Place and Personality

Vlad Tanasescu, Christopher B. Jones, Gualtiero Colombo, Martin J. Chorley, Stuart M. Allen, Roger M. Whitaker

School of Computer Science & Informatics,
Cardiff University
Cardiff, United Kingdom
{v.tanasescu, c.b.jones, g.colombo, m.j.chorley, stuart.m.allen, r.m.whitaker}@cs.cardifff.ac.uk

Abstract

An individual's personality according to the Five-Factors ('Big Five') model of personality can be described in terms of *openness, conscientiousness, extraversion, agreeableness* and *neuroticism*. This work adapts human personality traits research to places, using keywords extracted from reviews as indicative of personality dimensions.

Keywords: place, personality, reviews, data mining, big five, five factor model.

1   Introduction

Personality research in psychology states that personality characteristics that are most important in peoples' lives will eventually become a part of their language. Reviews of places are texts reflecting the 'perception' or 'feeling' of a place, usually accompanied by a rating. This research aims to capture this underlying 'feeling' of a place, or personality of a venue, in a way related to personality research in human psychology.

The Five Factor Model (FFM), also called 'Big 5' personality traits, uses five dimensions usually represented by roman numerals (I-V). For mnemonic purposes the acronym OCEAN is also used. OCEAN stands for descriptions of the five dimensions as *Openness* (inventive or curious vs. consistent or cautious), *Conscientiousness* (efficient or organized vs. easy-going or careless), *Extraversion* (outgoing or energetic vs. solitary or reserved), *Agreeableness* (friendly or compassionate vs. cold or unkind) and *Neuroticism* (sensitive or nervous vs. secure or confident).

Beside humans, personality traits research has been applied to non-human animals [1]. To the best of our knowledge, this constitutes the only extension of personality traits research beyond human personality, with no attempts towards other realms such as objects or places.

The research presented here applies personality research to geographic entities by assuming that venues can also be described using human personality dimensions, due to the personality of the individuals that frequent the venue, or because of characteristics of the venues. This usage is also embedded in language, which sometimes uses the same vocabulary to describe place and human personality. For example an "arty" place is expected to have a higher Openness (O) value than a "traditional" pub. A nightclub would be expected to score high on Extraversion (E), while a "quiet" coffee place should score lower on that personality dimension.

2   Approach

We follow a 'bag of words' approach in which both venues and personality dimensions are represented by unigrams (individual words). Unigrams describing places are extracted from online sources providing reviews for those places, while keywords (adjectives) representing dimensions of personality are gathered from psychology literature [2].

2.1   Venues Keywords

The sources used for review extraction are: *Foursquare, Google Places, Yelp, Qype* and *Yell*. Review text was stripped of punctuation and stopwords, uncapitalised, and part of speech (POS) tagged to filter verbs and adverbs. A negation detection algorithm is applied to the list of words to tag adjectives used in a negative sense. The algorithm is a modified version of the one presented in [3], itself inspired by NegEx [4].

2.2   Inventory of Personality Adjectives

The inventory of personality adjectives used was presented by Saucier and Goldberg in [2]. The 435-adjective inventory was obtained in their work by annotating personality related adjectives with familiarity ratings, in order to filter out the least familiar English trait personality descriptors. Following principal component analysis Saucier and Goldberg identified five clusters corresponding to five personality dimensions. Correlation scores for a selection of adjectives are given in Table 1.

Table 1 Examples of personality adjectives with FFM correlation values.

Adjective	II	I	III	IV-	V
excitable	0.22	0.3	-0.07	0.31	-0.1
friendly	0.37	0.39	-0.03	-0.17	-0.16
generous	0.4	0.15	-0.03	-0.15	0.04
independent	-0.14	0.2	0.18	-0.2	0.3
kind	0.6	0.07	0.06	0.02	0
playful	0.2	0.41	-0.12	-0.02	-0.09
quiet	0.15	-0.64	0.15	-0.09	0.12
reasonable	0.38	-0.06	0.25	-0.25	0.12
relaxed	0.21	0.17	0.04	-0.48	-0.02
traditional	0.14	-0.14	0.28	0.02	-0.36

The mapping between dimensions I to V and the OCEAN traits characterisation is given in Table 2. IV is usually inversed to represent neuroticism (IV-), i.e. high on IV- corresponds to low emotional stability. Following this correspondence, the adjective *reasonable* strongly correlates (0.38) with the OCEAN A dimension (Agreeableness) as well as with C (Conscientiousness) at 0.25, while *friendly* also correlates with A (0.37) but even more with E (Extraversion) with 0.39 as well as negatively (-0.17) with N (Neuroticism) and Openness (-0.16).

Table 2 Familiar adjectives dimensions to OCEAN correspondence.

I	Extraversion
II	Agreeableness/Benevolence
III	Conscientiousness
IV	Emotional Stability (neuroticism)
V	Intellect/Imagination (openness)

3   Matching

12 venues were selected in Cardiff, Wales, UK according to the criteria of review availability and diversity. To assign dimensions of personality to a venue, terms obtained from the review extraction were matched to the list of familiar personality adjectives.

The negation algorithm is applied to venues keywords, to obtain a negation marker (1 or -1) for each term. Venues keywords extracted from reviews as well as familiar adjectives are stemmed using the Porter stemmer. Stemmed review terms are then matched to the list of 435 stemmed familiar adjectives. In case of a match, the correlation values of the matching adjective is added, possibly negated if a negation has been detected. The resulting sum is then averaged, multiplied by the range (5), and added to the average value for a dimension (2.5).

Formally, for every adjective $a_i \in A$ found in reviews for a venue v, the number of found adjectives n, the all-ones vector, and the corresponding correlation vector c_i as established in Saucier, and the sign s_a produced by the negation detection algorithm, the resulting OCEAN review vector or_v is obtained according to Equation 1.

$$or_v = \left(\frac{\sum_1^n s_a c_i}{n} * 5\right) + (1 * 2.5)$$

Equation 1 Ocean score vector for automated personality extraction.

4   Results and Discussion

The following comments may be made on the results:
- An independent coffee place is ranking high in O, but not the highest. This is due to an abundance of agreeableness terms present in reviews, such as 'friendly' (stem: 'friendli'), which also affect the O dimension (-0.16 on O).
- A traditional pub also ranks high in O due to the term 'reasonable', qualifying quality and prices, and correlating with O (0.12) and with A (0.38)
- Night clubs rank the highest on E, although a fast food ranks the highest, due to a majority of 'friendly' (0.39 on E)
- Neuroticism is constantly below average.

5   Conclusions

This work constitutes a first attempt at a qualification of venues according to their personality using an automated matching method between reviews and personality related adjectives. It provides an experimental approach of venue personality annotation by human subjects, as well as an automated venue personality extraction approach based on reviews. A larger scale evaluation of the automated approach is needed, requiring more human experiment samples.

References

[1]   T. A. R. Weinstein, J. P. capitanio, and S. D. Gosling, "Personality in animals," in *Handbook of Personality: Theory and Research*, no. 12, O. P. John, R. W. Robins, and L. A. Pervin, Eds. 2008.

[2]   G. Saucier and L. R. Goldberg, "Evidence for the Big Five in analyses of familiar English personality adjectives," *Eur. J. Pers.*, vol. 10, no. 1, pp. 61–77, 1996.

[3]   B. Ohana, "Opinion mining with the SentWordNet lexical resource," Dublin Institute of Technology, 2009.

[4]   W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, "A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 34, no. 5, pp. 301–310, Oct. 2001.