Party-based Sentiment Analysis Pipeline for the Legal Domain

Sahan Jayasinghe, Lakith Rambukkanage, Ashan Silva, Nisansa de Silva, Amal Shehan Perera

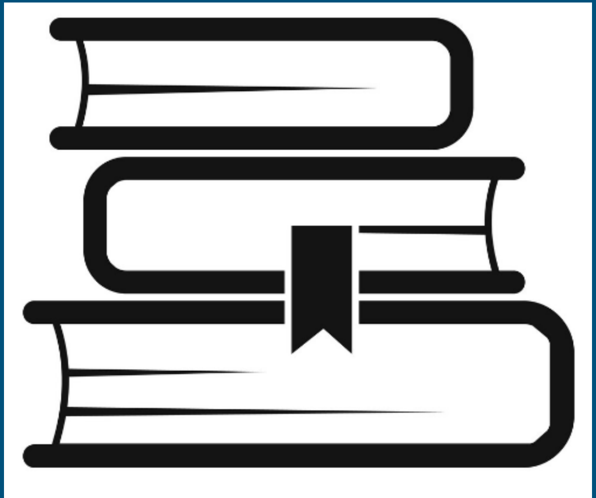
Table of Contents

Introduction

1. NLP
2. NLP in the Legal domain
3. Important notions in the Legal domain
4. Problem Statement

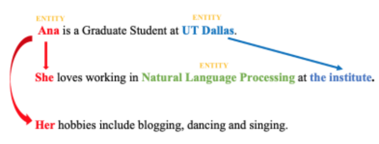
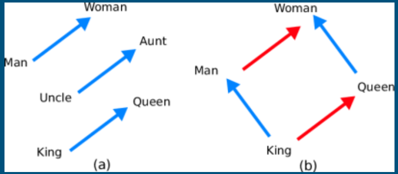
Natural Language Processing (NLP)

- The amount of text data is growing rapidly.
- So is the need for text based analysis.
- Analysis of unstructured data is non trivial
- Therefore the requirement for NLP arises



Natural Language Processing (NLP)

Techniques used in this research

<table>
<tr><th>Named Entity Recognition</th><th>Coreference Resolution</th><th>Sentiment Analysis</th><th>Embeddings</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
</table>

NLP in the legal domain

Legal domain is an area with growing amounts of textual data.

Also these texts are read and analyzed by professionals on a daily basis.

Interpreting legal texts is an analytically demanding task

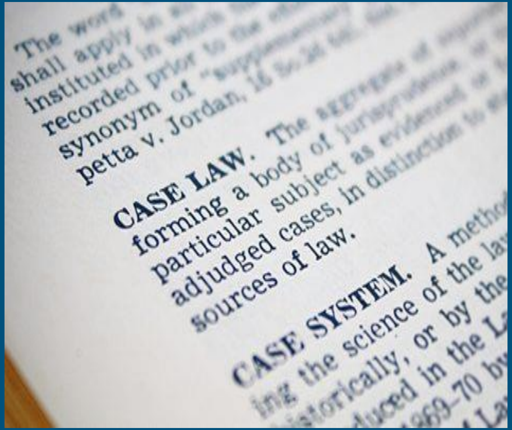
NLP in the legal domain : Research

- Party Based Sentiment Analysis
- Party Identification
- Domain Specific Embeddings
- Identifying Discourse Relations among sentences
- Critical Sentence Identification
- Winning Party Prediction

Important notions in the Legal domain

<u>Case Law</u>

- Past court cases are documented for future use.
- They are used to support ongoing court cases.
- Accounts to a large amount of text data in the legal domain.
- Legal professionals use them on a daily basis.
- Serves as good training data for applying NLP



Important notions in the Legal domain

Legal Party

Two main parties :

1. Petitioner
2. Defendant

Extracting these parties is not the same as Name Entity Recognition (NER)

- Other entities are present in the text : Requires Party Extraction
- Use of different names : Requires coreference resolution

Problem Statement

- Party Identification and Party Based Sentiment Analysis problems have been approached independently.
- Still there exists a lot of manual work that need to be done to use them sequentially.
    a. Manual Party labeling for PBSA input.
    b. Non uniform representation of parties.
- Therefore the need for a unified pipeline exists.

Related work

1. Party Identification in the Legal Domain
2. Party Based Sentiment Analysis (PBSA)

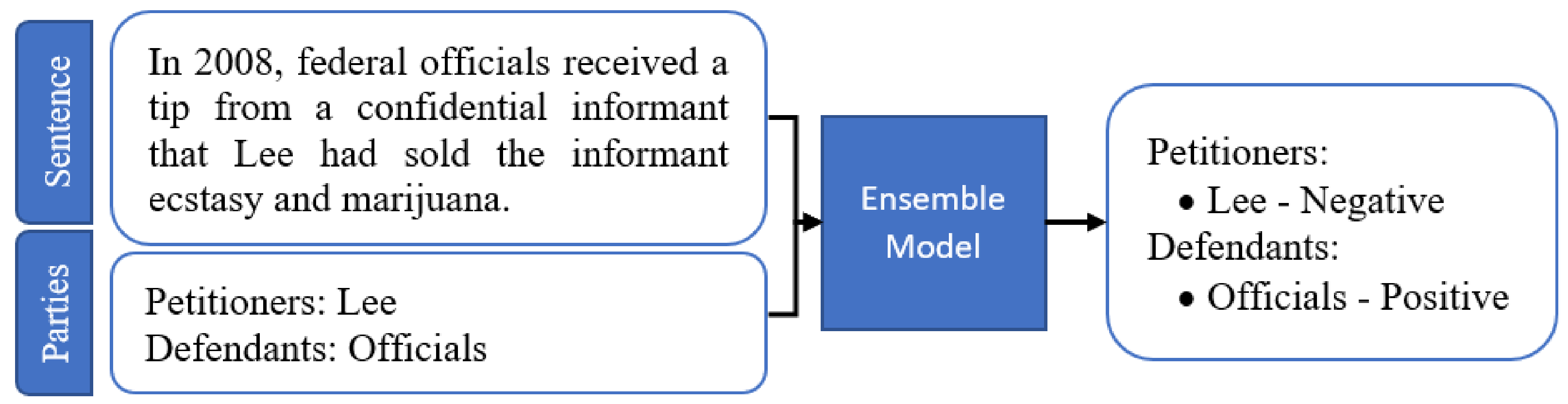
Party Identification in the legal domain

Identifying legal party members from legal opinion texts using natural language processing [1]

[1] C. Samarawickrama, M. de Almeida et al., “Identifying legal party members from legal opinion texts using natural language processing,” EasyChair, Tech. Rep., 2021.

Party-based sentiment analysis

SigmaLaw PBSA - A deep learning model for Aspect based sentiment analysis for the legal domain [2]

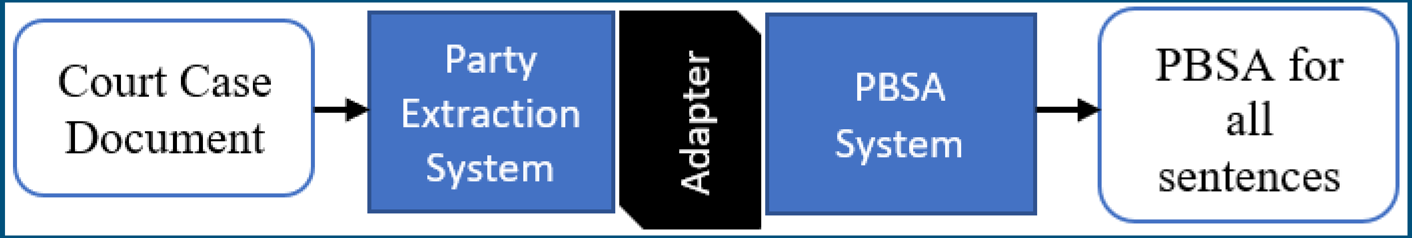


[2] I. Rajapaksha, C. R. Mudalige et al., "SigmaLaw PBSA-A Deep Learning Model for Aspect-Based Sentiment Analysis for the Legal Domain," in International Conference on Database and Expert Systems Applications. Springer, 2021, pp. 125–137.

Methodology

1. Pipeline Overview
2. *Baseline* model
3. *nuRef* model: updated co-reference
4. *nuRefGRU* model: updated co-ref + re-trained PE model
5. Pipeline Implementation 1
6. Drawbacks of Implementation 1
7. Pipeline Implementation 2

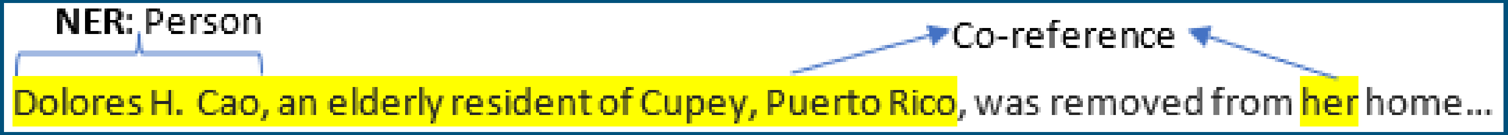
Pipeline Overview



- Party extraction system is based on Seq-to-Seq GRU model with 512 output nodes.
- Adapter generates the input for PBSA system using the output of party extraction system.

Baseline model

- GRU 512 model defined by Samarawickrama et al. for party identification.
- Model outputs probabilities for each input token.
- An algorithm for party extraction is defined using output probability sequence and stanford co-reference.
- There were unexpected results for certain scenarios with above setup.



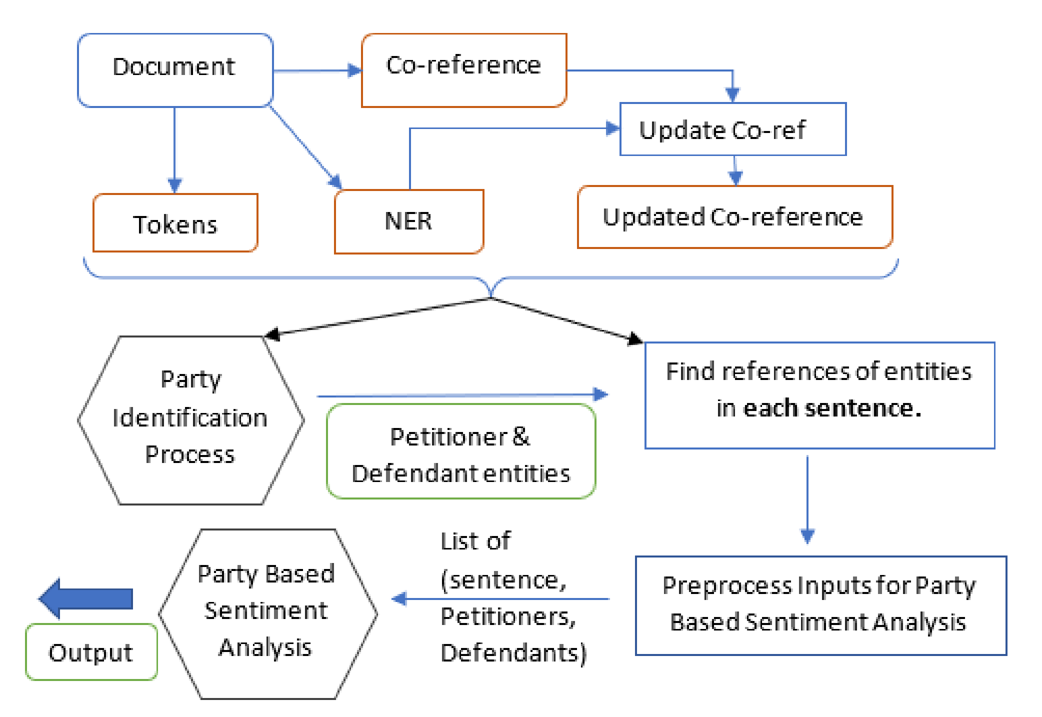
nuRef model: updated co-reference

- An improvement for the co-reference annotation of party members due to the issues occurred in Baseline approach.
- An algorithm is defined to update the Stanford co-reference output to remove the tailing description after an entity using NER.
- New setup of updated co-reference + Baseline model is evaluated.

nuRefGRU model: updated co-ref + re-trained PE model

- Evaluation results of *nuRef model* emphasized the requirement to re-train the GRU 512 model from data pre-processed using updated co-reference.
- At the pre-processing step, an additional dimension is added to word vectors which marks it as one of “Person”, “Organization” or “None” category.
- This process is referred to as masking by Samarawickrama et al.

Pipeline Implementation 1

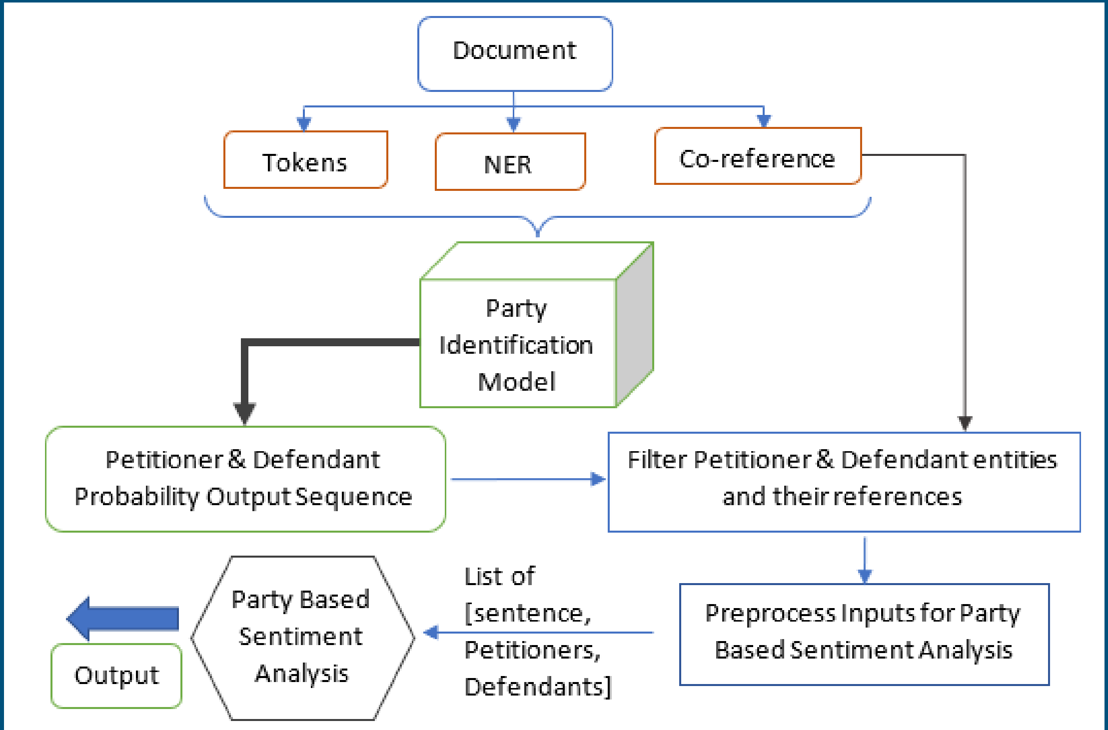


Drawbacks in Pipeline Implementation 1

- Depends on the accuracy of party extraction model and Stanford NER, co-reference at the party extraction process.
- Stanford NER is not capable of identifying domain related words as entities. (Ex: plaintiff, petitioner, defendant, respondent)

Plaintiffs are Ecuadorian crew members of a fishing boat . The United States Coast Guard saw their boat ... smuggling drugs. The Coast Guard stopped Plaintiffs' boat and boarded it. ... Plaintiffs then sued the United States for ...

Pipeline Implementation 2



Experiments and Results

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Petitioner	baseline	99.78	85.03	82.12	83.11
	nuRef	99.58	81.14	76.70	77.42
	nuRefGRU	**99.98**	**88.12**	**88.12**	**88.12**
Defendant	baseline	99.70	70.56	65.40	67.00
	nuRef	99.55	64.58	63.58	63.14
	nuRefGRU	**99.94**	**74.75**	**74.45**	**74.58**

Conclusion and Future work

- ➔ Party-based sentiments is useful to analyze the impact on each party in a court case.
- ➔ The PBSA Pipeline automates the work of analyzing the arguments brought forward in a court case and provides an overall outcome for each argument.
- ➔ We will be focusing on integrating the pipeline to more sophisticated tasks such as:
    - ◆ predicting the winning party of a court case based on arguments.

Thank you!

Q & A