



Information Extraction & Entailment of Common Law & Civil Code

John Hudzina,
Kanika Madan,
Dhivya Chinnappa,
Jinane Harmouche,
Hiroko Bretz,
Andrew Vold,
& Frank Schilder

November 17, 2020



THOMSON REUTERS®

Overview

COLIEE 2020 Tasks

- Task 1: Legal Case Retrieval
- Task 2: Legal Case Entailment
- Task 3: Civil Code Retrieval
- Task 4: Civil Code Entailment

Case Law Retrieval & Entailment

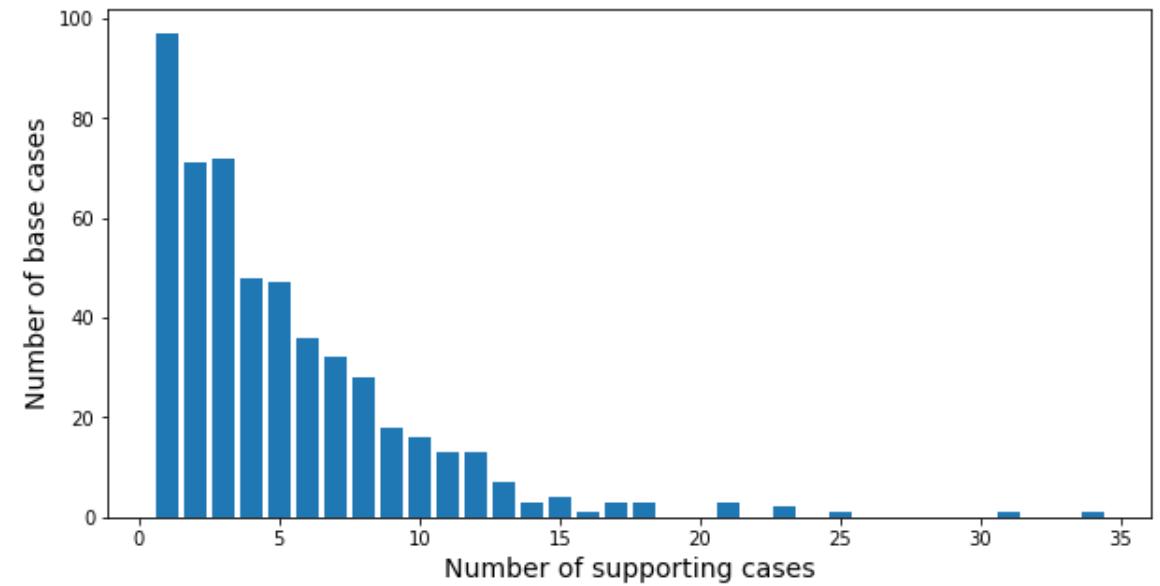
Task 1: Legal Case Retrieval: Task Description

For a given base case, identify supporting cases from a pool of candidate cases.

- Each base case has 200 candidate cases
- Most bases cases have less than 10 supporting cases out of the 200 candidate cases

Methodology:

- Rank candidate cases
- Build 3 classifiers for candidate cases based on their ranking



Task 1: Legal Case Retrieval: Task Approach

Ranking task:

Rank candidate cases based on similarity measures

- Jaccard similarity
- Bag of words similarity
- GloVe embedding similarity
- Word2Vec embedding similarity
- Combination of similarity measures (Jacc. + w2v, Jacc. + BOW, etc.)

Classification task:

Build 3 classifiers for candidate cases with

- rank < 50,
- rank > 50 and <150
- rank > 150
- Classifier: xgboost

Ranking task Inference:

- Jaccard similarity combinations consistently ranked supporting cases at top
- Chosen similarity measures based on their ranking performance on supporting cases
 - Jaccard
 - Jaccard + Word2Vec
 - Jaccard + GloVe

Classification task Inference:

Results on validation set:

- < 50 ~0.5 F1
- > 50 and <150 ~0.33 F1
- > 150 ~0.15 F1

Results on test set: 0.38

Overfitting

Imbalanced data distribution

Task 2: Legal Case Entailment: Task Description

- Given a decision Q of a new case and a relevant case R , a specific paragraph that entails the decision Q needs to be identified.
- The case R is a relevant case to Q , many paragraphs in R can be relevant to Q regardless of entailment.
- A specific entailment method is required which compares the meaning of each paragraph in R and Q in this task.

Task 2: Legal Case Entailment: Task Approach

- **Approach:** A ranking algo uses an aggregate function of individual rankings from a set of sentence vectorizers.
Ranking results for the rank of the entailing paragraphs in the training dataset:

Number of Paragraphs	Frequency	Rank at 80 percentile
$N \leq 30$	57%	3
$30 < N \leq 50$	26%	5
$N > 50$	17%	8

- **Features:** Text Similarity features between the decision paragraph and each of the top K paragraphs
- **Classifier:** Random Forest
- Classification cross-validation results corresponding to the ranking from the table above:

Number of Paragraphs	Frequency	Precision	Recall	F1-score
$N \leq 30$	57%	0.71	0.66	0.68
$30 < N \leq 50$	26%	0.66	0.68	0.67
$N > 50$	17%	0.37	0.39	0.38

- Final test score was lower on the unseen dataset (F1: 0.4107) – Possibly from sensitivity of parameters from the ranker

Civil Code Retrieval & Entailment

Task 3 Description

Task: Retrieve most relevant civil code articles to answer given exam question

Strategy:

- Language used: Japanese
- TFIDF/cosine similarity (Kim et. Al. 2019 Statute law information retrieval and entailment)
- Custom idiom-level tokenization (e.g. 第三者 ("third party") as opposed to 第三 ("third") and 者 ("party"))
- Additional training data: Japanese Wikibook articles of civil code

Results:

Run	lang	F2	Prec.	Recall	MAP	R_5	R_{10}	R_{30}
LLNTU	E/J	0.6587	0.6875	0.6622	0.7604	0.8071	0.8571	0.9214
JNLP.tbe	E/J	0.5532	0.5766	0.5670	0.6618	0.6857	0.7143	0.7786
cyber1	E/J	0.5290	0.5058	0.5536	0.5540	0.5500	0.6929	0.8000
HUKB-1	J	0.5160	0.4196	0.5908	0.5687	0.6714	0.7214	0.8143
CUBERT1	E/J	0.5139	0.5402	0.5193	0.5848	0.6429	0.6857	0.7071
TRC3_1	J	0.5011	0.4561	0.5357	0.5978	0.6929	0.7714	0.8429
OVGU_bm25	E/J	0.4768	0.4003	0.5342	0.5095	0.5929	0.6143	0.7214
TAXI_R3	E/J	0.4546	0.4393	0.5089	0.5057	0.5714	0.6143	0.6786
GK_NLP	E/J	0.4273	0.2857	0.4985	0.4982	0.5571	0.6357	0.7214
UA.tfidf	E/J	0.3913	0.4286	0.3869	0.4777	0.5429	0.6071	0.6714
HONto_hybrid	E/J	0.2822	0.2545	0.2991	0.0142	0.0071	0.0071	0.0500

Idiom-Level Tokenizer

Japanese Language Systems:

- **Hiragana** (Japanese original. Phonetical. Used for particles, adverbs, postpositions)
- **Kanji** (Chinese Characters, each characters carry meanings)
- **Kantakana** (used mostly for foreign/imported words)

Example:

私権は、公共の福祉に適合しなければならない。

Private rights must be congruent with the public welfare.

Tokens:

- 私権 (private rights)
- 公共 (public)
- 福祉 (welfare)
- 適合 (congruent with)

Task 4 Description

Kim et al. (2019) model of the entailment task

Given a set of relevant articles as a premise $P = [P_1, P_2, \dots, P_n]$ and a hypotheses bar exam question H , determine if P entails H .

- Each article in P contains conditions, exceptions, and conclusions
- H contains facts and a conclusion
- The system must apply the facts to the articles conditions and determine if H can lead to the correct conclusion

Sent.	Text
P_1	A mandate shall terminate when the mandator or mandatory dies.
H	The mandate terminated upon the mandator's death.

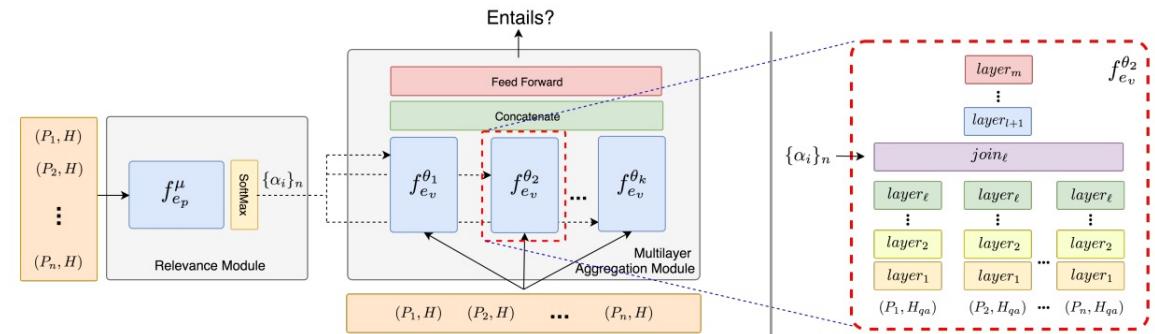
Task 4 Approach

Observations

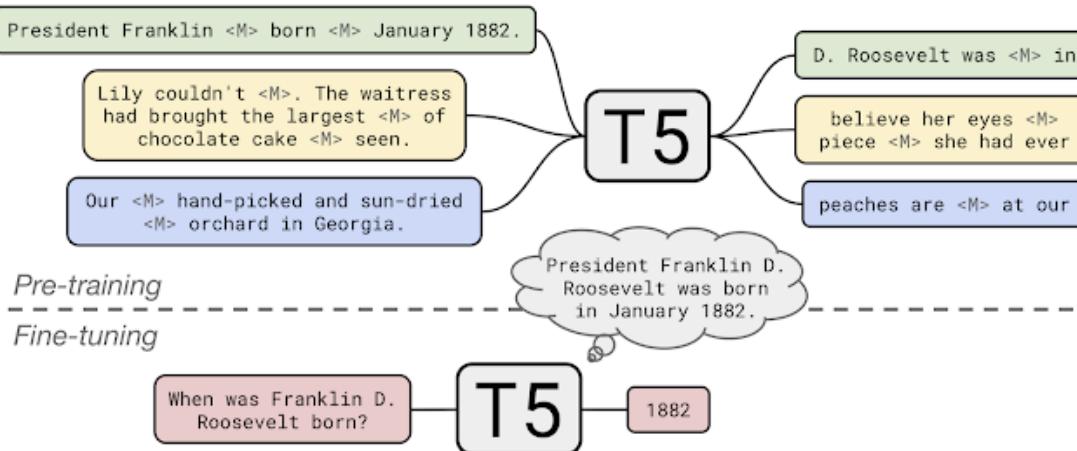
Approach must

- Match relevant phases in the P & H
- Deal with multiple conditions, exceptions and conclusions in P
- Leverage transfer learning if machine learning is used
- Handle legal terminology

Multee: Multi-Hop Question & Answering



T5: Text to Text Transfer Transformer



Naïve Feature-Based Approach

Baseline

- TF-IDF Vectors from Task 3
- Gradient Boosting Classifier

Task 4 - Training Datasets

No Pre-Training

Muttee Dataset

- Pre-Trained Embeddings
 - [GLoVe Embeddings](#)
 - Common Crawl (860B,300D)
- Relevance Module
 - [Multi-Genre NLI Corpus](#)
 - [Stanford NLI Corpus](#)
- Multi-Level Aggregator
 - [Modified OpenBookQA Dataset](#)
 - Converted multiple choice to yes/no
 - COLIEE
 - Train: H18-H29
 - Val: H30

T5 Tasks & Datasets

- Pre-Trained Embeddings
 - T5-base (<https://huggingface.co/t5-base>)
 - C4 (<https://huggingface.co/datasets/c4>)
- Fine Tuning
 - Denoise Task (predict missing spans)
 - Japanese Civil Code Test
 - Japanese Civil Code Titles
 - [Wikibook articles](#) on the Japanese Civil Code
 - Entailment Task
 - COLIEE
 - Train: H18 to H29
 - Val: H30

Task 4 Results

Run	Correct	Val. Accuracy	Accuracy
Baseline	59/122	-	0.5268
JNLP.BERTLaw	81	-	0.7232
TRC3mt	70	0.5990	0.6250
TRC3t5	70	0.6714	0.6250
UA_attention	70	-	0.6250
UA_roberta	70	-	0.6250
TRC3A	56	0.6400	0.5000

Discussion

- Multee
 - Weighing article relevance provide an improvement over last years attention-based model (from 0.5612 to 0.6250)
- T5
 - Multi-task fine tuning didn't preform as well compared to winning pretraining approach
 - Tasks only focused on unsupervised denoising and not explainability tasks
 - Future Work
 - Additional tasks related to statutory interpretation
 - Multi-task pretraining