Athens Journal of Sciences- Volume 9, Issue 2, June 2022 – Pages 91-106

Sentiment Analysis of International Relations with Artificial Intelligence

By Dadhichi Shukla[*] & Stephan Unger[±]

Geopolitical strategy is characterized by a dynamic and complex structure of entity relationships, geo-spatial data and human decisions. We employ machine and deep learning techniques to retrieve the sentiment between countries through scraping and analyzing news articles. The change in the sentiment score between countries allows to analyze historic developments of international relations as well as to evaluate the primary and secondary network effects of potential events and policy decisions on the global relationship structure. We find that the key for the most accurate real mapping of the sentiment score between countries is the maximization of the quantity of news while simultaneous minimization of the noise added by the news. Moreover, we show the potential of Artificial Intelligence (AI) to improve and forecast international relations.

Keywords: *Natural language processing, international relations, sentiment analysis, geo-political forecasting.*

Introduction

Strategic decision making depends on a variety of aspects. These aspects are very often a diffuse formation and inter-dependencies of historic formations, relations, internal political structures, economic situations, and strategic goals. These factors are most often given like an externality to the decision making process. On top of that comes the individual characteristic of the decision maker, which depends on preferences, experience and personal goals. These factors might very often be in conflict with the existing externalities. But not only individuals can be in conflict with the external structure, also whole political or economic bodies such as of governmental groups or institutions are getting regularly in conflict with either given structures, or forces which are counteracting their goals. This makes the decision making process sometimes to a random process and thus, outcomes hard to predict.

First, it is important to clarify who the decision maker is. Very often, a country or the country's leaders are seen to be as the decision makers on a geo-political level. But the subject is far deeper reaching as one might think in the first moment. Since all geo-political structures are built up like a hierarchical pyramid, the key influencing factors are very often and very likely to find below the top level, e.g., in forms of advisors, institutions, think tanks, foundations, influential corporations, or other entities.

In order to better understand which influence the underlying structures will

[*]Senior Data Scientist, STRG.at, Austria.
[±]Associate Professor, Saint Anselm College, USA.

https://doi.org/10.30958/ajs.9-2-1 doi=10.30958/ajs.9-2-1

have on the leader's final decision, it is often very useful to be able to model the goals and preferences of each of these entities separately, as well as the structure of their relationship to each other, including their significance and relevance in the decision making process.

The problem is that with the increasing granularity of the sub-structure, the signals become more and more uncertain, or disturbed, meaning that institutional choices or preferences for certain decisions will vary increasingly the more individual factors are considered. On top of that comes that at a certain point it not only gets too complex to model the individual inter-relationships but also too diffuse and the marginal utility of the gained insights decrease.

Therefore, it makes sense to follow a certain high-level approach for modeling purposes. Nevertheless, underlying factors should be comprised and available to the system according to their relevance and significance to the overall decision making process.

An important point to stress is the factor that in a globalized world, political structures face a layover of independent structures, which can be seen as external, or hybrid structures, such as international organizations such as the United Nations, Warsaw Treaty Organization, North Atlantic Treaty Organization (NATO), etc. This means that the analysis of the individual actors, e.g., countries, must be conducted under the light of the interweaving of international Non-Governmental Organizations (NGOs) with national entities.

In this paper we apply Natural Language Processing (NLP), entity classification, and sentiment analysis to major geo-political news headlines involving countries to map geo-political relationships between countries. We then construct and feed a network structure with the extracted sentiment values and conduct several tests for reliability and robustness. Our key hypothesis is that the sentiment of news articles is a valid measure of the relationship between two countries. Furthermore, the sentiment analysis between two countries allows to deduct a network sentiment structure in such a way that it reflects reality. This in turn allows to conduct predictions about future relationship developments between countries.

In the literature, research on NLP and text mining is being conducted on several levels, from conceptual/theoretical modeling to technical optimization. One common denominator amongst most recent literature is its reference to artificial intelligence and machine learning. Fehlmann and Kranich (2019) analyze how AI systems can be tested, depending on the system's goals and objectives. Razek (2021) investigates the theoretical interdependence between the two evaluation notions of operational observation and mathematical modeling, two concepts which are crucial in the fundamental conception of an NLP machine learning framework. He finds that mathematical modeling needs operational observation simply to be credible and that the second needs the first for deeper research.

Other, more technology-applied approaches extend regular NLP frameworks and incorporate new database structures which allow faster accessing and processing of generated data, i.e., Burdack et al. (2018) introduce a specialized, lightweight in-memory database management system which perfectly fits to the characteristics of time series sensor data. They show that time series sensor data can be stored efficiently using a new table structure. Their storage logic leads to an efficient data access of the compressed in-memory data structure, thus, every reporting or analysis task accesses the data efficiently and fast.

Regarding the theoretical analysis of International Relations there exist several approaches to geo-political decision making analysis and prediction. A key-fundamental analysis of the different approaches is presented by Allison Graham (1971), who distinguishes three different types of models, which are used by most analysts to explain and predict the behavior of national governments: 1. The Classical model (the rational actor model), 2. The Organizational Process model, and 3. The governmental (bureaucratic politics model).

All these types of theoretical models can nowadays be learned and trained in AI models using NLP and text mining techniques, paired with game theoretical models. Traditionally, learning has been studied either in the unsupervised paradigm (e.g., clustering, outlier detection) where all the data are unlabeled, or in the supervised paradigm (e.g., classification, regression) where all the data are labeled (Zhu and Goldberg 2009). Named entity classification problems have been studied by various researchers and is an ever increasing topic. Collins and Singer (1999) show that the use of unlabeled data can reduce the requirements to train a classifier for unlabeled data by making use of the leverage of natural redundancy in data. Yogatama et al. (2015) use embedding methods for fine grained entity type classifications and show that these outperform state-of-the-art methods on benchmark entity classification data sets. Niu et al. (2003) apply a bootstrapping approach to named entity classification and show that this method approaches supervised name entity performance. Cimiano and Voelker (2005) address the unsupervised classification of named entities with regard to large sets of classes which are specified by a given ontology.

We use the FLAIR NLP framework, provided by Akbik et al. (2019), to facilitate training and distribution of state-of-the-art sequence labeling, text classification and language models, which provides an interface for conceptually very different types of word and document embeddings.

For context-dependency we use Bidirectional Encoder Representations from Transformers, or short-called: BERT (Alammar 2019), which is an open-source neural network-based technique for natural language processing pre-training. It is a major force behind Google Search.

The contribution of this paper to the existing literature is the application of NLP and sentiment analysis to geo-political news in order to evaluate and predict geo-political sentiment. We are the first to our knowledge who apply NLP and sentiment analysis to country classifications.

In the next section we give an overview about our system architecture and explain the data and methodology we use. Then we present the results obtained from our sentiment analysis. Following, we discuss the limitations of geo-political sentiment analysis and afterwards we provide an outlook about future developments. Finally, we conclude in the Conclusions section.

The Architecture

NLP Framework

Sentiment analysis of a news event among countries involves two steps: (1) filter country pairs, (2) analyze the sentiment of the news event. To perform text analysis we adapt FLAIR, state-of-the-art NLP framework. The flair framework provides access to numerous machine learning techniques, specifically deep learning models, for text analysis of news headlines, speech, and embeddings. Embeddings are important since they characterize the context in which a word should be understood. Standard methods produce different embeddings for the same word depending on its contextual usage. The string "Washington" for instance would be embedded differently depending on whether the context indicates this string to be a last name or a location. While shown to be highly powerful, especially in combination with classic word embeddings, such methods require an architecture in which the output states of a trained language model (LM) are concatenated with the output of the embedding layer, thus adding architectural complexity.

Classic approaches combine classic word embeddings with character-level features trained on task data (Ma and Hovy 2016, Lample et al. 2016). To accomplish this, they use a hierarchical learning architecture in which the output states of a character-level convolutional neural network (CNN) or recurrent neural network (RNN) are concatenated with the output of the embedding layer.

A deep learning model Peters et al. (2018) is used for named entity recognition (NER), which can identify person, location, and organization from a piece of text. A semantic similarity search is performed to identify affiliation of the recognized entities with their respective countries. Finally, a sentiment analysis deep learning model is used to estimate the sentiment of the news, and categorize it as positive or negative. The results of the two steps are integrated to learn about the relations among countries.

Data and Methodology

We take daily global news by utilizing newsapi.org. Due to limitations in news access we are limited to 100 news inquires per day. For our purpose we are want to focus on the 20 most geo-politically active countries in recent times. Therefore, we create our own G20-list of countries, which includes: USA, Russia, China, UK, Ukraine, Germany, India, Iran, Israel, Turkey, North Korea, South Korea, Japan, Australia, Saudi Arabia, France, Italy, Greece, Pakistan, and Indonesia.

The time frame we analyze spans from March 24, 2021, to May 23, 2021. In total, we analyze more than 7,000 news articles directly addressing the country pairs. Moreover, we analyze not only the news headline, but also the news description. This provides a much more accurate sentiment score because more words can be analyzed. The sentiment score ranges from 0 to 1. For negative sentiments we multiply the sentiment score with -1. Since it lies in the nature of political news articles that headlines and descriptions are very often similarly formulated, while BERT is trained on a variety of linguistics, the sentiment scores are very often very narrow to each other. To better distinguish the sentiment scores from each other, we perform a simple mapping using the 'tanh' function, which still returns values between 0 and 1, but stretches the values in such a way that they become more distinguishable while not changing their explanatory power.

It is important to note that the sentiment score we obtain does not reflect the severity or strength how good or bad a news is, but displays the probability that the particular news has a positive or negative sentiment. In that sense, we can derive the severity or strength of a certain news, assuming that severe events are mostly formulated in a very clear and strong way, while uncertain events or situations will most likely lead to very vague statements and news situations. Therefore we can interpret the obtained sentiment score not only as probability of its accuracy, but also as strength indicator for a positive or negative situation.

The methodology is chosen to follow a sequential process. The first step involves the filtering of news associated with a certain country pair, e.g., all news on a particular day between USA and Russia. Then this news is being scraped and analyzed by BERT to figure out the sentiment associated with this news. In case of multiple news on a day we take the average sentiment score. This value is then being mapped through the 'tanh' function. In order to generate a rolling time window we then take the average sentiment score over the past 30 days. We apply this methodology to all country pair combinations.

Results

Table 1. *Sentiment Scores for All Analyzed Country Pairs from 03/24/2021-05/23/2021*

		Positive			**Negative**			**Negative**
Country 1	Country 2	Sentiment	Country 1	Country 2	Sentiment	Country 1	Country 2	Sentiment
UK	USA	0.12	USA	Russia	-0.07	Israel	Turkey	-0.30
Germany	USA	0.40	USA	China	-0.18	Israel	North Korea	-0.76
India	USA	0.34	USA	Ukraine	-0.76	Israel	South Korea	-0.26
Israel	USA	0.42	USA	Iran	-0.22	Israel	Australia	0.00
Japan	USA	0.20	USA	Turkey	-0.76	Israel	Greece	-0.41
Australia	USA	0.73	USA	South Korea	-0.11	Israel	Pakistan	-0.58
France	USA	0.68	USA	Italy	-0.59	Israel	Indonesia	-0.63
Greece	USA	0.76	USA	Pakistan	-0.74	Turkey	Saudi Arabia	-0.58
Israel	Russia	0.06	USA	Indonesia	-0.76	Turkey	France	-0.27
South Korea	Russia	0.05	Russia	China	-0.30	Turkey	Greece	-0.30
Japan	Russia	0.12	Russia	UK	-0.19	Turkey	Indonesia	-0.11
France	Russia	0.09	Russia	Ukraine	-0.54	North Korea	South Korea	-0.39
Italy	Russia	0.30	Russia	Germany	-0.01	North Korea	Japan	-0.05
Greece	Russia	0.75	Russia	India	-0.20	North Korea	Australia	-0.76
Indonesia	Russia	0.72	Russia	Iran	-0.38	South Korea	Japan	-0.06
Italy	China	0.04	Russia	Turkey	-0.46	South Korea	Australia	-0.17
Greece	China	0.76	Russia	North Korea	-0.03	South Korea	Indonesia	-0.48
Germany	UK	0.10	Russia	Australia	-0.23	Japan	Australia	-0.25
Japan	UK	0.19	Russia	Saudi Arabia	-0.04	Japan	Pakistan	-0.71
Saudi Arabia	UK	0.31	Russia	Pakistan	-0.53	France	Italy	-0.08
France	UK	0.04	China	UK	-0.43	France	Greece	-0.06
Italy	UK	0.31	China	Ukraine	-0.17	France	Pakistan	-0.72
Australia	Ukraine	0.52	China	Germany	-0.12			
Italy	Ukraine	0.53	China	India	-0.15			
Greece	Ukraine	0.75	China	Iran	-0.07			

Indonesia	Ukraine	0.72	China	Israel	-0.01			
Iran	Germany	0.04	China	Turkey	-0.24			
Israel	Germany	0.11	China	North Korea	-0.36			
Turkey	Germany	0.35	China	South Korea	-0.02			
South Korea	Germany	0.73	China	Japan	-0.14			
Japan	Germany	0.48	China	Australia	-0.45			
Australia	Germany	0.42	China	Saudi Arabia	-0.19			
Saudi Arabia	Germany	0.75	China	France	-0.14			
Italy	Germany	0.11	China	Pakistan	-0.58			
Greece	Germany	0.04	China	Indonesia	-0.18			
Pakistan	Germany	0.75	UK	Ukraine	-0.24			
Turkey	India	0.13	UK	India	-0.22			
South Korea	India	0.17	UK	Iran	-0.63			
France	India	0.32	UK	Israel	-0.32			
Italy	Iran	0.75	UK	Turkey	-0.45			
Greece	Iran	0.75	UK	North Korea	-0.70			
Pakistan	Iran	0.19	UK	South Korea	-0.09			
Indonesia	Iran	0.67	UK	Australia	-0.07			
Japan	Israel	0.06	UK	Greece	-0.48			
Saudi Arabia	Israel	0.50	UK	Pakistan	-0.72			
France	Israel	0.05	UK	Indonesia	-0.70			
Italy	Israel	0.45	Ukraine	Germany	-0.48			
South Korea	Turkey	0.75	Ukraine	India	-0.71			
Japan	Turkey	0.75	Ukraine	Iran	-0.74			
Australia	Turkey	0.36	Ukraine	Israel	-0.69			
Italy	Turkey	0.33	Ukraine	Turkey	-0.55			
Pakistan	Turkey	0.58	Ukraine	France	-0.44			
France	North Korea	0.49	Germany	India	-0.05			
Pakistan	North Korea	0.17	Germany	France	-0.16			
Saudi Arabia	South Korea	0.75	India	Iran	-0.19			
France	South Korea	0.75	India	Israel	-0.04			
Italy	South Korea	0.75	India	North Korea	-0.48			
Saudi Arabia	Japan	0.06	India	Japan	-0.05			
France	Japan	0.44	India	Australia	-0.17			
Italy	Japan	0.70	India	Saudi Arabia	-0.19			
Indonesia	Japan	0.28	India	Italy	-0.14			
Saudi Arabia	Australia	0.75	India	Greece	-0.19			
France	Australia	0.31	India	Pakistan	-0.26			
Italy	Australia	0.75	India	Indonesia	-0.25			
Greece	Australia	0.76	Iran	Israel	-0.55			
Pakistan	Australia	0.74	Iran	Turkey	-0.70			
Indonesia	Australia	0.32	Iran	North Korea	-0.60			
France	Saudi Arabia	0.75	Iran	South Korea	-0.41			
Italy	Saudi Arabia	0.75	Iran	Japan	-0.60			
Greece	Saudi Arabia	0.19	Iran	Australia	-0.18			
Pakistan	Saudi Arabia	0.21	Iran	Saudi Arabia	-0.40			
Greece	Italy	0.18	Iran	France	-0.41			
Indonesia	Pakistan	0.72						

Figure 1. *30 Day Moving Average Sentiment Score between USA and Russia from 04/23/2021-05/23/2021*

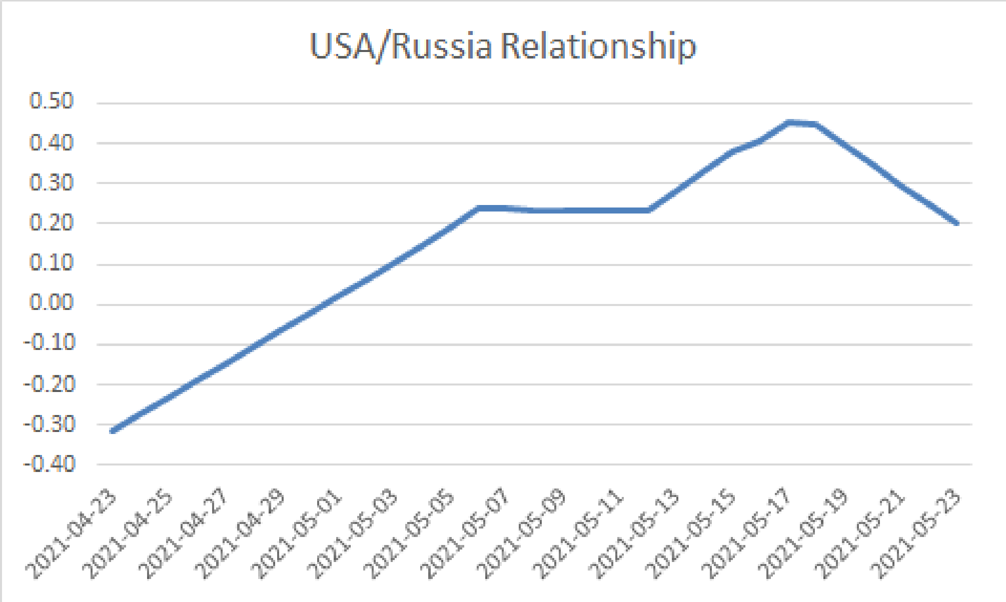


Figure 2. *30 Day Moving Average Sentiment Score between USA and China from 04/23/2021-05/23/2021*

Figure 3. *30 Day Moving Average of Sentiment Score between USA and Germany from 04/23/2021-05/23/2021*

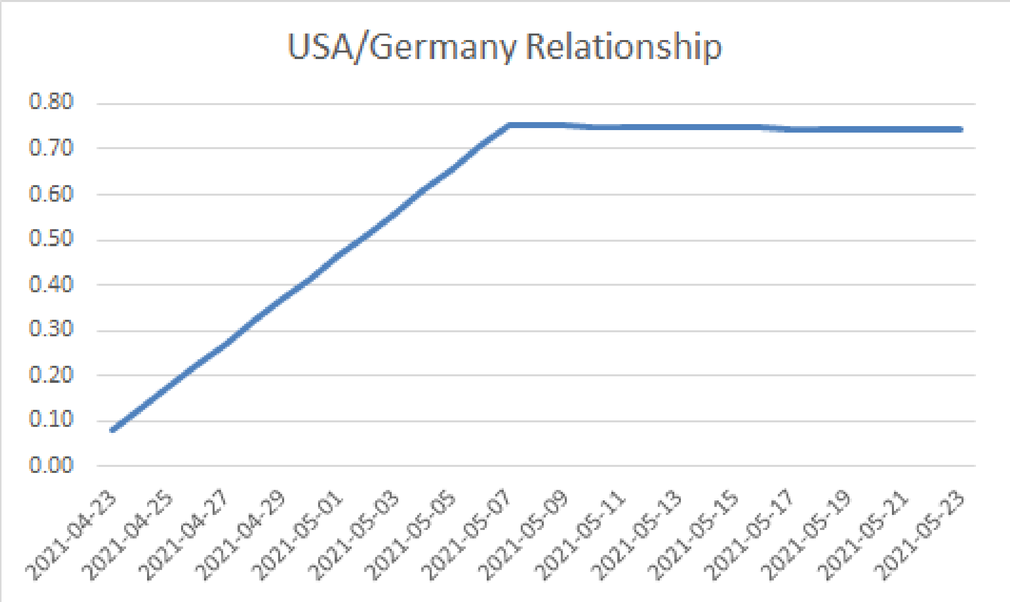


Figure 4. *30 Day Moving Average Sentiment Score between Turkey and Greece from 04/23/2021-05/23/2021*

Figure 5. *30 Day Moving Average Sentiment Score between Russia and Ukraine from 04/23/2021-05/23/2021*

Figure 6. *30 Day Moving Average Sentiment Score between Russia and Saudi Arabia from 04/23/2021-05/23/2021*

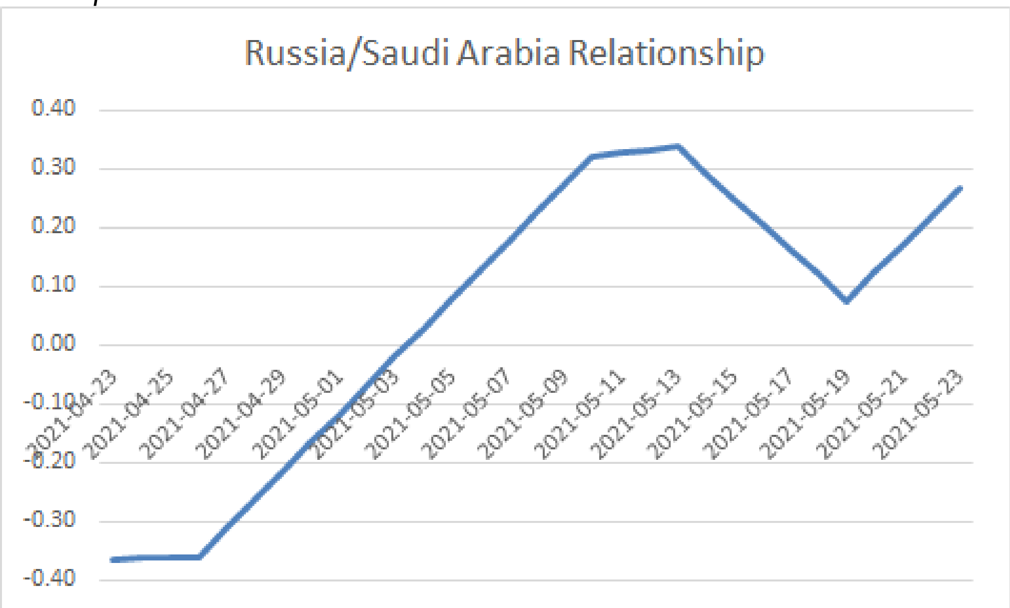


Figure 7. *30 Day Moving Average Sentiment Score between Iran and Israel from 04/23/2021-05/23/2021.*

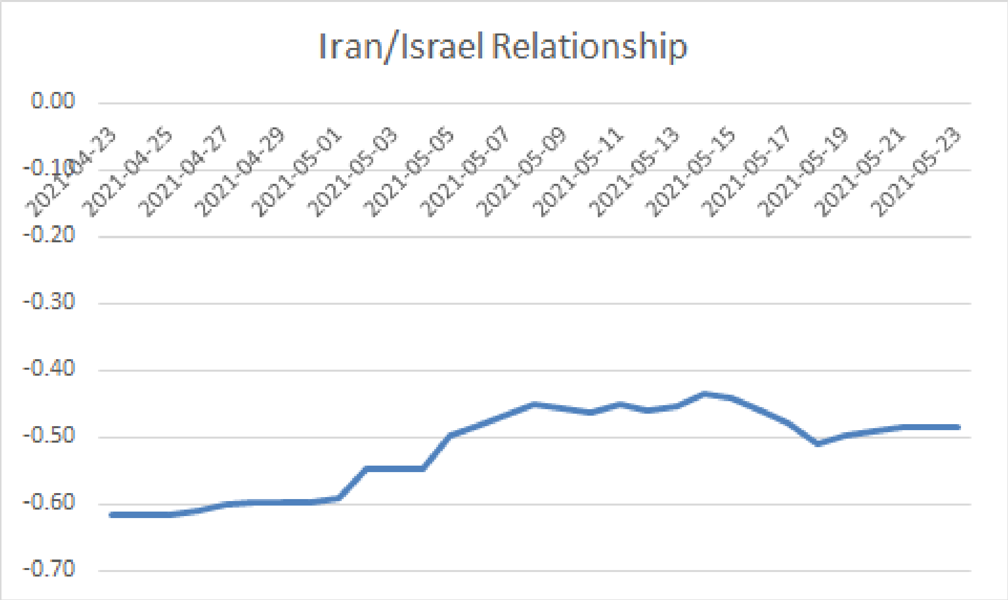


Figure 8. *30 Day Moving Average Sentiment Score between Iran and Saudi Arabia from 04/23/2021-05/23/2021*

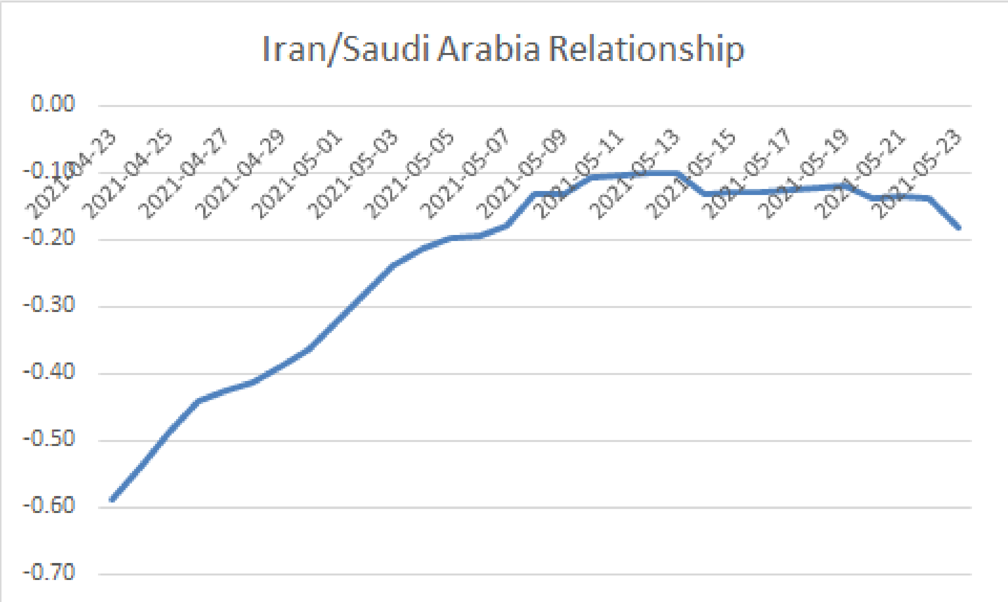
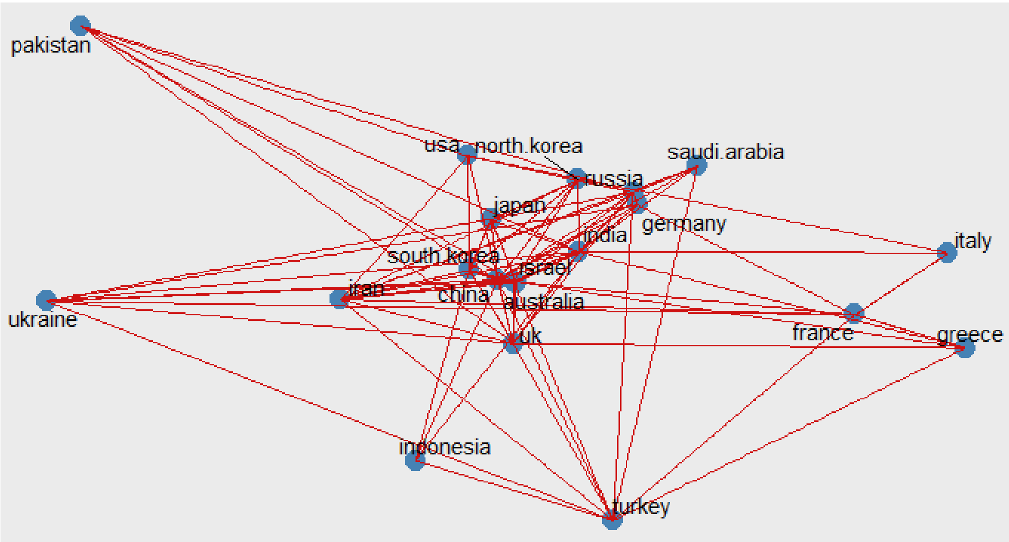


Table 1 reports the average sentiment scores for all 20 country combinations for the time period between 03/24/2021-05/23/2021. We can see that the overall sentiment across all news concerning the relationship between all country pairs was 59.32% more negative than positive. Some of the news were close to a value of 0, indicating a high uncertainty of being correctly interpreted. However, a closer analysis of the news being scraped and processed reveals the high complexity of entity recognition and correctly associating the sentiment score with the relevant entities. The high uncertainty can be mainly attributed to the complexity of semantic of language.

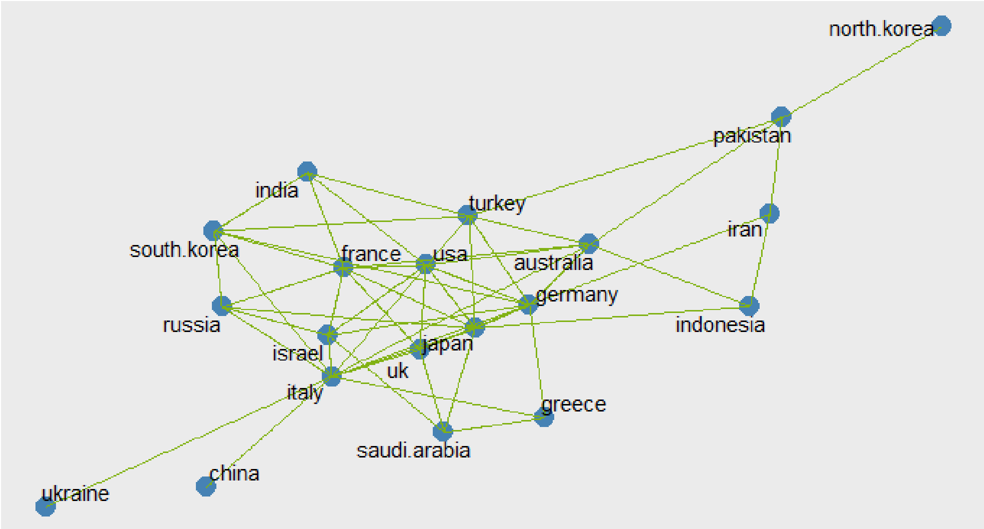
Figures 1-8 show the 30 day moving average development of the sentiment score between some randomly selected country pairs. We can see that for certain country pair combinations, such as Russia/Ukraine, Iran/Israel, and USA/China, the system seems to capture the correct sentiment and to reflect reality. For other country pair combinations, the sentiment is subject to a much higher uncertainty, meaning that the development of the sentiment between these country pair combinations can't be fully trusted, e.g., between USA/Germany and Russia/Saudi Arabia. One of the main reasons, which can be taken from the graphs, is the mostly linear development of the sentiment scores, which indicates that not much news was being analyzed, making the prevailing sentiment score more vulnerable to wrongly interpreted news. Nevertheless, most developments seem to reflect reality in terms of catching the positive or negative sentiment. It is very important to note again that the values do not reflect the actual condition of the sentiment between two countries, but reflect the probability, or uncertainty, that the sentiment is either positive or negative.

Figure 9. *Network Diagram of Negative Sentiment Relationships from 04/23/2021 -05/23/2021*



Note: Countries close to each other have a lower magnitude of negative sentiment.

Figure 10. *Network Diagram of Positive Sentiment Relationships from 04/23/ 2021-05/23/2021*



Note: Countries close to each other have a higher magnitude of positive sentiment.

Figures 9 and 10 show the network diagram of all negative, respectively positive sentiments between countries for our testing period from 04/23/2021-05/23/2021. The countries with the closest distance to each other are of higher sentiment than the country pairs which are further apart. So, countries in close neighborhood are in better relationship to each other than countries with a greater distance. We can see that network diagrams are very useful to visualize the relative distance, and therefore the status of the international sentiment relation between countries. It is important to note that only country pairs which are directly connected with each other care subject of measurement. For our analyzed time period we find the relationships between e.g., Pakistan/Russia, Pakistan/USA, or Turkey/Greece to be worse than the relationships between e.g., USA/Russia, Japan/Russia, or China/USA. At the same time we can see that the relationships between e.g. France/USA, USA/Australia, or USA/Japan are better than between e.g., Ukraine/Italy, Japan/Indonesia, or Turkey/Pakistan.

Limitations

General limitations of this analysis are the constraint data sources and reduced ability of entity recognition, as well as linguistic barriers coming along with the interpretation of news headlines. Improvements of the quality of the results can definitely be achieved by inclusion of more news sources and APIs, as well as the integration of more countries. Moreover, by recognition of political players such as politicians and institutions, the information basis can also be extended. We also tried to include entity name recognition, but due to linguistic barriers, such as in-direct speech, irony, special characters, multiple entities, etc., the sentiment score was rather getting more distorted than improved. One key element in improving international sentiment relationship measurement is the quantity of analyzed news sources. Due to the high degree of noise in news headlines and complexity of language, the degree of imperfection in news interpretation is rather high. Therefore, a high number of news inputs will smooth out the random noise effect and provide rather accurate trends and relative sentiment differences, following the logic of the law of large numbers.

To provide an example: The relationship between USA and Russia was negatively determined with a sentiment score of -0.5621 on 04/03/2021 by the following news: "USA vs. Russia: Who would have won a Cold War naval conflict?", by Robert Farley on nationalinterest.org. We can see that this news was certainly not related to an active development, but rather reflected the prevailing negative sentiment be- tween the two countries. However, to determine the current sentiment, we would need more news. The problem is that by broadening the news basis dilutes the relevance of the posts, while a stricter filtering reduces the number of potentially analyzed articles, which will again have a negative effect on the accuracy of deter- mining the current sentiment score. Nevertheless, a more accurate filtering seems to dominate the need of increasing the quantity of news.

However, there are certainly single news headlines which are being analyzed correctly, e.g., on 05/02/2021, the system returned a negative sentiment score of -0.6884 between Russia and Ukraine by the following news: "How Russia tested power grid attacks in Ukraine", published by CBS news. The question in determining the accuracy of the current sentiment between two countries depends therefore very heavily on the ratio between correct, or relevant news to incorrect, or irrelevant news.

Future Developments

NLP is not a new concept for various kinds of applications. However, in the sphere of international relations, it hasn't been well established yet due to its technical limitations. Nevertheless, the technological development is progressing at an unprecedented level and can be expected to increase at an exponential rate. Recent developments allow insights in future developments, and thus, what role NLP will play in the future development of international relations itself.

Interestingly, Schrodt (1991) already mentioned in 1991 that the AI/ International Relations (IR) community is characterized by a healthy level of internal debate. He lays out some perspective on the concepts used in AI/IR. O'Connor et al. (2013) describe a new probabilistic model for extracting events between major political actors from news corpora. They recover expert-assigned event class valences, and detect real-world conflicts to evaluate the model's performance on political science benchmarks. Their research shows that the direction NLP might be heading towards will be an Intelligence Augmentation-related one. In contrast to Artificial Intelligence works to augment human intelligence and support it in their decision-making functions. This means that while NLP will become more and more powerful in its speed and capacities to capture the content and meaning of published news, statements, and political texts, human experts will always need to be at the last stage to interpret and set the insights gained from NLP into policies or action. This is at the same time not only a sufficient but also a necessary condition due to security reasons as it might not be in humanity's interest if countries' defense systems react to a machine-read and interpreted news article without human supervision.

Another future development of NLP concerning International Relations is the sphere of pattern-based biomedical relation extraction systems. After the occurrence of a world-wide pandemic, the necessity of global communication and inter- action got in the focus of countries' governments and authorities. NLP will play a crucial role in developing faster response systems and promotion of international co- operation, not only in the bio-medical sphere, but also in the coordination of global policies. Peng et al. (2014) already provided a novel framework to facilitate the development of a pattern-based biomedical relation extraction system which aims to identify designated relations among biological entities reported in literature. Such kind of functionality will be brought up to a level such that it can be incorporated into direct policy coordination for inter-governmental responses.

Moreover, NLP has a huge potential to capture trends and deduct future impacts of social developments such as preferences, concerns, and latent wishes of society. Already nowadays, bots and other tools are able to communicate with humans and process their feedback and reactions. This data is then being used to optimize search profiles, detect and reveal social networks, and utilize it for commercial or national defense purposes. Wiedemann (2016) lays out potential text Mining applications for Qualitative Data Analysis in the Social Sciences. In International Relations, such detection of network profiles will be made possible by progressing NLP technology, using machine learning algorithms, smart technology and real time interactive surveillance and response protocols, such as enabled by 5G.

In general, future technological advances will enable governments and other political institutions to utilize NLP and text mining in such a way that the management of the political process will be optimized.

Conclusion

In this paper we present a deep learning neural network NLP approach for analyzing geo-political news and building a relational network structure to interpret and measure international relations. Our approach is to search for 20x20 country pair-related news through newsapi.org and then utilize BERT, the NLP engine used to optimize google search, to retrieve the sentiment between each country pair. We track each day's sentiment score for all country pair combinations and aggregate it for all country pairs over a 30-day rolling time window for 30 days.

We find that measuring international relations through NLP has a huge potential as already simple entity classifications and sentiment measurements seem to capture the real sentiment of international relations. The question is of course, what the real sentiment is. There is no precise answer to this question, but observers of international relations have a sense if certain relationships are more positive or negative, and which development each relationship takes. Besides very useful information that can be extracted, such as the relation between positive to negative news world-wide, or country-pair specific, the certainty of the sentiment and the relative level of sentiment compared to other countries can also be revealed by construction of a network structure. The implications of such insights are very useful to any player in international or domestic politics, as certain trends can be made visible, and the impact of possible policy or event scenarios can be measured. Despite the impressive realistic mapping of international scenarios using a very limited news access, certain general limitations could be detected such as the trade-off between broadening the news basis and narrowing it down. Broadening the news basis allows the system to analyze more news. Due to the high error rate in measuring single news headlines, due to linguistic and semantic reasons, as well as the high complexity of detecting the relationship between multiple entities mentioned in one and the same news headline, a high rate of incoming news is necessary to capture the correct trend. Contrary, with increasing the inflow of scraped news headlines, the degree of noise increases proportionally, meaning that more irrelevant news is being processed, which can significantly skew the sentiment score.

Therefore, our key finding is that the driver of quality of international relationship sentiment forecasting is the relationship between quantity and quality of scraped news. While the increase of quantity of scraped news increases the probability for detecting the correct trend, it adds noise at the same time. This calls for an increase in quality of scraped news. But an increase in the quality of news automatically reduces the quantity of relevant news and makes therefore the sentiment score more vulnerable to changes in semantics. The most accurate real mapping of the sentiment score at a time t of international relations can thus be defined by

$$SC_t = max_Q\, min_N\, (Q_i, N_i), \qquad (1)$$

where Q_i is the quantity of news being analyzed for country pair i, N_i is the noise generated by the news of country pair i, measured by its relevance.

References

Akbik A, Bergmann T, Blythe D, Rasul K, Schweter S, Vollgraf R (2019) FLAIR: an easy-to-use framework for state-of-the-art NLP. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, 54–59.

Alammar J (2019, November 26) *A visual guide to using BERT for the first time*. Retrieved from: https://jalammar.github.io/a-visual-guide-to-using-bert-for- the-first-time/. [Accessed 25 May 2021]

Allison Graham T (1971) *Essence of decision, explaining the Cuban missile crisis*. 12th Printing. Harvard University.

Burdack M, Roessle M, Kuebler R (2018) A concept of an in-memory database for IoT sensor data. *Athens Journal of Sciences* 5(4): 355–374.

Cimiano P, Voelker J (2005) Towards large-scale, open-domain and ontology-based named entity classification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*.

Collins M, Singer Y (1999) Unsupervised models for named entity classification. In *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*.

Fehlmann T, Kranich F (2019) Testing artificial intelligence by customers' needs. *Athens Journal of Sciences* 6(4): 265–286.

Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C (2016) *Neural architectures for named entity recognition*. In arXiv preprint arXiv:1603.01360.

Ma X, Hovy E (2016) *End-to-end sequence labeling via bi-directional LSTMs-CNNsCRF*. In arXiv preprint arXiv:1603.01354

Niu C, Li W, Ding J, Srihari RK (2003) A bootstrapping approach to named entity classification using successive learners. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 335-342.

O'Connor B, Steward BM, Smith NA (2013) Learning to extract international relations from political context. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 1094-1104. Sofia, Bulgaria, 4-9 August 2013.

Peng Y, Torii M, Wu CH, Vijay-Shanker K (2014) A generalizable NLP framework for fast development of pattern-based biomedical relation extraction systems. *BMC Bioinformatics* 15(Aug): 285.

Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. (2018) *Deep contextualized word representations*. In arXiv preprint arXiv:1802.05365.

Razek A (2021) Pragmatic association of the two evaluation concepts of operational observation and mathematical modeling. *Athens Journal of Sciences* 8(1): 23–36.

Schrodt PA (1991) *Artificial intelligence and international relations: an overview, artificial intelligence and international politics*. 1st Edition. Routledge.

Wiedemann G (2016) Introduction: qualitative data analysis in a digital world. In *Text Mining for Qualitative Data Analysis in the Social Sciences*. Kritische Studien zur Demokratie. Wiesbaden: Springer VS.

Yogatama D, Gillick D, Lazic N (2015) Embedding methods for fine grained entity type classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, volume 2: Short Papers, 291–296.

Zhu X, Goldberg AB (2009) Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 3(1): 1–130.