ZREC architecture for textual sentiment analysis

Martin Pavlíček, Tomáš Filip, and Petr Sosík

Institute of Computer Science - Faculty of Philosophy and Science - Silesian University in Opava
martin.pavlicek@fpf.slu.cz

Abstract: We present recent results of the research project ZREC aimed at psycho-social phenomena (group polarization, belief echo chamber and confirmatory bias) analysis based on bio-inspired computing methods. We present two updated pipeline solutions to work with bio inspired AI methods and data gathering tools integrated in a complex (but simple to implement) vertical information system. The scope of the investigated phenomena is reduced to the aspect based sentiment analysis with an integration of methods covering named entity recognition and relation extraction. We present a simple ontology addition to group polarization in the last year due to COVID pandemic and stress the importance of project in the social and IT sphere and multi-tier cooperation. We also provide introductory results based on test data using several deep learning architectures and demonstrating that the presented approach is robust and functional.

1 Introduction

In the recent years we can see dramatic increase in interaction between individuals and groups in cyberspace [7] together with news dissemination [2] and real time reporting, as well as increasingly polarized groups presenting their narrative and beliefs [13] in the cyberspace.

We can also see processes of regulation [12] [3] and specific narrative information enforcement, which are not only due to the novel COVID situation worldwide. Together with cybersecurity, national interests are aligned with acceptation of information as weapons and information warfare battlefield [5] [6] [4] [10].

These premises motivate us to investigate and build tools to understand the flow of information in cyberspace in a more open and rigorous manner. To keep the project manageable, we restrict our investigation to information about event exposures and specific sentiment reactions (positive, negative, neutral) which rise in an individual and which can be traced to a group behavior. We focus on three phenomena – group polarization [1], belief echo chamber [18] [22] and confirmatory bias [21]. Besides the interaction we monitor world events through the GDELT dataset which is viewed as a trigger of sentiment response.

The goal is to investigate these phenomena and maintain an open system ZREC (`www.zrec.org`) and its cornerstones – algorithms, research community and methods which can be used for further work both in the scope of IT and in an applied research. We focus on understanding these phenomena within specific ecosystem - nation, language, a selected group of sources and other parameters. In a simple way, we can analyze approval or disapproval with world events which occurred as information in cyberspace or within interaction of individuals who act on the surface Internet.

The paper is organized as follows: in the next section, we describe a novel project architecture based on pipelined tasks. Data pre-processing phase is described in Section 3. Section 4 presents details of the key project component - aspect based sentiment analysis, and experimental result we have obtained with our architecture using three different deep learning models. The two last sections contain discussion and conclusions.

2 Project description

2.1 Pipeline

In a pipeline view of the system we introduce two pipeline solutions which cover both data and the AI methodology integration model. This division is needed to track changes, to track learning data and their ability to create a narrative bias and to share these metadata within developers community.

The first pipeline covers the implementation and training of ML methods for the NLP analysis. In this pipeline we store and train specific models of our live data and we also store pre-trained models and analyze the results. At any time we can access a specific version of the model together with specified data which can provide feedback and a possible rollback in the system's development.

The second pipeline focuses on data gathering, cleanup and storage. To exploit different sources and different social networks like Facebook, GAB, Twitter, Parler and others, we maintain a set of tools which are used to gather data from predefined sources within a defined algorithm. The data are cleaned, meta-annotated and stored in the system.

Further work with the data is possible within the common batch analysis framework (described bellow) which is available to the users (Figure 1).

2.2 Architecture

We can describe the state of the system as a scalable vertical architecture which has emerged from the initial phase.

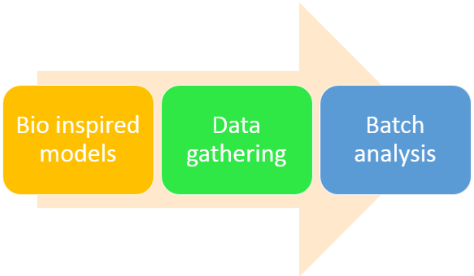

Figure 1: A pipeline view of the system architecture with an optional batch analysis

In the scope of technology, we work with scripting languages for creating the application part of the system, relational (SQL) and graph databases are used to store the data and to provide the basic architecture. For presentation we use the concept of web information system and use a library of visual front-end framework to simply present the front-end of the system to end users.

Our goal is to create a complex yet relatively simply implementable system (Figure 2). The architecture can be divided into two parts. The first part is an administrative and methodical system. The second part is the data part combined with AI methods. The key components of the system represent data collection and tagging, NLP methods and dataset warehouse, group and individual ontology graphs, common system analytical tasks scheduler.

Bio inspired methods training ground is used to store specific (mostly deep learning) AI methods [33, 30], with pre-selected training data and specific iteration of pre-trained methods as an essential part of our system. This part of system gives us the ability to strongly support the integration of new bio inspired learning models for emergence of update both models and specific data which were used to train these models. From our experiments we see a strong trend to gain a specific bias when training our models on live data from certain sources. This is, e.g., the effect of echo chambers present in the sources we gather data from. The ability to snapshot model training data and model definition is essential.

Data gathering and tagging is a part of the system focusing on definition of selected sources and individuals, as well as selected methods and algorithms to gather the predefined text data. We focus on simple definition of selectors and the ability to self heal within error spaces.

For a survey of possible methods we refer the reader to, e.g. [11]. For instance we use Twint tools[1] - Twitter Intelligence Tool to collect data from Twitter using Python language and bypassing the need to use Twitter API. With the help of this tool we can select queries for specific users and specify the time period for which we want to collect all available data. Our gathered data includes posts, comments, and user interactions, including related metadata. The advantage of this tool is the ability to process data without using Twitter´s API.

Ontology is used as the main data structure to define groups and individuals. A comprehensive definition of ontology of captions is a strong tool to solve complex situation of similarity and anomaly detection. We use a relation database to store a predefined a specific static ontology of captions transformed into graph network [16] which is then used for computational purposes.

Batch analysis defines framework of methods of analysis in the system. The system is built to handle multiple tasks from multiple users on multiple data sources. Batch analysis provides a robust system of common analytical queries which can be used as a simple batch scheduler. This definition of tasks gives us the ability to store specific combinations of data, users and methods which altogether control the analysis. In the user scenario this gives us the ability to cache and speed-up processes and to have a pool of results which can be used for further comparison and cross-check.

Information system core is the meta programming language we use to build the system. Base of the information system has the ability to render data pages, to check global and parametric permissions, to define users and their roles and their history. The core gives us an ability to tweak the system, to view it with permissions of other roles and users, and to give a transparent model of accessing all the data and all subsystems.

Specific data module interrelates data sources and events gathered from the surface Internet. Information about events are obtained through the GDELT[2] dataset in a CAMEO format. Further specific datasets (textual and numerical) are being integrated to the system - currently the storage of COVID cases from authoritative sources (Johns Hopkins University[3]).

Translation module define roles of translators which can access the system and proceed with translation from/to different languages, increasing system's accessibility.

3 Data pre-processing

In this section we describe a series of recent known methods for text feature extraction which are (or will be) used in our architecture to pre-process input data for the experiments described in the next section.

3.1 Creating a dataset

It is necessary to label the collected data for further processing. Manual data labeling is time consuming. Tomáš Mikolov et al. (2013) [17] introduced a method Word2Vec

[1] https://github.com/twintproject/twint

[2] https://www.gdeltproject.org/

[3] https://github.com/owid/covid-19-data/tree/master/public/datag

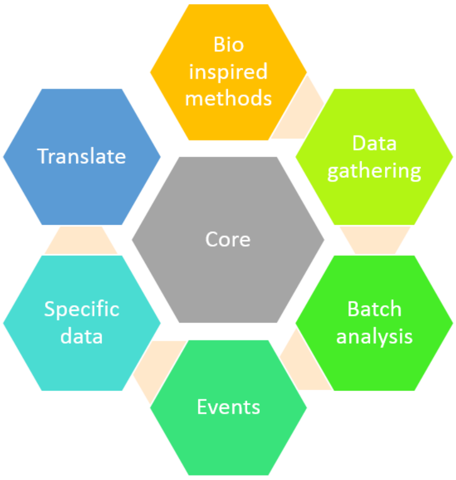

Figure 2: Architecture of the system

which project the word into a multidimensional feature vector. This projection allows us to use vector algebra tools to measure the distance between words. If we are able to determine how semantically similar individual words are, we can use this technique to measure the relevance of texts. One such tool for document similarity metric is Word Mover´s Distance (WMD Kusner et al. 2015) [14]. WMD finds the minimum distance to transport all words from a source document to a destination document. Because this method uses pre-trained embedding, WMD allows us to find a relationship between texts that do not share same words but has a similar meaning. Relaxed word mover´s [14] distance further reduces the time consuming of WMD from $O(p^3 logp)$ to $O(p^2)$, where p denotes the number of unique words in the texts. This technique allows to find the most relevant texts for the given query and thus to streamline the process of creation of a training dataset.

3.2 Named Entity Recognition (NER)

One of the essential functions of natural text processing models is to correctly predict name entities and the relationships between them. This capability is important for tasks that use named entities such as Question Answering (QA) or entity Relation Extraction (RE). Models handling contextual information have brought significant improvement for NER. Yamada et al. (2020) [31] added the entity-aware self-attention mechanism and entity type embedding to its model. He also added a pre-training task where he replaced a certain number of entities with a special $\langle$MASK$\rangle$ token in order to predict these entities. This model has achieved the most accurate results on tasks working with entities: NER, relation classification and entity typing. Wang et al. (2021) [27] used a search engine to find texts semantically similar to the input text. To evaluate similar texts they used BertScore (Zhang et al. 2020) [35], which measures cosine similarity between tokens of the given texts. A concatenated input document and documents returned from search engine are used together to train the model. The assumption is that both outputted distributions should be similar. This is done by updating loss function. This model achieved the highest score on 8 different NER datasets from different domains.

3.3 Relation Extraction

Apart from the NER, another important task for text comprehension is to classify the relationship between entities. Xu et al. (2021) [29] added the Structured Self-Attention Network (SSAN) to the Transformer deep learning architecture. The SSAN model incorporates the Biaffine Transformation or Decomposed Linear Transformation which creates the structure $S_{i,j}$. This structure represents the connection between words w_i and w_j and makes it possible to classify the type of link between entities and discover coreference structures. Wadden et al. (2019) [26] introduced the multi-task framework DYGIE++, for three tasks of information extraction: RE, NER and event extraction. The basis is a pre-trained NLP model. Its outputs are sent to the graph propagation module. It then modifies the representation by integrating the current representation with previous representations using the gating function. The resulting predictions are obtained from the re-contextualized representation using a scoring function. It contains two feedforward neural nets (FFNN). The final outputs are equal to $FFNN(g_i)$ for NER and $FFNN([g_i g_j])$ for RE, where g_i and g_j are the representations for span i and j. A different approach was used by Zhang et al. (2021) [34] where they applied U-Net (Ronneberger et al. 2015) [20] model known from computer vision to find global relationships between entities. First, they created an entity-level relation matrix. Entity similarity was calculated using similarity-based method (concatenating cosine similarity, elementwise similarity and bilinear similarity) or context based method (entity-aware attention). The feature vectors form a matrix $M^{i \times j \times d}$, where i and j indicate a relation between i-th and j-th entity, d is the size of feature vector. This matrix is put to the U-Net model where d serves as a feature channel. The resulting relational type probability are obtained using feedforward network, entity pair embedding and output from the U-net model.

4 Experimental results

4.1 Aspect Based Sentiment Analysis (ABSA)

ABSA is a method for classifying text polarity. In contrast to aspect analysis, it makes it possible to determine sentiment in a fine-grained detail. The analyzed document may be related to several independent aspects and each of these aspects may have different sentiment. Thus ABSA can be divided into two separate tasks. First, finding all aspects which occur in the sentence. Second, predict sentiment to each aspects.

Various methods have been proposed to solve this task. One of the classic solutions is the formation of a dependency tree. Devlin et al. (2019) [9] introduced the BERT model built on the Transformer architecture (Vaswani et al. 2017) [25]. BERT was created to capture the right and left context of a word and it was used as a backbone in many ABSA models. The BERT was pre-trained to predict tokens in the sentences that were artificially corrupted. Some randomly selected words from sentence were replaces by special $\langle$MASK$\rangle$ token. A disadvantage of this pre-training task is the loss of the context between masked words.

This problem is solved by the XLNet model (Yang et al. 2020) [32] learning contextual information from all permutations of the factorization order. This method ensures that contextual information from all possible positions of the right and left context are used.

Liu et al. (2019) [15] introduced the Robustly optimized BERT approach (RoBERTa), which has been pre-trained on a more robust data corpus than BERT using larger batch sizes. However, the pre-training tasks do not directly incorporate text sentiment determination.

Tian et al. (2020) [24] introduced the self-supervised SKEP method for pre-training the BERT model. Instead of randomly selected words as in BERT, words related to sentiment or aspects are selected for replacement with the $\langle$MASK$\rangle$ token. The model predicts the words polarity and the masked sentiment words. Models pre-trained using this method achieved better performance than baseline models.

Dai et al. (2021) [8] used the Pertubed masking method which searches for syntactic connections in a pre-trained BERT model to create an induced tree.

Finally, Sun et al. (2019) [23] used two different inputs for the pre-trained BERT model. The first input is a sentence from the dataset and the second input is an auxiliary sentence. The auxiliary sentence contains the target and the aspect. Using these two inputs, the model predicts the resulting polarity. This method transforms the ABSA task to a QA task.

4.2 Experiments with BERT, XLNet and RoBERTa

To evaluate the capabilities of our pipeline architecture, we performed a series of experiments based on the test dataset Sentihood which is publicly available at `https://github.com/uclnlp/jack/tree/master/data/sentihood` as a part of the project Jack the Reader (JACK) [28]. The Sentihood dataset contains opinions about living in various locations in London, UK. In particular, there are 2480 training samples (opinions) with positive sentiment and 921 with negative sentiment, i.e., 3401 in total. Instead of processing the whole ABSA pipeline, we used a predefined subset of aspects which we wanted to predict in the collected data and we created an appropriate set of auxiliary sentences. The disadvantage of using auxiliary sentence for predicting polarities is the need for repeated predictions for each aspect.

Model	Parameters	Precision	Recall	AUC
BERT	5 701 889	0.9996	0.9504	0.9918
XLNet	6 368 001	0.9996	0.9460	0.9939
RoBERTa	5 701 889	0.9987	0.9383	0.9917

Table 1: Training results of the ML models BERT, XLNet and RoBERTa on the Sentihood dataset with targeted auxiliary sentences.

In our experiments we tested our architecture with the pre-trained BERT, XLNet and RoBERTa deep learning models. Hyperparameters of the models were set as follows: No. of training epochs 150–200, batch size 48, learning rate 1e-5, optimizer: Adam Weight Decay. The numerical scores of training of the three models is summarized in Table 1, where AUC stands for the Area Under Curve Score.

Graphical comparison of results of the three models is presented at Fig. 3 and 4. We can conclude that all three models provided rather impressive results and that the textual analysis in our ZREC architecture proves applicable to real world data which we are now collecting.

5 Discussion

5.1 Retrospective look

The project ZREC defines two areas of importance – for society and for IT. Both can be achieved by creating an open distributed ecosystem which can be used to understand emerging phenomena. This is now even more important as in the last year we saw a world transforming via COVID restrictions, and so the need to understand cyberspace phenomena and their influence on society is still more urgent as the communication is moving to cyberspace. We see this like a clear trend and motivation for the project.

The IT research side is more profound since we want to develop, integrate and implement state-of-the-art AI methods aimed at natural language understanding in specific areas. Hence the function of the project as a strongly defined sandbox which is an integration tool for plethora of specific methods from the NLP filed is both effective and promising.

5.2 Trends and main ontology themes

In our previous publication [19] we defined a main ontology based on basic polarization which defined entities in our information ecosystem, like sentiment towards: Czech Republic, United States, Russia, Israel, Ukraine, political figures from the United States, Russia and also Czech Republic, intelligence agencies like CIA, FSB, GRU, BIS

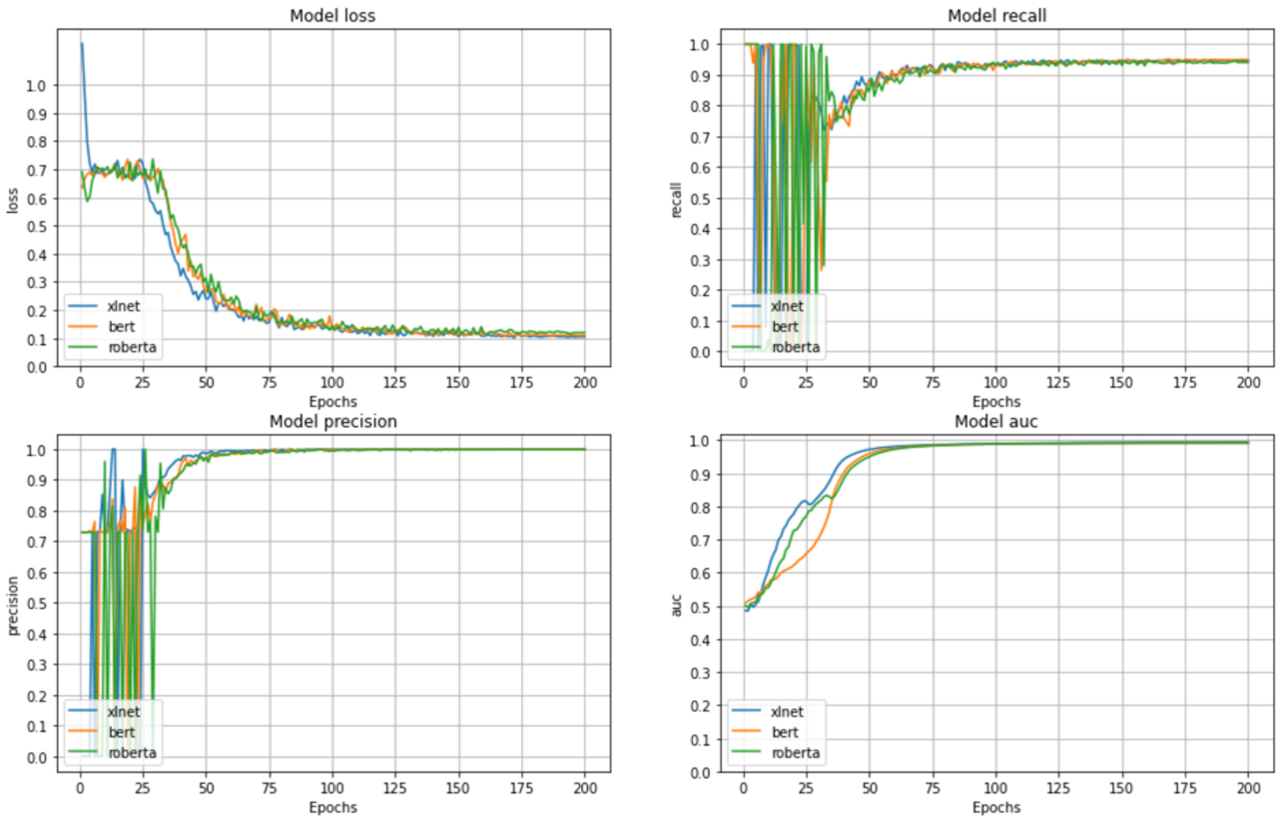

Figure 3: Training progress of the ML models BERT, XLNet and RoBERTa on the Sentihood dataset during 200 training epochs.

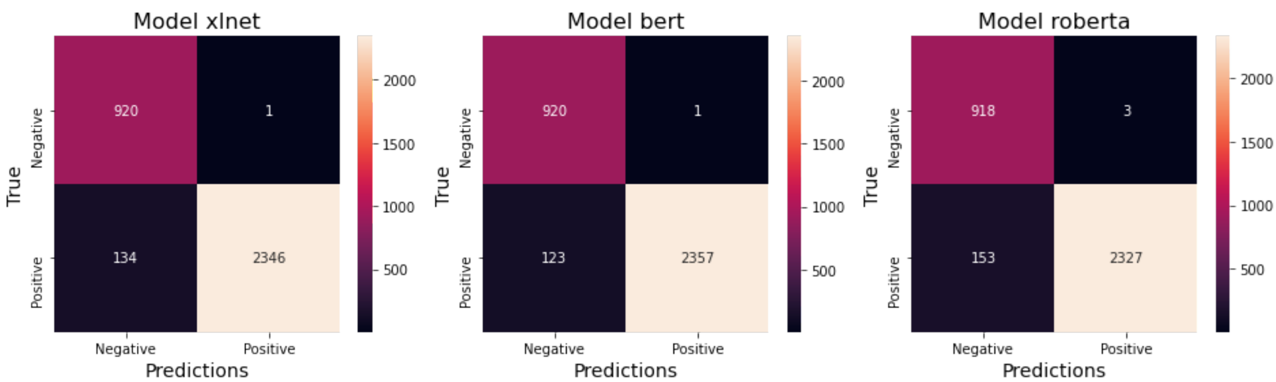

Figure 4: Prediction success of the ML models BERT, XLNet and RoBERTa on the Sentihood dataset.

and others. This ontology together with the used sources is a key factor in creating an individual or group profile.

A new communication topic with an enormous socio-economic impact and an adequate amount of hoax and fake news has emerged: vaccination, COVID restrictions and COVID pandemic acknowledgment. These topics are (together with topics covering national security and politics) in the center of interaction covering basic events emerging in the cyberspace. With our modular architecture we can continue to follow individual and group responses and polarizations based on the interaction in the field of vaccination narrative with just an addition of new terms to our existing ontology. In accordance, we added to our ontology sentiment to specific vaccines (Pfizer, Moderna, Astra Zeneca, Sputnik, NovaVax), specific medical terms like SARS, Spike-protein, RNA, sentiment towards the efficiency and need of vaccination.

5.3 Industry and research feedback

Our system is not scaled for harvesting all available data on social networks and surface Internet. We stress that we focus on specific datasets and specific ecosystems that are used like a main observation point for the phenomena we model and try to understand. To be more specific we find a value in a transparent definition of dataset and sources description – both in the system and internally within research community. We thus see the system ZREC also as a tool presenting some basic methodologies to select and describe sources which are used to get data.

We still assume the creation of a universal AI crawler which can process data collections from various sources as very important, but in the core development we focus more on the creation of the NLP AI pipeline which can be used to understand the phenomena.

We expect that our project would benefit from multi-tier cooperation with research centers, universities and industry partners. This is confirmed by the response of potential benefiters, and we use the academic space also as a call for a join initiative incorporating people, IT resources and internal information and ecosystem knowledge.

5.4 Progress and upcoming tasks

The ZREC system is being developed under the SCRUM methodology. The complexity of the development was reduced due the clustering of the system into mentioned subsystems. An efficient way of dealing with data and models was the introduction of the two pipeline solutions providing an open tool set.

Incorporation of AI models suitable for NLP tasks is human-intensive within the scope of acquiring state of the art ideas, and the NLP training is demanding also in IT resources. Due to this fact we focus on the integration of the ontology based solution with prepared data, which can be used as best cost effective way to achieve results. As a next step we will focus on development and incorporation of new self-pretraining methods specifically designed for sentiment classification. Promising solutions for ABSA can be based on auxiliary sentences and attention model usage.

6 Conclusion

We have presented an updated ZREC project (`www.zrec.org`) whose aim is the analysis of psycho-social phenomena (group polarization, belief echo chamber and confirmatory bias) in the surface Internet. These phenomena are analyzed in the context of reactions (positive, negative) to information about local and world events. Our primary sources are social networks, and discussions and comment boards within web pages. A part of the project focuses on analysis, visualization and dissemination of information about events at the surface Internet.

We have also presented a novel architecture in the scheme of two pipeline solutions. The first pipeline covers AI methods used for NLP tasks, training and data management. The second pipeline covers data gathering, storage, cleaning and simple meta-annotation. Main tasks run in a batch mode via an open analytic toolbox. First experimental results based on test dataset Sentihood proved efficiency of our architecture which is now prepared to process larger-scale datasets acquired from Internet.

Our recent research focuses on the task of aspect based sentiment analysis (ABSA). We see a clear promise in building a strong ontology of entities and relation which can detect both standard narratives related to key topics (national security, politics, COVID. . .) and anomalies.

Further research work is seen mainly in the development and implementation of new ABSA methods, and in definition of new data transformation into multi-dimensional spaces allowing for their better understanding. Finally, the crucial step is the data acquisition focusing on current active narratives in cyberspace which are in the center of our studies.

Acknowledgements

The research was supported by the Silesian University in Opava under the Student Funding Scheme, project SGS/9/2019, by the Student Grant Foundation - SGF/5/2020, and by European Union under European Structural and Investment Funds Operational Programme Research, Development and Education project “Zvýšení kvality vzdělávání na Slezské univerzitě v Opavě ve vazbě na potřeby Moravskoslezského kraje” CZ.02.2.69/0.0/0.0/18_058/0010238.

References

[1] C. A. BAIL ET AL., *Exposure to opposing views on social media can increase political polarization*, Proceedings of the National Academy of Sciences, 115 (2018), pp. 9216–9221.

[2] D. BAR-TAL, *Group beliefs: A conception for analyzing group structure, processes, and behavior*, Springer Science & Business Media, 2012.

[3] BBC, *Twitter hides Trump tweet for 'glorifying violence'*, 2020. `https://www.bbc.com/news/technology-52846679`.

[4] *Annual report of the security information service for 2016*, 2017. `https://www.bis.cz/public/site/bis.cz/content/vyrocni-zpravy/en/ar2016en.pdf`.

[5] *Annual report of the security information service for 2017*, 2018. `https://www.bis.cz/public/site/bis.cz/content/vyrocni-zpravy/en/ar2017en.pdf`.

[6] *Annual report of the security information service for 2018*, 2019. `https://www.bis.cz/public/site/bis.cz/content/vyrocni-zpravy/en/ar2018en.pdf`.

[7] V. BLAZEVIC, C. WIERTZ, J. COTTE, K. DE RUYTER, AND D. I. KEELING, *Gosip in cyberspace: Conceptualization and scale development for general online social interaction propensity*, Journal of Interactive Marketing, 28 (2014), pp. 87 – 100.

[8] J. DAI, H. YAN, T. SUN, P. LIU, AND X. QIU, *Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta*, 2021.

[9] J. DEVLIN, M.-W. CHANG, K. LEE, AND K. TOUTANOVA, *Bert: Pre-training of deep bidirectional transformers for language understanding*, 2019.

[10] *A Europe that protects: The EU steps up action against disinformation*, 2018. `http://europa.eu/rapid/press-release_IP-18-6647_en.htm`.

[11] A. GIACHANOU AND F. CRESTANI, *Like it or not: A survey of twitter sentiment analysis methods*, ACM Computing Surveys (CSUR), 49 (2016), pp. 1–41.

[12] T. HATMAKER, *Youtube bans david duke, richard spencer and other white nationalist accounts*, 2020. https://techcrunch.com/2020/06/29/youtube-ban-stefan-molyneux-david-duke-white-nationalism/.

[13] J. KAPUSTA, P. HÁJEK, M. MUNK, AND L'UBOMÍR BENKO, *Comparison of fake and real news based on morphological analysis*, Procedia Computer Science, 171 (2020), pp. 2285 – 2293. Third International Conference on Computing and Network Communications (CoCoNet'19).

[14] M. KUSNER, Y. SUN, N. KOLKIN, AND K. WEINBERGER, *From word embeddings to document distances*, in Proceedings of the 32nd International Conference on Machine Learning, F. Bach and D. Blei, eds., vol. 37 of Proceedings of Machine Learning Research, Lille, France, 07–09 Jul 2015, PMLR, pp. 957–966.

[15] Y. LIU, M. OTT, N. GOYAL, J. DU, M. JOSHI, D. CHEN, O. LEVY, M. LEWIS, L. ZETTLEMOYER, AND V. STOYANOV, *Roberta: A robustly optimized bert pretraining approach*, 2019.

[16] P. MIKA, *Ontologies are us: A unified model of social networks and semantics*, Journal of web semantics, 5 (2007), pp. 5–15.

[17] T. MIKOLOV, K. CHEN, G. CORRADO, AND J. DEAN, *Efficient estimation of word representations in vector space*, 2013.

[18] C. T. NGUYEN, *Echo chambers and epistemic bubbles*, Episteme, 17 (2020), p. 141–161.

[19] M. PAVLÍČEK, T. FILIP, AND P. SOSÍK, *Zrec.org - psychosocial phenomena studies in cyberspace*, in ITAT 2020: Information Technologies – Applications and Theory, 2020, pp. 209–216.

[20] O. RONNEBERGER, P. FISCHER, AND T. BROX, *U-net: Convolutional networks for biomedical image segmentation*, 2015.

[21] K. SHU, A. SLIVA, S. WANG, J. TANG, AND H. LIU, *Fake news detection on social media: A data mining perspective*, SIGKDD Explor. Newsl., 19 (2017), p. 22–36.

[22] C. SINDERMANN, J. D. ELHAI, M. MOSHAGEN, AND C. MONTAG, *Age, gender, personality, ideological attitudes and individual differences in a person's news spectrum: how many and who might be prone to "filter bubbles" and "echo chambers" online?*, Heliyon, 6 (2020), p. e03214.

[23] C. SUN, L. HUANG, AND X. QIU, *Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence*, arXiv preprint 1903.09588, (2019).

[24] H. TIAN, C. GAO, X. XIAO, H. LIU, B. HE, H. WU, H. WANG, AND F. WU, *Skep: Sentiment knowledge enhanced pre-training for sentiment analysis*, 2020.

[25] A. VASWANI, N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES, A. N. GOMEZ, L. U. KAISER, AND I. POLOSUKHIN, *Attention is all you need*, 2017.

[26] D. WADDEN, U. WENNBERG, Y. LUAN, AND H. HAJISHIRZI, *Entity, relation, and event extraction with contextualized span representations*, 2019.

[27] X. WANG, Y. JIANG, N. BACH, T. WANG, Z. HUANG, F. HUANG, AND K. TU, *Improving named entity recognition by external context retrieving and cooperative learning*, 2021.

[28] D. WEISSENBORN ET AL, *Jack the Reader — a machine reading framework*, in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL) System Demonstrations, July 2018.

[29] B. XU, Q. WANG, Y. LYU, Y. ZHU, AND Z. MAO, *Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction*, 2021.

[30] A. YADAV AND D. K. VISHWAKARMA, *A comparative study on bio-inspired algorithms for sentiment analysis*, Cluster Computing, 23 (2020), pp. 2969–2989.

[31] I. YAMADA, A. ASAI, H. SHINDO, H. TAKEDA, AND Y. MATSUMOTO, *Luke: Deep contextualized entity representations with entity-aware self-attention*, 2020.

[32] Z. YANG, Z. DAI, Y. YANG, J. CARBONELL, R. SALAKHUTDINOV, AND Q. V. LE, *Xlnet: Generalized autoregressive pretraining for language understanding*, 2020.

[33] L. ZHANG, S. WANG, AND B. LIU, *Deep learning for sentiment analysis: A survey*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8 (2018), p. e1253.

[34] N. ZHANG, X. CHEN, X. XIE, S. DENG, C. TAN, M. CHEN, F. HUANG, L. SI, AND H. CHEN, *Document-level relation extraction as semantic segmentation*, 2021.

[35] T. ZHANG, V. KISHORE, F. WU, K. Q. WEINBERGER, AND Y. ARTZI, *Bertscore: Evaluating text generation with bert*, 2020.