engines often return a different number of results. This could vary depending on user screen real estate, or the speed at which the results are returned. To normalize their dataset, they collected only the first eight results from each search engine, and pruned users who had less than eight

results. And using Kendall Tau Rank Distance (KTD) [11] to count the number of pairwise disagreements between two ranking lists. They also used a weighted clustering coefficient [12].

A low clustering coefficient (0) for a search result means that, within a given search engine, there is little Personalization within the search results, while a high engine personalization. They used the overall average and standard deviation to describe the amount of agreement within a group of search results. A low average suggests minimal disagreement for a given search term, while a high average suggests there is significant disagreement.

4.3 Results [9]

The result shows the amount of clustering that exists within the system for high numbers of people. In figure shown below, they analyzed that there are two clear outliers – people (user identifiers are in node labels) with high KTD from others. These individuals are "bubbled" compared to the rest of their peers. The Figure shows two columns of five search results, the first column being from the Bing search engine and the second being from the Google search engine. By comparing across the columns one can see the effect search engine algorithms have on the same query. By comparing down the columns one can see the effect different queries have on the filter bubble within a given search engine. This allows for cross search engine comparison. For example, the difference in Kendall's Tau Distance values is much larger between asthma and jobs in Google than in Bing. Does this mean that Goggle "turns off" the filter bubble in some searches? If so, is it possible to identify which searches?
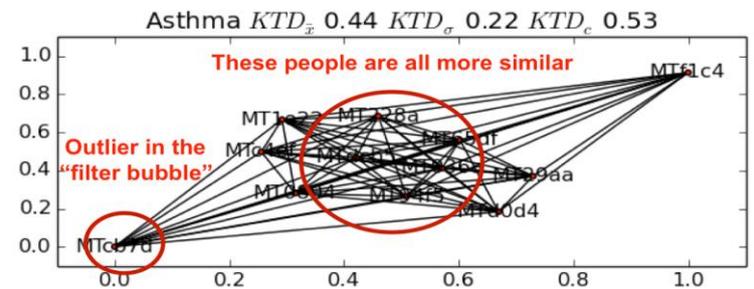

Figure 1: Analysis of search results

5. Bursting Filter Bubble

Filter bubbles or personalization actually aims to help people who deals with large amount of data by sorting the data according to their interest .the problem is that sometimes this may cause over personalization when they try to avoid content overload. Over personalization means that the browser will show data that it thinks user likes. Due to over personalization user have the chance to miss some important information.

There are different ways to burst filter bubble.in this paper I will mention the best 10 ways to burst bubble according to Eli pariser [7].

5.1 methods to pop filter bubble [15]

5.1.1 Burn your cookies

What Are Cookies? What is a Cookie? [13]

Cookies are small files which are stored on a user's computer. They are designed to hold a modest amount of data specific to a particular client and website, and can be accessed either by the web server or the client computer. This allows the server to deliver a page tailored to a particular user, or the page itself can contain some script which is aware of the data in the cookie and so is able to carry information from one visit to the website (or related site) to the next.

Are Cookies Enabled in my Browser?

To check whether your browser is configured to allow cookies, visit the Cookie checker. This page will attempt to create a cookie and report on whether or not it succeeded.

Personalization [14]

Cookies can be used to remember information about the user in order to show relevant content to that user over time. For example, a web server might send a cookie containing the username last used to log in to a website so that it may be filled in automatically the next time the user logs in. Many websites use cookies for personalization based on the user's preferences. Users select their preferences by entering them in a web form and submitting the form to the server. The server encodes the preferences in a cookie and sends the cookie back to the browser. This way, every time the user accesses a page on the website, the server can personalize the page according to the user's preferences. For example, the Google search engine once used cookies to allow users (even non-registered ones) to decide how many search results per page they wanted to see.

The use of cookies for personalization is the reason why we need to burn our cookies. We can disable the cookies in our browsers settings.

5.1.2 Erase your web history

Web history

It is the history of your searches. We want to erase history because it is based on our web searching history the personalization is applied. The personalization websites keep track of our history.

To delete your browsing history. Click the Internet Explorer icon on the taskbar to open Internet Explorer. Click the Tools button, point to Safety, and then click Delete browsing history. Select the types of data you want to remove from your PC, and then click Delete**.**

5.1.3 Tell Facebook to Keep Your Data Private

While using Facebook most people keep our data public. If we keep our data public then whole world can access our data .this data about us will help personalization web sites to track our interest for applying personalization.

So tell Facebook to keep your data private. You can do this in your privacy setting option of Facebook.

5.1.4 it's your birthday you can hide if you want to

You can tell face book to hide your birthday notification because birthday + name is you .birthday notification can help personalization web sites to identify you.

5.1.5 Turn off the target ads and tell sneakers to buzz off

This will disable all the target ads so the ads will stop tracking your interests.

5.1.6. Go incognito

It means go for private browsing. You can either use browsers which especially use for private browsing or you can enable private browsing in your browser.

5.1.7 go anonymous

You can go anonymous using browsers like tor browser [16] or anonymous.com.

5.1.8 Depersonalize your browser

We have to de personalize our browser if we want to avoid personalization because there is a technology called browser finger printing which will help personalization algorithms to identify our patterns.

Browser finger prints [17]

Browser fingerprinting is an increasingly common yet rarely discussed technique of identifying an individual user by the unique patterns of information visible whenever a computer visits a website. The information collected is quite comprehensive and often includes the browser type and version, operating system and version, screen resolution, supported fonts, plugins, time zone, language and font preferences, and even hardware configurations.

These identifiers may seem generic and not at all personally identifying, yet typically only one in several million people have exactly the same specifications as you.

A quick look here (https://panopticlick.eff.org) provides a glimpse of the type of information any website can see about you, and also shines a light on the uniqueness of your individual configuration.

How to disable browser finger prints

Best Method for Avoiding Browser Fingerprinting: use Tor Browser

Great Method: Privacy Badger and Disconnect

The Electronic Frontier Foundation's Privacy Badger extension is a good way to help protect your privacy online, but it won't completely block fingerprinting. It also requires some additional configuration.

Disconnect is a service that blocks most advertising and tracking domains. This extension, in addition to a great ad blocker, will help you block the domains that are trying their hardest to fingerprint and track you. However, it still doesn't have the full effectiveness offered by Tor.

Good Method: Disable JavaScript and Flash

Most fingerprinting methods (at least, deep-level fingerprinting methods) use JavaScript or Flash to get the extra information required to make a more complete fingerprint.

Disabling JavaScript and Flash is a good way to circumvent a lot of (but not all) browser fingerprinting, but unfortunately, remains incomplete. While Flash can usually be disabled just fine without breaking all but the oldest websites, JavaScript remains a key part of many website functions. Disabling JavaScript will negatively impact your browsing experience at some point.

Okay Method: Random User Agent Extensions

While these are supposedly the most effective methods of preventing browser fingerprinting aside from the Tor Browser, there is a reason why these extensions for Chrome and Firefox are ranked so low.

Like JavaScript, using Random User Agent extensions means that on some web pages your browsing experience is going to get broken or otherwise adversely effected.

The reason here is that the extension is reporting false information about your browser. While this is great for preventing you from getting fingerprinted, pages with low compatibility range outside of a single browser can end up giving you some trouble

5.1.9. Tell google and Facebook to make it easier to see and control your filters.

Keeping your Facebook info private is getting harder and harder all the time—mostly because Facebook keeps trying to make it public.so when logging in tell Facebook and google to control your filters

6. Pros and cons of filter bubble

6.1 Pros
• For younger kids they can't be scared of what they don't know
• From the user's point of view, personalization helps cut through content overload and it takes
Users right to content they're likely to want to consume.
• From a company's point of view, it increases the amount of time a user spends on a website, users return more often and higher engagement levels are seen by delivering personalization.

6.2 Cons
• They filter out bad things this is a good and bad thing if they filter out the bad they will never know what is going on in the real world?
• Also filter bubbles block things so for example kidnappings were happening in your neighborhood you wouldn't know if there were filter bubbles
• Lastly if someone had filter bubbles and was doing a current event project it would block all the important things that are happening in the world
• Filter bubble has a hand in your decision. And in turn, shape who you become
• The filter bubble reduces creativity and learning. -

7. Conclusion

Personalization is on the rise: Google showing personalized search results marked the beginning of the era of personalization. Google is not alone: Facebook, Yahoo, Microsoft and others all use personalization. This trend is rising and the big companies predict that practically all information will be personalized in the near future. In this paper I tried to explain what a filter bubble is and how it effects our searches. And also I tried to explain the different methods to pop our filter bubble. Filter bubbles are having both pros and cons but according to me I absorb Eli pariser's opinion that internet websites must not hide anything from users hence it is the right of user to see all the information's even though they are not useful.

References

[1] Masher, B., "Web Usage Mining and Personalization", in Practical Handbook of Internet Computing, M.P. Singh, Editor. 2004, CRC Press. p. 15.1-37.

[2] Bamshad mobasher,Robert cooley,jaideep Srivastava,"Automatic personalization based on web usage mining" August 2000/Vol. 43, No. 8 COMMUNICATIONS OF THE ACM

[3] Beatric,ryan Fernandez,leo j peo,nikhila kamat,sergius Miranda ," New Approaches to Web Personalization UsingWeb Mining Techniques" Beatric et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2195-2201

[4] Eirinaki, M.,Vazirgiannis, M.,"Web mining for web personalization", ACM Transactions on Internet Technology, Vol.31, pp. 1-27, 2003.

[5] Eirinaki, M., Vazirgiannis, M.,Varlamis, I.,"SEWeP:using site semantics and a taxonomy to enhance theWeb personalization process", Proceedings of theninth ACM SIGKDD international conference onKnowledge discovery and data mining, WashingtonDC, USA, pp. 99-108, 2003

[6] Faten Khalil Jiuyong LiHua Wang," Integrating Recommendation Models for Improved Web Page Prediction Accuracy", Conferences in Research and Practice in Information Technology (CRPIT), 2008, Vol. 74.

[7] Pariser, Eli. 2011. Beware online filter bubbles. TED talk http://www.ted.com /talks/eli_pariser_ beware_online _filter _ bubbles.html.

[8] Pariser, Eli. 2011. The Filter Bubble. Viking, London.

[9] Tawanna R. Dillahunt, Christopher A. Brooks, Samarth Gulati, "Detecting and visualizing filter bubbles in Google and Bing"in ACM digital library, http://dl.acm.org/ citation. jcfm?id=2732850 &dl=ACM

&coll=DL&CFID=804595149&CFTOKEN=9646252
4.

[10] Top 15 most popular search engines | January 2015.The eBusiness Guide.http://www.ebizmba.com/articles/search-engines.

[11] Kendall's Tau Distance. (n.d.). In Wikipedia. http://en. wikipedia.org / wiki/ Kendall_tau_distance.

[12] Saramäki, J., Kivelä, M., Onnela, J.P., Kaski, K,Kertesz, J., 2007. Generalizations of the clustering coefficient to weighted complex networks. Physical Review E Stat. Nonlinear Soft Matter Phys. 75 027105. http://jponnela.com/web_documents/a9.pd

[13] http://www.whatarecookies.com/

[14] https://en.wikipedia.org/wiki/HTTP_cookie

[15] https://www.powtoon.com/online-presentation/eDZXU3VIpoA/10-ways-to-pop-your-filter-bubble/

[16] https://www.torproject.org/projects/torbrowser.html.en

[17] http://www.networkworld.com/article/2884026/security0/browser-fingerprints-and-why-they-are-so-hard-to-erase.html

Author Profile

Sneha Prakash Received B.C.A and M.C.A from Mahatma Gandhi University Kottayam in 2009 and 2012 respectively. Presently working as Asst.Professor at DePaul Institute of science and technology angamaly. She is interested in the area of Data mining. Published a paper named Web Personalization using web usage mining: applications, Pros and Cons, Future in the journal IJCSIT in 2015.