Best Matching Algorithm to Identify and Rank the Relevant Statutes

Kayalvizhi S[a], Thenmozhi D[b] and Chandrabose Aravindan[c]

[a]SSN College Of Engineering, Chennai
[b]SSN College Of Engineering, Chennai
[c]SSN College Of Engineering, Chennai

Abstract

Automated identification of relevant documents or statutes for the given query is a time efficient process. Artificial Intelligence in Legal Assistance task (AILA) task is about using the artificial intelligence to identify the suitable prior documents or statute for the given query and semantic segmentation of the legal documents. We have identified the suitable statute for the given query using Best Matching (BM-25) ranking algorithm. We have calculated the scores using BM25 algorithm and sorted the statutes according to the scores and then identified the suitable statutes with better MAP score of 0.2975 than most of the other approaches on AILA@FIRE-2020 dataset.

Keywords

Artificial Intelligence, Best Matching, Identify, Ranking, Sorting, Statute

1. Introduction

Referring the prior cases is highly desired to in order to give judgement to the case to ensure justice. Identifying the suitable prior cases and statutes from the vast number of documents is a time consuming process for which artificial intelligence can be made use for effective identification. Artificial Intelligence for Legal Assistance (AILA) task of Forum for Information Retrieval Evaluation (FIRE) 2020 [1] is a series of shared tasks from 2019. AILA@FIRE-2019 [2] is the task of identifying statutes and cases whereas AILA@FIRE-2020 is with additional task of classifying the statements into semantic segments.

Different approaches like unsupervised approach of vectorizing followed by ranking using cosine similarity [3], cosine similarity with embeddings [4, 5, 6], information retrieval model [7], BM25[8], auto summarization[9], etc. have been reported for retrieving the prior documents and statute on AILA@FIRE-2019 dataset.

2. Task Description

Artificial Intelligence in Legal Assistance task is about identifying the relevant cases or statutes for the given query and semantic segmentation. There are two tasks namely Task 1: Precedent

Forum for Information Retrieval Evaluation 2020, December 16-20, 2020, Hyderabad, India

AILA@ FIRE-2020		Training set	Testing set
Task 1	No. of queries	50	10
	Object case_docs	2914	2914
	Object_statutes	197	197
Task 2	No. of case documents	50	10

Table 1
Data set description

& Statute retrieval and Task 2: Rhetorical Role Labeling for Legal Judgements. Task 1 has sub-tasks with Task 1a for identifying the precedent case document and Task 1b for identifying the relevant statutes. Task 2 is the classification task of classifying the legal statements in one of the seven semantic segments/rhetorical roles.
The seven classes include:
Facts : the events that led to the case.
Ruling by Lower Court : the decision of the lower court
Argument : arguments
Statute : relevant prior statute
Precedent : relevant prior case
Ratio of the decision : statement given for judgement
Ruling by Present Court : the decision of the Supreme court

3. Dataset Description

The AILA@FIRE-2020 data set has 2914 case documents and 197 case statutes from which the relevant case documents and statutes has to be identified. The training set had 50 queries along with their relevant case documents and statutes. 10 queries were given as a testing set for identifying the relevant case document and statutes. For Task 2 of Rhetorical Role Labeling for Legal Judgements, 50 case documents along with their seven labels of rhetorical roles were given as training set and the test set is of 10 case documents which has to be classified according to the semantic segments which is explained in Table1.

4. Proposed Methodology

We have presented a ranking algorithm for the task 1b of identifying the relevant statutes among the tasks provided by the Artificial Intelligence for Legal Assistance. BM25 [5] is the Best Matching ranking algorithm that ranks the documents according to its relevance to the given search query. It is a bag-of-words retrieval function where the documents are ranked on the basis of appearance of query terms in each document.

4.1. Task 1b: Identifying relevant statutes - BM25

In this approach, the statutes are pre-processed, vectorized and ranked according to the relevant queries.

4.1.1. Pre-processing

All the given statutes are written in a single file by removing the characters that are not letters or numbers, the stop words in the NLTK corpus stop words and converting the words in the sentences to a lower case.

4.1.2. Ranking

For ranking the statutes, initially a word dictionary is built with the words of the statutes except “for, a, of, the, and, to, in" and its frequency. Then, the document length, average document length , inverse term frequency are also calculated to calculate the BM25 score. The default parameters are taken which is 1.5 for k (saturation parameter) and 0.75 for b(length parameter) to calculate the score.

4.1.3. Sorting

After calculating the score of individual statute document vs queries, the statutes are sorted according to their scores.

5. Results

The results of Task 1b of identifying the relevant statutes is shown in Table 2. The performance of our team is denoted by SSN_NLP. Our approach has attained a MAP score of 0.2975, BPREF score of 0.2531, recip_rank of 0.4769 and P@10 score of 0.15. Our approach performed better than other approaches and is comparable to the top ranking approaches.

6. Conclusion

Thus, artificial intelligence can be used for the identification of precedent documents to identify the documents effectively. We have used Best Matching ranking algorithm (BM25) to rank the relevant statute with respect to the given query which have attained better results than most of the other approaches with a MAP score of 0.2975, BPREF score of 0.2531, recip_rank of 0.4769 and P@10 score of 0.15. The performance can further be improved by altering the values of hyper parameters such as saturation parameter and length parameter.

Acknowledgments

We would like thank the Department Of Science and Technology (DST)-SERB funding scheme and HPC laboratory for providing the resources and space for our research.

Team Name	MAP	BPREF	recip_rank	P@10
SSN_NLP	***0.2975***	***0.2531***	***0.4769***	***0.15***
SSNCSE_NLP	0.3423	0.136	0.3423	0.07
IMS_UNIPD	0.3383	0.279	0.5349	0.17
Uottawa_NLP	0.2506	0.186	0.3144	0.12
UB	0.3134	0.2633	**0.5787**	0.15
LAWNICS	0.2962	0.2812	0.4607	0.13
TUW-informatics	0.2619	0.2033	0.4855	0.13
nlpninjas	0.00917	0.024	0.1204	0.07
scnu_1	**0.3851**	**0.3054**	0.5615	**0.18**
fs_hit_1	0.2139	0.1587	0.3371	0.13
fs_hu	0.235	0.198	0.3581	0.08
fs_hit_2	0.2003	0.1587	0.3452	0.1

Table 2
Final evaluation for Test Data - TASK 1b

References

[1] P. Bhattacharya, P. Mehta, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, Overview of the FIRE 2020 AILA track: Artificial Intelligence for Legal Assistance, in: Proceedings of FIRE 2020 - Forum for Information Retrieval Evaluation, 2020.

[2] P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, P. Mehta, A. Bhattacharya, P. Majumder, Overview of the fire 2019 aila track: Artificial intelligence for legal assistance., in: FIRE (Working Notes), 2019, pp. 1–12.

[3] S. Mandal, S. D. Das, Unsupervised identification of relevant cases & statutes using word embeddings., in: FIRE (Working Notes), 2019, pp. 31–35.

[4] S. Kayalvizhi, D. Thenmozhi, C. Aravindan, Legal assistance using word embeddings., in: FIRE (Working Notes), 2019, pp. 36–39.

[5] S. Robertson, H. Zaragoza, The probabilistic relevance framework: BM25 and beyond, Now Publishers Inc, 2009.

[6] S. Renjit, S. M. Idicula, Cusat nlp@ aila-fire2019: Similarity in legal texts using document level embeddings., in: FIRE (Working Notes), 2019, pp. 25–30.

[7] J. Gao, H. Ning, H. Sun, R. Liu, Z. Han, L. Kong, H. Qi, Fire2019@ aila: Legal retrieval based on information retrieval model., in: FIRE (Working Notes), 2019, pp. 64–69.

[8] B. Gain, D. Bandyopadhyay, A. De, T. Saikh, A. Ekbal, Iitp at aila 2019: System report for artificial intelligence for legal assistance shared task., in: FIRE (Working Notes), 2019, pp. 19–24.

[9] Y. Shao, Z. Ye, Thuir@ aila 2019: Information retrieval approaches for identifying relevant precedents and statutes., in: FIRE (Working Notes), 2019, pp. 46–51.