Report
Group 52: Personality Recognition

A.Anuraag
Navya
Nithin
Sagar Balwani

1. Problem-statement:

To automatically classify the personality traits of the author based on his/her posts in social media.

- **Personality traits :**
  It is evident that there is a strong correlation between user's personality and the user's status text on the social media. Human personality can be summarized by five personality traits as mentioned below .

  i)Extrovert vs. Introvert (sociable, assertive, playful vs. aloof, reserved, shy)
  ii) Emotional stability vs. Neuroticism (calm, unemotional vs. insecure, anxious)
  iii)Agreeableness vs. Disagreeable (friendly, cooperative vs. antagonistic, faultfinding)
  iv) Conscientiousness vs. un-conscientiousness (self-disciplined, organized vs. inefficient, careless)
  v)Openness to experience (intellectual, insightful vs. shallow, unimaginative)
  Hence we represent human personality as a vector of the above five traits with boolean values (i.e yes/no).

- **Motivation :**
  In recent years the interest of the scientific community towards Personality Recognition has grown incredibly, since there are many applications that can take advantage of personality recognition, including social network analysis, social computing, recommendation systems, deception detection, authorship attribution, sentiment analysis/opinion mining, and others.

- **Dataset :**

There are two gold standard labeled datasets: Essays and MyPersonality .
Essays is a large dataset of stream-of-consciousness texts (about 2400, one for each user), collected between 1997 and 2004 and labeled with personality classes . This dataset can be used for identifying the feature set of the personality traits .
MyPersonality consists of about 10000 Facebook status updates of 250 users, plus Facebook network properties (including network size, betweenness centrality, density and transitivity) labeled wxith personality. This dataset can be divided into training data and test data .

2. Approach:

- **Feature selection for each of the traits :**

  A set of unigrams specific to each of the traits can be predetermined based on some metric and these unigrams can be included in the feature set for the trait . These unigrams can be extracted from the facebook data along with essays dataset. The feature set of the trait is heavily accountable for the correctness of the classifier we build .

- **Classification of the traits :**

  We build a different classifier model for each of the 5 personality traits from the training data which does the following :

  1) It distinctively classifies a user into 2 subclasses yes/no assigning a value to that trait in the personality vector of the user using **SVM**.
  2) It considers the appropriate features which carry distinctive information for personality traits recognition .

3. Challenges and Issues:

The feature set of the trait is heavily accountable for the correctness of the classifier we build . So knowing the important features related to a personality trait is the major challenge. The features related to each personality trait are different and one that is an important feature of a trait may not be that important to the other traits, in fact it may act as an overfit for those traits. As an example Network features are important for predicting the extrovert nature of a person but not necessary for agreeableness of that person. So again features should be selected with care to

overcome this issues.

4. Tools:

As of now we are only using **SVM** as a tool.

5. **Final Deliverables:**

The product takes a facebook user's statuses in a text format as input and tells what is the personality of the person!!
1) Feature set for each of the personality traits .
2) A classifier model which labels the personality traits of a user as yes/no , given a set of posts of the user. The user's personality is given as a vector of 5 dimensions with boolean traits.