Linking UN Speeches to GDP Growth and the Financial Crisis of 2007 to 2009

Bono Lardinois 11713364.
Jan Janiszewski 10004378.
Maxim Vitalis 11303719.

University of Amsterdam, Fundamentals of Data Science – Assignment 1

Abstract. This research analyzes the impact of the financial crisis from end of 2007 to 2009 on UN speeches, using explorative methods such as Word Clouds, Valence Aware Dictionary and sEntiment Reasoner, and Latent Dirchlet Analysis. Moreover, it analyzed the impact of UN speeches on the GDP growth of their respective country using the Bidirectional Language Representation Transformer (BERT). The results showed a decrease in sentiment but not that speeches influenced their country's GDP growth. Indicating that UN speeches do not reflect the economic crisis.

1 Introduction

The economic crisis, which started in the end of 2007 and ended in 2009 is one of the most severe financial crises since the Great Depression. Financial markets worldwide were crumbling, banks were on the brink of bankruptcy, and major Wall Street giants were brought to their knees. The global economy was estimated to have declined by two trillion dollars. The financial crisis not only left an economic scar on the world but also affected the lives of billions who lost their jobs and their financial stability (Chang et al., 2013).

It would seem that a major crisis like this is on the political agenda of different countries and is, therefore, discussed during the annual United Nations (UN) General Assembly. During this, UN member states can send their representatives to give speeches in the General debate. Speakers during the general debate address topics that reflect on issues important to the UN member state and address globally important topics (Baturo et al., 2017).

Not only should the words of representatives be impacted by a financial crisis, but they should also shape a crisis, with good representatives choosing their words wisely enough to impact their country's economy positively. Therefore, it is essential to know how far representatives and speeches influence their current country's economic well-being.

Given both considerations, we, therefore, specified a general research question, which was then split into two sub-questions:

– General question: How did the most common words in speeches during the economic crisis evolve, and did this impact GDP growth?

– How did the economic crisis of end of 2007-2009 impact the speaking patterns of UN delegates in their speeches?
– How do UN speeches impact their respective country's GDP growth?

2 Methodology

2.1 Data sources

Two different major sources were used for the analyses:

First, we used the GDP statistics provided by the Sustainable Development Solutions Network (a UN initiative), which collected data from 166 countries on different facets supposed to reflect a country's happiness in the years 2005 to 2021 ($N = 2089$; Helliwell et al., 2021). The transformed variable *relative GDP growth* ($\mu = 1.88\%, sd = 5.88\%$; based on the original variable *logarithmic GDP per capita*; $\mu = \$20375; sd = \$19894$) was finally used in our research.

Second, we used the UN General Debate Corpus which contained 8481 speeches by UN delegates from 201 countries from the years 1970 to 2020 (Jankin Mikhaylov et al., 2017).

Both data sets were joined together using a mapping from country name (as provided in Happiness Report data) to ISO3 country code (as provided by UN data; Lukes, 2022); non-matched entries were dropped. This resulted in a final data set containing speeches as well as relative GDP growth ($\mu = 1.91\%, sd = 4.66\%$) information for the years 2006 until 2020 for 128 countries ($N = 1376$).

2.2 Data cleaning

Before analyzing the data, we preprocessed it to allow the explorative analysis as well as the machine learning models to focus on relevant information (rather than e.g., stop words). The data preprocessing steps included:

1. Lowering capital letters: Capital letters were lowered (e.g., s.t. 'A' > 'a')
2. Tokenizing sentences: Sentences were split into tokens using the NLTK tokenize function (Loper and Bird, 2002).
3. Dropping sentence indicators: Sentence indicators (.e.g, *, ; .*) were dropped from the speeches.
4. Dropping numbers: Numbers that were not written out (e.g. *500*) have been dropped from the speeches.
5. Dropping stop words: Typical stop words (e.g., *or, and*) as well as politeness expressions (e.g., *honorable*) were dropped from the speeches.

Additional preprocessing steps were done as part of the LDA data preparation:

6. Lemmatizing words: Words have been lemmatized using the WordNet lemmatizer before using them in the model (Fellbaum, 1998).
7. Creating bigrams: Words that appeared together for at least 10 times have been categorized as bigrams (i.e., a combination of two words which appears as one token in the LDA model; Touw and Janiszewski, 2022).

2.3 Sentiment Analysis

Sentiment analysis was used to investigate whether the financial crisis affected the sentiment of the speeches at the UN General Assembly (2007 - 2009). This analysis, categorized texts into *positive* or *negative* sentiment.

For this, the NLTK implementation of the Valence Aware Dictionary and sEntiment Reasoner (VADER; Loper and Bird, 2002). VADER uses a lexicon with words that are labeled either positive or negative to generate a score based on the analyzed text. Because VADER is especially powerful in analyzing small texts like Twitter, we divided each speech into sentences and analyzed these. Then, we took the average compound score of all the speeches from that year. We used the normalized compound score [between -1 (very negative) and 1 (very positive)], which we deemed better to compare the sentiment of texts over the years (Hutto and Gilbert, 2014). Finally, we performed a t-test to test whether the sample means of the years are significantly different.

2.4 Topic Analysis

We analyzed in how far topics generated by a Latent Dirichlet Allocation (LDA) cover the economic crisis and compared those to topics covered in non-crisis years (2005-2006, 2010-2020; for more information on the LDA, see Blei et al., 2003; for code template used, see Touw and Janiszewski, 2022).

2.5 GDP Prediction Using BERT

In addition to the exploratory analysis methods discussed previously, a predictive analysis was performed. We assumed that it would be potentially possible to predict the growth or decline of GDP based on the speeches delivered at the UN. For this, we retrained a pre-trained Bidirectional Language Representation Transformer (BERT). With its 12 layers and 12 self-attention heads for bidirectional context interpretation (768 hidden nodes) pre-trained on unlabeled text-data, it is surprisingly easy to adjust in order to predict GDP growth with only little retraining of only the last layers.

A number of variations of the BERT model exist and are used for different applications. To meet our needs, finBERT was chosen as it was modified specifically to analyze financial sentiment and can easily be retrained to our needs (Araci, 2019a; Yang et al., 2020a).

To represent the growth or decline of the GDP, we processed the GDP growth to be contained within the three mutually exclusive categories ”negative”, ”neutral,” and ”positive”. Based on Fernald et al. (2019), we consider a GDP growing with a rate between 0% and 3% as normal (i.e., neutral). Although a number of emerging economies had shown that this rule may not be universally true for all countries, individual extremes do not affect global growth figures enough to deviate from this rule; we therefore find this limitation acceptable for the goals of this analysis. We therefore defined the GDP growth categories as follows:

- negative ($GDP - growth < 0\%$)

- neutral: ($0\% \leq GDP - growth < 3\%$)
- positive: ($3\% \leq GDP - growth$)

Although finBERT is a pretrained model, it still needs to be fine-tuned for the intended purpose and on the selected data set. A training, validation, and test data set was sampled from the original data with a 80% training, 10% validation, and 10% test split, respectively. These data sets contain the speeches in combination with the categorical change in the GDP growth category.

Sun et al. (2019) describe different strategies for the usage of BERT for text classification and experiment extensively with the different parameters that are modifiable. For our analysis, we drew great inspiration from these experiments as they are imperative for the best results and overcoming problems such as catastrophic forgetting (McCloskey and Cohen, 1989). We decided on fine-tuning the model over the course of 4 epochs, with each of these cycles consisting of 24 steps in batch sizes of 32. The number of hidden layers and encoder size was, after referencing Yang et al. (2020a), kept at the default values of 12 and 768 respectively. Of the hidden layers, 6 were retrained during fine-tuning. As a basis for our model, a modified version of the original notebooks was used. (Touw and Janiszewski, 2022)

3 Results

In this section the results of the exploratory analysis are presented.

3.1 Wordclouds

As first explorative method, word clouds were generated which showed the 50 most frequent words in the UN speeches by year, with the word size increasing with its frequency. This analysis did not reveal any speech pattern changes between the crisis and non crisis years (for the direct comparison of 2007 and 2008, see Figure 1; for word clouds for all years, see Appendix C).

3.2 Sentiment Analysis

Our sentiment analysis indicates a large decline in speech sentiment between 2007 and 2008 when the crisis fully developed (see Figure 2). To check whether the decline is significant, a t-test was performed. Due to the large sample size, we assume normal distribution of the sample means without especially testing for it (Rosenblatt, 1956). We tested for different hypotheses, namely sentiment scores being equal in 2007 and 2008 (H_0) vs. sentiment scores not being equal in 2007 and 2008 (H_a).

The t-test revealed a significant difference between the scores of 2007 and 2008 ($t(14136) = 3.96; p < 0.01$), implying that there was indeed a different sentiment in the year 2008 compared to the year 2007, which could have been caused by the financial crisis.

(a) 2007 (crisis)

(b) 2008 (non-crisis)

Fig. 1: Word cloud comparison for years 2009 (crisis year) vs. 2010 (non-crisis year)

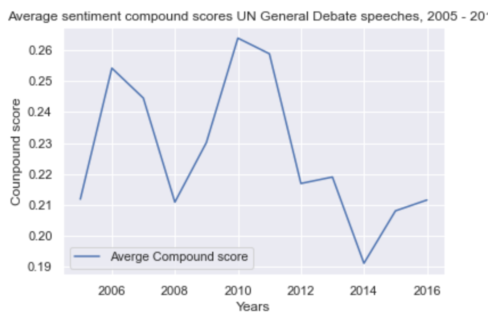

Fig. 2: Graph of the average compound score of speeches given at the general debate between 2006 - 2026.

3.3 Topic Analysis

As previous methods were not clear on the potential impact of the financial crisis on UN speech patterns, we also conducted two LDAs. The LDA topic analysis of crisis year (2007-2009) UN speeches did not reveal any strong focus of UN speeches on the economy in the crisis years (2007-2009). One topic (i.e., topic 8) included certain words related to economic issues (e.g., ıpercent, price, market, production); however, it grouped those words to other words which related to the UN millennium development goals (MDG's) and not to the financial crisis. Additionally, a second (non-crisis) LDA on UN speeches for non crisis years (i.e., before 2007, after 2009) included a topic similar to the topic in the previous crisis LDA, which is unlikely for a topic related to the financial crisis. Therefore, we conclude that there is no evidence from the topic analysis that the financial crisis was a "hot topic" in UN speeches in 2007-2009 (see table 1 for the topic comparison). Moreover, from the analysis of all ten LDA topics of both LDAs, we deduct, that economical matters are not relevant to UN speeches in general (for a table of all topics, see Appendix A and Appendix B).

LDA: crisis years, topic 8	LDA: non crisis years, topic 2
per cent	per
price	cent
market	per cent
production	food
water	trade
mdgs	market
central	health
america	energy
drug	eduction
health	billion

Table 1: Comparison of LDA topics related to economy for LDA on crisis years vs. LDA on non-crisis years

3.4 GDP prediction

Results indicate that the finBERT model did not provide meaningful predictions of GDP growth based on the speeches alone (weighted $\mu_{accuracy} = 0.4$).

Although none of the categories scored well, there is noticeable difference in performance between them. The model performs better on either negative or positive data compared to those labeled neutral. The f1-scores further underlines the models poor performance (see table 2). From these results, we conclude that speeches did not have an influence on GDP growth of their respective countries.

Category	Negative	Neutral	Positive
Accuracy	0.35	0.11	0.51
Recall	0.11	0.65	0.21
f1 score	0.17	0.19	0.30

Table 2: Performance metrics for finBERT model

4 Discussion

The analyses on the first research question on the impact of the financial crisis on UN speeches revealed different results. The sentiment analysis indicated that overall speech sentiment dropped from 2007 to 2008. However, results from other methods (i.e., word clouds, LDA) did not support the sentiment analysis. Therefore it is not clear if UN speech sentiment dropped due to the financial crisis (as suggested by the sentiment analysis) or due to other influences (e.g., political issues in Africa, as suggested by our word cloud).

The analysis on the second research question on the impact of UN speeches on GDP growth was done using the finBERT and the BERT models. Although finBERT is a highly advanced and proven model, it did not perform well on this specific task. Based the research by Araci (2019b); Yang et al. (2020b); Liu et al. (2021), we deduce the following possibilities for the poor results as being most likely:

First, our training corpus was very small in comparison to other research done using BERT and finBERT and and it might be insufficient to train a model for our task.

Second, due to computational power, the training of our BERT model was limited to very few training epochs, contrary to what was suggested by Sun et al. (2019).

Third, GDP growth could indeed not be related to general debate UN speeches as they rather cover general societal and political topics (.e.g, human rights, wars) and not specific financial topics (those are more likely to be discussed in specific expert groups). This hypothesis is also supported by our word clouds and the LDA, which both found that economy as a topic doesn't appear often UN speeches.

Unfortunately, we cannot be sure of the true reason for the bad performance of finBert as research on BERT for text classification is limited and we therefore don't have much research to fall back on in order to argue in any direction of thought (Yang et al., 2020a). Therefore, we strongly suggest that future research using either BERT of finBERT focuses on understanding the workings of BERT with small datasets, so that it becomes clear if BERT performance really drops significantly with limited data and computing power.

5 Conclusion

In this report, we analyzed UN general debate speeches using a number of exploratory and predictive methods.

First, the exploratory models revealed mixed evidence for an influence of the financial crisis on UN speeches, depending on the method with which they were analyzed.

Second, UN speeches' impact on their respective country's GDP growth was not supported by the finBERT model, with various reasons being the possible explanation.

Therefore, we can finally state that the financial crisis of end 2007-2009 could have impacted the UN general debate speeches during that time (and especially their sentiment), but we cannot conclude that the topics discussed were linked to the crisis, nor can we conclude that UN general debate speeches have any impact on the GDP of a country.

Bibliography

Araci, D. (2019a). FinBERT: Financial Sentiment Analysis With Pre-Trained Language Models. *arXiv preprint arXiv:1908.10063*.

Araci, D. (2019b). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Baturo, A., Dasandi, N., and Mikhaylov, S. J. (2017). Understanding state preferences with text as data: Introducing the un general debate corpus. *Research & Politics*, 4(2):2053168017712821.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Chang, S.-S., Stuckler, D., Yip, P., and Gunnell, D. (2013). Impact of 2008 global economic crisis on suicide: time trend study in 54 countries. *Bmj*, 347.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.

Fernald, J., Li, H., et al. (2019). Is slow still the new normal for gdp growth? *FRBSF Economic Letter*, 17:1–5.

Helliwell, J. F., Layard, R., Sachs, J. D., and Neve, J.-E. D. (2021). World happiness report 2021.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Jankin Mikhaylov, S., Baturo, A., and Dasandi, N. (2017). United Nations General Debate Corpus.

Liu, Z., Huang, D., Huang, K., Li, Z., and Zhao, J. (2021). Finbert: A pre-trained financial language representation model for financial text mining. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 4513–4519.

Loper, E. and Bird, S. (2002). Nltk: The natural language toolkit. *CoRR*, cs.CL/0205028.

Lukes (2022). Iso-3166-countries-with-regional-codes. https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv.

McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the national Academy of Sciences*, 42(1):43–47.

Sun, C., Qiu, X., Xu, Y., and Huang, X. (2019). How to fine-tune bert for text classification? In *China national conference on Chinese computational linguistics*, pages 194–206. Springer.

Touw, D. and Janiszewski, J. (2022). Nlp. https://github.com/JanJJaniszewski/NLP.

Yang, Y., Uy, M. C. S., and Huang, A. (2020a). Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.

Yang, Y., Uy, M. C. S., and Huang, A. (2020b). Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.

A Top ten words for topics for 10-topic LDA trained on non-crisis years speeches

topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7	topic 8	topic 9	topic 10
african	per	european	iraq	bosnia	european	israel	america	japan	nuclear
africa	cent	european union	syria	herzegovina	europe	iran	people	afghanistan	weapon
african union	per cent	spain	humanitarian	bosnia herzegovina	humanitarian	pakistan	drug	energy	pandemic
mali	food	relation	terrorist	indonesia	syria	azerbaijan	mexico	water	peacekeeping
continent	trade	serbia	refugee	asean	crime	georgia	latin	nuclear	disarmament
partner	market	cyprus	freedom	slovenia	want	nuclear	american	asia	nepal
south	health	europe	violence	southeast	weapon	afghanistan	solidarity	thailand	treaty
governance	energy	greece	canada	partnership	freedom	ukraine	crime	central	covid19
particularly	education	international law	ireland	strengthen	international law	palestinian	colombia	kyrgyzstan	comprehensive
sudan	billion	negotiation	palestinian	ocean	european union	territory	public	assistance	inclusive

B Top Ten words for topics for 10-topic LDA trained on crisis years speeches

topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7	topic 8	topic 9	topic 10
nuclear	honduras	humanitarian	war	african	japan	victim	per cent	republic	african
european	military	oil	power	republic	india	nuclear	price	georgia	sierra
weapon	republic	republic	want	chad	nuclear	nicaragua	market	territory	delegation
afghanistan	freedom	response	freedom	poor	disarmament	latin	production	international law	african union
peacekeeping	coup	palestinian	value	particular	weapon	american	water	south	continent
european union	constitutional	price	let	effect	nuclear weapon	remember	mdgs	turkey	implementation
convention	give	growing	israel	continent	technology	purpose	central	sovereignty	partner
negotiation	may	military	much	burundi	korea	public	america	integrity	arm
treaty	solidarity	market	history	concern	republic	let	drug	territorial	strategy
mission	let	per cent	century	pakistan	nuclear disarmament	unity	health	justice	somalia

C Word Clouds For Years 2005-2020

(a) 2005

(b) 2006

(a) 2007

(b) 2008

(a) 2009

(a) 2010

(a) 2011

(b) 2012

(a) 2013

(b) 2014

(a) 2015

(b) 2016

(a) 2017

(b) 2018

(a) 2019

(b) 2020