



UNIVERSITY OF
OXFORD



Predicting Legal Outcomes With Deep Learning

Alina Petrova, Thomas Lukasiewicz and John Armour

Alina Petrova



alina.petrova@cs.ox.ac.uk



<http://www.cs.ox.ac.uk/people/alina.petrova>



WP3: Frontiers of AI in Legal Reasoning

The screenshot shows a dark-themed website for the University of Oxford Faculty of Law. In the top left corner are the logos for the University of Oxford and the Faculty of Law. The top navigation bar includes links for Admissions, Research, Centres & Institutes, News, Events, People, About us, and Vacancies. A breadcrumb trail 'Research >' is visible above the main title. The main title is 'Unlocking the Potential of Artificial Intelligence for English Law' in large, white, serif font. Below the title is a paragraph of text about the project's funding by the Industrial Strategy Challenge Fund's (ISCF) Next Generation Services Research Programme and UK Research and Innovation (UKRI). To the right of the main content area is a sidebar with a stylized icon of a person with a circuit board head holding a balance scale, followed by a list of links: 'In Artificial Intelligence and English Law', 'Blog', 'People', 'Blog', 'Blog', and 'About us'. The main content area contains two sections of text: one about the potential and limitations of AI in legal services, and another about legal services contributing to the economy through revenues of service providers and through legal services.

UNIVERSITY OF OXFORD
FACULTY OF LAW

Admissions ▾ Research ▾ Centres & Institutes ▾ News Events ▾ People ▾ About us ▾ Vacancies

Research >

Unlocking the Potential of Artificial Intelligence for English Law

A project funded by the Industrial Strategy Challenge Fund's (ISCF) Next Generation Services Research Programme and UK Research and Innovation (UKRI), run by researchers in the Oxford departments and faculties of Law, Economics, Computer Science, Education and the Said Business School, in 2019 and 2020.

The proposed research will explore the potential and limitations of using artificial intelligence (AI) in support of legal services. AI's capabilities have made enormous recent leaps; many expect it to transform how the economy operates. In particular, activities relying on human knowledge to create value, insulated until now from mechanisation, are facing dramatic change.

Legal services contribute to the economy both through revenues of service providers and through

In Artificial Intelligence and English Law

Blog

People

Blog

Blog

About us

<https://www.law.ox.ac.uk/unlocking-potential-artificial-intelligence-english-law>

Modern Legal NLP Tasks



- text summarization
- information extraction: identify specific clauses, find citations, named entities etc.
- (outcome) prediction: charges, fines, length of sentences

Prominent works:

Alet16, Katz17, Sule17, Chal19, Lim19, Medv20, Chal20
Luo17, Hu18, Zhon18, Yang19, Chen19

Advances in Legal NLP

Data types:

- legal contracts,
- **court cases** (proceedings or legal reports)

Sources:

- French Supreme Court,
- the European Court of Justice,
- the European Court of Human Rights,
- Chinese Criminal Law dataset,
- **Caselaw Access Project**,
- BAILII library, etc.

*Extracting Outcomes from
Appellate Decisions in US State Courts*

JURIX 2020

Caselaw Access Project (CAP)



- run by Harvard Law School: <https://case.law>
- 7 mio cases from State and Federal Courts across all states
- 360k appellate cases from 4 State Courts publicly available
- **Caselaw4**: 360k judicial decisions from CAP, from which 250k automatically annotated with outcome labels

Data Annotation

- **Caselaw4:** 360k case reports from New Mexico, North Carolina, Illinois, and Arkansas Courts of Appeal
- **Possible outcomes:**
 - the previous ruling is kept as is (*Affirm*)
 - the previous ruling is changed/annulled (*Reverse*)
 - some parts of the previous ruling are kept and some are changed (*Mixed*)
 - the appeal is dismissed (a type of *Affirm*)
- **Bootstrap annotation:**
 - **500 cases** from the New Mexico Court of Appeals were **manually annotated** with the *Affirm*, *Reverse*, or *Mixed* outcome label
 - **250k cases** were further annotated automatically

Manual Annotation



OPINIONDONNELLY, Judge: Worker Peter Herrera appeals from the Workers' Compensation Judge's (WCJ's) order reducing Worker's compensation to be paid by Employer Quality Imports based upon Worker's receipt of a college degree after the original compensation order was entered. Worker raises a single issue on appeal: whether the WCJ erred in changing the NMSA 1978, 52-1-26.3 (1990, effective Jan. 1, 1991), statutory modifier based upon a change in Worker's educational status that occurred after the trial. For the reasons discussed herein, we reverse.
FACTS On March 29, 1996, Worker was injured while working as a sales representative for Respondent. On December 10, 1997, following a trial on the merits, the WCJ awarded permanent partial disability benefits (PPD) based upon a 17% PPD rating. The 17% rating determined by the WCJ was based upon a 5% permanent impairment, plus twelve additional points for age, education, experience, training, and residual physical capacity. The WCJ determined that, as of the date of the trial, Worker had a high school education and was entitled to no points for an education modifier pursuant to Section 52-1-26.3(B)(3).{3} On September 29, 1998, Employer filed a motion under NMSA 1978, 52-1-56 and 52-5-9 (1989), to reduce the 17% disability rating based upon evidence that Worker had recently obtained a college degree from the University of New Mexico. Employer argued that because Worker had

Annotator(s) AG

Document Classification ⓘ

Complex Dismiss
 Reverse Affirm

Outcome 2 annotations ⏷

For the reasons discussed herein, we reverse. ⏷

The WCJ's order modifying compensation order is rever... ⏷

[Clear all](#)

500 cases were manually annotated with the *Affirm*, *Reverse*, or **Mixed outcome labels** as well as with the **outcome sentences**

Outcome extraction

We split the process of extracting the outcome from a case into **two steps**:

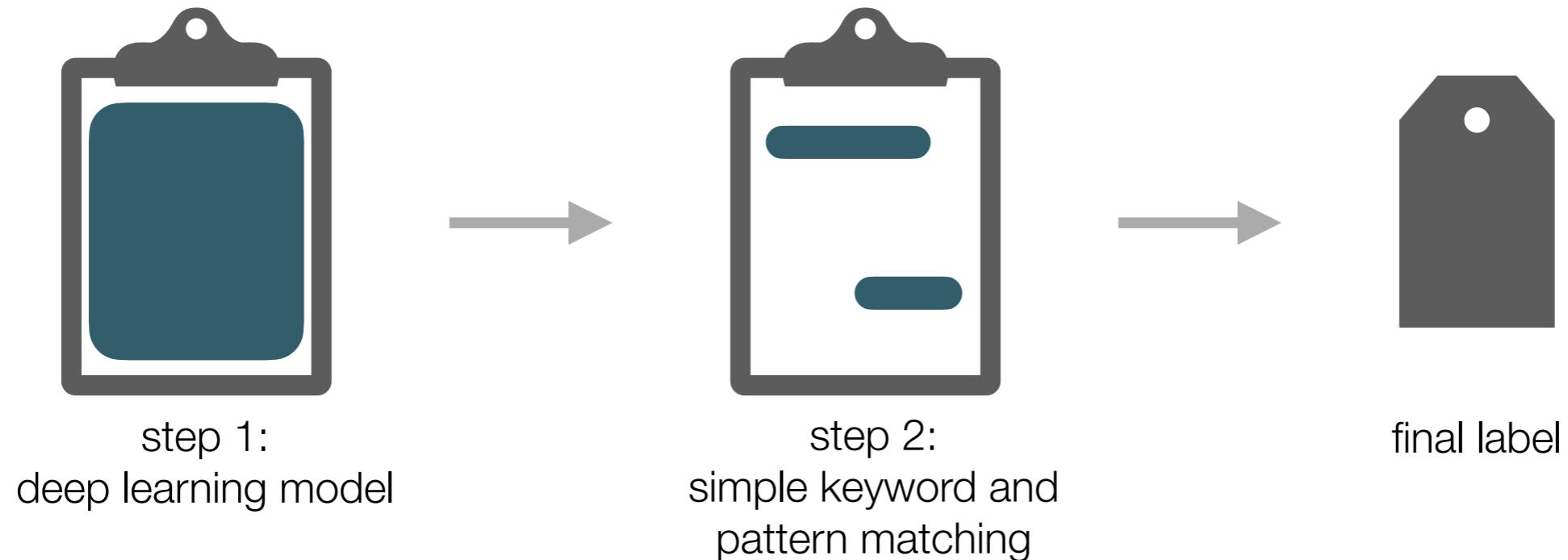
1. we select all the sentences that contain the outcome description:

The chancellor's order for alimony will be continued until final decree is entered on remand of the cause. In other respects the decree will be affirmed.

2. we decide on the outcome label based on the sentences only: *Affirm*



Extraction Methodology



Main constraint: amount of annotated data

- cannot do full-text DL extraction from just 500 cases
- keyword search on full text is inaccurate (best effort 81.32% F1)
- sentence extraction from 1.5k outcome sentences (+1.5 non-outcome sentences) is still possible with a simple neural network
- keyword search from pre-filtered sentences works fine (and DL is not possible over 500 instances)

Step 1: Sentence Extraction

- From 500 cases, we extracted 1455 outcome sentences (and 92k non-outcome sentences).
- With dataset re-balancing, we got 3k sentences for binary classification task: *outcome* vs. *non-outcome*.
- We trained a simple bi-LSTM network (single hidden layer with 50 neurons, embedding size 200, input size 100). Such network has few parameters, and in a sense **mimics keyword search**, but with better results:

	Precision	Recall	F-measure
OUTCOME	97.90	95.89	96.89
NON-OUTCOME	96.49	98.21	97.35
Total	97.15	97.13	97.13

Step 2: Outcome Extraction

Simple keyword matching:

- if the pre-filtered sentences contain a token ‘affirm’ or ‘dismiss’, the outcome is *Affirm*;
- if they contain a token ‘reverse’, the outcome is *Reverse*;
- if both ‘affirm’/‘dismiss’ and ‘reverse’ are present, the outcome is *Mixed*.

Main issue is transferring the model to other jurisdictions:

- out-of-vocabulary words
- grammatical and stylistic variability
- recall drops: some *Affirm* and *Reverse* cases are missed and labelled as *Mixed*

Step 2: Outcome Extraction

- We used a combination of sentence-level and word-level patterns to annotate all 360k cases with outcome labels based on outcome sentences.
- **Label validation:** manual validation of 100 labels, 25 per jurisdiction.

	Precision	Recall	F-measure
AFFIRM	93.18	78.85	85.42
REVERSE	100.00	80.77	89.36
MIXED	54.29	86.36	66.67
Total	86.40	81.00	82.32

Performance of the outcome extraction model

True label	Predicted label			
		AFFIRM	REVERSE	MIXED
AFFIRM	AFFIRM	41	0	11
REVERSE	REVERSE	0	21	5
MIXED	MIXED	3	0	19

Confusion matrix

Automatic Annotation

- We tried predicting the label for all 360k cases, then kept only *Affirm* and *Reverse* labels in the final annotation.
- Annotated subset of Caselaw4: 170k *Affirm* cases, 84.5k *Reverse* cases
- The whole dataset, with both manual and automatic annotations, is publicly available: <https://github.com/chinmusique/outcome-prediction>

Challenges of outcome prediction

Next Step: Outcome Prediction?

Suppose outcome extraction is solved. Can we directly move to outcome prediction?

Next Step: Outcome Prediction?

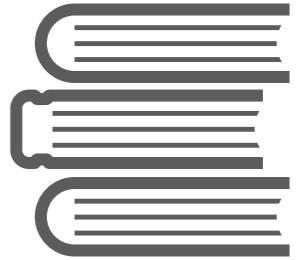
Suppose outcome extraction is solved. Can we directly move to outcome prediction?

Not yet: using case reports as is does not make sense (yet they are our main source of data).

- we need to remove all outcome sentences
- we need to separate original facts from previous rulings and from legal reasoning

OPINION OF THE COURT. PARKER, J.

This is an appeal from the District Court of Bernalillo County dismissing a bill of complaint for divorce. The court made findings of fact which would seem to present a serious question as to the correctness of the decree. Appellee claims, however, that the findings were inadvertently made by the court without notice to his counsel, and that they do not correctly represent the actual state of facts proved. **Under the circumstances, the validity of the decree being in doubt, we deem, it advisable to reverse the decree and remand the cause with instructions to proceed further, and it is so ordered.**



Intermediate Takeaways

Data bottleneck:

- availability of legal data to the community
- data amount
- metadata, data structure and annotation
- data homogeneity



Thank you!

References

- [Chal19]** Chalkidis I, Androutsopoulos I, Aletras N. *Neural legal judgment prediction in English*. ACL 2019.
- [Alet16]** Aletras N et al. *Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective*. PeerJ Computer Science 2016.
- [Medv20]** Medvedeva M, Vols M, Wieling M. *Using machine learning to predict decisions of the European Court of Human Rights*. Artificial Intelligence and Law 2020.
- [Katz17]** Katz DM, Bommarito MJ, Blackman J. *A general approach for predicting the behavior of the Supreme Court of the United States*. PloS one 2017.
- [Sule17]** Sulea OM et al. *Predicting the law area and decisions of French Supreme Court cases*. RANLP 2017.
- [Lim19]** Lim C. *An evaluation of machine learning approaches to Natural Language Processing for legal text classification*. Imperial College London 2019.
- [Luo17]** Luo B, Feng Y et al. *Learning to predict charges for criminal cases with legal basis*. EMNLP 2017.
- [Zhong18]** Zhong H, Guo Z et al. *Legal judgment prediction via topological learning*. EMNLP 2018.
- [Yang19]** Yang W et al. *Legal judgment prediction via multi-perspective bi-feedback network*. IJCAI 2019.
- [Hu18]** Hu Z, Li X et al. *Few-shot charge prediction with discriminative legal attributes*. COLING 2018.
- [Chen19]** Chen H, Cai D et al. *Charge-based prison term prediction with deep gating network*. EMNLP 2019.
- [Chal20]** Chalkidis I, Fergadiotis M et al. *LEGAL-BERT: The Muppets straight out of Law School*. arXiv 2020.
- [Elwa19]** Elwany E, Moore D, Oberoi G. *BERT Goes to Law School: Quantifying the Competitive Advantage of Access to Large Legal Corpora in Contract Understanding*. arXiv 2019.

Images Used

- <https://unsplash.com/photos/DZpc4UY8ZtY>
- <https://unsplash.com/photos/6sl88x150Xs>