

CSE 3504: Project - Part 2

Joseph Sweeney and Soumya Kundu

April 22, 2017

Practice 1:

3. Changing the damping factor from 0.85 to 0.95:

- (a) Increasing the damping factor does change our ranks.
- (b) The new rank of the page that was ranked first when using a damping factor of 0.85 is still first. Therefore, the rank of the top page does not change.
- (c) The new rank of the page that was ranked last when using a damping factor of 0.85 is still last. Therefore, the rank of the last page does not change.

4. Changing the damping factor from 0.85 to 0.50:

- (a) Lowering the damping factor does change our ranks.
- (b) The new rank of the page that was ranked first when using a damping factor of 0.85 is still first. Therefore, the rank of the top page does not change.
- (c) The new rank of the page that was ranked last when using a damping factor of 0.85 is still last. Therefore, the rank of the last page does not change.

Practice 2:

1. The name of the algorithm is *autocomplete*.
2. Autocomplete exists to improve the user's experience when using Google's search feature and increase the efficiency of searching for information using Google. In order to make that possible, autocomplete tries to make the process of searching for information much faster. This is accomplished using an algorithm that performs search prediction.

According to Google, search predictions are possible search terms that one can use that are related to the terms that they're typing and that other people are searching for [1]. As one starts typing up a search using Google, the autocomplete algorithm predicts what the user wants to search and immediately starts suggesting possible searches. These search predictions are generated using a variety of different factors and are based on what the user has already typed.

The autocomplete algorithm looks at the text that has already been typed and then finds the most frequently searched items that contain that text. This information about the most frequent searches come from two sources. One source is the user's search history. If the user has already performed searches containing the current text that they have entered so far, then the algorithm suggests those previous searches, as it is very likely that the user would be searching for the same thing again. Another source is trending stories, which indicates what other people are searching for. Trending stories are popular topics in one's area at the current time and are not related to the user's search history. The intuition here is that if a lot of other people are searching for topics that contain the text that the user has typed in so far, then there is a high probability that the user is also trying to search for one of the same topics.

If the user does not want to search for any of the suggestions presented by autocomplete, then they can simply ignore the suggestions and proceed with their search using the standard process of typing out the entire search text. However, if the user

finds what they wanted to search for within the autocomplete suggestions, which does happen quite often, then they can save a lot of time by simply selecting the suggestion instead of manually typing out their entire search text.

3. The mathematical concept underlying the autocomplete algorithm is Bayes' Theorem. Bayes' Theorem is a very useful and powerful theorem in probability that has numerous applications in various different fields. Bayes' Theorem states that if a complete list of mutually exclusive events B_1, B_2, \dots, B_n have prior probabilities $Pr(B_1), Pr(B_2), \dots, Pr(B_n)$, and if the likelihood of the event A given event B_i is $Pr(A|B_i)$ for each i , then:

$$Pr(B_i|A) = \frac{Pr(A|B_i)Pr(B_i)}{\sum_j Pr(A|B_j)Pr(B_j)}$$

Therefore, the posterior probability of B_i given A , written as $Pr(B_i|A)$, is proportional to the product of the likelihood $Pr(A|B_i)$ and the prior probability $Pr(B_i)$, where the normalizing constant $Pr(A) = \sum_j Pr(A|B_j)Pr(B_j)$ is the prior probability of A [2].

This theorem allows autocomplete to calculate the probability of a certain text being searched conditioned on the text that has already been entered and based on the prior probability of making that search. Once these probabilities are calculated, the algorithm can then suggest the search predictions with the highest probabilities and continuously update the predictions based on new text that the user enters.

4. We can use our knowledge of the autocomplete algorithm to actively modify our startup website in ways that will enhance the ranking of our website in Google's search results. We know that Google's autocomplete algorithm suggests search predictions, which users are very likely to use in order to automatically complete their searches, using publicly available information from trending stories and topics that show us what people are currently searching for most frequently. Therefore, we can look through the most popular topics that are relevant to the content of our website on a weekly

or monthly basis and then incorporate the relevant most frequently searched terms into the text present in our website. This will increase the probability that more users search for terms that place our website higher in the search results, as we are actively selecting our website to appear among the results for the most frequently searched terms.

5. The references for this report are as follows:

- [1] Search using autocomplete. (2017). Retrieved from <https://support.google.com/websearch/answer/106230?hl=en>
- [2] Larget, B. (2008). Bayesian Phylogenetics. Retrieved from <http://www.stat.wisc.edu/~ane/bot940/bayes.pdf>