

Predicting Personality and Psychological Distress Using Natural Language Processing

Authors : Jihee Jang, Seowon Yoon, Gaeun Son, Minjung Kang, Joon Yeon Choeh, Kee-Hong Choi

Abstract : Background: Self-report multiple choice questionnaires have been widely utilized to quantitatively measure one's personality and psychological constructs. Despite several strengths (e.g., brevity and utility), self-report multiple-choice questionnaires have considerable limitations in nature. With the rise of machine learning (ML) and Natural language processing (NLP), researchers in the field of psychology are widely adopting NLP to assess psychological constructs to predict human behaviors. However, there is a lack of connections between the work being performed in computer science and that psychology due to small data sets and unvalidated modeling practices. Aims: The current article introduces the study method and procedure of phase II, which includes the interview questions for the five-factor model (FFM) of personality developed in phase I. This study aims to develop the interview (semi-structured) and open-ended questions for the FFM-based personality assessments, specifically designed with experts in the field of clinical and personality psychology (phase 1), and to collect the personality-related text data using the interview questions and self-report measures on personality and psychological distress (phase 2). The purpose of the study includes examining the relationship between natural language data obtained from the interview questions, measuring the FFM personality constructs, and psychological distress to demonstrate the validity of the natural language-based personality prediction. Methods: The phase I (pilot) study was conducted on fifty-nine native Korean adults to acquire the personality-related text data from the interview (semi-structured) and open-ended questions based on the FFM of personality. The interview questions were revised and finalized with the feedback from the external expert committee, consisting of personality and clinical psychologists. Based on the established interview questions, a total of 425 Korean adults were recruited using a convenience sampling method via an online survey. The text data collected from interviews were analyzed using natural language processing. The results of the online survey, including demographic data, depression, anxiety, and personality inventories, were analyzed together in the model to predict individuals' FFM of personality and the level of psychological distress (phase 2).

Keywords : personality prediction, psychological distress prediction, natural language processing, machine learning, the five-factor model of personality

Conference Title : ICC 2023 : International Conference on Convergence Computing

Conference Location : London, United Kingdom

Conference Dates : August 17-18, 2023