

CHAPTER 1

INTRODUCTION

Social networking on the web has grown dramatically over the last decade. Social networks have become widely used and popular mediums for information distribution as well as for social interactions. User activities on social networking websites provide valuable insight into individual behavior, experiences, opinions and interests. It is quite similar how a person interacts socially with the human nature and behavior.

Personality is the most intricate human attribute and it also describes the uniqueness of a person. Personality is one of the fundamental aspects, by which we can understand behavioral personalities. It has been a long-term goal for psychologists to understand human personality and its impact on human behavior. Behavior involves an interaction between a person's underlying personality traits. The situation, that a person finds himself or herself in, plays a major role on his or her reaction. However, in most of the cases, people respond with respect to their underlying personality traits. It possible to access and analyze large amounts of text samples in order to automatically identify personality types of authors and predict potential reactions and behaviors.

Humans have the tendency to understand others' behavior on the basis of the observation of their everyday behavior. Huge number of researchers around the world has been attracted to work on this research domain from different fields especially researchers in computational linguistics, psychology, artificial intelligence, natural language processing, human-machine interaction, behavioral analytics, and machine learning.

Use of social networking websites has been increased exponentially in recent times. Millions of users create profiles about themselves on social media platforms such as Twitter and use the services to connect with their friends and relatives all around the world. At Twitter, 288 million active users per month express themselves with short informal text messages called tweets which amounts to 500 million tweets in a single day.

Twitter profiles became an important source of information used to form impressions about others. In the process of creating social networking profiles, users reveal a lot about themselves both in what they share and how they say it. Through self-description, status updates, photos, and interests, much of a user's personality comes out through their profile. For example, people examine other people's Twitter profiles when trying to decide whether to start dating them, and they are also used when assessing job candidates.

1.1 Objectives

The objectives of the thesis are as follows:

- To understand sentiment analysis.
- To investigate the predictability of the personality traits of social media users base on text.
- To understand Support Vector Machine (SVM).

1.2 Field Background

The relationship between real-world social networks and personality has been usually studied using a personality test called “The Big Five”.

1.2.1 Big Five Model

The “Big Five” model of personality dimensions has emerged as one of the most well-researched measures of personality structure in recent years. Personality is defined as the coherent patterning of affect, behavior, cognition and desire over time and space, which are used to characterize unique individuals. The most widely used personality traits model in the literature is the “Big-Five” model, five broad personality dimensions (Matthews et al., 2003). It describes the human personality as a vector of five values corresponding to bipolar traits. This is a popular model among the language and computer science researchers and it has been used as a framework for both personality traits identification and simulations.

The Big-5 personality traits model is defined as follows:

- O (Openness): Artistic, curious, imaginative, curious, intelligent, and imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences.
- C (Conscientiousness): Efficient, organized, responsible, organized, and persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners.
- E (Extraversion): Energetic, active, assertive, outgoing, amicable, assertive. Friendly and energetic, extroverts draw inspiration from social situations.
- A (Agreeableness): Compassionate, cooperative, cooperative, helpful, nurturing. People who score high in agreeableness are peace-keepers who are generally optimistic and trusting of others.
- N (Neuroticism): Anxious, tense, self-pitying, anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions

1.3 Overview of the system

In this thesis, the system uses learning algorithms and advanced data mining concepts to mine user characteristics data and learn from the patterns. This thesis will come across areas where it has access to large amounts of person behavioral data. This data can be helpful to classify persons using personality prediction and classification. There are areas where there is access to large amounts of person behavioral data. This data can help us classify persons using automated personality classification.

Five characteristics of different individuals commonly known as big five characteristics namely, openness, neuroticism, conscientiousness, agreeableness and extraversion are stored in a dataset and used for training. Based on this training, the personality of individuals are predicted using data mining concepts. Before testing the dataset, it is pre-processed using different data mining concepts like handling missing values, data discretization, normalization etc. This pre-processed data can then be used to classify/predict user personality based on past classifications. The system analyses user characteristics and behaviors. System then predicts new user personality based on personality data stored by classification of previous user data. Model used to predict test dataset is “Support Vector Machine (SVM)” because SVM is an effective model to predict output class labels for dependent categorical data.

1.4 Organization of the Thesis

In this thesis, there are four chapters.

Chapter 1 describes introduction, objectives of the thesis, field background, and overview of the system.

Chapter 2 discusses about theory background in detail. It represents Support Vector Machines (SVM) model. And then it explains SVM and how it works.

Chapter 3 provides the design of the system, the flow of the system, and the implementation of the system by using SVM, and then results the output of the system.

Chapter 4 describes the conclusion, limitations and further extensions of the thesis.