PROGRAM BOOKLET

Student Poster Session
June 26, 2020

Table of contents

Blanca Calvo (GRONINGEN)

Explaining the dynamics between the issues mass media emphasize and the salience audiences attribute to them is the main concern of agenda-setting research. Previous studies suggest that the capacity of media to influence the salience of issues depends on four para- meters, namely: (1) obtrusiveness; (2) duration; (3) abstractness; and (4) dramatism. Using a typology based on these parameters (Soroka, 2002), we ask whether and how issue-types give rise to different life patterns of news stories. We thus identify three measurable dimensions: (a.) lifespan (i.e., the number of days over which articles of a news story are published); (b.) intensity (i.e., the number of articles per day per story); and (c.) burstiness (i.e., the speed with which stories go from emer- ging to their climax).

We collected a corpus of 50.385 political articles from major Spanish newspapers in 2018. News stories were generated by aggregating articles with K-means clustering. For each week, we iden- tified the number of clusters (i.e. stories) using the elbow method and removed general clusters using silhouette analysis. Then, articles relating to the clusters were tracked in the colliding weeks. Our approach is language independent and can be applied to study news issue-types across media platforms and countries.

The quantitative dimensions show that the issue-types differentiate only to some extent. Sensational stories have the highest intensity, shortest lifespan, and highest burstiness (i.e., they appear "out of the blue"). Prominent stories last for long periods (14 days

on average) but have low intensity and burstiness (i.e. they slowly grow to their climax). Governmental stories lay in the middle between these two latter.

Gaetana Ruggiero (GRONINGEN)

Authorship Attribution on Italian Personal Writings

Authorship Attribution (AA) is the study of identifying the author of a text. Over the past few years, determining the authors of online content has played a crucial role in many fields, such as online security [8], plagiarism detection [4] and fake news identification [1]. While extensive research has been done for English [5, 6], little investigation has focused on Italian, with the only outstanding case being the study on Elena Ferrante's true identity [7]. Moreover, existing research on AA focuses on texts for which a lot of data is available (i.e novels, articles), and which are not necessarily influenced by an author's personal writing style due to editorial intervention.

This study approaches the AA task in terms of Authorship Verification (AV), a binary classification task where, given two texts, the goal is to decide whether or not they are written by the same author. A significant contribution of this work is the collection of two datasets, one containing blog posts and one containing diary fragments. Following Hürlimann et al. [2] and inspired by the work on blogger identification of Mohtasseb et al. [3], the GLAD AV system [2] is run on the two datasets. The study focuses on the interaction between four different variables: genre, topic, authors' gender and number of words taken into account per author. Since a first set of experiments highlighted the possible introduction of a topic bias when creating the text pairs, the second set of

experiments focuses on eliminating this artifact, which hopefully will lead to an increase in interpretability.

References

[1] Sadia Afroz, Michael Brennan, and Rachel Greenstadt. 2012. Detecting hoaxes, frauds, and deception in writing style online. In 2012 IEEE Symposium on Security and Privacy, pages 461–475. IEEE.

[2] Manuela Hürlimann, Benno Weck, Esther van den Berg, Simon Suster, and Malvina Nissim. 2015. Glad: Groningen lightweight authorship detection. In CLEF (Working Notes).

[3] Haytham Mohtasseb, Amr Ahmed, et al. 2009. Mining online diaries for blogger identification.

[4] Andi Rexha, Mark Kröll, Hermann Ziak, and Roman Kern. 2018. Authorship identification of documents with high content similarity. Scientometrics, 115(1): 223–237.

[5] Efstathios Stamatatos, Walter Daelemans, Ben Verhoeven, Martin Potthast, Benno Stein, Patrick Juola, Miguel A Sanchez-Perez, and Alberto Barrón-Cedeño. 2014. Overview of the author identification task at pan 2014. In CLEF 2014 Evaluation Labs and Workshop Working Notes Papers, Sheffield, UK, 2014, pages 1–21.

[6] Efstathios Stamatatos, Martin Potthast, Francisco Rangel, Paolo Rosso, and Benno Stein. 2015. Overview of the pan/clef 2015 evaluation lab. In International Conference of the Cross-Language Evaluation Forum for European Languages, pages 518–538. Springer.

[7] Arjuna Tuzzi and Michele A Cortelazzo. 2018. Drawing Elena Ferrante's Profile: Workshop Proceedings, Padova, 7 September 2017. Padova UP.

[8] Min Yang and Kam-Pui Chow. 2014. Authorship attribution for forensic investigation with thousands of authors. In IFIP International Information Security Conference, pages 339–350. Springer.

Inge Salomons (GRONINGEN)

Improving Impaired Speech using Non-Parallel Voice Conversion

Impaired speech can result in reduced intelligibility, causing communication problems and other (social) difficulties for patients. To overcome this, Voice Conver- sion (VC) methods have been proposed to improve the intelligibility of impaired speech. VC aims to modify a speech signal uttered by a source speaker to sound as if it was uttered by a target speaker, while maintaining the linguistic content of the utterance. In a non-parallel approach, the linguistic information of the source and target data is unrelated. One of the advantages of this approach is that the data set of unimpaired speech can be relatively small. We adopt a state-of-the-art model that has improved the intelligi- bility of impaired speech resulting from an articulation disorder in Mandarin, and aim to determine to what extent this model can be applied to improve the intelligibility of esophageal speech. This type of impaired speech is produced by means of vibrations in the esophagus after the larynx has been removed. Due to this difference in speech production, esophageal speech differs from natural speech in a wide range of acoustic features. The model is a Generative Adversa- rial Network (GAN) and is trained on a large data set of unimpaired and a small data set of esophageal Spanish speech. The results show that the generated speech is closer to unimpaired speech in naturalness, but not in intelligibility. This indicates that a non-parallel approach might be suitable for other

types of impaired speech but that esophageal speech requires a more data-specific model.

Marion Bartl (GRONINGEN)

Gender bias in BERT: does context make it better?

Just like traditional word embeddings ([1], [2]), contextualized word embeddings ([3], [4]) capture social biases such as gender bias ([5], [6]). Even though the quick adaptation of contextualized embeddings led to improvements in many NLP tasks, previous techniques for quantifying or mitigating gender bias cannot be adapted as smoothly. Therefore, this work aims to find a reliable method of quantifying gender bias in BERT and to subsequently test the findings against real-world statistics. Finally, I examine whether applying Counterfactual Data Substitution (CDS) to fine-tuning data can mitigate bias.

Following the approach of Kurita et al. (2019)[5] and inspired by sentence templates in the Equity Evaluation Corpus (EEC) ([7]), I built templates that contain both a gendered word (Target) and a profession (Attribute) with either low, equal or high female participation [8]. Comparing target probabilities assigned by the BERT language model, I assume that in a non-biased setting, the target probability should not be influenced by the profession word in the sentence. Moreover, I am planning to mitigate gender bias in BERT by using CDS, which manipulates text by switching the gender of pronouns and person words ([9], [10]). After fine-tuning on a dataset treated with CDA I am expecting to observe a decrease in the intensity of bias.

References

[1] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

[2] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805.

[4] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

[5] Keita Kurita, Nidhi Vyas, Ayush Pareek, Alan W Black, and Yulia Tsvetkov. 2019. Measuring bias in contextualized word representations. arXiv preprint arXiv:1906.07337.

[6] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Ryan Cotterell, Vicente Ordonez, and Kai-Wei Chang. 2019. Gender bias in contextualized word embeddings. arXiv preprint arXiv:1904.03310.

[7] Svetlana Kiritchenko and Saif M Mohammad. 2018. Examining gender and race bias in two hundred sentiment analysis systems. arXiv preprint arXiv:1805.04508.

[8] Bureau of Labor Statistics (BLS). 2020. Labor force statistics from the current population survey. [Online; accessed 16-March-2020].

[9] Kaiji Lu, Piotr Mardziel, Fangjing Wu, Preetam Amancharla, and Anupam Datta. 2018. Gender bias in neural natural language processing. arXiv preprint arXiv:1807.11714.

[10] Rowan Hall Maudslay, Hila Gonen, Ryan Cotterell, and Simone Teufel. 2019. It's all in the name: Mitigating gender bias with name-based counterfactual data substitution. arXiv preprint arXiv:1909.00871.

Marta Carletti (SAARLAND)

Argumentation is defined as any methodological and inferential reasoning going from a set of premises to a conclusion with the purpose of justifying acts, beliefs, attitudes and values. An argument represents an opportunity for a human or a system to convince an interlocutor of the veracity of its own beliefs or modify the interlocutor's opinions. This perspective on argumentation as persuasive technique where different (or similar) systems of beliefs interact with each other implies the necessity of the construction and formalization of different models. In particular, other than a system for the argument structure itself and the operations available in an argument, a model of beliefs for both the speaker and addressee is needed. Models of belief not only represent the grounding of an argument from a speaker A, but also how the argument itself is likely to be received by an interlocutor B. Moreover, dealing with conversations involving two speakers, beliefs and arguments need to be inserted and integrated in a system that accounts for the flow of a debate where more argument structures interact and beliefs update continuously on the basis of a feedback. This thesis proposes a modular architecture to formalize argumentation in dialogue, and, in particular, it attempts to model the interface between (i) the argument structure (AS), (ii) the speaker/addressee systems of beliefs (BS) and (iii) the dialogue structure (DS). The goal is to reach an architecture that enables to stock the beliefs of both interlocutors and detect whether a belief has changed or not after an

10/18
[scroll back to beginning]

argumentative dialogue and the respective operation (contraction, expansion, revision of belief) that took place.

Simon Preissner (SAARLAND)

Similar, but Different: Unsupervised Detection of Semantic Shifts in Diachronic Word Embeddings

Human language differs across many dimensions: situations, domains, culture etc. For this reason, a large body of research in NLP aims to identify and measure these factors, one of the most researched of these being diachronic language change, i.e. change of language over time. NLP usually investigates semantic shifts within word embeddings, which represent lexical semantics of a large body of natural language in the form of multidimensional spaces where words are expressed as vectors whose proximity is interpreted as semantic similarity.

I present work-in-progress on the automatic detection of diachronic meaning shifts. The starting point are word embeddings obtained from various time intervals of the Royal Society Corpus (RSC), a collection of English scientific texts that spans from 1665 to 1929. Previous work on language change within the RSC obtained measurements for each space separately and compared them across time intervals. Differently to this, I align two spaces via Gromov-Wasserstein Optimal Transport (GWOT) to make them directly comparable to each other. GWOT aligns spaces via common constellations of elements, without the need for a bilingual signal. By restricting the alignment to the semantically most stable concepts, I preserve the semantic shifts.

Next, I detect these shifts. Most approaches search for shifts in predefined sets of words, which are prone to

incompleteness. Therefore, I combine similarity measures and clustering methods to detect systematic shifts independently from human intuition. I validate the methods for alignment and shift detection and relate the findings to previous work with the RSC.

Siyana Pavlova (SAARLAND)

Going from UD towards AMR

This poster presents our efforts to develop a system that takes sentences in natural language, parses them in the Universal Dependencies (UD) framework and applies a set of rewrite rules on the UD parses to produce Abstract Meaning Representations (AMRs) of the sentence. The rewriting system is supported with a lexical resource containing predicates from the PropBank dataset, a part of which is enriched with semantic role information. We evaluate our system on the first 100 sentences of the AMR Bank and show that both our rulebase and lexicon increase F1-score on said dataset.

Xiao Xu (BASQUE COUNTRY)

Predicting Legal Judgments with Multilingual Classification Model

With the developing trend of globalization, an increasing amount of documents are now available in multiple languages, and automatically processing these documents with Natural Language Processing techniques (NLP) has been an emerging challenge. Text classification is one of the fundamental tasks in NLP with broad applications, such as spam detection, document labeling and sentiment analysis. Within the legal domain, text classification has been applied to predict the outcome of case law, based on ever-growing databases for legal data. The outcome prediction of the European Court of Human Rights (ECtHR) is the focus of many researchers (Aletras et al. 2016, Medvedeva et al. 2019) who have been building models for automatically predicting its judicial decisions. However, as well as many other European institutions, the ECtHR works in a multilingual way as it adopts both English and French as its working language, and its documents are available in either languages or both. Since these earlier classification models were built for English data only, which fail to cover many legal cases and do not make full use of the available resources, a multilingual classification model is proposed in this study to work on multiple different languages simultaneously.

Yu-Wen Chen (SAARLAND)

Previous research has shown that translated texts, in any language, can be distinguished from originally-written languages through several linguistic-associated features. These shared features form a universal phenomenon termed as 'Translationese', making the translated texts a certain 'dialect' of the language. Such type of discrepancy is found in human translation from scratch (HT) and machine translation (MT), and also in Post-editing (PE), within which it is named 'Post-editese'.

However, as the recent MT development aims for human parity, the difference of translationese characteristics between HT and MT still remains unclear as the quality of MT systems is currently evaluated according to automatic metrics in most of the available research. This thesis aims to bridge this gap by exploring four aspects of the translation universals - simplification, normalization, explicitation, and interference. Under this framework, we implement computational approaches based on Supervised Machine Learning with several syntactic and lexical features extracted from dependency parsing and POS tags. We intend to answer the following questions: which aspects of the translationese are valuable to distinguish between HT and MT, and moreover, whether PE also exhibits distinct attributes compared with HT while the human parity is claimed. Along this line, we hope to contribute a deeper understanding of the unique characteristics of

the machine translationese and Post-editese with their relations to the human parity.