Overview

This work contributes to the study of self-presentation in online dating systems. The larger goal is to understand effects of self-presentation via free-text on perceived attractiveness.

Motivating Question

What are the similarities and differences between how men and women self-present in online dating profiles?

Data and Tools

To date, study of free-text components has been limited to few 10s of profiles. We extend study on a dataset of:

- 500 male and 500 female profiles
- restricting reported ages between 18 and 60
- restricting to those looking for people of the opposite sex

Data was crawled from Yahoo! Personals, a popular, paid online dating site.

We used 80 features from Linguistic Inquiry Word Count (LIWC) spanning Linguistic, Psychological, Personal, Paralinguistic and Punctuation categories.

All statistical analysis were performed using XLSTAT, a statistical software for MSExcel

Analysis – Part 1

Exploratory Factor Analysis to identify systematic co-occurrence patterns among LIWC features in profiles.

Finding 1: Many similarities in overall factor structures between male and female profiles.

Analysis – Part 2

Calculated factor scores for all profiles. High factor scores imply variables in profiles have high values in a factor pattern. A profile may load anywhere on the three factors.

Performed Cluster Analysis using profile factor scores as predictors. Profiles grouped in any cluster are intended to be maximally similar in their use of the features, while different clusters are maximally distinguished.

Finding 3: Differences in male and female profile types

A k-means cluster analysis uncovered three distinct clusters for each dataset highlighting some contrasts between the sexes in their combinations of word types.

Analysis – Part 3

This research was sponsored in part by NSF DHR-IIS-0624356

We also conducted an analysis of the words used by the two genders.

Finding 4: Gender similarities in word-usages across LIWC categories

- For each LIWC category, the proportion of words used from that category versus all words in the corpus were comparable for male and female profiles.

- For almost all LIWC categories, 7 out of 10 high frequency words used by the gender groups were the same. High similarities were also seen in the usage of closed-class words (pronouns, prepositions etc.) and open-classed word categories (affect, emotion words, verb groups).

- Possible hypothesis: Self-expression might tend toward attempting homophily in online dating profiles.

This research was sponsored in part by NSF DHR-IIS-0624356