RYAN L. BOYD AND PAUL KAPOOR

PSYCHOLOGICAL PROFILING AND EVENT FORECASTING USING COMPUTATIONAL LANGUAGE ANALYSIS

Psychologists have long believed that we can discern what makes a person tick by analysing their language. The modern study of language has become a highly sophisticated area of research that leverages computational modelling, objective measures of language, and extensive empirical rigor.

The links between a person's mental processes and the words that they say or write have been extensively studied, validated, and applied to fields as diverse as computer science, medicine, sociology, and anthropology, to name just a few. The ability to 'get inside a person's head' by analysing their language patterns from a distance has tremendous appeal and several practical applications, ranging from the patently obvious to the surprisingly nuanced.

SUBSTANCE VERSUS STYLE

In research on the psychology of language, most scientists have traditionally focused on the substance or content of language – words that have an explicit meaning (e.g., house, friend, bomb, etc.). Unsurprisingly, there are several direct links between what a person talks about and what they are thinking. Extroverts tend to use more words related to social processes and use more positive language. Neurotic people tend to use more words indicative of anxiety, and so on.

These (perhaps obvious) patterns are grounded in psychological theory and can therefore be extrapolated to a broader understanding of the individual. The content of a person's language is reliably diagnostic of their intelligence, political orientation, personality characteristics, and even how long they live. The things that occupy a person's mind are not merely diagnostic of their thoughts – they are indicative of deeply ingrained patterns in their life.

Perhaps more interesting, however, is the style of a person's language. An abundance of research in recent years has found that the small, throw-away words in language like articles, pronouns, conjunctions, and so on, are deeply revealing of lower-level psychological processes. People whose language is highly self-focused (e.g., high rates of words like 'I', 'me', and 'my') tend to be relatively insecure and depressed. People who use more articles ('the', 'a', 'an') and prepositions ('in', 'by', 'across') in their writing tend to be more analytic in their thoughts, and factors such as someone's social status and authenticity tend to be reflected in a person's linguistic style more than its content.

By pairing the analysis of what a person says with how they say it, we can often paint a remarkably detailed picture of a person's mental and social universe. Such analyses can be performed extremely quickly and objectively using computational tools, and many psychological phenomena can be reliably estimated using relatively simple statistical models.

PSYCHOLOGICAL PROFILING

Much of the work in computational psychological profiling is founded on research demonstrating that linguistic patterns are relatively stable across time and contexts, particularly the stylistic components of language. The quality of language-derived psychological profiles can range from speculative to unbelievably strong, sometimes allowing us to identify an author with near-perfect accuracy using only their language. A language-driven approach to profiling allows us to understand the person behind a given text rather than just the text itself.

Rather than simply taking a threat of violence at face value, we can computationally evaluate the speaker's language for deeper clues. Are they at-risk for a future depressive or schizophrenic episode? Are they obsessive-compulsive, or perhaps prone to conspiratorial thinking? Statistical estimation of these types of psychological vulnerabilities can help to highlight critical intervention strategies.

Language-based psychological profiles can also be applied at the group level, revealing fundamental differences in how group members think and engage with the world. Recent research found that Islamic State, as a group, shows greater authoritarianism and religious fervour in their psychological profile (revealed by markers such as low rates of present-focused and tentative language, plus high rates of religious language) relative to al-Qaeda. Moreover, study participants who scored high on authoritarianism and religiosity reported more favourable attitudes towards the language of Islamic State compared to the language of al-Qaeda. Understanding such group differences can provide insights into how a group functions, as well as what types of people might find these groups appealing.

Psychological profiles can also be built for broader communities and monitored over time. The psychological health of a community can easily be tracked following a tragedy using various data sources, such as newspapers or social media. Research from several labs has investigated the psychological impact of calamitous events ranging from the 9/11 attacks to mass shootings, finding unique patterns of coping as they unfold in response to major upheavals.

BEHAVIOURAL FORECASTING

In a vacuum of information about an author, the statistical analysis of language can give us important clues about a person's future behaviours. Often, this approach relies on the relationship between language and general behavioural patterns. For instance, we find that the language representative of someone's core values (e.g., family, work ethic, empathy) are strongly related to their regular behaviours, such as attending religious functions, donating time/money to a cause, or even playing games online.

Most of the recent work in language-based behavioural forecasting focuses on interpersonal behaviours. Messages written prior to events like suicide or spree killings show distinct psycholinguistic fingerprints. When soliciting sex from minors online in sting operations, individuals who exhibit high certainty and planning markers in their language are at high risk for repeat-offending in the same crime categories (e.g., acquisition of child pornography, future attempts to solicit minors).

Similarly, research on group processes finds that linguistic cues related to planning decrease immediately before the betrayal of an ally (along with an increase in positivity and politeness). A failure to linguistically adapt to a changing group membership tends to precede members exiting a group, and changes in interpersonal linguistic coordination can foreshadow the initiation, stability, and dissolution of a relationship.

WHERE DO WE GO FROM HERE?

The implications of language-based profiling and behavioural forecasting are far-reaching and can represent a double-edged sword. The same language data can be leveraged for multiple purposes, and care must be taken to protect the words of vulnerable or high-profile individuals. Individuals working in the security sector who are psychologically vulnerable can be identified from their language patterns, resulting in a non-negligible risk for targeted exploitation.

Sources who have insight into future plans require particularly high discretion. A person who knows of impending policy changes or upcoming events may show extremely subtle changes in their language patterns. Such changes are often not discernible to an untrained observer yet can still be detected using modern computational techniques. Compartmentalisation of knowledge may serve as insulation from detection, yet this approach may not be feasible (or desirable) in many situations.

Future research will continue to discover still-unknown links between language and psychology, meaning that language data from any period can be revisited extensively and mined for new insights. Language data is one piece of the puzzle, and recent work that integrates language-derived profiles with other known factors (such as age, political affiliation, and images) show significant promise for advancing the field. Reliable obfuscation techniques remain to be developed and will likely be reactive (rather than proactive) as new methods for language analysis continue to emerge.

Dr Ryan L. Boyd is a computational social scientist and behavioural scientist at the University of Texas at Austin. His research involves the inference of motives and psychological patterns from verbal behaviour. Paul Kapoor is a Senior Principal Systems Engineer at the Northrop Grumman Corporation and a former US Navy Civil Servant.