A Hybrid Method for Finding Relevant Case Law

Winkels, R.G.F.

A HYBRID METHOD FOR FINDING RELEVANT CASE LAW

Radboud Winkels

Associate Professor, Leibniz Center for Law, University of Amsterdam
P.O. Box 1030, 1000 BA Amsterdam, NL
winkels@uva.nl; http://www.leibnizcenter.org/~winkels/

Keywords: ***network analysis, LDA, bag of words, recommender system, topicalisation***

Extended Abstract:

Within the OpenLaws.eu project, we attempt to suggest relevant new sources of law to users of legal portals based on the documents they are focussing on at a certain moment in time, or those they have selected [Winkels, 2015]. In the future we attempt to do this both based on 'objective' features of the documents themselves and on 'subjective' information gathered from other users ('crowd sourcing'). At this moment we concentrate on the first method.

In earlier work [Winkels e.a. 2014] we have described results of experiments using analysis of the network of references or citations to suggest these new documents. Based on the current document in focus, i.e. an article in a Dutch law, we built a local network of sources of law based on the references from and to this focus document. This network contained both case law and legislation. Now we describe two experiments where we mix the use of a network analysis with similarity measures based on the comparison of the actual text of documents.

The first experiment concerns case law within the Dutch tax domain, about 6,000 documents. Bag-of-words combined with TFIDF weighting and cosine similarity has been used to find the textual similarity between two case law documents; this serves as a baseline for evaluating. The same algorithms are used to calculate the reference structure similarity between two documents. The references are extracted from the documents by a parser that has a precision of 0.55 in this domain. Experts evaluated the recommendations done by the baseline and the text similarity combined with the reference similarity algorithm.

The evaluation resulted in the conclusion that adding a similarity measure on reference structures is not performing as well as text similarity alone if not all references are identified by the parser. The parser can easily be improved, but perhaps similarity in references is no indication for relevant cases.

The second experiment concerns case law in the Dutch immigration law domain. Suggestions for new documents are generated by comparing the (pre-processed) texts of the 13,500 cases using a topic model. The unsupervised learning method La-

tent Dirichlet Allocation (LDA, Blei e.a. 2003) is used for this, extended with n-grams.[1] *The topic model represents the cases as mixtures of topics, after which the most similar ones are found by calculating the similarity between the topic mixtures.*

The topic similarity based suggestions that this project generated, were evaluated by legal experts and novices, who ranked three suggestions from best to worst, and stated whether the suggestion is good enough for a recommender system. One of the three suggestions was based on topic similarity, another was based on references to legislation and the other was based on a combination of these two methods.

Both legal experts and novices showed significant preference for the suggestions based on topic similarity. The legal experts wanted to see 85% of the suggestions based on topic similarity in a recommender system, for the novices this was 87%. The legal experts ranked the suggestions based on topic similarity as best suggestion 80% of the time, while the novices always ranked the suggestion based on topic similarity as best suggestion. This suggests that suggestions based on topic similarity can give useful suggestions within Dutch case law.

Acknowledgements*: Part of this research is co-funded by the Civil Justice Programme of the European Union in the OpenLaws.eu project under grant JUST/2013/JCIV/AG/4562.*

References

Blei, D.M., Ng, A.Y., Jordan, M.I., Latent Dirichlet Allocation. Journal of Machine Learning Research 3, 2003, pp. 993-1022.

Winkels, R.G.F., The OpenLaws project: Big Open Legal Data. Proceedings of IRIS 2015. Feb 2015, Salzburg, pp. 189-196.

Winkels, R.G.F., Boer, A., Vredebregt, B. & Someren, A. Van, Towards a Legal Recommender System. In: Hoekstra, R. (ed). Proceedings of JURIX 2014, IOS Press, Amsterdam, 2014, pp. 169-178.

[1] LDA was performed using MALLET with number of topics set to 150.