Legal Knowledge and Information Systems
M. Palmirani (Ed.)

doi:10.3233/978-1-61499-935-5-200

Towards Explainable Semantic Text Matching

Jörg LANDTHALER, Ingo GLASER and Florian MATTHES
Software Engineering for Business Information Systems, Department of Informatics, Technical University of Munich, Germany

Abstract. The growing amount of textual data in the legal domain leads to a demand for better text analysis tools adapted to legal domain specific use cases. Semantic Text Matching (STM) is the general problem of linking text fragments of one or more document types. The STM problem is present in many legal document analysis tasks, such as argumentation mining. A common solution approach to the STM problem is to use text similarity measures to identify matching text fragments. In this paper, we recapitulate the STM problem and a use case in German tenancy law, where we match tenancy contract clauses and legal comment chapters. We propose an approach similar to local interpretable model-agnostic explanations (LIME) to better understand the behavior of text similarity measures like TFIDF and word embeddings. We call this approach eXplainable Semantic Text Matching (XSTM).

Keywords. Semantic Text Matching, Explainable AI, Word Embeddings, TFIDF, Text Similarity Measure, German Tenancy Law

1. Introduction

The amount of textual data relevant in the legal domain is continuously growing. This leads to a demand for text analysis tools that capture more and more of the semantics of the textual data. Automated semantic processing of texts requires an adequate representation of texts. Many scientific applications of NLP for legal information systems leverage word embeddings, for example, question answering by Adebayo et. al [1], information extraction by Chalkidis et. al [2] or argumentation mining by Rinott et. al [3]. It is unclear when TFIDF or word embeddings is the superior technology. While word embeddings' characteristics are intriguing, to date it is not understood why certain structures occur in the embedding spaces nor efficient and effective quality measures for word embeddings are available.

Explainable artificial intelligence (XAI) is a research area that aims to better understand the behavior of algorithms. Waltl and Vogl [4] elaborate on the importance of XAI approaches for the legal domain. Waltl et al. [5] investigated the application of the particular XAI method LIME [6] to explain the behavior of supervised machine learning algorithms on classification tasks. In this paper we focus on text similarity measures to solve Semantic Text Matching (STM) problems. STM is the general problem identifying implicit semantic or logical relationships among text fragments. We propose an explanation approach similar to LIME for an unsupervised machine learning pipeline that we call eXplainable semantic text matching (XSTM). XSTM performs a sensitivity analysis for the words that are part of a text similarity measurement. This allows to investigate the contribution of the individual words to the text similarity measurement. We show preliminary results of XSTM on a German tenancy law use case, where text fragments of tenancy contracts and legal comments are matched.

The remainder of this paper is organized as follows: Section 2 recapitulates STM and briefly summarizes results of our previous research. We introduce our XSTM approach in Section 3. In Section 4 we set our work into the context of related work. Section 5 concludes the paper with a short summary and an outlook to future work.

2. Semantic Text Matching (STM)

STM has been introduced in [7]. STM is the general problem of identifying implicit links among text fragments where text fragments stem from documents of one, two or more different document types. For example, in argumentation mining premises need to be matched against claims. Problems of this kind are often tackled with text similarity methods. A high text similarity indicates a high probability for a link. A benefit of text similarity measures based on TFIDF or word embeddings is the unsupervised nature, i.e. no labeled data is required. STM can be seen as generalization of several problems that can be solved using text similarity measure technologies. In contrast to explicit citation networks, where references to other text fragments are present in the text, STM is an approach to identify implicit semantic or logical references among text fragments. STM is related to general information retrieval. The more different document types are involved, the more STM approaches a general search problem. A restriction to one or two document types leads to a problem easier to solve than general search. This enables a deeper investigation of the involved text similarity measures with less side effects.

Our use case is an envisioned support tool for lawyers that analyze or edit contracts. For our novel human-computer interaction method lawyers interactively

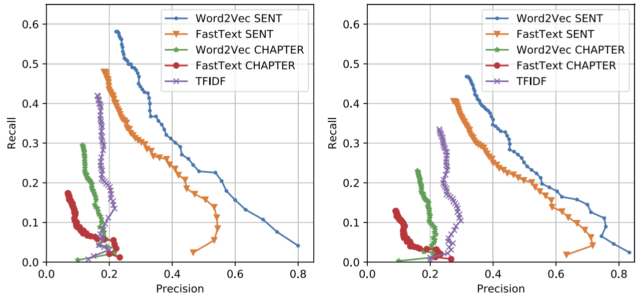

Figure 1. Precision/Recall curves for our STM use case that matches tenancy contract clauses and legal comment chapters. The SENT approach significantly performs better than TFIDF and CHAPTER approaches. word2vec performs better than FastText. TFIDF performs better than word embeddings with the CHAPTER approach. Results shown for the *narrow* tagging in the left column and for the *broad* tagging in the right column.

select a text fragment of interest and see relevant results of a suitable corpus. We implemented this as a web application. The dataset encompasses six tenancy contracts (37 cosmetic repairs related clauses) and three legal comments (1800 chapters). The identification of related legal comment chapters can be seen as a STM problem. We compared three approaches to recommend users relevant legal comment chapters:

- **TFIDF:** We encode the corpus of all contracts and legal comment chapters using the traditional TFIDF[1] representation and calculate the cosine similarity to rank all chapters.
- **CHAPTER:** Contract clauses and legal comment chapters are represented with vectors where a text fragment is the sum of all word embedding vectors representing the contained words[8]. We rank results again with cosine similarity.
- **SENT:** Equal to the CHAPTER approach except that the legal comment chapters are further segmented into sentences and the chapters with the most similar sentences are retrieved.

From the dataset described in [7] we only show the results for the cosmetic repair contract clauses where our ground truth for evaluation contains 223 links for the broad tagging and 127 links for the narrow tagging among contract clauses and legal comments. The differences among the broad and narrow tagging are explained in [7]. The word embeddings have been trained with word2vec[2][9] and FastText[3][10] with standard parameters except size is set to 300, iterations to 100 and min-count is set to zero. We use the CBOW model. Fig. 1 shows that our SENT approach performs best. However, for our use case the user needs some training to use our system effectively and the ground truth can only capture a subset of potential queries.

3. eXplainable Semantic Text Matching (XSTM)

We propose the XSTM approach to investigate the effect of individual words in text similarity applications. XSTM draws from ideas of LIME [6]. The idea is to perform a sensitivity analysis with the text fragments words as input features and the text similarity score among text fragments as output. XSTM can be applied to any text similarity application with different text similarity technologies, for example word embeddings. In order to assess the impact of an individual word we remove the word from one text fragment and re-calculate the text similarity between two text fragments. The difference among the original text-similarity and the newly calculated text similarity can be seen as the contribution for this particular word for the similarity among the two involved text fragments. This can be extended to all words of two text fragments by subsequently removing all words one after the other. Fig. 2 illustrates the contributions for all words among two text fragments of our tenancy law use case. XSTM can be further extended

[1] https://radimrehurek.com/gensim/, version 3.1.0, last accessed September 2018
[2] https://github.com/kzhai/word2vec, version 0.1c (for OSX), last accessed September 2018
[3] https://fasttext.cc/, version 0.1.0, last accessed September 2018

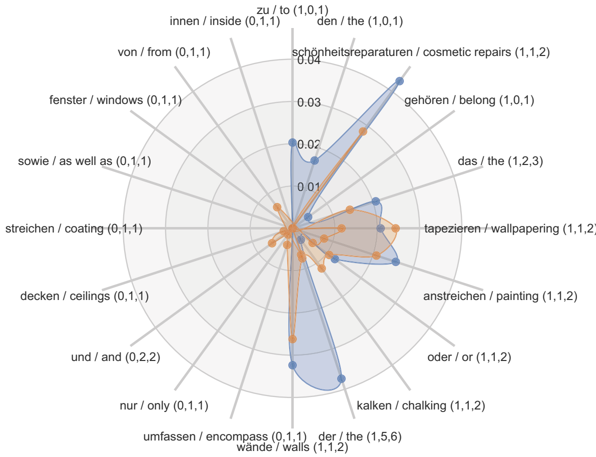

Figure 2. The contribution of individual words to a single match can be visualized with a radar chart for both: the query (contract clause) and the matching sentence (from a legal comment chapter). The query (orange) is *Zu den Schönheitsreparaturen gehören das Tapezieren, Anstreichen oder Kalken der Wände.* (To the cosmetic repairs belongs the wallpapering, painting or chalking of the walls.). The matching sentence (blue) using our SENT approach is *Schönheitsreparaturen umfassen nur das Tapezieren, das Anstreichen der Wände und Decken sowie das Streichen der Fenster von innen.* (The cosmetic repairs encompass only the wallpapering, painting of walls and ceilings as well as the coating of windows from the inside.). The axes display the contribution for the participating words (in brackets: German, English translation, frequency in query, frequency in matching sentence, occurrence frequency in both, query and matching sentence). A simplified example was chosen to facilitate visualization and it is not representative for the dataset. An interesting observation is that nouns seem to contribute most to the similarity among query and matching sentence. A structured evaluation of several matches will be necessary.

to assess the contribution of words among text fragments that are part of several links.

4. Related Work

LIME, proposed by Ribeiro et. al [6], is a XAI approach that performs a sensitivity analysis on black box machine learning classifiers. The effect of input variations on the output can serve as 'explanation' of the importance of different features to a specific classification result. In contrast to that, we focus on an unsupervised application. Waltl et. al [5] compare a rule-based approach and machine learning classifiers to classify sentences of the tenancy law part of the German Civil Code. On one concrete classification result they showed, using LIME, that the machine learning classifiers most significant features are similar to the features of the manually crafted rules. Semantic Text Matching [7] is the general problem of identifying implicit semantic or logical links among text fragments and is related to or present in several legal domain applications, for example, information retrieval and argumentation mining. A subtask of argumentation mining is to identify premises that support a claim. Rinott et. al [3] use TFIDF and word embeddings to identify evidence for claims in debates. However, it is rarely investigated why one text similarity measure surpasses another. Qureshi and Greene [11] present an unsupervised explainable word embeddings technique (EVE) that modifies the training of word embeddings in way so that individual dimensions of the word embeddings are clamped to specific concepts of a knowledge base such as Wikipedia. EVE is a constructive approach to build explainable embedding models by nature. In contrast to that, our approach is an attempt to investigate the characteristics of native embedding methods like word2vec or FastText from the outside.

5. Conclusion

We recapitulated STM as a general problem that also occurs in legal domain specific applications such as argumentation mining. We compare TFIDF and word embeddings as text similarity measures to solve a particular STM in German tenancy law. We propose XSTM as an approach to assess the impact of individual features (words) when used in a text similarity application. We hope that XSTM will enable us to deeper investigate the behavior of the different text representation and text similarity methods TFIDF and word embeddings.

References

[1] G. B. Adebayo Kolawole John, Luigi Di Caro and C. Bartolini, "- an approach to information retrieval and question answering in the legal domain," in *Proceedings of the 10th International Workshop on Juris-informatics (JURISIN 2016)*, 2016.

[2] I. Chalkidis and I. Androutsopoulos, "A deep learning approach to contract element extraction," in *Legal Knowledge and Information Systems - JURIX 2017: The Thirtieth Annual Conference, Luxembourg, 13-15 December 2017.*, 2017, pp. 155–164.

[3] R. Rinott, L. Dankin, C. Alzate, M. M. Khapra, E. Aharoni, and N. Slonim, "Show me your evidence–an automatic method for context dependent evidence detection," in *Proceedings of the 2015 Conference on Empirical Methods in NLP (EMNLP), Lisbon, Portugal*, 2015, pp. 17–21.

[4] B. Waltl and R. Vogel, "Explainable artificial intelligence the new frontier in legal informatics," *Jusletter IT 22*, 2018.

[5] B. Waltl, G. Bonczek, E. Scepankova, and F. Matthes, "Semantic types of legal norms in german laws: Classification and analysis using local linear explanations," *Artificial Intelligence and Law*, 2018.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 2016, pp. 1135–1144.

[7] J. Landthaler, I. Glaser, E. Scepankova, and F. Matthes, "Semantic text matching of contract clauses and legal comments in tenancy law," in *Tagunsband IRIS: Internationales Rechtsinformatik Symposium*, 2018.

[8] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents." in *ICML*, vol. 14, 2014, pp. 1188–1196.

[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR*, vol. abs/1301.3781, 2013.

[10] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[11] M. A. Qureshi and D. Greene, "Eve: explainable vector based embedding technique using wikipedia," *Journal of Intelligent Information Systems*, Jun 2018.