Improving Sentence Retrieval from Case Law for Statutory Interpretation

Jaromir Savelka[1a,2], Huihui Xu[1a], Kevin D. Ashley[1a,1b,1c]

[1a]Intelligent Systems Program, [1b]Learning Research and Development Center, [1c]School of Law, University of Pittsburgh PittTweet

[2]Gravity Stack LLC gravitystacklaw

jas438@pitt.edu

ICAIL 2019, Montreal
June 18, 2019

This work was supported in part by a National Institute of Justice Graduate Student Fellowship (Fellow: Jaromir Savelka) Award # 2016-R2-CX-0010, "Recommendation System for Statutory Interpretation in Cybercrime," and by a University of Pittsburgh Pitt Cyber Accelerator Grant entitled "Annotating Machine Learning Data for Interpreting Cyber-Crime Statutes."

Presentation Overview

Motivation

> 29 U.S. Code 203 - Definitions
> "Enterprise" means the related activities performed (either through unified operation or common control) by any person or persons for a common business purpose, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

Suppose there is a Thai restaurant in one part of the city and an Indian restaurant in another part both having a single owner.

Are these restaurants an "enterprise" within the meaning of the definition?

Motivation

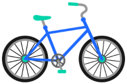
> No vehicles in the park.

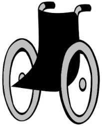
Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

$\Rightarrow$

Legislators use vague,[1] open textured terms,[2] abstract standards,[3] principles, and values.[4]








When there are doubts about the meaning of the provision they may be removed by interpretation.[5]

1. Endicott 2000; 2. Hart 1994; 3. Endicott 2014; 4. Daci 2010; 5. MacCormick & Summers 1991.

Motivation

Interpretation involves an investigation of how the term has been referred to, explained, recharacterized or applied in the past.

“ Example Uses of the Term

i. Any mechanical device used for transportation of people or goods is a vehicle.
ii. A golf cart is to be considered a vehicle.
iii. To secure a tranquil environment in the park no vehicles are allowed.
iv. The park where no vehicles are allowed was closed during the last month.
v. The rule states: “No vehicles in the park.” ”

Going through the sentences is labor intensive because many sentences are useless and there is a large redundancy.

29 U.S. Code 203 - Definitions

"Enterprise" means the related activities performed (either through unified operation or common control) by any person or persons for a common business purpose, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

List of Interpretive Sentences

The "common business purpose" requirement is not defined in the Act.

Appellants common "business purpose" is the operation of an institution primarily engaged in the care of the sick or aged.

The utilization of a common service does not by itself establish a common business purpose shared by the owners of separate businesses.

The common business purpose of this enterprise was framing construction in the construction of single and multi-family homes.

The Fifth Circuit has held that the profit motive is a common business purpose if shared.

Presentation Overview

Task

Given a statutory provision, user's interest in the meaning of a phrase from the provision, and a list of sentences ...

... we would like to rank more highly the sentences that elaborate upon the meaning of the statutory phrase of interest, such as:

- definitional sentences (e.g., a sentence that provides a test for when the phrase applies)
- sentences that state explicitly in a different way what the statutory phrase means or state what it does not mean
- sentences that provide an example, instance, or counterexample of the phrase
- sentences that show how a court determines whether something is such an example, instance, or counterexample.

Presentation Overview

Motivation

Task

Statutory Interpretation Data Set

Experiments
- Direct Retrieval
- Smoothing with Context
- Query Expansion
- Novelty Detection
- Compound Models

Conclusion

Statutory Term Interpretation Data Set

Court decisions are an ideal source of sentences interpreting statutory terms.

For our corpus we selected three terms from different provisions of the United States Code:

1. "independent economic value" (18 U.S. Code § 1839(3)(B))
2. "identifying particular" (5 U.S. Code § 552a(a)(4))
3. "common business purpose" (29 U.S. Code § 203(r)(1))

For each term we have collected a set of sentences by extracting all the sentences mentioning the term from the court decisions retrieved from the Caselaw access project data.[1]

In total we assembled a small corpus of 4,635 sentences.

1. The President and Fellows of Harvard University 2018 (`https://case.law`)

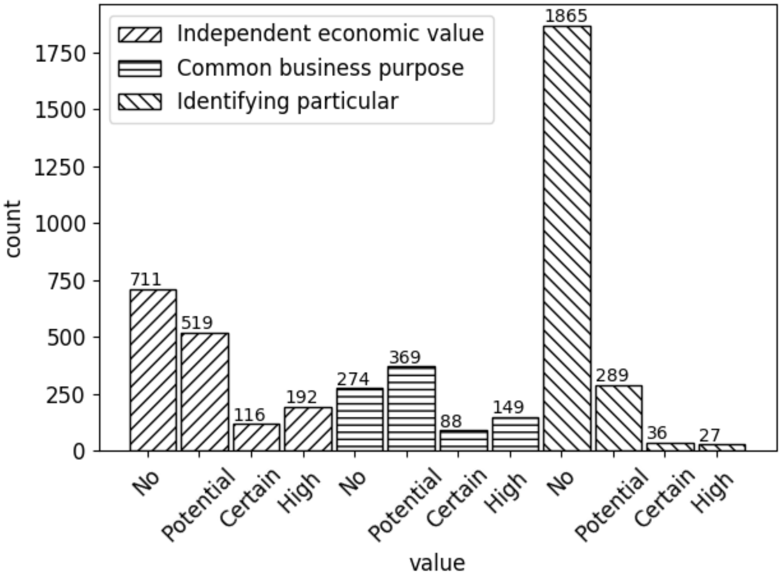
Statutory Term Interpretation Data Set

Three annotators (authors) classified the sentences into four categories according to their usefulness for the interpretation:

1. high value – sentence intended to define or elaborate on the meaning of the term
2. certain value – sentence that provides grounds to elaborate on the term's meaning
3. potential value – sentence that provides additional information beyond what is known from the provision the term comes from
4. no value – no additional information over what is known from the provision

inter-annotator agreement (alpha): .79

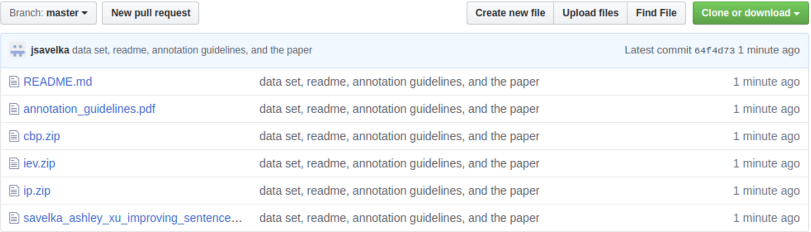
Statutory Term Interpretation Data Set



Statutory Term Interpretation Data Set

The complete data set including the annotation guidelines is publicly available.

https://github.com/jsavelka/statutory_interpretation



Presentation Overview

Research Questions

Is it possible to retrieve sentences directly by measuring similarity between the query and a sentence?

Does utilization of sentences' contexts improve performance?

Does query expansion improve performance?

Does novelty detection improve performance?

Could the context-aware methods be integrated with query expansion and novelty detection to yield an even better model?

Evaluation

We use normalized discounted cumulative gain (NDGC) to evaluate the performance of different approaches.

We chose to evaluate the rankings at $k = 10$ and 100.

$$NDGC(S_j, k) = \frac{1}{Z_{jk}} \sum_{i=1}^{k} \frac{rel(s_i)}{log_2(i+1)}$$

$$rel(s_i) = \begin{cases} 3 & \text{if } s_i \text{ has high value} \\ 2 & \text{if } s_i \text{ has certain value} \\ 1 & \text{if } s_i \text{ has potential value} \\ 0 & \text{if } s_i \text{ has no value} \end{cases}$$

Retrieving Sentences Directly

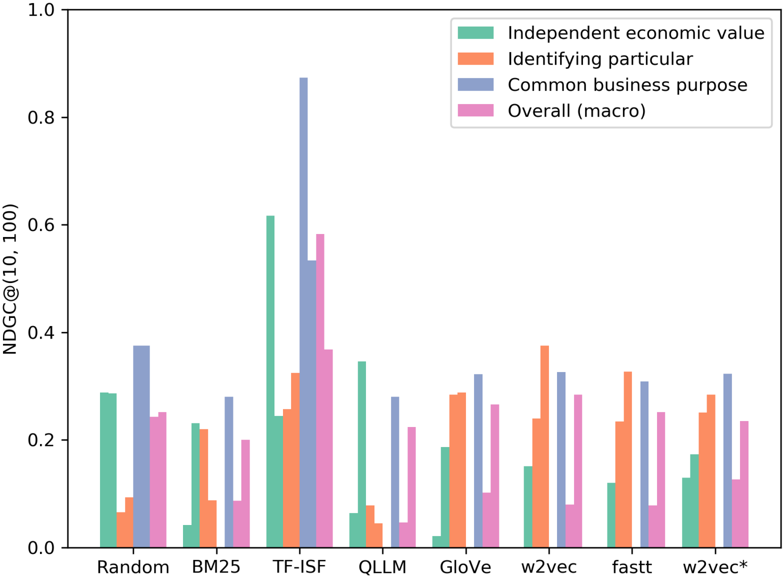
Based on computing similarity between the the term of interest and sentences mentioning it.

Using different strategies for measuring similarity of a sentence and a query, sentences are ranked from the most similar to the least:

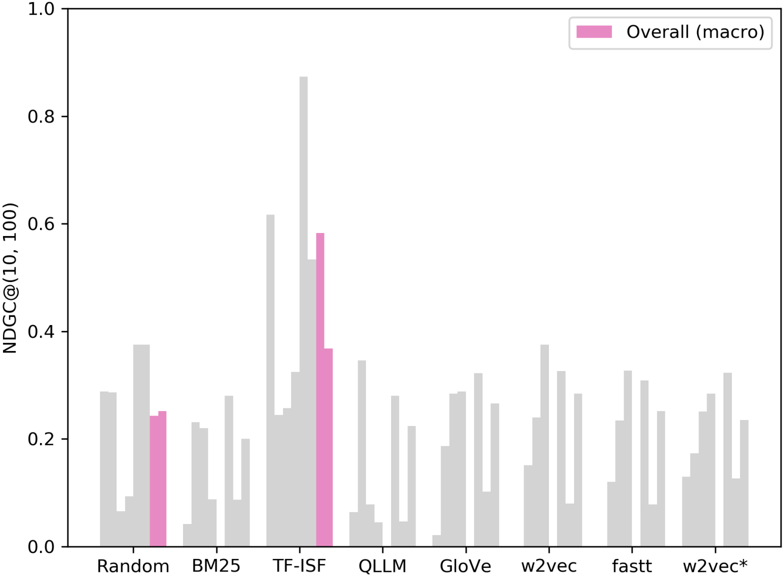
- Okapi BM25[1]
- TF-ISF[2]
- Query likelihood language model[3]
- Cosine similarity between word embeddings
  - word2vec[4]
  - GloVe[6]
  - FastText[7, 8]

1. Manning et al. 2008; 2. Allan et al. 2003; 3. Ponte & Croft 1998; 4. Mikolov et al. 2013; 6. Pennington 2014; 7. Bojanowski et al. 2017; 8. Joulin et al. 2016a.

Retrieving Sentences Directly



Retrieving Sentences Directly



Retrieving Sentences Directly

The root cause of the low performance appears to be the preference of the systems for short over long sentences, e.g.:

> *The "Independent Economic Value" Test*
>
> *It was Altavion's burden to show independent economic value.*
>
> *But again, the Producers' evidence of independent economic value was more theoretical than real.*
>
> *As the formula for an unreleased product, it has independent economic value.*

The relative success of the TF-ISF method could be explained by the deliberate omission of the normalization based on a document length in the TF part.

$$\text{TF-ISF} = \sum_{t \in q} log(tf_{td} + 1) \cdot log\left(\frac{N+1}{\frac{1}{2} + df_t}\right) \cdot log(tf_{tq} + 1)$$

Smoothing Sentences with Context

Based on considering the context of a sentence when deciding about its ranking.

The sentences are ranked based on their similarity to the query as well as the similarity of their varying contexts to the query.

We experiment with the following context sizes:

- whole case
- opinion
- paragraph
- recursive[1]

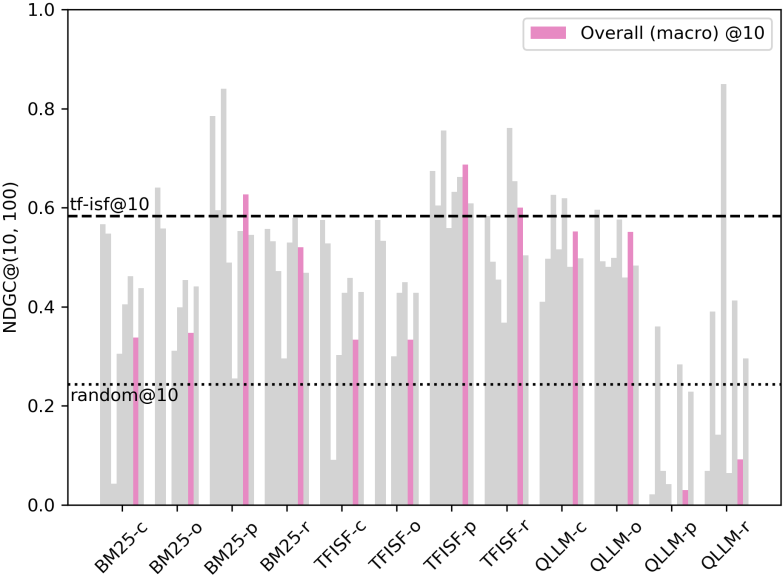
We use the following approaches to incorporate the context:

- linear interpolation
- recursive interpolation[1]

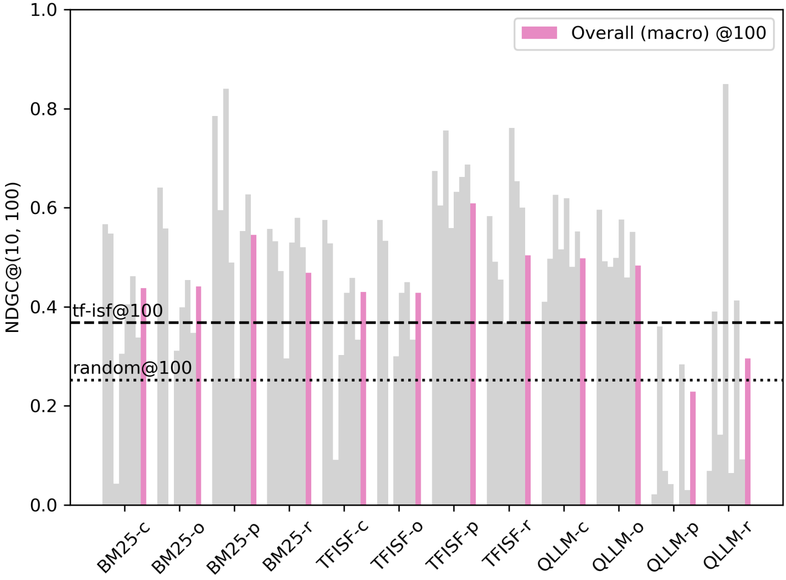
1. Doko et al. 2013.

Smoothing Sentences with Context

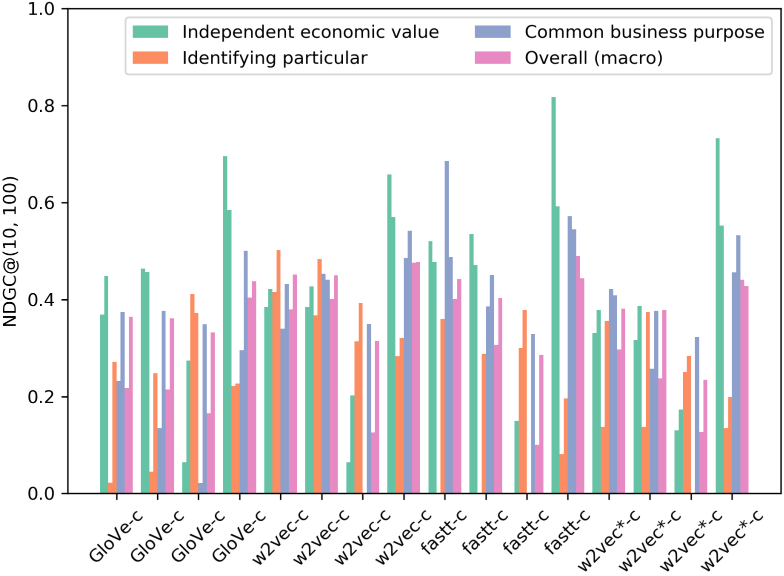
Smoothing Sentences with Context



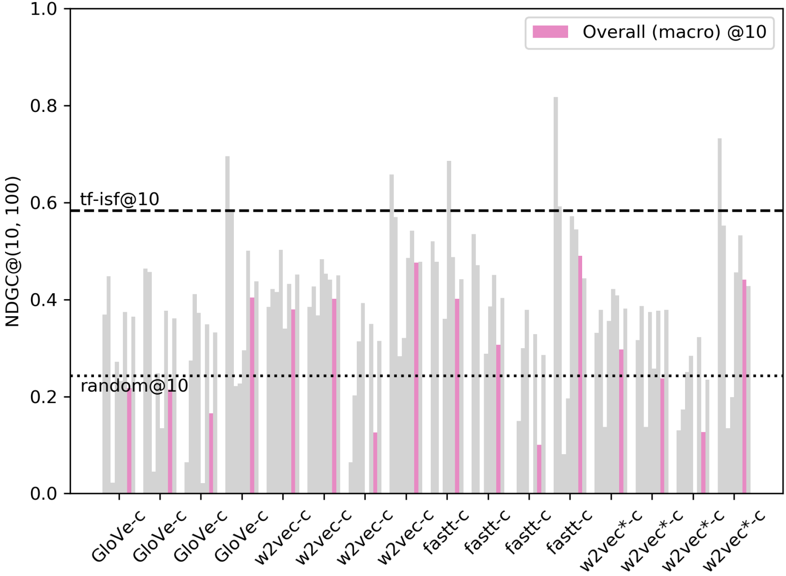
Smoothing Sentences with Context



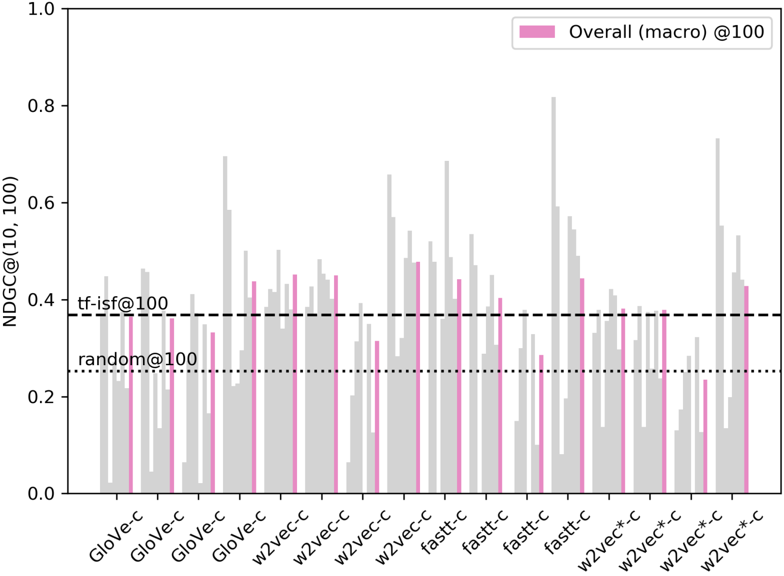
Smoothing Sentences with Context



Smoothing Sentences with Context



Smoothing Sentences with Context



Smoothing Sentences with Context

The "context-aware" models perform surprisingly well, especially the two from the TF-IDF family.

There are two systematic problems:

1. A verbatim citation of the source provision that is included in a local context that heavily discusses the term.
2. Cases that frequently mention the term which comes from a different domain.

Methods operating on smaller contexts are more susceptible to problem 1; those using larger contexts are more harmed by issue 2.

Query Expansion

We extend the query with:

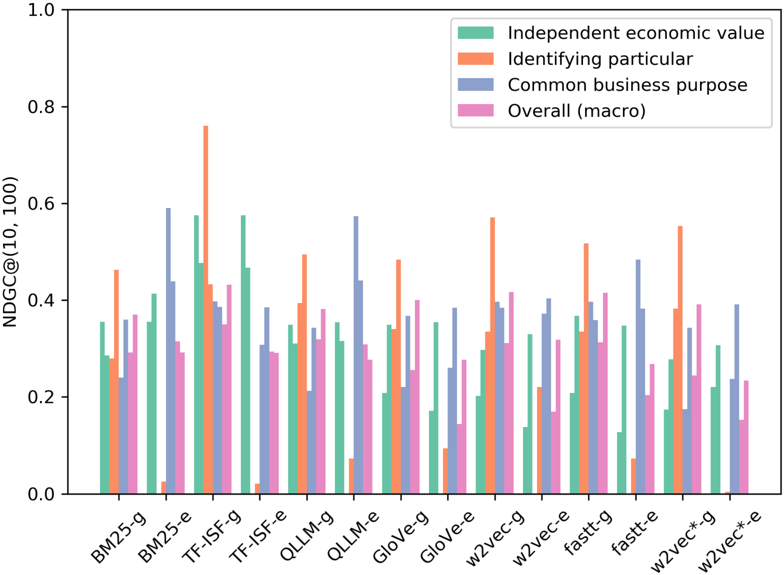
- other words from the source provision on top of those that are part of the term of interest
- the top 500 most similar words produced by the word2vec embeddings trained on our corpus

Measuring similarity between sentences and expanded queries turned out not to be effective.

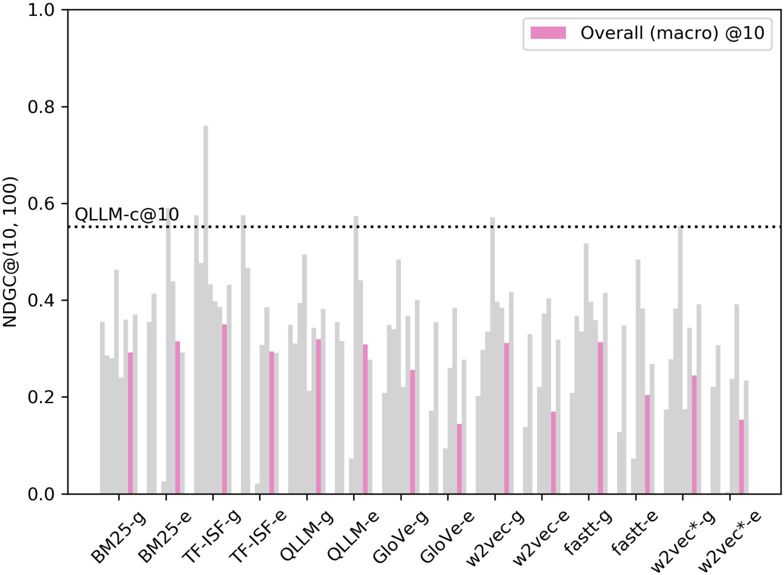
Comparing expanded queries to the whole cases identifies decisions that are focused on:

- the source provision (in case the query is expanded with the words from the source provision)
- the term of interest (in case the query is expanded with the top 500 similar words)

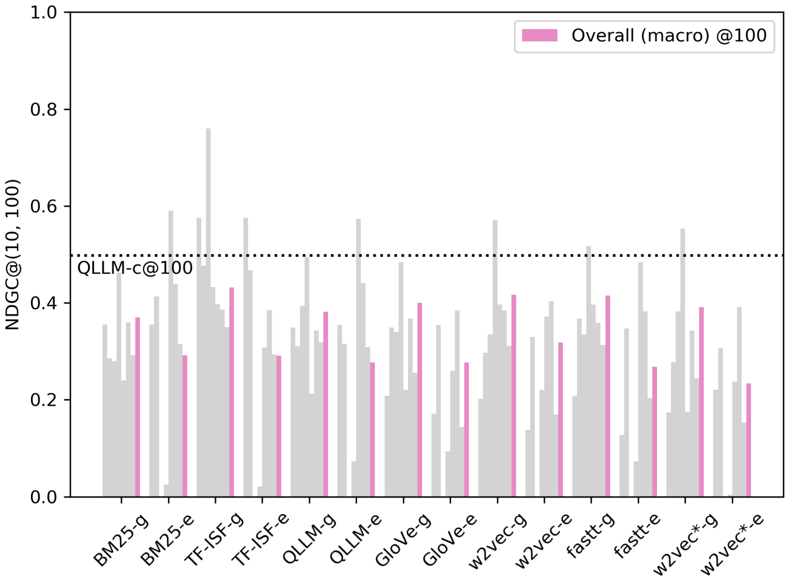
Query Expansion



Query Expansion



Query Expansion



Expanding Query with the Provision

Expanding the query with the source provision or similar words generated by the word2vec model does not improve performance.

A model based on the expansion with the words from the source provision prefers sentences that cite pieces of the provision.

As for expansion with the most similar words the poor performance is surprising since the similar words look promising:

> *overall, profitability, competitive, ... revenue, projection, status*

The models have an interesting property of preferring cases that discuss the source provision or the actual term. (domain problem)

Novelty Detection

Based on measuring the amount of information a sentence provides over what is known from the provision.

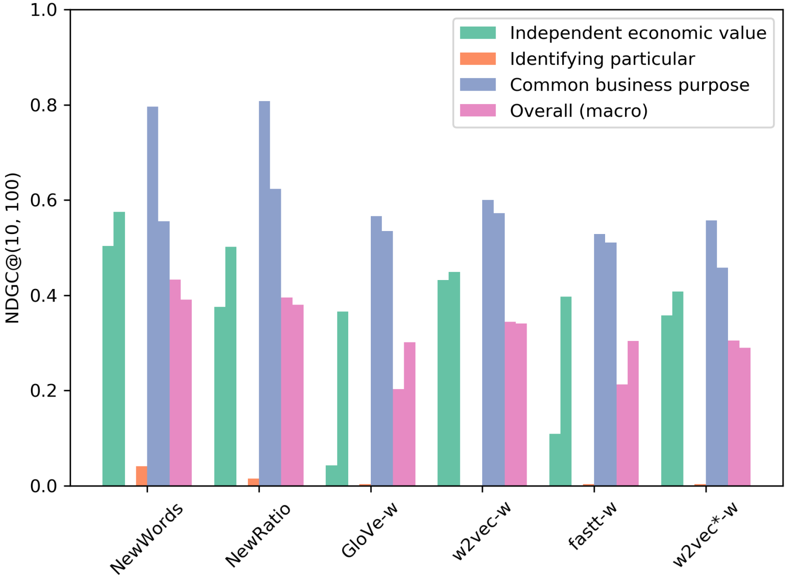
The focus on novelty is interesting because sentences that do not provide additional information are deemed as having no value.

We use the following approaches to measure novelty:

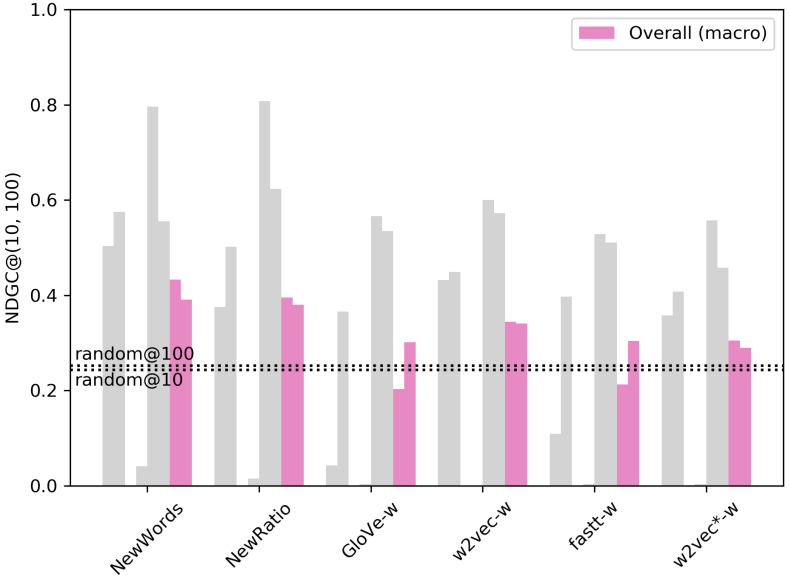
- number of new words the sentence includes over the provision[1]
- ratio of the new words in a sentence to the sentence's length[1]
- Word Mover's Distance (WMD) on the embedding of a sentence and the provision[2]

1. Allan et al. 2003; 2. Kusner et al. 2015.

Novelty Detection



Novelty Detection



Novelty Detection

The models based on novelty likely benefit from focusing on one aspect of the relevance definition and handle that aspect well.

The novelty models have an interesting property of identifying sentences that are verbatim citations of the source provision.

(verbatim citations problem)

Compound Models

We compute scores based on the most successful models that:

1. utilize the sentence's context
2. work with an expanded query
3. measure the sentence's novelty with respect to the provision

The context-aware models are improved with signals coming from the models based on query expansion and novelty detection.

We propose a task specific sentence retrieval model that has the following general form:

$$\text{CMP-rank}(q, s, sp, \mathbf{C}) = \textit{Sim-i}(q, s, \mathbf{C}_i) \cdot \textit{DDI}(q, \mathbf{C}_j) \cdot \textit{NI}(s, sp)$$

Compound Models (Example of Tp–Tg–Nr)

$$\text{Tp-Tg-Nr} = \textit{TF-ISF-p} \cdot \textit{TF-ISF-g} \cdot \textit{NWR}$$

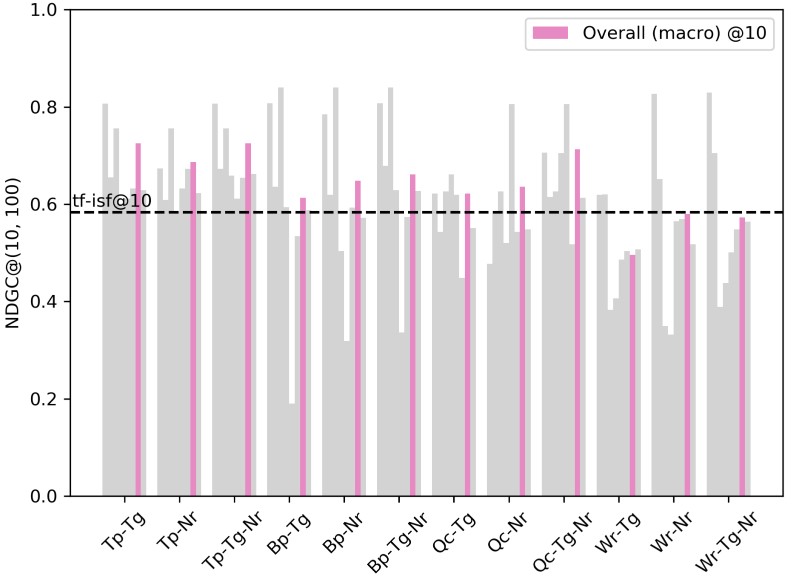
$$\text{TF-ISF-p} = \sum_{t \in q} \left[(1 - \lambda_1) \cdot log(tf_{ts} + 1) \cdot log\left(\frac{S+1}{\frac{1}{2} + sf_t}\right) + \right.$$

$$\left. + \lambda_1 \cdot log(tf_{tp} + 1) \cdot log\left(\frac{P+1}{\frac{1}{2} + pf_t}\right)\right] \cdot log(tf_{tq} + 1)$$

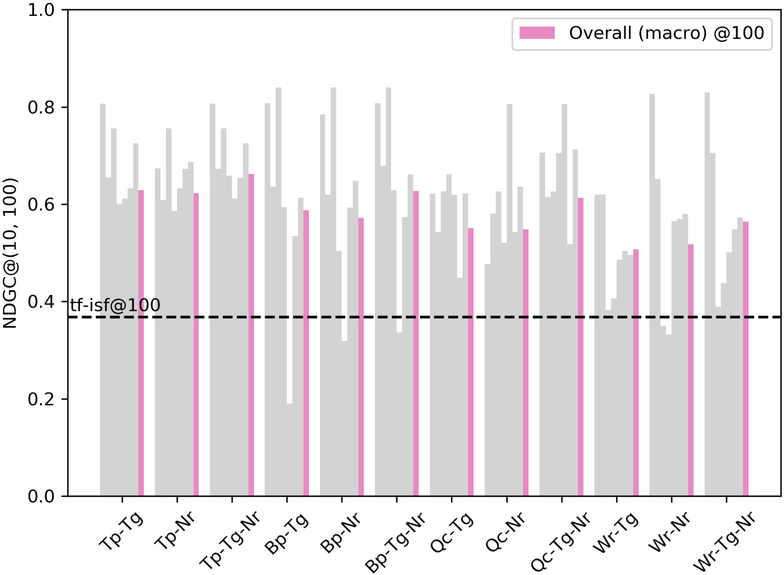
$$\textit{TF-ISF-g} = \begin{cases} 1 & \text{if } \sum_{t \in g} log(tf_{tc} + 1) \cdot log\left(\frac{C+1}{\frac{1}{2} + cf_t}\right) \cdot log(tf_{tg} + 1) \geq \lambda_3 Avg_{10} \\ 0 & \text{if } \sum_{t \in g} log(tf_{tc} + 1) \cdot log\left(\frac{C+1}{\frac{1}{2} + cf_t}\right) \cdot log(tf_{tg} + 1) < \lambda_3 Avg_{10} \end{cases}$$

$$NWR = \begin{cases} 1 & \text{if } \frac{NW}{|\{w_i | w_i \in \mathbf{S}_j|} \geq \lambda_2 \\ 0 & \text{if } \frac{NW}{|\{w_i | w_i \in \mathbf{S}_j|} < \lambda_2 \end{cases}$$

Compound Models


Overall (macro) @10
tf-isf@10
NDGC@(10, 100)
1.0
0.8
0.6
0.4
0.2
0.0
Tp-Tg
Tp-Nr
Tp-Tg-Nr
Bp-Tg
Bp-Nr
Bp-Tg-Nr
Qc-Tg
Qc-Nr
Qc-Tg-Nr
Wr-Tg
Wr-Nr
Wr-Tg-Nr

Compound Models



Independent economic value

High	[...] testimony also supports the independent economic value element in that a manufacturer could [...] be the first on the market [...]
High	[...] the information about vendors and certification has independent economic value because it would be of use to a competitor [...] as well as a manufacturer
Certain	[...] the designs had independent economic value [...] because they would be of value to a competitor who could have used them to help secure the contract
Potential	Plaintiffs have produced enough evidence to allow a jury to conclude that their alleged trade secrets have independent economic value.
Certain	Defendants argue that the trade secrets have no independent economic value because Plaintiffs' technology has not been "tested or proven."

Compound Models

Identifying particular

High	In circumstances where duty titles pertain to one and only one individual [. . .], duty titles may indeed be "identifying particulars" [. . .]
Potential	Appellant first relies on the plain language of the Privacy Act which states that a "record" is "any item . . . that contains [. . .] identifying particular [. . .]
High	Here, the district court found that the duty titles were not numbers, symbols, or other identifying particulars.
Potential	[. . .] the Privacy Act [. . .] does not protect documents that do not include identifying particulars.
High	[. . .] the duty titles in this case are not "identifying particulars" because they do not pertain to one and only one individual.

Common business purpose

High	[. . .] the fact of common ownership of the two businesses clearly is not sufficient to establish a common business purpose.
Potential	Because the activities of the two businesses are not related and there is no common business purpose, the question of common control is not determinative.
High	It is settled law that a profit motive alone will not justify the conclusion that even related activities are performed for a common business purpose.
High	It is not believed that the simple objective of making a profit for stockholders can constitute a common business purpose [. . .]
High	[. . .] factors such as unified operation, related activity, interdependency, and a centralization of ownership or control can all indicate a common business purpose.

Presentation Overview

Future Work

We plan to significantly increase the size of the data set.

We plan to utilize features such as:

- a presence of a reference to the source provision
- syntactic importance of the term of interest
- structural placement of the sentence
- attribution

A deeper semantic analysis of the sentences (perhaps, focused on finding typical patterns) appears to be a promising path forward.

term_of_interest "(such as" _ ")"

Conclusion

We performed a study on a number of retrieval methods for the task of retrieving sentences for statutory interpretation.

We confirmed that retrieving the sentences directly by measuring similarity between the query and a sentence yields mediocre results.

Taking into account sentences' context turned out to be the crucial step in improving the performance of the ranking.

Query expansion and novelty detection capture information that is useful as an additional layer in a ranker's decision.

We integrated the context-aware ranking methods with the components based on query expansion and novelty detection.

Thank you!

Questions, comments and suggestions are welcome now or any time at jsavelka@gravitystack.com, hux16@pitt.edu, or ashley@pitt.edu.

References I

Kevin D Ashley and Vern R Walker, *Toward constructing evidence-based legal arguments using legal decision documents and machine learning*, Proc. of the 14th International Conference on Artificial Intelligence and Law, ACM, 2013, pp. 176–180.

James Allan, Courtney Wade, and Alvaro Bolivar, *Retrieval and novelty detection at the sentence level*, Proc. of the 26th international ACM SIGIR conference on Research and development in informaion retrieval, ACM, 2003, pp. 314–321.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov, *Enriching word vectors with subword information*, Transactions of the Association for Computational Linguistics **5** (2017), 135–146.

Gerlof Bouma, *Normalized (pointwise) mutual information in collocation extraction*, Proceedings of GSCL (2009), 31–40.

Hang Cui, Min-Yen Kan, Tat-Seng Chua, and Jing Xiao, *A comparative study on sentence retrieval for definitional question answering*, SIGIR Workshop on Information Retrieval for Question Answering (IR4QA), 2004, pp. 383–390.

Jordan Daci, *Legal principles, legal values and legal norms: are they the same or different?*, Academicus International Scientific Journal **02** (2010), 109–115.

Emile de Maat, Kai Krabben, Radboud Winkels, et al., *Machine learning versus knowledge based classification of legal texts.*, JURIX, 2010, pp. 87–96.

Emile de Maat and Radboud Winkels, *A next step towards automated modelling of sources of law*, Proc. of the 12th ICAIL, ACM, 2009, pp. 31–39.

Alen Doko, Maja Stula, and Darko Stipanicev, *A recursive tf-isf based sentence retrieval method with local context*, IJMLC **3** (2013), no. 2, 195.

References II

Timothy Endicott, *Vagueness in law*, Oxford University Press, 2000.

______, *Law and Language the stanford encyclopedia of philosophy*, `http://plato.stanford.edu/`, 2014, Accessed: 2016-02-03.

Alejandro Figueroa and John Atkinson, *Contextual language models for ranking answers to natural language definition questions*, Computational Intelligence **28** (2012), no. 4, 528–548.

John Rupert Firth, *A synopsis of linguistic theory 1930–1955*, Studies in Linguistic Analysis (1957).

Ronald T Fernández, David E Losada, and Leif A Azzopardi, *Extending the language modeling framework for sentence retrieval to include local context*, Information Retrieval **14** (2011), no. 4, 355–389.

Ingo Glaser, Elena Scepankova, and Florian Matthes, *Classifying semantic types of legal sentences: Portability of machine learning models*, JURIX, 2018.

Zellig S. Harris, *Distributional structure*, WORD **10** (1954), no. 2-3, 146–162.

Herbert L. Hart, *The concept of law*, 2nd ed., Clarendon Press, 1994.

Stefan Höfler, Alexandra Bünzli, and Kyoko Sugisaki, *Detecting legal definitions for automated style checking in draft laws*, Tech. report, Department of Informatics, University of Zurich, 2011.

Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov, *Fasttext. zip: Compressing text classification models*, arXiv preprint arXiv:1612.03651 (2016).

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov, *Bag of tricks for efficient text classification*, arXiv preprint arXiv:1607.01759 (2016).

References III

- Matjaz Juršic, Igor Mozetic, Tomaz Erjavec, and Nada Lavrac, *Lemmagen: Multilingual lemmatisation with induced ripple-down rules*, Journal of Universal Computer Science **16** (2010), no. 9, 1190–1214.
- Klaus Krippendorff, *Computing krippendorff's alpha-reliability*.
- Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger, *From word embeddings to document distances*, International Conference on ML, 2015, pp. 957–966.
- Jörg Landthaler, EGILHMF Scepankova, Ingo Glaser, Hans Lecker, and Florian Matthes, *Semantic text matching of contract clauses and legal comments in tenancy law*, IRIS: Internationales Rechtsinformatik Symposium, 2018.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, *Efficient estimation of word representations in vector space*, arXiv:1301.3781 (2013).
- Saeedeh Momtazi, Matthew Lease, and Dietrich Klakow, *Effective term weighting for sentence retrieval*, International Conference on Theory and Practice of Digital Libraries, Springer, 2010, pp. 482–485.
- C.D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*, Cambridge University Press, 2008.
- D. N. MacCormick and R. S. Summers, *Interpreting statutes*, Darmouth, 1991.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, *Distributed representations of words and phrases and their compositionality*, Advances in neural information processing systems, 2013, pp. 3111–3119.
- Vanessa G Murdock, *Aspects of sentence retrieval*, Tech. report, Massachusetts University Amherst Department of Computer Science, 2006.

References IV

María-Dolores Olvera-Lobo and Juncal Gutiérrez-Artacho, *Question-answering systems as efficient sources of terminological information: an evaluation*, Health Information & Libraries Journal **27** (2010), no. 4, 268–276.

Jay M Ponte and W Bruce Croft, *A language modeling approach to information retrieval*, Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 1998, pp. 275–281.

The President and Fellows of Harvard University, *Caselaw access project*, `https://case.law/`, 2018, Accessed: 2018-12-21.

Jeffrey Pennington, Richard Socher, and Christopher Manning, *Glove: Global vectors for word representation*, Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

Radim Řehůřek and Petr Sojka, *Software Framework for Topic Modelling with Large Corpora*, Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks (Valletta, Malta), ELRA, May 2010, pp. 45–50 (English).

Jaromir Savelka and Kevin D Ashley, *Extracting case law sentences for argumentation about the meaning of statutory terms*, Proceedings of the Third Workshop on Argument Mining (ArgMining2016), 2016, pp. 50–59.

J Savelka and Kevin D Ashley, *Using conditional random fields to detect different functional types of content in decisions of united states courts with example application to sentence boundary detection*, Workshop on Automated Semantic Analysis of Information in Legal Texts, 2017.

Jaromír Savelka and Kevin D Ashley, *Detecting agent mentions in us court decisions.*, JURIX, 2017, pp. 39–48.

References V

Jaromir Savelka and Kevin D. Ashley, *Segmenting u.s. court decisions into functional and issue specific parts*, JURIX, 2018.

Jaromir Savelka, Vern R Walker, Matthias Grabmair, and Kevin D Ashley, *Sentence boundary detection in adjudicatory decisions in the united states*, Traitement automatique des langues **58** (2017), no. 2, 21–45.

Stephan Walter, *Definition extraction from court decisions using computational linguistic technology*, Formal Linguistics and Law **212** (2009), 183.

Vern R Walker, Parisa Bagheri, and Andrew J Lauria, *Argumentation mining from judicial decisions: The attribution problem and the need for legal discourse models*, Workshop on Automated Detection, Extraction and Analysis of Semantic Information in Legal Texts (ASAIL-2015), 2015.

Bernhard Waltl, Florian Matthes, Tobias Waltl, and Thomas Grass, *Lexia: A data science environment for semantic analysis of german legal texts*, Jusletter IT **4** (2016), no. 1, 4–1.

BM25 and TF-ISF

$$\text{BM25} = \sum_{t \in q} TF \cdot IDF \cdot QTF$$

$$\text{TF} = \frac{(k_1 + 1) \cdot tf_{td}}{k_1 \cdot \left(1 - b + b \cdot \frac{L_d}{L_{avg}}\right) + tf_{td}}$$

$$\text{IDF} = log\left(\frac{N - df_t + \frac{1}{2}}{df_t + \frac{1}{2}}\right) \qquad \text{QTF} = \frac{(k_3 + 1) \cdot tf_{tq}}{k_3 + tf_{tq}}$$

$$\text{TF-ISF} = \sum_{t \in q} log(tf_{td} + 1) \cdot log\left(\frac{N + 1}{\frac{1}{2} + df_t}\right) \cdot log(tf_{tq} + 1)$$

Query Likelihood Language Model

$$P(d|q) \propto P(d) \prod_{t \in q} ((1-\lambda)P(t|M_c) + \lambda P(t|M_d))$$

Smoothing Sentences with Context

$$Sim\text{-}i = (1 - \lambda_1) Sim_s + \lambda_1 Sim_i$$

$$\text{BM25-i, TF-ISF-i} = \sum_{t \in q} [(1 - \lambda_1) TF_s \cdot IDF_s + \lambda_1 TF_i \cdot IDF_i] \cdot QTF$$

$$P(d|q) \propto P(d) \prod_{t \in q} [(1 - \lambda_1 - \lambda_2) P(t|M_c) + \lambda_1 P(t|M_i) + \lambda_2 P(t|M_d)]$$

$$\text{COS-i} = (1 - \lambda_1) COS_s + \lambda_1 COS_i$$

$$Sim\text{-}r_s = (1 - \lambda_1) Sim_s + \lambda_1 [Sim\text{-}r_{s-} + Sim\text{-}r_{s+}]$$

Novelty Detection

$$NW = |\{w_i | w_i \in \mathbf{S}_j \setminus \{w_i | w_i \in \mathbf{P}\}\}|$$

$$NWR = \frac{NW}{|\{w_i | w_i \in \mathbf{S}_j|}$$

$$\min_{\mathbf{T} \geq 0} \sum_{i,j=1}^{n} \mathbf{T}_{ij} c(i,j)$$

$$\text{subject to: } \sum_{j=1}^{n} \mathbf{T}_{ij} = d_i \qquad \forall i \in \{1, \ldots n\}$$

$$\sum_{i=1}^{n} \mathbf{T}_{ij} = d'_i \qquad \forall j \in \{1, \ldots n\}$$

Evaluation

$$NDGC(S_j, k) = \frac{1}{Z_{jk}} \sum_{i=1}^{k} \frac{rel(s_i)}{log_2(i+1)}$$

$$rel(s_i) = \begin{cases} 3 & \text{if } s_i \text{ has high value} \\ 2 & \text{if } s_i \text{ has certain value} \\ 1 & \text{if } s_i \text{ has potential value} \\ 0 & \text{if } s_i \text{ has no value} \end{cases}$$

$$NDGC(\mathbf{S}, k) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} NDGC(S_j, k)$$

Sentences' Context Importance

	BM25	TF-ISF	QLLM	GloVe	w2vec	fastt	w2vec*
case	1.0	1.0	.6	.9	1.0	1.0	1.0
opinion	1.0	1.0	.7	.9	1.0	1.0	1.0
paragraph	1.0	.9	.1	.5	.2	.2	0.0
recursive	.6	.6	.6	.6	.6	.6	.6