Personality and Individual Differences 134 (2018) 119–124

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Linguistic markers of moderate and absolute natural language

Mohammed Al-Mosaiwi*, Tom Johnstone

Department of Psychology, School of Psychology and Clinical Languages, University of Reading, UK

ARTICLE INFO

Keywords:
Extreme responding
Absolutism
Text analysis
Natural language
Machine learning

ABSTRACT

In social, personality and mental health research, the tendency to select absolute end-points on Likert scales has been linked to certain cultures, lower intelligence, lower income and personality/mental health disorders. It is unclear whether this response style reflects an absolutist cognitive style or is merely an experimental artefact. In this study, we introduce an alternative, more informative, flexible and ecologically valid approach for estimating absolute responding, that uses natural language markers. We focussed on 'function words' (e.g. particles, conjunctions, prepositions) as they are more generalizable because they do not depend on any specific context.

To identify such linguistic markers and test their generalizability, we conducted a text analysis of online reviews for films, tourist attractions and consumer products. All written reviews were accompanied by a rating scale (akin to Likert scale), which allowed us to label text samples as absolute/moderate. The data was split into independent 'training' and 'test' sets. Using the training set we identified a rank order of linguistic markers for absolute and moderate text, which were evaluated in a classifier on the test set. The top three markers alone ("but", "!" and "seem") produced 88% classification accuracy, which increased to 91% using 31 linguistic markers.

1. Introduction

In social, personality and mental health research, absolute responding (or 'extreme' responding) is a response style estimated using Likert type scales. Where selecting the absolute endpoints of a scale (e.g. 1 and 5 on a 5-point scale) corresponds to absolute responding, while selecting any point in-between corresponds to non-absolute or moderate responding. This study aimed to identify linguistic markers which act as surrogates to absolute and moderate responding on Likert scales. These markers could expand our understanding of both the language and cognition related to absolute and moderate responding. The language we use has previously been shown to relate to the way we think (e.g., Al-Mosaiwi & Johnstone, 2018). In measuring absolute and moderate responding, linguistic markers are also a more informative and ecologically valid alternative/addition to using Likert scales.

1.1. Absolute responding using Likert scales and the limitations

Absolute responding on Likert-scales has been linked to a number of cognitive, social and cultural factors. Lower IQ and less education (e.g., Light, Zax, & Gardiner, 1965; Marin, Gamba, & Marin, 1992) have been associated with more absolute responding, as have personality characteristics such as intolerance of ambiguity and simplistic thinking (e.g., Naemi, Beal, & Payne, 2009).

Greater absolute responding has also been linked to 'black' and 'Hispanic' cultures (e.g., Bachman, O'Malley, & Freedman-Doan, 2010; Hui & Triandis, 1989; Marin, Gamba, & Marin, 1992); while lower absolute responding (more moderate responding) is linked to Japanese, Chinese (e.g., Chen, Lee, & Stevenson, 1995) and Korean cultures (e.g., Chun, Campbell, & Yoo, 1974). On closer inspection, these cultural findings often depend on the size of the scale used; an observed effect on a 5-point scale may not be apparent on a 10-point scale (e.g., Clarke, 2000; Hui & Triandis, 1989). Such inconsistencies naturally raise doubts about the veracity of the results.

Additionally, a series of studies with depressed participants reveal that both positive and negative absolute responses on Likert scales predicted future relapse (e.g., de Graaf, Huibers, Cuijpers, & Arntz, 2010; Peterson et al., 2007; Teasdale et al., 2001). However, other studies have failed to find the effect (Ching & Dobson, 2010), or raised methodological concerns regarding the use of Likert scales, specifically in reference to the effect of item content on response style (Forand & DeRubeis, 2014). That is, the content of the questions and the labelling of the end-points (e.g. "Mostly agree"), could compromise the absolute nature of an end-point response. This moderating effect would not be accounted for when simply measuring the number of end-point responses.

These previous findings have exclusively relied on observing an absolute response style on Likert scales. This simplistic method cannot

* Corresponding author at: University of Reading, Harry Pitt Building, Earley Gate, Reading RG6 6AL, UK.
E-mail address: m.a.a.almosaiwi@pgr.reading.ac.uk (M. Al-Mosaiwi).

https://doi.org/10.1016/j.paid.2018.06.004
Received 18 November 2017; Received in revised form 13 May 2018; Accepted 2 June 2018

be applied to qualitative data, it lacks ecological validity, and there is no evidence as to whether the findings generalize beyond Likert scales. That is, it is not clear whether the absolute responding of some groups relates to meaningful differences in absolutist thinking, or simply an experimental artefact specific to using Likert scales.

Our proposed method of measuring absolute responding through linguistic markers in natural language presents an alternative that avoids some of the limitations inherent to Likert scales. Being based on complex, naturalistic data (natural language), it offers greater flexibility and ecological validity because it is not reliant on structured response formats and can be used in an observational study of data acquired from a wide variety of sources.

1.2. Function word linguistic markers

To be generalizable, linguistic markers cannot depend on the content of any given subject (e.g. nouns, verbs, adjectives), as these will differ from one subject to another. Therefore, we restrict our feature selection to include only 'function words', which have a grammatical and structural role, but convey little to no content (e.g. particles, conjunctions, prepositions). Ordinarily, we attend to the content of language and have little conscious awareness of its functional style. For this reason, function words have previously been examined as implicit measures, particularly for differences in writing style (for review see Tausczik & Pennebaker, 2010).

Text analysis studies have associated specific classes of function words with certain writing styles. For example, conjunctions, negations, articles and prepositions have been associated with a categorical or formal language style (Chung & Pennebaker, 2007). Exclusive words (e.g. "but", "except", "without"), conjunctions and prepositions have been shown to be markers of greater 'cognitive complexity' (Pennebaker & King, 1999). Increased use of auxiliary verbs, pronouns and adverbs are characteristic of a narrative language style (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014). Personal pronouns predictably indicate a self-focus; while it is suggested that third person pronouns (they, he, she) are a sign of wellbeing (Chung & Pennebaker, 2007).[1] We aim to extend this literature by identifying function words which correlate with absolute and moderate responding on Likert-type scales.

1.3. Machine learning classification

Text analysis combined with machine learning has regularly been used to classify natural language text linked to positive vs. negative ratings (e.g., Feldman, 2013); this is referred to as 'sentiment analysis'. In this study, we followed the same process, except we were interested in absolute/moderate ratings differences, rather than positive/negative. The purpose of building a classifier, similar to those previously used for valence classification, was to demonstrate the predictive accuracy of the linguistic markers we identified in the training set.

2. Methods and data analysis

2.1. Data collection

The internet is increasingly being used as a source of naturalistic writing for research in linguistics and psychology. Many websites allow users to leave lengthy comments in the form of personal narratives, requests for help, or reviews. In this study, we collected natural language text posts from three popular websites; IMDB, TripAdvisor and Amazon. All three websites combine a star rating system (akin to a Likert scale) with written natural language reviews about films, holiday destinations or products respectively. Reviews paired with the lowest or highest (end-point) ratings were labelled absolute, and all other reviews were labelled non-absolute (or moderate). The valence of the reviews (positive or negative) was not factored into the analysis. This means that absolutely positive reviews were grouped with absolutely negative reviews as they were both absolute. Convergent validity in absolute responding between Likert scales and natural language was therefore estimated using the star rating scales and the text posts of these websites.

We selected the websites IMDB, TripAdvisor and Amazon as they were large enough to provide sufficient data for training and testing with our classifier approach. All three websites currently have the most web traffic in their respective domains of 'Arts and Entertainment', 'Travel' and 'Shopping' as shown by *www.similarweb.com*. We selected websites from three completely different industries, so that the linguistic markers identified would be less dependent on any particular context. In IMDB, users commented on films, for TripAdvisor they wrote about tourist destinations and on Amazon they reviewed everyday products. From each website we selected 18 films, tourist attractions and products, respectively. Generally, our selection procedure was to first identify the films, tourist attractions and products with the most overall number of reviews. We then singled out those that had the broadest ratings distributions (i.e. not predominantly positive or negative). This was to ensure a reasonable sample size could be collected at each level of the star rating scale. Additionally, we were keen to select films, tourist attractions and products from wide mix of different genres, countries and categories (respectively[2]).

For each film, tourist attraction and product, we gathered the written text accompanying each star rating. We aimed to collect 15,000 words for each level of star rating for all films, attractions and products. Where this was not possible, we simply collected all the available reviews, ensuring a minimum of 3000 words were sampled. These were copied and pasted into a single text file. For TripAdvisor and Amazon, reviews are rated on a 5-point scale, this resulted in 90 text files ($18 * 5$) from each website. IMDB was a slight exception, where the star rating scale ranges from 1 to 10 (not 1–5), so we generated 180 text files ($18 * 10$) for this website.

To reduce the IMDB 10-point scale to match with the Amazon and TripAdvisor 5-point scales, we first aligned the absolute end-points. For both scales, 1-star meant absolute negative. Absolute positive is 10-stars for IMDB but was reassigned to 5 to match the TripAdvisor and Amazon 5-point scale (i.e. 1-star - > 1-star; 10-stars - > 5-stars). We next determined that the central values on the 10-point scale (that corresponding to '3' on a 5-point scale) were between 5 and 6, these were reassigned as 3 (i.e. 5-stars - > 3-stars; 6-stars - > 3-stars). This meant that 2–4 stars on a 10-point scale, which are neither absolutely negative, nor central, corresponded to 2-stars on a 5-point scale. Similarly, ratings 7–9 stars on a 10-point scale, which are neither absolutely positive, nor central, corresponded to 4-stars on a 5-point scale. This realignment achieved our main objective of preserving the integrity of the absolute end-points (e.g. not combining 9-stars with 10-stars, as 9-stars is not an absolute).

2.2. Data-analysis in R

We used R programming language (R Development Core Team, 2010) to conduct the text analysis and measure function word usage by dividing text into unigrams (single words). For our training set, we identified unigrams which best differentiated between absolute and moderate natural language. These would then be used in machine learning classification, on an independent test set, to automatically label text as either absolute or moderate.

[1] More information on the grammatical and structural role of particular classes of function words is provided in the Supplementary Material.

[2] All data are available at 10.6084/m9.figshare.6199235