<u>AODA Case Study</u>

Regan Meloche - Dec 18, 2020

This is a summary of an **article by Serena Peruzzo** detailing a case study on using NLP processes to extract info from government legislation.

The 'Accessibility for Ontarians with Disabilities Act' (AODA) was passed in 2005, and defines rules and requirements for accessibility and sets processes for eliminating barriers for people with disabilities.

Goals of the case study

- Simplify the extraction of rules and obligations
- Understand entities responsible for compliance
- Link different pieces of legislation

Challenges

- Language parsing is HARD (formatting, abbreviations, specific legal doc mentions, etc)
- Vocab is limited and specialized (good), but very context-sensitive
- No pre-trained industry models. No labelled set
- Syntax is complex and non-linear

Summary

The Framework combined a custom rule-based system with elements of a standard NLP pipeline and unsupervised ML. The NLP pipeline consists of the following components: Tokenizer, Lemmatizer, POS Tagger, dependency parser. The aim was to extract info in terms of rules defined in the legislation, entities responsible for compliance, and organize rules into groups (clustering).

The rule-based system focused on scanning the text of the law and extracting rules related to burdens (i.e. who is responsible for supporting the legislation).

This was done by identifying words that expressed some sort of obligation (shall, must, oblige…). It is acknowledged to be a coarse approach, but since the entire vocabulary is limited, it works quite well.

Once the fact of obligation is established, the next need is to identify the entities responsible for fulfilling that obligation. NLP has robust libraries for extracting parts-of-speech from text and establishing lexical dependencies, so this can be done without too much trouble. Finally, they wanted to differentiate the responsible entities into clusters based on whether it was a private or public responsibility.

Next steps

- This research can help with extraction and summarizing so that rules and requirements can be made more accessible.
- Can help lawmakers by highlighting parts of legislation that are ambiguous and could be re-written or adapted to be more clear and accessible.

Application to a PDE

This is a very valuable case study, since it highlights the possibilities and limitations of NLP within the context of government rules. We may run into the same roadblock of the lack of pre-labelled data to train statistical models on.

The proof of concept for using NLP practices to classify different documents into categories is promising, as this may be applicable to building a search engine for government rules, aligning with the rules module of the PDE.