Persgroep: Urgent, or can it wait? Personalising push for Algemeen Dagblad

Participants: Marlies van der Wees (company representative), Menno van Zaanen (TU, academic leader), Daniel Berecz, Anne Dirkson, Chris Emmery, Lieke Gelderloos, Chris Kamphuis, Mozhgan Karimi, Radhesh Krishnan Konoth, Tudor Paisa, Gustavo Penha, Susan Vermeer.

Context and problem

Every day, the Algemeen Dagblad (AD) produces over 500 news articles. When the author of an article is finished writing their article, they typically publish it online, through websites and apps. Currently, that is the moment a team of editors decides whether the article should be pushed to our users through an app, highlighting the existence of the article.

These editors have a difficult decision to make, for at least two reasons. The first reason relates to user preferences. The AD app users are currently subscribed to very broad categories for which they wish to receive push notifications. For instance, when users indicate that they want to receive push notifications for “Sport”, they will receive notifications for all sports articles that editors decide should be pushed. This includes articles on “Formula 1” even though the user is only interested in “Soccer”. This problem can be tackled by inferring much more fine grained user preferences from user behaviour and by recommending only those articles that match these inferred preferences.

The second reason relates to the more general question whether an article should be pushed at all. Intuitively, articles are push-worthy because they are urgent and only relevant for a short period of time. For example, for a user interested in “Soccer”, this may be the outcome of a match of their favorite soccer club. In this sense, articles like food recipes or travel stories are not considered to be push-worthy. Deciding whether something is urgent or whether it can wait is a hard and unsolved problem that we want to tackle in this challenge.

The goal of the project is to investigate whether it is possible, given an article (with its metadata), to predict whether the article is push-worthy. In a natural setting, figuring out push-worthiness of an article probably depends on the user (and their preferences) as well. However, we want to separate user preferences from the urgency aspect in this context and only concentrate on push-worthiness.

Research approach

To fully understand what the notion of push-worthiness really entails, several directions can be taken. To get a better idea of the current process of deciding whether an article should be pushed or not, we took a qualitative approach. The decision to push a particular article is made by “pushers”, people who are heavily involved with the publishing process. By interviewing these people, we may gain insight in the reasoning behind each decision. Next, we may computationally model these decisions. To investigate the performance of computational implementations of the push decision process, we took two quantitative approaches. First, we analyze the available dataset that consists of a set of articles. For each article, the article text and corresponding metadata as well as the information on whether the article had been pushed is given. This provides information on the distribution of pushed articles related to several metadata properties. Even though the pushers may be able to describe their reasoning behind push decisions, there may be underlying patterns in their behavior that they do not realize personally. Finally, computational models that learn from the dataset are built. A wide range of features, properties from the dataset, can be extracted from the dataset and be used as the input to machine learning methods. These methods can be used to predict aspects of push-worthiness. By applying the computational models on unseen data (of which the true answers are known), we can measure the performance of the system by comparing the output of the computational models against the true information from the dataset.

Results

To fully understand the reasoning behind the process of pushing certain articles, we talked to a "pusher" who was willing to answer a range of questions on the pushing process. It turns out that the decision of pushing is very intuitive. The general reason for deciding when to push an article depends on common sense. A wide range of topics can be pushed, national and international news, show news, sports, also regional news. There are essentially three reasons why articles are pushed: important news that many people should know, warnings (e.g., weather, transport), and notifications that something important is going to happen. The idea is that readers will then go to the website at the time of the event itself.

The actual factors that "pushers" use to decide on pushing an article are difficult to define. It requires a "news feeling". Factors like impact and (geographic) distance are used, but also topics that people talk about (for example, at the coffee machine). Sometimes "pushers" are discussing whether to push an article or not. If in doubt, "pushers" typically push these articles as more pushing is considered better. There is no upper limit on the number of articles that can be pushed per day.

With respect to timing, pushing an article as soon as possible is best, in particular when the news is also available for other news providers. As different news providers are direct competitors of each other, being a bit late with pushing an article now and then is not too bad, but this should not happen too regularly.

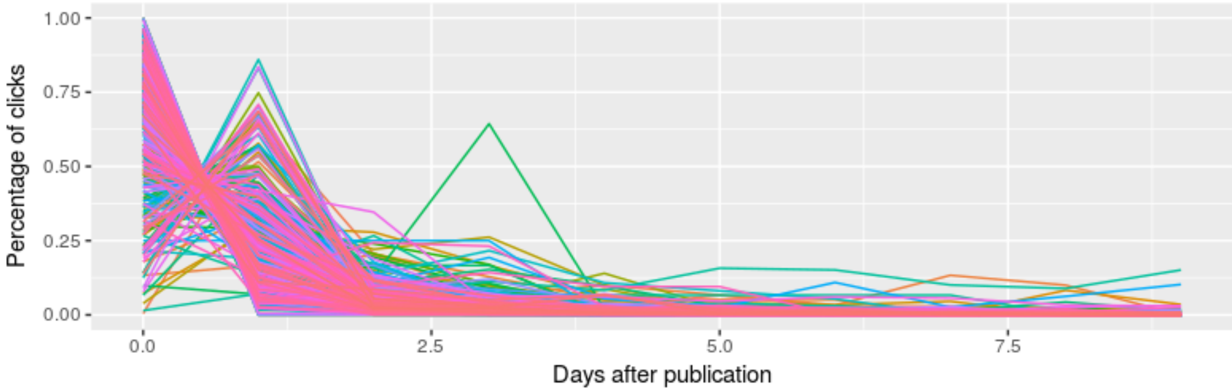
When asking the "pushers" on the usefulness of automating the push process, they expected it to help for personalization (so they do not expect help with the actual push decisions for articles). At the moment, users can only indicate that they are interested in a few very large topics. If selected, they will receive all push notifications for these topics. If the push process is automated, then the topics may be more fine-grained.

In addition to the qualitative results from the interview of the "pusher", we looked at the available dataset. This dataset consisted of three parts: the text part, containing the title and text of news articles, meta-data, such as named entities, IPTC tags, sentiment and objectivity, complexity, date and time of publishing. The click part that contains information on how often articles are clicked on (i.e., being red). Each click is represented by article id, and date and time information. The pushed part is a list of article ids that indicate that these articles have been pushed.

Within the text part, we see that articles from 18 distinct newspapers can be found. Of these, three are national newspapers, the rest regional. There are over 2.2 million articles in this part, of which 25% are published in the AD, 7% in Trouw, and 0.6% in Volkskrant. For all of these articles, we have information on over 3.5 billion clicks.

Unfortunately, we only had information on the articles that were pushed from mid-August until the end of January. The push information came from the national newspapers, so we decided to concentrate on the three national newspapers and this date restriction only. This left us with over 48,000 articles and 324 million clicks.

To get an idea of when people click on articles, we investigated the distribution of clicks over time. The results can be seen in the figure, which shows the percentage of all clicks over time (per day) for a range of articles. Overall, a consistent pattern is found with only a few outliers (such as an article that peaks on the third day). Some articles peak on the second day, which may have to do with the time of publication (at the end of the day).



Looking at the information of the pushed articles, we see that only 770 articles of the 48,000 articles in the dataset have been pushed. This information, combined with the apparent randomness of the selection by the “pushers” and the fact that pushed articles are clicked on more frequently (pushed articles receive around five times more clicks) , we decided to change the task from predicting push-worthiness to predicting the number of expected clicks for an article. This property may be related to the decision of pushing an article or at least provide useful information for “pushers” when they need to make that decision.

Finally, we built a range of computational regression models that predict the number of clicks based on properties from the article. We implemented feature extraction methods for existing meta-data from the dataset, such as complexity, topics, entities, sentiment, but also occurrence counts, TF*IDF weights, and word embeddings of words in the articles and titles. Using these features, we ran a range of regressors (linear regression, support vector regression, random forest) and we compared all these results against a baseline system. The baseline always predicted the most often occurring click count throughout the dataset.

To evaluate the performance of the systems, we used mean absolute error. The performance of the baseline for AD is 11520 (meaning that on average, this model is 11520 clicks off from the correct amount). For Trouw, the baseline mean absolute error is 3181. When running a support vector regressor with a linear kernel using only word embeddings of the title leads to mean absolute errors of 6321 and 3024 for AD and Trouw respectively. A support vector machine with an RBF kernel, using sentiment, complexity, topics, entities, and a range of title properties, mean absolute errors of 6287 and 3014 for AD and Trouw are found.

Future work

For future work, a range of directions can be considered. Firstly, what needs to be investigated is the relation between the click prediction and the push-worthiness of an article. This requires a more indepth analysis of the process of the “pushers” as well as a computational comparison that shows the relationship between clicks, predicted clicks, and whether the “pusher” would push the article.

Secondly, more computational approaches can be taken. For example, we implemented a novel feature that measures the similarity of the article under consideration and the recently published articles. This could be used to measure recurring topics (on-going news) or novel news topics. The impact of this metric is currently not yet investigated. Obviously, other features can be considered as well and these can be used in alternative classifiers or regressors as well. This may give us more information on the effectiveness or usefulness of these features.

Finally, this work has been done to support the push process (or the “pushers”). We can only understand how the “pushers” will use this information and how it will impact their work or their working process once the models can be used. For this, an actual working system with user interface will need to be developed. Based on, the system can then also be expanded into the direction of personalization. This allows for more fine-grained push messages to be sent.