Summary
Predicting Personality from Twitter

Submitted by:
Pallavit Aggarwal
Intelligent Systems – 22333721

The paper utilizes twitter feed, and analyzes the textual data from the tweets to identify user personalities.

It sets the base by discussing multiple studies that show that the personality can be broken down into the BIG 5 personality traits which are Extraversion vs. introversion, Agreeableness vs. antagonism, Conscientiousness vs. lack of direction, Neuroticism vs. emotional stability, Openness vs. closedness. This is important to mention as it is a generic model that is used by multiple studies to judge and categorize personality types and is known to be generic across race, age and culture.

The authors considered 50 subjects for their study on predicting personality using twitter feeds, and asked each subject 45 questions from the Big 5 inventory of 45 questions and took all or 2000 of their most recent tweets.

A tweet was limited to 140 characters, and they joined all the tweets together for text analysis, but this fell short on properly evaluating personality conclusively as these tweets were a chain of disconnected and time-spaced thoughts.

The heuristics that were collected include – number of followers, number of people user is following, number of mentions, replies hashtags etc. To generate more insights and usable parameters, Linguistic tools like LIWC and the General Inquirer were also used.

LIWC produces 81 distinct features, but they excluded 2 features which they considered irrelevant and noisy. The General inquirer produced a score on the -1 to 1 scale, which was the average sentiment score of the tweets that they collected.

A Pearson Correlation analysis next reveals interesting patterns, like aggregable people tend to use the word “you” a lot that others, and that people considered conscientious didn’t use words with negative connotations, but these correlations were not considered deeply as they deviate from the element of prediction of personality, and tend more towards nit-picking individual variables that define an element in the personality and its attribution to other elements in the personality traits.

Using Gaussian Process and ZeroR on the above generated feature set, they were able to predict the five personality traits within 11-18% of their actual values. The traits of openness were easily identifiable owing to the fact that openness could be analysed through friendliness sentiments in the tweets and the number of followers and other soft metrics that showed an agreeable and friendly demeanor. And, neuroticism which might be convoluted to assess, as some tweets could be facts, or random chain of thoughts, or even sarcastic comments, was the hardest to compute.

Authors concede that a bigger sample space could've helped improve the results, and state that the research was set in the right direction and the tools used were apt and generated the right kind of data to assess personality traits.

There are multiple use cases of this type of prediction namely – to improve recommendation systems, web and app interfaces could be tweaked according to individual's personality type to make it more appeasing for the users, and finally advertising could benefit hugely by identifying the personality traits of a user.

Two other papers that shed more light into above subject matter are "Using linguistic cues for the automatic recognition of personality in conversation and text" (runs classification and regression on essay corpus to indentify the mood and personality characteristic elements) and "Language use as an individual difference" (that evaluates linguistic styles to determine personality traits).

References

1) General inquirer

2) Big five inventory

3) LIWC tool

4) F. Mairesse, M. Walker, M. Mehl, and R. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research, 30(1):457–500, 2007.

5) J. Pennebaker and L. King. Linguistic styles: Language use as an individual difference. Journal of personality and social psychology, 77(6):1296–1312, 1999.