Personality Extraction Through LinkedIn

Frédéric Piedboeuf[1(✉)], Philippe Langlais[1], and Ludovic Bourg[2]

[1] RALI, Université de Montréal, Montreal, Canada
frederic.piedboeuf@umontreal.ca, felipe@iro.umontreal.ca
[2] LittleBIGJob, Montreal, QC H3B 4W5, Canada
ludovic.bourg@lorenzandhamilton.com
http://rali.iro.umontreal.ca/rali/

Abstract. LinkedIn is a professional social network used by many recruiters as a way to look for potential employees and communicate with them. In order to facilitate communication, it is possible to use personality models to gain a better understanding of what drives the person of interest. This paper first looks at the possibility of collecting a corpus on LinkedIn labelled with a personality model, which has never been done before, then looks at the possibility of extracting two different personalities from the user. We show that we can achieve results going from 73.7% to 80.5% of precision on the DiSC personality model and from 80.7% to 86.2% of precision on the MBTI personality model. These results are similar to what has been found on other social networks such as Facebook or Twitter, which is surprising given the more professional nature of LinkedIn. Finally, an analysis of the significance of the results and of the possible sources of errors is presented.

1 Introduction

With the advent of social networks and their more recent popularization through mainstream culture, the internet has now become a considerable source of information on people habits, opinions, and personalities. Research focusing on using this information has taken a great importance in the field of natural language processing, and more recently, personality extraction has started to grab the attention of researchers.

Personality extraction is the task of assessing the personality of a user through his social network interactions. The personality of an individual is defined as the set of responses to external stimuli [7]. It is usually described according to one of several models of personality that are used in psychology, such as the MBTI or the DiSC, both explained in Sect. 2. Personality extraction has many uses, from personalizing the user experience to using the personality type as input to other machine learning problems, e.g., by using it to better extract influential communities on Twitter [6].

The most popular social networks for such tasks are probably Twitter and Facebook, because they provide easy platforms for collecting the corpus and allow easy testing of the personality with the help of external web applications

M.-J. Meurs and F. Rudzicz (Eds.): Canadian AI 2019, LNAI 11489, pp. 55–67, 2019.
https://doi.org/10.1007/978-3-030-18305-9_5

that administer the tests. However, in the professional world, Facebook and Twitter are rarely consulted and most often, the professional social network LinkedIn will be favoured for evaluating a potential candidate when hiring [2]. To our knowledge, the only other study that has looked at personality assessment on LinkedIn is the one of [20], but the evaluation of the personality was made by humans and not by computers, even though the latter have been shown to achieve a better precision score when classifying personality based on digital footprints [21].

This paper looks at the possibility of automatically extracting the personality of a LinkedIn user and shows, somewhat surprisingly given the peculiar nature of the texts on LinkedIn, that our results are comparable to those reported on other types of texts. The automatic extraction of personality on LinkedIn could become an important tool for many companies, as it could diminish the cost of employee selection processes. It could also help improve communications with potential employees by understanding what drives them and what kind of communication they better respond to.

2 Personality Models

The personality model that has probably been the most studied in psychology is the Big-5 [4]. It characterizes a person with the help of 5 traits, usually on a scale of 1 to 5: Openness to experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. Although this model has been used extensively in the scientific community, the continuous nature of the test results makes it difficult for the non-scientific community to discuss their results and so other, categorical models, such as the Myer-Briggs, have become more popular.

The Myer-Briggs model, also called MBTI, evaluates the personality along 4 different axes, representing the way one processes the surrounding information [11]. The four axes are:

- Introversion (I) vs. Extroversion (E), or does one focus on internal stimuli or external ones.
- Sensing (S) vs. Intuition (N), or does one process the events as facts happening or by trying to find patterns and meaning.
- Thinking (T) vs. Feeling (F), or does one make decisions by thinking over the consequences or by following their instincts.
- Judging (J) vs. Perceiving (P), or does one prefer the outside world to be structured or to be flexible and spontaneous.

The combination of these four characteristics form 16 distinct types of personality that are identified with a code such as “ENFP” or “ISFJ”.

In our case, we were interested in DiSC, which is another categorical test that was developed as a way to describe how people interact with each other [15]. The model gives a score to 4 different traits: Dominance, Influence, Steadiness, and Conscientiousness, which indicate what drives the person in their professional environment. Dominant people are driven by results, Influent people by

relationships with others and influencing others, Steady people by cooperation with other people, and Conscientious people by the quality and accuracy of the work done [17].

Since the DiSC personality model has not been as widely studied as some other personality models, our study also looks at the MBTI personality model, which is used extensively by companies and has also been considerably studied by researchers [8,9,12]. The goal of this research is to:

1. Assess whether a LinkedIn profile has enough information to successfully extract the personality of its author;
2. Compare the results of DiSC extraction with the results of MBTI extraction in order to determine how significant the DiSC extraction results are.

3 Related Work

Given the nature of personality models, most researchers break down the problem in as many traits of personality as the model has. For example, it is common practice to break the Big-5 model into 5 different classifiers, one for each of the traits, and the final result is then the prediction of all 5 classifiers.

In [14], the authors were among the firsts to enter the field by predicting the personality of Twitter users based on their number of following, followers, the number of times they had been listed, and the number of Facebook social contacts. They achieved a Normalized Root Square Error (NRSE) of 0.2 on average over the different traits of the Big-5 model, showing that we can successfully extract personality based on very little information - if that information is pertinent.

Also using the Big-5 model, the authors of [10] trained classifiers on a Facebook dataset of 250 profiles containing both texts from the users as well as some other information, such as the number of groups the person belongs to or the number of likes. The corpus was provided in the context of a shared task, and therefore 7 other teams also had access to the dataset [3]. They extracted a total of 725 features and performed feature selection with a Support Vector Machine (SVM) using the top 5 to 16 features, achieving better results for each of the traits than the other participating teams.

In [13], a way to obtain a corpus automatically by selecting users that voluntarily put their personality test results on social networks was proposed. The study was carried out using the MBTI model, which has the double advantage of being popular with the general population and easy to lookup using regular expressions, since the code associated with the different personality models can be searched for directly. Overall, they harvested 1.2 million tweets belonging to 1500 different users and used logistic regression with feature selection to predict each trait of the MBTI model, but achieved poor results, with a precision below the majority baseline for the Intuition-Sensing and Judging-Perception traits.

While these studies focused on the MBTI and Big-5 personality models, the two most popular ones, it is important to also study other personality models that are commonly used to avoid restricting the playing field. The authors

of [1] chose Twitter for a data-mining study with the DiSC personality model. Using a list of keywords associated with each of the 4 dimensions of the model, they downloaded tweets containing these keywords. They then performed data analysis in order to extend the vocabulary related to each of the traits by finding the most common words used by each of the types, excluding stop words. Even if they did not perform classification on the data, the study still showed a correlation between the vocabulary used and the personality type, and laid the foundation for other personality studies using the DiSC personality model.

In [12] the starting dataset was a Dutch corpus of essays written by 145 students. Each essay was also labelled with the MBTI personality of the student. They used a centroid-based model to classify each essay according to each of the personality traits, achieving precision scores ranging between 76% and 85%.

The authors of [5] obtained a corpus labelled with the MBTI personality test results from the internet forum personalitycafe.com, a forum based on personality types. Working only with the text, they used a recurrent neural network with pre-trained word embeddings to predict the personality of the users. They broke down the MBTI model into its 4 dimensions and obtained precision figures ranging from 62% to 78%, which is lower than other studies on the MBTI model, suggesting that the use of neural networks on personality extraction has not been developed enough yet to outperform classic algorithms.

4 Corpus Preparation

Since there is no other study on automatic personality extraction on LinkedIn, no corpus was available. In order to obtain one, we followed the method of [13], who devised a way to extract social network profiles of users who had completed a personality test and posted the results on their social networks.

The GoogleSearch Library for Python allowed us to access approximately 81 million public LinkedIn profiles. We initially looked for all profiles containing the words “DiSC” and “personality”, which gave us a total of 19,000 profiles. However, an examination of 100 randomly selected profiles found 4 of them having an actual DiSC personality and the rest of them being noise. The noise was introduced mostly by team-building coaches and musicians and although the words “DiSC” and “personality” were present on their LinkedIn page, there was no mention of the user personality test results.

Finding LinkedIn profiles with a DiSC personality by directly researching the personality types was not doable due to the fact that the DiSC personality model uses a very common vocabulary subset (e.g., the Dominant personality can be expressed as dominant, driver, achiever, inspirational, etc.) and because there are many ways for a user to express the same personality, as shown in Fig. 1.

In order to capture LinkedIn profiles labelled with the DiSC personality test results, we surmised that people who would write their MBTI personality would be more likely to also write their DiSC personality in their profiles. A corpus following that assumption would have both personality expressed in their profiles,

Disc Test: D55% I8% S10% C26%	Disc Assessment: (5|1|3|7)	Disc Profile: [Natural D Adaptive I]	My DiSC assessment reveals I am a persuader
High DI and above the line C on DISC	Disc: "SDIC"	Disc: (D) Driver followed by (C) Conscientious	DiSC: Result-Oriented Pattern

Fig. 1. Different ways to express the Dominant trait of the DiSC personality model.

and we would then be able to split them into 2 different corpora, one labelled with the MBTI personality model and the other labelled with the DiSC personality model. A search of the public LinkedIn profiles containing the words "DiSC", "personality", and one of the 16 MBTI codes yielded a total of 1253 profiles, that we then saved in JSON format. To eliminate the noise and keep only the profiles containing a DiSC personality test result, we eliminated all profiles containing the tokens "HR", "certified", "jockey", "administering", "workshop", "golf", "spine", and "coaching" close to the token "DiSC".

This produced a total of 841 profiles for the DiSC corpus, which is the set of profiles that we could label with one of the DiSC personalities, and we kept the original 1253 profiles for the MBTI corpus. The small size of the corpora can be explained by two factors: First, the use of Linkedin, which is a more professional social network than Twitter or Facebook, may discourage people from sharing their personality results on their profile and second, the use of DiSC, a less popular model, yields fewer results than if we had done corpus extraction solely based on MBTI.

An example of a profile, slightly modified for anonymity, is shown in Listing 1.2. We can see that the data will still need a lot of preprocessing before being usable by a classifier.

With the DiSC personality model, each user is expected to have one main trait and one or two supporting traits. However, on LinkedIn, users often express their personality test results as "I am high D and I", or "I am equal part I, S and C" and so it is hard to discern between the main trait and the secondary traits. We decided that we wanted a system that would return the personality test results the users say they have (e.g. DI for "I am high D and I"), and so our labeling did not discern between the main and secondary traits.

Initial attempts to capture the DiSC personality with regular expressions generated errors due to other types of personality tests that had a similar vocabulary to DiSC and to the fact that DiSC uses very common words to describe the different personalities. For example, "Achiever" and "Developer" are both in the vocabularies of the StrengthFinder personality test[1] and the DiSC personality model, which also includes common words such as "Result", "Creative",

[1] https://www.gallupstrengthscenter.com/home/en-us/cliftonstrengths-themes-domains.

and "Agent". The list of regular expressions for the Steadiness trait is reported in Listing 1.1 and shows the complexity of the task of capturing the personality type automatically. Although these were ultimately not used for the labeling process, they were used to remove the words that would give away the personality before classifying.

Listing 1.1. Regular expression for the Steadiness trait

```
S=["High[- ]S", "\"S\"",  "[sS]econdary S", "[Ss]tead[a-zA-Z]*",
    " S ", "[/-] S ", " S[CIiD][CIiD ]", " [CIiD]S", " [CDIi][
    CIiD]S ", "S[CIiSD]?/", " S,","\([CIiD]?S\)", "\(S[CIiD]?\)",
    "\[S\]", "[Aa]chiever", "[Aa]gent", "[Cc]ounselor", "[Ii]
    nvestigator", "[pP]erfectionist", "[pP]ractitioner", "[sS]
    pecialist", "[sS]-oriente", "S[0-9]", "&S",  "S&"]
```

Since we wanted to be sure to have a corpus that was correctly labelled to avoid any mistakes before starting the classification task, we looked at the $-150/+150$ window of characters around the word "DiSC" and labelled the profiles manually. This also gave us the opportunity to check that no noise was in the corpus. The labeling of the MBTI corpus was straightforward, since the MBTI codes could be looked up with simple regular expressions.

Listing 1.2. Linkedin profile slightly modified for anonymity, after being saved in JSON.

```
{
 "Summary": "I am recognized by my peers as a tech-savvy
     intuitive problem-solver. I thrive in environments where
     rapid change and the need to constantly adapt and learn new
     information are considered the norm. My DiSC personality
     is dominant",
  "Personal Branding Claim": "Problem Solver",
  "Connections": "500+",
  "Followers": "0",
  "Skills": ["Leadership 110 endorsements 99+ who are skilled 2
      colleagues", "Strategy 30 endorsements"],
  "Recommendation": "Received (12) Given (3)",
  "Voluntary experiences": {
    "Guest Speaker": {
      "To_date": "Feb 2005",
      "From_date": "Feb 2005"
      "Description": "Guest Speaking at East High School"}},
  "Educations": {
    "Northeastern College of Professional Studies":{
      "Field of study": "Organizational Leadership",
      "From_date": "1987",
      "To_date": "1993",
      "Description": "4 best invested years of my life"}},
  "Experiences": {"Teacher":{
    "To_date": "Present",
    "From_date": "2015",
    "Employer": "Neverends Education",
    "Description": "Teaching software to kids"}},
```

```
"Interests": ["Profyle Tracker 349 followers", "Leoprino Foods
    18802 followers"],
"Achievements": {
  "Course": {
    "Name": "Team management",
    "Description": "null"},
  "Test": {
    "Name": "MBTI",
    "Result": "ENFP"}}}
```


The distribution of the data is represented in Table 1. The classes are imbalanced, but at least for the MBTI results, the distribution is similar to the one found in [13] and to the statistics given by the Myers-Briggs Foundation[2]. Both are also shown in Table 1.

According to the DiSC Profile website[3], the main traits are expected to be roughly equally distributed, with 25% of users in each personality type. However, we found a radically different distribution. This could be explained by two factors. Firstly, the fact that we did not discern between main and supporting traits changes the statistics. For instance, it might be possible that the Dominant trait appears more often as a supporting trait than the Steadiness trait or the Conscientious one. Secondly, it is possible that Dominant and Influent users are more likely to put their personality test results on LinkedIn than Steady and Conscientious users.

Table 1. Percentage of the different traits for the MBTI and DiSC personality

MBTI	Linkedin	Plank 2015	MBTI foundation	DiSC	Linkedin	Disc Profile website (https://www.discprofile.com/)
I/E	35.7	36.0	50.7	D	60.5	24.8
N/S	72.0	73.0	73.3	I	57.2	25.1
T/F	55.6	58.0	59.8	S	27.3	25.7
P/J	34.0	41.0	45.9	C	36.9	24.4

5 Feature Engineering

For both corpora, we extracted textual and non textual features. Forty non textual features were hand-crafted and then extracted from the LinkedIn profiles.

[2] Found at https://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.htm?bhcp=1.

[3] Found at https://www.discprofile.com/what-is-disc/faq/.

5.1 Non-Textual Features

The non-textual features can be broken down in 7 categories.

1. **General information:** Information that concerns the user in general, such as the number of connections, whether the user has something written in the description, etc.
2. **Work:** Information about the work experience, such as the number of jobs, the average duration of those jobs, etc.
3. **Volunteer experience:** These features are related to the volunteer experiences of the user, which contains the number of volunteer experiences, the average duration of those, etc.
4. **Education:** Features related to number of diplomas, average length, etc. Very similar to the work and volunteer experience categories with the added feature of the highest level of education.
5. **Skills:** Features related to the skills section, such as the number of skills, average number of endorsements, number of endorsements from colleagues, etc.
6. **Accomplishments:** The total number of accomplishments, as well as the number of accomplishments broken down in categories (titles, languages, courses, etc.)
7. **Interests:** Features related to the user's interests, such as the number of interests, the number of followers of these interests, etc.

5.2 Textual Features

We also extracted general text from the LinkedIn profiles of each user, which was taken from 1- the introduction written by the user, and 2- the description of schools, jobs, and volunteering experience when these were available. The profiles have an average of 830.9 words, and each word is of an average length of 6.4 characters. We derived 415 textual features from this text. These features included 370 features from the General Inquirer[4] as well as the number of different POS (part of speech) tags used in the text. The General Inquirer is a Java program that takes a text as input and returns both percentages and total counts of words pertaining to each category included in its dictionary. Categories include, but are not limited to: Anxiety, Family, Health, Sadness, etc. [16]. We also used Term Frequency Inverse Document Frequency (Tf-Idf) in our experiments, using the 5000 most frequent words. For each classifier, we tried a combination of Tf-Idf and of the 415 features, with and without stopwords. The best results for each is reported in Sect. 6.

6 Experiments

Before classifying, we were careful to remove any words giving away directly the personality of either the DiSC or the MBTI models by using sets of regular

[4] http://www.wjh.harvard.edu/~inquirer/.

expressions, as shown in Listing 1.1. For the MBTI model, we simply looked for and removed the 16 personality codes.

The same procedure was used for the DiSC corpus and the MBTI corpus, and gave similar results. We trained a classifier for each trait, using 20-fold cross validation. We tested an SVM classifier with a feature ranking algorithm and optimization, a Random Forest with AdaBoost, and a simple Naives Bayes. Several architectures of neural networks were also tried, but due to the small size of the corpora, couldn't learn to generalize efficiently. The baseline chosen was a majority class rule, which seemed appropriate, since the classes are often imbalanced.

The feature ranking algorithm for the SVM started with all features and tried them k at a time, until there were only the k best features, discarding the worst c features at each iteration. Two different measures of efficiency for the features were tried: the difference between the precision with and without the feature, as well as simply the precision with the feature. The latter measure yielded better results and was therefore used here, k and c being meta-parameters of the algorithm. We set them at 30 and 15, respectively, which seemed to give the best results in a timely fashion. As with the final algorithms, we also used a 20-fold cross validation on each loop of the algorithm.

Without the feature ranking algorithm, the SVM performed at the same level or at best a few percents higher than the baseline. With the feature ranking algorithm, the SVM performed on average 11% better than the baseline, and with the fine tuning of the meta parameters through a grid search, the SVM slightly improved in precision. Adaboost wasn't used with the SVM, since it showed no improvements but instead a decrease in performance.

Naive Bayes (NB) and Random Forest (RF) were also tried. Random Forest worked best with a mix of Tf-Idf features and hand-crafted features, presented in Sect. 5. The mix contained only the n most frequent words, and the m hand crafted features used were those having the highest correlation with the trait being predicted. Different numbers of n and m were tried, but at the end, made little difference to the performance for the Random Forest due to its capacity to pick the most useful features for classification. For the Naive Bayes, we used a grid search to find the optimal number of n and m for each of the traits. Since it is a linear classifier, the features were ranked with the help of a Pearson correlation coefficient, and those that were the most correlated were introduced to the algorithm first. AdaBoost was then used to boost performance, allowing to gain on average 1% on precision for both NB and RF. Stop-words removal and lemmatization were also tried, but only hurt the performance and therefore were not kept.

The complete results can be seen in Table 2. An interesting thing to note here is that the SVM and the Random Forest capture some very distinct phenomenons. While the SVM finds itself, after the feature ranking algorithm, using mostly the non-textual features, as well as some of the POS tags, the RF mostly uses the textual features from the Tf-Idf. Despite several studies finding the features extracted by the General Inquirer very useful for classification [10,16], our

algorithm learned to perform mostly without their help, maybe due to the fact that LinkedIn profiles are often written in a way which seems professional and not emotional.

Table 2. Precision (%) over the different traits for the MBTI (top) and the DiSC (bottom) personality models. NT stands for Non-Textual, T for Textual, NB for Naive Bayes, and RF for Random Forest.

Traits	Baseline	NB	RF NT	RF T	RF	SVM NT	SVM T	SVM
IE	64.3	72.5	75.2	82.6	**84.4**	77.5	71.8	77.9
NS	72.0	78.0	79.1	81.5	**86.2**	82.0	77.1	83.7
TF	55.6	65.5	68.9	78.3	**83.5**	69.4	68.3	71.7
JP	66.0	67.1	72.4	76.5	**80.7**	77.4	73.6	78.4
D	60.5	66.1	70.9	72.7	72.5	72.3	66.5	**73.7**
I	57.2	63.1	71.8	74.5	**75.1**	73.2	66.7	73.8
S	72.7	72.5	78.2	79.3	79.3	80.5	77.7	**80.5**
C	63.1	66.2	72.1	74.1	74.1	75.5	74.1	**76.7**

7 Analysis

Although there is no other study on personality classification on LinkedIn, we can still compare our results to other studies that have used the MBTI model for classification. Since the data is imbalanced, we also report the improvement over the majority baseline when available, which helps giving a better idea of the performance of the algorithm. In Table 3, we show the results of our MBTI classifier, of the study of [12] (Study 1) and the study of [5] (Study 2). In parentheses is the improvement over the majority baseline, which was available for Study 1, but not for Study 2.

Table 3. Comparison between our results and results of other studies for the MBTI personality model

Traits	Our results	Study 1	Study 2
IE	84.4 (+18.1)	76.4 (+21.2)	67.0
NS	86.2 (+14.2)	76.5 (+22.7)	62.0
TF	83.5 (+27.9)	80.5 (+8.1)	77.8
JP	80.7 (+14.7)	84.5 (+3.8)	63.7

We can see that, overall, we get rather good results over all traits. Although we cannot surpass Study 2 for the Judging-Perceiving trait, we get the best score

overall for all other three traits. However, since the classes are very imbalanced, it is hard to say how much this really means.

The errors can be explained by several factors. Firstly, the method used, which takes profiles of people who have already passed the personality test, does not allow us to check how accurate the tests are. Personality tests often range from being given by trained psychologists to tests having been created on the internet by someone having no qualifications to do so. It is then very possible that some of these tests are not entirely accurate and introduce a bias in our data.

The choice of LinkedIn also makes the task more difficult. Since LinkedIn is a social network meant to display the users' professional qualities, not only will people tend to be more self conscious about what they publish [19], but they may also try to guess what type of personality an employer would want and "bend" their results towards the ideal personality. Even if the user did not actively change the results that they obtained on the personality test, the simple mental image of what a good employee is, and the desire to be a good employee at the moment of the test, may bias the outcome.

The results obtained here are nevertheless promising. The two goals of this study were to find out if LinkedIn had enough information to correctly classify personality, as well as if DiSC could be considered a valid personality model compared to those that have been more studied. Our initial assumption was that it would be harder to classify personalities on LinkedIn due to the peculiar nature of the social network, requiring that the user be more reserved in what is written since the user expects to be read by eventual employers.

The results show, interestingly, that it is possible to classify LinkedIn profiles to the same extent as profiles from other social networks. However, it should be mentioned that the LinkedIn profiles we obtained are profiles that had mostly been filled out thoroughly, and so it should be interesting to see how well this generalizes to profiles that lack information.

This study also gives some validity to the DiSC personality model. We collected a fairly small corpus and so it could be argued that the correlation found between profiles and personality traits are due to pure luck, but it would be highly improbable.

8 Conclusion and Future Work

LinkedIn is a professional social network used by many employers as a means to screen potential employees and communicate with them. In order to facilitate contacts between employers and employees, it would be easier to know the personality of the person contacted. However, it is very difficult, for a human, to detect the personality of someone on LinkedIn. This is due to the fact that descriptions on LinkedIn are often written in a way that the employees think the employers want, and so is not in accordance to their natural writing style [19].

Our study shows that a computer is able to extract the personality from a LinkedIn profile with a reliable precision. This was done using two personality

models that are often used in professional settings as a way to better understand employees or coworkers: the MBTI personality model and the DiSC personality model.

Although studies have been done on correlating job success to personality traits [18], only neuroticism has been found to be negatively correlated with job performance, and so the application of personality extraction in the trimming of candidates must be done carefully and employers must work to avoid putting their personal biases in the selection process.

This study opens the door, however, to many interesting applications, such as facilitating conversations between employers and employees, helping recruiters find people that would fit the job better, or helping LinkedIn users to themselves find the next career step that would be best suited for their talents and personality.

In the future, studies on personality extraction through LinkedIn could focus on either getting a larger dataset so that better machine learning algorithms may be applied, finding a way to transpose with accuracy the linguistic manual features to other languages, or working on the anonymization of a LinkedIn dataset so that researchers may compare their results on a common benchmark.

References

1. Ahmad, N., Siddique, J.: Personality assessment using Twitter tweets. Proc. Comput. Sci. **112**, 1964–1973 (2017)
2. Caers, R., Castelyns, V.: Linkedin and Facebook in Belgium: the influences and biases of social network sites in recruitment and selection procedures. Soc. Sci. Comput. Rev. **29**(4), 437–448 (2011)
3. Celli, F., Pianesi, F., Stillwell, D., Kosinski, M.: Workshop on computational personality recognition (shared task). In: Proceedings of the Workshop on Computational Personality Recognition (2013)
4. Goldberg, L.R.: The structure of phenotypic personality traits. Am. Psychol. **48**(1), 26 (1993)
5. Hernandez, R.K., Scott, I.: Predicting Myers-Briggs type indicator with text (2017)
6. Kafeza, E., Kanavos, A., Makris, C., Vikatos, P.: T-PICE: Twitter personality based influential communities extraction system. In: 2014 IEEE International Congress on Big Data (BigData Congress), pp. 212–219. IEEE (2014)
7. Kaushal, V., Patwardhan, M.: Emerging trends in personality identification using online social networks' a literature survey. ACM Trans. Knowl. Discov. Data (TKDD) **12**(2), 15 (2018)
8. Lima, A.C.E., de Castro, L.N.: Predicting temperament from Twitter data. In: 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 599–604. IEEE (2016)
9. Ma, A., Liu, G.: Neural networks in predicting Myers Brigg personality type from writing style (2017)
10. Markovikj, D., Gievska, S., Kosinski, M., Stillwell, D.: Mining Facebook data for predictive personality modeling. In: Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA, pp. 23–26 (2013)

11. Myers, I.B., McCaulley, M.H., Quenk, N.L., Hammer, A.L.: MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator, vol. 3. Consulting Psychologists Press, Palo Alto (1998)
12. Noecker Jr., J., Ryan, M., Juola, P.: Psychological profiling through textual analysis. Literary Linguistic Comput. **28**(3), 382–387 (2013)
13. Plank, B., Hovy, D.: Personality traits on Twitter—or—how to get 1,500 personality tests in a week. In: Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 92–98 (2015)
14. Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our Twitter profiles, our selves: predicting personality with Twitter. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), pp. 180–185. IEEE (2011)
15. Reynierse, J.H., Ackerman, D., Fink, A.A., Harker, J.B.: The effects of personality and management role on perceived values in business settings. Int. J. Value Based Manage. **13**(1), 1–13 (2000)
16. Stone, P.J., Bales, R.F., Namenwirth, J.Z., Ogilvie, D.M.: The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information. Behav. Sci. **7**(4), 484–498 (1962)
17. Sugerman, J.: Using the disc® model to improve communication effectiveness. Ind. Commercial Training **41**(3), 151–154 (2009)
18. Tett, R.P., Jackson, D.N., Rothstein, M.: Personality measures as predictors of job performance: a meta-analytic review. Pers. Psychol. **44**(4), 703–742 (1991)
19. Van Dijck, J.: 'You have one identity': performing the self on Facebook and Linkedin. Media Cult. Soc. **35**(2), 199–215 (2013)
20. van de Ven, N., Bogaert, A., Serlie, A., Brandt, M.J., Denissen, J.J.: Personality perception based on Linkedin profiles. J. Managerial Psychol. **32**(6), 418–429 (2017)
21. Youyou, W., Kosinski, M., Stillwell, D.: Computer-based personality judgments are more accurate than those made by humans. Proc. Nat. Acad. Sci. **112**(4), 1036–1040 (2015)