ALBERTA MACHINE INTELLIGENCE INSTITUTE

UNIVERSITY OF ALBERTA

Entailment Detection Techniques Applied for Argument Mining

COLIEE 2018

Juliano Rabelo

Postdoctoral Fellow

rabelo@ualberta.ca

Outline

- The COLIEE case law entailment task
- Our approach
- Results
- Argument mining
- Application of our case law model to argument mining
- Other possible approaches

COLIEE

- Until 2017 had tasks on Statute Law
  - Given a legal bar exam question, retrieve related Civil Code articles
  - Augment above task by answering if the articles entail or not the given question
- In 2018, two new tasks focusing on Case Law
  - Identify which cases should be noticed in respect to a given base case
  - Identify which paragraph(s) of a related case entail a base case decision
  - Data drawn from an existing collection of Canadian case laws provided by vLex Canada

Entailment Task Description

Given:

a base case b, represented by its decision d, summary s and facts F , and

a related case r, represented by its paragraphs $P = \{p_1, p_2, \ldots, p_n\}$, such that $noticed(b, r)$ is true

The task is to find the set

$$E = \{p_1, p_2, \ldots, p_m \mid p_i \in P \land entails(p_i, d)\}$$

where $entails(p_i, d)$ denotes a relationship which is true when p_i entails d

Data Overview

- 181 cases, each with a noticed case
- XML file describing the dataset
- Each base case consists of its summary, a fact file and its decision
- Each candidate is a list of paragraphs of the original case
- 8,794 candidates, with 239 true positives (2.71%)
- 1.3 entailment paragraphs per case (stddev 0.7)

Our Method

- Ideally:
  - In depth domain + common sense knowledge
  - Reasoning capability
- More feasible:
  - Extract features which correlate with the entailment relationship
- Problem:
  - Severe data scarcity and class imbalance

Preprocessing

- Regular techniques (stop words, accents, digits removal, etc)
- Language detection to remove French paragraphs
  - 6 expected answers “lost”
  - Reasonable limitation

Pairwise Paragraph Comparison

- Leverages the idea that the main concept of a case law is encoded in specific paragraphs
- The comparison step produces a matrix
- We create features based on the similarities:
  - Histogram of similarities
  - Standard deviation
  - Paragraph similarities considering only the “legal terms”
- Data fed to a Random Forest classifier

Embeddings Augmented Pairwise Paragraph Comparison

- Similar concepts may be expressed with different words
- Two approaches:
  - Pre-trained, general purpose embedding
  - Embedding trained on public available Canadian Supreme Court Case Laws
- Document similarity calculated by the Word Mover's Distance
- Similar feature matrices generated and fed to a classifier
- Computationally expensive

Similarity Based

- Similarity calculated between the candidate paragraph and the base case:
  - Decision
  - Summary
  - Paragraphs (histogram)
- Tested with two classifiers: Random Forest and Gradient Boosting
- Post processing
  - Adding “context” to the method by considering the priors
  - Established a safe range for each base case result

The Imbalance Problem

- Less than 3% of the samples are true entailment cases
- To cope with that:
  - Undersampling on the negative class
  - Oversampling on the positive class
  - Cost sensitive classifier

Task 2 Results

id	Precision	Recall	F-measure
Smartlaw	0.046512	0.150943	0.071111
UA	0.238095	0.283019	0.258621
UA-100estimators	0.190476	0.226415	0.206897
UA-500estimators	0.238095	0.283019	0.258621
UBIRLED-1	0.048352	0.830189	0.091381
UBIRLED-2	0.049495	0.924528	0.093960
UBIRLED-3	0.046667	0.792453	0.088143
UNCC0	0.032967	0.056604	0.041667

Baseline: precision : 0.0291, recall: 1.0, f-measure: 0.0428

Argument Mining

Argument Mining

- Task of identifying argumentative structures in natural language texts
- Usual model
  - Premises
  - Conclusions (claims)
- Pipeline
  - Segmentation
  - Identification of argumentative sentences
  - Classification (premises/claims)

Usual Approaches

- Sentence segmentation
- Hand crafted rules
  - Specific lexicon
  - Specific text structure
  - Part of speech
  - Parse trees
- Machine learning
  - SVM, Naïve Bayes, Maximum Entropy…
  - Careful feature engineering
  - Deep learning seldom used

Similarity Based Models on AM

- The overall framework could be applied:
  - Tagged corpus
  - Identification of argumentative sentences
  - Pairwise similarity calculated between sentences
    - Word embeddings
  - Classifier learns the relationship
- Reasonable results might be achieved

Other Machine Learning Approaches

- Challenge: not many tagged corpora available
  - AraucariaDB: 641 documents, 3,798 sentences
  - ECHR: 47 documents, 2,571 sentences
    - 3 annotators for more than 1 year
    - 75% agreement
- LUIMA
  - Law-specific semantic extraction toolbox

Deep Learning Approaches

- DL-based NLP traditional approaches
  - Pre-trained word embeddings as a representation layer
  - Subsequent layers trained on a specific domain
- Useful, but limited
  - Data x specific hand crafted mechanisms

Transfer Learning

- "NLP's ImageNet Moment"
  - ~1.3M images
  - ULMFit: leverages pre-training based on language modeling
- LM has limitations
  - Anaphora/coreference resolution
  - Common sense knowledge

ALBERTA MACHINE INTELLIGENCE INSTITUTE

UNIVERSITY OF ALBERTA

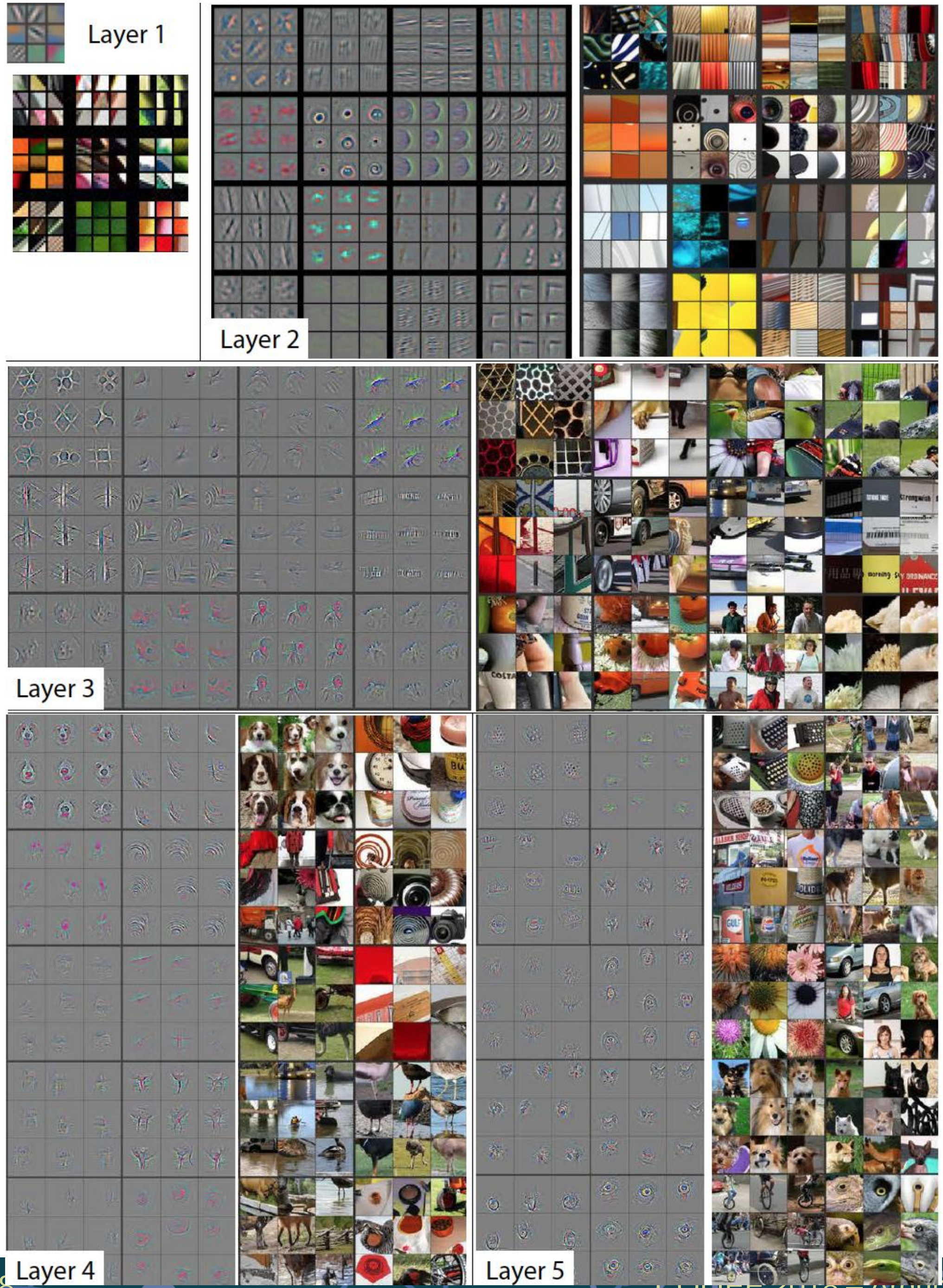
Entailment Detection Techniques Applied for Argument Mining

COLIEE 2018

Juliano Rabelo

Postdoctoral Fellow

rabelo@ualberta.ca



Visualizing and Understanding Convolutional Networks
Matthew D. Zeiler and Rob Fergus