Machine Learned Ranking

Information Retrieval, Extraction & Integration

Course 2022/23

Víctor Morcuende Castell

Guillermo Nájera Lavid

1. Introduction

It is clear that Machine Learned Ranking (MLR) has gained a significant amount of attention during the last years, due to its application in information retrieval, e-commerce, healthcare and more. In this sense, the purpose of MLR approaches is to optimize the ranking of search results based on user queries.

In this report, we will explain first the type of MLR approach selected (in our case, the listwise), how the training dataset has been built to fulfill the 3 queries provided (Glucose in Blood, Bilirubin in Plasma and White Blood Cells Count) and finally how the implementation of an AdaRank algorithm model has been created, along with its performance.

2. Type of MLR approach

The judgement that has been chosen for the completion of this assignment is based on a multi-level rating or listwise approach. The AdaRank algorithm is a machine learning technique used for ranking that is based on the AdaBoost algorithm, which is an ensemble learning algorithm which combines weak learners to create a strong one. Regarding the ranking, the AdaRank algorithm works by training multiple weak ranking models on the training dataset and combining them to create a strong ranking model.

Therefore, the goal was to use the AdaRank algorithm to create a list of documents (or in this case markers) ordered by relevance based on a certain query (as previously said, the "Glucose in Blood", "Bilirubin in Plasma" and "White Blood Cells Count" queries). With respect to the relevance, we decided to create 4 different rating levels: High > Mid > Low > 0.

3. Building the appropriate training set

The idea was to classify the degree of relevance for each marker in each query. To do so, we made use of the SearchLOINC tool, to get a better understanding of each marker and the correlations it could potentially have with the query marker. Once the SearchLOINC tool was used, we added the relevance rating for each marker, combining the information recently acquired and our knowledge/judgement, carefully justifying the election of the relevance rating under the "comments" column in the loinc_dataset-v2 excel file. Therefore, for better understanding of our reasoning regarding the rating classification please look at the referenced column.

Furthermore, for the task of extending the datasets and consequently improving the model's performance, we decided to use the 20 first results that appeared in SearchLOINC when the queries are being searched.

4. Implementing the model

Once the training dataset was completed/classified, the implementation of the model started.

To implement the AdaRank algorithm, we first loaded the training dataset for the 3 queries ("Glucose in Blood", "Bilirubin in Plasma" and "White Blood Cells Count"). After that, we performed several data pre-processing operations, such as separating the "relevant" column from each query to be the target label for the model as well as dropping irrelevant information (being precise, the comments). To conclude this phase, the remaining features were one-hot encoded for the purpose of properly preparing them to be fed to the model.

Then, the AdaRank model was trained on each dataset using a DecisionTreeRegressor as the base estimator and 100 estimators. Finally, we evaluated the performance of the model on the test set by applying the NDGC score, which is a quality measure that considers both the relevance of each item and its position in the ranking, assigning higher scores to models that place the most relevant items at the top of the ranking. This was used to determine if the model had reached an NDCG score of 0.9 or higher, meaning that its performance would be reasonably adequate. Once we got to this stage, the model was considered trained, and the number of iterations required to reach the target score were shown.

Moreover, after this whole process we decided to add the extended datasets to the model. For achieving this, the only thing that we had to do was repeating the same steps that we performed for the initial implementation of the model, but in this case, for the extended datasets.

5. Results

As it can be appreciated in the code file, we printed the results for each query using the NDCG score. Looking at them, generally the results for the "Glucose in Blood" query achieve the 0.9 NDCG score much faster than the others, remarking that usually "Bilirubin in Plasma" achieves them faster than "White Blood Cells Count".

Moreover, when we did the same with the extended datasets, and as it was supposed, the results from the extended code were substantially better than the ones from the initial ones, also maintaining the general achievement rate tendency.

6. Conclusion

To conclude this report, we will summarise the main tasks performed: first, we applied the listwise MLR approach, particularly, the AdaRank algorithm, to a dataset of medical markers for three queries: glucose in blood, bilirubin in plasma, and white blood cells count. Moreover, we performed the same process to an extended dataset. After that, we showed how the AdaRank algorithm is significantly useful for machine learning ranking tasks, and how to build a simple yet effective appropriate training set, along with the implementation the corresponding model. Finally, our results on the test set demonstrated the effectiveness of the AdaRank algorithm for these tasks related to medical markers.