Predicting Brazilian court decisions

André Lage-Freitas, Héctor Allende-Cid, Orivaldo Santana, and Lívia de Oliveira-Lage

[1]Universidade Federal de Alagoas, Brazil.
[2]Pontificia Universidad Catlica de Valparaso, Chile.
[3]Universidade Federal do Rio Grande do Norte, Brazil.
[4]Prosecutor at State's Attorney's Office (Procuradoria Geral do Estado de Alagoas), Brazil.
lage@laccan.ufal.br, hector.allende@pucv.cl, orivaldo.santana@ect.ufrn.br, livia.oliveira@pge.al.gov.br

***Abstract*—Predicting case outcomes is useful but still an extremely hard task for attorneys and other Law professionals. However, it is not easy to extract valuable information from cases as this requires dealing with huge and complex data sets such as myriads of textual legal decisions. This paper introduces an approach for predicting Brazilian court decisions which is also able to predict whether the decision will be unanimous. We developed a working prototype which performs 82% of accuracy (F1-score) on a data set composed of $4,043$ cases from a Brazilian court. To our knowledge, this is the first study to forecast judge decisions in Brazil.**

***Index Terms*—legal outcome forecast, predictive algorithms, jurimetrics**

I. Introduction

Since Code of Hammurabi[1], we have been trying to improve legal certainty in human relationships by making public the law and the rulings of courts. In addition to publicizing the laws, legal systems usually provide further support to legal certainty through judicial decisions. These decisions are useful not only for judging specific situations, but also to influence society behavior by exposing the legal consequences of our actions. Thereby, predicting legal decisions is fundamental to understanding the consequences of our actions as well as for supporting law professionals to improve the quality of their work.

In Brazil for example, lower court judges decisions might be appealed to Brazilian courts (*Tribiunais de Justiça*) to be reviewed by second instance court judges. In an appellate court, judges decide together upon a case and their decisions are compiled in Agreement reports named *Acórdãos*. Similar to lower court decisions, Acórdãos include Report, Legal Principles (*Fundamentos*), Votes[2], and further metadata such as judgment date, attorneys, judges, etc. These Agreements documents are very useful for understanding jurisprudence thus guiding lawyers and court members about the decisions. For instance, attorneys often use these documents to prepare cases while judges should take them into account – or even use them as guidelines – for forthcoming decisions.

In order to understand Acórdão decisions, one has to read the subject at the summary, read the decision Report, how each judge voted in this case (Votes), and the final decision which can be unanimous or not. Moreover, each Acórdão might have more than one decision – regarding one or more appealed case claims – which can increase the Acórdão complexity. This problem becomes harder as there usually are hundreds – and sometimes thousands – of Acórdãos related to the case on which a Law professional is working.

A very common and extremely important task for Law professionals is to speculate how a specific court would decide given the ideas and the facts which compose the case [7]. For example, this is useful for preparing and tuning a case to have a favourable decision. Hence, attorneys can rely on substantial assumptions on how judges will decide based on their arguments. Although this information can be found in public Acórdãos, the myriads of available documents make this task very complex and error prone, even for experiment lawyers.

In addition to Brazil, several other legal systems in the world share the very same problem of predicting legal decisions. The challenge is hence generalized as *how to automatically predict legal decisions with a satisfactory level of accuracy* to support the work of attorneys, judges, and other professionals such as counters and real state offices. By satisfactory, we mean that the quality of the prediction in terms of accuracy should be comparable – or even higher – than Law experts.

Nevertheless, it still is very hard to perform any legal decision prediction with satisfactory accuracy, even though computers have been used for such challenge for decades [7]. For instance, Ashley and Brüninghaus [2] proposed a method for classifying and predicting cases which is able to meet 91.8% of accuracy, however the evaluation takes only into account a small data set (146 cases). Katz et al. [6] used historical data for predicting USA Supreme Court decisions by classifying decisions in two and three classes and by presenting judge profiles. That approach reaches 70.2% of accuracy for predicting case decisions and is assessed by on a data set with $28,000$ cases. Also using data from the USA Supreme Court, Ruger et al. [11] exposed how the prediction of Law experts performs in comparison to a trained statistical model for different Law fields by using less than 200 cases. In [1], Aletras et al. used Support-vector Machine (SVM) for predicting if cases from the European Court of Human Rights violate Articles 3, 6, and 8 of the European Convention on Human Rights. Their results achieved 78% of accuracy on 584 European Court cases separated by subjects.

Other related work takes advantage of machine learning

April 21st, 2019. Authorship: A.L.F. proposed the research problem, supervised this work, and executed the experiments. H.A.C. proposed and developed NLP and machine learning algorithms. O.S. verified the methodology. L.O.L. supervised and reviewed the Law aspects of this work. All authors discussed the results and contributed to the final manuscript.

[1]https://en.wikipedia.org/wiki/Code_of_Hammurabi

[2]C.f. Brazilian Law: *Art. 489, Lei 13105/15*.

techniques to support further legal tasks. In [8], the authors proposed a framework for automatically judging legal decisions by using Attention neural network models. They applied the approach for divorce decisions in China. In [13], Shulayeva et al. separated legal principles from case facts on legal documents by using a Naive Bayses Multimodal classifier. In [4], the authors proposed the use of transfer learning to recognize the same words which have different meanings in different contexts, i. e., name-entity linking task. In [3], the authors used Bayesian networks to classify legal decisions from a Brazilian Labor court and concluded that both employees and employers are approximately successful in their litigation. Last, Ruhl et al. [12] exposed perspectives on how complex systems are useful for supporting policy-makers on legal-related topics such as appellate jurisprudence and tax policy analysis.

We are also motivated by recent results that show that intelligent systems can perform better than Law experts[3]. Our hypothesis is that *by taking advantage of Natural Language Processing (NLP) and Machine Learning techniques it is possible to build a system that meets high quality legal decision predictions*. Differently from the closest related works of this paper [1], [2], [6] which addressed United States and European courts, we propose an approach for legal decision prediction for Brazilian courts which also predicts whether the court decision will be unanimous. Moreover, in contrast to [1], [2], we trained a model at thousand-scale data set with $4,043$ cases. Further, in contrast to [1], our approach does not only rely on a binary classification problem nor requires that case data set should be separated by specific Law articles. In this context, the generality of our contribution is greater then the approach presented in [1].

The reminder of this paper is structured as follows. In Section II, we present details on the aforementioned problem such as the case study and the methodology employed. Section III exposes the results while Section IV concludes our investigation and proposes future directions on this subject.

II. Material and Methods

The research question which guides our study is *how to predict legal decisions with a satisfactory level of accuracy for Brazilian courts by including the prediction of the court unanimous behavior*. Next sections provide further information about our assumptions and the proposed methodology.

A. A generic approach

We focus on Brazilian courts as Brazil legal system is not trivial. We believe that if we are able to solve this problem for such complex legal system, our approach would also fit other simpler or it could be straightforwardly adapted for more complex legal systems which share similarities. Nevertheless, it is worth to state that other legal systems also rely on similar documents. For instance, in Indiana Court of Appeal (United States), the Appellate Court decisions are composed by a group of three judges whose decisions (opinions) are divided in Case Summary, Facts, Procedural History, and the court conclusion at the end of that document. In France, the Appellate Court (*Cour d'appel*) also renders decisions coming from the agreement of three judges. That decision is called an *Arrêt* whose structure is also composed of legal basis for the appeal, case history, and the final decision.

Further, we share the same assumption of Aletras et. al [1]: "there is enough similarity between (at least) certain chunks of the text of published judgments and applications lodged with the Court and/or briefs submitted by parties with respect to pending cases".

B. Decision labels and data set

Regarding the flow process of a Brazilian appeal, when lawyers lodge an appeal at a court it is analyzed by a group composed of three judges to check whether the appeal is able to be judged by the court. If the appeal does not meet the formal requirements, the appeal is identified as not cognized (*recurso não conhecido*) hence not judged by the court. Otherwise, the appeal is judged and classified in various categories. We therefore assumed that court decisions can be classified by using the following labels:

- `not-cognized`, when the appeal was not accepted to be judged by the court;
- `yes`, for full favourable decisions;
- `partial`, for partially favourable decisions;
- `no`, when the appeal was denied;
- `prejudicada`, to mean that the case could not be judged for any impediment such as the appealer died or gave up on the case for instance;
- `administrative`, when the decision refers to a court administrative subject as conflict of competence between lower court judges.

In addition to the decision labels, an orthogonal concern of Brazilian court decisions – as well as for other legal systems – refers to its unanimous aspect, being labeled as:

- `unanimous` which means that the decision was unanimous among the three judges that voted in the case; and
- `not-unanimous` by meaning that one of the judges disagreed on the decision.

With respect to the data set, we relied on $4,762$ decisions (*Acórdãos*) from a State higher court (appellate court), the *Tribunal de Justiça de Alagoas*. From this data set, we removed the decisions that had repeated descriptions to not bias the sample thus resulting $4,332$ examples. Repeated decision descriptions occur owing to very similar cases which share the same description. Moreover, for the sake of predictability, we removed all the decisions classified as `prejudicada`, `not-cognized` and `admnistrative` as these labels refer to very peculiar situations which are not useful for prediction purposes addressed by this paper. Finally, the total amount of examples were $4,043$ cases.

C. Methodology

Figure 1 depicts an overview of our methodology. From the legal decision data set, we extracted and separated the

[3]https://www.bbc.com/news/technology-41829534

texts which hold information about the case description, the decision, and their unanimous aspect. This Natural Language Pre-processing task includes removing stop-words and word suffixes for improving the capacity of word representation. Then, we took advantage of Term Frequency-inverse Document Frequency (tdf-df) statistics to increase the importance of relevant words while decreasing the importance of general repetitive words not relevant to the addressed problem. As follows, we used texts which refer to decisions and unanimity to classify them to one of the possible labels (c.f. Section II-B). As a result, we built a structured training data set depicted by Table I.

Next, we used 80% of the data set to train a Machine Learning model which was then assessed by using the latter 20% of the data set (c.f. Figure 1). To train a model means to automatically find out which parameters are the most suitable for predicting decisions based on the training data set. Because we address decision and unanimous predictions, it requires to train two models to address both predictions. Once trained, the models can be used to predict decision and unanimity given a case description. Last, to evaluate our approach, we used the F1-score metrics and performed 5-fold cross-validation to improve the feasibility of our approach. Results are thus exposed as success accuracy rate in percentage.

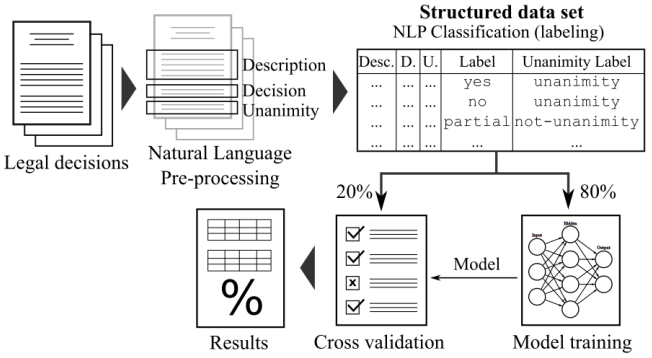

Fig. 1. Methodology. Specific data from case outcomes are extracted then classified to different labels. A machine learning model is hence trained and assessed by using 5-fold cross-validation.

Furthermore, in order to assess the exposed methodology, we developed a working prototype in Python. We used the NLTK framework [9] for Natural Language Processing in such a way that our prototype is easily configurable for various languages in addition to Portuguese. The prototype also provides a graphical user interface which can be accessible from any Web browser.

III. Results

Our approach was able to score **78,99%** F1-score ($\sigma^2 = 0.000017$) when predicting legal decision for the *Tribunal de Justiça de Alagoas* Brazilian court by using $4,043$ judge decisions. The number of samples for each label is exposed in Table II.

In order to analyze our prediction over a more uniformly distributed data set, we randomly removed $1,549$ `no`-labeled decisions to have the same number of `partial`-labeled decisions. Table III depicts the distributions of each decision label. Our assumption was that this data set would strongly bias the model. However, this assumption was not validated as the accuracy of case outcome prediction in this situation was **74.07%** ($\sigma^2 = 0.00029$) for the F1-score metrics.

Furthermore, we also evaluated our approach when reducing the prediction problem to a binary case outcome forecast. Hence, we considered all `partial`-labeled decisions as `yes` resulting in the data set depicted by Table IV. In this scenario, the proposed approach scored **82.03%** ($\sigma^2 = 0.0001$) F1-score metrics.

With respect to predicting the unanimous behaviour of the *Tribunal de Justiça de Alagoas* Brazilian court, our approach scored **98.46%** ($\sigma^2 = 0.000031$) for the F1-score metrics. This assessment was performed in a data set with $2,274$ cases. From the $4,332$ data set – which had no repeated decision descriptions –, we removed the samples that either our classifier did not managed to label or the decision itself did not had any information about unanimity. The resulting data set had $2,289$ samples. As follows, we removed from this data set the decision whose labels were `prejudicada`, `not-cognized` and `admnistrative` – as these labels are not relevant for the predictive addressed problem – resulting in a data set with $2,274$ examples. The distribution of `unanimous` and `not-unanimous` labels are depicted by Table V.

The very-high unanimous predictive accuracy of 98.46% is explained by the fact that most of decisions are unanimous, therefore the model was biased to this label. We indeed expected that most decisions were unanimous since this is well known by law experts. However, the great difference between unanimous and non unanimous decision is a surprising result. In order to understand how our approach would perform when predicting unanimity by using a more uniformly distributed data set, we therefore performed another evaluation by randomly removing decisions whose label was `unanimous` to meet the same number of `not-unanimous` decisions. The resulting data set had 90 examples, half of them labeled as `unanimous` and the other half `not-unanimous`. With this configuration, our prototype reached **76.94%** ($\sigma^2 = 0.015$) F1-score accuracy.

IV. Conclusion

This paper proposes a methodology for predicting Brazilian court legal decisions which is able to reach 82% of accuracy when employed for a Brazilian court data set with $4,043$ cases. In addition to considering a binary predictive problem, i.e. `no` and `yes` predictive results, our approach is also able to predict case outcomes by also predicting `partial` favorable decisions. In this context, our approach accuracy is 79%. Moreover, the proposed method also predicts whether the decision will be unanimous, which fits not only Brazil legal system, but also several others whose decisions are judged by more than one judge. The unanimity prediction performance of our approach is 77% of accuracy. To our knowledge, this is the first study to predict Brazil legal decisions.

Furthermore, our approach is easy to use as it only requires that users provide the description of their litigation and the

TABLE I
TRAINING DATA SET INCLUDES DECISION TEXTS AND LABELS WHICH WERE CLASSIFIED ACCORDING TO RESPECTIVE DECISION TEXTS. E.G., IN SAMPLE 1, *provido* WAS CLASSIFIED AS A FAVORABLE (YES) DECISION AND *Unanimidade* WAS CLASSIFIED AS UNANIMOUS.

Data	Decision description	Decision	Unanimity	Label	Unanimity label
Sample 1	Direito Processual Civil...	Recurso conhecido e provido	Unanimidade	yes	unanimous
Sample 2	Apelação criminal...	Recurso conhecido e parcialmente provido	Decisão unnime	patial	unanimous
Sample 3	Apelação Cível em Ação Ordinária...	Recurso conhecido e não provido	Decisão unânime	no	unanimous

TABLE II
DISTRIBUTION OF DECISIONS ACCORDING TO THEIR LABELS.

Labels	no	partial	yes
N. of decisions	2,415	866	762

TABLE III
DISTRIBUTION OF DECISIONS ACCORDING TO THEIR LABELS WHEN RANDOMLY REMOVING NO-LABELED DECISION SAMPLES TO CREATE A REGULAR DISTRIBUTED DATA SET.

Labels	no	partial	yes
N. of decisions	866	866	762

TABLE IV
DISTRIBUTION OF DECISIONS ACCORDING TO THEIR LABELS WHEN CONSIDERING A BINARY PREDICTION PROBLEM (NO AND YES LABELS ONLY).

Labels	no	yes
N. of decisions	2,415	1,628

TABLE V
DISTRIBUTION OF DECISIONS ACCORDING TO JUDGE UNANIMOUS BEHAVIOR.

Labels	not-unanimous	unanimous
N. of decisions	45	2,229

output will be one of the aforementioned case outcome predictive label along with its predictive unanimity label. This information is very relevant for attorneys, judges, and other Law professionals as it provides practical support for their work. Moreover, our contribution also includes a working prototype which can be configured to other languages as well as for different data sets.

Although we believe that our contribution is quite satisfactory given the accuracy rate aforementioned, future investigations might consider comparing our results with Law experts, as performed in [11] and by current Lawtech products such as Case Crunch and LawGeex[4]. Other future work includes to investigate whether taking advantage of existent Named-entity recognition data sets for Brazilian law documents [10] improves the prediction quality. Furthermore, the assessment of the proposed method can be performed on other data sets, such as the European Court of Human Rights for instance. Ultimately, despite the various directions one might take to take advantage of our work, we believe that Mireille Hildebrandt's viewpoint on "agnostic machine learning" and its consequences to the Rule of Law [5] should be taken into account when designing and using a legal predictive system.

[4]https://www.artificiallawyer.com/2018/02/26/lawgeex-hits-94-accuracy-in-nda-review-vs-85-for-human-lawyers/

V. ACKNOWLEDGMENTS

We thank the anonymous reviewers who provided significant and constructive critics about this manuscript. We declare no competing interests and that we received no funding for this study.

REFERENCES

[1] N. Aletras, D. Tsarapatsanis, D. Preoiuc-Pietro, and V. Lampos. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. *PeerJ Computer Science*, 2:e93, oct 2016.

[2] K. D. Ashley and S. Brüninghaus. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17(2):125–165, jun 2009.

[3] R. Barros, F. Lorenzi, and L. K. Wives. *Recent Trends and Future Technology in Applied Intelligence*, volume 10868 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham, 2018.

[4] A. Elnaggar, R. Otto, and F. Matthes. Named-Entity Linking Using Deep Learning For Legal Documents: A Transfer Learning Approach. Technical report, Technische Universität, München, 2018.

[5] M. Hildebrandt. Algorithmic regulation and the rule of law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128):20170355, sep 2018.

[6] D. M. Katz, M. J. Bommarito, and J. Blackman. A general approach for predicting the behavior of the Supreme Court of the United States. *PLOS ONE*, 12(4):e0174698, apr 2017.

[7] L. Loevinger. Jurimetrics: The Methodology of Legal Inquiry. *Law and Contemporary Problems*, 28(1):5, 1963.

[8] S. Long, C. Tu, Z. Liu, and M. Sun. Automatic Judgment Prediction via Legal Reading Comprehension. Technical report, Tsinghua University, Beijing, sep 2018.

[9] E. Loper and S. Bird. NLTK. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics* -, volume 1, pages 63–70, Morristown, NJ, USA, 2002. Association for Computational Linguistics.

[10] P. H. Luz de Araujo, T. E. de Campos, R. R. R. de Oliveira, M. Stauffer, S. Couto, and P. Bermejo. LeNER-Br: A Dataset for Named Entity Recognition in Brazilian Legal Text. In *Proceedings of the International Conference on the Computational Processing of Portuguese*, pages 313–323, Canela, 2018.

[11] T. W. Ruger, P. T. Kim, A. D. Martin, and K. M. Quinn. The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking. *Columbia Law Review*, 104(4):1150, may 2004.

[12] J. B. Ruhl, D. M. Katz, and M. J. Bommarito. Harnessing legal complexity. *Science*, 355(6332):1377–1378, mar 2017.

[13] O. Shulayeva, A. Siddharthan, and A. Wyner. Recognizing cited facts and principles in legal judgements. *Artificial Intelligence and Law*, 25(1):107–126, mar 2017.