Argument Identification

Identifying Texts Containing Reasoning

by

Jeremy Sakovich

Introduction: Reasoning Systems

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

Introduction: Reasoning Systems

A highly robust Artificial Intelligence system would be able to:

1) Identify human reasoning in text

Introduction: Reasoning Systems

A highly robust Artificial Intelligence system would be able to:

1) Identify human reasoning in text

2) Parse out components of the structure of the reasoning present in the text

Introduction: Reasoning Systems

A highly robust Artificial Intelligence system would be able to:

1) Identify human reasoning in text

2) Parse out components of the structure of the reasoning present in the text

3) Engage with reasoning in texts by offering:

- Counter points
- Counter arguments
- Fact checking

IBM’s Solution: Project Debater

IBM’s Project Debater team is engaged in research to develop an Artificial Intelligence system capable of carrying a variety of reasoning tasks.

IBM’s Solution: Project Debater

IBM’s Project Debater team is engaged in research to develop an Artificial Intelligence system capable of carrying a variety of reasoning tasks.

Project Debater aims to create a reasoning system which can debate meaningfully with highly skilled human debaters.

IBM’s Solution: Project Debater

IBM’s Project Debater team is engaged in research to develop an Artificial Intelligence system capable of carrying a variety of reasoning tasks.

Project Debater aims to create a reasoning system which can debate meaningfully with highly skilled human debaters.

It is thought such technology holds a number of useful applications in a variety of fields.

The Problem: Manual Tagging of Texts

One of the most costly tasks in the kind of research needed to make Project Debater a success is the time it takes to manually tag texts appropriate for use in training machine learning models.

The Problem: Manual Tagging of Texts

One of the most costly tasks in the kind of research needed to make Project Debater a success is the time it takes to manually tag texts appropriate for use in training machine learning models.

It can take countless hours of human work in reading simply to identify texts which contain arguments.

My Solution:

Create a machine learning model capable of automating the costly step of identifying arguments.

My Approach:

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

My Approach:

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

My Approach:

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

Results:

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

Limitations:

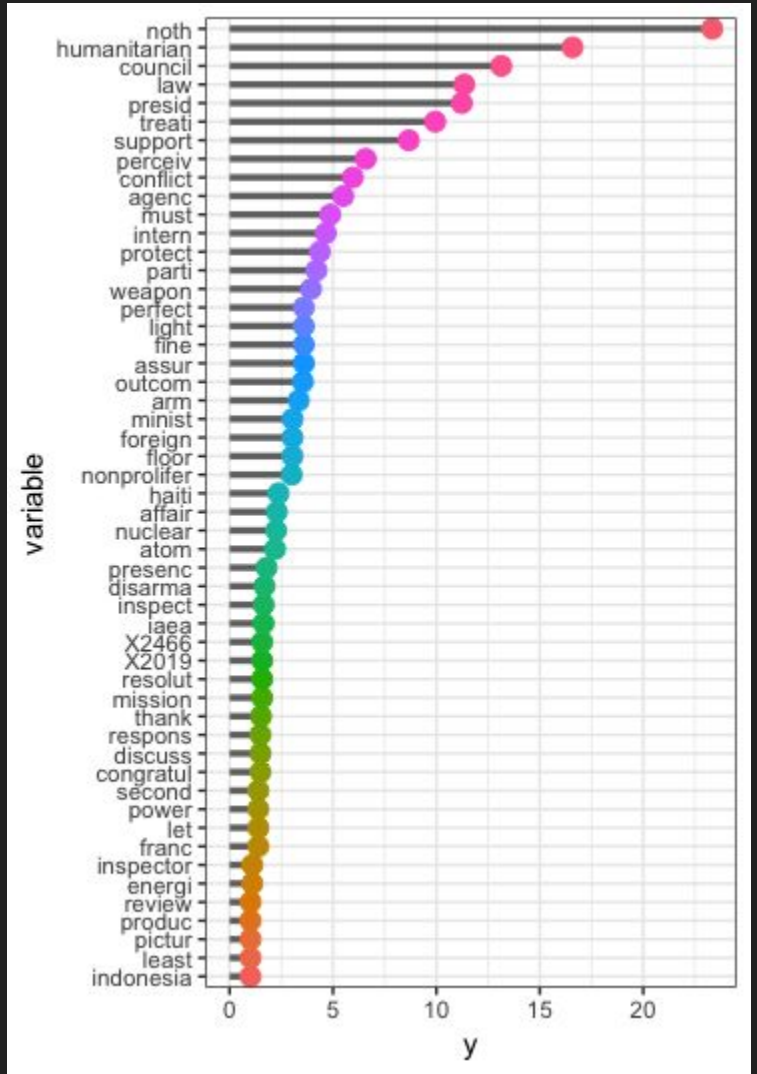
There are some limitations to the model:

Limitations:

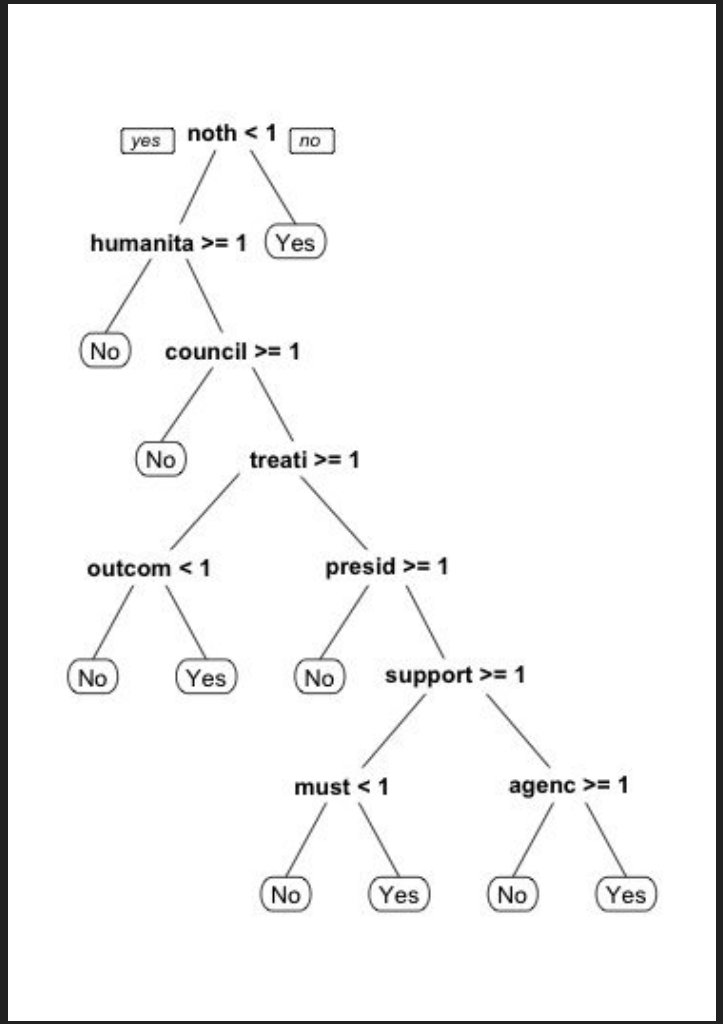
The primary limitation to the model:

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

Importance of the stem words used in the trained model



Decision Tree



Recommendations:

1) Retrain a model with a more varied dataset.

Recommendations:

1) Retrain a model with a more varied dataset.

Recommendations:

1) 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 study shows that it is possible to train a model capable of identifying texts which contain arguments.

Conclusion:

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

With further iteration and work, it could be a significant cost saver in terms of time to utilize such a model to aide IBM’s Project Debater team.