POLITECNICO DI TORINO
Repository ISTITUZIONALE

TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning

Original
TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning / Carducci, G., Rizzo, G., Monti, D., Palumbo, E., Morisio, M.. - In: INFORMATION. - ISSN 2078-2489. - ELETTRONICO. - 9:5(2018), pp. 127-127. [10.3390/info9050127]

Availability:
This version is available at: 11583/2708024 since: 2018-05-22T16:30:32Z

Publisher:
MDPI

Published
DOI:10.3390/info9050127

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

05 November 2026

information

MDPI

Article

TwitPersonality: Computing Personality Traits from Tweets Using Word Embeddings and Supervised Learning

Giulio Carducci [1,2,†], Giuseppe Rizzo [1,*], Diego Monti [2], Enrico Palumbo [1,2,3] and Maurizio Morisio [2]

1 Istituto Superiore Mario Boella (ISMB); Via Pier Carlo Boggio, 61, 10138 Turin, Italy; giulio.carducci@studenti.polito.it (G.C.); palumbo@ismb.it (E.P.)
2 Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy; diego.monti@polito.it (D.M.); maurizio.morisio@polito.it (M.M.)
3 EURECOM, Sophia Antipolis, Campus SophiaTech, 450 Route des Chappes, 06410 Biot, France
* Correspondence: giuseppe.rizzo@ismb.it; Tel.: +39-011-227-6244
† Work done while doing the internship in ISMB.

Received: 15 February 2018; Accepted: 15 May 2018; Published: 18 May 2018

Abstract: We are what we do, like, and say. Numerous research efforts have been pushed towards the automatic assessment of personality dimensions relying on a set of information gathered from social media platforms such as list of friends, interests of musics and movies, endorsements and likes an individual has ever performed. Turning this information into signals and giving them as inputs to supervised learning approaches has resulted in being particularly effective and accurate in computing personality traits and types. Despite the demonstrated accuracy of these approaches, the sheer amount of information needed to put in place such a methodology and access restrictions make them unfeasible to be used in a real usage scenario. In this paper, we propose a supervised learning approach to compute personality traits by only relying on what an individual tweets about publicly. The approach segments tweets in tokens, then it learns word vector representations as embeddings that are then used to feed a supervised learner classifier. We demonstrate the effectiveness of the approach by measuring the mean squared error of the learned model using an international benchmark of Facebook status updates. We also test the transfer learning predictive power of this model with an in-house built benchmark created by twenty four panelists who performed a state-of-the-art psychological survey and we observe a good conversion of the model while analyzing their Twitter posts towards the personality traits extracted from the survey.

Keywords: affective computing; personality dimensions; Big5; word embeddings; machine learning

1. Introduction

Nowadays, social media platforms are the largest mines of personal information, since they continuously record people's habits, interactions, interests in musics, movies and shopping. Such information regarding individuals is so comprehensive that has become essential for many applications targeting the final customer; industries such as (digital) goods retailers and advertising agencies are now researching and implementing strategies to acquire this data to better profile their customers and, therefore, to customize their offers about products and services. Acquiring massively and processing such data to derive people's cues has been popular for numerous years and has triggered numerous research studies aiming to profiling human beings computationally and accurately.

In a popular study [1], authors demonstrate the degree to which relatively basic digital records of human behavior such as Facebook likes can be used to automatically and accurately estimate

a wide range of personal attributes that people would typically assume to be private and profile the individual along five dimensions, namely Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. The acquisition of such information requires an explicit consent by the individuals resulting in an extreme complexity and thus making it unfeasible to be applied in real usage scenarios. This is also due to the prevention of sensitive information leaks that can potentially harm or discriminate individuals.

Social media platforms are also used as channels where individuals deliver thoughts, emotions, and essays. Usually, they are delivered in a public fashion meant to reach a very large audience. Twitter, for instance, is among the most popular channels used by people and tweets are tools that ultimately characterize the individuals who have written them or shared tweets of others. We formulated the following research question to leading our investigation: to what extent tweets can be processed in order to derive meaningful traits of an individual? Profiling a person by using words he or she uses has been extensively studied in language psychology. In fact, a central assumption in language psychology is that the words people use reflect who they are. For instance, in [2,3] authors investigated and observed that the way people speak is related to their physical and mental health problems. With the increasing computer advances, efforts have been made to capture psychological themes or people's underlying emotional states that might be reflected in the words they used (e.g., [4]). Other psychological and linguistic studies have supported the intuition that the personality traits of individuals are implicitly encoded in the words used to shape a sentence [5,6].

In this paper, we propose a supervised learning approach to compute five personality traits by only relying on what an individual tweets about his or her thoughts. The approach segments tweets in tokens, then it learns word vector representations as embeddings that are then used to feed a supervised learner classifier. We demonstrate the effectiveness of the approach by measuring the mean squared error of the learned model using an international benchmark of Facebook status updates. We also tested the transfer learning capability and feasibility of this model with an in-house built benchmark created by twenty four panelists who performed a state-of-the-art psychological survey and we observed a good conversion of the model while analyzing their Twitter posts towards the personality traits extracted from the survey.

The remainder of this paper is structured as follows: in Section 2, we compare our approach with state-of-the-art learning algorithms and in Section 3 we introduce the so-called Five Factor Model. In Section 4, we present both the gold standard used to train the learning model presented in this paper and the Twitter gold standard used as benchmark. In Section 5, we detail the approach and in Section 6 we report the experimental results we have achieved when extracting personality traits from tweets. We discuss the threats to validity of the approach in Section 7 and we conclude with insights and future work in Section 8.

2. Related Work

Numerous studies have demonstrated the correlation between personality traits and many types of behavior. This includes job performance [7], psychological disorders [8,9], and even romantic success [10]. Significant correlations were also found between personality and preferences. Studies showing connections between personality and music taste are well established in literature [11–14], while Jost et al. [15] found that knowing the personality traits of an individual allowed to predict whether they would be more likely to vote for McCain or Obama in the presidential election of 2008 in the United States. Cantador et al. [16] presented a preliminary study in which they analyzed preferences of roughly 50,000 Facebook users who expressed their interests about sixteen genres in each of these domains: movies, TV shows, music and books. Such results can be very valuable to enhance personalization services in several domains.

The volume of research on computational personality recognition has grown steadily over the last years [17,18]. The ubiquity of social media platforms has inspired researchers to move towards these platforms to seek useful information to be used for personality prediction. Many studies showed

correlation between personality and online behavior [19–21], and there is an excellent corpus in literature about inferring personality traits from those platforms. Quercia et al. [22] were the first who explored at large the relationship between personality and use of Twitter; they also proposed a model to infer users' personality based on just following, followers, and listed count numbers. Similarly, Jusupova et al. [23] used demographic and social activity information to predict personality of Portuguese users, whereas Liu et al. [24] proposed a deep-learning-based approach to build hierarchical word and sentence representations that is able to infer personality of users from three languages: English, Italian, and Spanish. Van de Ven et al. [25] based their analyses on LinkedIn, a job-related social media platform, yet they did not find strong correlations between personality traits and user profiles, except for Extraversion. YouYou et al. [26] demonstrated that computer-based judgments about an individual can be more accurate than those made by friends, spouse, and even the individual himself, if sufficient data is available.

The first approaches on personality prediction were mainly based on SVM, using syntactic and lexical features [6,27]. It was just in recent years that researchers moved towards deep learning. Kalghatgi et al. [28] and Su et al. [29] used neural networks by feeding them a number of meticulously hand-crafted features, the first about syntax and social behavior while the latter regarding grammar and LIWC annotations extracted from a dialogue. Majumder et al. [30] use a CNN to derive a fixed-length feature vector starting from word2vec word embeddings [31], which they extend with eighty four additional features from Mairesse's library [32]. For classification, the so computed document vectors are fed both to a multi-layer perceptron (MLP) and to a polynomial SVM classifier. In contrast to this work, we adopt a dictionary of existing word representations generated by the skip gram model implemented in FastText. We then combine word representations using geometric manipulations of the vectors and, ultimately, we feed them into a radial SVM classifier. In Section 5, we explore some of the available word embeddings models in literature and explain, in detail, how tweets are transformed into corresponding distributed representations in the vector space using those models.

Distributed representations of words are capable of successfully capturing meaningful syntactic and semantic properties of the language and it has been shown [33] that using word embeddings as features could improve many NLP tasks, such as information retrieval [34,35], part-of-speech tagging [36] or named entity recognition (NER) [37]; Kuang and Davidson [38] learned specific word embeddings from Twitter for classifying healthcare-related tweets. Since learning those word representations is a slow and non-trivial task, already trained models can be found in literature; state-of-the-art embeddings are mainly based on deep-learning [31,39], but other techniques have been previously explored, for instance spectral methods [40,41]. If correctly trained, word embeddings could also improve performances on sentiment analysis. In fact, existing models typically ignore the sentiment information; words like *good* and *bad* are mapped to close vectors, due to their similar usages and grammatical roles. Tang et al. [42] overcome this issue by learning sentiment specific word embeddings. They extend the model from Collobert et al. [43] by incorporating sentiment information in the neural network.

There is a wide number of studies that do not use lexical features but are instead based on social media platform metadata. Kosinski et al. [1] apply singular value decomposition (SVD) on a user-like matrix containing associations between users and Facebook pages, then predict a wide range of personal attributes implementing a regression model with the top 100 SVD components; Golbeck et al. [44] collected everything available from users' Facebook profile and activities, which they extended with a number of composite features; Quercia et al. [22] based their analyses only on three numbers publicly available on Twitter: following, followers and listed count. Based on these features, users are classified as popular, listeners, highly-read, and two types of influencers, each of these groups having different personality characteristics. In Section 6 we compare such results with those obtained by our analysis using the mean squared error as same evaluation metric.

Social media platforms are rich sources of information not only for personality prediction. Sentiment analysis has attracted increasing interest in recent years [45–48]. It consists of classifying the

sentiment polarity of a sentence, as positive, negative or neutral. Dai et al. [49] explored numerous feature engineering techniques for detecting adverse drug reactions from Twitter posts. Studies on demographics are also worth of note, in particular for those sites (e.g., Twitter) where demographic information is mostly unavailable. Chamberlain et al. [50] and Zhang et al. [51] successfully inferred users' age based on their interaction on the social media platform. Other studies on enhancing social data with demographic attributes include gender [52], location [53], ethnicity [54], and political affiliation [55].

3. Five Factor Model

The Five Factor Model (FFM), also known as the Big Five Model (Big5), is the most widely accepted model to describe personality traits. This is because it integrates a wide array of personality constructs in an efficient and comprehensive manner [56]. For decades, several independent groups of researchers put their effort on deriving a general and comprehensive personality construct, and they all achieved similar results, suggesting that, in fact, there could be a general model describing human personality. The credit for advancing the first model is generally given to Tupes and Christal [57]. Later, Digman [58] and Goldberg [59] proposed further advancements and refinements.

The five traits defined by the model are:

- *Openness to Experience* (inventive/curious vs. consistent/cautious): it measures a person's curiosity, intelligence, appreciation for art, seeking for new experiences and adventures. People low in Openness tend to be more conservative and close-minded.
- *Conscientiousness* (efficient/organized vs. easy-going/careless): it measures the capability of a person to be organized, reliable and consistent. Those who have a high score on Conscientiousness prefer planning rather than having a spontaneous behavior, and they opt for seeking long-term goals, while low Conscientiousness individuals tend to be more tolerant and less bound by rules and plans.
- *Extraversion* (outgoing/energetic vs. solitary/reserved): it measures the tendency to seek stimulation in the external world, the company of others and being more talkative. Those low in extraversion are reserved and solitary.
- *Agreeableness* (friendly/compassionate vs. challenging/detached): it measures the tendency of being altruistic and tender-minded. Those who have a high score in Agreeableness are trusting, altruistic, tender-minded, and are motivated to maintain positive relationships with others. Low Agreeableness is related to being suspicious, challenging and antagonistic towards other people.
- *Neuroticism* (sensitive/nervous vs. secure/confident): it measures the tendency to be impulsive and emotionally unstable, experiencing mood swings and negative emotions. Those who have a low score in Neuroticism are more calm and stable.

Some may argue that the FFM is too general and the five traits are not nearly as powerful as predicting and explaining actual behavior because of the numerous facets on primary traits [60,61]. According to McCrae et al. [62], the measurement of the five factors gives a complete characterization of a person only at a global level.

3.1. Assessing Personality with Surveys

Personality traits are generally determined using a questionnaire, which presents a variable number of questions that describe common situations and behaviors, and the individual should indicate the extent to which each statement describes himself using a Likert scale. Questions are positively or negatively related to a specific trait; based on the answer, a specific value is assigned to each of them. Finally, the trait score is computed by aggregating all the values of its related answers. It is generally indicated with a continuous value in the range of $[1, 5]$.

An example of personality questionnaire is *The Revised NEO Personality Inventory* from Costa and McCrae (NEO PI-R, [63]). It consists of 240 items on a five-level Likert scale, and can accurately measure the five personality traits plus six underlying facets for each of them. The International Personality Item Poll ([64,65]) features a public collection of items and scales that can be used for personality tests. It also includes two questionnaires, the 300-items IPIP NEO [65] and a shorter version with 120 items [66]. Although they are both based on the NEO PI-R, they are not equivalent.

3.2. Social Media Platforms as Sources to Deriving Personality Traits

There is a large body of research on predicting personality from language features. This is motivated by the *lexical hypothesis*, which states that personality characteristics of individuals will come to be encoded in terms of words choice. Many studies on the topic have delivered significant results [5,6,32,67,68].

Social media platforms are a rich source of textual data, and users' behavior on these platforms tends to reflect many aspects of real life, including personality [19–21]. People widely share their feelings, moods, and opinions providing a rich and informative collection of personal data that could be used for a variety of purposes. This is often referred to as subconscious crowdsourcing, as implied by the term subconscious, which indicates that people are unaware and are subconsciously sharing information [69] to their network. Thus, the enormous growth of those social media platforms in the last decade makes those users perfect candidates for personality prediction. However, as pointed out by Chin and Wright [70], each platform has its own characteristics that will likely affect the effectiveness of predictions. This includes the number and length of entries, whether the author is identified or not, grammar and word choice, and others. Therefore, differences and biases of the various platforms should be taken into account when analyzing textual data from those sources as observed in our experimental setting (see Section 5.4).

Tweets, for example, are usually very short and present a peculiar language usage, full of abbreviations and uncommon forms with respect to other textual sources, such as blog posts. This makes the task of personality profiling from text more challenging. Previous researches have shown that Extraversion and Neuroticism were found to correlate with number of friends in the real world as well as on Facebook [44]. Individuals high in Extraversion and low on Neuroticism tend to maintain persistent connections with their friends [71,72]. Extroverted people also find those platforms easy to use [73]. Generally, users tend to be friend of people with higher Agreeableness and select contacts with similar Openness, Extraversion and Agreeableness [74]. This is an example of the homophily principle, which states that people with related attributes, such as age, interests, and personality, form similar ties [75]. Analyses of [76–80] are just additional examples of the excellent body of work carried out by researchers in order to better understand the existing correlation between personality and social media platforms.

4. Gold Standards

In order to quantitatively test our approach, we rely on the dataset created by the myPersonality project [1]. The dataset has been collected for research purposes by David Stillwell and Michael Kosinski using a Facebook application that administered a personality test and collected a wide range of personal and activity information from Facebook's profile of users mainly from US and UK under their consent. Participants answered a variable length (20–100 or 300 items) proxy test derived from Costa and McCrae NEO-PI-R [63] and elaborated by the International Personality Item Pool (IPIP) project [81]. In practice, users were asked to answer to a set of questions that were positively or negatively correlated with certain personality traits, such as "Do you have frequent mood swings?", using a $[1, 5]$ Likert scale and scores were then computed averaging the results corresponding to a specific personality trait. The myPersonality application has been active from 2007 to 2012 and collected a huge amount of data, which is available upon request.

We base our study on a sample of the original myPersonality dataset that has been made publicly available [82]. The gold standard contains 9913 English status updates of 250 users (anonymized), annotated with personality traits scores and basic statistics describing the user social network, such as number of friends and betweenness, that is, a measure of the centrality of a node within a graph [83]. In this work, we use as input for our predictive models the raw text coming from status updates. Tables 1 and 2, and Figure 1 illustrate some statistics of the dataset. It contains approximately 15,000 distinct terms and 300 stopwords. A natural language word is generally considered a stopword when it has very little meaning for the sentence. Examples of stopwords are articles and conjunctions. We used the following list of stopwords as reference [84]. However, stopwords account for almost half of the total, lowering the predictive power of the dataset.

Table 1. Statistics of the myPersonality gold standard in terms of number of users, status updates and words. We highlight the effect of pre-processing on those dimensions.

Metrics	Count	Lowest	Average	Highest
Total users	250	-	-	-
Total status updates	9913	-	-	-
Status updates per user	-	1	39	223
Total words	146,128	-	-	-
Total words after preprocessing	72,896	-	-	-
Unique words	15,470	-	-	-
Unique words after preprocessing	15,185	-	-	-
Word per status update	-	1	14	113
Word per status update after preprocessing	-	0	7	57

Table 2. Statistics about the personality traits of the myPersonality gold standard. We report standar deviation, highest, lowest and average value across all 250 users. Where OPE = Openness, CON = Conscientiousness, EXT = Extraversion, AGR = Agreeableness, NEU = Neuroticism.

	OPE	CON	EXT	AGR	NEU
max	5	5	5	5	4.75
min	2.25	1.45	1.33	1.65	1.25
avg	4.0786	3.5229	3.2921	3.6003	2.6272
std	0.5751	0.7402	0.8614	0.6708	0.7768

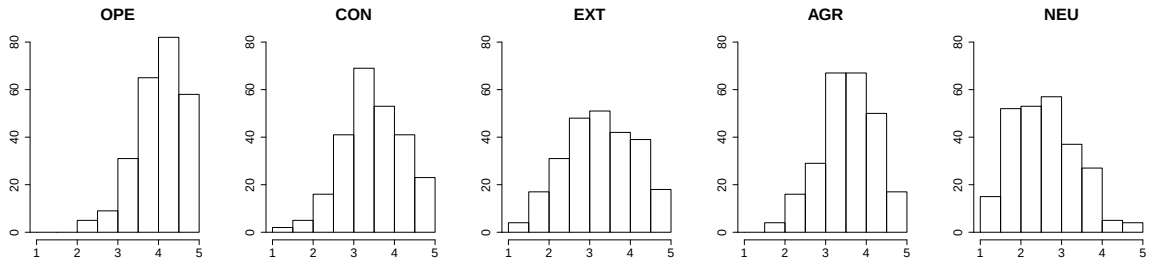


Figure 1. Five histograms representing the distributions of the Big5 personality traits of the users available in the myPersonality gold standard.

We then created an in-house gold standard of Twitter users, English posts, and personality trait scores, to empirically validate our transfer learning approach. In the remainder of this paper, we refer to this gold standard as Twitter sample. We asked twenty four panelists to provide their Twitter handle and to perform the Big Five Inventory (BFI) [85] personality test.

The BFI test consists of forty four short phrases based on the trait adjectives that represent the prototype of each Big5 trait. For example, the adjective *original* is associated with the trait of Openness and its corresponding BFI item is the phrase "is original, comes up with new ideas". Each Big5 trait is described by 8–10 phrases; for each phrase, the subject is asked to express, on a Likert scale from 1 to 5, if he or she agrees or not with that description of his/her personality. The outcome is then computed by averaging the normalized scores associated to each trait. Twenty six panelists took part to the creation of the Twitter sample. Finally, two were discarded since they have an insufficient number of tweets. Tables 3 and 4 report some statistics about the Twitter sample and Figure 2 shows the histograms of the personality scores obtained from the survey of the five personality traits.

Table 3. Basic statistics about our Twitter user sample. All the values reported are calculated after having applied all the preprocessing steps.

Metrics	Count	Lowest	Average	Highest
Total users	24	-	-	-
Total tweets	18,473	-	-	-
Tweets per user	-	9	769.7	2252
Total words	263,984	-	-	-
Total words after preprocessing	77,900	-	-	-
Unique words	173,867	-	-	-
Unique words after preprocessing	34,651	-	-	-
Word per status update	-	1	13.36	33
Word per status update after preprocessing	-	0	8.8	25
Avg words per tweet per user	-	5	6.8	8.8
Number of followers	-	12	1375.5	20,800

Table 4. Statistics about the personality traits of the Twitter user sample. We report standard deviation, highest, lowest and average value across all twenty four users. Where OPE = Openness, CON = Conscientiousness, EXT = Extraversion, AGR = Agreeableness, NEU = Neuroticism.

	OPE	CON	EXT	AGR	NEU
max	4.8	4.78	4.38	4.33	3.63
min	2.5	2.33	1.75	2.78	1.5
avg	3.8917	3.5513	3.2208	3.6438	2.6642
std	0.5763	0.5682	0.5449	0.3707	0.5250

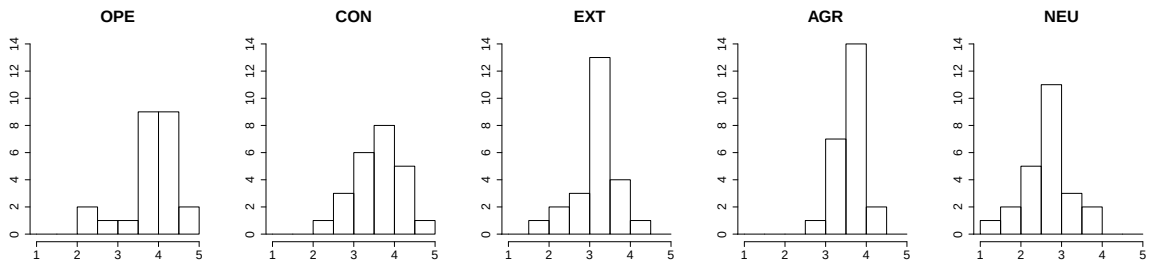


Figure 2. Five histograms representing the distributions of the Big5 personality traits of the Twitter sample.

5. Approach

Using the myPersonality gold standard described in Section 4, we derive a predictive model that is able to infer personality traits of Twitter users. Since the dataset we intend to use consists of

status updates extracted from Facebook, our approach is based on transfer learning, namely modeling an algorithm for a context and applying it on a (slightly) different, yet related, context. In our case, we derive a machine learning model for predicting personality of Facebook users based on their status updates, then we test it on Twitter users' tweets. In the next subsections, we show how data is processed in order to derive a predictive model, and how the model is tested using data from Twitter.

5.1. Text Processing

We convert textual data into numeric signals. To do this, we treat each Facebook post individually to obtain a corresponding vector space representation, inspired by the approach of Li et al. [86]. We first apply some basic pre-processing steps to clean the documents from noise and less-informative data. We choose not to perform stemming, because we may lose the semantic meaning of a word. Without preprocessing, data used both for training and testing the predictive models would be messy and contaminated with irrelevant pieces of information, which may harm their accuracy.

In detail, the processing stages we apply sequentially are:

Conversion to lowercase: "Today is a #sunny day!" → "today is a #sunny day!".
Stop-word removal: "today is a #sunny day!" → "today #sunny day!".
Punctuation removal: "today #sunny day!" → "today sunny day".
Tokenization: "today sunny day" → [today] [sunny] [day].
Short post removal: it removes status updates with $n_{tokens} < 3$.

For stop-words removal, we use the list of stop words provided by the module *CountVectorizer* from scikit-learn [84]. We remove status updates that result in less than three tokens, which, in total, are 15.78% of the initial set.

We do not learn word embeddings but instead we test some pre-trained models available in literature. In Section 6, we describe those models and report their statistics and performances with the dataset. All those are based on feature word vectors of dimension 300. Landauer and Dumais [87] showed, with empirical experiments, that 300 is the optimal number of features for distributed word representations, although they based their approach on LSA and word co-occurrence rather than neural networks.

We look up each token in the pre-trained word embeddings dictionary to obtain its word vector representation. If it does not exist, we simply ignore it (in Section 6 we report word coverage statistics). At this point we combine all the dense real-value vectors into a single one. Li et al. [86] explore different methods for representing a tweet from word embeddings, and test these approaches on two tasks: sentiment analysis and topic classification. We choose to represent status updates with concatenation of maximum, minimum and average, as in the approach of [86]. To compute the maximum, we consider for each one of the 300 dimensions the maximum value among all word vectors for that component. Similarly, we derive also minimum and average. Therefore the vector resulting from the concatenation has a dimension of 900. Figure 3 illustrates the process of deriving a vector space representation from a list of tokens. The choice of concatenation is verified empirically (Section 6); in fact, this method appears to be more stable than other approaches, having the least variance in terms of mean squared error.

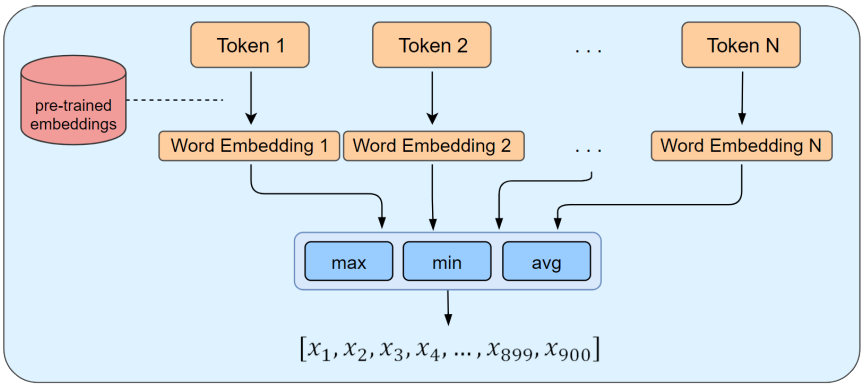


Figure 3. Mapping of a post to the corresponding representation in the vector space. For each token we obtain the corresponding word embedding, then we concatenate max, min and avg computed on all word vectors.

5.2. Learning Model

We train five different models, one for each personality trait built using the gold standard, and as learning model a support vector machine (SVM) [88]. The choice of this learning algorithm is motivated by the experimental results detailed in Section 6. We use the status updates representations as the sole learning features of the model.

An SVM linear model constructs one or more separators (hyperplanes) as a decision surface so that the distance of datapoints of different classes is maximized. Classes are defined as the list of the expected categorical values produced by the classification task. When this approach is applied to a regression problem, it is called Support Vector Regression (SVR). The goal of SVR is to compute y_i from the observation x_i with a margin of tolerance ϵ. The objective function is the following:

$$f(\mathbf{x}) = \beta\mathbf{x} + b, \qquad \min J(\beta) \tag{1}$$

$$J(\beta) = \frac{1}{2}\|\beta^2\| + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \tag{2}$$

where ξ_i and ξ_i^* indicate the distance of datapoints that lie outside the margin ϵ, in both sides of the hyperplane, respectively. C is a hyperparameter regulating the penalty for errors. We use the following convention, that is valid throughout the paper: bold letters stand for vectors (lowercase) and matrices (uppercase), while scalars and constants are indicated with light letters. The optimization problem is computationally simpler to solve in its Lagrangian dual formulation. The dual formula is derived from the primal by introducing non-negative multipliers α_n and α_n^* for each observation x_n, and it is expressed as:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_i - a_i^*)(a_j - a_j^*)\langle x_i \cdot x_j\rangle + \epsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i) \tag{3}$$

The parameter β can be described as a linear combination of the training observations with the following equation:

$$\beta = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*)(x_n \cdot x) + b \tag{4}$$

Some regression problems cannot be accurately described by a linear model. The dual formulation allows us to extend the approach to non-linear functions by substituting the dot-product $\langle x_i \cdot x_j\rangle$ in the formula with a non-linear kernel function $G(x_i, x_j) = \sum_n \phi_n(x_i)\phi_n(x_j)$, which is a transformation function that maps $\mathbf{x}$ to a high-dimensional space [89]. The classifier is again a hyperplane in the

transformed space, though it may not be linear in the original space. Table 5 reports some common kernel functions. In our approach we experimented and validated empirically the use of the rbf kernel (Section 6).

Table 5. Common SVM kernels. γ, α and c are parameters of the functions that need to be tuned for the application. γ is often set to $1/2\sigma^2$.

Kernel Name	Kernel Function
Linear	$\mathbf{G}(x_j, x_k) = x_j \cdot x_k$
Polynomial	$\mathbf{G}(x_j, x_k) = (x_j \cdot x_k + 1)^n, \quad n = \{2, 3, \dots\}$
Radial basis function (rbf)	$\mathbf{G}(x_j, x_k) = \exp(-\gamma \|x_j - x_k\|^2), \quad \gamma > 0$
Hyperbolic tangent (sigmoid)	$\mathbf{G}(x_j, x_k) = \tanh(\alpha x_j \cdot x_k + c)$

5.3. Model Optimization

We defined the hyperparameters of the SVM by minimizing the loss function of the entire problem when studying the performance of the approach with the gold standard. Those parameters are:

Kernel: linear, polynomial, radial basis function, hyperbolic tangent.
C: the penalty factor that regulates the trade-off between misclassification and simplicity of the decision surface.
Gamma: indicates how far the influence of a single training example reaches, with low values meaning far and high values meaning close, only meaningful for the rbf kernel.
Degree: of the polynomial equation, only meaningful for a polynomial kernel.

The loss function we use is the mean squared error (MSE), which is the mean squared difference between predicted and actual values:

$$MSE(P, A) = \frac{1}{n} \sum_{i=0}^{n-1} (p_i - a_i)^2 \tag{5}$$

where with $P = (p_1, p_2, \dots, p_n)$ are the predicted values and $A = (a_1, a_2, \dots, a_n)$ are actual values of the training samples that we are evaluating the model on. MSE is a positive score in the range $[0, (max(p_i) - min(a_i))^2]$ where max is a function that selects the maximum value of the list and min a function that selects the minimum value of the list. Smaller values of MSE indicate better accuracy.

Since we do not have a test set, we estimate the predictive power of our models using a 10-fold cross-validation strategy over the gold standard. We repeat the process ten times with different cross-validation splits to evaluate the model on the entire dataset and for different configurations of the hyper-parameter values for a total of twenty values chosen randomly. The number of values has been defined empirically when minimizing the MSE while keeping the task feasible at computational level. We then select the hyper-parameters that minimize the overall MSE of the approach.

5.4. Predicting Personality Traits from Tweets

Our approach is thus based on transfer learning: we train the models on Facebook status updates and we apply them on predicting the personality traits using tweets as inputs. To test our model we collect a test set of Twitter users who gave their consensus on analyzing their profiles and at the same time answered a psychological questionnaire to measure their personality trait scores.

Given a Twitter handle (e.g., @username), we crawl the Twitter API to retrieve all available tweets of that user. Since tweets are slightly different from Facebook status updates, we define additional preprocessing steps that are executed after those listed in Section 5.1. In detail, they are:

Pure retweet removal: When retweeting a tweet, if no personal text content is added, it will be represented as *RT@tweet-owner: tweet-text tweet-url*. We refer to this case as pure retweet. Since we base

the prediction only on the textual features created by a user, all retweets without additional content are removed.
URL removal: URLs may appear in retweets, sharing from other sources (e.g., YouTube video), or other cases. In the context of our study they are not informative, so they are removed from the tweets.
Mentions removal: Users often mention their friends in their tweets, adding some noise to the text data; mentions are also found in retweets. We remove all mentions from our data.
Hashtags removal: We choose not to remove the whole word, as in the previous cases. Instead, we keep the original word without the hash (#) symbol. This is because some tags consist of meaningful words (e.g., "I love #star #wars") which we do not want to lose. This approach effectively addresses also the case of non-existing hashtag words (e.g., "#netNeutarlity2017") that are ignored in the next step, when a corresponding word embedding will not be found.

All the processing steps detailed above are carried out using regular expressions, with the exception of hashtags removal. In fact, hash symbols are automatically deleted by punctuation removal. Apart from those further refinements, we process tweets the same way as status updates in order to derive a feature vector of 900 dimensions, as in Figure 4. We separately feed each tweet vector to the trained model to obtain a prediction, and we average all the values to compute the final personality trait score.

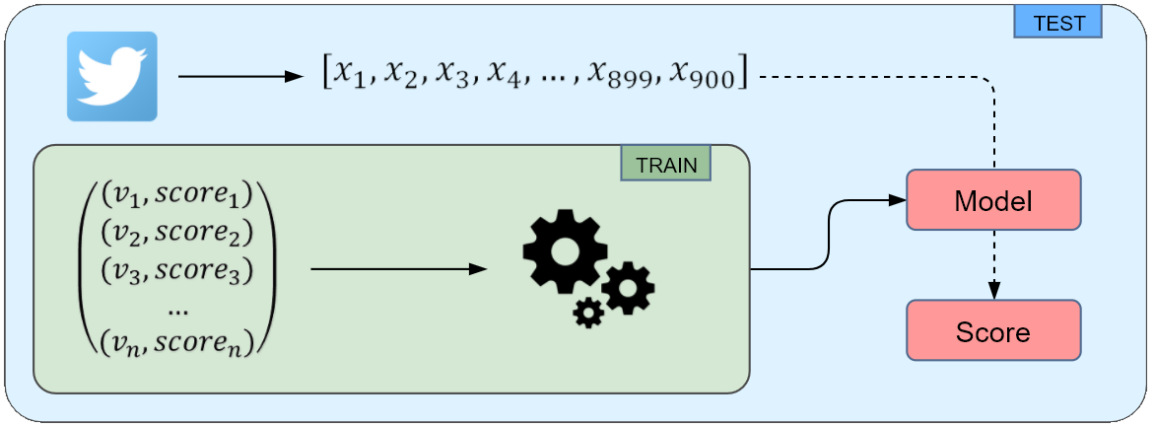


Figure 4. Each pair of transformed status updates and corresponding personality trait score is a training sample that we feed to the algorithm. We train five different models, one for each personality trait, and test them with tweets. A tweet is transformed into a vector the same way as status updates.

To assess the accuracy of the models we compare the predicted personality scores with an in-house built gold standard. We first compute the mean squared error for the single traits, then the mean of the five MSEs for the individual. This way we can inspect which traits are easier or harder to predict while at the same time having a general estimate of the model performance.

6. Experimental Results

We first evaluate our approach on the myPersonality dataset comparing it with two baseline algorithms; we then verify the transfer learning capability of the model with an in-house built gold standard created from twenty four panelists.

6.1. Model Verification

To derive the best performing predictive model we explore different machine learning algorithms and configurations and we evaluate them on the training set minimizing the mean squared error used as loss function. We also compare the learning model used in our approach, namely SVM, with two baseline algorithms, namely Linear Regression [90] and LASSO [91], both used in state-of-the-art approaches for personality prediction [1], and both trained with the same configuration setup of our

approach. We now report the inner-workings of Linear Regression and LASSO in order to understand how the optimization has been derived.

Linear regression is a model used to predict continuous values using a linear combination of a set of features, that in our case are the word vectors. Variables $\mathbf{X}$ and $\mathbf{y}$ are related by an unobserved parameter β_t (where t stands for *true*); the goal is fitting a predictive model to the observed data $\mathbf{X}$ and $\mathbf{y}$ by estimating a set of parameters β. The parameter estimation is done with the ordinary least squares method, which minimizes the sum of squared residuals of the predictions, defined as follows:

$$\mathbf{y} = \beta\mathbf{X} + \epsilon, \qquad y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{900} x_{i900} + \epsilon_i, \qquad i = 1, 2, \ldots, N \tag{6}$$

$$\beta = (\mathbf{X} \cdot \mathbf{X})^{-1}\mathbf{X} \cdot \mathbf{y} = \left(\sum \mathbf{x_i} \cdot \mathbf{x_i}\right)^{-1} \left(\sum \mathbf{x_i} y_i\right) \tag{7}$$

since β is merely an approximation of β_t, the prediction of $\mathbf{y} = \{y_1, y_2, \ldots y_N\}$ includes a certain error, which is denoted with ϵ.

Least absolute shrinkage and selection operator (LASSO) is also a regression analysis technique that performs both features regularization and subset selection over the available features, improving the accuracy of the prediction while providing more interpretable models. The optimization objective for LASSO is:

$$\min_{\beta} \frac{1}{2n_{samples}} \|\mathbf{X}\beta - \mathbf{y}\|_2^2 + \alpha\|\beta\|_1 \tag{8}$$

where α is again a hyperparameter of the model that controls the degree of sparsity of the estimated coefficients. With $\alpha = 0$, we have again an OLS linear regression model.

For both SVM and LASSO we perform a tuning phase to select the best configuration of the hyperparameters for our application. We explore a subset of all possible combinations, for a total of thirty different models for each personality trait. The parameters that we considered, along with their different values, are reported in Table 6. We observe that SVM performs better than other algorithms. For each personality trait there exists an SVM configuration with a minimum MSE lower than that of other learning models. Table 7 reports those configurations with the relative error value, that is computed using 10-fold cross-validation on the training set. All the best performing models use a rbf kernel and similar values for *C* and *gamma*. Openness presents the lowest error while Extraversion has the highest. Optimal configuration for linear and LASSO regression are reported in Table 8. We can also observe that LASSO performs slightly worst than SVM, and all the best performing models have the same value for α, while linear regression, as expected, has the highest MSE values overall, and thus performs worse. Table 9 sums up the mean squared error of the different models and highlights the margin of improvement of SVM over other algorithms.

Table 6. Hyperparameters and their values of SVM and LASSO models. For SVM, rather than using an exhaustive grid search approach, we only test a subset of all the possible combinations, for a total of 20.

Algorithm	Parameter	Value
SVM	Kernel	*linear*, *rbf*, *poly*
	C	1, 10, 100
	Gamma	0.01, 0.1, 1, 10
	Degree	2, 3
LASSO	Alpha	$10^{-15}, 10^{-10}, 10^{-8}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 1, 5, 10$

Table 7. Best-performing SVM configurations of kernel and hyperparameter values used in our approach segmented per trait and measured by mean squared error.

Trait	Kernel	C	Gamma	MSE
Openness	rbf	1	1	0.3316
Conscientiousness	rbf	10	1	0.5300
Extraversion	rbf	10	1	0.7084
Agreeableness	rbf	10	1	0.4477
Neuroticism	rbf	10	10	0.5572

Table 8. Mean squared error values for Linear Regression (LReg) and LASSO regression for the five factors. In LASSO, the best-performing configuration always results for $\alpha = 0.0001$.

Trait	Model	Configuration	MSE
Openness	LReg	-	0.3915
	LASSO	alpha = 0.0001	0.3345
Conscientiousness	LReg	-	0.6200
	LASSO	alpha = 0.0001	0.5363
Extraversion	LReg	-	0.8210
	LASSO	alpha = 0.0001	0.7106
Agreeableness	LReg	-	0.5407
	LASSO	alpha = 0.0001	0.4500
Neuroticism	LReg	-	0.6625
	LASSO	alpha = 0.0001	0.5595

Table 9. Summary of mean squared error values for all models. LASSO performs almost as well as SVM, while linear regression (LReg) has a more consistent margin of error. Those differences are somewhat consistent across all personality traits. Both for Linear Regression and LASSO we also report the increment in terms of mean squared error with respect to SVM.

Model	OPE	CON	EXT	AGR	NEU
SVM	0.3316	0.5300	0.7084	0.4477	0.5572
LASSO	0.3345 (0.0029)	0.5363 (0.0063)	0.7106 (0.0022)	0.4500 (0.0023)	0.5595 (0.0023)
LReg	0.3915 (0.0599)	0.6200 (0.0900)	0.8210 (0.1126)	0.5407 (0.0930)	0.6625 (0.1053)

As we mentioned in Section 5, we choose to use concatenation for representing tweets and status updates in the vector space. We motivate this choice with the experimental results presented in Table 10, which shows an average value and a standard deviation of MSE across all nineteen tested SVM configurations and five personality traits. In detail, the configurations are: linear kernel with C = 1; poly kernel with all combinations without repetition of degree = [2,3] and C = [1,10,100]; rbf kernel with all combinations without repetition of degree = [0.01, 0.1, 1, 10] and C = [1, 10, 100]. We note that although other methods sometimes achieve a lower MSE, standard deviation for concatenation is the lowest for each personality trait, suggesting that this method is more stable and consistent, hence more reliable.

We test and evaluate our models using pre-trained word vectors. We explore different embeddings and how they affect the predictive power of the models in terms of mean squared error. FastText is an open source library from Facebook AI research [92]. There are numerous pre-trained word embeddings in many languages, though we are only interested in English for this experimentation. We use a model with one million word vectors trained on Wikipedia 2017, UMBC webbase corpus and statmt.org news dataset. Li et al. [86] explored different sources and text processing combinations to train word embeddings. They learned word vectors both from tweets and from general text data available on the Web, including "spam" tweets (less informative ones), and considered both words and words

plus phrases for the models. The combination of those approaches led to ten different datasets. Although learning embeddings from tweets should improve the overall performance of the model, this highly depends on the context of the application and on the data it is tested on, and we observe that in our case the model from FastText performs better. Table 11 shows some statistics of the datasets we explored. We notice that despite being trained on 2.8 times more words on average, those datasets are able to achieve only up to 3.14% more word coverage than FastText.

Table 10. Average MSE value and standard deviation of the nineteen SVM models across all the different configurations that we take into account for tuning the hyperparameters. Mean values, except for sum, are almost equivalent, though standard deviation for concatenation is the lowest for all the five factors.

Method	OPE		CON		EXT		AGR		NEU	
	Mean	std	Mean	std	Mean	std	Mean	std	Mean	std
sum	0.459	0.308	0.723	0.542	0.939	0.655	0.6	0.429	0.75	0.497
maximum	0.352	0.014	0.547	0.023	0.735	0.031	0.461	0.018	0.58	0.028
minimum	0.352	0.016	0.546	0.024	0.732	0.032	0.462	0.017	0.579	0.026
average	0.355	0.015	0.547	0.025	0.735	0.03	0.464	0.019	0.582	0.034
concatenation	0.352	0.01	0.548	0.016	0.736	0.026	0.463	0.018	0.583	0.024

Table 11. Statistics of the pre-trained word embeddings models we use in our experiments. Although some of the other datasets achieve lower MSE for one or more personality traits, the mean value across all traits is still higher than the one relative to FastText. Mean squared error is computed on the models trained with configuration reported in Table 7. Datasets 1–10 differ in terms of text source (general data vs. tweets), tweets filtering (spam tweets included or not) and text processing (words vs. words and phrases).

Dataset	Word Coverage	# of Word Vectors	avg OCEAN MSE
FastText	95.08%	1M	0.517
[86] Dataset 1	96.02%	1.9M	0.532
[86] Dataset 2	95.31%	2.9M	0.551
[86] Dataset 3	96.36%	2.7M	0.558
[86] Dataset 4	95.69%	4M	0.541
[86] Dataset 5	96.49%	1.4M	0.539
[86] Dataset 6	95.91%	3.1M	0.518
[86] Dataset 7	98.05%	1.7M	0.52
[86] Dataset 8	97.45%	3.7M	0.541
[86] Dataset 9	98.22%	2.2M	0.533
[86] Dataset 10	97.65%	4.4M	0.537

6.2. Transfer Learning Assessment

We evaluate the predictive power of our approach using the best performing models assessed with the myPersonality dataset over the gold standard of Twitter users. We download and process the tweets as described in Section 5.4, and consider each of them as a single test sample that is then fed to the five SVM models. We compute the final personality trait score as the mean of the predicted values for all the tweets of a given user, and compare the results with the actual values derived from the questionnaire and reported in the Twitter sample. In Table 12, we compare the mean squared error values of the actual (expressed by the panelists though the survey) and predicted values (computed by our approach) with the myPersonality dataset as the mean value over the 10 cross-validation iterations,

reported in Table 7. Our approach achieves significantly better results when applied to these newly created in-house gold standard for Conscientiousness, Extraversion, Agreeableness, Neuroticism; it only performs lower for Openness with a delta difference that is rather small. Anyway, our approach tested with the myPersonality and Twitter gold standards outperforms the one of Quercia et al. [22], which we use as a reference for assessing the performance of our model. Nevertheless, the modest size of the Twitter sample cannot account for enough variation in the personality traits, whose distribution is rather biased, and this may affect the interpretation of the predictions; this will be further investigated in future work.

Table 12. Comparison of mean squared error for myPersonality dataset and for the Twitter sample. In both cases the models are trained with the configurations reported in Table 7. We also report the results obtained by Quercia et al. [22] in their study. It should be noted that in the original paper the evaluation metric is root mean squared error (RMSE), which we converted into mean squared error for the means of comparison.

	OPE	CON	EXT	AGR	NEU
myPersonality	0.3316	0.5300	0.7084	0.4477	0.5572
Twitter sample	0.3812	0.3129	0.3002	0.1319	0.2673
Quercia et al. [22]	0.4761	0.5776	0.7744	0.6241	0.7225

7. Discussion

Results reported in Table 12 show that our approach is able to predict the personality of Twitter users with a certain degree of accuracy, and a minimum mean squared error that is lower than the one of [22]. However, the predictive models that we trained with the techniques introduced in the previous sections may lack of discriminative power, and tend to predict personality values that are close to the average score in the myPersonality Gold Standard. This is probably due to the fact that personality scores in the training set are not equally distributed along their range, as it is possible to see from the histograms of Figure 1. Similarly, the Twitter sample that we collected for this experiment is composed by people belonging to the same social and working environment, hence having very similar personality characteristics. This, together with its modest size, makes of our Twitter sample not well representative of the social network, and the mean squared error is likely to increase when our approach is extended to a greater and more diverse sample of users.

Personality traits are usually a representation of ourselves. (Semi-)public distribution channels of posts, such as social media platforms, however, tend to frighten individuals and to let them present as others would they look like. This is the typical phenomenon of creating a digital representation that differs from the original one to some extent. In such a context, deriving personality traits of an individual might not be well representative. However, there is some research in literature [93] that proves this hypothesis wrong. In addition, language features and word choices that are the only information on which we base our analysis should be less affected by this issue. The interaction context affects our typical self-manifestation, increasing or decreasing the expression of certain parts of our personality, and thus it influences the ability to determine how well our approach can predict personality.

8. Conclusions and Future Work

In this paper, we presented a supervised learning approach to compute personality traits by only relying on what an individual tweets about his thoughts publicly. The approach segments tweets in tokens, it does ad-hoc preprocessing of tweets, then it learns word vector representations as embeddings that are then combined and used to feed a radial SVM classifier. Our approach has developed five different learning models, one for each personality trait. We tested the convergence of our approach using an international benchmark of Facebook status updates in a controlled experiment

and we observed low mean squared errors from the predicted vs. the actual values. We then applied the learned models with a Twitter sample that we generated in-house. The generation protocol of the sample followed a well-defined psychological test and we collected actual values of personality traits from twenty four panelists who gave their consent in collecting and processing their tweets. We then tested our five learning models and observed lower MSE values than those obtained with the Facebook international benchmark dataset. The source code is publicly available and accessible at [94].

In the approach we proposed, Twitter posts are treated individually. An experiment we are running aggregates the posts from the same user by computing the average for each one of the 900 dimensions. In this way, a single training sample would carry a lot more information and we expect to observe an increase of the accuracy. However, the relatively small size of the dataset we are using does not allow us to conduct such analysis. We thus plan to experiment with the larger myPersonality dataset and to extend, meanwhile, the Twitter sample by acquiring more panelists running our questionnaire addressing the overspecialization introduced by analyzing a homogeneous network of people. We are also studying the effect of the interaction context in our approach. Having a larger gold standard, we plan to compute word representations using a Convolutional Neural Network to capture the context of a set of tweets written by the same person. The goal is to tune the embedding generation according to the chosen optimization function and thus minimizing the mean square error over the actual personality trait values and being able to update the initial FastText embeddings. We aim to capture also a richer semantic representation of words so that we can correlate vectors to personality dimensions in a continuous learning fashion. The current approach has been tested on English content only, we also plan to extend it to other languages exploiting the similarity of word meanings in the vector space.

Author Contributions: G.C. performed the literature review, he ran the experiments and documented them. G.R. designed the approach and lead both the experiments and the reporting part. D.M. contributed in the creation of the Twitter sample and its description. E.P. elaborated the myPersonality gold standard description and supported the creation of the Twitter sample. M.M. contributed to the writing and supported with idea brainstorming.

Acknowledgments: A special thank to the myPersonality project for having shared with us the dataset used for training our learning model and conducting the experimentation. Authors also thank the twenty six Twitter users who took part to the survey and allowed us to collect and process their tweets.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kosinski, M.; Stillwell, D.; Graepel, T. Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5802–5805. [CrossRef] [PubMed]
2. Gottschalk, L.A.; Gleser, G.C. *The Measurement of Psychological States through the Content Analysis of Verbal Behavior*; University of California Press: Oxford, UK, 1969.
3. Graham, D.T.; Stern, J.A.; Winokur, G. Experimental investigation of the specificity of attitude hypothesis in psychosomatic disease. *Psychosom. Med.* **1958**, *20*, 446–457. [CrossRef] [PubMed]
4. Mergenthaler, E. Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *J. Consult. Clin. Psychol.* **1996**, *64*, 1306–1315. [CrossRef] [PubMed]
5. Pennebaker, J.; King, L.A. Linguistic Styles: Language Use as an Individual Difference. *Personal. Soc. Psychol.* **1999**, *77*, 1296–1312. [CrossRef]
6. Argamon, S.; Dhawle, S.; Koppel, M.; Pennebaker, J. Lexical predictors of personality type. In Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America, Cincinnati, OH, USA, 24–28 July 2005.
7. Barrick, M.; Mount, M. The Big Five personality dimensions and job performance: A meta-analysis. *Pers. Psychol.* **1991**, *44*, 1–26. [CrossRef]
8. Saulsman, L.; Page, A. The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clin. Psychol. Rev.* **2004**, *23*, 1055–1085. [CrossRef] [PubMed]

9. Huang, Y.; Wei, L.; Chen, Y. Detection of the Prodromal Phase of Bipolar Disorder from Psychological and Phonological Aspects in Social Media. *arXiv* **2017**, arXiv:1712.09183. [CrossRef]
10. Shaver, P.; Brennan, K. Attachment styles and the "Big Five" personality traits: Their connections with each other and with romantic relationship outcomes. *Personal. Soc. Psychol. Bull.* **1992**, *18*, 536–545. [CrossRef]
11. Rentfrow, P.; Gosling, S. The do re mi's of everyday life: The structure and personality correlates of music preferences. *J. Personal. Soc. Psychol.* **2003**, *84*, 1236–1256. [CrossRef]
12. Dollinger, S. Research note: Personality and music preference: Extraversion and excitement seeking or openness to experience? *Psychol. Music* **1993**, *21*, 73–77. [CrossRef]
13. Hansen, C.; Hansen, R. Constructing personality and social reality through music: Individual differences among fans of punk and heavy metal music. *J. Broadcast. Electron. Media* **1991**, *35*, 335–350. [CrossRef]
14. Rawlings, D.; Ciancarelli, V. Music preference and the five-factor model of the NEO Personality Inventory. *Psychol. Music* **1997**, *25*, 120–132. [CrossRef]
15. Jost, J.; West, T.; Gosling, S. Personality and ideology as determinants of candidate preferences and Obama conversion in the 2008 US presidential election. *Du Bois Rev.* **2009**, *6*, 3–124. [CrossRef]
16. Cantador, I.; Fernandez-Tobias, I.; Bellogín, A.; Kosinski, M.; Stillwell, D. Relating Personality Types with User Preferences Multiple Entertainment Domains. In Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization (UMAP 2013), Rome, Italy, 10–14 June 2013; p. 997.
17. Celli, F.; Lepri, B.; Biel, J.; Gatica-Perez, D.; Riccardi, G. The workshop on computational personality recognition 2014. In Proceedings of the 22nd ACM International Conference on Multimedia (MM '14), Orlando, FL, USA, 3 November 2014; pp. 1245–1246.
18. Tkalčič, M.; de Carolis, B.; de Gemmis, M.; Odić, A.; Košir, A. Preface: EMPIRE 2014-2nd Workshop Emotions and Personality in Personalized Services; In Proceedings of the 22st Conference on User Modeling, Adaptation, and Personalization (UMAP 2014), Aalborg, Denmark, 7–11 July.
19. Hughes, D.; Rowe, M.; Batey, M.; Lee, A. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Comput. Hum. Behav.* **2011**, *28*, 561–569. [CrossRef]
20. Bachrach, Y.; Kosinski, M.; Graepel, T.; Kohli, P.; Stillwell, D. Personality and patterns of Facebook usage. In Proceedings of the 4th Annual ACM Web Science Conference 2012 (WebSci'12), Evanston, IL, USA, 22–24 June 2012.
21. Gosling, S.D.; Augustine, A.A.; Vazire, S.; Holtzman, N.; Gaddis, S. Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychol. Behav. Soc. Netw.* **2011**, *14*, 483–488. [CrossRef] [PubMed]
22. Quercia, D.; Kosinski, M.; Stillwell, D.; Crowcroft, J. Our twitter profiles, our selves: Predicting personality with twitter. In Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), Boston, MA, USA, 9–11 October 2011; pp. 180–185. [CrossRef]
23. Jusupova, A.; Batista, F.; Ribeiro, R. Characterizing the Personality of Twitter Users based on their Timeline Information. In Proceedings of the Atas da 16 Conferência da Associacao Portuguesa de Sistemas de Informação, Porto, Portugal, 22–24 December 2016; Volume 16, pp. 292–299.
24. Liu, F.; Perez, J.; Nowson, S. A Language-independent and Compositional Model for Personality Trait Recognition from Short Texts. *arXiv* **2016**, arXiv:1610.04345. [CrossRef]
25. Van de Ven, N.; Bogaert, A.; Serlie, A.; Brandt, M.J.; Denissen, J.J. Personality perception based on LinkedIn profiles. *J. Manag. Psychol.* **2017**, *32*, 419–429. [CrossRef]
26. YouYou, W.; Kosinski, M.; Stillwell, D. Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. USA* **2014**, *112*, 1036–1040. [CrossRef] [PubMed]
27. Nowson, S.; Oberlander, J. The Identity of Bloggers: Openness and gender in personal weblogs. In Proceedings of the AAAI Spring Symposium, Computational Approaches to Analysing Weblogs, Boston, MA, USA, 16–20 July 2006; pp. 163–167.
28. Kalghatgi, M.P.; Ramannavar, M.; Sidnal, N.S. A neural network approach to personality prediction based on the bigfive model. *Int. J. Innov. Res. Adv. Eng.* **2015**, *2*, 56–63.
29. Su, M.; Wu, C.; Zheng, Y. Exploiting turn-taking temporal evolution for personality trait perception in dyadic conversations. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 733–744. [CrossRef]
30. Majumder, N.; Poria, S.; Gelbukh, A.; Cambria, E. Deep Learning-Based Document Modeling for Personality Detection from Text. *IEEE Intell. Syst.* **2017**, *32*, 74–79. [CrossRef]

31. Mikolov, T.; Chen, K.; Corrado, G.S.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, *arXiv:1301.3781*. [CrossRef]
32. Mairesse, F.; Walker, M.; Mehl, M.; Moore, R. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *J. Artif. Intell. Res.* **2007**, *30*, 457–500.
33. Turian, J.; Ratinov, L.; Bengio, Y. Word representations: A simple and general method for semi-supervised learning. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10), Uppsala, Swede, 11–16 July 2010.
34. Pasca, M.; Lin, D.; Bigham, J.; Lifchits, A.; Jain, A. Names and similarities on the web: Fact extraction in the fast lane. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, (ACL-44), Sydney, Australia, 17–18 July 2006.
35. Manning, C.; Raghavan, P.; Schtze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, MA, USA, 2008.
36. Shutze, H. Distributional part-of-speech tagging. In Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics (EACL '95), Dublin, Ireland, 27–31 March 1995; pp. 141–148.
37. Ratinov, L.; Roth, D. Design challenges and misconceptions in named entity recognition. In Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL '09), Boulder, CO, USA, 4–5 June 2009; pp. 147–155.
38. Kuang, S.; Davison, B. Learning Word Embeddings with Chi-Square Weights for Healthcare Tweet Classification. *Appl. Sci.* **2017**, *7*, 846. [CrossRef]
39. Pennington, J.; Socher, R.; Manning, C. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Volume 14, pp. 1532–1543.
40. Lebret, R.; Legrand, J.; Collobert, R. Is deep learning really necessary for word embeddings? In Proceedings of the NIPS 2013 Deep Learning Workshop, Lake Tahoe, CA, USA, 5–10 December 2013.
41. Dhillon, P.S.; Foster, D.; Ungar, L. Multi-view learning of word embeddings via cca. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*; MIT Press Ltd.: Cambridge, MA, USA, 2011; p. 24.
42. Tang, D.; Wei, F.; Yang, N.; Zhou, M.; Liu, T.; Qin, B. Learning sentiment-specific word embedding for twitter sentiment classification. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 22-27 June 2014; Volume 1, pp. 1555–1565.
43. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M. Kavukcuoglu, K.; Kuksa, P. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
44. Golbeck. J.; Robles, C.; Turner, K. Predicting personality with social media. In Proceedings of the CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11), Vancouver, BC, Canada, 7–12 May 2011; Volume 10, pp. 253–262.
45. Jiang, L.; Yu, M.; Zhou, M.; Liu, X.; Zhao, T. Target-dependent twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11), Portland, Oregon, 19–24 June 2011; Volume 1, pp. 151–160.
46. Hu, X.; Tang, J.; Gao, H.; Liu, H. Unsupervised sentiment analysis with emotional signals. In Proceedings of the 22nd International Conference on World Wide Web (WWW '13), Rio de Janeiro, Brazil, 13–17 May 2013; pp. 607–618.
47. Mohammad, S.M.; Kiritchenko, S.; Zhu, X. Nrc-Canada: Building the state-of-the-art in sentiment analysis of tweets. In Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013), Atlanta, GA, USA, 13–15 June 2013.
48. Kanavos, A.; Nodarakis, N.; Sioutas, S.; Tsakalidis, A.; Tsolis, D.; Tzimas, G. Large Scale Implementations for Twitter Sentiment Classification. *Algorithms* **2017**, *10*, 33. [CrossRef]
49. Dai, H.; Touray, M.; Jonnagaddala, J.; Shabbir, S.A. Feature Engineering for Recognizing Adverse Drug Reactions from Twitter Posts. *Information* **2016**, *7*, 27. [CrossRef]
50. Chamberlain, B.P.; Humby, C.; Deisenroth, M.P. Probabilistic inference of twitter users' age based on what they follow. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Skopje, Macedonia, 18–22 September 2017; pp. 191–203.

51. Zhang, J.; Hu, X.; Zhang, Y.; Liu, H. Your Age Is No Secret: Inferring Microbloggers' Ages via Content and Interaction Analysis. In Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016.
52. Burger, J.D.; Henderson, J.; Kim, G.; Zarrella, G. Discriminating gender on Twitter. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11), Edinburgh, UK, 27–31 July 2011; pp. 1301–1309.
53. Conover, M.D.; Gonalves, B.; Ratkiewicz, J.; Flammini, A.; Menczer, F. Predicting the political alignment of Twitter users. In Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), Boston, MA, USA, 9–11 October 2011; Volume 10, pp. 192–199.
54. Cheng, Z.; Caverlee, J.; Lee, K. You are where you tweet: A content-based approach to geo-locating Twitter users. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10), Toronto, ON, Canada, 26–30 October 2010; pp. 759–768.
55. Pennacchiotti, M.; Popescu, A.M. A machine learning approach to twitter user classification. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM-11), Catalonia, Spain, 17–21 July 2011; Volume 11, pp. 281–288.
56. McCrae, R.R.; Oliver, P.J. An Introduction to the Five-Factor Model and Its Applications. *J. Personal.* **1992**, 60, 175–215. [CrossRef]
57. Tupes, E.; Christal, R. Recurrent Personality Factors Based on Trait Ratings. *J. Personal.* **1992**, *60*, 225–251. [CrossRef]
58. Digman, J. Personality Structure: Emergence of the FiveFactor Model. *Ann. Rev. Psychol.* **1990**, *41*, 417–440. [CrossRef]
59. Goldberg, L. The Structure of Phenotypic Personality Traits. *Am. Psychol.* **1993**, *48*, 26–34. [CrossRef] [PubMed]
60. Mershon, B.; Gorsuch, R.L. Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *J. Personal. Soc. Psychol.* **1988**, *55*, 675–680. [CrossRef]
61. Paunonen, S.V.; Ashton, M.S. Big Five factors and facets and the prediction of behavior. *J. Personal. Soc. Psychol.* **2001**, *81*, 524–539. [CrossRef]
62. McCrae, R.R.; Costa, P.T., Jr.; Busch, C.M. Evaluating comprehensiveness in personality systems: The California Q-Set and the five-factor model. *J. Psychol.* **1986**, *54*, 430–466. [CrossRef]
63. Costa, P.T.; McCrae, R.R. *Revised NEO Personality Inventory (NEO Pl-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*; Psychological Assessment Resources: Odessa, FL, USA, 1992.
64. International Personality Item Pool. Available online: http://ipip.ori.org (accessed on 18 May 2018).
65. Goldberg, L.R. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personal. Psychol. Eur.* **1999**, *7*, 7–28.
66. Johnson, J.A. Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *J. Res. Personal.* **2014**, *51*, 78–89. [CrossRef]
67. Pennebaker, J.; Niederhoffer, K.; Mehl, M. Psychological Aspects of Natural Language Use: Our words, Our Selves. *Ann. Rev. Psychol.* **2003**, *54*, 547–577. [CrossRef] [PubMed]
68. Oberlander, J.; Gill, A.J. Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discour. Process.* **2006**, *42*, 239–270. [CrossRef]
69. Chang, C.; Saravia, E.; Chen, Y. Subconscious Crowdsourcing: A feasible data collection mechanism for mental disorder detection on social media. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 18–21 August 2016; pp. 374–379.
70. Chin, D.N.; Wright, W.R. Social Media Sources for Personality Profiling. In Proceedings of the 22nd Conference on User Modeling, Adaptation, and Personalization, Aalborg, Denmark, 7–11 July 2014; Volume 1181, pp. 79–85.
71. Anderson, C.; John, O.; Keltner, D.; Kring, A. Who attains social status? Effects of personality and physical attractiveness in social groups. *J. Personal. Soc. Psychol.* **2001**, *81*, 116. [CrossRef]
72. Berry, D.S.; Willingham, J.K.; Thayer, C.A. Affect and personality as predictors of conflict and closeness in young adults' friendships. *J. Res. Personal.* **2000**, *34*, 84–107. [CrossRef]

73. Rosen, P.; Kluemper, D. The impact of the Big Five personality traits on the acceptance of social networking website. In Proceedings of the Americas Conference on Information Systems (AMCIS 2008), Toronto, ON, Canada, 14–17 August 2008.
74. Schrammel, J.; Köffel, C.; Tscheligi, M. Personality traits, usage patterns and information disclosure in online communities. In Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology (BCS-HCI), Cambridge, UK, 1–5 September 2009; pp. 169–174.
75. McPherson, M.; Smith-Lovin, L.; Cook, J.M. Birds of a Feather: Homophily in Social Networks. *Ann. Rev. Sociol.* **2001**, *27*, 415–444. [CrossRef]
76. Buffardi, L.E.; Campbell, W.K. Narcissism and social networking Web sites. *Personal. Soc. Psychol. Bull.* **2008**, *34*, 1303–1314. [CrossRef] [PubMed]
77. Mehdizadeh, S. Self-presentation 2.0: Narcissism and self-esteem on Facebook. *Cyberpsychol. Behav. Soc. Netw.* **2010**, *13*, 357–364. [CrossRef] [PubMed]
78. Orr, E.S.; Sisic, M.; Ross, C.; Simmering, M.G.; Arseneault, J.M.; Orr, R.R. The influence of shyness on the use of Facebook in an undergraduate sample. *Cyberpsychol. Behav.* **2009**, *12*, 337–340. [CrossRef] [PubMed]
79. Ross, C.; Orr, E.S.; Sisic, M.; Arseneault, J.M.; Simmering, M.G.; Orr, R.R. Personality and motivations associated with Facebook use. *Comput. Hum. Behav.* **2009**, *25*, 578–586. [CrossRef]
80. Sheldon, P. The relationship between unwillingness-to-communicate and students' Facebook use. *J. Media Psychol.* **2008**, *20*, 67–75. [CrossRef]
81. Goldberg, L.R.; Johnson, J.A.; Eber, H.W.; Hogan, R.; Ashton, M.C.; Cloninger, C.R.; Gough, H.G. The international personality item pool and the future of public-domain personality measures. *J. Res. Personal.* **2006**, *40*, 84–96. [CrossRef]
82. Celli, F.; Pianesi, F.; Stillwell, D.; Kosinski, M. *Workshop on Computational Personality Recognition: Shared Task*; AAAI: Menlo Park, CA, USA, 2013.
83. Barthelemy, M. Betweenness centrality in large complex networks. *Eur. Phys. J. B* **2004**, *38*, 163–168. [CrossRef]
84. List of Stopwords Used by Scikit-Learn. Available online: https://github.com/scikit-learn/scikit-learn/blob/master/sklearn/feature_extraction/stop_words.py (accessed on 18 May 2018).
85. John, O.P.; Naumann, L.P.; Soto, C.J. Paradigm shift to the integrative Big Five trait taxonomy. In *Handbook of Personality: Theory and Research*; History, Measurement, and Conceptual Issues; John, O.P., Robins, R.W., Pervin, L.A., Eds.; Guilford Press: New York, NY, USA, 2008; pp. 114–158.
86. Li, Q.; Shah, S.; Fang, R.; Liu, X.; Nourbakhsh, A. Data Sets: Word Embeddings Learned from Tweets and General Data. *arXiv*, **2017**, arXiv:1708.03994. [CrossRef]
87. Landauer, T.K.; Dumais, S.T. A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychol. Rev.* **1997**, *104*, 211–240. [CrossRef]
88. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.J.; Vapnik, V.N. Support vector regression machines. In *Advances in Neural Information Processing Systems 9, NIPS*; MIT Press: Cambridge, MA, USA, 1996; pp. 155–161.
89. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory (COLT '92), Pittsburgh, PA, USA, 27–29 July 1992.
90. Kenney, J.F.; Keeping, E.S. Linear regression and correlation. *Math. Stat.* **1962**, *15 Pt 1*, 252–285.
91. Tibshirani, R. Regression shrinkage selection via the LASSO. *J. R. Stat. Soc. Ser. B* **2011**, *73*, 273–282. [CrossRef]
92. fastText English Word Vectors. Available online: https://fasttext.cc/docs/en/english-vectors.html (accessed on 18 May 2018).
93. Back, M.; Stopfer, J.; Vazire, S.; Gaddis, S.; Schmukle, S.; Egloff, B.; Gosling, S. Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychol. Sci.* **2010**, *21*, 372–374. [CrossRef] [PubMed]
94. TwitPersonality. Available online: https://github.com/D2KLab/twitpersonality (accessed on 18 May 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).