Ain Shams University
Faculty of Computer and Information Sciences
Computer Science Department

Personality Traits Prediction on Social Networks

Thesis submitted as a partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences

By

Marwa Salah-Eldin Salem Khalefa

Teaching Assistant at Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.
B.Sc. of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University.

Under the Supervision of

Prof. Dr. Mostafa Aref
Professor of Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

Dr. Sally Saad Ismail
Lecturer of Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

May – 2019
Cairo

Acknowledgment

In the name of Allah the Merciful

First, I am very thankful to God Almighty for providing me with the health, patience and knowledge to complete this work.

I would like to thank the supervisors of this work, Prof. Dr. Mostafa Aref and Dr. Sally Saad for their contribution in ending this work. I am especially grateful for the encouragement and guidance given to me by Prof. Dr. Mostafa Aref. Thanks and special appreciation to my dear husband who helped me to finish this work and my dear daughter for her patience.

I convey my thanks and appreciation to my parents, my family and my friends for their patience and moral support throughout this work and for their prayers, love, patience and encouragement. Finally, I would like to thank everyone who contributed to the data compilation.

Abstract

Each individual has his own distinct character, making his own decisions, which is based on his personality. Gaining insight of a web user's personality can be very useful for many applications like recommender systems, personalized advertising and on-line marketing. Psychologists are interested in predicting the personality traits of individuals, which costs them great effort. Researchers in computer science field have tried to reach a model for extracting personality traits relying on user's profiles on social network sites as an input. Content created by users such as text posts, photos and even shared activities in social network sites are considered as a huge source of data. Researches all-over the world tried to use that rich source of data to predict personality traits in many different languages, but none of them have tried to work with the Arabic language. Even Arabic speakers represent a large sector of social media users.

In this thesis, I introduce a new-labeled dataset for Egyptian dialect twitter users (AraPersonality). It is contains timeline feed of about 92 twitter users along with their profile information. I discuss the method of gathering, annotating, properties, and statistics of the dataset. I present set of benchmark experiments using different types of machine learning algorithm, features and preprocessing to reach a best model. I reached a model with average accuracy is 66.4% and 35.8% for two different representations of the dataset.

Keywords: Social media, Personality recognition, Big five personality traits, Machine learning techniques, Artificial Intelligent.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	Agreeableness
AI	Artificial Intelligent
AV	Audio-visual features
BF	Big Five
BLR	Bayesian Logistic Regression
C	Conscientiousness
CNN	Convolutional Neural Network
DT	Decision Tree
E	Extraversion
ERC	Ensemble of Regressor Chains
ERCC	Ensemble of Regressor Chains Corrected
FFM	Five Factor Model
GRU	Gated Recurrent Unit
KNN	K-Nearest Neighbors
LDL	Label Distribution Learning
LIWC	Linguistic Inquiry and Word Count
LSTM	Long Short Term Memory
MBTI	Myers-Briggs Type Indicator
MCQ	Multiple Choice Questions
MLP	Multiple Layer Perceptron
MNB	Multinomial naive Bayes
MORF	Multi-objective random forest
MTS	Multi-target stacking
MTSC	Multi-target stacking corrected
N	Neuroticism
NB	Naïve Bayes
O	Openness
OSNs	Online Social Networks

POS	Part Of Speech
S	Sentiment features
SMO	Sequential Minimal Optimization
SNA	Social Network Analysis
SNSs	Social Networking Sites

List of Publications

1. Marwa S. Salem, Sally S. Ismail, Mostafa Aref, ”Personality Traits Recognition Methods on Facebook, Twitter and YouTube”, 3rd International Conference for Computer Science and Information System, Canadian International Collage’s Multidisciplinary International Conference, Cairo, Egypt, 2018.
2. Marwa S. Salem, Sally S. Ismail, Mostafa Aref, ”Personality Traits for Twitter Users in the Egyptian Dialect Writing Dataset”, accepted for 8th International Conference on Software and Information Engineering (ICSIE), Cairo, Egypt, 2019.
3. Marwa S. Salem, Sally S. Ismail, Mostafa Aref, ”Preprocessing The Egyptian Arabic Dialect For Personality Traits Prediction”, accepted for International Journal of Intelligent Computing And Information Sciences, Cairo, Egypt, 2019.

Chapter 1

Introduction

Chapter 1: Introduction

Social networking sites (SNSs) have gained high popularity among individuals in the past few years. It enables them to share their thoughts, feelings and daily activities, as they contain a huge amount of information about its users. This source of data cannot be underestimated. From the psychologist's point of view, words of individuals reflect who they are. This information can be used in predicting gender, age [1], personality traits [2–5], first impressions [6] and interests [7]. In this thesis, I am interested in predicting personality traits.

I can define personality as the set of habitual behaviors, emotional patterns and cognitions that evolve from environmental factors and biological, which can be used to characterize a unique individual. Many studies have been conducted by psychologists to reach a method to define human's behavior. As a result, they reached a collection of human features that identifies it, which are then called personality traits. Some researchers in psychology divided these features into four groups: sensation (S), intuition (N), feeling (F) and thinking (T) called Myers-Briggs Type Indicator (MBTI) [8]. Others divided them to five groups: openness, conscientiousness, extraversion, agreeableness and neuroticism called The Big Five (BF) or the Five Factor Model (FFM) [9], which is the most widely used nowadays.

The prediction of personality traits, which is carried out using social media websites as a source of data can be very accurate, effective and useful for many business models like virtual assistants, healthcare such as mood detection, detection of personality disorder, social network analysis, and customer profiling. Therefore, the computer science researchers aim to reach an artificial intelligent system that can make a use of the data extracted from the social media users to predict their personality traits in different languages as English [1, 10], Spanish [1, 10], Italian [1, 10], Dutch [1, 10], Chinese [11, 12], Filipino [13] and Portuguese [14].

However, there is no Arabic personality traits prediction intelligent model has been published yet, although Arabic is one of the six official languages of the United Nations Organization, which is the native language for 467 million of the world's population. However, less attention is paid to it in the field of scientific research, as it is classified within the most difficult languages in dealing with in scientific researches. In this thesis, a new-labeled dataset for Egyptian dialect twitter users is provided. I discuss the method of gathering, annotating, properties, and statistics of the dataset. I present a set of benchmark experiments.

1.1 Motivation

Nowadays, personalization of intelligent systems gains the interest of investors in many fields. As they add value in many business models, such as virtual assistants, healthcare like mood detection, detection of personality disorder, inter-personal relations, job satisfaction, professional and romantic relationship success. Predict personality for social media users can be very useful for many applications like recommender systems, personalized advertising and on-line marketing. Marketers believe in the importance of profiling the interests of those they consider as their targeted customers to achieve the best prospects for purchasing the product/service. It can be applicable by predicting the customer's personality.

1.2 Problem Definition

Despite of the importance and the massive contribution of the Arabic users in different social media platforms as they are estimated to reach 11.1 million in March 2017, almost doubling up from 5.8 million three years ago [15]. I cannot find any research on extracting the personal characteristics of the Arab social media users or a well-trained dataset that can help researchers to carry on such research.

1.3 Research Objectives

As Arabic language suffers from lack of interest by researchers. The research objectives are:

- Create the first labeled personality traits prediction dataset for Egyptian dialect twitter users.
- Provide Arabic users with means to predict their personality traits through their twitter feed.
- Help building Artificial Intelligent (AI) system for personality traits detection.