ANALYSIS OF SENTIMENT DIRECTION FROM THE HANSARD DEBATE ARCHIVE ACROSS TWO CENTURIES

MORENO VARDANEGA

MSc in Big Data

UNIVERSITY *of* STIRLING

SUMMARY

This Project concerns the following central research question: "can machine learning techniques be used to automatically examine the change in aggregate levels of sentiment polarity within a defined topic in the UK Parliament speeches?"

We address the question by making use of methods of natural language processing (NLP). This project evaluates whether opinion-mining techniques can successfully be applied in Parliament Debates text analysis. As Data we use transcriptions of the speeches of the plenary meetings in the Hansard Archive (digitised debates of UK Parliament from 1803 to present). A Sentiment algorithm has been implemented to retrieve meaningful patterns that are based on the classification results. A corpus of recent parliamentary proceedings has been collected and a gold standard annotation is created on both subjectivity and orientation.

Main Idea

An automatic sentiment analysis model can be very useful to visualize quickly how the opinion of the Parliament, that usually reflects the National sentiment, changed over a period of time. This enables the user to get an overall picture of the position on a specific issue and to pick up the underlying trends without having to go through the entire sequence of the Hansard speeches. The visualisations have to be clear with no ambiguities and be easy to understand.

Implementation

The sentiment algorithm has been built on top of Nalytics search engine, the core product of Nalanda Technology. Nalytics provides a unique experience, searching the inherent textual structure of documents and unstructured data content to enable fast, accurate searches.

The Sentiment Analysis Model was the outcome of a series of experiments that includes exploring the performances of different supervised learning algorithms such as Support Vector Machines, Linear Regression, Naive Bayes, etc. The classifier has been trained and tested with different parameters on various corpora, as well as dealing with different feature sets. These techniques and methodologies are written in Python using NLTK and SciKit-Learn. Given the complex nature of the parliamentary discourse, the initial results that are presented here are promising.

We created visualisations based on cumulative Line Charts as a means to provide a historical polarity of the selected topic across the years, allowing the user to find out the prevailing opinions on a matter quickly. The underlying technology used for implementing the visualisations is Python library Matplotlib.

Additionally, by looking at the history of a line graph, a sudden increase in negativity or positivity of a particular issue reminds the user that during that period an important event has occurred. For example, if there is an increase in 'Iraq War' as a topic then perhaps it was a new phase in foreign relations regarding that country, or a change in 'NHS' could be caused by new privatisation laws.

Sentiment Analysis Visualisation
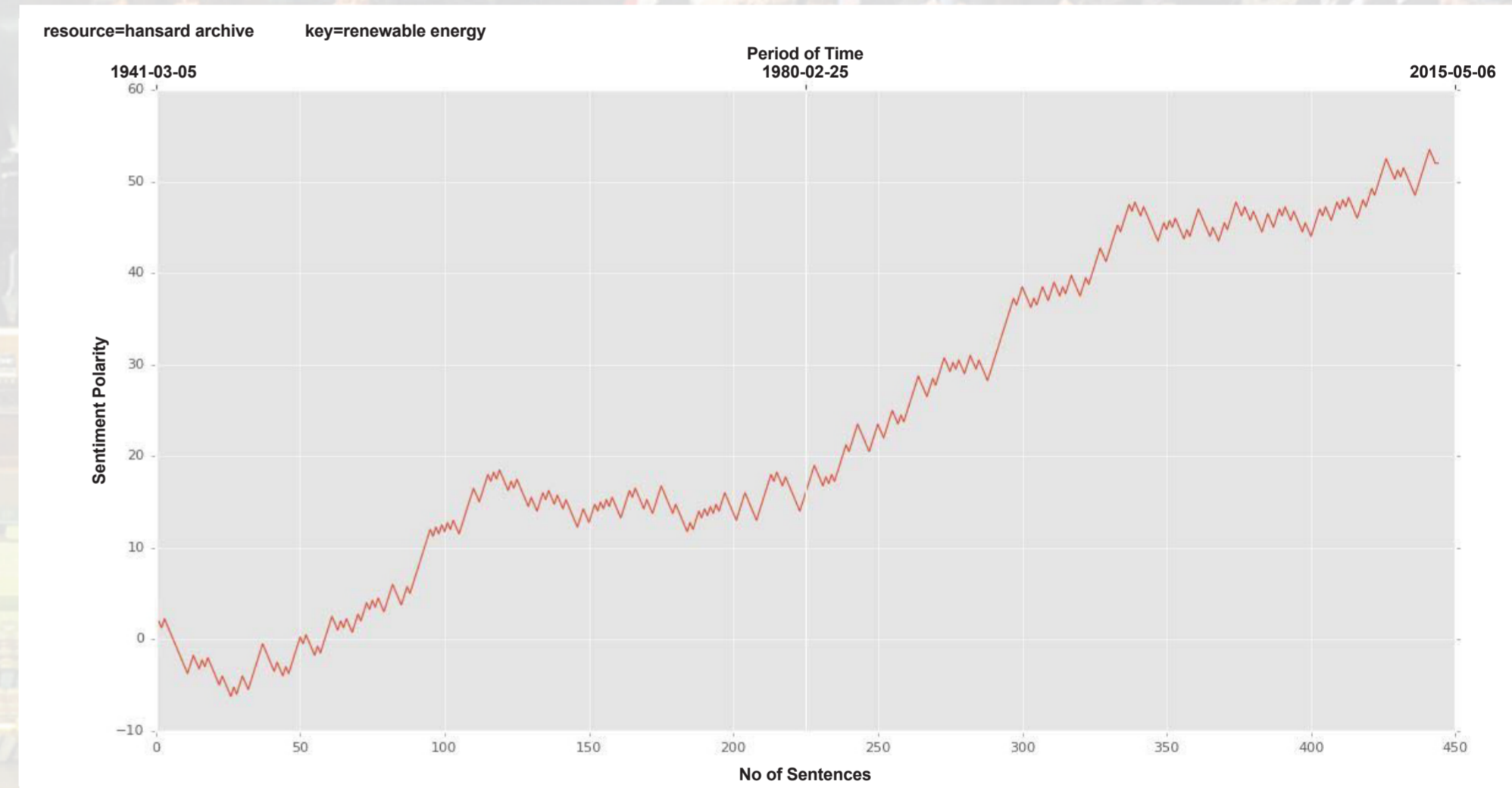
This graph shows the changes in polarity of 'Renewable Energy' across the period of 1941 to 2015