Techniques for Agenda Detection

Anonymous ACL submission

Abstract

The behavior and decision making of groups or communities can be dramatically influenced by individuals pushing particular agendas. In the examination of online influence campaigns, particularly those related to important political and social events, scholars often concentrate on identifying the sources responsible for setting the agenda (e.g., public media). In this article we present a methodology for detecting specific instances of agendas in situations where annotated data is limited or non-existent. By using a modest corpus of Twitter messages centered on the 2022 French Presidential Elections, we carry out a comprehensive evaluation of various approaches and techniques that can be applied to this problem. Our findings demonstrate that by treating the task of text classification as a textual entailment problem, it is possible to overcome the requirement for a large annotated dataset.

1 Introduction

An agenda, being a collection of items to be attended to in a certain order, can have a significant impact on the actions of a group, especially in the context of interpersonal communication and relationships. The individual who establishes and directs the agenda often exercises considerable control and influence over the group. In the sociolinguistics of group behavior, the concept of agenda control is widely recognized as a strong indicator of both leadership and influence, as evidenced by numerous studies in the field (Wang et al., 2018).

When studying the impact of online influence campaigns, such as those surrounding significant political and social events (e.g., elections), researchers often focus on evidence of agenda-setting activity emanating from particular sources. These sources could be traditional public media or clandestine online groups that wish to shape public opinion. According to social science literature, there are three distinct levels of agenda setting. At level one, the public is told explicitly what to think and do in a given situation, for example, to vote for a particular candidate. In the second level of agenda setting, rather than prescribing specific beliefs or actions, the influencers emphasize certain aspects of their targets (e.g., political candidates) as either positive or negative, outwardly leaving the public to form their own opinions (McCombs et al., 1997; Balmas and Sheafer, 2010; Meraz, 2011). At the third level, multiple targets are associated to one another through direct comparison or juxtaposition (Guo et al., 2012) thus imparting apparent preferences onto the public.

In this study, we are interested in both level one and level two agenda setting activities and how to detect their presence in social media, with a specific focus on the 2022 French Presidential Elections. Our goal is to detect specific instances of agendas being actively promoted via Twitter messages (tweets), including retweets, responses, and mentions, posted in multiple languages during the relevant time period. The objective is to automatically tag each tweet with appropriate agenda labels, from the set of labels curated by experts. Given the practical limitations of obtaining a large quantity of annotated training data, a phenomenon true to most real-world applications, our approach focuses on utilizing either a small, expert annotated sample or relying on zero-shot and few-shot methods. Our proposed framework provides a solution for general agenda detection that operates on a set of ad-hoc agenda labels.

We propose to cast the agenda detection task as a textual entailment problem. In previous studies (Yin et al., 2019), text classification has been viewed through the lens of textual entailment with promising results. This approach imitates how humans make decisions while annotating text examples, picking the correct label among all possible labels. Human annotators are often given a task description, as well as label definitions that explain the meaning of each candidate label. Equipped with these definitions, a human can understand the problem and mentally construct a hypothesis by picking a candidate label to fill in the blank: *"This text is about ___"*. Then they ask themselves if this hypothesis is true given the text example.

We treat agenda detection as a textual entailment problem so that our model can gain knowledge from entailment datasets (Bowman et al., 2015; Williams et al., 2017; Dagan et al., 2006; Bentivogli et al., 2009).

2 Related Work

In this section, we present a review of select literature in the domains of traditional agenda detection and text classification through the lens of textual entailment, as our proposed models draw inspiration and incorporate elements from these fields.

2.1 Agenda Detection

The impact of various agendas being pushed through the media (both official and unofficial) have on shaping public opinion has been widely studied, as has the interplay between the news outlets and the social media. For example, (McCombs et al., 1997) attempts to understand how media agendas shape or influence the public's opinion on political candidates, and (Vargo et al., 2014) studies how the public selectively accepts media agendas. Additionally, the effect that news media and social media have on each other is closely examined in (Su et al., 2020).

In the absence of large annotated data sets, scholars often perform manual analysis to detect agendas. (McCombs et al., 1997) hand-coded news articles and surveys, while (Su and Borah, 2019) conducted a manual analysis of a sample of collected tweets. Automated methods such as those used in (Vargo et al., 2014), (Ceron et al., 2016) and (Haim et al., 2018) utilized sets of keywords to detect topics and sentiment associated with the target agendas, rather than the agendas themselves.

More recently, several studies explored machine learning methods for the detection of agendas in big data. In (Su et al., 2020; Su, 2022; Guo, 2019), the authors first utilize topic modeling to identify the topics within their datasets. Human experts manually develop agenda labels associated with each topic. Subsequently, multiple annotators tag a subset of the data using the agenda labels developed in the previous step. The labeled data is then used to train a set of Support Vector Machine (SVM) (Boser et al., 1992) classifiers, one per each agenda label. If the average performance of the classifiers was not satisfactory (e.g., F1 < 0.7), more data is annotated and the training is repeated. The total number of items annotated in (Guo, 2019) was 2000 news articles, in (Su et al., 2020) it was 3000 tweets and 500 news articles, and in (Su, 2022) it was 1500 news articles and 5000 tweets. The authors of (Chen et al., 2019) used a similar approach, but deployed different classifiers, trained on 2500 annotated microblog messages. We note that all the above approaches are costly and impractical, particularly in novel and rapidly evolving situations.

2.2 Textual Entailment Text Classification

In their work, (Yin et al., 2019) introduced a framework for text classification by formulating it as a series of premise-hypothesis pairs, where the premise is the text to be classified and the hypotheses represents the candidate labels, essentially transforming the problem into a textual entailment challenge. Their study demonstrated the efficacy of this method, and released a benchmark dataset for 0-shot text classification. Subsequently, this approach has been widely adopted and expanded for many 0-shot text classification tasks (Shu et al., 2022; Zhang et al., 2020; Wang et al., 2021; Seoh et al., 2021). Our work builds upon this basic methodology by applying textual entailment to the task of agenda detection, and we evaluate and compare various approaches that could also be used to solve this problem, including conventional text classification methods. By doing so, we demonstrate the utility of this framework for addressing the task of agenda detection in the absence of large annotated datasets.

3 Data

Our proposed model makes use of pre-existing textual entailment datasets, described in the following paragraphs, to gain general knowledge. Then, the model is trained on agenda specific data.

3.1 Pre-training Data

To teach our model how to solve the textual entailment task, we deploy three widely used datasets into an early fine-tuning training step. These datasets are i) the Stanford Natural Language Inference (SNLI) dataset (Bowman et al., 2015), ii) the

Multi-genre Natural Language Inference (MNLI) dataset (Williams et al., 2017), and iii) the Recognizing Textual Entailment (RTE) dataset [1] (Dagan et al., 2006; Bentivogli et al., 2009), which is also part of the GLUE benchmark (Wang et al., 2018).

We convert all datasets to represent binary classification problems, where for three-class datasets, we collapse neutral and contradiction into not entailment, such that our model learns to distinguish entailment versus not entailment. All three datasets come with predefined training examples, so we merge all three training partitions into a single training set. Since our downstream task of Agenda detection contains messages that are in English and French, we automatically translate[2] 30% of the combined SNLI/MNLI/RTE training dataset from English to French.

In our experiments, discussed further into this article, we test whether including this data collection into the fine-tuning process improves the performance of the textual entailment approach.

3.2 Fine-tuning Data

To fine-tune our model to the task of agenda detection, we use the publicly available data set of tweets[3]. This collection of Twitter messages contains posts on the topic of the 2022 French Presidential Elections that were made on the platform between 8 November, 2021 and 3 April, 2022. The posts were filtered by keywords including the candidate names and their associated official Twitter account, however, they do not have any agenda annotations.

The agenda labels under consideration and their definitions in English are presented in Table 1. We should note that the "Violent Action" agenda specifically encourages its (non-government) audience to engage in violence, where violence is understood to be essentially physical rather than e.g. verbal.

To assemble our Agenda dataset, we leverage machine translation to obtain the French version of our previously established Agenda definitions. Given that our data is primarily French but also includes English messages, we utilize a multi-lingual sentence embedding model and calculate the cosine similarity between the Agenda definition embeddings and the embedding of each tweet. This enables us to bootstrap the annotation process by sorting the un-tagged messages based on their relevance to our task.

To ensure representation of all Agenda classes in the final dataset, we collect a minimum of 500 messages with the highest similarity score for each class. It should be noted that the number of messages retrieved for each Agenda class varies, with some classes yielding more messages than others. To counterbalance any potential biases introduced by our bootstrapping method, we randomly selected an additional 500 messages for annotation. Then, two annotators independently tag the 4096 messages with the appropriate Agenda labels. If a message cannot be categorized into any of the Agenda classes, it is designated as "Other". After the two annotators annotate the tweets, we calculate the Inter Rater Reliability. The annotators achieve 97.5% of agreement and Cohen's Kappa of 0.89. Disagreements are resolved by discussion between the annotators.

To train and evaluate our models, we create a training, a dev and a test set. Overall, we include 96-120 messages per class from the annotations, except the "Other" class for which we randomly pick 506 (the sum of the other 5 classes) messages. Therefore, our final dataset has 1012 annotated messages in total, 35 of which have more than one label (Table 2). We make our annotated dataset available for public use so that it can be of benefit to future research.

This data collection is prepared to be used in both textual entailment and traditional text classification approaches, that serve as our baselines. Any model trained on this set is not considered to be a 0-shot method, but rather to be few-shot, as each agenda class does not have more than a small number of training examples.

4 Method

In our approach, we cast the problem of text classification as a textual entailment problem. This enables our system to gain further knowledge from entailment datasets, essentially learning how to imitate the human decision-making process of categorizing text.

Following similar methods to (Yin et al., 2019), we depart from the traditional text classification methods where labels are denoted as indices and models lack any understanding of their specific

[1] RTE dataset: https://dl.fbaipublicfiles.com/glue/data/RTE.zip

[2] Machine Translation model: https://huggingface.co/Helsinki-NLP/opus-mt-en-fr

[3] Un-tagged Twitter corpus: https://www.kaggle.com/datasets/jeanmidev/french-presidential-online-listener

Agenda Labels	Agenda Definitions (EN)
Online Solidarity	The message encourages readers to share information relevant to a cause, promote or magnify the positions of specific individuals, use symbols or language in online profiles to demonstrate support for a specific position on an issue.
Engagement	The message encourages readers to engage in the formal political process, either by voting, attending public government meetings, assemblies, etc., to support or oppose a candidate, party, law, political position, or (nominally) collective action by a government.
Disengagement	The message encourages readers to disengage from a normal political, economic, or social process in order to demonstrate opposition to the status quo on a specific issue or to highlight the importance of a specific stance.
Peaceful Protest	The message encourages readers to protest peacefully, to attend rallies, marches, and other forms of mass political demonstration, etc. in support of or opposition to a cause. The action or demonstration urged by the document must be non-violent in nature.
Violent Action	The message encourages readers to engage personally in violent or destructive action (bombing, destruction of property, formation of militias, fighting in foreign countries in a mercenary capacity, etc.).
Other	The text is about something else.

Table 1: Agenda labels with definitions.

Agenda Labels	Train	Dev	Test	Total
Online Solidarity	77	10	10	97
Engagement	96	12	12	120
Disengagement	96	12	12	120
Peaceful Protest	86	11	11	108
Violent Action	76	10	10	96
Other	404	51	51	506
Total	808	102	102	1012

Table 2: Number of messages per class in the Train, Dev and Test sets.

interpretation or meaning. Instead, the labels are transformed into a set of natural language hypotheses that the input messages will be paired against and the truth value of the label can be decided. This way the system can understand the described task and the meaning of the labels by associating the input text and the context of the hypotheses.

4.1 Models

Our proposed approach leverages the cutting-edge T5 (Raffel et al., 2020) language model and its variants, mT5 (Xue et al., 2020) and T5v1.1[4]. T5 stands out for its exceptional performance, owing to a number of key factors, such as its encoder-decoder architecture, the corrupting span denoising objective, and the utilization of an extensive pre-training dataset. Furthermore, T5v1.1 and mT5 are further enhanced by the integration of GeLU (Shazeer, 2020) activation. mT5, in particular, has been pre-trained on over 120 languages, including French, which is of particular interest in the context of our task.

[4]T5 v1.1: `https://huggingface.co/google/t5-v1_1-base`

4.2 Pre-training

For models trained under the Textual Entailment framework, we first train using the binarized SNLI/MNLI/RTE pre-training dataset discussed above, such that our models learn to distinguish entailment versus not entailment when a premise and a hypothesis are given as inputs. This step has been shown (Yin et al., 2019) to improve the robustness of the model on 0-shot text classification tasks. For traditional classification approaches, we do not include a similar pre-training step.

4.3 Fine-tuning

To adapt our Agenda dataset into a format suitable for Textual Entailment, we convert each unique Agenda label into a hypothesis following the process described in the following section. Then, we consider each input message as a premise that has positive hypotheses (entailment) corresponding to the ground truth label, while negative labels provide negative hypotheses (not entailment). During fine-tuning, we form all possible positive (i.e., entailment) message-to-label examples (where a message may be associated with more than one Agenda), and in addition, two negative examples (i.e., non-entailment) for each positive. This mimics the distribution of entailment/not entailment found in our pre-training dataset. For traditional classification approaches, each message is associated with its ground truth labels in a multi-label fashion.

4.3.1 Generating Hypotheses from Agenda Labels

An integral part of our approach is the construction of hypotheses representing the Agenda classes. Here, we use the definition of each class, see Table 1, as a guide to write succinct hypotheses in natural language. We first write the hypotheses in English and use machine translation to obtain their French versions. The hypotheses we used in our experiments are listed for each class in Table 3.

Initially, in our experiments we used the Agenda class definitions as our hypotheses without trying to shorten the text. The longer hypotheses, even though they are more comprehensive, resulted in slightly poorer performance during testing. This led us to experiment with the shortened versions listed in Table 3. We only report evaluation results using the shorter hypotheses.

4.3.2 Interpreting Agenda Predictions from Textual Entailment

Finally, textual entailment classification results can be interpreted into one or more Agenda classes. As our base case, when the model predicts non-entailment for all possible hypotheses for an input example, we resolve it as the "Other" Agenda class and output it as the final prediction for that example. In the cases where there are one or more entailment predictions for some input text, all Agenda classes corresponding those hypotheses form the final output prediction.

When outputting confidence scores, for each class we look at the probability of the corresponding hypothesis being entailed. For generative models, we use the probability of the related token.

5 Experimental Set-up & Results

Our textual entailment based approach is evaluated against a range of baseline techniques, such as conventional text classification and semantic search. We adopt a multi-class multi-label approach evaluation process, as our textual entailment approach predicts the entailment of a premise (tweet message) with respect to each of the hypotheses, one class at a time. We perform the evaluation on a static test set by varying the decision threshold of each model based on the macro-averaged F1-score, as optimizing the threshold on the micro-average tends to favor the most frequent class, in this case "Other". The micro-averaged F1-scores are reported after the optimal thresholds are determined.

This evaluation method is cost-effective as it can be performed after obtaining predictions on the test set without updating the underlying model. The decision threshold value reflects the model's confidence level, with higher values indicating greater confidence and lower values indicating that predictions with low confidence are accepted, which can result in increased False Positives. We set the minimum possible threshold to 0.3, as we are not interested in trivial scenarios where predictions include all of the available agenda labels.

5.1 0-shot Agenda Detection

In this experiment, we evaluate the performance of our proposed model for detecting agendas in social media on the 0-shot setting. To provide a comprehensive evaluation, we compare our model with several baselines, including BERT (Devlin et al., 2018), SBERT (Shazeer, 2020) and mnli-BART[5].

Our semantic search baselines use SBERT to obtain sentence embeddings of the messages and the hypotheses, but we also test a variant using the label's text. Then, we compare the message embeddings to each hypothesis embedding using cosine-similarity. The computed score serves as the confidence that the message belongs to the agenda class specified by each hypothesis. This baseline approach yields four models, two comparing the English-only (all-mpnet-base-v2) versus the multilingual model (paraphrase-multilingual-mpnet-base-v2), and two comparing the use of hypotheses versus the labels themselves.

For our proposed approach in the 0-shot setting, we pre-train the models on the combination of the SNLI, MNLI, and RTE datasets. We compare the performance of the T5 model with BERT and pre-train on either the English-only or bilingual ver-

[5]mnli-BART: https://huggingface.co/facebook/bart-large-mnli

Agenda Labels	Hypotheses (EN)	Hypotheses (FR)
Online Solidarity	The author encourages readers to share information relevant to a cause, promote the positions of individuals and show support for a position on an issue.	L'auteur encourage les lecteurs à partager des informations pertinentes pour une cause, à promouvoir les positions des individus et à montrer leur soutien à une position sur une question.
Engagement	The document encourages readers to engage in the formal political process, by voting and attending public government meetings , to support or oppose a candidate, party, law or a political position.	Le document encourage les lecteurs à s'engager dans le processus politique formel, en votant et en assistant aux réunions publiques du gouvernement, pour soutenir ou s'opposer à un candidat, un parti, une loi ou une position politique.
Disengagement	The author wants the readers to disengage from a normal political process in order to demonstrate opposition to the status quo on an issue or to highlight the importance of a stance.	L'auteur souhaite que les lecteurs se désengagent d'un processus politique normal afin de manifester leur opposition au statu quo sur une question ou de souligner l'importance d'une position.
Peaceful Protest	The message motivates the readers to protest peacefully in support of or opposition to a cause.	Le message motive les lecteurs à manifester pacifiquement pour soutenir ou s'opposer à une cause.
Violent Action	The author rallies the audience to engage personally in violent or destructive action.	L'auteur rallie le public à s'engager personnellement dans une action violente ou destructrice.
Other	The text is about something else.	Le texte parle d'autre chose.

Table 3: Agenda labels with hypotheses.

Models		EN		FR		Overall		Thresh
		micro	macro	micro	macro	micro	macro	
Semantic Search	hypotheses-SBERT	0.31	0.30	0.27	0.32	0.30	0.30	0.39
	hypotheses-mSBERT	0.28	0.32	0.32	0.33	0.30	0.32	0.52
	labels-SBERT	0.27	0.29	0.11	0.13	0.21	0.24	0.30
	labels-mSBERT	0.28	0.33	0.31	**0.36**	0.30	0.34	0.39
Pre-trained	rte-en-BERT	**0.50**	0.18	0.34	0.28	0.39	0.26	0.44
	rte-bi-mBERT	0.46	0.14	**0.50**	0.22	0.48	0.19	0.31
	rte-en-T5	0.49	0.36	0.42	0.33	0.45	**0.36**	0.31
	rte-bi-mT5	**0.50**	0.21	**0.50**	0.27	**0.50**	0.24	0.30
	mnli-BART	0.32	**0.39**	0.35	0.23	0.33	0.32	0.37

Table 4: 0-shot evaluation results. We report both micro- and macro-averaged F1-scores. The decision threshold was optimized based on the overall macro-average F1-score. An "m" in the models name indicates the use of the multi-lingual underlying model, while "rte-en" and "rte-bi" refer to the use of English-only or bi-lingual pre-training using our combined RTE dataset.

sion of the combined RTE dataset. For instance, the "t5-v1.1-base" model is fine-tuned using the English-only version, while the "mt5-base" model is fine-tuned using the bilingual version. The models are then applied directly to the Agenda test set to generate predictions.

The availability of in-domain data is critical for training a robust classification model. Our 0-shot evaluation results (Table 4) show that a lack of in-domain data leads to low model confidence and a correspondingly low threshold. However, the models pre-trained on our combined RTE dataset using the textual entailment framework exhibit significant improvements, with an overall micro F1-score of 0.5, demonstrating the potential of this approach.

5.2 Textual Entailment Agenda Detection

Our proposed model leverages the power of textual entailment by combining general RTE pre-training and Agenda-specific fine-tuning to robustly detect agendas in short-form social media messages. In addition to our main model, we also train and evaluate BERT models and compare against variants that do not include the RTE pre-training.

Fine-tuning our models on the Agenda data while following the textual entailment framework resulted in the highest performing models, as can be seen in the detailed results presented in Table 5. For English-only scenarios, direct fine-tuning of BERT proves to be the optimal approach. However, in a multilingual setting, the mT5 model, which was pre-trained for textual entailment on our bilingual combined RTE data set (with 30% of the examples translated into French) and then fine-tuned on the Agenda data, outperforms all other baselines, including the conventional multi-label multi-class classification techniques.

This highlights that pre-training the model on the combined RTE dataset has a major impact on performance, as it gives the model strong task-specific knowledge, allowing it to tackle the textual entailment problem with ease. By fine-tuning on in-domain data, we observe even better results.

5.3 Conventional Text Classification for Agenda Detection

To detect agendas in social media, we also explore traditional text classification techniques. To this end, we train classifiers for the multi-label, multi-class task using Support Vector Machines (SVM) (Boser et al., 1992). The textual features are extracted through the TF-IDF vectorization yielding 1600 features. Additionally, we deploy BERT and T5 models as baselines, trained for sequence classification. Given T5's text-to-text architecture, we train the model using the tweet message as the source sequence and as the target sequence we use the agenda labels, represented as comma-delimited strings. During testing, T5 generates up to 32 tokens for a given text, which are then parsed and converted into agenda predictions.

The results from this experiment are presented in Table 6. Only slightly better than our 0-shot baselines, we see relatevley low confidence values and performance scores. In a surprising turn of events, mlc-BERT showed a significant decline in performance compared to agenda-BERT despite being trained on the same data. Our hypothesis is that by exploiting BERT's architecture, which is inherently suited for textual entailment, led to agenda-BERT's superior performance. The limited size of our Agenda training data, which has far fewer examples than what is typically required to produce robust models, could also have contributed to the suboptimal results. However, the SVM model serve as a noteworthy exception to the trend of comparatively low performance.

5.4 Discussion

To gain insight into the limitations of our model and identify areas that may require improvement, we calculate the confusion matrix based on the predictions generated by our best-performing model (agenda-rte-bi-mT5) on the French test set (Figure 1). The multi-class multi-label nature of the task results in not only false positives and false negatives, but also introduces the presence of extra labels, where a false positive does not have a corresponding false negative, and missed labels, where a false negative does not have a corresponding false positive.

The agenda-rte-bi-mT5 model tends to over-label, resulting in more predicted labels than actual true labels, with the exception of "Peaceful Protest". Most of the excessive predictions for "Online Solidarity", "Engagement", "Disengagement", and "Violent Action" under the "Other" class. But the extra predictions for "Other" are dispersed among various classes. However, this can be improved by using a better stated "Other" hypothesis or by using of out-of-domain "Other" messages. Conversely, most missed labels occur when the model lacks confidence in assigning any label to the input message.

Our model incorrectly classified two instances of "Violent Action" as "Peaceful Protest". One such message, translated into English, states *"... to get what we want, peaceful demonstrations are no use! You have to do as in Corsica or as in the suburbs!!!"*. We believe that by integrating external knowledge into the textual entailment process, e.g. knowledge regarding the violent incidents that

Models		EN		FR		Overall		Thresh
		micro	macro	micro	macro	micro	macro	
Fine-tuned	agenda-BERT	**0.71**	**0.66**	0.61	0.59	0.66	0.63	0.68
	agenda-mBERT	0.65	0.59	0.71	0.68	0.68	0.64	0.74
	agenda-T5	0.49	0.11	0.49	0.11	0.49	0.11	0.30
	agenda-mT5	0.50	0.11	0.48	0.11	0.49	0.11	0.56
Pre-trained +Fine-tuned	agenda-rte-en-BERT	0.69	0.66	0.67	0.62	0.68	0.64	0.58
	agenda-rte-bi-mBERT	0.65	0.57	0.69	0.66	0.67	0.62	0.81
	agenda-rte-en-T5	0.67	0.66	0.65	0.60	0.66	0.63	0.77
	agenda-rte-bi-mT5	0.69	0.66	**0.76**	**0.72**	**0.72**	**0.69**	0.80

Table 5: Trained textual entailment evaluation results. We report both micro- and macro-averaged F1-scores. The decision threshold was optimized based on the overall macro-average F1-score. An "m" in the models name indicates the use of the multi-lingual underlying model, while "rte-en" and "rte-bi" refer to the use of English-only or bi-lingual pre-training using our combined RTE dataset.

Models		EN		FR		Overall		Thresh
		micro	macro	micro	macro	micro	macro	
Fine-tuned	tfidf-SVM	**0.68**	**0.64**	**0.64**	**0.58**	**0.66**	**0.61**	0.30
	mlc-BERT	0.49	0.12	0.45	0.11	0.47	0.11	0.51
	mlc-mBERT	0.49	0.11	0.49	0.11	0.49	0.11	0.30
	mlc-T5	0.42	0.18	0.42	0.21	0.42	0.20	0.30
	mlc-mT5	0.48	0.11	0.49	0.11	0.49	0.11	0.30

Table 6: Evaluation results for models trained via the traditional multi-class multi-label classification approach. We report both micro- and macro-averaged F1-scores. The decision threshold was optimized based on the overall macro-average F1-score. An "m" in the models name indicates the use of the multi-lingual underlying model, and "mlc" stands for Multi-label Classification.

occured in Corsica during the French Election of 2022, could lead to further improvement in the model's performance.

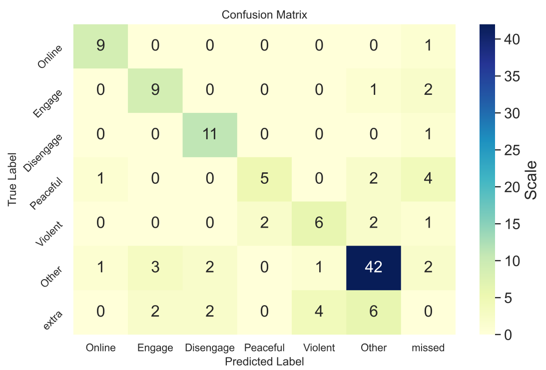

Figure 1: Confusion matrix of agenda-rte-bi-mT5 French results. The extra labels are in the bottom row and the missed labels are in the rightmost column.

6 Conclusion

The methodology we have presented for detecting agendas with limited or non-existent labeled examples demonstrates that it is possible to overcome the need for a vast amount of annotated data. This is evident from the superior evaluation results observed by our models. Through an extensive evaluation of various techniques and approaches applied to a small corpus of annotated Twitter messages centered on the 2022 French Presidential Elections, we have shown that treating the task of text classification as a textual entailment problem produces promising results that could not have been achieved through equivalent conventional sequence classification methods.

Our proposed model offers the advantage of not being limited to a set of predefined labels and allows for the testing of an arbitrary number of hypotheses to uncover a multitude of agendas. This versatility makes it an effective tool in detecting new and emerging influence campaigns in social media. However, the spread of agendas is not limited to just Twitter and can also occur through other media such as news articles and blogs, leading us to the next step of studying the applicability of our techniques in longer forms of text and discovering what new insights can be learned.

References

Meital Balmas and Tamir Sheafer. 2010. Candidate image in election campaigns: Attribute agenda setting, affective priming, and voting intentions. *International journal of public opinion research*, 22(2):204–229.

Luisa Bentivogli, Peter Clark, Ido Dagan, and Danilo Giampiccolo. 2009. The fifth pascal recognizing textual entailment challenge. In *TAC*. Citeseer.

Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152.

Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.

Andrea Ceron, Luigi Curini, and Stefano M Iacus. 2016. First-and second-level agenda setting in the twittersphere: An application to the italian political debate. *Journal of Information Technology & Politics*, 13(2):159–174.

Zhuo Chen, Chris Chao Su, and Anfan Chen. 2019. Top-down or bottom-up? a network agenda-setting study of chinese nationalism on social media. *Journal of Broadcasting & Electronic Media*, 63(3):512–533.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The pascal recognising textual entailment challenge. In *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment: First PASCAL Machine Learning Challenges Workshop, MLCW 2005, Southampton, UK, April 11-13, 2005, Revised Selected Papers*, pages 177–190. Springer.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Lei Guo. 2019. Media agenda diversity and intermedia agenda setting in a controlled media environment: A computational analysis of china's online news. *Journalism Studies*, 20(16):2460–2477.

Lei Guo, Hong Tien Vu, Maxwell McCombs, et al. 2012. An expanded perspective on agenda-setting effects. exploring the third level of agenda setting. *Revista de comunicación*, 11(1):51–68.

Mario Haim, Gabriel Weimann, and Hans-Bernd Brosius. 2018. Who sets the cyber agenda? intermedia agenda-setting online: the case of edward snowden's nsa revelations. *Journal of Computational Social Science*, 1:277–294.

Maxwell McCombs, Juan Pablo Llamas, Esteban Lopez-Escobar, and Federico Rey. 1997. Candidate images in spanish elections: Second-level agenda-setting effects. *Journalism & Mass Communication Quarterly*, 74(4):703–717.

Sharon Meraz. 2011. The fight for 'how to think': Traditional media, social networks, and issue interpretation. *Journalism*, 12(1):107–127.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough. 2021. Open aspect target sentiment classification with natural language prompts. *arXiv preprint arXiv:2109.03685*.

Noam Shazeer. 2020. Glu variants improve transformer. *arXiv preprint arXiv:2002.05202*.

Lei Shu, Hu Xu, Bing Liu, and Jiahua Chen. 2022. Zero-shot aspect-based sentiment analysis. *arXiv preprint arXiv:2202.01924*.

Yan Su. 2022. Networked agenda flow between elite us newspapers and twitter: A case study of the 2020 black lives matter movement. *Journalism*, page 14648849221092521.

Yan Su and Porismita Borah. 2019. Who is the agenda setter? examining the intermedia agenda-setting effect between twitter and newspapers. *Journal of Information Technology & Politics*, 16(3):236–249.

Yan Su, Jun Hu, and Danielle Ka Lai Lee. 2020. Delineating the transnational network agenda-setting model of mainstream newspapers and twitter: A machine-learning approach. *Journalism Studies*, 21(15):2113–2134.

Chris J Vargo, Lei Guo, Maxwell McCombs, and Donald L Shaw. 2014. Network issue agendas on twitter during the 2012 us presidential election. *Journal of communication*, 64(2):296–316.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. 2018. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*.

Sinong Wang, Han Fang, Madian Khabsa, Hanzi Mao, and Hao Ma. 2021. Entailment as few-shot learner. *arXiv preprint arXiv:2104.14690*.

Adina Williams, Nikita Nangia, and Samuel R Bowman. 2017. A broad-coverage challenge corpus for sentence understanding through inference. *arXiv preprint arXiv:1704.05426*.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2020. mt5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*.

Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *arXiv preprint arXiv:1909.00161*.

Jian-Guo Zhang, Kazuma Hashimoto, Wenhao Liu, Chien-Sheng Wu, Yao Wan, Philip S Yu, Richard Socher, and Caiming Xiong. 2020. Discriminative nearest neighbor few-shot intent detection by transferring natural language inference. *arXiv preprint arXiv:2010.13009*.