Automated Identification of Argument Structure in Natural Text

B.Tech. Project.

Arkanath Pathak
Roll No. 12CS10007
4th Year B.Tech. Student
Dept. of Computer Science and Engineering
IIT Kharagpur

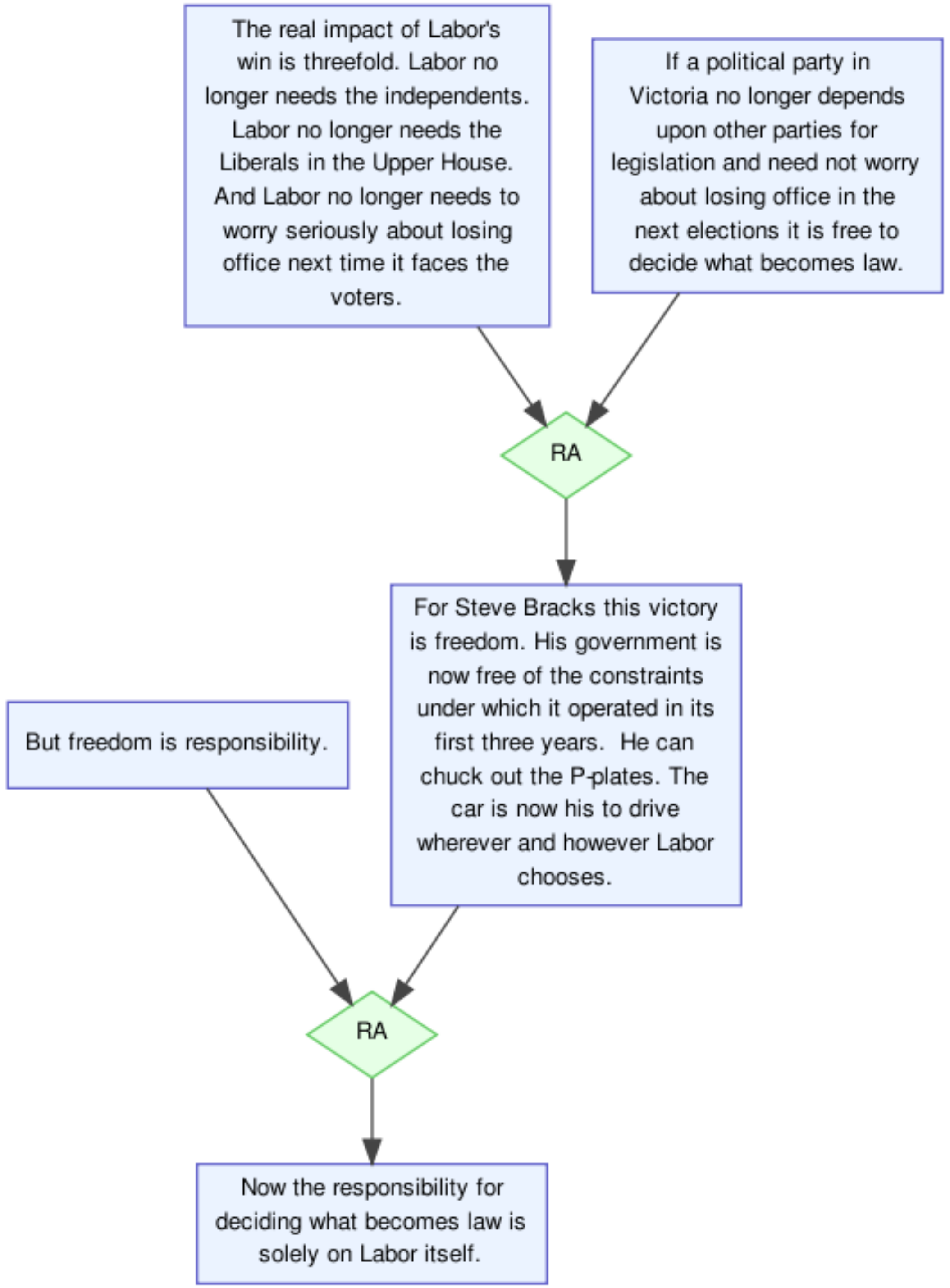
Project Advisors:
Dr. Pawan Goyal
Dr. Plaban Bhowmick

Introduction

- Argumentation mining is a relatively new challenge in corpus-based discourse analysis with the ultimate goal of identifying argumentative structures within a document.
- An argument can be considered as a set of **conclusions** that can be reached through logical reasoning based on **premises**.
- We assume the argument structure to be a tree to support the inherent hierarchy in the problem, i.e. an sub-argument can act as a premise for a conclusion.
- The various subtasks involved in the field are identification of the premises, conclusion, and argumentation scheme of each argument.

Example Argument

- This is an example argument map from the dataset that we are working on.
- The RA nodes in the diagram stand for the relation of inference.



The 3 steps widely encountered

1. Detection of argumentative propositions
   - **Moens et al., 2011 (AIL)** : Accuracy of 73% - 80%
2. Classification of propositions into premises and conclusions
   - **Moens et al., 2011 (AIL)** : Accuracy of 68% - 74%
3. Identifying the structure of argument by adding edges between the propositions
   - **Lawrence et al., 2014 (ACL)** : Accuracy of 33%

Our Goal: Jointly solve steps 2 & 3, construct the structure given the non-classified propositions.

Problem Formulation

1. Statement:
   Given a set of argumentative propositions (unstructured, english), find the structure of the argument by joining all the propositions to form a directed tree.
2. The main conclusion can be treated as the root of the tree.
3. There will be a set of premises for the main conclusion which can themselves be conclusions for a deeper level of premises.
   - *Hence, the hierarchy.*

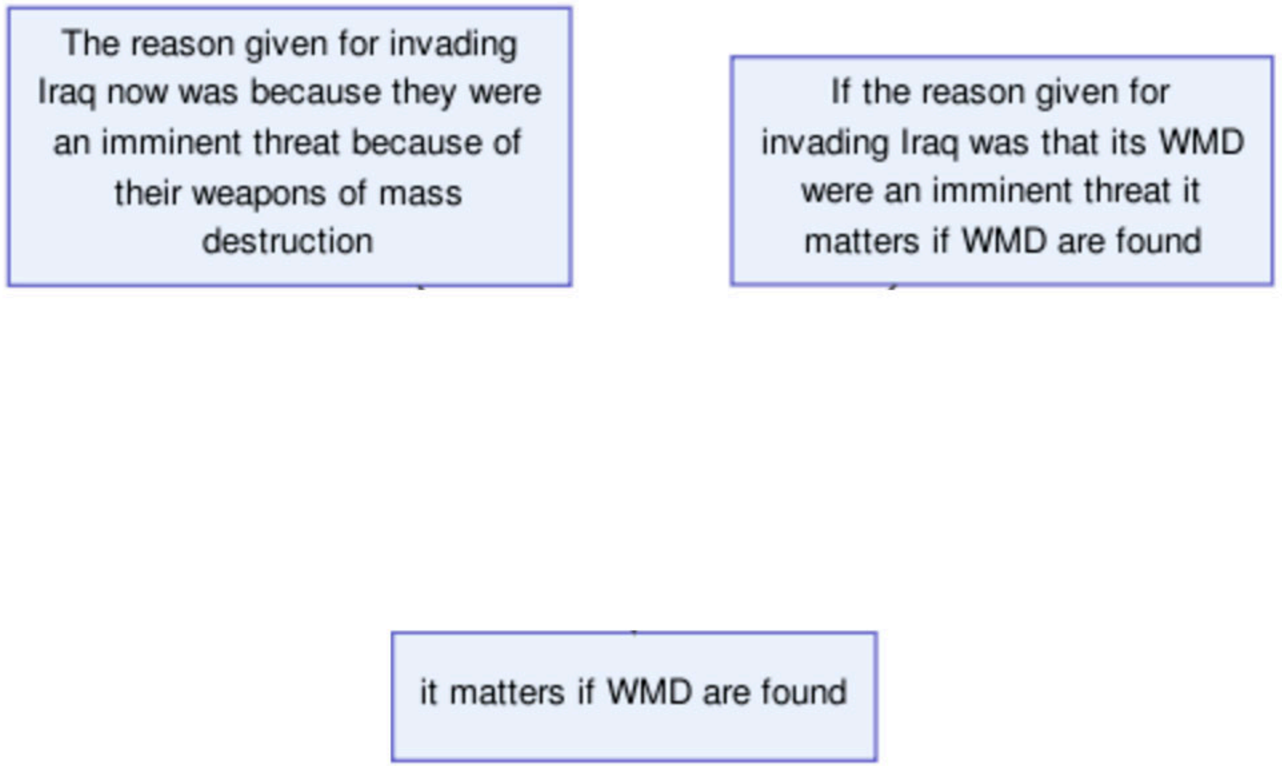
Brief review of literature

1. Cabrio et al., (ACL 2012)
   - work on online debates, automated identification by using textual entailment as the first stage of joining propositions and then using argumentation theory to reject invalid arguments.
2. Lawrence et al., (ACL 2014)
   - work on 19th Century Philosophical Texts, formed bidirectional edges between propositions based on Euclidean distance between topic measures by a generating a topic model.
3. Peldszus et al., (EMNLP 2015)
   - Similar problem formulation to ours. First perform the task of attachment classification, finding if there is an argumentative attachment or not. Then they assume there is a central claim to which each proposition would either support or attack.

Approach

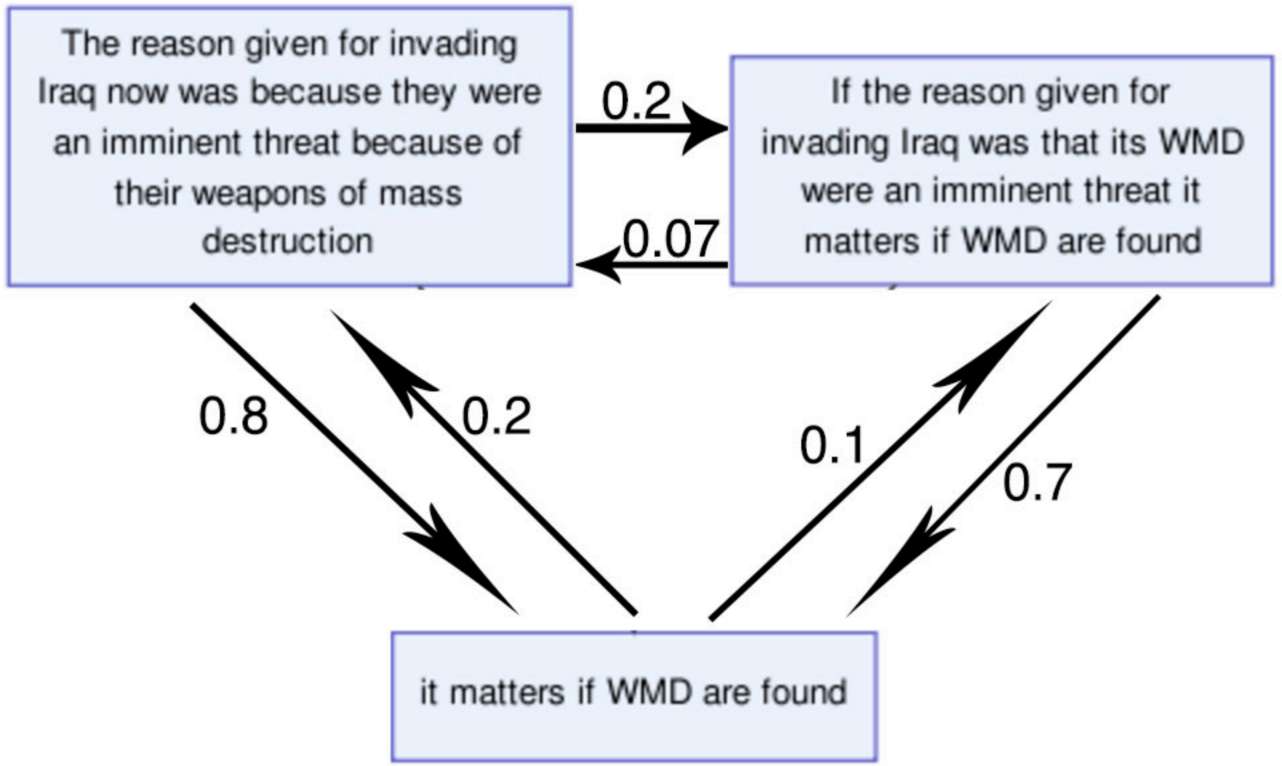
1. Find the edge weights for each possible ordered pair of nodes.
   - Edge weights account for the degree of support between an ordered pair of propositions
   - For example, the edge weight between a pair of nodes might just be a number between 0 to 1 representing the degree of support.
   - The most crucial step. We focus on this step throughout this presentation.
2. Construct the tree structure using the edge weights found in Step 1.
   - Possible approach of using some simple MST decoding algorithm.
3. Find out the accuracy for our approach using some scoring model.
   - Possible approach to consider some graph edit distance measure to find similarity to actual structure

Example (values not correct)



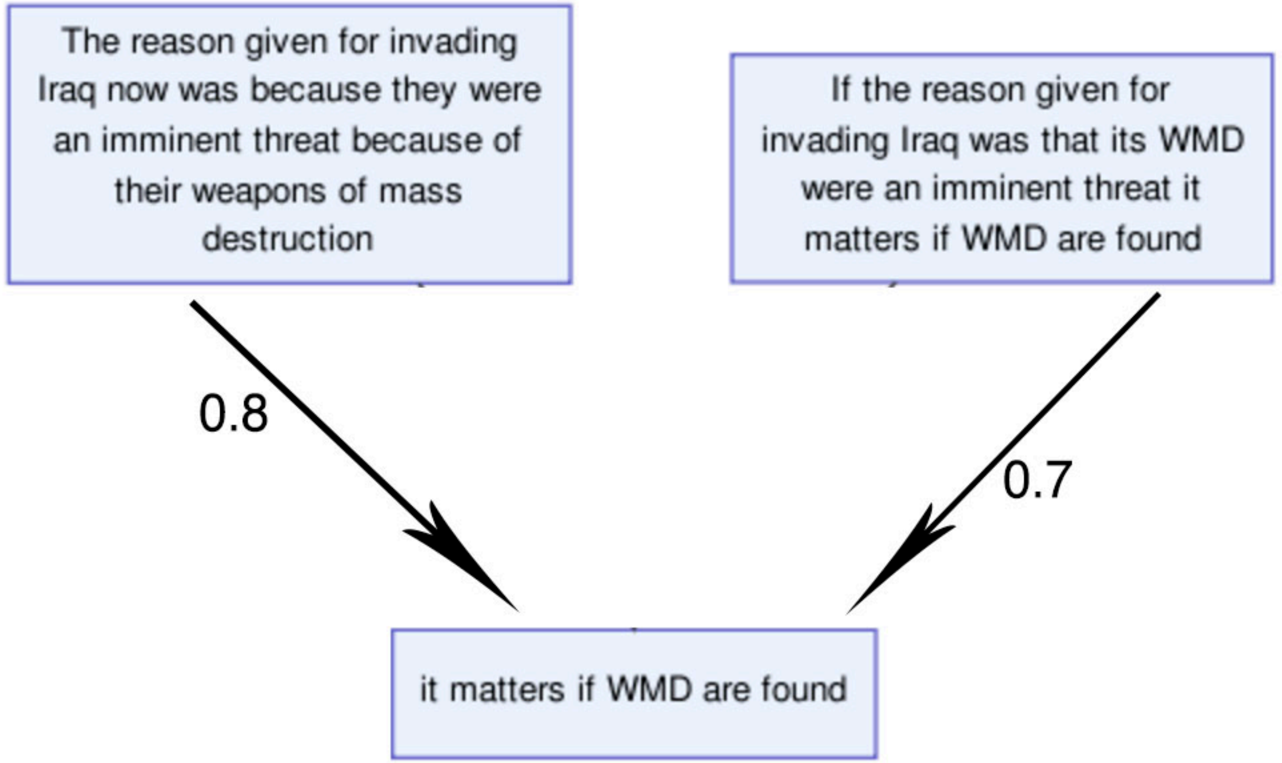
Example (values not correct)

1. Find the edge weights for each possible ordered pair of nodes



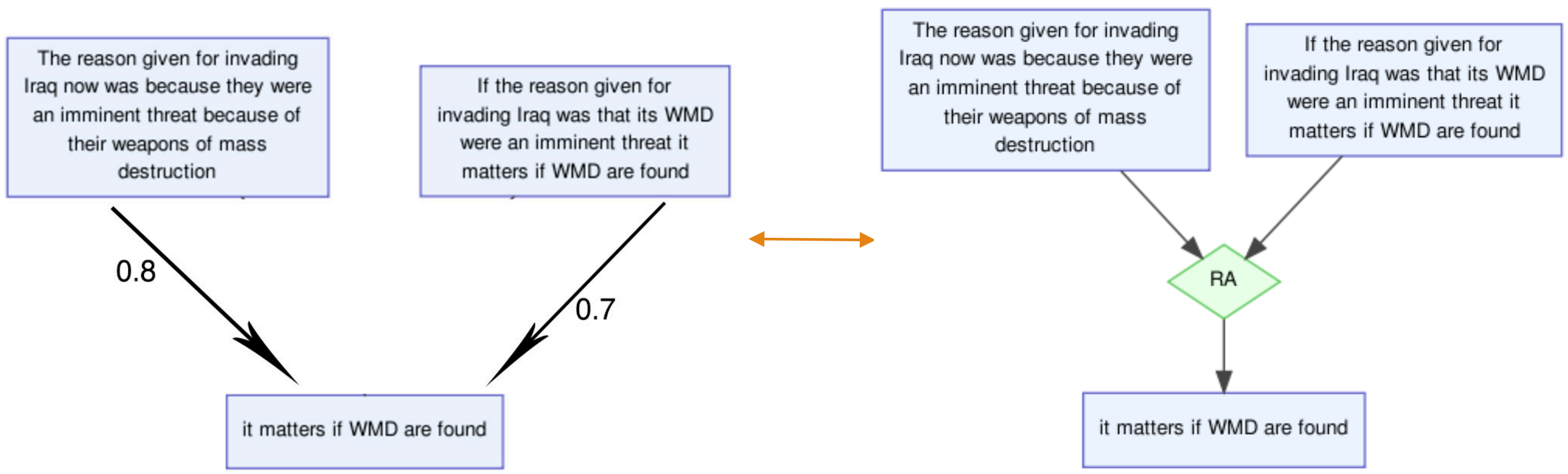
Example (values not correct)

1. Find the edge weights for each possible ordered pair of nodes
2. Construct the tree structure using the edge weights found in Step 1.



Example (values not correct)

1. Find the edge weights for each possible ordered pair of nodes
2. Construct the tree structure using the edge weights found in Step 1.
3. Find out the accuracy for our approach using some scoring model.



Initial Setup: Classifier Settings

1. SVM with an rbf kernel (radial basis function, a.k.a. Gaussian kernel) with the task of binary classification
2. Input as an ordered pair of texts **(text and hypothesis)**.
3. About Araucaria DB:
   - Araucaria DB provided by AIFdb (http://www.arg.dundee.ac.uk/aif-corpora/)
   - The database consists of 661 argument maps.
   - Take all the support relation pairs as the input for the support labels.
   - Take all possible ordered pairs which don't have a support edge as a valid neutral pair
     - Down sample to have a balanced train data.

Initial Setup: Features

1. Discourse Markers (e.g. “if”, “therefore”)
2. Modal Features (e.g. “would”, “could”)
3. Wikipedia entity distance between the text and hypothesis
4. Count of all possible word bi-grams from the train set
5. Count of all possible POS bi-grams from the train set
6. Avg. vector over the words in text and hypothesis, found by using the Google News trained word vectors [8]. (https://code.google.com/p/word2vec/)

Initial Setup: Results

1. Support Pairs Labeling: **523** labeled as Support, 466 labeled as Neutral.
2. Neutral Pairs Labeling: **3957** labeled as Neutral, 2511 labeled as Support.
3. Accuracy: 0.604
4. Macro-averaged precision: 0.53
5. Macro-averaged recall: 0.565
6. There is no baseline because all the previous approaches have been different, meaning they have either found out the undirected edges or they have performed only step 3 (after successful step 2)
7. EDITS Entailment Tool results for our train set:
   Support Pairs Labeling: **192** labeled as Support, 768 labeled as Neutral.
   Neutral Pairs Labeling: **6623** labeled as Neutral, 626 labeled as Support.

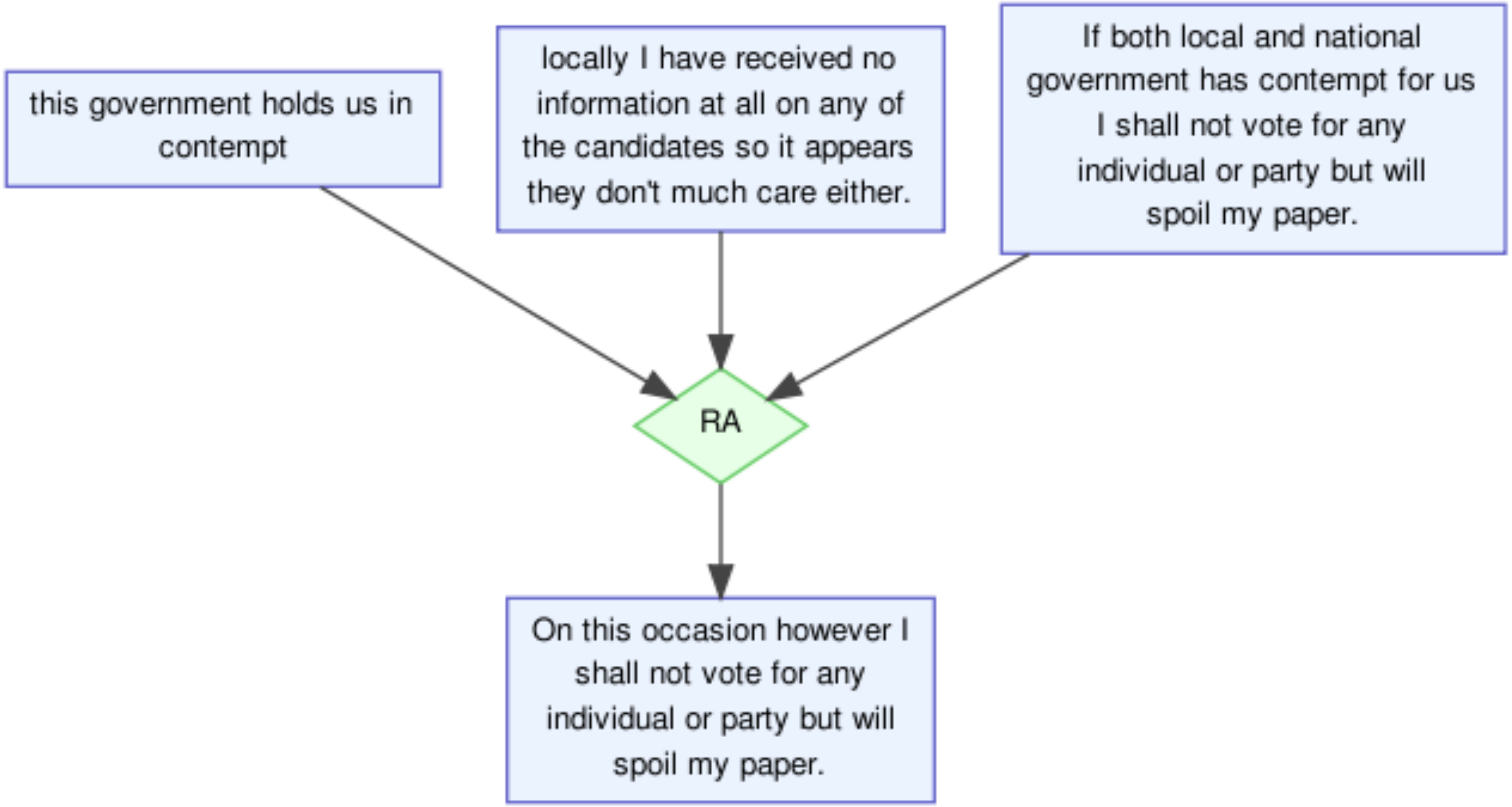
Present Investigation

1. After realizing the results to be not so promising, we tried to debug the false classifications manually.
2. Observed was that it was in fact because of the poorly chosen features.
3. Decided to dive deeper into the given dataset to identify the features causing the argumentation.
4. Though many arguments involved what we identified as “Complex Logic” which can not be captured by numerical features so easily, we still identified some features that can help us improve the accuracy.

Present Investigation: Plausible Features

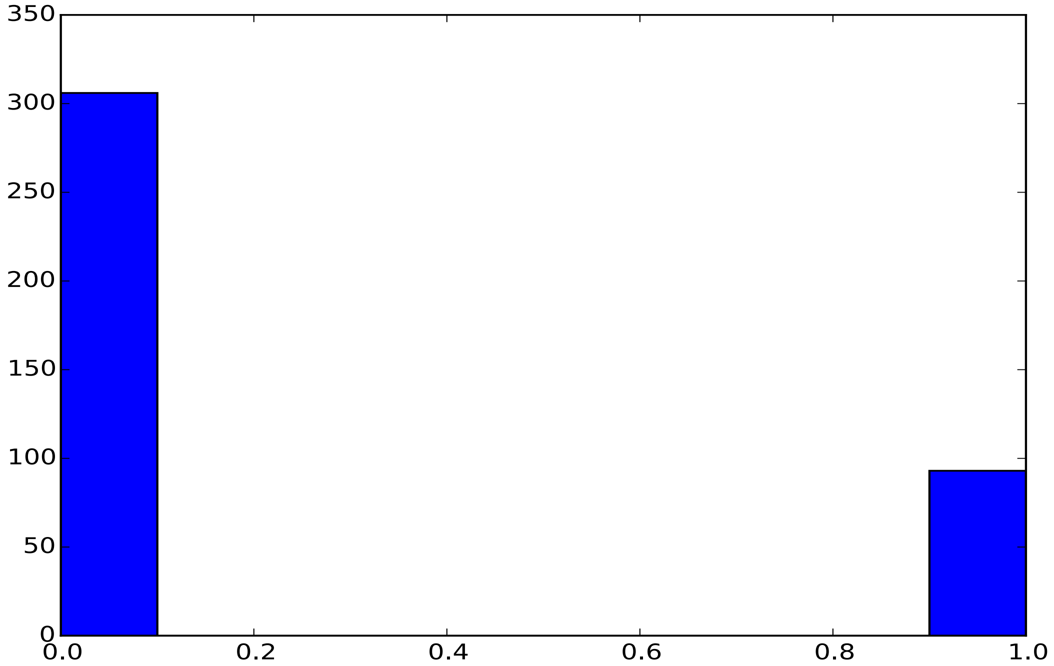
1. Relative length of the text between argument nodes and conclusions.
2. POS of the word which followed the verb, e.g. “We should” should usually occur in an affirmative sentence, giving rise to a possible conclusion.
3. Identifying effective bigrams and unigrams may be more effective than using all possible bigrams.
4. Containment/Linking criteria of conclusion in premise (might involve context). Described by example in next slide.
5. Finding paraphrase similarities.
6. POS/Tense of the verbs involved.
7. More similarity measures between text and hypothesis (other than the common Wikipedia Entities measure), e.g. verb context.

Plausible Feature: Containment/Linking



Plausible Feature: Max Length Role

- Relative length between the texts and the hypothesis.
- Max length nodes are usually more probable to be text than hypothesis.



Role distribution for nodes with maximum length in the respective argument. 0 stands for text, and 1 stands for hypothesis

Plausible Feature: Effective Bigrams

- Bigram **“we should” 7.56** times more to belong in the hypothesis set than the text set.
- Bigram “**if the**” is **2.45** times more probable to be in the text set than the hypothesis set

Bigrams in Text (cutoff=3)

1	Bigram	Count
2	of,the	213
3	in,the	138
4	to,the	94
5	is,to	81
6	for,the	81
7	the,best	77
8	it,is	77
9	is,not	72
10	that,the	62
11	if,the	62
12	is,a	61
13	the,world	58
14	to,be	57
15	means,to	49
16	and,the	48
17	there,is	48
18	will,be	47
19	can,not	45
20	of,a	45
21	if,a	44
22	is,the	44
23	on,the	43
24	best,means	42
25	in,a	38
26	do,not	31
27	one,of	31
28	at,the	30
29	if,we	30

Bigrams in Hypothesis (cutoff=3)

1	Bigram	Count
2	of,the	69
3	in,the	50
4	is,not	47
5	there,is	36
6	it,is	34
7	not,be	32
8	should,not	32
9	for,the	29
10	to,the	28
11	we,should	27
12	to,be	24
13	that,the	22
14	should,be	22
15	is,a	22
16	can,not	20
17	the,world	19
18	is,the	19
19	of,a	18
20	will,be	17
21	from,the	15
22	does,not	15
23	he,is	14
24	you,should	14
25	and,the	13
26	at,the	13
27	we,are	13
28	this,is	13
29	is,no	12

Conclusion and Future Work

1. Our approach is novel and we hope to arrive at some interesting results.
2. In future, hope to implement these complex features into the classifier.
3. Also play with different classifiers available
   - The work done for Natural Language Inference by Bowman et al. uses a neural network model centered around a Long Short-Term Memory network to achieve the state of the art efficiency
4. Once we have the classifier ready, we will move on to steps 2 and 3 as described in the approach.

Important References

1. Mochales, Raquel, and Marie-Francine Moens. "Argumentation mining." Artificial Intelligence and Law 19.1 (2011): 1-22.
2. Lawrence, John, et al. "Mining arguments from 19th century philosophical texts using topic based modelling." ACL 2014 (2014): 79.
3. Peldszus, Andreas, and Manfred Stede. "Joint prediction in MST-style discourse parsing for argumentation mining." Proc. of the Conference on Empirical Methods in Natural Language Processing. 2015.
4. Bowman, Samuel R., et al. "A large annotated corpus for learning natural language inference." arXiv preprint arXiv:1508.05326 (2015).