1. TFIDF model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Baseline** | **P@10** | **Recall** | **MAP** | **NDCG5** | **MRR** |
| B1 - VSM, no pre-processing | 0.01 | 0.04 | 0.01 | 0.01 | 0.01 |
| B2 - VSM, english stopwords | 0.01 | 0.04 | 0.01 | 0.01 | 0.01 |
| B3 - VSM, lowercase words | 0.07 | 0.19 | 0.04 | 0.06 | 0.01 |
| B4 - VSM, bigrams and unigrams | 0.02 | 0.04 | 0.01 | 0.01 | 0.01 |
| B5 - VSM, only bigrams | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 |
| B6 - VSM, lemmatization, lowercase | 0.07 | 0.18 | 0.05 | 0.08 | 0.01 |
| B7 - VSM, lowercase, bi and unigrams | 0.06 | 0.19 | 0.04 | 0.06 | 0.01 |

These three were chosen, because they have the best metrics. For these three baselines, precision-recall curves were created to compare these models:

Chart

Description automatically generated

1. LMJM model

To choose a lambda value, 10 experiments were run. Basing on these experiments, a graph of P@10 and recall vs lambda was created. The model in these experiments did not use any preprocessing.

Chart, scatter chart

Description automatically generated

The best value of lambda seems to be between 0.3 and 0.4 – combining the highest P@10 with a pretty high recall. To confirm this, another experiment has been run – for the lambda values between 0.3 and 0.4. The results are below:

Chart, scatter chart

Description automatically generated

Using this graph, we can choose lambda = 0.32. This is reasonable because we are dealing with long queries.

Now, this is a place for a table with the same baselines as in TFIDF but with LMJM model. Three best baselines will be used to plot precision-recall curves. Basing on these curves, the best baseline can be chosen.

**ALL THE LISTED METRICS WERE OBTAINED FOR THE TRAINING DATASET**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| B1 - LMJM, no pre-processing | 0.02 | 0.04 | 0.01 | 0.01 | 0.01 |
| B2 - LMJM, english stopwords | 0.02 | 0.04 | 0.01 | 0.01 | 0.01 |
| B3 - LMJM, lowercase words | 0.06 | 0.19 | 0.04 | 0.06 | 0.01 |
| B4 - LMJM, bigrams and unigrams | 0.02 | 0.05 | 0.01 | 0.01 | 0.01 |
| B5 - LMJM, only bigrams | 0.02 | 0.05 | 0.01 | 0.01 | 0.01 |
| B6 - LMJM, lemmatization, lowercase | 0.06 | 0.18 | 0.05 | 0.08 | 0.01 |
| B7 - LMJM, lowercase, bi and unigrams | 0.05 | 0.18 | 0.04 | 0.05 | 0.01 |

Chart

Description automatically generated

1. Comparison of the models

Here, I would evaluate the best baselines of both models. I would plot the precision-recall curves obtained by these baselines on a test set. I think it’s also reasonable to calculate all the required metrics obtained by these two baselines on the test set. We can present that in a table and discuss the differences.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model, baseline** | **P@10** | **Recall** | **MAP** | **NDCG5** | **MRR** |
| LMJM, B6 | 0.05 | 0.11 | 0.04 | 0.07 | 0.01 |
| VSM, B6 | 0.06 | 0.14 | 0.03 | 0.08 | 0.01 |

Chart

Description automatically generated