CuSINeS: Curriculum-driven Structure Induced Negative Sampling for Statutory Article Retrieval

Santosh T.Y.S.S, Kristina Kaiser, Matthias Grabmair

School of Computation, Information, and Technology; Technical University of Munich, Germany

Introduction

Statutory Article Retrieval : Finding relevant statutes for a legal question

- Prior works used BM25, TF- IDF, Indri, Word Movers' Distance, BERT and their ensembles.
- Recently Dense retrieval methods [1], synthetic query generation and legal domain-oriented pre-training [2]
- Graph neural networks to enrich article representations [2] - Using interdependencies among articles within the topological structure of legislation
- A key aspect that has been overlooked : **How to construct high-quality negative samples for training SAR models ?**
- Prior works rely on BM25-based semantic similarity to derive hard negatives.

Contributions

CuSINeS, a model-agnostic negative sampling to train dense retriever for SAR.

- A curriculum-based scheduling of negative samples
- Utilizes topological structure of legislation to mine hard negatives
- Dynamically assess semantic difficulty with the retrieval model being trained.

Our Method: CuSINeS

- **Semantic-based ranking:**
- Unlike static model independent BM25, we dynamically compute semantic difficulty of negative articles using the model being trained
- Captures model's learning dynamics
- **Structure-based ranking:** Two views based on structure
- Hierarchical view: Difficulty of each negative article is measured by its proximity to the set of positive articles within the hierarchical graph.
- Sequential View: Treats statutes as a linearized sequence. Calculates the distance by considering the relative position in the sequential enumeration of articles.
- Semantic difficulty captures the interplay between queries and negative articles
- Structural difficulty reflects the relationship between positive and negative articles,
- To capture the complementary information, unify these rankings through reciprocal rank fusion (RRF).
- **Curriculum Scheduler:**
- Based on cumulative difficulty ranking, categorize negatives into various difficulty-level buckets, ranging from easy to difficult.
- In initial training, more samples from the easier buckets with lesser from the difficult ones.
- As training progresses, the ratio gradually shifts, with more of difficult samples

Dataset & Metrics

- BSARD dataset [1] : 1108 french legal questions, Corpus of 22,600 Belgian legal articles. A query can have multiple relevant legal articles.
- Metrics: Recall@k (R@k) (k=100,200,500), Mean Average Precision (MAP), Mean R-Precision (MRP)

Base Models

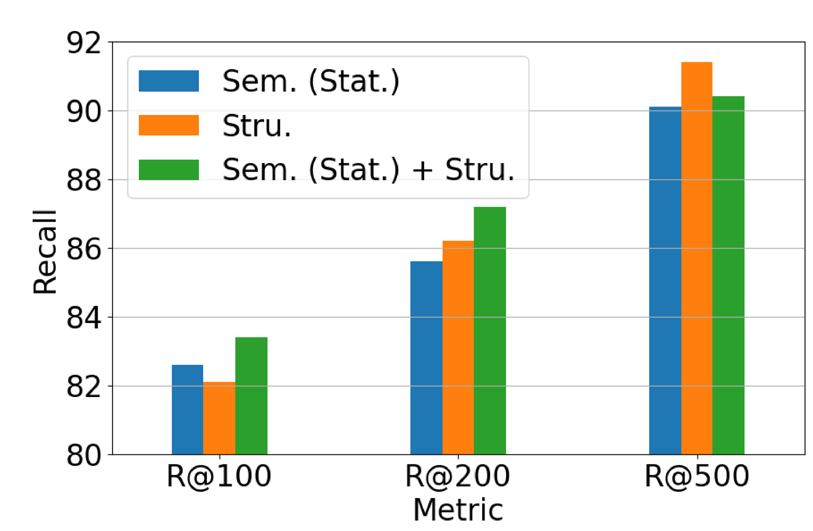
- BM25 [4]
- Dense Retrieval (DR) model: Query encoder - BERT, Article encoder - Hierarchical version of BERT
- DR+GNN [2]: DR model with Graph Attention Network to enrich article representations by using legislative graph topology constructed from hierarchical organization of statutes.
- Two initializations of DR: CamemBERT [3] and LegalCamemBERT [1]

Results

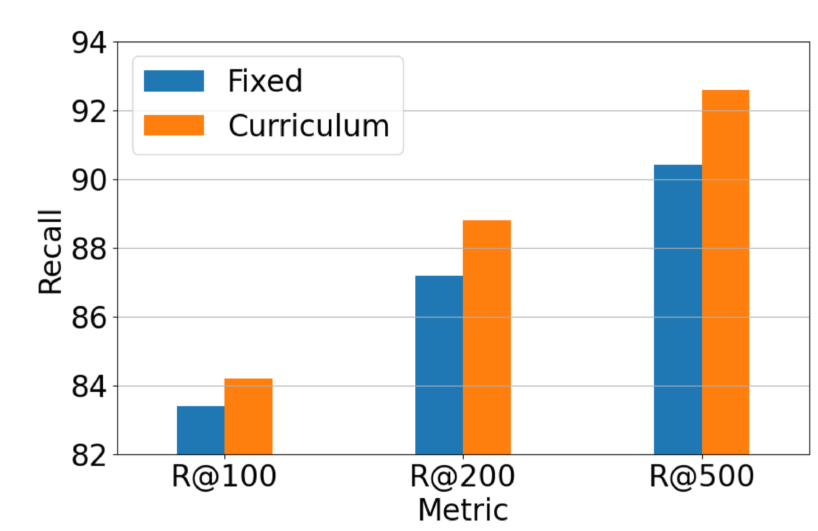
Method		**R@**			**MAP**	**MRP**
		100	**200**	**500**		
BM25	Baseline	49.3	57.3	63	16.8	13.6
DR CB	Baseline	77.1	81.8	86.7	35.6	28.8
	CuSINeS	82.6	86.6	91.6	38	29.1
DR+GNN CB	Baseline	80.2	83.2	88.6	39.2	32.6
	CuSINeS	83.2	88.1	92.6	42.2	33.4
DR LCB	Baseline	79.8	83.9	88.9	39.5	31.3
	CuSINeS	83.7	87.5	92.3	41.2	32.1
DR+GNN LCB	Baseline	82.6	85.6	90.1	44.6	35.8
	CuSINeS	84.9	89.6	93.3	46.2	36.2

Table 1: Comparison of CuSINeS with Baseline negative sampling strategy on four dense models. (L)CB denote (Legal)CamemBERT as encoder model.

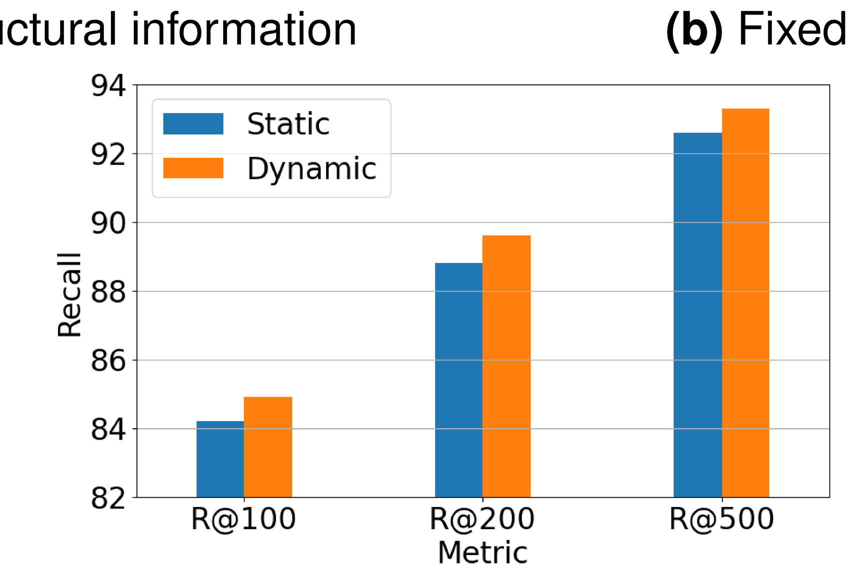
CuSINeS improvement can be attributed to (i) Legislative Structure topology based negative mining (ii) Curriculum-based negative schedule (iii) Dynamic criterion of semantic-based difficulty ranking



(a) Influence of Structural information



(b) Fixed vs Curriculum



(c) Semantic: Static vs Dynamic

Figure 1: Analysis of sub-components of CuSINeS

- Structure-based negatives are more informative than semantic ones with improvements on R@k at higher k-values.
- Combining both validates their complementary nature at lower k-values.
- Easy-to-difficult curriculum helps to learn coarse-grained distinctions between the articles initially and progressively move towards finer-grained nuances.
- Computing difficulty dynamically helps to design adaptive curricula based on its competence at the current training step.

References

[1] A. Louis and G. Spanakis. A statutory article retrieval dataset in french. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6789–6803, 2022.

[2] A. Louis, G. Van Dijck, and G. Spanakis. Finding the law: Enhancing statutory article retrieval via graph neural networks. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2753–2768, 2023.

[3] L. Martin, B. Muller, P. O. Suarez, Y. Dupont, L. Romary, É. V. De La Clergerie, D. Seddah, and B. Sagot. Camembert: a tasty french language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, 2020.

[4] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, M. Gatford, et al. Okapi at trec-3. *Nist Special Publication Sp*, 109:109, 1995.