Lex Rosetta: Transfer of Predictive Models Across Languages, Jurisdictions, and Legal Domains

Jaromir Savelka[1], Hannes Westermann[2], Karim Benyekhlef[2], Charlotte S. Alexander[3], Jayla C. Grant[3], David Restrepo Amariles[4], Rajaa El Hamdani[4], Sébastien Meeùs[4], Aurore Troussel[4], Michał Araszkiewicz[5], Kevin D. Ashley[6], Alexandra Ashley[6], Karl Branting[7], Mattia Falduti[8], Matthias Grabmair[9], Jakub Harašta[10], Tereza Novotná[10], Elizabeth Tippett[11], Shiwanni Johnson[11]

[1]Carnegie Mellon University, [2]Université de Montréal, [3]Georgia State University, [4]HEC Paris, [5]Uniwersytet Jagielloński, [6]University of Pittsburgh, [7]MITRE Corporation, [8]Libera Università di Bolzano, [9]Technische Universität München, [10]Masarykova univerzita, [11]University of Oregon

International Conference on Artificial Intelligence and Law

June 23, 2021

1

Overview

- Task relevant across domains, jurisdictions, languages, legal systems
- Pre-trained multilingual embeddings enable ML models to generalize
- Release of type schema, annotated dataset, code

Introduction

Machine learning models are usually trained and evaluated on data with the same distribution.

Dataset assembled -> Split -> Used for training and testing model.

Can it generalize? Learned specific information? Important in practice.

Is it possible to build models that generalize between contexts?

Is it possible to build models that generalize between contexts?

Languages,
Jurisdictions
Legal domains?

I. Kasační stížnost se zamítá. II. Žalovaná nemá právo na náhradu nákladů řízení o kasační stížnosti.

III. [...]uplně žalobkyně Mgr. Monice Ipserové, advokátce, se sídlem Sladkovského 505, Pardubice, se přiznává odm[...]iní náhrada hotových výdajů ve výši 960 Kč, která je splatná do 60 dnů od právní moci tohoto rozsudku z účtu [...]du. Žalobkyně (d[...] "stěžovatelka") se kasační [...]nností domáhá zrušení v záhlaví označeného usnesení krajského soudu v Hradci Králové - po[...]ky v [...] soudu pod sp. zn. 50 Cad 6/2008.

Krajský soud konstatoval, že stěžovatelka již je [...] usnesením ze dne 14. 10. 2010, č. j. Na 1056/2010 - 30.

Opětovnému rozhodování o tomto návrhu tedy brání překážka věci rozsouzené.

Na závěr krajský soud stěžovatelku upozornil, že má-li k dispozici nové lékařské posudky, může podat novou žádost o přiznání plného invalidního důchodu.

V kasační stížnosti stěžovatelka namítá, že ve svém druhém návrhu na obnovu řízení uplatnila skutečnosti a důkazy, které nemohla uplatnit v řízení původním.

Jde zejména o závěry endokrinologické a ortopedické (padající klenba chodidel) a rovněž o skutečnosti relevantní pro posouzení snížení schopnosti soustavné výdělečné činnosti.

Stěžovatelka rovněž tvrdí, že krajský soud neměl důvod nevyhovět jejímu prvnímu návrhu na obnovu řízení s odkazem na rozhodnutí Nejvyššího správního soudu ze dne j. 4 As 25/2008 - 107, neboť se v jejím případě skutečně jedná o ochranu před zásahem správního orgánu, a návrh na obnovu řízení tedy měl být podle § 114 zákona č. 150/2002 Sb., soudní řád správní, (dále jen "s. ř. s."), přípustný. Žalovaná ve svém vyjádření ke kasační stížnosti uvedla, že plně souhlasí s odůvodněním krajského soudu, podle něhož se v daném případě jedná o překážku věci rozsouzené.

Poté, co Nejvyšší správní soud zhodnotil, že se jedná o kasační stížnost přípustnou, zabýval se v rozsahu uplatněného kasačního důvodu podle § 103 odst. 1 písm. e) s. ř. s. její důvodností.

Pro posouzení této kasační stížnosti je stěžejní především tvrzení stěžovatelky, že se v případě nyní odmítnutého návrhu na obnovu řízení jedná o návrh odlišný od toho, o němž krajský soud rozhodl dne 14. 10. 2010.

Totožnost návrhu je podle dosavadní judikatury zdejšího soudu dána totožností účastníků řízení a napadeného rozhodnutí (srov. např. rozsudek Nejvyššího správního soudu ze dne 16. 6. 2011, č. j. 2 As 63/2011 - 60, všechna citovaná rozhodnutí Nejvyššího správního soudu jsou dostupná na www.nssoud.cz).

V posuzovaném případě vzešly oba návrhy od téže osoby - stěžovatelky - a v obou případech šlo o návrh na obnovu řízení o žalobě proti rozhodnutí žalované ze dne 16. 7. 2008, č. X.

V souladu s výše uvedeným se tedy o totožný návrh nepochybně jedná.

Podle § 46 odst. 1 písm. a) s. ř. s. přitom soud usnesením odmítne návrh, jestliže o téže věci již rozhodl nebo o téže věci již řízení u soudu probíhá.

V nyní posuzované věci je zřejmé, že překážka věci rozsouzené nastala.

Na tom nic nezmění ani tvrzení stěžovatelky, že ve druhém návrhu na obnovu řízení uplatnila nové skutečnosti.

Stěžovatelka soudu společně s tímto návrhem doručila asi dvacet dokumentů, převážně lékařských zpráv, které se datují z let 1992 - 2010. Část z nich tedy mohla doložit již společně se svou žádostí o plný invalidní důchod, kterou podala dne 7. 5. 2008 (stěžovatelka minimálně v návrhu neuvedla žádnou konkrétní skutečnost, která by svědčila o tom, že je k žádosti připojit nemohla).

Ostatní dokumenty jsou pozdějšího d[...] než 7. 5. 2008, kdy podala žádost o důchod.

Nejvyšší správní soud tedy nemohl [...] postupu krajského soudu.

Nejvyšší správní soud poukazuje [...]form[...] že má-li stěžovatelka za to, že u ní došlo ke zhoršení zdravotního stavu, je oprávněna požádat o plný invalidní důchod opětovně, a tím zahájit nové řízení o přiznání této důchodové dávky.

Stěžovatelka současně uplatnila námitku vůči rozhodnutí ze dne 14. 10. 2010, č. j. Na 1056/2010 - 30, jímž krajský soud rozhodl o jejím prvním návrhu na obnovu řízení.

Tvrdí, že krajský soud neměl důvod tomuto návrhu nevyhovět s odkazem na § 114 s. ř. s., neboť se v daném případě jednalo o ochranu před zásahem správního orgánu a podmínky pro obnovu řízení tak byly splněny.

Tuto námitku však Nejvyšší správní soud nepovažuje za relevantní, neboť nyní projednávaná kasační stížnost směřuje proti jinému rozhodnutí krajského soudu (ze dne 9. 5. 2011, č. j. 50 Cad 6/2008 - 136).

Pouze na okraj však zdejší soud odkazuje na svůj rozsudek ze dne 29. 4. 2005, č. j. 4 As 17/2004 - 55, publikovaný pod č. 626/2005 Sb. NSS, z něhož vyplývá, že „pod pojem zásahu spadá velké množství faktických činností správních orgánů, ke kterým jsou různými zákony oprávněny.

Jde o úkony neformální, pro které mohou a nemusí být stanovena pravidla, např. faktické pokyny (typické v dopravě), bezprostřední zásahy (při ohrožení, při demonstraci, příkazy ke zjednání nápravy), zajišťovací úkony atd., tedy obecně úkony, které nejsou činěny formou rozhodnutí, ale přesto jsou závazné pro osoby, vůči nimž směřují, a ty jsou povinny na jejich základě něco konat, nějaké činnosti se zdržet nebo nějaké jednání strpět, a to na základě jak písemného, tak i faktického (ústního či jinak vyjádřeného) pokynu či příkazu. "

Z výše uvedeného lze dovodit, že rozhodnutí žalované o nepřiznání dávky plného invalidního důchodu není možné považovat za nezákonný zásah, neboť jde o typické rozhodnutí správního orgánu, proti němuž lze podat žalobu ve smyslu § 4 odst. 1 písm. a) s. ř. s. Byť uvedené rozhodnutí stěžovatelka mohla považovat za citelný zásah do svého života, není toto rozhodnutí nezákonným zásahem ve smyslu § 82 s. ř. s., neboť dikce uvedeného ustanovení předpokládá, že zásah, pokyn nebo donucení správního orgánu není rozhodnutím.

Ze shora uvedených důvodů

Nejvyšší správní soud nepovažoval kasační stížnost za důvodnou a v souladu s § 110 odst. 1 poslední větou s. ř. s. ji zamítl.

O náhradě nákladů řízení o kasační stížnosti Nejvyšší správní soud rozhodoval podle § 60 s. ř. s. ve spojení s § 120 s. ř. s. Stěžovatelce, jako neúspěšnému účastníku řízení, právo na náhradu nákladů řízení podle § 60 odst. 1 s. ř. s. nepřísluší. Žalované, jako správní orgán ve věcech důchodového pojištění, pak nemá na náhradu nákladů řízení právo podle § 60 odst. 2 s. ř. s.

Stěžovatelce byla usnesením krajského soudu č. j. 50 Cad 6/2008 - 155 ze dne 15. 6. 2011 ustanovena pro řízení o kasační stížnosti zástupkyně Mgr. Monika Ipserová, advokátka, se sídlem Sladkovského 505, Pardubice.

Na základě výzvy k vyčíslení nákladů zastoupení tato zástupkyně požaduje odměnu celkem 2400 Kč za dva úkony právní služby (první porada s klientkou včetně převzetí a přípravy zastoupení a sepis kasační stížnosti).

Vzhledem k tomu, že [...] složit [...] by [...] za převzetím zastoupení a přípravou kasační stížnosti sest[...] se [...]rat [...]ávní [...]i prvního úkonu právní služby - sepis kasační stížnosti - podle § 11 odst. [...] advokátního tarifu [...] o poskytování právních služeb advokátní tarifu.

Odměna za jeden úkon právní služby činí podle § 9 odst. 2 ve spojení s § 7 citované vyhlášky 500 Kč.

Dále náleží ustanovené zástupkyni stěžovatelky paušální náhrada hotových výdajů ve výši 300 Kč podle § 13 odst. 3 advokátního tarifu.

Vzhledem k tomu, že je ustanovená zástupkyně plátkyní DPH, přiznal jí Nejvyšší správní soud na odměně za zastupování a náhradě hotových výdajů celkem 960 Kč a pro zaplacení odměny stanovil přiměřenou lhůtu.

Poučení: Proti tomuto rozsudku nejsou opravné prostředky přípustné.

This case arises under the employee protection provision of the Consumer Financial Protection Act of 2010, Section 1057 of the [Dodd-Frank Wall Stre]et Reform and Consumer Protection Act of 2010, 12 U.S.C. § 5567 (2010) (CFPA). [On ...], [J]on[ath]an D. Brofford filed a complaint with the Occupational Safety and Health Administration (OSHA) [alleging that P]NC [Inve]stments LLC (PNC) fired him in November 2012 for engaging in CFPA-protected activity.

OSHA dismissed the complaint as untimely.

Brofford requested a hearing on his complaint before an Administrative Law Judge (ALJ).

While Brofford's case was pending before the ALJ, PNC filed a Motion to Dismiss (Motion), with exhibits, arguing that the complaint was untimely.

Brofford, appearing pro se, responded to the Motion by submitting a document entitled "Response to PNCI 's Motion for Summary Dismissal," with exhibits.

On September 21, 2017, the ALJ issued a Decision and Order Dismissing Complaint (D. & O.).1 Brofford has appealed the ALJ's ruling to the Administrative Review Board (ARB)

The Secretary of Labor has delegated to the ARB the authority to issue final agency decisions under the CFPA and its implementing regulations at 29 C.F.R. Part Responsibility to the Administrative Review Board, 77 Fed. Reg. 69,378 ; 69,379 (Nov. 16, 2012); 29 C.F.R. § 1985.110(a).

The ARB reviews an ALJ's grant of summary decision de novo under the same standard the ALJ applies.

Summary decision is permitted where "there is no genuine dispute as to any material fact and the movant is entitled to decision as a matter of law." 29 C.F.R. § 18.72(a) (2018).

The ARB views the record on the whole in the light most favorable to the non-moving party.

Stroud v. Mohegan Tribal Gaming Auth., ARB Nos. 13-079, 14-013, ALJ Nos. 2013-ACA-003, 2013-CFP-003, slip op. at 2 (ARB Nov. 26, 2014).

Under the CFPA, a retaliation complaint must be filed within 180 days of the alleged adverse action. 12 U.S.C. § 5567(c)(1)(A).

In determining whether the Board should permit the adjudication of an otherwise untimely complaint, we have recognized four principal situations in which equitable modification of filing deadlines may apply: (1) respondent has actively misled the complainant regarding the cause of action; (2) complainant has in some extraordinary way been prevented from filing his or her action; (3) complainant has raised the precise statutory claim in issue but has done so in the wrong forum; and (4) respondent's own acts or omissions have lulled the complainant into foregoing prompt attempts to vindicate his or her rights.

Turin v. AmTrust Financial Services, Inc., ARB No. 11-062, ALJ No. 2010-SOX-00018, slip op. at 8 (March 29, 2013). motion for summary decision. See, e.g., D. & O. at 2 ("I find there is no genuine issue of material fact relative to Complainant's filing of his claim, that it is time barred, and that Respondent is entitled to judgment as a matter of law.").

If matters outside a motion to dismiss and any responsive pleading are presented to and not excluded by the administrative law judge, it is appropriate to treat the original motion as one for summary decision, see Fed.

R. Civ. P. 12(d), and analyze the evidence under the framework provided in 29 C.F.R. § 18.72 (2018).

A review of the record reveals that Brofford filed his complaint over three years and ten months after PNC terminated his employment.

The ALJ took into consideration not only that Brofford was appearing pro se but also that whistleblower limitations periods are subject to modification (i.e., equitable tolling) when the untimeliness is the result of circumstances beyond the complainant's control.

But ultimately the ALJ held that Brofford had failed to establish that he was entitled to an extension of the filing period.

Brofford focuses on what he asserts is PNC'S pretextual basis for his termination.

Specifically, he asserts that the "reason [for his termination] offered by the respondent is utterly false and misled me to falsely believe that PNC fired me for a legitimate company policy violation with a universally applied pre-prescribed penalty of termination."

Complainant's Petition for Review at 1.

We are not persuaded that allegations of a pretextual termination, without more, are sufficient to constitute the compelling circumstances that would justify departure from the statutory filing deadlines applicable to complaints under the CFPA.

Many whistleblower complaints allege pretextual termination, and to hold that such allegations are sufficient to excuse untimely filing would create an exception that would largely swallow the rule requiring timely filing of whistleblower complaints.

For these reasons, we agree with the ALJ that there is no genuine dispute as to any material fact concerning Brofford's untimely filing and PNC is entitled to decision as a matter of law.

Our review of the record discloses no ground for an equitable extension of the statutory filing deadline of 180 days.

We therefore AFFIRM the ALJ's decision and order and this case is hereby DISMISSED. SO ORDERED.

This case arises under the employee protection provision of the Consumer Financial Protection Act of 2010, Section 1057 of the D_____ ____m and Consumer Protection Act of 2010, 12 U.S.C. § 5567 (2010) (CFPA).

O_____ ___an D. Brofford filed a complaint with the Occupational Safety and Health Administration (OSHA) a_____stments LLC (PNC) fired him in November 2012 for engaging in CFPA-protected activity.

OSHA dismissed the complaint as untimely.

Brofford requested a hearing on his complaint before an Administrative Law Judge (ALJ).

While Brofford's case was pending before the ALJ, PNC filed a Motion to Dismiss (Motion), with exhibits, arguing that the complaint was untimely.

Brofford, appearing pro se, responded to the Motion by submitting a document entitled "Response to PNCI 's Motion for Summary Dismissal," with exhibits.

On September 21, 2017, the ALJ issued a Decision and Order Dismissing Complaint (D. & O.).1 Brofford has appealed the ALJ's ruling to the Administrative Review Board (ARB)

The Secretary of Labor has delegated to the ARB the authority to issue final agency decisions under the CFPA and its implementing regulations at 29 C.F.R. Part Responsibility to the Administrative Review Board, 77 Fed. Reg. 69,378 ; 69,379 (Nov. 16, 2012); 29 C.F.R. § 1985.110(a).

The ARB reviews an ALJ's grant of summary decision de novo under the same standard the ALJ applies.

Summary decision is permitted where "there is no genuine dispute as to any material fact and the movant is entitled to decision as a matter of law." 29 C.F.R. § 18.72(a) (2018).

The ARB views the record on the whole in the light most favorable to the non-moving party.

Stroud v. Mohegan Tribal Gaming Auth., ARB Nos. 13-079, 14-013, ALJ Nos. 2013-ACA-003, 2013-CFP-003, slip op. at 2 (ARB Nov. 26, 2014).

Under the CFPA, a retaliation complaint must be filed within 180 days of the alleged adverse action. 12 U.S.C. § 5567(c)(1)(A).

In determining whether the Board should permit the adjudication of an otherwise untimely complaint, we have recognized four principal situations in which equitable modification of filing deadlines may apply: (1) respondent has actively misled the complainant regarding the cause of action; (2) complainant has in some extraordinary way been prevented from filing his or her action; (3) complainant has raised the precise statutory claim in issue but has done so in the wrong forum; and (4) respondent's own acts or omissions have lulled the complainant into foregoing prompt attempts to vindicate his or her rights.

Turin v. AmTrust Financial Services, Inc., ARB No. 11-062, ALJ No. 2010-SOX-00018, slip op. at 8 (March 29, 2013). motion for summary decision. See, e.g., D. & O. at 2 ("I find there is no genuine issue of material fact relative to Complainant's filing of his claim, that it is time barred, and that Respondent is entitled to judgment as a matter of law.").

If matters outside a motion to dismiss and any responsive pleading are presented to and not excluded by the administrative law judge, it is appropriate to treat the original motion as one for summary decision, see Fed.

R. Civ. P. 12(d), and analyze the evidence under the framework provided in 29 C.F.R. § 18.72 (2018).

A review of the record reveals that Brofford filed his complaint over three years and ten months after PNC terminated his employment.

The ALJ took into consideration not only that Brofford was appearing pro se but also that whistleblower limitations periods are subject to modification (i.e., equitable tolling) when the untimeliness is the result of circumstances beyond the complainant's control.

But ultimately the ALJ held that Brofford had failed to establish that he was entitled to an extension of the filing period.

Brofford focuses on what he asserts is PNC'S pretextual basis for his termination.

Specifically, he asserts that the "reason [for his termination] offered by the respondent is utterly false and misled me to falsely believe that PNC fired me for a legitimate company policy violation with a universally applied pre-prescribed penalty of termination."

Complainant's Petition for Review at 1.

We are not persuaded that allegations of a pretextual termination, without more, are sufficient to constitute the compelling circumstances that would justify departure from the statutory filing deadlines applicable to complaints under the CFPA.

Many whistleblower complaints allege pretextual termination, and to hold that such allegations are sufficient to excuse untimely filing would create an exception that would largely swallow the rule requiring timely filing of whistleblower complaints.

For these reasons, we agree with the ALJ that there is no genuine dispute as to any material fact concerning Brofford's untimely filing and PNC is entitled to decision as a matter of law.

Our review of the record discloses no ground for an equitable extension of the statutory filing deadline of 180 days.

We therefore AFFIRM the ALJ's decision and order and this case is hereby DISMISSED. SO ORDERED.

Cross-context prediction

What happens if we train a model on multiple datasets?

What happens if we train a model on multiple datasets, including data from the target context?
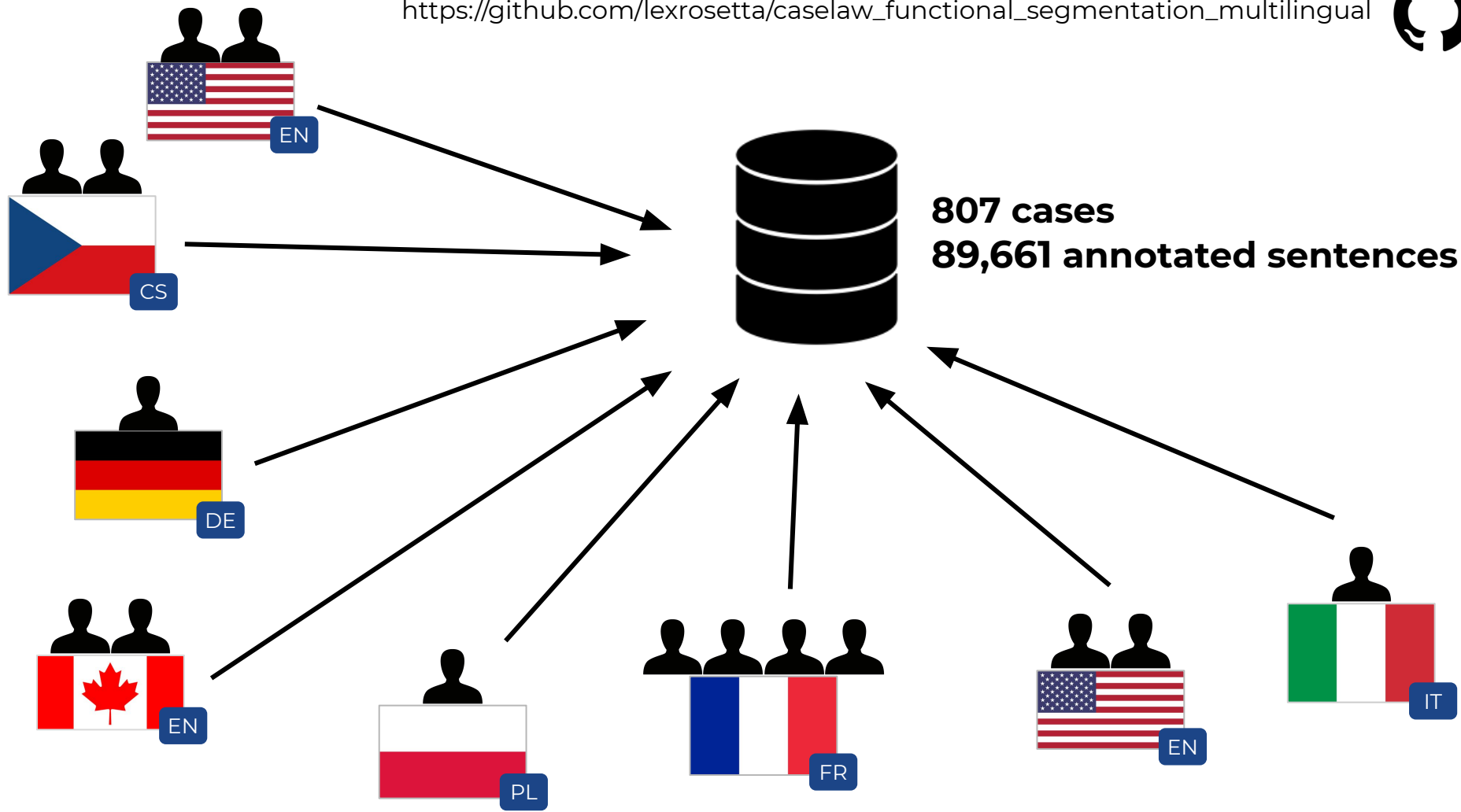
- Robust, generalizable models
- Rapid adaptation to new contexts
- International collaboration of researchers
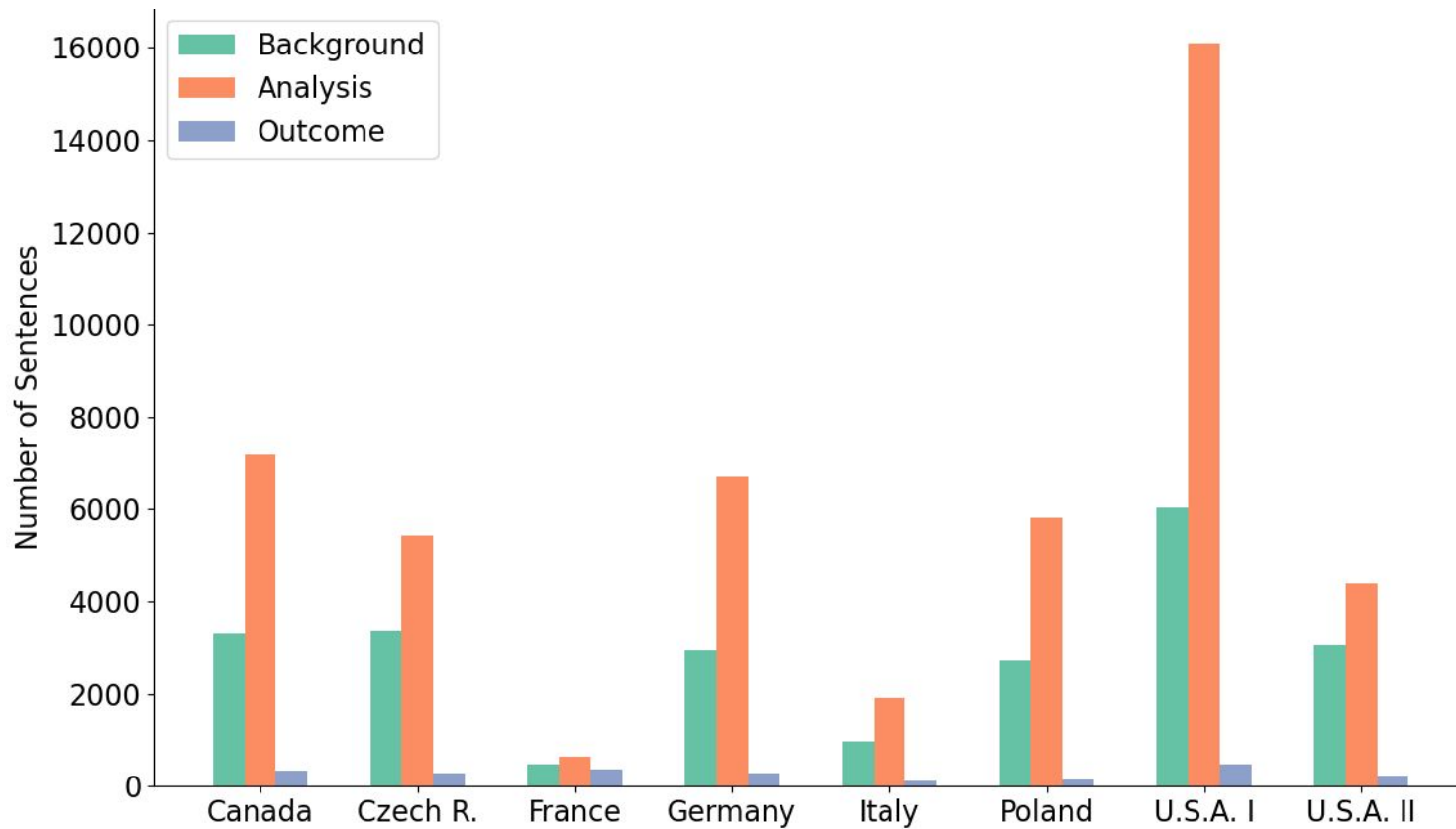
Dataset

Functional Segmentation of Adjudicatory Decisions

- **Out of Scope**
- **Heading**

- **Background**
- **Analysis**
- **Outcome**

807 cases
89,661 annotated sentences
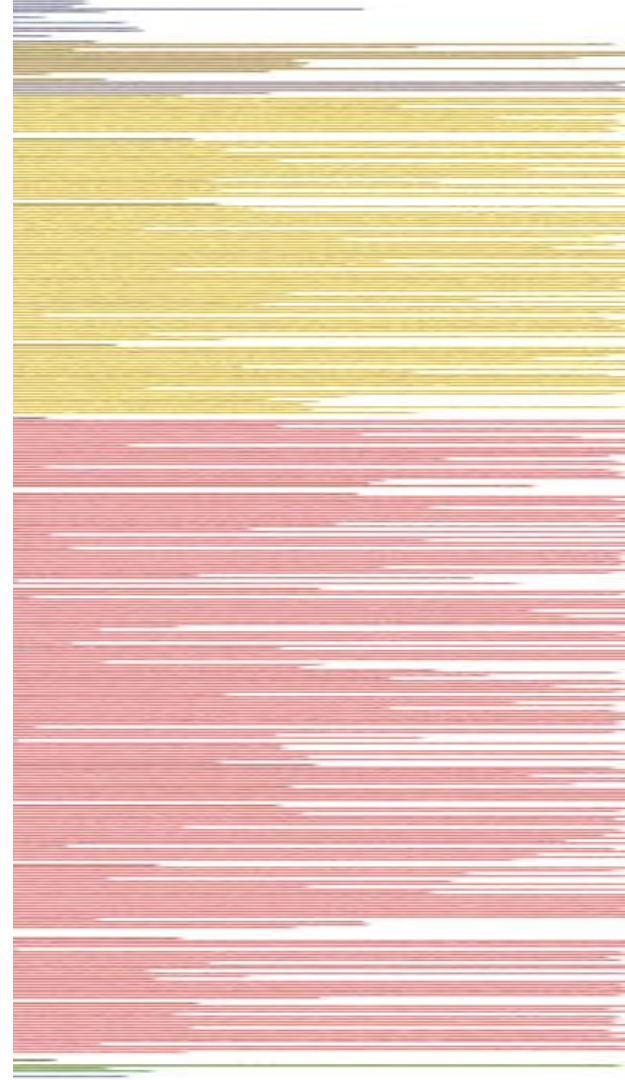
Data Set Statistics

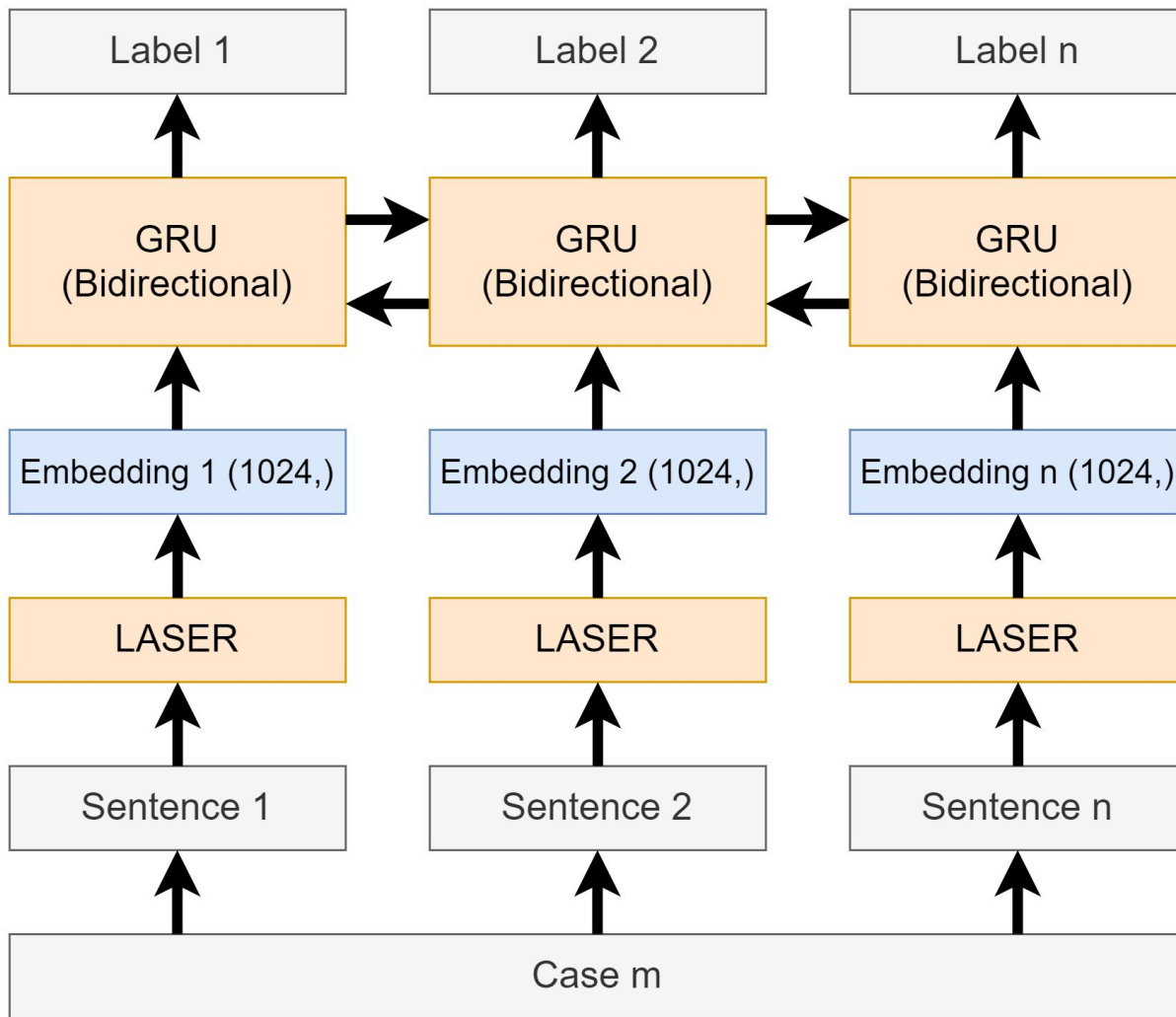# Models

Experiments

Experimental Design

- Three hypotheses
- Differ in what is used as training data - single or multiple datasets
- Compared against a baseline
- Evaluation: F1-score, 10 K-fold splits

H1 - Out-Context Experiment

Out-Context Experiment

Can a model trained on a single context generalize to other contexts?

H1 - Out-Context Experiment

Train

Test

H1 - Out-Context Experiment

Train

Test

H1 - Out-Context Experiment

Train

Test

Baseline: Stratified random prediction

Out-Context Experiment

Trained on:	Avg F1-score
Baseline	*.51*
Canada	.76
Czech Republic	.77
France	.68
Germany	.67
Italy	.67
Poland	.74
U.S.A. I	.71
U.S.A. II	.74

Results

-> Models able to transfer knowledge between contexts

H2 - Pooled Out-Context Experiment

Pooled Out-Context Experiment

Does adding additional contexts to the training pool increase robustness?

H1 - Out-Context Experiment

Train

Test

H2 - Pooled Out-Context Experiment

Train

Test

H2 - Pooled Out-Context Experiment

Train

Test

Baseline: Best performing Out-Context model per context. *Competitive*

Pooled Out-Context Experiment

Target context:	*Baseline F1*	Model F1
Canada	**.83**	**.83**
Czech Republic	.83	**.87**
France	.64	**.66**
Germany	.85	**.90**
Italy	.83	**.85**
Poland	.83	**.88**
U.S.A. I	**.87**	.81
U.S.A. II	.89	**.92**

Results

-> Training on pool of datasets leads to more robust models. Better adaptation to unseen contexts.

H3 - Pooled with In-Context Experiment

Pooled With In-Context Experiment

Does combining in-context data with out-context data improve performance?

H2 - Pooled Out-Context Experiment

Train

Test

H3 - Pooled With In-Context Experiment

Train

Test

H3 - Pooled With In-Context Experiment

Train

Test

Baseline: In-Context models. *Competitive*

Pooled With In-Context Experiment

| Target context:	*Baseline F1*	Model F1
Canada	.82	**.88**
Czech Republic	.91	**.94**
France	**.86**	.82
Germany	.88	**.96**
Italy	**.95**	.94
Poland	.93	**.94**
U.S.A. I	.91	**.92**
U.S.A. II	.94	**.96**

Results

-> Training on In-Context data pooled with Out-Context data does improve performance.
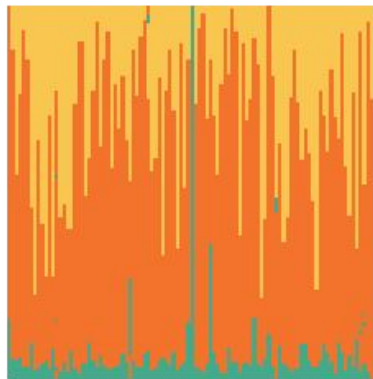
Discussion

The experiments for H1 and H2 show that:

Multilingual embeddings appear to capture the semantics of a sentence, rather than superficial linguistic features.
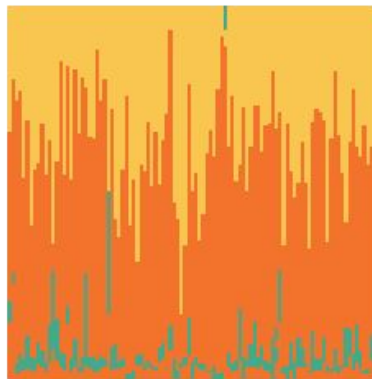
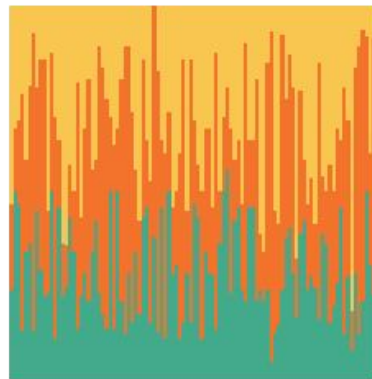# Relationships Among Contexts
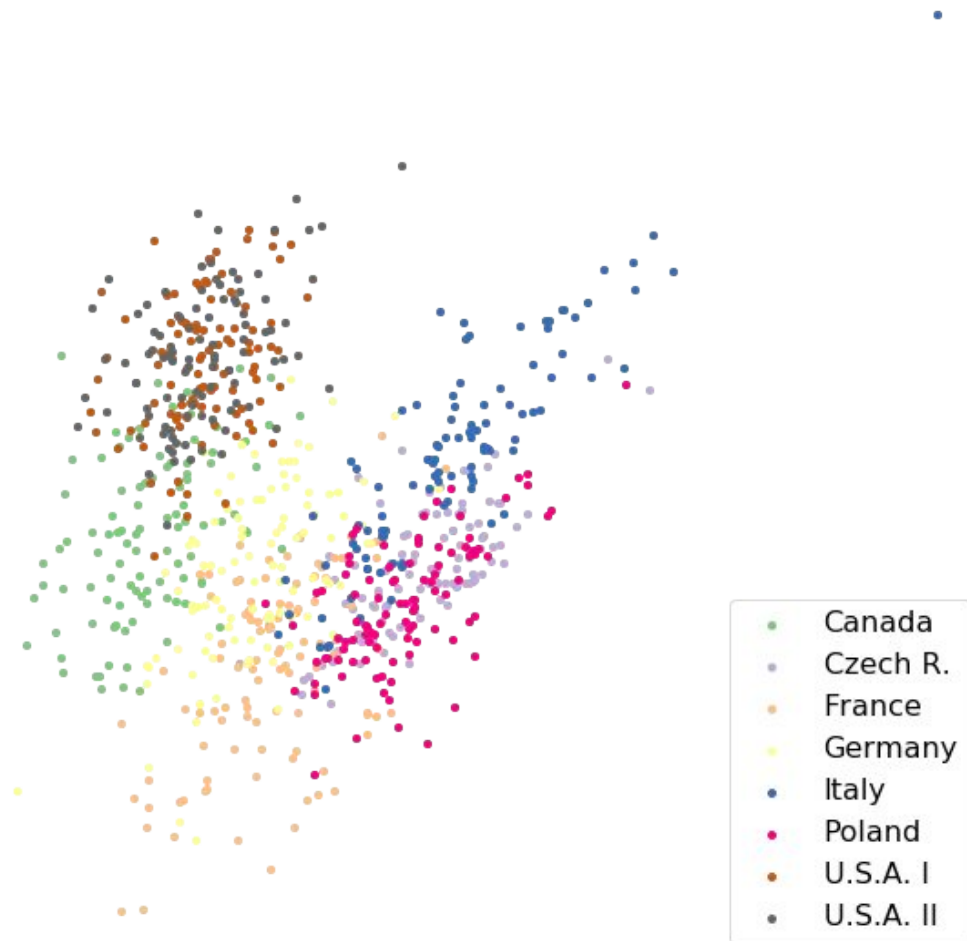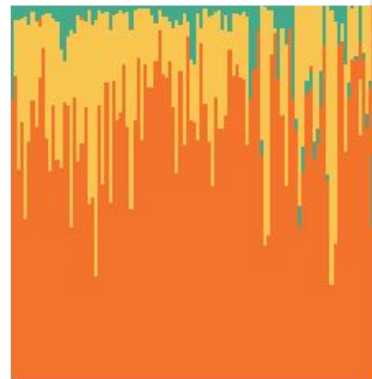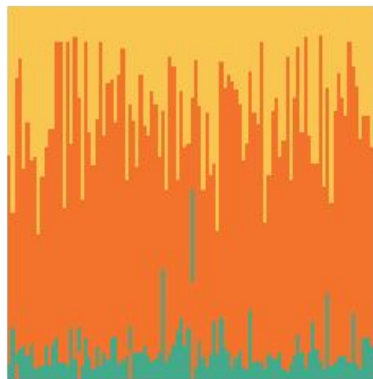
Sentence Types Distributions
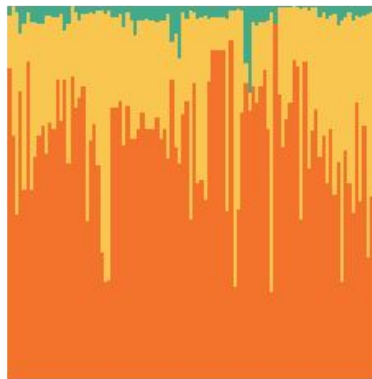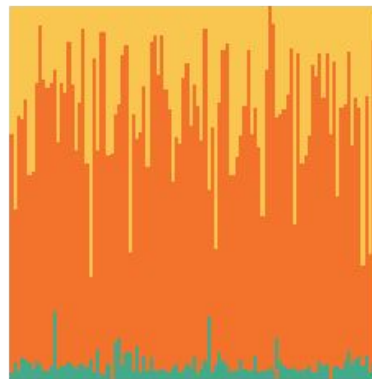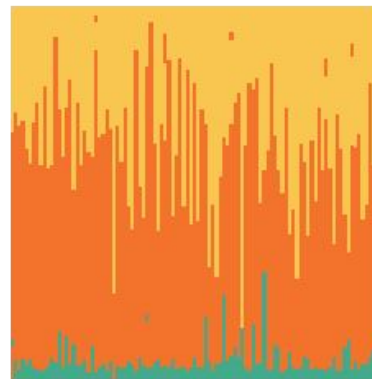
Conclusions

Conclusions

- Models are able to generalize across contexts
- Pooling increases robustness/performance
- And it helps even when in-domain data is available

jsavelka Merge branch 'master' of https://github.com/lexrosetta/caselaw_functi... ...   2d510d7 yesterday   ⟳ 5 commits

📁 annotation_guidelines	Data, data preprocessing, and dataset statistics code	4 days ago
📁 data	Data, data preprocessing, and dataset statistics code	4 days ago
📄 .gitignore	Add notebook with code for experiments, description in readme.	2 days ago
📄 ICAIL2021_experiments.ipynb	Add notebook with code for experiments, description in readme.	2 days ago
📄 README.md	Merge branch 'master' of https://github.com/lexrosetta/caselaw_functi...	yesterday
📄 dataset_ICAIL2021.py	Data, data preprocessing, and dataset statistics code	4 days ago
📄 dataset_clean.py	Data, data preprocessing, and dataset statistics code	4 days ago
📄 dataset_statistics.ipynb	Data, data preprocessing, and dataset statistics code	4 days ago
📄 ia_agreement.ipynb	Data, data preprocessing, and dataset statistics code	4 days ago
📄 ia_agreement.py	Data, data preprocessing, and dataset statistics code	4 days ago
📄 utils.py	Data, data preprocessing, and dataset statistics code	4 days ago

▤ README.md

Lex Rosetta: Transfer of Predictive Models Across Languages, Jurisdictions, and Legal Domains

This is an accompanying repository to the ICAIL 2021 paper entitled "Lex Rosetta: Transfer of Predictive Models Across Languages, Jurisdictions, and Legal Domains". All the data and the code used in the experiments reported in the paper are to be found here.

Data

The data set consists of 807 adjudicatory decisions from 7 different countries (6 languages) annotated in terms of the following type system:

- **Out of Scope** - Parts outside of the main document body (e.g., metadata, editorial content, dissents, end notes, appendices).
- **Heading** - Typically an incomplete sentence or marker starting a section (e.g., "Discussion," "Analysis," "II.").
- **Background** - The part where the court describes procedural history, relevant facts, or the parties' claims.
- **Analysis** - The section containing reasoning of the court, issues, and application of law to the facts of the case.
- **Introductory Summary** - A brief summary of the case at the beginning of the decision.
- **Outcome** - A few sentences stating how the case was decided (i,e, the overall outcome of the case).

The country specific subsets:

- **Canada** - Random selection of cases retrieved from www.canlii.org from multiple provinces. The selection is not limited to any specific topic or court.
- **Czech Republic** - A random selection of cases from Constitutional Court (30), Supreme Court (40), and Supreme Administrative Court (30). Temporal distribution was taken into account.
- **France** - A selection of cases decided by Cour de cassation between 2011 and 2019. A stratified sampling based

---

ICAIL2021 experiments.ipynb ☆

File Edit View Insert Runtime Tools Help   Last edited on June 20

+ Code + Text   Connect ▾   ✏ Editing

Lex Rosetta: Transfer of Predictive Models Across Languages, Jurisdictions, and Legal Domains

This notebook contains the code for the paper:

Jaromir Savelka, Hannes Westermann, Karim Benyekhlef, Charlotte S. Alexander, Jayla C. Grant, David Restrepo Amariles, Rajaa El Hamdani, Sébastien Meeùs, Aurore Troussel, Michał Araszkiewicz, Kevin D. Ashley, Alexandra Ashley, Karl Branting, Mattia Falduti, Matthias Grabmair, Jakub Harašta,Tereza Novotná, Elizabeth Tippett, and Shiwanni Johnson. 2021. Lex Rosetta: Transfer of Predictive Models Across Languages, Jurisdictions, and Legal Domains. In Eighteenth International Conference for Artificial Intelligence and Law (ICAIL'21), June 21–25, 2021, São Paulo, Brazil. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3462757.34661491

The notebook contains the code necessary to load data from annotated multicontextual legal cases. Then, it embeds the individual sentences into a multilingual vector embedding, using Language-Agnostic SEntence Representations. Finally, it trains a gated recurrent unit network to take a new case and predict the label of each sentence. The code to run the experiments presented in the paper and create the visualizations used in the discussion section is included. Notably, that inincludes an evaluation fo how well the model performs when evaluated on a context different from the one it is trained on.

This notebook can be run either locally or in the cloud with Google Colab.

- On Google Colab **(easier)**. Follow this link: https://colab.research.google.com/drive/1zSsKIPZXp3JdlU5E5GVZox-FiapYm4oD?usp=sharing The notebook will download the data from github. Note that RAM restrictions on colab means that the experiments for H2 and H3 are likely to crash.
- Locally: It is recommended to set up a new python environment and run the first cell to install the necessary requirements. Instructions to enable CUDA training, which significantly speeds up execution, is available at https://www.tensorflow.org/install/gpu and https://pytorch.org/get-started/locally/. Certain cells (such as loading the data from github) are only required on Google Colab and can be skipped.

Created by Hannes Westermann, Cyberjustice Laboratory

▾ Installation

Only required first time, or when running online. Can take a while, especially when running locally, since it downloads a number of large packages.

```
[ ]  ## Only required when running locally, these are already installed on Google Colab
     !pip install tensorflow
     !pip install pandas
     !pip install matplotlib
     !pip install scikit-learn
```

```
[ ]  #Install laserembeddings and load the model. Required both on colab and locally.

     from IPython.display import import clear_output
     !pip install laserembeddings
     !python -m laserembeddings download-models
     clear_output()

     ## RESTART RUNTIME
```

Future Work

- Dataset extension
- Use data in other applications
- Transfer learning strategies
- Comparative analysis?

Thank you for your attention!