Social Media Unrest Prediction during the COVID-19 Pandemic: Neural Implicit Motive Pattern Recognition as Psychometric Signs of Severe Crises

Dirk Johannßen
MIN Faculty
Dept. of Informatics
Universität Hamburg
& Nordakademie

Chris Biemann
MIN Faculty
Dept. of Informatics
Universität Hamburg
22527 Hamburg, Germany

`http://lt.informatik.uni-hamburg.de/`
`{biemann, johannssen}@informatik.uni-hamburg.de`

Abstract

The COVID-19 pandemic has caused international social tension and unrest. Besides the crisis itself, there are growing signs of rising conflict potential of societies around the world. Indicators of global mood changes are hard to detect and direct questionnaires suffer from social desirability biases. However, so-called implicit methods can reveal humans intrinsic desires from e.g. social media texts. We present psychologically validated social unrest predictors and replicate scalable and automated predictions, setting a new state of the art on a recent German shared task dataset. We employ this model to investigate a change of language towards social unrest during the COVID-19 pandemic by comparing established psychological predictors on samples of tweets from spring 2019 with spring 2020. The results show a significant increase of the conflict-indicating psychometrics. With this work, we demonstrate the applicability of automated NLP-based approaches to quantitative psychological research.

1 Introduction

The COVID-19 pandemic and the reactions to it have led to growing social tensions. Guitérrez-Romero (2020) studied the effects of social distancing and lockdowns on riots, violence against civilians, and food-related conflicts in 24 African countries. The author found that the risk of riots and violence have increased due to lockdowns. Resistance against national health regulations such as the duty to wear masks are partially met with resistance by movements such as anti-maskers or anti-obligation demonstrations.[1] Even anti-democratic alterations of e.g. services offered by the US Postal Service (USPS) of delivering mail-in ballots for the US presidential elections 2020, which are essential for social distancing measures amidst the pandemic, are being utilized amidst this international crisis and foster social unrest and potential outbursts of violence, civil disobedience or uprisings.[2]

Social media has become an important reflection of nationally and internationally discussed topics, and is a predictor of e.g. stock markets, disease outbreaks or political elections (Kalampokis et al., 2013). The majority of human-produced data exists in textual form and broadly in social media and thus, an the investigation of social unrest and conflict situations from social media becomes a worthwhile application area for natural language processing (NLP) problems (Gentzkow et al., 2019).

When speaking about such global phenomena such as a rise in international social unrest and possible occurrences of conflict reflected in text, the detection of specific keywords or utterances have not been successful in past research. Mueller et al. (2017) utilized Laten Dirichlet Allocation (LDA, (Blei et al., 2003)) topic modelling on war-related newspaper items and were not able to improve predictability from other multi-factor models that take into account e.g. GDP figures, mountainous terrain or ethnic

---

[1]`https://firstdraftnews.org/latest/coronavirus-how-pro-mask-posts-boost-the-anti-mask-movement/`

[2]`https://www.businessinsider.com/trump-walks-back-threat-block-covid-relief-over-usps-funding-2020-8?r=DE&IR=T`

polarization. Furthermore, Chadefaux (2012) showed that news reports on possible war situations alone did not function as good predictors but identified sharp frequency increases before war emerged, possibly helping with just-in-time safety measures but likely failing to avoid war situations altogether.

Alternatively, the risks of escalation could be determined based on politician's personalities and the current mood and tone of utterances (Schultheiss and Brunstein, 2010, p. 407). However, intrinsic desires and personality can hardly be measured directly (see Section 3). Intrinsic or subconscious desires and motivation would more likely correlate with personalities, tone, and thus possibly social unrest.

We hypothesize that the frequency of social unrest predictors have significantly changed in social media textual data during the COVID-19 pandemic drawn from the Twitter 1 percent stream[3] in early 2019 and 2020, whilst linguistic features stay comparably stable and unchanged. With this, we aim to demonstrate a possible transition from laborsome manual psychological research to automated NLP approaches.

After presenting and discussing related work in Section 2, we will first introduce the concept of implicit motives and self-regulating levels in more details in Section 3 and the social unrest predictors thereafter in Section 4. The data utilized for experiments is described in Section 5 and the methodology in Section 6. Thereafter, we will present the results in Section 7 and discuss their impacts in Section 8. Lastly, we will draw a conclusion in Section 9.

2 Related Work

Conflict predictions from natural language are rarely encountered applications and have mainly been about content analysis and less about crowd psychology. Kutuzov et al. (2019) used one-to-X analogy reasoning based on word embeddings for predicting previous armed conflict situations from printed news. Johansson et al. (2011) performed named entity recognition (NER) and extracted events via Hidden Markov Models (HMM) and neural networks, which were combined with human intelligence reports to identify current global areas of conflicts, that, in turn, were utilized mainly for world map visualizations.

Investigation of personality traits has mainly been focussing on so-called explicit methods. For these, questionnaires are filled out either by interviewers, through observations, or directly by participants. One of the most broadly utilized psychometrics is the Big Five inventory, even though its validity is controversial (Block, 1995). The five-factory theory of personality (later named Big Five) identifies five personality traits, namely *openness to experiences*, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism* (McCrae and Costa Jr., 1999; Goldberg, 1981). This Big Five inventory was utilized by Tighe and Chegn (2018) for analyzing these five traits of Filipino speakers.

Some studies perform natural language processing (NLP) for investigating personality traits. Lynn et al. (2020) utilized an attention mechanism for deciding upon important parts of an instance when assigning the five-factor inventory classes. The Myers-Briggs Type Indicator (MBTI) is a broadly utilized adaption of the Big Five inventory, which Yamada et al. (2019) employed for asserting the personality traits within tweets.[4]

The research field of psychology has moved further towards automated language assertions during the past years. One standard methodology is the utilization of the tool linguistic inquiry and word count (LIWC), developed by Pennebaker et al. (1999). The German version of LIWC was developed by Wolf et al. (2008). It includes 96 target classes, some of which are rather simple linguistic features (word count, words longer than six characters, frequency of punctuation), and psychological categories such as anxiety, familiarity, or occupation. Even though the tool appears rather simple from an NLP point of view, it has a long tradition to be utilized for content research in the field of behavioral psychology. Studies utilizing LIWC have shown that function words are valid predictors for long-term developments such as academic success (Pennebaker et al., 2014). Furthermore, it has been shown that LIWC correlates with the Big Five inventory (McCrae and Costa Jr., 1999). Importantly, the writing style of people can be considered a trait, as it has shown high stability over time, which means that it is not dependent on

---

[3]https://developer.twitter.com/en/docs/labs/sampled-stream/overview

[4]A *Tweet* is a short message from the social network microblogging service Twitter (https://www.twitter.com/) and consists of up to 240 characters.

one's current mood, the time of day, or other external conditions (Pennebaker and King, 2000). Hogen-raad (2003) utilized an implicit motive (see Section 3) dictionary approach to automatically determine risks of war outbreaks from different novels and historic documents, identifying widening gaps between the so-called power motive and affiliation motive in near-war situations.

Overall, the work on automated classification of implicit motive data or the use of NLP for the assertion of psychological traits in general is rather sparse or relies on rather outdated methods, as this application domain can be considered a niche (Schultheiss and Brunstein, 2010; Johannßen and Biemann, 2019; Johannßen and Biemann, 2018; Johannßen et al., 2019). One recent event in this area was the GermEval 2020 Task 1 on the Classification and Regression of Cognitive and Motivational Style from Text (Johannßen et al., 2020). The best participating team reached a macro f1 score of 70.40 on the task of classifying implicit motives combined with self-regulating levels, resulting in 30 target classes. However, behavioral outcomes from automatically classified implicit motives have – to our knowledge – not yet been researched.

3 Implicit Motives and Self-regulatory Levels

Implicit motives can reveal intrinsic, unconscious human desires, and thus avoid social desirability biases, which usually are present when utilizing direct questionnaires. They originated from the Thematic Apperception Test (TAT) by Murry et al (1943). Participants are confronted with ambiguous images of multiple people that interact with each other as displayed in Figure 1, and are asked to answer four questions: i) who is the main person, ii) what does that person feel? iii) why does the person feel that way, and iv) how does the story end? From these questions, trained psychologists can assign one of five motives: affiliation (A), freedom (F), achievement (L), power (M), and zero (0). The psychologists follow some rules, one being the so-called primacy rule, where the very first identifiable motive determines the whole instance, despite what follows (Scheffer and Kuhl, 2013). These motives have shown to be behavioral predictors and allow for long-term statements of e.g. group dynamics or success (McClelland and Boyatzis, 1982; Schultheiss and Brunstein, 2010). Implicit motives have been broadly utilized in the 1980s but at the cost of laborsome manual annotating processes. It takes about 20 hours of training for an annotator to encode one of the implicit motives. Skilled human annotators take up to 50 hours per 100 participants. This costly annotating process has hampered this once-promising psychometric (Schultheiss and Brunstein, 2010, p. 140).

Whilst the classification performance of implicit motive models have been explored and achieved high results (e.g. (Johannßen and Biemann, 2019; Johannßen et al., 2020)), behavioral consequences and mass phenomena from automated labeled textual instances have barely been researched.

In addition to the implicit motives, the data set from the GermEval20 Task 1 comes with so-called levels per textual instance. The levels were developed by Kuhl (2001). They describe the self-regulatory enactment in five dimensions. According to Scheffer and Kuhl (2013) the 1st level is the ability to self-regulate a positive affect, the 2nd is the sensitivity for positive incentives, the 3rd self-regulates a negative affect, the 4th is the sensitivity for negative incentives and the 5th level describes the passive coping with fears. In other words, these levels help to identify the type of the participant's emotional response according to the identified implicit motive.

As with many psychometrics, the reliability of implicit motives, and especially their predecessor (the TAT) is controversial. One main point of criticism is that implicit motives do not correlate significantly with so-called explicit motives. Whilst implicit motives try to measure intrinsic desires indirectly by asking participants associative questions, explicit motives try to measure desires via direct questionnaires. In psychology, reliability means, that personality traits revealed by one measure may not conflict with personality traits measured by other, well-established measures. Since the measured desires of implicit and explicit motives generally do not match, the reliability of implicit motives is said to be weak.

Schultheiss et al. (2010) explain this lack of reliability and correlation with the fact that explicit implicit motives are by definition of different measurements that can not be directly compared. Whilst implicit motives measure intrinsic desires that are subconscious, explicit motives are more influenced by a social expectation bias (i.e. what do participants think is a socially sound and accepted answer to a

question) and thus are closer connected to behaviorism. Nonetheless, reliability in psychological research demands different observable results of metrics to be coherent (Reuman, 1982) when the descriptions of what a psychometric is supposed to measure matches (i.e. desires).

Another point of criticism is the way the TAT images are selected. They emerge from an empirical study, where participants are shown different images. Only when past frequencies of motives are achieved with an image, this image gets added to the available testing stock. With this, however, the very first selected implicit motive images could not have been validated. Nowadays, many scholars argue that the amounts of positive evidence legitimize this methodology, but it has yet to be resolved (Hibbard, 2003).
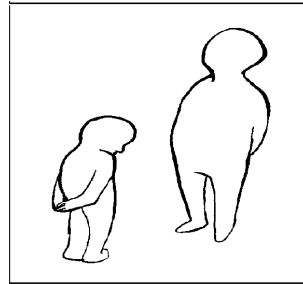

Figure 1: Exemplary image to be interpreted by participants utilized for the operant motive test (OMT). Identifiable motives are the affiliation motive (A), the power motive (M), achievement (L), and freedom (F). A 0 represents the unassigned motive (Kuhl and Scheffer, 1999).

The transition from natural language to intrinsic desires and motivation is not trivial, as humans do not express intrinsic and unconscious desires unfiltered and directly. As soon as a direct questionnaire is involved, social desirability biases (i.e. thoughts of publicly expected answers) alter an uninfluenced introspection (Brunstein, 2008). Such direct questionnaires are called explicit methods, in contrast to implicit methods, such as e.g. the TAT and subsequent tests produced by image descriptions.

4   Social Unrest Predictors

Times of severe social unrest are reflected by distinct patterns of implicit motives and linguistic features. Winter (2007) surveyed multiple prior studies, identifying three main predictors: *responsibility*, *activity inhibition*, and *integrative complexity*, displayed in Table 1. In this study, the author identified and analyzed 8 occurrences of crises and social unrest by examining influential political speeches of this time. Thereafter, the outcomes of these crises – whether they ended peacefully or in conflict – were projected on indicators from earlier research.

Winter and Barenbaum (1985) found that the power motive (M) has a moderating effect of responsibility. In other words, responsibility determines, how vast amounts of power motivated expressions are behaviorally enacted. If a high responsibility score is measurable, power motivated individuals act pro-social. On the contrary, if the responsibility score is low, aggression and lack of leadership are to be expected.

Activity inhibition is reflected, according to by McClelland et al. (1972) as the frequency of "not" and "-n't" contradictions in TAT or other verbal texts. Activity inhibition functions as motivational and emotional regulation. The authors identified a negative correlation between activity inhibition and male alcohol consumption. Combined with a high power motive (M) and low affiliation motive (A), subsequent research by McClelland and his colleagues revealed a so-called leadership motive pattern (LMP) (McClelland and Boyatzis, 1982; McClelland, 1988). The higher this LMP, the more responsible leaders act. As for *integrative complexity* it was observed, that the lower the frequency of utterances in accordance to the 7-point score was (see Table 1), the more likely escalations became.

Category	Measure	Example or Explaination
Responsibility		
i) moral standards	observable, if people, actions, or things are described with either morality or legality	'she wants to do the right thing'
ii) obligation	means, that a character in a story is obliged to act because of a rule or regulation	'he broke a rule'
iii) concern for others	emerges, when a character helps or intends to help others or when sympathy is shown or thought	'the boss will understand the problem and will give the worker a raise'
iv) concerns about consequences	can be identified when a character is anxious or reflects upon negative outcomes	'the captain is hesitant to let the man on board, because of his instructions'
v) self-judgment	scores when a character critically judges his or her value, morals, wisdom, self-control, etc. and has to be intrinsic	'the young man realizes he has done wrong'
Activity inhibition		
linguistic negation	in English terms, the authors describe activity inhibition as the frequency of "not" and "-n't"	responsibility measure, e.g. a variable negatively correlated with male alcohol consumption
leadership motive pattern (LMP)	combined with a high power motive (M) and low affiliation motive (A)	predicts responsible leadership power behaviors instead of profligate impulsive expressions of power
Integrative complexity		
7-point continuum range score from simplicity to	1: no sign of conceptual and differentiation or integration can be observed	only one solution is considered to be legitimate
differentiation and integration	7: overreaching	viewpoints are expressed, involving different relationships between alternate perspectives

Table 1: According to Winter (2007), some distinct psychometrics and their combinations predict social unrest – namely responsibility, activity inhibition, and integrative complexity. The table shows their categories, measurements and offer examples or explanations. Responsibility is measured with a dedicated TAT, activity inhibition (AI) and integrative complexity is determined via content analysis. Especially the combination of low responsibility, high activity inhibition and little integrative complexity (e.g. high frequency of the power motive combined with the self-regulatory 4th level) predict situations of social unrest with negative escalatory outcomes.

5 Training and experimental data

For testing the proposed hypothesis (Section 1), we first train a classification model and utilize this model for testing social network textual data. In this section, we will describe the two different data sources for training and the experiments.

5.1 Model Training Data

The data utilized for training models were made available by the organizers of the *GermEval-2020 Task 1 on the Classification and Regression of Cognitive and Motivational Style from Text*[5][6] (Johannßen et al., 2020). The training set consists of 167,200 unique answers, given by 14,600 participants of the OMT (see Section 3). The training data set is imbalanced. The power motive (M) is the most frequent class, covering 41.02% of data points. The second most frequent class, achievement (L) only accounts for 19.63% and thus is half as frequent as M. The training data was assembled and annotated by the University of Trier, reaching a pairwise annotator intraclass correlation of r = .85. With only 22 words on average per training instance (i.e. a participant's answer) and a standard deviation of 12 words, training a classifier on this data is a short text classification task (Johannßen et al., 2020).[7]

5.2 Experimental Data

The experimental data was collected before this work by crawling the Twitter API and fetching 1 percent of the worldwide traffic of this social network (Gerlitz and Rieder, 2013). We sample posts over the time window from March to May of both, 2019 and 2020. There are no apparent linguistic differences between the two samples. The average word count, part-of-speech (POS) tags, sentence length, etc. are comparable.

Thereafter, we extracted the *text* and *date time* fields of posts marked as German. From those files hashtags, name references (starting with '@'), corrupted lines, and any post shorter than three content words were removed. The resulting files for 2019 and 2020 contained more than 1 million instances. Lastly, the instances were randomly shuffled. We drew and persisted 5,000 instances per year for the experiments, as this data set size is large enough for producing statistically significant results. The posts on average consist of 11.97 (2019) and 11.8 (2020) words per sentence, and thus are very short. During the experiments, further pre-processing steps were undertaken, which are described in Section 6. By stretching out the data collection time window and by comparing the same periods in two subsequent years, we aim to reduce any bias effect that might impact Twitter user behavior over short periods, e.g. the weather, any sports event, or short-lived political affairs.

6 Methodology for Implicit Motive Classification Social Unrest Prediction

For constructing a model of sufficient quality to test our hypothesis, we follow Johannßen and Biemann (2019) and train a long short-term memory network (LSTM, (Hochreiter and Schmidhuber, 1997)) combined with an attention mechanism.

An LSTM is a special type of recurrent neural network (RNN). An RNN not only has connections between units from layer to layer but also between units of the same layer. Furthermore an LSTM also has a mechanism called the *forget gate*, allowing the structure to determine which information to keep and which information to forget during the training process. The attention mechanism (Young et al., 2018) can capture the intermediate importance of algorithmic decisions made by the network. It can be employed for enhanced results but also investigated for researching algorithmic decisions. However, it is debated upon, whether this algorithmic importance can serve as an explanation. Even though oftentimes, the algorithmic importance is correlated with an explanation for the task (i.e. does a model for image recognition of animals *look* at the animals or the backgrounds of the images?), there are cases, where algorithmic importance and explanation for the task differ (Jain and Wallace, 2019; Wiegreffe and Pinter, 2019)). Since automatically labeling implicit motives is a sequential problem revolving around identifying the first verbal enactment of a motive (see Section 3, we decided to employ a Bi-LSTM with an attention mechanism (Schuster and Paliwal, 1997).

---

[5]GermEval is a series of shared task evaluation campaigns that focus on Natural Language Processing for the German language.

[6]https://www.inf.uni-hamburg.de/en/inst/ab/lt/resources/data/germeval-2020-cognitive-motive.html

[7]The data can be retrieved via https://www.inf.uni-hamburg.de/en/inst/ab/lt/resources/data/germeval-2020-cognitive-motive.html

We decided against additional features such as part of speech (POS) tags or LIWC features like in our previous work (Johannßen and Biemann, 2019), as we did not reach the best results with these additional features. The maximum token length was set to 20, as determined by preliminary experiments (Johannßen and Biemann, 2019), and reflects the primacy rule of the implicit motive theory described in Section 3. The average answer length of the training data set was 22 tokens (see Section 5). With this decision to limit the considered tokens, we aim to closely replicate the implicit motive coding practices manually performed by trained psychologists (Kuhl and Scheffer, 1999). Accordingly, it is preferable to assign the 0 motive (i.e. no clear motive could be identified) than to falsely assign a motive that is not the very first one in the sequence.

Some standard pre-processing steps were applied to reduce noise, which was to remove the Natural Language Toolkit (NLTK) German corpus stop words[8], to lowercase the text, remove numbers, normalize special German letters (i.e. umlaute). Emojis were removed as well, since Twitter offers a selection of 3,348 emojis[9] , that in turn mainly do not capture sufficient informational gain per textual answer for the task at hand. To remove stop words has to be an informed choice when it comes to performing NLP on psychological textual data. For example, function words are said to predict academic success (see Section 2). However, during our experiments, we saw an increase in model performance when stop words were removed.

After the training, we utilize the model on the two sampled data sets described in Subsection 5.2. According to our hypothesis in Section 1 and following the theories in Section 4, we investigate the frequency of the power motive with the self-regulatory level 4, which we expect to be higher. At the same time, we will also analyze the other motives and levels to see which ones are now less frequent and to what extent. Furthermore, we compare different linguistic features and statistics from 2019 to 2020 to see, if any of these show differences that might indicate possible biases in the data.

Our Bi-LSTM model was set to be trained within 3 epochs and with a batch size of 32 instances. The model was constructed having 3 hidden layers and utilized pre-trained fasttext embeddings (Bojanowski et al., 2017), as this character-based or word fragment-based language representation has shown to be less prone to noisy data and words that have not been observed yet like e.g. spelling mistakes or slang – both often observable in social media data. The fasttext embeddings had 300 dimensions and were trained on a Twitter corpus, ideally matching the task at hand.[10] Explorative experiments with different parameter combinations have shown that a drop-out rate of .3 and step width of .001 produced good results.

The cross-entropy loss was reduced rather quickly and oscillated at 1.1 when we stopped training early during the second epoch. After each epoch, the model was evaluated on a separate development test. After the training was finished, the model was tested once on the GermEval20 Task 1 test data and with the official evaluation script. This provides the chance to compare the achieved results with the best-participating team. Schütze et al. (2020) achieved a macro f1 score of 70.40, which our Bi-LSTM model was able to outperform with an f1 score of 74.08, setting a new state of the art on this dataset.

7 Results

After having trained the Bi-LSTM model and sampled the experimental data, we will describe the results and findings of the conducted Twitter COVID-19 experiments in this section. An overview of all results is displayed in Table 2.

To investigate the main predictor for social unrest *activity inhibition* (see Section 4), the power motive (M) in combination with level 4 was counted. The self-regulatory level 4 describes the sensitivity for negative incentives (see Section 3). These measures are collected for all four data sets. Our Bi-LSTM model assigned power 4 in 33.76% of all cases for the Twitter sample from March to May of 2019, making this the most frequent label. However, for the data sample from 2020, power 4 is as frequent as

---

[8]The Natural Language Toolkit (NLTK) is a collection of python libraries for NLP `https://www.nltk.org/`.

[9]`https://emojipedia.org/twitter/twemoji-13.0.1/`

[10]The fasttext model was obtained from Spinningbytes at `http://spinningbytes.ch/resources/wordembeddings`

Metric	2019	2020	Percentage delta	Significance
Activity inhibition and responsibility				
Power 4	33.76	37.40	10.97	p<.01***
LIWC Family	.08	.05	-37.60	p<.05*
LIWC insight	.23	.17	-26.09	p<.05*
Implicit motives				
Power motive	65.84	68.24	3.64	p<.01***
Freedom motive	20.28	17.72	-12.63	p<.01***
Achievement motive	6.80	7.00	2.94	p>.05
Affiliation motive	2.00	1.86	-7.00	p>.05
Null motive	5.10	5.10	.00	p>.05
Self-regulatory levels				
Level 1	6.50	6.01	-7.54	p>.05
Level 2	2.76	3.26	18.12	p>.05
Level 3	27.20	25.58	-5.96	p<.01***
Level 4	42.78	45.20	5.67	p<.01***
Level 5	15.86	14.92	-5.93	p<.05*
Linguistic statistics				
Average words	11.97	11.80	-1.42	p>.05
Verbs	1.19	1.22	2.52	p>.05
Adjectives	.43	.43	.00	p>.05
Words >6 letters	38.65	38.86	.54	p>.05

Table 2: Overview of the different psychometric and statistical results. * represents significant results, *** represents highly significant results. All combinations of motives and levels have been examined. Note that most motives, levels, and statistical values stay constant. However, power 4 is more frequent, whilst the freedom motive is less. As the linguistic statistic metrics stay relatively stable, this indicates no observable sampling bias.

37.4%, making this an increase of 10.97%. For calculating the significance of this rise, we perform a t-test on the label confidences for the power motive with self-regulatory level 4 for both, 2019 and 2020 with the 5,000 samples from each year (see Section 5).

The two-sample t-test on the confidence levels shows, that the rise in frequency is statistically significant ($p < 0.05$ with $\bar{x}_1 = .27$, $\bar{x}_2 = .29$, $\sigma_1 = .28$, $\sigma_2 = .28$, $N_1 = 5,000$ and $N_2 = 5,000$).

The affiliation motive (A) is barely classified, covering only 2% (2019) and 1.89% (2020) of all instances. The slight decrease is not statistically significant ($p > .05$). The frequency of self-regulatory level 4 is elevated by 6.7%. The whole of all assigned power motive labels has only risen by 3.64%, both having risen less than the combination of the power motive and level 4 combined. The strongest decline in frequency can be measured for the freedom motive with -12.63%. The other motives of affiliation, achievement, and null have barely changed in comparison to 2019 with 2020. The same holds for the average amounts of words per sentence, verbs, adjectives, and words containing at least 6 letters, all of which have barely changed, not indicating sampling biases. An overview of the class frequencies is provided in Table 3.

Since both, responsibility and integrative complexity can only be measured by employing a specific TAT and a questionnaire, which would have to be performed with each Twitter user, we can only investigate activity inhibition as a combination of the power motive with the self-regulatory level 4. However, we will review some psychological LIWC categories, that follow a close description as the five categories

Implicit motive	Frequency	Self-regulatory level	Frequency
2019			
Power	3,251	1	492
Affiliation	141	2	193
Achievement	414	3	1,487
Freedom	9622	4	1,872
Zero	232	5	724
		0	232
2020			
Power	3,433	1	316
Affiliation	90	2	151
Achievement	203	3	1,259
Freedom	923	4	2,233
Zero	761	5	780
		0	261

Table 3: Overview of the class frequencies.

of Winter's responsibility scoring system (Winter and Barenbaum, 1985). Relevant LIWC categories for the responsibility is the combination of *family*, which are terms connected to expressions like 'son' or 'brother', and *insight*, which contain expressions such as 'think' or 'know', representing self-aware introspection. Family shows a significant decrease from 2019 (0.08) to 2020 (0.05) of -37.5%. The frequency of insight terms fell from 2019 (.23) to 2020 (.17) by -26%, all of which are statistically significant changes ($p < 0.05$ for both categories).

8 Discussion

We hypothesized that the social unrest predictors by Winter (2007), namely *activity inhibition*, *responsibility*, and *integrative complexity* are automatable and reveal changes in natural language and signs of social unrest observable through the use of social media textual data connected to the COVID-19 pandemic.

The main research objective of this work is to find novel approaches to automatically provide the community with red flags for growing tensions and signs of social unrest via social media textual data. For this, *activity inhibition* is the main predictor. It consists of a distinct shift in implicit motives. It is present when the frequency of the power motive with the self-regulating level 4 (sensitivity for negative incentives, see Section 3) is elevated and the affiliation motive is suppressed – even though Winter (2007) did not find clear evidence of the latter. The comparable rise by 10.97% ($p < 0.01$) is an indicator of the social tension of COVID-19 related social media posts.

Since other linguistic statistics, such as the average amounts of adjectives, verbs, words per sentence, or words containing at least 6 letters have barely changed, this indicates that the measurable differences in social unrest predictors are content-based and not due to linguistic biases.

It is remarkable, that whilst the power motive has been labeled more frequently, the frequency of the labeled freedom motive has declined by -12.63% from 2019 to 2020. This freedom motive has barely been researched yet but has a close connection to the power motive. Whilst power-motivated individuals desire control over their fellow humans and their direct surrounding for the sake of control, freedom-motivated individuals seek to express themselves and want to avoid any restraining factors. Motives are said to be rather stable but can change over time (Schultheiss and Brunstein, 2010). This change in motive direction could indicate a roughening of verbal textual content and interpersonal communication. Example utterances classified as freedom and power from 2019 compared with 2020 are displayed in Table 4.

The change of responsibility, as reflected in LIWC categories, retreated by roughly 30% from 2019 to 2020. This responsibility indicates a personal involvement in topics and decisions, that we feel are relevant for our surroundings. If this involvement diminishes, our interest in participating in constructive solutions to problems does as well.

```
M 'RT @FrauLavendel: ist es wahr        M Corona-Regeln im Saarland sind zum
  dass schulleitungen den                 Teil absurd und unverhältnismäßig
  schüler*innen drohen                   F RT @kattascha: In den USA bekommen
F RT @UteWeber: Nach einem relativ        viele Menschen keine Lohnfortzahlung
  unfeierlichen,                          im Krankheitsfall. Das bedeutet:
  regionalen Offline-Tag aufs Sofa        Selbst bei Verdacht auf #COVID19 w...
  sinken, wie von der Tarantel           A Wer einen Discord-Server sucht,
  gestochen aufspringen und zur...         um entspannt mit seinen Kollegen
A Weltbestseller "P.S. Ich liebe          zu zocken oder gemeinsam abzuhängen
  dich" bekommt einen zweiten Teil        ist hier genau...
  https://t.co/9Ifl5CrNAP               ------- Translation -----
  ------- Translation -----             M the Corona rules for the Saarland
M 'RT @FrauLavendel: is it true that      are partially absurd and dis-
  principals threatens students           proportionate
F RT @UteWeber: after a relatively      F RT @kattascha: in the US a lot
  un-celebrational, regional offline      of people don't receive continued
  day, as bitten by a tarantula           pay in case of illness. That means:
  jumping up                              even in case of suspected #COVID19
A world best-selling book "P.S. Ich    A Whoever is looking for a Discord
  liebe dich" gets a second part          server for enjoyably game with
  https://t.co/9Ifl5CrNAP                 their colleagues or chill together,
                                          is in the right place...
```

Table 4: Some example tweets from 2019 (left) compared with 2020 (right). Whilst the power motive (M) is more frequent in 2020, the freedom motive (F) became less frequent. The affiliation motive (A) was very infrequent in both, 2019 and 2020. This signature indicates increased social unrest in 2020 in comparison with 2019.

9 Conclusion and Outlook

With this work, we conducted a first attempt at automating psychometrics for investigating social unrest in social media textual data. The Bi-LSTM model combined with an attention mechanism of this work achieved an f1 score of 74.08 on 30 target classes, making it state of the art on a respective recent shared task dataset. With this model, we measured a statistically significant rise in the power motive with self-regulating level 4, which reflects the social unrest predictor of *activity inhibition* in the direct comparison of the samples from March to May of 2019 vs. 2020.

Furthermore, we investigated *responsibility*, which shows significant reductions during the COVID-19 pandemic, hinting at negative outcomes of interpersonal and verbal communication on the social media platform Twitter.

This first approach most likely does not qualify for a real-world social prediction system. Predictions of such a system can not yet be reliable enough for deriving necessary actions from them. On the upside, implicit motives do not only qualify for examining general socio-economic tensions, but can be applied on an individual or small group scale. As an example, detecting tensions within a small group can help to shape the group and guiding it into a better fit. Furthermore, we advocate for combining implicit motives with sufficiently many complementary psychometrics and content-based analysis e.g. sentiment analysis, topic modeling, or emotion detection.

Besides those combinations with other information sources for future work, different sampling approaches and larger data set sizes should be utilized for reproducing findings and research correlations with other social unrest predictors and indicators. In this work, we have made the first steps towards understanding the automation of psychological findings. Since only 5,000 samples were drawn from a single social network platform, we advocate for broadening this approach to include many more samples from wider time windows paired with mixing the data sources. In addition to that, deeper investigations into the linguistic variances between times of so-called social unrest and more peaceful times should be performed, as those could reveal patterns and characteristics of time-specific utterances.

Even though this work is only introductory, the observed correlations and social unrest patterns are in line with an intuitive assumption of how language in social media data changes amid a pandemic. Future work arises in the application of this methodology on other events and crises, eventually providing a quantitative basis for implicit motive research.

References

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

Jack Block. 1995. A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117(2):187–215.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Joachim C. Brunstein. 2008. Implicit and explicit motives. *Motivation and Action*, pages 227–246.

Thomas Chadefaux. 2012. Early Warning Signals for War in the News. *Journal of Peace Research*, 51(1).

Matthew Gentzkow, Bryan Kelly, and Matt Taddy. 2019. Text as Data. *Journal of Economic Literature*, 57(3):535–574.

Carolin Gerlitz and Bernhard Rieder. 2013. Mining One Percent of Twitter: Collections, Baselines, Sampling. *M/C Journal*, 16(2), nr.620.

Lewis R. Goldberg. 1981. Language and individual differences: The search for universals in personality lexicons. *Review of personality and social psychology*, 2(1):141–165.

Roxana Gutiérrez-Romero. 2020. Conflict in Africa during COVID-19: social distancing, food vulnerability and welfare response. *SSRN Electronic Journal*. ID 3616421.

Stephen Hibbard. 2003. A critique of Lilienfeld et al.'s (2000) "The scientific status of projective techniques". *Journal of Personality Assessment*, 80(3):260–271.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-term Memory. *Neural computation*, 9:1735–1780.

Robert Hogenraad. 2003. The Words that Predict the Outbreak of Wars. *Empirical Studies of the Arts*, 21:5–20.

Sarthak Jain and Byron C. Wallace. 2019. Attention is not Explanation. *arXiv:1902.10186 [cs]*.

Dirk Johannßen and Chris Biemann. 2018. Between the Lines: Machine Learning for Prediction of Psychological Traits - A Survey. In *Proceedings of the International Cross-Domain Conference*, pages 192–211, Hamburg, Germany. Springer.

Dirk Johannßen and Chris Biemann. 2019. Neural classification with attention assessment of the implicit-association test OMT and prediction of subsequent academic success. In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Long Papers*, pages 68–78, Erlangen, Germany. German Society for Computational Linguistics & Language Technology.

Dirk Johannßen, Chris Biemann, and David Scheffer. 2019. Reviving a psychometric measure: Classification of the Operant Motive Test. In *Proceedings of the Sixth Annual Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*, pages 121–125, Minneapolis, MN, USA. Association for Computational Linguistics.

Dirk Johannßen, Chris Biemann, Steffen Remus, Timo Baumann, and David Scheffer. 2020. GermEval 2020 Task 1 on the Classification and Regression of Cognitive and Motivational style from Text. In *Proceedings of the GermEval 2020 Task 1 Workshop in conjunction with the 5th SwissText & 16th KONVENS Joint Conference 2020*, pages 1–10, Zurich, Switzerland (online). German Society for Computational Linguistics & Language Technology.

Fredrik Johansson, Joel Brynielsson, Pontus Hörling, Michael Malm, Christian Mårtenson, Staffan Truvé, and Magnus Rosell. 2011. Detecting Emergent Conflicts through Web Mining and Visualization. In *Proceedings - 2011 European Intelligence and Security Informatics Conference, EISIC 2011*, pages 346 – 353, Athens, Greece. Springer.

Evangelos Kalampokis, Efthimios Tambouris, and Konstantinos Tarabanis. 2013. Understanding the Predictive Power of Social Media. *Internet Research*, 23(5).

Julius Kuhl and David Scheffer. 1999. *Der operante Multi-Motiv-Test (OMT): Manual [The operant multi-motive-test (OMT): Manual]*. Impart, Osnabrück, Germany: University of Osnabrück.

Julius Kuhl. 2001. *Motivation und Persönlichkeit: Interaktionen psychischer Systeme*. Hogrefe Verlag, Göttingen, Germany.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2019. One-to-X Analogical Reasoning on Word Embeddings: a Case for Diachronic Armed Conflict Prediction from News Texts. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 196–201, Florence, Italy. Association for Computational Linguistics.

Veronica Lynn, Niranjan Balasubramanian, and Hansen A. Schwartz. 2020. Hierarchical Modeling for User Personality Prediction: The Role of Message-Level Attention. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5306–5316, Online. Association for Computational Linguistics.

David C. McClelland and Richard Boyatzis. 1982. Leadership Motive Pattern and Long-Term Success in Management. *Journal of Applied Psychology*, 67:737–743.

David C. McClelland and William N. Davis. 1972. *The Drinking Man: Alcohol and Human Motivation*. Free Press, New York.

David C. McClelland. 1988. *Human Motivation*. Cambridge University Press.

Robert R. McCrae and Paul T. Costa Jr. 1999. A Five-Factor theory of personality. In *Handbook of personality: Theory and research, 2nd ed*, pages 139–153. Guilford Press, New York, NY, US.

Hannes Mueller and Christopher Rauh. 2017. Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. *American Political Science Review*, 112:1–18.

Henry A. Murray. 1943. *Thematic Apperception Test*. Harvard University Press.

James W. Pennebaker and Laura King. 2000. Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77:1296–312.

James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 1999. Linguistic inquiry and word count (LIWC). *Software manual*.

James W. Pennebaker, Cindy K. Chung, Joey Frazee, Gary M. Lavergne, and David I. Beaver. 2014. When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLOS ONE*, 9(12):e115844.

David A. Reuman. 1982. Ipsative behavioral variability and the quality of thematic apperceptive measurement of the achievement motive. *Journal of Personality and Social Psychology*, 43(5):1098–1110.

David Scheffer and Julius Kuhl. 2013. *Auswertungsmanual für den Operanten Multi-Motiv-Test OMT*. sonderpunkt Verlag, Münster, Germany.

Oliver C. Schultheiss and Joachim C. Brunstein. 2010. *Implicit Motives*. Oxford Univ Pr, Oxford ; New York, 1 edition.

Mike Schuster and Kuldip K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.

Henning Schäfer, Ahmad Idrissi-Yaghir, Andreas Schimanowski, Michael R. Bujotzek, Hendrik Damm, Jannis Nagel, and Christoph M. Friedrich. 2020. Predicting Cognitive and Motivational Style from German Text using Multilingual Transformer Architecture. In *Proceedings of the GermEval 2020 Task 1 Workshop in conjunction with the 5th SwissText & 16th KONVENS Joint Conference 2020*, pages 17–22, Zurich, Switzerland (online). German Society for Computational Linguistics & Language Technology.

Edward Tighe and Charibeth Cheng. 2018. Modeling Personality Traits of Filipino Twitter Users. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 112–122, Louisiana, LA, USA. Association for Computational Linguistics.

Sarah Wiegreffe and Yuval Pinter. 2019. Attention is not not Explanation. *arXiv:1908.04626 [cs]*. arXiv: 1908.04626.

David G. Winter and Nicole B. Barenbaum. 1985. Responsibility and the power motive in women and men. *Journal of Personality*, 53(2):335–355.

David G. Winter. 2007. The Role of Motivation, Responsibility, and Integrative Complexity in Crisis Escalation: Comparative Studies of War and Peace Crises. *Journal of personality and social psychology*, 92:920–37.

Markus Wolf, Andrea B. Horn, Matthias R. Mehl, Severin Haug, James W. Pennebaker, and Hans Kordy. 2008. Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count. *Diagnostica*, 54(2):85–98.

Kosuke Yamada, Ryohei Sasano, and Koichi Takeda. 2019. Incorporating Textual Information on User Behavior for Personality Prediction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 177–182, Florence, Italy. Association for Computational Linguistics.

Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. *IEEE Computational Intelligence Magazine*, 13:55–75.