CS447 Literature Review: Natural Language Processing and the Psychological State

James I. Garijo-Garde,
jamesig2@illinois.edu

December 6, 2021

Abstract

Our internal psychological states can provide useful insight into natural language, but it is only beginning to be explored in the context of natural language processing (NLP). Research centered around word frequencies and "words as attention" form the basis of current language-based assessments of psychological state, yet are flawed in that they don't incorporate the context of the use of those words. Two studies investigating NLP for signs of stress and emotion and assessments of needs, motives, and emotional states mitigated the issue of context by gathering detailed contextual information, leading to successful classifications of these traits. A third study ignored context while leveraging the advantages of social media to yield a more accurate model incorporating psychological state. When NLP can effectively incorporate evaluations of psychological state featuring contextual information without placing a significant burden on the researchers, its abilities will be significantly expanded.

1 Introduction

The field of natural language processing (NLP) employs many different facets of language to form a representation of what utterances mean. The structure, distribution, meaning, and parts of speech of words, for example, are all broadly investigated as means of understanding the underlying meaning of a string of words; grammar rules are invoked to get a deeper understanding of how different words interact in a passage. But, there is further meaning that cannot be captured in analyses of words or sentences.

Words are used for specific reasons – conscious or subliminal – because of our internal states. In fact, the choice to speak in the first place can even provide insight into what we are really saying and why we are saying it. The traditional points of analysis in NLP lack the context of our internal states, and (perhaps as a result of this) models employing only traditional NLP inputs are ill-suited for predicting the internal states of human sources.

Currently, NLP is challenged by this deficiency. Consumer NLP products (e.g., personal assistants) can't handle certain statements that appear inappropriate in certain situations without the full context of internal psychological state and lack the ability to craft nuanced responses tailored to the immediate thought process of an individual. An example of this shortcoming is discussed by Jaiswal et al. (2020) in the context of a chatbot designed to provide cognitive behavioral therapy (CBT) to human patients. Therapists are trained to respond differently to patients that are stressed versus calm (Jaiswal et al., 2020). For a CBT chatbot to provide proper counseling in this scenario, it would have the difficult challenge of determining whether a patient is stressed or calm based only

on the language sent to it with each reply. While a human might also struggle with this task, humans can pick up on certain subtle cues that traditional NLP analyses alone would overlook. Using innovative, cutting edge techniques, the publications discussed below make strides in incorporating data representing the psychological state motivating utterances into the NLP pipeline.

2 Analysis of Natural Language and the Psychology of Verbal Behavior

Psychological language analysis is the behavioral science practice of extracting psychological insights from a person's verbal behavioral data (Boyd and Schwartz, 2021). Natural language processing techniques attempting to incorporate psychological state in analyses are built on the assumptions and methodologies of psychological language analysis. In 1999, the Linguistic Inquiry and Word Count (LIWC) software was released (Boyd and Schwartz, 2021). The LIWC was the first widely used psychological text analysis software, and it was guided by the idea that word frequencies represent allocations of attention, an idea termed "words as attention" by Boyd and Schwartz (2021) that soon became a fundamental tenant of psychological language analysis.

Use of the LIWC and similar software employing the "words by attention" paradigm have provided several insights into natural language as an indicator of psychological state, and vice versa. For example, people who use high rates of articles like "the", "a", or "an" and prepositions like "next" or "above" tend to focus on formal or concrete concepts and their inter-relations more than someone who uses those words at a lower rate, while people with higher social status and confidence use "you" and "royal *we*" words more than "I" words (Boyd and Schwartz, 2021). The words people use when describing their homes tell us whether they are focused on the relaxing or stressful properties of their residences (Boyd and Schwartz, 2021). LIWC analyses allow insight into subtle differences in what people pay attention to, their personalities, their life experiences, culture, and society beyond what traditional NLP methods can provide.

Word frequencies can also shed light on how much attention two people are paying to each other or themselves (Boyd and Schwartz, 2021). Personality research shows that social attention is the core, defining feature of extroversion (Boyd and Schwartz, 2021). "Words by attention" research seems to back this up: extroverts use more "social" words than others (Boyd and Schwartz, 2021). However, the social aspects of verbal behavior have traditionally been downplayed in computerized language analysis due to the difficulty in securing and analyzing quality data of interactions between two or more people (social language analysis is often confined to mass media, which is not always representative of how we interact with others in real life situations) (Boyd and Schwartz, 2021).

A strength of the "words as attention" interpretation is its extensible nature (Boyd and Schwartz, 2021). An example mentioned by Boyd and Schwartz (2021) is that once "cognitive" words have been identified (words relating to problem-solving and stress), a person's stress levels can be estimated by counting words. Put more broadly, LIWC and similar algorithms can be used as a tool to investigate a wide range of aspects of a person's psychology, and can even prove useful across other domains of scientific research.

Yet, there are drawbacks to the "words as attention" paradigm. LIWC and similar software are not language models in the traditional NLP sense, but rather merely psychometric statistical analysis tools (Boyd and Schwartz, 2021). While these tools can provide insights into psychological state, they are not designed to accommodate the full complexity and functionality of verbal behavior or language. And, "words as attention" is what Boyd and Schwartz (2021) label a "rocking chair theory:" we continue to learn things ("we keep moving"), but without great scientific process ("but

we do not move forward"). The findings of "words as attention" research is also often highly formulaic: much of the published findings fit the format of "talking about *X* is correlated with *Y*" (Boyd and Schwartz, 2021). Yet, despite these shortcomings, it is undeniable that "words as attention" has proven critical to fostering early inclusion of psychological state in NLP.

Another sentiment of value to NLP offered by Boyd and Schwartz (2021) is the importance of context to the evaluation of text for psychological insight. Boyd and Schwartz (2021) observe that data scraped from social media often lacks valuable context prompting the post. Being aware of this is critical when using social media data for investigations, they write. It's also worth mentioning that word counting algorithms like LIWC don't take into account this context in the first place, discarding valuable psychological clues and further insight into verbalizations (Boyd and Schwartz, 2021). For example, these algorithms would consider a "river bank" and a "financial bank" as the same word if they were represented merely as "bank." Observations taken from word usage counts should be taken with a grain of salt without consideration of the underlying contexts. Boyd and Schwartz (2021) posit that contextual word embeddings are a promising possible mitigation to this shortcoming.

3 Natural Language Processing for Stress and Emotion

As mentioned in the example in the introduction, sometimes insight into a person's emotional state and stress level is valuable. Jaiswal et al. (2020) sought to establish a corpus of utterances analyzed for both of these traits to empower natural language processing applications that can recognize complex user states. Their dataset consists of recordings of 28 college students (9 female and 19 male) across two sections: one with exposure to an external stressor (during finals season), and one with the stressor removed (after finals season). Jaiswal et al. (2020) exposed each subject to a series of emotional stimuli (both short videos and monologue questions) that differed across the two sessions to avoid repetition, but otherwise invoked the same dimensions of emotions. They recorded video of the participant's face, video of the participant's upper body, thermal video of the participant's face, recordings of the participant's voice, various physiological measurements (heart rate, breathing rate, skin conductance, and skin temperature), annotations from self-reports for emotion/stress and from analyses outsourced to Amazon Mechanical Turk, and Big-5 personality scores (except for one participant). Previous studies have shown that thermal recordings and physiological measures like the ones collected by Jaiswal et al. (2020) correlate with stress symptoms. Researchers manually segmented the voice recordings into utterances (Jaiswal et al., 2020).

An analysis of self-reported stress revealed a desired impact of finals season producing higher levels of stress than the after finals period (Jaiswal et al., 2020).

Multiple separate unimodal deep neural network models were used for predicting emotional valence and activation; the emotion recognition analysis used acoustic, lexical (LIWC, detailed above, was used for these features), thermal, and close-up video features (Jaiswal et al., 2020). The acoustic modality seemed to carry the most information about activation, while the lexical modality seemed to carry the most information about valence (Jaiswal et al., 2020). When Jaiswal et al. (2020) merged all the networks together using late voting on each modality (decision fusion), they determined that the combination of all modalitys performed the best for predicting activation. A combination of acoustic, lexical, visual, and thermal features was best at predicting valence (Jaiswal et al., 2020).

Deep neural networks performing binary classification were used for predicting stress (Jaiswal et al., 2020). Acoustic, lexical (LIWC again used), thermal, close-up video, upper body video, and

physiological features were used in the stress analysis (Jaiswal et al., 2020). Stressed vs. non-stressed was best differentiated with audio and physiological features analyzed together, but the modalies for predicting emotion were also found to be good predictors of stress (Jaiswal et al., 2020).

The speech datasets were recorded to capture both stress and emotion separately and do not account for any interdependence (Jaiswal et al., 2020).

4 Modeling Naive Psychology of Characters in Simple Commonsense Stories

Understanding a story requires reasoning about the causal links between the events in the text and the mental states of the characters, even when these relationships are not explicitly elaborated (Rashkin et al., 2018). While humans can do this quite naturally, this reasoning is very hard for statistical and neural machines – most powerful language models have been designed to effectively learn local fluency patterns and generally lack the ability to abstract beyond the surface patterns of text in order to model more complex implied dynamics (Rashkin et al., 2018). In their study, Rashkin et al. (2018) attempt to produce a NLP solution capable of densely labeling commonsence short stories in terms of the mental states of the characters involved. Their resulting manually-annotated dataset (spanning 15,000 stories) provides a fully-specified chain of motivations and emotional reactions for each story character as pre- and post-conditions of events. The dataset annotations, developed with Amazon Mechanical Turk, include state changes for entities even when they are not mentioned directly in a sentence, and encompass formal labels from multiple theories of psychology as well as open text descriptions of motivations and emotions (Rashkin et al., 2018).

Rashkin et al. (2018) used Maslow's hierarchy of needs with Reiss's basic motives acting as subcategories of Maslow's categories for their motivational classifications. Plutchik's wheel of emotions was used for emotional classifications (Rashkin et al., 2018). Binary labels were predicted (using logical regression) with state classifiers for each of Maslow's needs, Reiss's motives, and Plutchik's categories (Rashkin et al., 2018).

All models used by Rashkin et al. (2018) out-performed the random baseline, and a performance boost was observed after adding entity-specific contextual information, indicating that the model learned to condition on a character's previous experience. Pretraining the encoder parameters using free response annotations from the training set also offered a clear performance boost for all models on all three prediction tasks (Rashkin et al., 2018). The best performing models were most effective at predicting physiological needs, food motives, and joy reactions (Rashkin et al., 2018). A simple model could map open text responses to categorical labels (Rashkin et al., 2018).

5 Leveraging Hashtags to Provide Insight into Motivational States

Language allows an individual to express information about their cognitive state, desires, opinions, and motivations (Tomlinson et al., 2014). A "motivational act" is an utterance by an individual that either informs their motivation for action or affects the motivation of another individual (Tomlinson et al., 2014).

Tomlinson et al. (2014) chooses to evaluate social media because of the new forms of discourse and enormous quantities of communications available, believing that Twitter especially provides a rich source of evidence about how people express their personal states. Tomlinson et al. (2014) also

note that Tweets, available in massive quantities, convey many different forms of information written by people across the spectrum of socio-economic and cultural backgrounds. Yet, Twitter brings with it its own unique set of challenges. Tweets often contain spelling mistakes, non-traditional grammatical usages, and shorthand (Tomlinson et al., 2014). Tomlinson et al. (2014) set out to model Tweets containing language expressing goals and rewards, tapping into research of private states: beliefs, opinions, sentiments, and desires of individuals.

Tomlinson et al. (2014) employed distant supervision for their research, using a small set of annotations linking to a larger knowledge base containing noisy instances of those annotations (Tomlinson et al., 2014). Hashtags were the subjects of these annotations, and were chosen because previous research showed hashtags can be used to signal complex personal states or clarify the meaning of a post and are often not subject to the afformentioned challenges Twitter brings with it (Tomlinson et al., 2014). The model was then trained with the language contained in the Tweets marked by those hashtags (Tomlinson et al., 2014).

The types of motivational acts in the study were derived from work in psychology aimed at understanding how individuals perceive intentionality in others and what factors change an individual's motivation (Tomlinson et al., 2014). Tomlinson et al. (2014) looked at comments that express goals or indicated a goal orientation (defined as "goals"), evidence that an individual has (or thinks they have) skill or control to act within their environment (defined as "control"), and expressions that indicate positive social value for the individual's work (either as self-directed rewards or reward statements directed at other individuals; defined as "rewards"). Goals encode an individual's desire for an event or reward associated with an event's outcome (Tomlinson et al., 2014). Expressions of goals are important not just for understanding motivation, but for insight into the probability of success as well (Tomlinson et al., 2014). Control acts indicate skills and control over a situation, but can also cover expressions indicating difference in an individual's perception of their locus of control (e.g., a lack of control) (Tomlinson et al., 2014).[1] Control acts are subcategorized into expressions of control, expressions of skill, and expressions indicating a lack of control (Tomlinson et al., 2014). Rewards interact with goals and intentions through prospect theory (Tomlinson et al., 2014). There are three concepts affecting how individuals value a reward:

1. reference points: rewards are valued in how far they deviate above or below a given reference point;

2. loss aversion: avoiding a loss is treated as being more important than an equivalent gain; and

3. diminishing sensitiveity: the value of a change is not linear, but decreases as the point gets further from a referent (Tomlinson et al., 2014).

Rewards can also come from support (or, much less commonly, negativity) from community members (Tomlinson et al., 2014). Tomlinson et al. (2014) used four subtypes of reward for the purposes of their study: self-directed positive rewards, self-directed negative rewards, self-directed other rewards, and other-directed positive rewards. The ways in which individuals utilize these motivational acts reflects their underlying motivation to perform (Tomlinson et al., 2014).

Tomlinson et al. (2014) generated a set of hashtags used to mark language exhibiting one of the motivational acts and analyzed 7.5 million Tweets containing one of the hashtags or related to a hashtag of interest by being Tweeted by the same author as a previously analyzed Tweet. URLs,

[1] As an aside, individuals that perceive themselves as being in control of an event are more likely to expend more effort towards manifesting an event outcome (Tomlinson et al., 2014).

hashtags, and user mentions were removed from Tweets to prevent them from influencing the classifier (Tomlinson et al., 2014). A naive-Bayes language model for each motivational act was used to classify the data, with Tweets containing an appropriate hashtag used as positive data and a random sampling of Tweets containing inappropriate hashtags used as negative data (Tomlinson et al., 2014). The accuracy of the resultant classifier suggest the annotation procedure was accurate at identifying tweets that had strong similarities in the language that was used in the Tweet and the language expressing each motivational action could be accurately captured by the model (Tomlinson et al., 2014). The model even revealed new hashtags not originally considered as associated with the investigated acts, indicating that the model could be used to identify new hashtags associated with cognitive factors; an iterative distant-supervised annotation procedure would be extremely beneficial, as the system could propose new hashtags which could be evaluated by an annotator and reincorporated into the system, potentially improving accuracy (Tomlinson et al., 2014). That being said, sarcasm proved a thorn in the side of the classifier, often falsely classified as the oposite sentiment (Tomlinson et al., 2014).

Tomlinson et al. (2014) stress that their model can generalize beyond tweets and acknowledge that the surrounding context of Tweets is hard to define on Twitter.

6 Discussion

The research of Jaiswal et al. (2020) shows how machine learning can be employed to determine stress and emotional state, opening the door to natural language processing determining these states from analyzed text, while the research of Rashkin et al. (2018) show that machine learning can be employed to discern the needs, motives, and emotions of characters in passages and Tomlinson et al. (2014) leverage machine learning to identify motivational states. But, there are also notable areas of technology that need to progress before the methodologies of each of these studies are more practically deployable. Boyd and Schwartz (2021) make the clear point that contextual word embeddings (or another means of incorporating context) is important for ”words as attention” approaches to gain more value. The manual annotation of Rashkin et al. (2018) provided context to a model, but a shortcoming to this approach is the amount of work needed to prepare the data for the experiment. Similarly, while Jaiswal et al. (2020) (rather commendably) produced lexical data complimented with a broad array of other measures intended to provide holistic context, this process also required a high degree of manual processing of information, not to mention an intensive data collection process. Sure, Jaiswal et al. (2020) and Rashkin et al. (2018) intend for their corpora to be reusable by other researchers, but each of these studies established new corpora aligned around somewhat specific aspects of psychological state – a researcher looking to investigate a different aspect of psychological state would have to procure or generate an additional corpora of data geared at whatever new area of psychology they wished to investigate. It will take more pioneering studies like these to establish these dataset or a fundamental change eliminating the volume of manual preprocessing altogether for NLP investigations into psychological state and its impact on language to truely take off.

Boyd and Schwartz (2021) and Tomlinson et al. (2014) differ on their perspectives on using social media for NLP, allowing for a good conversation about the trade-offs of the medium. Both researchers agree that social media isolates posts from the context that inspired them – a significant drawback to the medium – but while Boyd and Schwartz (2021) seems to feel they are not as valuable for understanding psychological state, Tomlinson et al. (2014) is of the opinion the benefits outweigh the negatives. This divergence of opinion stems largely from the firm conviction of

Boyd and Schwartz (2021) that context will play a crucial role in the expansion of NLP to include psychological state in its evaluations. Tomlinson et al. (2014), meanwhile, primarily believe social media's value lies in its vast, richly diverse cache of expressions, presumably believing these qualities redeem the lack of context. While their is no "right" answer here, Boyd and Schwartz (2021) does concede that it's difficult to secure quality data of interactions between two or more people outside of the context of mass medias. Ultimately, the belief of Boyd and Schwartz (2021) that context is important for significant steps likely has merit, but the reality that social media remains a powerful and easy-to-leverage tool means it likely deserves to continue playing a role in NLP moving forward.

7 Conclusion

Ultimately, natural language processing will need to contain mechanisms to effectively assess psychological state in order to more effectively understand the meaning behind utterances. While current investigations are promising, they remain plagued with issues of omissions of context and intensive manual preprocessing or concurrent contextual data gathering. However, as more and more researchers turn their attention towards this untapped aspect of natural language, barriers to research into the field will grow lower and lower.

References

Ryan L. Boyd and H. Andrew Schwartz. 2021. Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present, and Future States of the Field. *Journal of Language and Social Psychology*, 40(1):21–41. Publisher: SAGE Publications Inc.

Mimansa Jaiswal, Cristian-Paul Bara, Yuanhang Luo, Mihai Burzo, Rada Mihalcea, and Emily Mower Provost. 2020. MuSE: a Multimodal Dataset of Stressed Emotion. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1499–1510, Marseille, France. European Language Resources Association.

Hannah Rashkin, Antoine Bosselut, Maarten Sap, Kevin Knight, and Yejin Choi. 2018. Modeling Naive Psychology of Characters in Simple Commonsense Stories. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2289–2299, Melbourne, Australia. Association for Computational Linguistics.

Marc Tomlinson, David Bracewell, Wayne Krug, and David Hinote. 2014. #mygoal: Finding Motivations on Twitter. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 469–474, Reykjavik, Iceland. European Language Resources Association (ELRA).