IRE-Major Project Personality Recognition

Faculty:Vasudeva Varma

Mentor:Santosh Kosgi

A.Anuraag
Navya Reddy
Nithin Bharadwaj
Sagar Balwani

Problem Statement

Given a Facebook profile we have to automatically classify author's personality traits.

Five Major personality traits :

- Extrovert vs. Introvert
- Emotional stability vs. Neuroticism
- Agreeableness vs. Disagreeableness
- Conscientiousness vs. un-conscientiousness
- Openness to experience

Motivation

- Recently Social networks have become widely used and popular mediums for information dissemination and social interactions.
- Users' contributions and activities provide a valuable insight into individual behavior, experiences, opinions and interests.

System Architecture
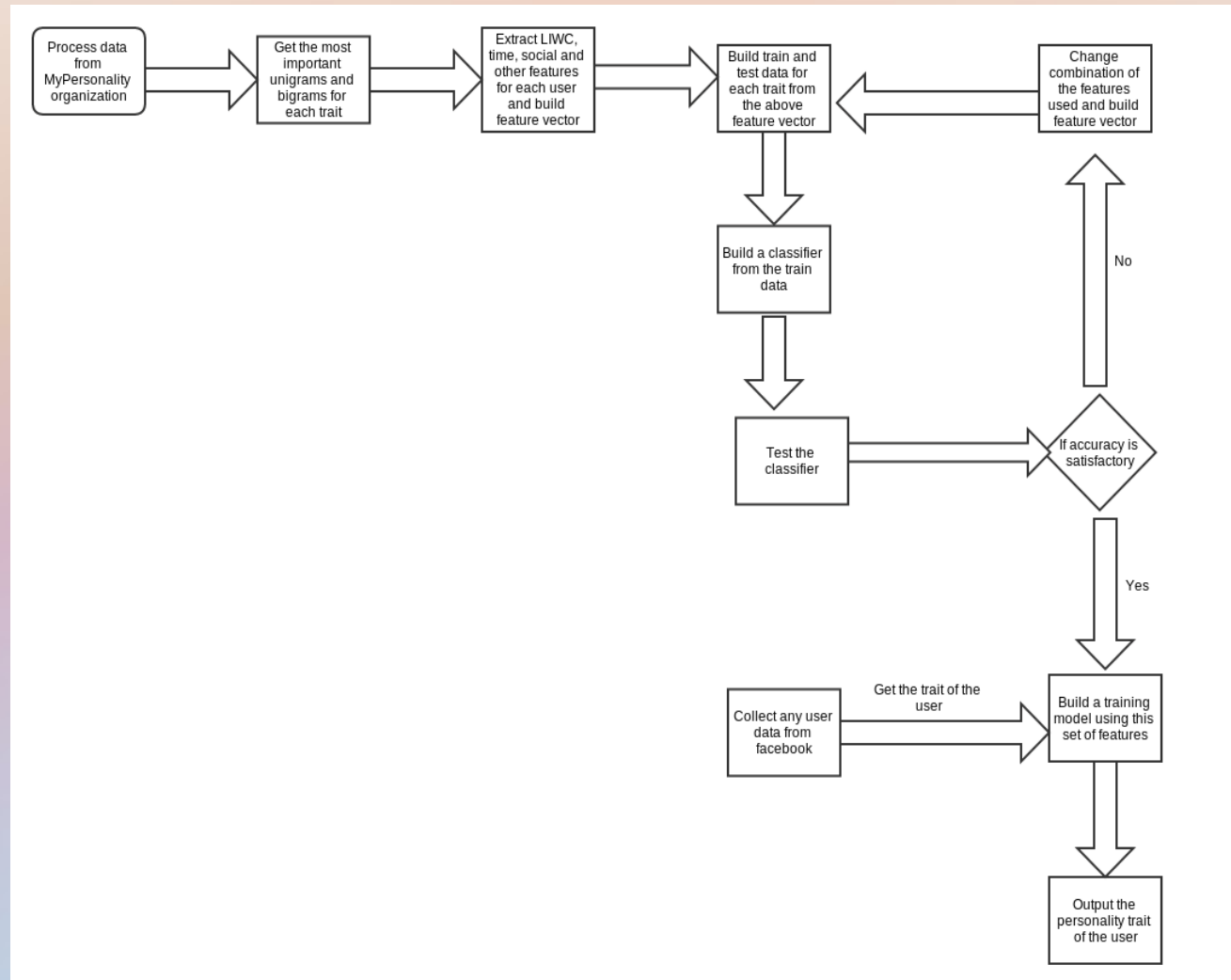
Approach

- Feature selection for each of the traits :

  A set of features specific to each of the traits can be determined based on some metric. These features can be extracted from the facebook data. The feature set of the trait is heavily accountable for the correctness of the classifier we build

Challenge

- Determining the best features related to a 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.

Features

- LIWC features: It consists of linguistic and word count features

- Unigrams & Bigrams: A set of unigrams & bigrams specific to each traits are determined from facebook and essays dataset.

- Meta Features: Features based on the writing style of the user.

- Social Features: Related to the network of the user.

Features

- Time Features: Features based on the time of the status updates of the user

- Other Features

Classification

- Training: Each user has one or more posts. All these posts are combined and a feature vector is produced and the category for each user is labelled. Using this data a classifier is built for each of the traits.

- Testing: The user data are now classified based on the classifier built. Testing is done on facebook statuses of various users.

Evaluation

- Based on the precision and recall, we repeat the process till a best combination of the features is decided for each trait.

- Finally, model for each trait is created from its corresponding feature set.

Results

Personality Trait	Best Feature set	Precision	Recall
Extrovert	unigrams+network features	0.671	0.54
Neuroticism	Affective+Language+ Network+Time+Others	0.617	0.97
Agreeableness	Language+affective+ Social+Smileys	0.619	0.8
Conscientiousn ess	Language+Biology+ Complexity+Smileys	0.617	0.80
Openness	Social+Affective+Meta-features	0.885	1.0