Self-learning search engines

©
ANNE SCHUTH
PhD student in the Information and Language Processing Systems group at the Informatics Institute, UvA.

→ How does a search engine such as Google know which search results to display? There are many competing algorithms that generate search results, but which one works best? We developed a new probabilistic method for quickly comparing large numbers of search algorithms by examining the results users click on. Our study was presented at SIGIR 2015, the leading international conference on information retrieval, held in Santiago (Chili) last summer.

Interleaving

Developers of web search engines constantly create hundreds of alternative search algorithms, all of which aim to find the best possible match between a user's information need and web pages. It is vital for both the search engine and the user to know which of these algorithms produces the best results. A common way to compare search algorithms is through interleaving, a method whereby the search engine analyses the users' click behaviour to determine a preference between two alternative algorithms. After the user has typed in a query, the unique results of two search algorithms (blue and red in the Figure) are interleaved alternatingly (from top to bottom, and displayed to the user as a single list. If the user then clicks on a result found by one search algorithm (red), the algorithm analysis infers that in this particular case the algorithm generating the selected result produces better results than the other one. By scaling up this type of inference to cover millions of users, the search engine automatically learns which algorithms yield the best results.

Multileaving 1.0

Interleaving is, however, limited by the fact that only two algorithms can be compared at a time, and thousands of comparisons may therefore be required to determine which one of hundreds of existing algorithms really works the most effectively. So-called multileaving methods, which have been developed at the University of Amsterdam, allow multiple algorithms to be compared simultaneously. In earlier work, we did so by combining the results from many lists of results at once (in the example of blue and red lists, imagine also adding orange and green lists, etc.). The multileaved result list that is shown to the user is then a mix of results originating from many search algorithms - a multicoloured list. We keep track of where each of the results came from (their colour), and, as with interleaving, we observe which search algorithm (colour) attracts most clicks from users. Again, the search algorithm that receives most clicks wins. Typically, once this has been established, the search engine will completely switch over to the victorious search algorithm for all its users and queries.

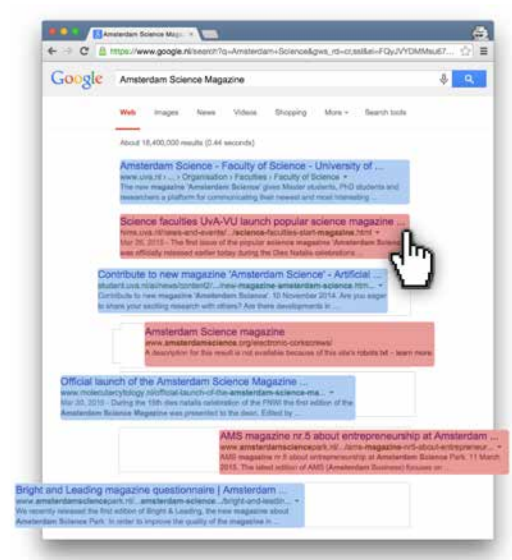
Next step: probabilistic multileaving

Our newest method takes multileaving a step further. While we still combine the results from many search algorithms into a single multileaved result list, we now do so probabilistically. Instead of alternatingly picking results from each of the lists, always working from the top-ranked downwards, we now define a (high) probability that the top-ranked result is picked, leaving a non-zero probability that a lower ranked result is selected instead. By making the multileaved list probabilistic, we ensure that any combination of search algorithms (coloured lists) could have resulted in the multileaved list that is shown to a user. This has the major advantage that we can retrospectively evaluate any search algorithm, using a multileaved result list that has already been shown to a user. In other words, it now becomes possible to reuse old combinations of multileaved result lists and users' clicks to keep evaluating new search algorithms. As can be expected, the search algorithms that originally contributed results to the multileaved result list, or algorithms that are very similar, can be evaluated with higher confidence than very different search algorithms. However, even working at lower confidence levels, it is a major advantage of our probabilistic multileaving method that new search algorithms that were not even invented when the multileaving took place can be evaluated retrospectively. This way, our method can identify the best search algorithms much faster, enabling search engines such as Google to self-improve much more efficiently. Ω

"the 'users' click behaviour determines the preference for the search algorithm"

→ Reference

A. Schuth, R.-J. Bruintjes, F. Büttner, J. van Doorn, C. Groenland, H. Oosterhuis, C.-N. Tran, B. Veeling, J. van der Velde, R. Wechsler, D. Woudenberg, M. de Rijke, Probabilistic Multileave for Online Retrieval Evaluation. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 955-958 (2015).



↑ Figure
An interleaved results list is generated by alternatingly selecting results from the results lists of two different algorithms (highlighted in red and blue).

Link to article:
http://bit.ly/probabilistic-multileave-pdf