Explaining Rankings
M.A. ter Hoeve

Abstract

Machine learning algorithms have become more complex over time and therefore it has become more difficult to understand the underlying decisions of these algorithms. In this research we investigate the explainability of ranking algorithms. In particular we focus on the ranking algorithm of Blendle, an online news kiosk that uses a ranking algorithm to make a personalized selection of news articles from a wide variety of newspapers for their users.

From a user study on 541 Blendle users we learn that users would like to receive explanations for their personalized news ranking, however, they do not show a clear preference as to how these explanations should be shown. Supported by these results we design LISTEN, a model-agnostic LISTwise ExplaNation method to explain the decisions of any ranking algorithm. Our method is model-agnostic, because it can be used to explain any ranking algorithm without the need to add additional information about the specific algorithm. Our method is listwise, because it explains the importance of features to the ranking by taking the influence of the features on the entire ranking into account. For rankings, existing pointwise approaches, where the importance of a feature is calculated by only looking at its influence on the item score, are not faithful. This is because the position of an item in the ranking is not only defined by its own score, but also by the score of the other items in the ranking. The new listwise approach is an important contribution of this work. The importance of features is found by gradually changing feature values and computing the effect on the entire ranking. The main intuition behind this approach is that if perturbations to features are able to change the ranking a lot, these are important features. If the perturbation of a feature does not change the ranking, this feature is not important for the ranking.

In order to allow our explanation model to run in production, where it needs to compute explanations for news articles on the fly, we implement two steps to increase the speed. First we divide the process of changing the feature values in two parts. In the first part we find the most disruptive feature values. In the second part we use only these values to find the most important features. As a second speed up, we train a neural network on the data that we made in the previous step. In production we only use the neural network to compute the explanations. We call this method Q-LISTEN. This speed up is another important contribution of this work.

We compare LISTEN and Q-LISTEN with two baselines: the already existing Blendle reasons (these are heuristic and therefore unfaithful explanations of the underlying ranking algorithm) and the reasons produced by LIME (Ribeiro et al., 2016), a local, pointwise explanation method. An offline evaluation shows that LISTEN produces faithful explanations and that the two speed up steps barely decrease the accuracy of the model. A large-scale online evaluation on all Blendle users who receive a personalized news selection shows that the type of explanation does not influence the number of reads of the users, which indicates that even though users find it important to receive explanations, they are less sensitive to the faithfulness of these explanations.