On Designing Content Recommender Systems for Online News Media

Abhijnan Chakraborty
Department of Computer Science and Engineering,
Indian Institute of Technology
Kharagpur, India - 721302

CHI'18 Extended Abstracts, April 21–26, 2018, Montréal, QC, Canada.
ACM ISBN 978-1-4503-5621-3/18/04.
http://dx.doi.org/10.1145/3170427.3173036

Introduction

Due to the enormous amount of information being carried over online systems today, no user can access all such information. Therefore, to help the users, all major online organizations deploy information retrieval (content recommendation, search or ranking) systems to find important information. Current information retrieval systems have to make certain design choices. For example, news recommendation systems need to decide on the quality of recommended news stories, how much emphasis to give to a story's long-term importance over its recency or freshness, and so on.

Similarly, recommendation systems over user generated contents (e.g., in social media like Facebook and Twitter) need to take into account the content posted by heterogeneous user groups. However, such design choices can introduce unintended biases in the contents presented to the users. For example, the recommended contents may have poor quality or less news value, or the news discourse may get hijacked by hyper-active demographic groups. In this thesis, we want to systematically measure the effect of such design choices in the content recommendation systems, and build alternate recommendation systems that mitigate the biases in the recommendation output.

Research Goals

1. Limit low quality content from being picked by recommendation systems.

To attract user attention in the crowded online media landscape, some media outlets come up with catchy headlines accompanying low quality articles, which lure users to click on the article links. Such headlines, known as Clickbaits, exploit the cognitive phenomenon *Curiosity Gap* [10], where the headlines provide forward referencing cues to generate enough curiosity among the readers such that they become compelled to click on the link to fill the knowledge gap. Examples of such clickbaits include *"This Rugby Fan's Super-Excited Reaction To Meeting Shane Williams Will Make You Grin Like A Fool", "15 Things That Happen When Your Best Friend Is Obsessed With FIFA"* or *"They Said She Had Cancer. What Happens Next Will Blow Your Mind"*.

These articles tend to attract a lot of users, and thus recommendation systems like Facebook Newsfeed or Twitter Timeline algorithms unwittingly promote such content. However, they often offer little news value, and thereby raise concerns regarding the role of *journalistic gatekeeping* with the prevalence of clickbaits [9]. In this work, we want to build automated classifier to distinguish between clickbait and traditional news headlines, and prevent such content from being picked by the recommendation systems.

2. Understand and address demographic biases in crowdsourced recommendations.

In social media, users are increasingly relying on crowdsourced recommendations called *Trending Topics* [13] to find important events and breaking news stories. Contents selected for recommendation indirectly give the initial users who promoted (by liking or posting) the content an opportunity to propagate their messages to a wider audience. Hence, it is important to understand the demographics of people who make a content worthy of recommendation, and explore whether they are representative of the media site's overall population. More importantly, we want to further explore whether certain demographic groups are systematically *under-represented* among the promoters of the trending topics. Finally, we want to design systems which would reduce the demographic bias, and ensure fairness and transparency in the recommendation outputs.

3. Optimize the recency-relevancy trade-off in online news recommendations.

The selection of 'front-page' stories on online news media sites usually takes into consideration several crowdsourced popularity metrics, such as number of views or shares by the readers [1]. In this work, we focus on automatically recommending front-page stories in such media websites. When recommending news stories, there are two basic metrics of interest: recency and relevancy. Ideally, recommender systems should recommend the most relevant stories soon after they are published. However, the relevancy of a story only becomes evident as the story ages, thereby creating a tension between recency and relevancy. We want to analyze how recommendation strategies in use today tackle this trade-off and propose a recommendation strategy which attempts to optimize on both the axes.

Our Contributions

The progress we have made so far are as listed below.
1. We compared clickbaits and traditional news headlines, and noticed that clickbait headlines use several language traits to attract users. For example, such headlines have more function words, more stopwords, more hyperbolic

words, more internet slangs, and more frequent use of possessive case, as compared to the traditional headlines where the title contains specific proper nouns and the reporting is in third person.

Based on these observations, we developed a clickbait classifier where given a news article headline, the classifier would classify it as clickbait or non-clickbait. Evaluating on a dataset of $15,000$ headlines, we observed around 93% cross-validation accuracy for the classifier. We also found that the headlines the users would like to block vary greatly across users. Hence, we proposed personalized clickbait blocking approaches. We finally built a browser extension, 'Stop Clickbait'[1], which warns the users about the possibility of being baited by clickbait headlines in different websites. The extension also offers the users an option to block certain types of clickbaits she would not like to see during future encounters.

2. Using extensive data collected from Twitter, we quantified the demographic biases in crowdsourced recommendations. Our analysis, focusing on the selection of trending topics, found that a large fraction of trends are promoted by crowds whose demographics are significantly different from the overall Twitter population. We found clear evidence of under-representation of certain demographic groups (female, black, mid-aged) among the promoters of the trending topics, with mid-aged-black-females being the most under-represented group. These observations suggest that the so called 'glass ceiling effect', usually used to describe the barriers that women face at the highest levels of an organization [8], may occur even in crowdsourced recommendations such as Twitter Trends.

[1]Available at chrome.google.com/webstore/detail/stop-clickbait/ iffolf-pdcmehbghbamkgobjjdeejinma

We further discovered that once a topic becomes trending, it is *adopted* (i.e., posted) by users whose demographics are less divergent from the overall Twitter population, compared to the users who were promoting the topic *before* it became trending. Our finding alludes to the influence and importance of trending topic selection on making users aware of specific topics. Therefore, there is a need for making the demographic biases of Twitter trend recommendations transparent. Hence, we developed and deployed a system 'Who-Makes-Trends'[2], where for any trend in the US, one can check the demographics of the promoters of that trend.

3. We analyzed the recency-relevancy trade-offs offered by the news recommendation strategies in use today. Our analysis, using real-world news stories datasets, showed that such strategies lead to poor trade-offs between recency and relevancy in practice. We proposed a simple yet previously overlooked strategy called *Future-Impact*-based recommendations, where news stories are selected based on how many views they are expected to receive in the future (and *not* in the past). Intuitively, future-impact of a story captures the extent to which the story is likely to be discussed in the future, and journalism studies have argued that it is a useful metric for selecting news stories in its own right [11, 12]. Additionally, two properties of the future-impact metric help achieving better trade-offs between recency and relevancy: (i) a highly relevant story has higher future-impact than a non-relevant story, and (ii) news stories have highest future-impact shortly after they are published, i.e., when they are very recent. We developed an optimization framework combining the predicted future-impact of the stories with the uncertainties in the predictions, which achieves good performance benefits.

[2]Available at twitter-app.mpi-sws.org/who-makes-trends

WHAT NEXT?

Our future research includes understanding the consumers in different social media sites who embrace different types of contents, which may help in exploring alternate approaches to limit the demographic biases we identified in crowdsourced recommendations. In addition to demographic biases, we also observed coverage bias [2] and dissemination bias in socially shared news [3]. We'll attempt to limit them in future work. Finally, we have designed recommendation strategies to optimize the recency and relevancy of recommended stories. However, diversity of the stories is another metric of interest. In future, we'll attempt to propose a news recommendation scheme which take recency, relevancy as well as diversity into account.

Related Publications

1. "Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media", in IEEE/ACM ASONAM, San Francisco, USA, August 2016 [6].
2. "Optimizing the Recency-Relevancy Trade-off in Online News Recommendations", in WWW, Perth, Australia, April 2017 [4].
3. "Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations", in AAAI ICWSM, Montreal, Canada, May 2017 [5].
4. "Tabloids in the Era of Social Media? Understanding the Production and Consumption of Clickbaits in Twitter", in ACM CSCW, New York, USA, November 2018 [7].

Acknowledgments: I sincerely thank my PhD supervisors, Prof. Niloy Ganguly (IIT Kharagpur, India) and Prof. Krishna P. Gummadi (MPI-SWS, Germany), for their constant guidance, motivation and support. I am a recipient of Google India PhD Fellowship and Prime Minister's Fellowship Scheme for Doctoral Research, a public-private partnership between Science & Engineering Research Board (SERB), Dept. of Science & Technology, Government of India and Confederation of Indian Industry (CII).

REFERENCES

1. Roja Bandari, Sitaram Asur, and Bernardo A Huberman. 2012. The Pulse of News in Social Media: Forecasting Popularity. In *AAAI ICWSM*.
2. Abhijnan Chakraborty, Saptasrshi Ghosh, Niloy Ganguly, and Krishna Gummadi. 2015. Can Trending News Stories Create Coverage Bias? On the Impact of High Content Churn in Online News Media. In *Computation and Journalism Symposium*.
3. Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2016. Dissemination Biases of Social Media Channels: On the Topical Coverage of Socially Shared News. In *AAAI Intl. Conf. on Web and Social Media (ICWSM)*.
4. Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2017a. Optimizing the Recency-Relevancy Trade-off in Online News Recommendations. In *Intl. Conference on World Wide Web (WWW)*.
5. Abhijnan Chakraborty, Johnnatan Messias, Fabricio Benevenuto, Saptarshi Ghosh, Niloy Ganguly, and Krishna P. Gummadi. 2017b. Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations. In *AAAI Intl. Conference on Web and Social Media (ICWSM)*.
6. Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *IEEE/ACM Advances in Social Networks Analysis and Mining (ASONAM)*.
7. Abhijnan Chakraborty, Rajdeep Sarkar, Ayushi Mrigen, and Niloy Ganguly. 2017. Tabloids in the era of social media? understanding the production and consumption of clickbaits in twitter. *Proc. ACM Human-Computer Interaction* 1, 2 (2017).
8. David A Cotter, Joan M Hermsen, Seth Ovadia, and Reeve Vanneman. 2001. The glass ceiling effect. *Social forces* 80, 2 (2001).
9. Jeffrey Dvorkin. 2016. Column: Why click-bait will be the death of journalism. pbs.org/newshour/making-sense/what-you-dont-know-about-click-bait-journalism-could-kill-you/. (2016).
10. George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116 (1994).
11. Max Novendstern. 2011. Why do we read the news? harvardpolitics.com/online/hprgument-blog/why-bother-to-read-the-news, *Harvard Political Review* (2011).
12. Phillip J Tichenor, George A Donohue, and Clarice N Olien. 1970. Mass media flow and differential growth in knowledge. *Public opinion quarterly* 34, 2 (1970).
13. Twitter. 2010. To Trend or Not to Trend. blog.twitter.com/2010/to-trend-or-not-to-trend. (2010).