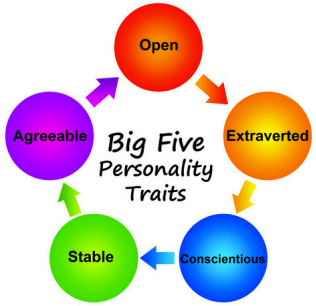
Mapping Profession and Culture to Personality:
Generating Virtual Characters from Text Data using NLP

Elissa Wolf, Camille Jwo, Xinying Xu, Supritha Bhanuvaram Ravikumar, Dr. Ani Nenkova and Dr. Norman Badler

Background

Virtual characters have an enormous range of applications including games, simulations, and training programs. Effective virtual characters must be believable, relatable, and human-like.

Previous work laid a framework for creating characters with individual differences in personality. The resulting appllication uses a character's personality (using the Big 5 model) to drive the animation of his/her motion. Additional traits such as culture and profession can be added if their relationship to personality is known.



Goals

- **Extract and quantify** personality associations that are latent in textual media
- **Determine a data-driven mapping** from culture and profession to Big 5 personality
- **Integrate the mapping** into a framework for animating realistic virtual characters
- Easily **create rich crowd simulations** that aptly represent a target population

Methods

Procedure

Use the **word2vec** tool to find associations between countries/professions and Big 5 personality descriptors.

- Word2vec tool uses a continuous bag-of-words model to learn word vectors
- Word distance: "closer" = more associated within the text
- Find closest personality words for each target profession/country
- Define an explicit country/profession -> Big 5 mapping, in which strength of each of the 5 factors is proportional to the factor's distance

Target Words

Build comprehensive lists of professions and countries, and a research-backed list of Big 5 personality words to find associations.

- **Professions**: FromWordNet synonym database
- **Countries**: CIA country/nationality factsheet
- **Big 5**: Base words from Goldberg's Big 5 Markers; extended with WordNet and human-generated synonyms

Data

- **News**: Associated Press articles
- **Blogs**: ICWSM Personal narratives

AP

ICWSM

Results

Preliminary Results

Word2vec has been successfully applied to a small set of target words for testing. Thus far, the results are promising.

Exerpt of the results:

```
scientist
   [('abstruse', '0.361806'),
   ('recluse',  '0.271816'),
   ('creative', '0.239914'),
   ('philosophical', '0.229408'),
   ('taxonomic', '0.222524'),
   ('disorganize', '0.201897')]
```

Each associated personality word is listed with its distance, or level of association.

Analysis

Objective "goodness" measures developed to track the quality of results as method is tested with different target words and data.

Analysis with preliminary professions + news data:

- Number of professions: **8**
- Number of unique words: **25**
- Number of repeated words: **11**
- Number of professions with no associated words: **0**
- Mean words per profession: **7**
- Max words per profession: **14**

Future Work

Continue applying word2vec, using expanded target word lists, and tracking quality of results. Then, develop mapping and integrate with animation framework.