Legal Knowledge and Information Systems
S. Villata et al. (Eds.)

doi:10.3233/FAIA200867

Judges Are from Mars, Pro Se Litigants Are from Venus: Predicting Decisions from Lay Text

Karl BRANTING, [1] Carlos BALHANA, Craig PFEIFER, John ABERDEEN, and Bradford BROWN
The MITRE Corporation, McLean VA, USA

Abstract. Access to justice could be significantly expanded if decision support systems were able to accurately interpret statements of fact by pro se (self-represented) litigants. Prior research, which has demonstrated that case decisions can often be predicted by machine-learning models trained on judges' statements of facts, suggests the hypothesis that these same learning algorithms could be effectively applied to pro se litigants' fact statements. However, there has been a dearth of corpora on which to test this hypothesis. This paper describes an experiment testing the ability to predict the outcome of pro se litigants' complaints on a corpus of 5,842 cases initiated by citizen complaints. The results of this experiment were strikingly negative, suggesting that fact statements by unguided pro se litigants are far less amenable to simple machine-learning techniques than judges' texts and appearing to disconfirm the hypothesis above.

1. Introduction

In many nations across the world, access to justice is increasingly elusive for the majority of citizens who are not wealthy [1] [2]. In the United States, for example, "more than 80 percent of people living below the poverty line and a majority of middle-income Americans receive no meaningful assistance when facing important civil legal issues, such as child custody, debt collection, eviction, and foreclosure" [3]. A widely-acknowledged factor in this inaccessibility is the complexity of legal rules and procedures. However, an equally important factor is the gap between the ordinary parlance used by laypersons and the specialized terminology and usage of legal discourse. This linguistic gap creates challenges both for decision support on behalf of pro se (self-represented) litigants and for decision support for the adjudicators who must handle the claims of litigants unfamiliar with legal language. Even when legal rules and procedures can be formalized in computer-interpretable and executable form, it is typically a formidable challenge to elicit case facts in language compatible with those rules and procedures.

This paper describes experiments in which techniques previously used to predict decisions from statements of facts in published decisions were applied to texts written

[1] Corresponding Author: E-mail: lbranting@mitre.org. The MITRE Corporation is a not-for-profit company, chartered in the public interest. This document is approved for Public Release; Distribution Unlimited. Case Number 20-2066.

by citizen complainants. The results of these experiments were strikingly negative, suggesting that a different approach is needed for eliciting and interpreting fact statements produced by pro se litigants.

2. Predicting Decisions from Complainant Texts

We gained access to a Complaint Data Set consisting of 5,842 attorney misconduct complaints processed by the bar association of a US state (*Bar Association*). A key step in the Bar Association's handling of these complaints is an initial determination of whether the case should be forwarded for full investigation or whether instead the case can be closed before investigation (*CBI*) because it fails to state a *prima facie* ("colorable") claim.

Each case in the Complaint Data Set consisted of information submitted by the complainant through an online complaint form, together with metadata including the following: the history of prior complaints filed against the attorney to whom the complaint was directed (the *respondent*); the legal services to have been provided by the respondent to the complainant; and allegation codes, which correspond to provisions of the state's code of professional responsibility and statutes regarding attorney misconduct and which are manually assigned by staff based on a reading of the complaint text at intake. The complainant information included the names of the complainant and respondent attorney or attorneys, a free-text description of the events justifying the complaint, a separate free text description of the relationship between the complainant and the respondent attorney (the "connection text"), and other information not relevant to the merits of the case. We supplemented this feature set with readability features, including Flesch Reading Ease and SMOG Index,[2] and mean per-sentence sentiment [4]. Each case was labeled as to whether it was closed before investigation or was investigated further. The categories were relatively balanced, with 55.65% closed at intake.

Our initial experiment explored how accurately CBI decisions could be predicted based on information available to the intake staff at the time the complaint was submitted. The complaint texts[3] were normalized by removing newlines and replacing each person's name with the token PERSON using the Stanford Named Entity Extraction (NER) tool.[4] We tested two feature representations for the texts:

- N-gram frequency vectors, for n = 2–4
- Vectors of 250 topic models[5]

We compared the performances of six machine-learning algorithms—Naive Bayes, Bayes Net, SMO, JRip, J48, and Random Forest—in 10-fold cross validation. The results of the highest-performing algorithms are shown in Table 1. Disappointingly, performance in predicting CBI decisions was only slightly higher than chance regardless of representation or algorithm. This result contrasts with the much better results obtained in other domains from text written by attorneys or judges, e.g., [5]. We hypothesize that the highly discursive, irregular, and inconsistent character of complaint texts is responsible for the much-lower predictive accuracy.

[2] https://pypi.org/project/readability/

[3] We appended the connection text, if any, to each complaint text.

[4] https://nlp.stanford.edu/software/CRF-NER.html

[5] The topic models were constructed using gensim (https://radimrehurek.com/gensim/about.html).

Features	Mean MCC	Frequency-weighted mean F1	Algorithm
n-gram frequency vectors	0.023	0.521	SVM
250 dim topic vector (gensim)	0.010	0.425	BayesNet

Table 1. The accuracy of decision predictions based on complaint text features.

Features	Mean MCC	Frequency-weighted mean F1
Case metadata only	0.116	0.525
Case metadata plus n-gram frequency vectors	0.153	0.551

Table 2. The accuracy of decision predictions based on a BayesNet model trained on metadata features, with and without complaint text.

Features	Mean MCC	Frequency-weighted mean F1
Allegation codes	0.376	0.695
Allegation codes plus text	0.376	0.695
Allegation codes plus metadata	0.377	0.696

Table 3. Prediction results based on a BayesNet model trained on allegation codes with and without text and metadata features.

We next evaluated the predictive value of the case metadata, which consisted of (1) non-narrative information provided by the complainant in the online form, (2) information from the Bar Association attorney database, (attorney history and prior complaints), and (3) the sentiment scores and various readability metrics calculated from the complaint texts.

Table 2 shows that the case metadata is somewhat more predictive of CBI decisions than complaint texts, and the combination of metadata and complaint texts is slightly more predictive than either individually, but even in combination these features are only weakly predictive of the CBI decisions.

We next explored the degree to which CBI decisions could be predicted after the intake staff had assigned allegation codes to each case, which occurs before the decision whether to send the case forward for investigation. As shown in Table 3, allegation codes standing on their own have moderate predictive value, with the MCC of 0.376 indicating that more than $1/3$ of the uncertainty about a complaint is eliminated if the allegation codes are known. Adding the text features didn't reduce the uncertainty further, indicating that the allegation codes capture most of the relevant, predictive information in the complaint text. Combining the allegation codes with metadata increases predictive accuracy by a very small amount.

This experiment indicated that allegation codes have a moderate predictive value for CBI decisions (an MCC of 0.376), so we turned to the question whether we could predict allegation codes from complaint texts. Assigning allegation codes to each new case is time-consuming for intake staff, so automating this process could be a useful form of decision support on its own, apart from helping identify cases that are likely to be closed on intake. We evaluated performance accuracy on prediction of the 10 most frequent allegation codes based on an n-gram representation of complaint texts, which collectively cover approximately 60% of all complaints. Only one allegation code could

be predicted with an MCC greater than 0.15, meaning that predictive accuracy for most allegation codes was only slightly higher than chance. We attempted to develop an annotation scheme for complaint texts so that we could apply the methodology described in [6] to the corpus, but the complaints' extreme variability and disorganization frustrated these annotation efforts.

In our view, these experimental results show that decision support systems that fail to support pro se litigants in expressing facts relevant and necessary for a claim create a high barrier to accomplishing the subsequent task of assessing whether the assertions state a prima facie case. the root problem is that pro se litigants seldom know what facts they need to establish or how to articulate and organize the facts in a manner that makes their claims amenable to evaluation.

In summary, our experiments with pro se complaint texts failed to replicate the predictive accuracy that we and others observed in previous work predicting decisions from judges' and other adjudicators' fact statements. We surmise that the characteristics of judges' and other adjudicators' language, including stylistic consistency and regularity, are critical to the ability of current machine-learning techniques to induce accurate predictive models from the statements of facts in published decisions.

3. Conclusion

This paper has presented an experiment in which the predictive accuracy previously demonstrated from judges' statements of facts could not be reproduced on fact statements written by pro se complainants. These results suggest that judges' statements of facts are a poor proxy for pro se litigants' narrative texts and that techniques suitable for prediction from judges' texts may not be appropriate for decision support for pro se litigants. We believe that a promising research direction is development of narrative elicitation techniques based on recent work on narrative schema induction [7]. Such techniques could help bridge the gap between the language of judges and the language of pro se litigants, which our experimental results suggest are as remote from one another as Mars is from Venus.

References

[1] Hadfield G. Rules for a Flat World: Why Humans Invented Law and How to Reinvent It for a Complex Global Economy. Oxford University Press; 2016.

[2] Himonas D, Hubbard T. Democratizing the Rule of Law. Stanford Journal of Civil Rights & Civil Liberties. 2020;16(2):261–282.

[3] Perlman AM. The Public's Unmet Need for Legal Services & What Law Schools Can Do about It. Daedalus. 2019;148(1):75–81.

[4] Hutto C, Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media; 2014. p. 00–00.

[5] Chalkidis I, Androutsopoulos I, Aletras N. Neural Legal Judgment Prediction in English. CoRR. 2019;abs/1906.02059.

[6] Branting LK, Pfeifer C, Brown B, Ferro L, Aberdeen J, Weiss B, et al. Scalable and explainable legal prediction. Artificial Intelligence and Law. 2020:1–26.

[7] Belyy A, Van Durme B. Script Induction as Association Rule Mining. In: Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events. Online: Association for Computational Linguistics; 2020. p. 55–62.