REVIEW

Exploring the Association Between Textual Parameters and Psychological and Cognitive Factors

Kadir Uludag 🆔

Shanghai Mental Health Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, People's Republic of China

Correspondence: Kadir Uludag, Shanghai Jiao Tong University, School of Medicine, 16 Wanping Nan Road, 600, Shanghai, 20030, People's Republic of China, Tel +86-18401653162, Email kadiruludag@mails.ucas.ac.cn

Background: Textual data analysis has become a popular method for examining complex human behavior in various fields, including psychology, psychiatry, sociology, computer science, data mining, forensic sciences, and communication studies. However, identifying the most relevant textual parameters for analyzing complex behavior is still a challenge.

Goal of Study: This paper aims to explore potential textual parameters that could be useful in analyzing behavior through complex textual data. Furthermore, we have examined the randomly generated text based on different textual parameters.

Methods: To achieve this goal, we conducted a comprehensive review of the literature on textual data analysis and identified several potential topics that could be relevant, such as sentiment analysis, discourse analysis, lexical analysis, and syntactic analysis. We discuss the theoretical background and practical implications of each parameter and provide examples of how they have been used in previous research. Furthermore, we highlight the importance of considering the context in which these parameters are applied and the need for interdisciplinary collaboration to gain a deeper understanding of complex behavior through textual data analysis. Furthermore, we have provided Python code in the Supplementary Materials to facilitate a comprehensive analysis of such behaviors. In addition, to generate the text for analysis, we utilized ChatGPT 3.5 Turbo by requesting it to generate a random text of 1000 words divided into five paragraphs. Afterwards, we applied the provided Python code to analyze the randomly generated text.

Conclusion: Overall, this paper provides a foundation for researchers to identify relevant textual parameters to analyze complex human behavior in their respective fields such as linguistics, sociology, psychiatry, and psychology.

Keywords: chatbot, behavioral analysis, textual parameters, textual data, psychology, text analysis, linguistics, textual patterns

Introduction

Textual data analysis has become an increasingly popular method for examining complex human behavior in various fields, including psychology, psychiatry, sociology, biology, computer science, linguistics, and communication studies.

The use of textual data analysis allows researchers to extract meaningful insights and patterns from large volumes of written or spoken language data, providing a deeper understanding of human behavior. There is a specific focus in studies on quantifying psychological characteristics.[1] Textual analysis has been conducted using two distinct approaches: theory-based, closed-vocabulary methods derived from social sciences, and data-driven methods rooted in computer science.[2] Furthermore, Corpus linguistics and conversation analysis method can reveal new insights into the relationship between interaction patterns and learning.[3] Additionally, emphasis is lacking on studying how people communicate in social sciences.[4] A study mentioned that, despite current methods have many disadvantages compared to more traditional methods of data collection, the increase of computational power a will make automated text analysis common.[1] For example, the automated linguistic analysis of consumer reviews proves beneficial in detecting deficiencies in service delivery.[5]

Textual analysis can be employed to meticulously examine the way individuals communicate and document it. For example, a previous project investigated the features of linguistic style that distinguish between true and false stories. Compared to truth-tellers, liars showed lower cognitive complexity, used fewer self-references and other-references, and used more negative emotion words.[6]

1139

Moreover, in a previous study, the words people use in disclosing a trauma were hypothesized to predict improvements in mental health and results from participants in previous writing studies showed that increased use of words associated with insightful and causal thinking was linked to improved physical but not mental health.

However, identifying the most relevant textual parameters for analyzing behavior can be challenging. While there are many potential textual parameters that could be useful for analyzing behavior through textual data, it is important to determine which parameters are most relevant to the research question and context. Psychological scales are frequently utilized to further elucidate the relationship between textual parameters and psychological phenomena.[7,8] Moreover, previous studies highlighted the capability of ChatGPT to proficiently conduct psychological analysis.[9,10]

Large Language Models and Textual Analysis

One of the major challenges in harnessing the true potential of computers is their limited understanding of language.[11] Large language models refer to advanced artificial intelligence models that have been trained on vast amounts of text data (eg, ChatGPT, LLaMA, Gopher, and Google Bard). The advancement in artificial intelligence, characterized by the availability of highly capable large language models has enhanced our ability to comprehend sentiment.[12] Nonetheless, the extent to which large language models resemble human text emotionally remains uncertain.

Through the utilization of few-shot learning, large language models have demonstrated exceptional performance.[13] The capacity to acquire knowledge efficiently from unprocessed text plays a vital role in reducing the reliance on supervised learning.[14] Numerous articles evaluating ChatGPT have examined its efficacy in various established tasks.[15] Moreover, a previous study[16] evaluated GPT-3's abilities in decision-making, information retrieval, contemplation, and causal reasoning through a series of experiments from previous literature. Findings indicate that GPT-3's performance is impressive.[16] Generative artificial intelligence is potentially useful in evaluating complex emotion dynamics.[17] Furthermore, it has potential to assess various texts. Additionally, it has the capacity to automatically assess a broad spectrum of texts without requiring human intervention.

In summary, the remarkable performance of large language models showcases their ability to efficiently acquire knowledge from unprocessed text, reducing the reliance on supervised learning and demonstrating efficacy in various established tasks, including psychological analysis, decision-making, information retrieval, contemplation, and causal reasoning.

Textual Analysis and Cognitive Psychology

There is a lack of consensus when it comes to the textual parameters associated with psychology. Researchers choose different parameters associated with texts, including sentence length and specific psychological factors. Moreover, textual analysis and cognition are intertwined fields that explore the intricate relationship between language, thought processes, and meaning extraction from written or spoken texts. Cognitive skills and psychological parameters are distinct from each other, and textual parameters may encompass knowledge pertaining to both of these aspects.

A fundamental question in cognitive studies of text production revolves around the mental processes involved when individuals generate a text.[18] At a broader scope, language can shed light on fluctuating patterns in how humans comprehend concepts like God and the self.[19] Consequently, it holds significance to identify textual parameters that could potentially have associations with psychological and cognitive factors.

Potential Textual Parameters

The relationship between textual parameters and psychology relates to the understanding that the content, structure, and style of language can provide valuable insights into an individual's psychological state, cognitive processes, personality traits, emotions, and social dynamics. By analyzing and interpreting these textual parameters, researchers can gain a deeper understanding of how language reflects and influences psychological phenomena.

Total Number of Words

The "total number of words" refers to the count of all individual words present in each text or document. It represents the sum of all distinct words, excluding repetitions and punctuation marks. This metric provides a quantitative measure of the

text's length or content, allowing for comparisons and analysis. Alternatively, we can measure total words not including the distinct words. Psychological parameters may be associated with the number of words typed within a specific period.

Average Sentence Length

The "average sentence length" refers to the average number of words in a sentence within a given text or document. It is a metric that provides insight into the overall structure and readability of the text. By calculating the average sentence length, one can gauge the complexity of the sentences and assess the overall flow of the writing. Average sentence length may be associated with psychological factors such as cognitive processing abilities, attention span, linguistic proficiency, and emotional state.

Average Sentence Standard Deviation

The "average sentence standard deviation" is a statistical measure that quantifies the variability or dispersion of sentence lengths within a given text or document. It provides an indication of how much the sentence lengths deviate from the average sentence length. This metric helps to understand the distribution and diversity of sentence lengths within the text. Certain emotional changes can be linked to the variation in sentence structure from the average.

Quantifying Sentence Count Within Paragraphs

Quantifying Sentence Count within Paragraphs refers to the process of measuring and analyzing the number of sentences present within each paragraph of a given text. It involves determining the average, range, or distribution of sentence counts to gain insights into the structural characteristics and organization of the text.

Paragraphs with an increased number of sentences often signify the presence of significant or essential information. The inclusion of additional sentences within a paragraph suggests a deeper exploration or elaboration of the subject matter at hand. Additionally, we can establish correlations between this parameter and specific psychological factors, such as attention.

Number of Adjectives in the Sentence

The number of adjectives in a sentence refers to the count of words that modify or describe nouns or pronouns. Adjectives provide additional information about the qualities, characteristics, or attributes of the nouns or pronouns they modify.

To determine the number of adjectives in a sentence, you would need to identify and count the words that fulfill this role.

For example, in the sentence "The big, black dog chased the small, brown cat", there are four adjectives: "big", "black", "small", and "brown". These adjectives modify the nouns "dog" and "cat" by describing their size and color.

Number of adjectives: [2, 2, 1, 2, 2], Mean: 1.8, Variance: 0.36, Standard deviation: 0.6

Average Number of Letters

The "average number of letters" refers to the average count of individual letters in each text or document. It provides a quantitative measure of the length or complexity of the text in terms of individual letters, excluding spaces and punctuation marks. Nevertheless, there is a scarcity of available literature regarding the average quantity of characters and the psychological factors connected to it.

Number of Letters Using Standard Deviation and Variance Among Sentences

The standard deviation of the number of letters is a statistical measure that quantifies the variability or dispersion of the lengths of words in each dataset. It provides a measure of how much the lengths of words deviate or differ from the average length.

Number of Letters Using Variance

The number of letters variance is a statistical measure that quantifies the variability or spread of word lengths in each dataset. It provides a measure of how much the lengths of words deviate or differ from the average length.

To calculate the variance of the number of letters, we typically follow these steps:

1. Compute the average (mean) length of the words in the dataset by summing up the lengths of all the words and dividing by the total number of words.
2. Calculate the difference between each word's length and the average length, and square the result for each word.
3. Compute the average (mean) of the squared differences obtained in step 2. This value represents the variance.

Polarity

In linguistics, polarity refers to the grammatical and semantic property of expressions that indicate whether a statement is positive or negative. It is a fundamental aspect of language that allows us to express affirmation or negation in our utterances.

In terms of grammar, polarity is often manifested using specific words or affixes that change the meaning of a sentence. For example, in English, the word "not" is commonly used to indicate negation. Consider the following examples:

1. Positive polarity: "I am happy".
2. Negative polarity: "I am not happy".

Previous paper[20] presents a new approach to phrase-level sentiment analysis that first determines whether an expression is neutral or polar and then disambiguates the polarity of the polar expressions. With this approach, the system can identify the contextual polarity for a large subset of sentiment expressions.[20] However, it is possible that we cannot precisely categorize sentences based on polarity. Furthermore, the results may change according to different programs. The variation in this regard may also be influenced by certain cultural and linguistic factors.

Subjectivity

The ability to differentiate between factual and subjective information would be advantageous for numerous language processing applications.[21]

It is possible to extract negative, neutral, positive, and compound values. It's essential to note that measuring subjectivity in writing is an ongoing research area, and there is no definitive method as it requires several approaches based on particular use cases and desired outcomes.

Counting Tenses in a Paragraph

Counting tenses in a paragraph refers to the process of identifying and enumerating the occurrences of different verb tenses within a given paragraph of text. The objective is to determine the frequency and distribution of verb forms that indicate different time references or aspects of actions and events. By systematically examining the paragraph's verbs, one can gain insights into the temporal characteristics and narrative structure of the written content.

Indeed, the way tenses are counted and identified can vary depending on the language being analyzed. Different languages may have unique verb forms and grammatical structures that convey tense and aspect.

- Simple Present Tense: The sun rises in the morning.
- Present Continuous Tense: She is currently studying for her exams.
- Present Perfect Tense: They have already completed their work.
- Simple Past Tense: He walked to the store yesterday.
- Past Continuous Tense: We were watching a movie when the power suddenly went out.
- Past Perfect Tense: She had already departed when I arrived.
- Simple Future Tense: I will meet you tomorrow.
- Future Continuous Tense: They will be traveling to Europe next week.
- Future Perfect Tense: By the time you arrive, we will have finished dinner.
- Present Perfect Continuous Tense: He has been working on this project for three hours.

In summary, the way we use grammar offers insights into our psychological state, giving us valuable hints about our thoughts, emotions, and cognitive abilities. Exploring the relationship between language and psychology allows us to gain a deeper understanding of human behavior and improve communication approaches in different situations. There is a dearth of research in literature regarding the connection between grammar and psychology.

Tenses and Average Length of the Sentences

The average length of sentences refers to the typical number of words or characters contained in a sentence. It is a measure used to gauge the complexity and readability of a text.

Punctuation in a Sentence

Counting punctuation in a sentence refers to the process of determining the number of punctuation marks present within the given sentence. Punctuation marks are symbols used in writing to enhance clarity, indicate pauses, convey meaning, or structure sentences. Common examples of punctuation marks include commas, periods, question marks, exclamation marks, colons, semicolons, quotation marks, and parentheses. By counting punctuation in a sentence, one can analyze the sentence's structure, style, or even assess its grammatical correctness. For example, we can calculate the number of words before and after comma using python.

a, an, the Frequency

In English grammar, "a", "an", and "the" are articles used to specify or indicate the presence of a noun. "A" and "an" are indefinite articles, used before singular nouns to refer to any member of a general category. "A" is used before nouns that begin with a consonant sound, while "an" is used before nouns that begin with a vowel sound. For example, we say "a cat" and "an apple". On the other hand, "the" is a definite article used before singular and plural nouns to refer to a specific person, thing, or group. It implies that the noun being referred to is already known or has been mentioned before. It could be associated with the focus on a particular subject matter.

Number of Subjects

The number of subjects in a sentence refers to the count of distinct subjects, which are the nouns or noun phrases that perform the action or are being described in the sentence. In English grammar, a sentence can have one or more subjects.

In a previous research, attempts were made to categorize deceptive emails based on different linguistic parameters.[22] Previous study indicates that a shift in language—silently referring to oneself using one's own name and non–first-person -singular pronouns—promotes emotion regulation.[23]

In conclusion, the number of subjects in a text holds significance as it can provide valuable information about individuals mentioned within the text. Future studies can investigate how the significance of the number of subjects in a text varies across different cultures. Grammatical rules and cultural influences can pose as potential factors that might complicate it.

Ratio of Subjects

Ratio of You and I

The "ratio of you and I" in a paragraph refers to the relative frequency or representation of the pronouns "you" and "I" within the given text. This ratio is determined by calculating the number of occurrences of the pronoun "you" and comparing it to the number of instances of the pronoun "I" in the paragraph. The ratio can be expressed as a fraction, percentage, or in other numerical forms.

In the literature, there are a few studies including ratio of subjects in a population. For example, in a previous study the utilization of the generic pronoun "you" in written examples encourages psychological detachment.[24]

Calculation of Location Sentences

The "calculation of location sentences" is not a standard term or concept in the field of linguistics or natural language processing. Therefore, I cannot provide a precise definition for it. However, based on the general understanding of the terms involved, it could refer to the process of analyzing or determining the location-related information within

sentences. Excessive use of location-related sentences may be related to psychological problems or traumatic experiences. In this study, we have calculated only geographic cities. This study has exclusively considered cities with a geographical location in its calculations.

Number of Numerical Keywords

The number of numerical words in a sentence refers to the count of words that represent numerical values or quantities. These words can include cardinal numbers (eg, one, two, three), ordinal numbers (eg, first, second, third), fractions (eg, half, third, quarter), decimals (eg, 0.5, 2.75), percentages (eg, 25%, 50%), and other numerical expressions.

Topic Modelling

Topic modelling can help us to understand the main characteristics of the given text. Topic models are clustering approach for text data.[25] Topic modeling is a commonly employed method for unveiling inherent thematic patterns within textual data.[26]

Topic modeling is a technique used in natural language processing and machine learning to identify and extract latent topics or themes from a collection of documents or texts.

A topic is a probability distribution over words. And a topic model specifies a probabilistic procedure by which documents can be generated. Each document is modeled as a mixture of multinomial distributions over words in different proportions.[27]

Most Frequently Used Words

The term "most frequently used words" refers to the words that appear most often in each language, text, or corpus. These words are commonly referred to as high-frequency words or commonly used words. Understanding someone's response can be facilitated by analyzing the words they use most frequently. Overused words can be used to deliver an emphatic message. Chatbots designed to provide cognitive textual therapy with a psychological focus can utilize overused words as an indicator or sign to gain insights into the user's mental and emotional state.

Most Frequently Used Sentences

"Most frequently used sentences" refers to the set of sentences that are commonly used or encountered in a particular context, language, or domain. These sentences are recurrent and widely employed due to their relevance, effectiveness, and familiarity to speakers of a language. It is probable that there are no duplicate sentences.

Use of Passive Voice

The use of passive voice in a paragraph refers to the construction of sentences where the subject receives the action of the verb, rather than performing the action. In other words, the object of the sentence becomes the subject, and the subject becomes the object, often with the use of the verb "to be" and a past participle.

Use of Passive Voice and Semantic Analysis

The use of passive voice and semantic analysis are related in that the choice of voice can affect the meaning and interpretation of a text.

In general, the active voice tends to be more direct and clear, making it easier for readers to understand the meaning of the text. In contrast, the passive voice can make the text seem more indirect and less engaging, which can make it more difficult for readers to comprehend the meaning. While defining passive voice can be challenging, it is important to understand its meaning and impact on writing.

Use of Passive Voice and Average Sentence Length

The relationship between the use of passive voice and average length refers to the correlation between the two textual features and how they can impact the readability and clarity of a text.

NLTK Package of Python

NLTK stands for Natural Language Toolkit. It is a popular open-source Python library designed for working with human language data. NLTK provides a wide range of tools, resources, and algorithms for tasks such as text classification, tokenization, stemming, parsing, semantic reasoning, and more. Python can be used in linguistic processing (Bird, Klein, and Loper, 2009). A previous paper[28] created the Reddit App to get the Client id and Secret identification of Creator then used python packages like praw and nltk, then we downloaded lexicon and then imported Sentiment Intensity Analyzer for identifying the sentiments and then assigned polarity_ scores to each headline as 0, 1, and −1 as a compound value.[28] Natural language processing includes the ability to draw insights from email data.[29] NLTK packages encourage researchers in natural language processing.[30] In summary, the availability of NLTK packages has significantly contributed to the advancement of research in natural language processing and artificial intelligence.

This paper aims to explore potential textual parameters that can be used to analyze behavior through textual data. We conducted a comprehensive narrative review of the literature on textual data analysis and identified several potential textual parameters that could be relevant, such as sentiment analysis, discourse analysis, lexical analysis, and syntactic analysis. We discuss the theoretical background and practical implications of each parameter and provide examples of how they have been used in previous research.

Furthermore, we highlight the importance of considering the context in which these parameters are applied and the need for interdisciplinary collaboration to gain a deeper understanding of behavior through textual data analysis. By identifying relevant textual parameters for analyzing behavior, researchers can better leverage the power of textual data analysis to extract meaningful insights from large volumes of written or spoken language data.

In summary, this paper establishes a fundamental framework that empowers researchers to recognize essential textual factors for examining human behavior within their specific domains. Furthermore, it serves as a steppingstone for future investigations in understanding and analyzing human behavior.

Methods

For the analysis of textual data, we employed the Jupyter Notebook environment with Python version 3.8.8 to execute the necessary Python code. Additionally, we have included the relevant Python code in the Supplementary Materials to facilitate a comprehensive analysis of subtle behaviors expressed through text.

To generate the text for analysis, we employed ChatGPT 3.5 Turbo, requesting it to create random text consisting of 1000 words spread across five paragraphs. Subsequently, we employed the mentioned code to analyze the randomly generated text. For reference and reproducibility, we have provided the code in the Supplementary Materials. We have used following prompt to create random text: Create a random text comprising 1000 words divided into 5 paragraphs:

In Table 1, we have compiled a collection of studies that employ textual analysis as their methodology, showcasing the diverse range of research conducted in this field.

Results

Using the ChatGPT 3.5 Turbo version, we generated five paragraphs of random text consisting of 1000 words. Subsequently, we employed the provided code to analyze the generated text as outlined in this study.

Potential Textual Parameters

In this section, we have calculated potential textual parameters in the given example. Total number of words, average sentence length, number of adjectives, average number of letters, standard deviation of number of letters, location words, and variance of number of letters were calculated.

Total Number of Words

The combined sentences contained a total of 394 words.

Average Sentence Length

The average sentence length was 16.41.

Table 1 Example Studies Using Text Analysis methods

Reference	Goal of Study	Results of Study
1. Snyder S, Milbury K, Wagner R, Cohen L. Words matter: The use of generic "you" in expressive writing in an oncology setting. Journal of Health Psychology. 2023:13591053231182218. doi:10.1177/13591053231182218[24]	The goal is to explore the potential development of psychological self-distancing prompts for use in writing interventions or as a clinical tool specifically designed for cancer patients.	The utilization of the generic pronoun "you" in written examples encourages psychological detachment
2. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: Predicting deception from linguistic styles. Personality and social psychology bulletin. 2003;29(5):665–675[6]	The study examined the characteristics of linguistic style that differentiate true narratives from false ones.	Compared to truth-tellers, liars showed lower cognitive complexity, used fewer self-references and other-references.
3. Kuperus W. Classifying deceptive emails with first and third person pronouns: Applying Support Vector Machine on. Tilburg University; 2017[22]	Goal of the thesis was to classify deceptive emails with first and third person pronouns.	The research presents proof that pronouns can serve as cues for constructing a predictive classification model.
4. Mehl MR, Raison CL, Pace TW, Arevalo JM, Cole SW. Natural language indicators of differential gene regulation in the human immune system. Proceedings of the National Academy of Sciences. 2017;114(47):12554–12559[31]	The goal of this study was to identify and investigate systematic individual differences in natural language use that exhibit a stronger correlation with CTRA gene expression compared to conventional self-report measures of stress, anxiety, or depression.	A previous study identified individual differences in natural language use that track CTRA gene expression more closely than do conventional self-report measures of stress, anxiety, or depression
5. Mesnil G, Mikolov T, Ranzato MA, Bengio Y. Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews. arXiv preprint arXiv:14125335. 2014[32]	Searching techniques for sentiment analysis of movie reviews.	The study achieved strong results on a dataset of IMDB reviews
6. Boals A, Klein K. Word Use in Emotional Narratives about Failed Romantic Relationships and Subsequent Mental Health. Journal of Language and Social Psychology. 2005/09/01 2005;24(3):252–268. doi:10.1177/0261927X05278386[33]	Investigating the relationship between emotional narratives and textual parameters.	Greater avoidance of the breakup predicted greater use of negative emotion, first person singular and third-person pronouns, and less use of cognitive words
7. Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology.* 2012/01/01 2012;63(1):163–173. doi:https://doi.org/10.1002/asi.21662[34]	The objective was to explore the detection of sentiment strength specifically for the social web.	Researchers discovered that as the number of observations increased, there were notable connections between the utilization of emotion words and self-reported emotions.
8. Manchaiah V, Swanepoel DW, Bennett RJ. Online consumer reviews on hearing health care services: A textual analysis approach to examine psychologically meaningful language dimensions. American Journal of Audiology. 2021;30(3):669–675[5]	The research investigated language dimensions that are psychologically significant, including social and emotional dimensions, health dimensions, and personal concerns, in online consumer reviews.	The automated linguistic analysis of consumer reviews proves beneficial in detecting deficiencies in service delivery

Average Sentence Standard Deviation
The average sentence standard deviation was 4.46.

Counting Location Words
Analyzed text does not include location words.

Number of Adjectives in the Sentence
The number of adjectives in the sentence was 32.

Average Number of Letters
The average number of letters per word was 5.34.

Standard Deviation of Number of Letters
Standard deviation of the number of letters was 2.74.

Variance of Number of Letters
The difference of the number of letters was 7.55.

Counting Tenses
We have classified tenses into three categories and observed that the majority of the text generated by the chatbot was in the past tense.

Past tense count: 34
Present tense count: 4
Future tense count: 0

Sentiment Analysis
If the polarity is greater than 0, it is considered positive sentiment. If it is less than 0, it is considered negative sentiment. Otherwise, it is considered neutral sentiment. In this study, the result indicated a positive sentiment.

Discussion
We conducted a comprehensive review of the literature on textual data analysis and identified several potential topics that could be relevant, such as sentiment analysis, discourse analysis, lexical analysis, and syntactic analysis. We discuss the theoretical background and practical implications of each parameter and provide examples of how they have been used in previous research. We conducted a comprehensive analysis of text samples generated by ChatGPT version 3.5 Turbo and compiled a summary of our key findings. Investigated textual parameters and psychological parameters can be used together to understand complex human behavior.

Textual data analysis has become a popular method for examining human behavior in various fields, including psychology, biology, sociology, social sciences, and communication studies. However, identifying the most relevant textual parameters for analyzing behavior is still a major challenge. Furthermore, studies are selecting diverse textual parameters for analysis, highlighting the variability in approaches to studying human behavior through text.

Various approaches were developed to analyze complex textual parameters. One of the approaches is thematic analysis. Thematic analysis aids in comprehending the underlying themes and topics within a given text. Thematic analysis is a commonly employed qualitative research method in the field of psychology.[35] Thematic analysis provides valuable insights into the overarching themes present within texts.

Furthermore, sentences can be separately analyzed emotionally since particles, which serve as the glue that holds nouns and verbs together, can be related to emotional state.[36] Moreover, discourse analysis might serve as a preliminary exposure to discourse analysis.[37]

Other study mentioned Coh-Metrix, analyzes texts on over measures of language-related parameters and its modules use lexicons, part-of-speech classifiers, and syntactic parsers.[38] Furthermore, LSA theory establishes a hidden semantic space wherein documents and words are depicted as vectors and it is feasible to compute term similarity, categorize terms and documents, and summarize extensive documents using automated process that imitate humans.[39]

Also, textual information from interviews can be evaluated using content analysis.[40]

Most research uses key phrase matching and word frequency to identify psychological cognitive change.[41] Collecting social media data may be more scalable than traditional approaches.[42] Also, it allows us to reach a wider audience.

Textual analysis can be further used in prediction studies. In a study, greater avoidance of the breakup predicted greater use of negative emotion, first person singular and third-person pronouns, and less cognitive words use.[33]

Textual analysis is advantageous for the field of biology as well. For example, a previous study identified individual differences in natural language use that track CTRA gene expression more closely than do conventional self-report measures of stress, anxiety, or depression.[31]

Sentiment Analysis

Sentiment analysis or opinion mining is the process of determining subjective information from text data such as positive, negative, or neutral. There is increasing interest in the affective dimension of the web.[34] Through simulations, researchers discovered that as the number of observations increased, there were notable connections between the utilization of emotion words on Facebook and self-reported emotions.[34] By applying semantic analysis with a syntactic parser and sentiment lexicon, prototype system achieved high precision in finding sentiments.[43]

One study using sentiment analysis achieved strong results on a dataset of IMDB reviews.[32] In this study, we have administered sentiment analysis as well. The result indicated a positive sentiment. The observed positive sentiment in the study could be attributed to the coding of ChatGPT (version 3.5 turbo), as it may lean towards generating responses with a positive sentiment bias.

Tag Clouds

A tag cloud, also known as a word cloud or a weighted list, is a visual representation of text data where the size or prominence of each word is determined by its frequency or importance within the given context. In a tag cloud, words are typically displayed in different font sizes, with more prominent or frequently occurring words appearing larger and smaller words indicating less frequency or significance.

TagClouds is a popular visualization for collaborative tags.[44] Ontology-based technologies, which can extract the relationships between concepts by using advanced text mining tools, can be applied to improve design information retrieval in the unstructured text.[45] The tag cloud, favored for its user-friendly interface.[46] In conclusion, tag clouds, also known as word clouds or weighted lists, provide a visual representation of text data based on the frequency or importance of each word within a given context.

Suggestions for Further Studies

Further studies can explore voice-related parameters to enhance our understanding of their impact on various aspects. By investigating voice-related parameters, researchers can gain insights into the influence of vocal characteristics on communication, emotion recognition, and speech disorders.

Moreover, the focus can be directed towards sample sizes comprising individuals with psychiatric disorders such as depression, schizophrenia, or personality disorders. This approach can prove valuable in the field of forensic sciences.

Textual data retrieved from social media can be a valuable resource for assessing the psychological health of individuals. By analyzing the content of social media posts, researchers and mental health professionals can gain insights into various aspects of an individual's well-being and psychological state.

It is important to recognize that utilizing textual parameters alone cannot establish a causal relationship. While analyzing textual data can provide valuable insights and correlations, it does not inherently determine a cause-and-effect relationship between variables.

Limitations

Assessing the appropriate intensity scores for cases with multiple intensifiers can be subjective and may vary across individuals. Furthermore, it is essential to acknowledge that our manuscript is a narrative review and does not adhere to systematic review methods.

Conclusion

This paper serves as a fundamental framework for researchers, enabling them to discern pertinent textual parameters and effectively analyze intricate human behavior within their specific domains. It explores a range of potential parameters, some of which have been utilized in the past while introducing new ones. This paper establishes a basis for researchers in fields such as linguistics, psychiatry, and psychology to identify pertinent textual parameters for analyzing intricate human behavior.

Abbreviations

NLTK, natural language toolkit.

Data Sharing Statement

Data can be made available upon a reasonable request.

Funding

Disclosure

The author reports no conflicts of interest in this work.

References

1. Iliev R, Dehghani M, Sagi E. Automated text analysis in psychology: methods, applications, and future developments. *Lang Cognit*. 2014;7 (2):265–290. doi:10.1017/langcog.2014.30
2. Eichstaedt JC, Kern ML, Yaden DB, et al. Closed-and open-vocabulary approaches to text analysis: a review, quantitative comparison, and recommendations. *Psychol Methods*. 2021;26(4):398. doi:10.1037/met0000349
3. O'Keeffe A, Walsh S. Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. *Corpus Linguist Linguist Theory*. 2012;8(1):159–181. doi:10.1515/cllt-2012-0007
4. Chung C, Pennebaker J. The psychological functions of function words. In: *Social Communication*. Psychology Press; 2011:343–359.
5. Manchaiah V, Swanepoel DW, Bennett RJ. Online consumer reviews on hearing health care services: a textual analysis approach to examine psychologically meaningful language dimensions. *Am J Audiol*. 2021;30(3):669–675. doi:10.1044/2021_AJA-20-00223
6. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: predicting deception from linguistic styles. *Personal Soc Psychol Bull*. 2003;29(5):665–675. doi:10.1177/0146167203029005010
7. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Comput Math Methods Med*. 2016;2016:1–8. doi:10.1155/2016/8708434
8. Sarzynska-Wawer J, Wawer A, Pawlak A, et al. Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Res*. 2021;304:114135. doi:10.1016/j.psychres.2021.114135
9. Rathje S, Mirea D-M, Sucholutsky I, Marjieh R, Robertson C, Van Bavel JJ. GPT is an effective tool for multilingual psychological text analysis; 2023.
10. Korkmaz A, Aktürk C, Talan T. Analyzing the user's sentiments of ChatGPT using twitter data. *Iraqi J Comput Sci Math*. 2023;4(2):202–214. doi:10.52866/ijcsm.2023.02.02.018
11. Turney PD, Pantel P. From frequency to meaning: vector space models of semantics. *J Artif Intell Res*. 2010;37:141–188. doi:10.1613/jair.2934
12. Roumeliotis KI, Tselikas ND, Nasiopoulos DK. LLMs in e-commerce: a comparative analysis of GPT and LLaMA models in product review evaluation. *Nat Lang Proces J*. 2024;6:100056. doi:10.1016/j.nlp.2024.100056
13. Chowdhery A, Narang S, Devlin J, et al. Palm: scaling language modeling with pathways. *J Mach Learn Res*. 2023;24(240):1–113.
14. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training; 2018.
15. Uludag K. Testing creativity of ChatGPT in psychology: interview with ChatGPT. Available at SSRN 4390872; 2023.
16. Binz M, Schulz E. Using cognitive psychology to understand GPT-3. *Proc Natl Acad Sci*. 2023;120(6):e2218523120. doi:10.1073/pnas.2218523120
17. Markowitz DM. Can generative AI infer thinking style from language? Evaluating the utility of AI as a psychological text analysis tool. *Behav Res Methods*. 2024;1–12. doi:10.3758/s13428-022-02040-x
18. Schilperoord J. It's about time: temporal aspects of cognitive processes in text production. In: *It's About Time*. Brill; 2022.
19. Jackson JC, Watts J, List J-M, Puryear C, Drabble R, Lindquist KA. From text to thought: how analyzing language can advance psychological science. *Perspect Psychol Sci*. 2021;17(3):805–826. doi:10.1177/17456916211004899
20. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis; 2005:347–354.
21. Riloff E, Wiebe J. Learning extraction patterns for subjective expressions; 2003:105–112.
22. Kuperus W. *Classifying Deceptive Emails with First and Third Person Pronouns: Applying Support Vector Machine on*. Tilburg University; 2017.
23. Orvell A, Vickers BD, Drake B, et al. Does distanced self-talk facilitate emotion regulation across a range of emotionally intense experiences? *Clin Psychol Sci*. 2020;9(1):68–78. doi:10.1177/2167702620951539

24. Snyder S, Milbury K, Wagner R, Cohen L. Words matter: the use of generic "you" in expressive writing in an oncology setting. *J Health Psychol*. 2023;13591053231182218. doi:10.1177/13591053231182218

25. De Battisti F, Ferrara A, Salini S. A decade of research in statistics: a topic model approach. *Scientometrics*. 2015;103(2):413–433. doi:10.1007/s11192-015-1554-1

26. Wang H, Prakash N, Hoang NK, Hee MS, Naseem U, Lee RK-W. Prompting Large Language Models for Topic Modeling. *IEEE*. 2023;1236–1241. doi:10.1109/JPROC.2023.3273517

27. McFarland DA, Ramage D, Chuang J, Heer J, Manning CD, Jurafsky D. Differentiating language usage through topic models. *Poetics*. 2013;41(6):607–625. doi:10.1016/j.poetic.2013.06.004

28. Khemani B, Adgaonkar A. A review on reddit news headlines with nltk tool. *SSRN Electr J*. 2021. doi:10.2139/ssrn.3834240

29. Farkiya A, Saini P, Sinha S, Desai S. Natural language processing using NLTK and wordNet. *Int J Comput Sci Inf Technol*. 2015;6(6):5465–5469.

30. Yogish D, Manjunath T, Hegadi RS. *Review on Natural Language Processing Trends and Techniques Using NLTK*. Springer; 2019:589–606.

31. Mehl MR, Raison CL, Pace TW, Arevalo JM, Cole SW. Natural language indicators of differential gene regulation in the human immune system. *Proc Natl Acad Sci*. 2017;114(47):12554–12559. doi:10.1073/pnas.1707373114

32. Mesnil G, Mikolov T, Ranzato MA, Bengio Y. Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews. *arXiv*. 2014. doi:10.48550/arXiv.1412.5335

33. Boals A, Klein K. Word use in emotional narratives about failed romantic relationships and subsequent mental health. *J Lang Soc Psychol*. 2005;24(3):252–268. doi:10.1177/0261927X05278386

34. Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web. *J Am Soc Inform Sci Technol*. 2012;63(1):163–173. doi:10.1002/asi.21662

35. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. *Qualit Psychol*. 2022;9(1):3–26. doi:10.1037/qup0000196

36. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Ann Rev Psychol*. 2003;54(1):547–577. doi:10.1146/annurev.psych.54.101601.145041

37. Widdowson HG. *Discourse Analysis*. Vol. 133. Oxford University Press Oxford; 2007.

38. Graesser AC, McNamara DS, Louwerse MM, Cai Z. Coh-Metrix: analysis of text on cohesion and language. *Behav Res Methods Instr Comput*. 2004;36(2):193–202. doi:10.3758/BF03195564

39. Evangelopoulos NE. Latent semantic analysis. *WIREs Cognit Sci*. 2013;4(6):683–692. doi:10.1002/wcs.1254

40. Kondracki NL, Wellman NS, Amundson DR. Content analysis: review of methods and their applications in nutrition education. *J Nutr Educ Behav*. 2002;34(4):224–230. doi:10.1016/S1499-4046(06)60097-3

41. Gu D, Li M, Yang X, et al. An analysis of cognitive change in online mental health communities: a textual data analysis based on post replies of support seekers. *Inform Proc Manag*. 2023;60(2):103192. doi:10.1016/j.ipm.2022.103192

42. Andreotta M, Nugroho R, Hurlstone MJ, et al. Analyzing social media data: a mixed-methods framework combining computational and qualitative text analysis. *Behav Res Methods*. 2019;51(4):1766–1781. doi:10.3758/s13428-019-01202-8

43. Nasukawa T, Yi J. Sentiment analysis: capturing favorability using natural language processing; 2003:70–77.

44. Chen Y-X, Santamaría R, Butz A, Therón R. *Tagclusters: Semantic Aggregation of Collaborative Tags Beyond Tagclouds*. Springer; 2009:56–67.

45. Shi F, Chen L, Han J, Childs P. A data-driven text mining and semantic network analysis for design information retrieval. *J Mech Design*. 2017;139(11). doi:10.1115/1.4037649

46. Khusro S, Jabeen F, Khan A. Tag clouds: past, present and future. *Proc Natl Acad Sci India Sec A Phys Sci*. 2021;91(2):369–381. doi:10.1007/s40010-018-0571-x