

Диагностика когнитивных искажений в публичных, групповых и персональных текстовых коммуникациях

Ариничева А.^{1*}, Колонин А.^{1, 2}

¹ Новосибирский государственный университет, Новосибирск, Россия

² Aigents, Новосибирск, Россия

* a.arinicheva@g.nsu.ru

Ключевые слова: когнитивные искажения; поиск оптимальных параметров; анализ данных

Мотивация и цель: Улучшение качества жизни среднестатистического человека, нестабильная ситуация в мире и широкая доступность интернет-ресурсов привели в первую очередь к резкому увеличению количества информации, которую потребляет человек в течение дня, включая как проверенные факты, так и ложные утверждения, в числе которых манипулятивные тексты, направленные на умышленное искажение картины мира и психического состояния их потребителя. Виртуальное пространство усугубляет трудности отличия достоверной информации от фальсификации из-за отсутствия возможности наблюдения за мимикой, жестами и интонацией собеседника, что характерно для личного общения. В такой ситуации человеку, столкнувшемуся с вызывающей сомнения новостью, способен помочь ассистент, который проанализирует содержание статьи и представит результат в виде вывода по тексту о наличии когнитивных искажений. Кроме того, XX в. дал нам существенные предпосылки к развитию психологии – труды Зигмунда Фрейда оказали большое влияние как на психоанализ, так и на людей, а после Второй мировой войны возникла проблема психологии посттравматического стрессового расстройства (ПТСР). В работе [1] показан тренд увеличения степени выраженности известных в когнитивно-поведенческой терапии когнитивных искажений в доступных публично текстах. В работах [2] и [3] показана выраженная связь между проявлением как эмоциональной составляющей, так и когнитивных искажений с движениями курсов криптовалют, что отражает степень воздействия манипулятивных воздействий на социум. Все это указывает на необходимость изучения психологии и степени влияния текстовых манипуляций на общество в целом, а также поиска методов противодействия этим манипуляциям. Сейчас в связи с доступностью большого количества информации психологические исследования стали открыты широкой общественности, благодаря чему люди начали обращать внимание на свое психологическое здоровье и, как следствие, чаще обращаться за помощью к специалистам. Таким образом, нагрузка на врачей возросла, а принципы диагностики остались прежними. Мы предлагаем алгоритм, который за счет анализа текста на наличие в них когнитивных искажений мог бы помочь в анализе публикаций в интернете и средах электронной коммуникации (мессенджерах), выявляя соответствующие манипулятивные воздействия, содержащиеся в этих текстах. Указанный алгоритм мог бы применяться для мониторинга общественного здоровья на основе публичных коммуникаций, как показано в работе [1], а также для персональной индикации потенциальных манипуляций в групповых коммуникациях (мессенджерах). Также в этой ситуации есть способ помочь

психотерапевту, создав ассистента, который, разговаривая с пациентом, проводил бы анализ когнитивных искажений в его речи и на основе полученных данных помогал бы специалисту ставить верный диагноз и отслеживать эффективность терапии, как предложено в работе [4]. В нашей работе мы представляем указанный алгоритм и проводим поиск и анализ оптимальных параметров для повышения его эффективности.

Методы и алгоритмы: Чтобы решить указанную задачу, мы создали набор словарей на английском языке, которые содержат в себе ключевые слова и фразы, характерные для каждого из выбранных нами когнитивных искажений на основе работы [1]. Далее описанный в работе [2] алгоритм был использован для анализа текста на предмет выраженности в нем соответствующих искажений. Он разбивает входные предложения на n -граммы и с учетом приоритетности, определяемой как порядок n -граммы N , проверяет их наличие в соответствующих словарях. Учет приоритетности представляет собой правило, по которому n -граммы меньшего размера, отвечающие одному когнитивному искажению и входящие в n -грамму большего размера, отвечающую другому когнитивному искажению, теряют свой вес, и выражение однозначно характеризуется когнитивным искажением большей n -граммы. После этого анализа алгоритм показывает коэффициент выраженности каждого искажения в тексте от 0 до 1. Затем мы делаем бинарную классификацию с порогами, установленными с шагом 0.1 и сравниваем полученные результаты при помощи F метрики с переменным параметром β с найденными по нашей тематике наборами данных.

Результаты и выводы: Мы выяснили, что лучше и универсальнее всего себя показывает среднее по всем искажениям с порогом 0.1 и F метрикой с параметром 0.1, что обусловлено наличием в датасетах большего числа текстов, содержащих искажения, чем нейтральных выражений. До порога 0.5 для F метрики с параметрами 0.1–0.5 также хорошо себя показывают искажения “negative statements”, “positive statements”, “rude statements” и “should statements”, что обусловлено неравномерным распределением выраженности когнитивных искажений в исследуемых наборах данных. С результатами можно ознакомиться по ссылке: <https://github.com/AnnaArinicheva/CognitiveProject.git>

Diagnosis of cognitive distortions in public, group, and personal text communications

Arinicheva A.^{1*}, Kolonin A.^{1, 2}

¹ Novosibirsk State University, Novosibirsk, Russia

² Aigents, Novosibirsk, Russia

* a.arinicheva@g.nsu.ru

Key words: cognitive distortions; search for optimal parameters; data analysis

Motivation and Aim: The enhancement of the average person's quality of life, the unstable global situation, and the broad availability of internet resources have primarily led to a sharp increase in the volume of information that people consume daily. This includes both verified facts and false claims, among which are manipulative texts aimed at deliberately distorting the world view and the mental state of their consumers. The virtual space exacerbates the difficulty of distinguishing authentic information from

falsification due to the absence of the ability to observe facial expressions, gestures, and intonation, which are characteristic of personal communication. In such situations, an assistant who can analyze the content of an article and present the results as an analysis of the text for cognitive distortions can help those faced with dubious news. Moreover, the 20th century provided significant groundwork for the development of psychology – the works of Sigmund Freud had a major impact on both psychoanalysis and people, and the issue of post-traumatic stress disorder (PTSD) emerged after World War II. Study [1] showed a trend of increasing expression of cognitive distortions known in cognitive-behavioral therapy in publicly available texts. Studies [2] and [3] demonstrated a pronounced correlation between the manifestation of both emotional components and cognitive distortions with the movements of cryptocurrency rates, reflecting the degree of manipulative impacts on society. All this indicates the necessity to study psychology and the degree of influence of textual manipulations on society as a whole, as well as searching for methods to counter these manipulations. Currently, with the availability of a large amount of information, psychological research has become accessible to the general public, as a result, people have started paying attention to their psychological health and, consequently, more often seek help from specialists. Thus, the burden on doctors has increased, while the principles of diagnosis have remained the same. We propose an algorithm that, through the analysis of text for cognitive distortions, could help in analyzing publications on the Internet and in electronic communication environments (messaging), identifying corresponding manipulative impacts contained in these texts. The specified algorithm could be used both for monitoring public health based on public communications, as shown in study [1], and for personal indication of potential manipulations in group communications (in messengers). Also, in this situation, there is a way to help the psychotherapist by creating an assistant who, by conversing with the patient, would analyze cognitive distortions in their speech and based on this data help the specialist make accurate diagnoses and monitor the effectiveness of therapy, as proposed in study [4]. In this work, we present the mentioned algorithm and conduct a search and analysis of optimal parameters to enhance its effectiveness.

Methods and Algorithms: Initially, we created a set of dictionaries in English that contain key words and phrases characteristic of each of the cognitive distortions we selected based on study [1]. Next, the algorithm described in study [2] was used to analyze the text for the degree of expression of the corresponding distortions. It breaks down input sentences into n-grams and, considering the priority determined as the order of n-gram N , checks their presence in the corresponding dictionaries. Priority consideration is a rule whereby smaller-sized n-grams corresponding to one cognitive distortion and included in a larger n-gram corresponding to another cognitive distortion lose their weight, and the expression is unequivocally characterized by the cognitive distortion of the larger n-gram. After this analysis, the algorithm displays the degree of expression of each distortion in the text from 0 to 1. We then perform binary classification with thresholds set at increments of 0.1, and compare the results using the F metric with a variable parameter β against datasets found in our theme.

Results and Conclusion: We have discovered that the average across all distortions with a threshold of 0.1 and an F metric with a parameter of 0.1 shows the best and most universal performance, which is due to the presence in the datasets of a larger number of texts containing distortions than neutral expressions. Up to the threshold of 0.5 for F metrics with parameters from 0.1 to 0.5, distortions such as “negative statements”,

“positive statements”, “rude statements”, and “should statements” also perform well, which is due to the uneven distribution of the expression of cognitive distortions in the studied datasets. You can view the results at the following link: <https://github.com/AnnaArinicheva/CognitiveProject.git>

Список литературы/References

1. Bollen J., Thij M., Breithaupt F., Barron A., Rutter L., Lorenzo-Luaces L., Scheffer M. Historical language records reveal a surge of cognitive distortions in recent decades. *PNAS*. 2021;118(30):e2102061118. doi 10.1073/pnas.2102061118
2. Raheman A., Kolonin A., Fridkins I., Ansari I., Vishwas M. Social Media Sentiment Analysis for Cryptocurrency Market Prediction. *arXiv*. 2022. doi 10.48550/arXiv.2204.10185
3. Kolonin A., Raheman A., Vishwas M., Ansari I., Pinzon J., Ho A. Causal Analysis of Generic Time Series Data Applied for Market Prediction. *arXiv*. 2022. doi 10.48550/arXiv.2204.12928
4. Kolonin A., Kurpatov A., Molchanov A., Averyanov. Cognitive Architecture for Decision-Making Based on Brain Principles Programming. *Procedia Comput Sci*. 2022;213:180-189. doi 10.1016/j.procs.2022.11.054