Explanation in Hybrid, Two-Stage Models of Legal Prediction

L. Karl Branting

The MITRE Corporation
McLean, VA, USA
lbranting@mitre.org

Abstract. This paper identifies a core set of legal decision support tasks requiring distinct forms of explanation, outlines a hybrid, two-stage model of legal prediction, describes how this model facilitates these explanation tasks, and outlines the development requirements of two-stage models.

1 Introduction

Explainability is a key requirement for AI systems to be understood, trusted, validated, and maintained. Recent research on explainability in AI has led to the recognition that there is no universal criterion for explanation utility and acceptability [12]. Instead, explanation acceptability and utility depend on the nature of the AI algorithm to be explained, the task to which the algorithm is applied, and the needs, expectations, and knowledge of the individuals for whom the explanations are produced.

Explicating the explanation requirements of *legal* AI systems is particularly challenging because there are many disparate stakeholders in legal systems, each with distinct objectives and levels of knowledge. As a result, legal problem solving encompasses a variety of legal tasks, each with distinct information-processing and explanation requirements. Approaches to explanation that focus on *transparency* [17] of algorithmic processes are generally useful for these tasks only if the individual algorithmic steps are based on concepts that are (1) legally meaningful or (2) grounded in case facts. Stated differently, these concepts must have an understandable connection to authoritative legal rules or to descriptions of possible states of the world, e.g., persons, actions, relationships, etc. Other concepts, such as connection weights, variable bindings, or decision surfaces might be useful for system verification but are unlikely in themselves meaningful to the stakeholders for whom decision support systems are developed. Post-hoc explanation approaches, like Local Interpretable Model-Agnostic Explanations (LIME) [16], may give rise to due process issues by misleading the user about the actual basis of a prediction or decision.[1]

[1] For another view see [20] (post hoc explanations for divisions of marital assets calculated by an opaque machine learning model).

This paper focuses the explanation needs two particular stakeholders—self-represented (*pro se*) litigants and the adjudicators who resolve claims by these litigants—in the context of legal prediction. Section 2 sets forth a model that distinguishes the individual tasks of these stakeholders. Section 3 describes a hybrid, two-stage models for legal prediction, and the use of such models for explanation is set forth in Section 4. The requirements for the development of hybrid, two-stage models is summarized in Section 5, and Section 6 summarizes and outlines future work.

2 Decision Support for Adjudicators and Pro Se Litigants

The objective of adjudicators is to resolve disputes, and the objective of litigants is to have their disputes resolved. This symmetry in objectives is reflected in a correspondence between the individual tasks that each must accomplish for a dispute to be resolved. Figure 1 illustrates how the overall case-adjudication process can be viewed from either perspective as consisting of three stages: initiation, assessment, and disposition.[2] Each of these stages can be further subdivided into individual tasks, each with separate information-processing requirements.

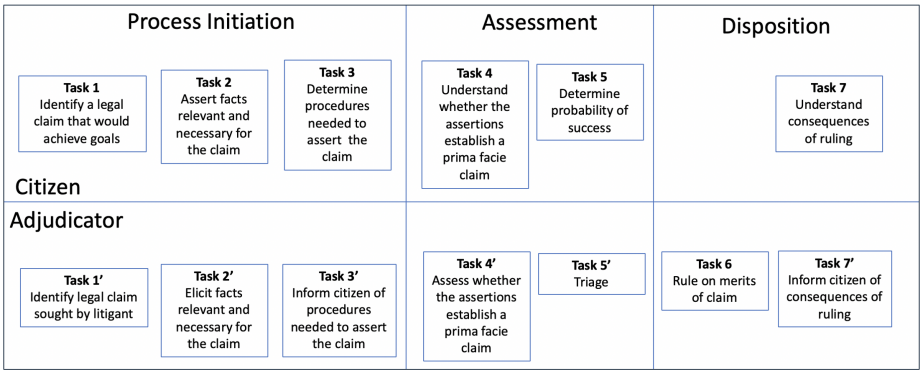

Fig. 1. Task decomposition for adjudication decision support. Citizens' and adjudicators' tasks are mirror images of one another.

To initiate the adjudication process, a claimant must first identify a legal claim that could potentially achieve the petitioner's goals (Task 1), then assert facts needed to establish the claim (Task 2), and finally determine what procedures are needed to move the claim forward (Task 3). To avoid being burdened

[2] This model does not address the various processes that may be required for hearings or trial, such as introducing evidence, testifying, or otherwise establishing the facts underlying the claim, which differ widely across different tribunals and causes of action.

with poorly-expressed claims that are difficult and time-consuming to understand and assess, an adjudication body must perform three corresponding tasks: identify the claim sought by the claimant (Task 1′), elicit the facts relevant to that claim (Task 2′), and inform the litigant of the procedure to assert the claim (Task 3′).

Once a case has been initiated, a litigant needs to understand whether the facts as asserted are sufficient to establish a prima facie case, that is, whether the facts if accepted would be sufficient to establish the claim (Task 4) and, if so, the probability of success should the claim be litigated (Task 5). It benefits both the litigant and the adjudicator for the litigant to have a realistic assessment of whether success on the claim is likely enough to justify continuing rather than abandoning the claim. After the claim is adjudicated, the litigant and adjudicator once again have corresponding tasks involving the adjudicator informing, and the litigant understanding, the consequences of the ruling.

Traditional legal decision-support systems generally focus on the pro se litigator's Tasks 2, 4, and sometimes 5 (the adjudicator's Tasks 2′, 4′, and 5′): eliciting facts, assessing whether a prima facie case has been established [4], and predicting the likelihood of success (typically in a highly-simplistic fashion). Generally, case facts are elicited through fillable web forms in the form of attribute/value pairs [10]. Legal reasoning in such systems is typically limited to simple propositional logic implemented in imperative programming constructs. This approach is conducive neither to verification nor to explanation, creating a significant risk of incorrect or incomprehensible legal advice [13].

As argued above, to be useful and comprehensible for the key tasks list above, explanations must be expressed in terms of concepts that are (1) legally meaningful or (2) grounded in facts. A family of predictive models that operate on such concepts is described in the next section.

3 Hybrid, Two-Stage Models of Decision Prediction

Many forms of legal argumentation and discourse are structured around precise rules that are modeled well by logic. Other aspects of legal problem solving, such as grounding the semantics of legal terms in the language of ordinary discourse, have no natural fit within the logical framework but are better suited to empirical analysis. The complementary role of logical reasoning with rules and semantic reasoning with case facts motivated a number of hybrid reasoning systems that combined rule-based with case-based reasoning [5] [19] [24] or other types of semantic analysis [18]. Hybrid approaches are intended to model the ability of human attorneys to create arguments that integrate arguments based both on prior cases and on rule-like norms, such as regulations and statutes.

For example, Figure 2 depicts a simplified hybrid model of VBA benefits determinations. Entitlement to benefits depends on four elements: military service; service-related injury' current disability; and a causal connection between the service-related injury and the current disability. Conceptually, these four elements are legal predicates, and entitlement to benefits requires establishing

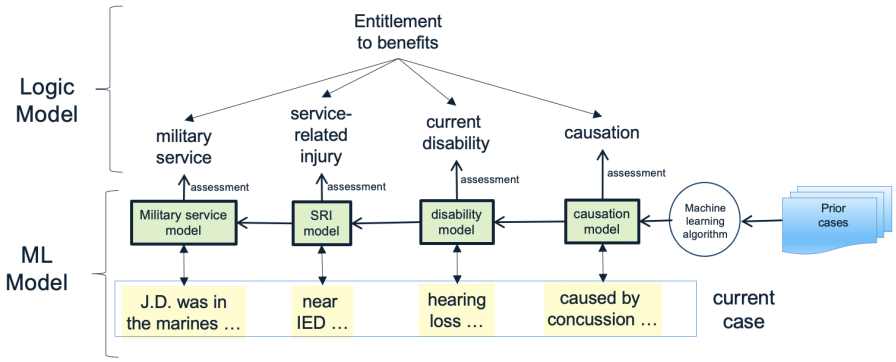

Fig. 2. A hybrid model of VBA benefits determination.

that each predicate is satisfied by the facts of the case. Each individual element in turn must be evaluated in terms of the facts of the case.

Early hybrid systems were able to attain an impressive level of explanatory capability, but they depended on manually represented case facts and were therefore are not scalable for practical systems in which case facts are expressed as text. In contrast, recent machine learning techniques have made it increasingly feasible to make legal predictions based on case facts expressed as text, but at the expense of explainability.

For example, machine learning models trained on fact statements have produced impressive levels of accuracy in predicting decisions of the European Court of Human Rights [1] [14] [9], US Board of Veterans Appeals cases, French Supreme Court decisions [22], UK court decisions [21], and World Intellectual Property Organization domain-name disputes [8]. However, outcomes predicted by these systems aren't justified in terms of legally relevant concepts or facts. Instead, the features on which the predictions are predicated are statistical features of the text, such as n-gram frequency vectors, metadata, or other features unrelated to the merits of the case. In some applications, such as litigation support, there can be significant strategic value in knowing the association between outcomes and factors unrelated to the merits, such as law firms and judges [23], but understanding predictions based on factors relevant to the actual merits of a case is vital both for pro se litigants themselves and adjudicators who have an institutional obligation to justify each decision regardless of whether a decision support tool assisted in analyzing or deciding the case.

One approach to enabling machine learning models to explain predictions in terms of legally relevant concepts is to conceptualize the process of prediction as consisting of two steps, depicted notionally in Figure 3.[3] The first model

[3] Another approach uses attention networks to identify the most relevant parts of case statements. One attempt at this approach proved ineffective for decision support in [7].

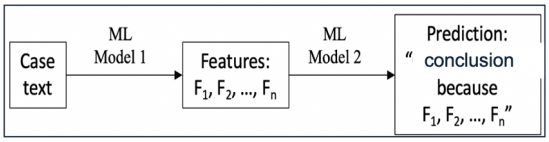

Fig. 3. A paradigm of two-stage legal prediction.

predicts the relevant concepts from the case text, and the second step predicts the decision based on the concepts predicted in the first step. Transparency in the second step's model can be the basis of explanation in terms that are legally meaningful, factually grounded, and useful for users' tasks.

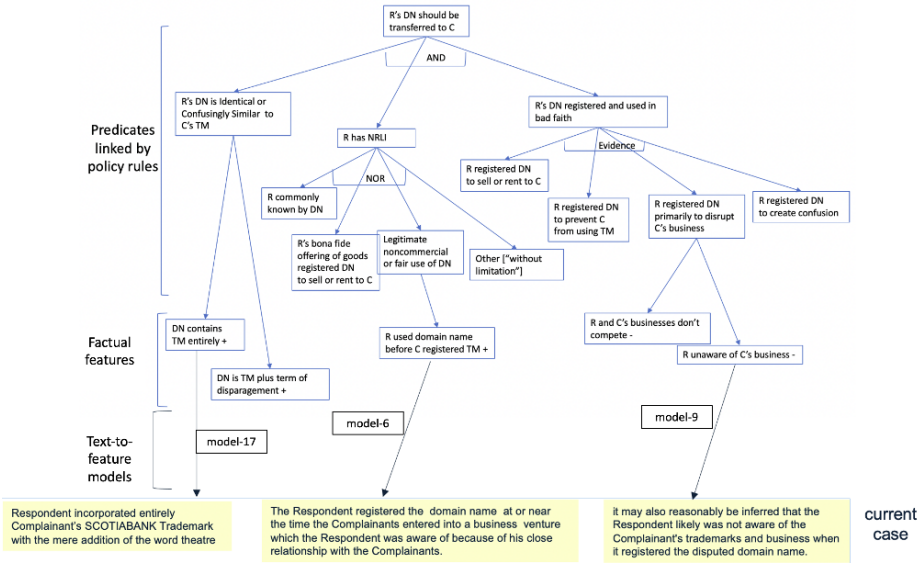

Fig. 4. A hybrid, two-stage architecture for explainable prediction of WIPO domain name disputes.

The hybrid approach can be combined with the two-stage model as illustrated in Figure 4, which shows an architecture for World Intellectual Property Organization (WIPO) domain name disputes that is hybrid in that it includes both policy rules and text-interpretation components, and two-stage in that legal predicates (e.g., "The domain name is confusingly similar to Complainant's trademark,") are factually grounded in the case text via intermediate factual features (e.g., "The domain name contains Complainant's trademark entirely"). The two-stage component of this architecture, which included a separate model for predicting each of 46 factual features from textual case statements, was im-

plemented in the SCALE system described in [7]. The motivation for predicting factual features which are then used to reason about the outcome is that factual features are meaningful in terms of the facts of the case and are therefore more likely to be comprehensible by and useful for users, as described in the next Section.

4 Explanation in Hybrid, Two-Stage Systems

The research literature on legal decision support systems is woefully deficient in human factors studies needed to explicate pro se litigants' cognitive assistance requirements. As a result, the explanation requirements of pro se litigants are known only anecdotally. In the absence of such empirical studies, this section is guided by the task model set forth in Section 2.

4.1 Process Initiation

Hybrid systems contain both explicit legal rules and mechanisms for grounding legal predicates in case facts. Tasks 1 and 2 from the decision-support model above are explainable from the rule component of a hybrid system. Identifying a legal claim that would achieve a litigant's goal, Task 1, requires finding a legal rule whose consequent matches the objective sought by the litigant. The primary explanatory challenge of this process is overcoming the linguistic gap between legal predicates and ordinary discourse [6].

Task 2, assisting the user in asserting the facts relevant and necessary for a claim, requires reasoning with explicit legal rules as well. A claimant must establish each of the elements required for a claim. All the claimant needs to know, however, are the predicates at the leaves of the rule tree, i.e., the predicates that must be grounded in case facts because the rules have "run out" [11]. In the WIPO domain, for example, the first element that a complainant must establish is that the domain name is "identical or confusingly similar to the complainant's trademark." In a two-stage model, this requirement could be explained by listing factual features that confirmed or rebut the predicate, e.g., the predicate above is confirmed if the "domain name contains the trademark entirely" or the "domain name is the trademark plus a term of disparagement." Each of these factual features can in turn be explained by examples of texts from the training set in which the factual features were definitely present or absent.

4.2 Assessment

Useful explanations are typically contrastive, that is, they identify how something differs from some reference or expected case [15]. Thus, the explanations most useful for Task 4, understanding whether the claimant's assertions are sufficient to establish a prima facie case, may generally be those that focus on how the assertions are insufficient, i.e., how they differ from the claimant's expectation that the claim would be sufficient. Such explanations should help the claimant

understand what additional factual assertions would be necessary to establish the claim. In principle this would require identifying the minimal set of sufficient additional facts. In practice, however, simply identifying the smallest set of predicates that failed in any traversal of the rule tree might be sufficient. As with Task 3, the factual features in a two-stage model would permit a meaningful explanation of what additional assertions would be needed, e.g., if a WIPO claim is insufficient because it fails to make assertions that, if true, would satisfy the first element, being "identical or confusingly similar," then a two-stage system could explain that the claim fails to assert that the domain name contains the trademark, is the trademark plus a term of disparagement, or any of the other known forms that being "identical or confusingly similar" can take.

For Task 5, determining the probability of success, the most meaningful explanation is again likely to be contrastive, i.e., "Why is the probability of success so low?" or "How can I make the probability higher?". The two-stage model permits such questions to be answered in terms of individual factual features, e.g., if there is a particular leaf predicate whose strongest support is from a weekly supported factual feature, then the explanation would be that there is only weak support for that predicate and that additional ad. As with Task 4, an explanation of this type could help the claimant understand what additional facts would need to be established to strengthen the case.

5 Engineering Two-Stage Models

This paper has argued for the utility of hybrid, two-stage legal prediction models from the perspective of the goals and needs of pro se litigants and adjudicators. The full details of how such two-stage models can be constructed are outside of the scope of this paper, but a brief discussion can clarify the basic requirements.

Two approaches have been explored for identifying factual features intermediary between legal predicates in the leaves of rule trees and case facts, such as those described in this paper. In the first approach, the factual features are developed by domain experts in the relevant area of law and correspond to fact patterns that can make a case stronger or weaker. CATO factors are features of this type [2]. Machine learning models for extracting such factors from case text for the purpose of factor-based case prediction were trained and evaluated in [3] and [25].

A second approach focuses on textual patterns that occur in explanations of case decisions. One approach to identifying such textual patterns is to annotate factual findings in a representative set of published decisions, then map those initial annotations onto an entire corpus based on proximity in semantic embedding space [7]. This approach has the potential benefit of leveraging a small set of annotations onto a much larger corpus, but has the limitation that it is applicable only when decisions that include text setting forth the findings or reasoning underlying the decision.

Factual features induced from case statements can be used for case-based dialectical reasoning, as in [3], or for supervised concept learning, as in [2] and

[7]. The latter approach lends itself to to explanations based on the presence or absence of features, which are themselves individually comprehensible.

6 Conclusion

This paper has identified a core set of legal decision support tasks requiring distinct forms of explanation, outlined a hybrid, two-stage model of legal prediction, described how the model facilitates these explanation tasks, and concluded with a brief description of the development requirements of two-stage models. Future work in decision support for pro se litigants and adjudicators who handle their cases should include human-factors analysis of (1) how pro se litigants conceptualize their claims and their interactions with decision forums to address the key questions, and (2) what forms of explanation are most beneficial to these litigants in terms of the rate of success in asserting claims, time efficiency, and overall satisfaction with the decision support process. Regardless of the outcome of this research, it seems very probable that useful decision support systems for explainable legal prediction must have a hybrid, two-stage design that permits explanation both in terms of legal predicates and in terms of factual features to span the gap between legal predicates and the language of ordinary discourse.

Acknowledgments

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 20-3101.

References

1. Aletras, N., Tsarapatsanis, D., Preotiuc-Pietro, D., Lampos, V.: Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. PeerJ CompSci (2016). Https://peerj.com/articles/cs-93/
2. Ashley, K., Aleven, V.: Reasoning symbolically about partially matched cases. In: Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, pp. 335–341. Morgan Kauffmann, San Francisco (1997)
3. Ashley, K.D., Brüninghaus, S.: Automatically classifying case texts and predicting outcomes. Artif. Intell. Law **17**(2), 125–165 (2009)
4. Branting, L.K.: An advisory system for pro se protection order applicants. International Review of Law, Computers & Technology **14**(3) (2000)
5. Branting, L.K.: Reasoning with Rules and Precedents: A Computational Model of Legal Analysis. Kluwer Academic Publishers, Dordrect/Boston/London (2000)
6. Branting, L.K.: Data-centric and logic-based models for automated legal problem solving. Artificial Intelligence and Law **25**(1), 5–27 (2017)
7. Branting, L.K., Pfeifer, C., Brown, B., Ferro, L., Aberdeen, J., Weiss, B., Pfaff, M., Liao, B.: Scalable and explainable legal prediction. Artificial Intelligence and Law pp. 1–26 (2020)

8. Branting, L.K., Yeh, A., Weiss, B., Merkhofer, E.M., Brown, B.: Inducing predictive models for decision support in administrative adjudication. In: AI Approaches to the Complexity of Legal Systems - AICOL International Workshops 2015-2017, Revised Selected Papers, *Lecture Notes in Computer Science*, vol. 10791, pp. 465–477. Springer (2017)
9. Chalkidis, I., Androutsopoulos, I., Aletras, N.: Neural legal judgment prediction in english. CoRR **abs/1906.02059** (2019)
10. Frank, J.: A2j author, legal aid organizations, and courts: Bridging the civil justice gap using document assembly. W. New Eng. L. Rev. **39**, 251 (2017)
11. Gardner, A.: An Artificial Intelligence Approach to Legal Reasoning. Bradford Books/MIT Press, Cambridge, MA (1987)
12. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z.: Xai—explainable artificial intelligence. Science Robotics **4**(37) (2019). https://doi.org/10.1126/scirobotics.aay7120. URL https://robotics.sciencemag.org/content/4/37/eaay7120
13. Lauritsen, M., Steenhuis, Q.: Substantive legal software quality: A gathering storm? In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ICAIL '19, pp. 52–62. ACM, New York, NY, USA (2019)
14. Medvedeva, M., Vols, M., Wieling, M.: Using machine learning to predict decisions of the european court of human rights. Artificial Intelligence and Law **28**(2), 237–266 (2020)
15. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence **267**, 1–38 (2018)
16. Ribeiro, M., Singh, S., Guestrin, C.: "why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pp. 97–101. Association for Computational Linguistics, San Diego, California (2016). https://doi.org/10.18653/v1/N16-3020. URL https://www.aclweb.org/anthology/N16-3020
17. Schmidt, P., Biessmann, F., Teubner, T.: Transparency and trust in artificial intelligence systems. Journal of Decision Systems **0**(0), 1–19 (2020)
18. Schraagen, M., Testerink, B., Oderkerken, D., Bex, F.: Argument-driven information extraction for online crime reports. In: Proceedings of International Workshop on Legal Data Analytics and Mining (LeDAM 2018). ACM (2018)
19. Skalak, D., Rissland, E.: Arguments and cases: An inevitable intertwining. Law and Artificial Intelligence **1**(1) (1992)
20. Stranieri, A., Zeleznikow, J., Gawler, M., Lewis, B.: A hybrid rule – neural approach for the automation of legal reasoning in the discretionary domain of family law in australia. Artificial Intelligence and Law **7**, 153–183 (1999)
21. Strickson, B., De La Iglesia, B.: Legal judgement prediction for uk courts. In: Proceedings of the 2020 The 3rd International Conference on Information Science and System, p. 204–209. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3388176.3388183
22. Sulea, O., Zampieri, M., Vela, M., van Genabith, J.: Predicting the law area and decisions of french supreme court cases. In: RANLP, pp. 716–722. INCOMA Ltd. (2017)
23. Surdeanu, M., Nallapati, R., Gregory, G., Walker, J., Manning, C.: Risk analysis for intellectual property litigation. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law, pp. 116–120. ACM, Pittsburgh, PA (2011)

24. Walker, R.: An expert system architecture for heterogeneous domains. Ph.D. thesis, Vrije University (1992)
25. Westermann, H., Walker, V.R., Ashley, K.D., Benyekhlef, K.: Using factors to predict and analyze landlord-tenant decisions to increase access to justice. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ICAIL '19, p. 133–142. Association for Computing Machinery, New York, NY, USA (2019)