Measuring the complexity of the law: the United States Code

Anonymous Author(s)
email

1 Summary

Legal complexity can be an important metric for policy discussions and research in law. Katz et al. [4] present a new empirical framework to measure and quantify legal complexity. They consider complexity as "the human capital expended by a society when an end user is required to review and assimilate a body of legal rules.". Based on that definition, they examine the three qualitative features structure, language, and interdependence that contribute to complexity and use various heuristics to quantify these features, namely:

- **Structure**:
  - Structural size: Number of structural elements (sections, subsections, etc...) contained in a Title.
  - Element depth distribution: Mean hierarchical level of the structural elements.
- **Language**:
  - Size: Number of tokens.
  - Word entropy: Shannon entropy of a Title.
- **Interdependence**:
  - Net flow: Citations of a Title (by other Titles) compared to number of citations in the Title, measuring interdependence across Titles.
  - Intra-Title citation: Proportion of citations that reference the same Title.

Based on these heuristics, they use weighted ranking to construct a composite measure for the complexity. They apply this framework to the forty-nine Titles of the U.S. Code.

2 Critique

In my opinion, an evaluation of their framework is completely missing. Their premise is that structure, language, and interdependence are good indicators for the legal complexity, but they never evaluate this claim quantitatively. Furthermore, they present various heuristics that can be used as a proxy for these indicators, but neither perform an evaluation if they really measure the intended indicator nor point to literature that supports these claims. For some of the indicators and heuristics, I doubt that they really contribute to complexity in the way they measure it. Is the element depth distribution really a good heuristic for complexity? For instance, if we hypothetically consider the thought experiment where a Title is revised into a more fine-grained hierarchy (with the same content), does that really increase the complexity (according to their definition, i.e. resulting in more human capital)? I would say it could also decrease the complexity because the text is more structured and therefore easier to read. Furthermore, the heuristics for measuring the complexity of the language are questionable in my opinion. Depending on the circumstances, a longer text can be understood faster than a short, condensed text. They also claim that "higher entropy indicates the end user will more often encounter new language and new concepts". While a higher entropy per definition indicates that a user will more

Submitted as response essay for the *Sequencing Legal DNA: NLP for Law and Political Economy* course at ETH Zurich. Do not distribute.

often encounter new language, the same does not necessarily hold for new concepts. For instance, when one replaces repetitive words in a texts with their synonyms, the entropy will be higher, but the amount of concepts will be the same and the readability might increase. Kontoyiannis [5] shows that entropy captures statistical strucure and descriptional complexity, but not the complexity that comes from the contextual and semantic meaning of the text. The usage of their language metrics are especially dangerous because they claim that their "(...) analysis naturally lends itself to studies of legal systems over time.". Language is constantly evolving and the word distributions change significantly over time [3], but their metrics do not incorporate these changes and would attribute them completely to complexity. It might arguably be that the language of legal texts developed differently over time (or not at all), but I would expect this issue to be part of such a paper, especially if one of the intended/anticipated usages is the study of legal systems over time. Finally, I have some doubts regarding the interdependence metrics. Do all citations add the same complexity (like they assume) or does the complexity also depend on the semantic context and is e.g. different when listings (which a reader potentially can skip and still understand the Title) or other rules are cited?

In theory, the dataset could be interesting for the evaluation of a new framework as the complexity among Titles differs according to the authors. But the authors only provide various analyses of the dataset according to their criterion, they do not provide data that confirms the diversity in the complexity among Titles. Moreover, as they do not perform an evaluation (they only mention that the Internal Revenue Code, which is considered complex, is far from the most complex in the ranking), the dataset is not used to answer the research question and the goal of the data analysis remains unclear. Because of this, the main research question, namely if their framework captures legal complexity well, remains unanswered.

3 Further Work

As pointed out in the previous section, an evaluation would improve the paper a lot. Because of their definition of legal complexity, a baseline could be established by giving the different Titles (or parts of it) to n end users (that are chosen according to their definition, i.e. lawyers, laypersons, public interest groups, and businesses) and letting them rate the complexity of the Titles/samples. This would show quantitatively that the dataset is really useful for the research question and would allow to evaluate the framework by measuring how well their scores align with the baseline. Furthermore, the different heuristics could be evaluated quantitatively (e.g., by measuring the correlation with the ground truth complexity).

Moreover, recent machine learning techniques for estimating text difficulty (e.g., [1, 2, 6]) could be incorporated into the framework (with potential modifications for legal texts) and replace the current, simple heuristics, especially the usage of Shannon entropy.

With a labelled dataset (where the ground truth complexity is collected as described previously), it would also be interesting to explore if the problem of predicting legal complexity could be turned into a supervised machine learning problem. It could be explored if, given the right features, relatively simple regression models perform well or if the relation between features and complexity is more complex and less interpretable models like random forests are required to get a good performance. With these results, the tradeoff between interpretability and accuracy could be explored explicitly and the authors could still decide to use a simpler, explainable model after analyzing the data.

References

[1] Renu Balyan, Kathryn McCarthy, and Danielle McNamara. “Comparing Machine Learning Classification Approaches for Predicting Expository Text Difficulty”. In: May 2018.

[2] Alfredo Cuzzocrea et al. “Multi-Class Text Complexity Evaluation via Deep Neural Networks”. en. In: *Intelligent Data Engineering and Automated Learning – IDEAL 2019*. Ed. by Hujun Yin et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 313–322. ISBN: 978-3-030-33617-2. DOI: 10.1007/978-3-030-33617-2_32.

[3] Adam Jatowt and Katsumi Tanaka. “Large Scale Analysis of Changes in English Vocabulary over Recent Time”. en. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management - CIKM ’12*. Maui, Hawaii, USA: ACM Press, 2012, p. 2523. ISBN: 978-1-4503-1156-4. DOI: 10.1145/2396761.2398682.

[4] Daniel Martin Katz and M. J. Bommarito. “Measuring the Complexity of the Law: The United States Code”. en. In: *Artificial Intelligence and Law* 22.4 (Dec. 2014), pp. 337–374. ISSN: 1572-8382. DOI: 10.1007/s10506-014-9160-8.

[5] I Kontoyiannisy. *The Complexity and Entropy of Literary Styles*. en. NSF Technical Report 97. Department of Statistics, Stanford University, p. 15.

[6] M. Zakaria Kurdi. “Text Complexity Classification Based on Linguistic Information: Application to Intelligent Tutoring of ESL”. In: *arXiv:2001.01863 [cs]* (July 2020). arXiv: 2001.01863 [cs].