The Gap between Deep Learning and Law: Predicting Employment Notice

Jason T. Lam [*1] **Rohan Bhambhoria** [*2] **David Liang** [3] **Xiaodan Zhu** [2] **Samuel Dahan** [3 4]

1. Introduction

In Canada, the amount of reasonable notice owed is predicated on four attributes decided in Bardal v. Globe & Mail. Known as the Bardal Factors, they state that the determination of reasonable notice should use the following four factors: 1) age of the employee, 2) length of service, 3) character of employment, and 4) availability of similar employment (Bardal v. Globe & Mail Ltd. (1960)). Although the courts issued rough criteria to be used for reasonable notice, there was no determination of the weight of each factor nor whether all, if any, factors should be used. This subjectivity in the law has since been echoed by Justice Dunphy, who wrote of the calculation of reasonable notice periods that "[it] is more art than science but must be one that is fair in all of the circumstances" (Fraser v. Canerector Inc.). In this study we build on our previous research in Dahan et al. (Dahan et al., 2020), where we investigated the predictability of reasonable notice using statistical machine learning. Furthering our previous findings, we use deep learning to predict the amount of reasonable notice a person would receive based on a free text summary of a case. Our work also comprised experimentation with a variety of domain adaptations in determining the efficacy of a pre-trained model's generalized understanding of the legal language.

Advances in pre-trained models have led to their use in almost all areas of Natural Language Processing. Devlin et al. (Devlin et al., 2018) first introduced this method with BERT, where they pre-trained a series of Transformers introduced by Vaswani et al. (Vaswani et al., 2017) on a large corpus prior to fine-tuning on downstream tasks. The pre-training enabled the model to have a generalized understanding of the language, which proved to greatly improve its ability to solve subsequent problems, prior to learning the downstream task. Furthering this pre-training technique, a common practice is domain-adapting a pre-trained BERT-esque model on the required language domain used by the final task. This allows BERT-esque models to begin with a generalized understanding of the language to learning the idiosyncrasies of language used in the final task. These domain adaptations have been shown to be effective in other domains, such as the scientific (SciBERT (Beltagy et al., 2019) and medical (BioBERT (Lee et al., 2020) fields. Just as lawyers begin by learning the language prior to learning the law and deciding cases, we used this approach to investigate whether we could domain-adapt a model to accurately predict the amount of reasonable notice a person would receive upon termination.

2. Problem Statement

Using free-text summaries of case law, we aim to predict the amount of reasonable notice a plaintiff would receive, expressed in months. We group quantities of more than 24 months greater into a singular class, leaving us with a classification task with 25 classes.

3. Data

Our training set for classification comprised 1,287 case summaries from the Westlaw Quantum Service [1], and 409 cases were used for testing. The summaries only contained objective facts of the case; any mention of the judgement or the conclusion was manually removed.

[*]Equal contribution [1]Queen's University [2]Department of Electrical and Computer Engineering, Queen's University [3]Faculty of Law, Queen's University [4]Cornell Law School, Cornell University. Correspondence to: Rohan Bhambhoria <r.bhambhoria@queensu.ca>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

[1]https://www.westlawnextcanada.com/quantums/

To understand the applications of deep learning models for law, we experimented with domain adaptations of RoBERTa models using the Harvard case law project. Language in cases prior to 1960 were determined to be linguistically different from present-day Canadian cases, and were removed. Our final dataset comprised approximately four million American cases.

4. Methods

We used a pre-trained RoBERTa (Liu et al., 2019) from Facebook AI[2] and benchmarked our models against two deep-learning models. Our first baseline is the Hierarchical Attention Network (Yang et al., 2016), which has shown success across multiple domains by separately learning the importance of each word and sentence. As each sentence of our summaries corresponds approximately to a fact of the case, we believe this hierarchical approach reflects the judicial decision-making process. Our second strategy was to implement a few-shot model (Hu et al., 2018) that has shown success in the prediction of criminal law, which used a stack of attention layers to predict attributes. We further augmented our few-shot model by using structured self-attention from Lin et al. (Lin et al., 2017), such that our attention mechanisms could attend to multiple parts of the text for the prediction of each attribute.

For our domain adaptation experiments, we used two domains to test effectiveness for learning the legal language. The first domain we used was the Harvard case law project as Liu et al. (Liu et al., 2019) stressed the importance of volume over quality for improving performance. The second domain was the full case text of the summaries we used to train our classification task.

5. Results and Discussion

Approach	Acc. (+/-2)
HAN	67%
Few-shot w/ Self-attention	51%
BERT+base	61%
BERT+full cases	49%
RoBERTa+full cases	63%
RoBERTa+Harvard	65%
RoBERTa+base	**69%**

Table 1. Summary of results

Results were reported as accuracy with a +/-2 window to help address the subjectivity of the law (i.e. our prediction was classified as correct if it was within +/-2 of the ground truth).

Interestingly, our $RoBERTa_{base}$ achieved the best performance at 69%, followed by our HAN at 67%. In addition, our experiments have shown that domain adaptations for law negatively impacted our performance, contrary to literature (Rietzler et al., 2019). The superior performance of domain-adapting RoBERTa with the Harvard Case Law dataset over the full case text is in line with the original conclusions of Liu et al. (Liu et al., 2019), which emphasized the importance of quantity.

6. Conclusion

In this work we summarize the previously completed statistical machine-learning approach applied to the task of determination of reasonable notice and present preliminary results on using deep-learning architectures to tackle the same problem. We also explore improvements in neural methods, resulting in the improved metric of accuracy. An obvious direction for future work would be exploring interpretable neural models to visualize the determining aspects in order to ensure alignment with a legal decision.

[2]https://github.com/pytorch/fairseq

References

Bardal v. Globe & Mail Ltd. (1960). 24 D.L.R. (2d).

Beltagy, I., Cohan, A., and Lo, K. Scibert: Pretrained contextualized embeddings for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.

Dahan, S., Touboul, J., Lam, J., and Sfedj, D. *Predicting Employment Notice with Machine Learning: Promises and Limitation*. 2020.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Fraser v. Canerector Inc., . CarswellOnt 4796, 2015 ONSC 2138 at para 32.

Hu, Z., Li, X., Tu, C., Liu, Z., and Sun, M. Few-shot charge prediction with discriminative legal attributes. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 487–498, 2018.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

Lin, Z., Feng, M., Santos, C. N. d., Yu, M., Xiang, B., Zhou, B., and Bengio, Y. A structured self-attentive sentence embedding. *arXiv preprint arXiv:1703.03130*, 2017.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Rietzler, A., Stabinger, S., Opitz, P., and Engl, S. Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification. *arXiv preprint arXiv:1908.11860*, 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480–1489, 2016.