

Argument Identification

Identifying Texts Containing Arguments

Jeremy Sakovich

Introduction: Computational Argumentation

One of the most distinctive features of human intelligence is the ability to reason through the use of logical arguments.

Introduction: Computational Argumentation

One of the most distinctive features of human intelligence is the ability to reason through the use of logical arguments.

Historically the study of arguments was conducted primarily by Logicians and Philosophers.

Introduction: Computational Argumentation

Today Computational Argumentation is an emerging subfield in Artificial Intelligence where researchers are using computers to analyse human reasoning and arguments which occur in natural language and text.

Introduction: Computational Argumentation

One current major Artificial Intelligence research project in the field of Computational Argumentation is being conducted by IBM's Project Debater team.

Introduction: Computational Argumentation

One current major Artificial Intelligence research project in the field of Computational Argumentation is being conducted by IBM's Project Debater team.

The Project Debater team is engaged in research to develop an Artificial Intelligence system capable of carrying a variety of reasoning tasks including the ability for an AI to engage in meaningful debate with humans.

Why Argument Identification?

Argument identification is a fundamental step in reasoning.

Why Argument Identification?

An Artificial Intelligence system capable of engaging in reasoning must be capable of distinguishing arguments from other sorts of discourse including: instructions, unsupported opinions, lyrics, fiction, descriptions, explanations, and so forth.

How to Identify Arguments with Computers?

How to Identify Arguments with Computers?

My Approach:

- 1) Create a corpus (a set of texts) tagged manually as arguments or non-arguments.

How to Identify Arguments with Computers?

My Approach:

- 1) Create a corpus (a set of texts) tagged manually as arguments or non-arguments.
- 2) Train a machine learning model on a subset of the corpus.

How to Identify Arguments with Computers?

My Approach:

- 1) Create a corpus (a set of texts) tagged manually as arguments or non-arguments.
- 2) Train a machine learning model on a subset of the corpus.
- 3) Test the model on the remaining data.

The Data

The Data

- LSAT practice argument questions.

The Data

- LSAT practice argument questions.
- Descartes's *Meditations*.

The Data

- LSAT practice argument questions.
- Descartes's *Meditations*.
- United Nations Security Council Transcripts.

The Data

Each text was broken into paragraphs.

The Data

Each text was broken into paragraphs.

Then they were all manually tagged as an argument, or not.

The Data

In total, 605 paragraphs were tagged for computational analysis.

Machine Learning

A decision tree model was trained on approximately 70% of the corpus.

Results:

When tested, the trained model successfully identified 78% of the argument texts.

Limitations:

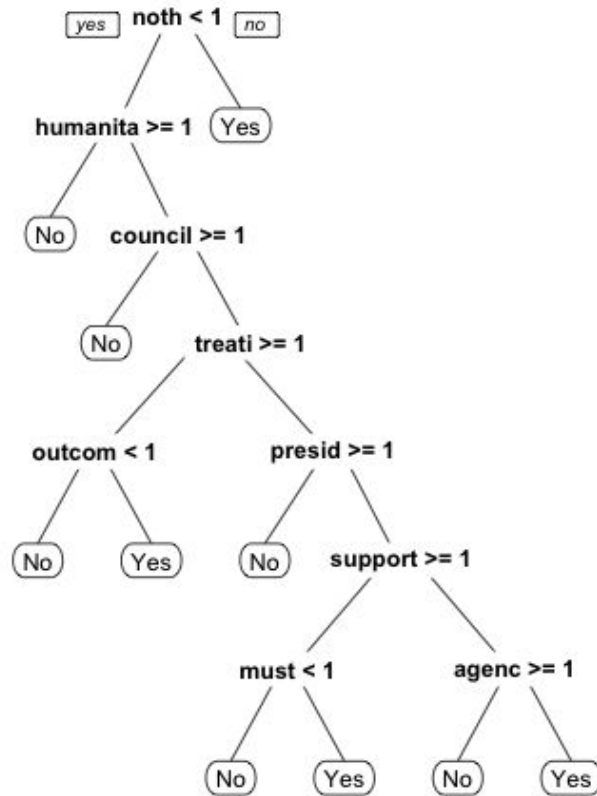
There are some limitations to the model.

Limitations:

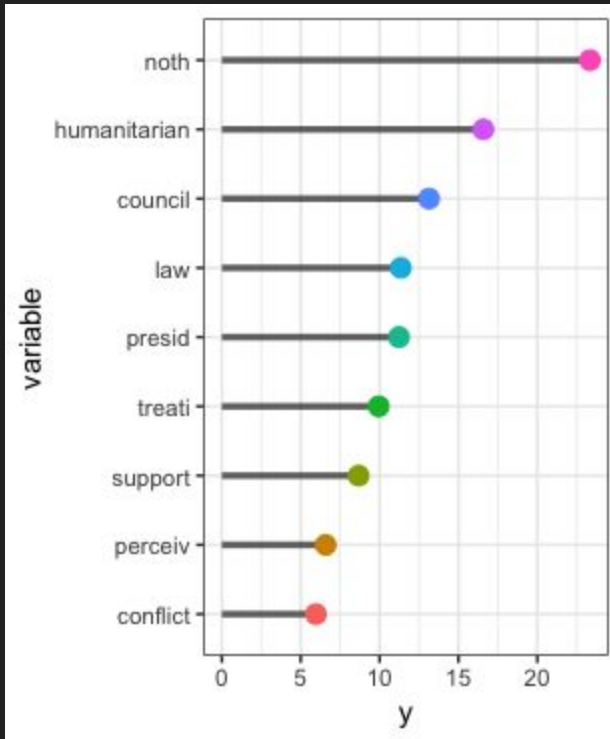
The primary limitation:

- A large portion of the training data was from United Nations transcripts.

Decision Tree



Importance of the
actual words
(stemmed for
analysis) which
were used by the
trained model



Recommendations:

- 1) Retrain the model with a more varied dataset.

Recommendations:

- 1) Gather more data and retrain a model with a more varied dataset.
- 2) Try an approach which would involve creating a dictionary of common argument “indicator” words: *because, therefore, consequently...*

Conclusion:

Overall, this preliminary approach shows that it is possible to train a model capable of identifying paragraph length texts which contain arguments.

Further study into this topic is merited and could yield better results.