

# How to Obtain Reliable Labels for MBTI Classification from Texts?

Sanja Štajner

Symanto Research  
Nuremberg, Germany

sanja.stajner@symanto.com

Seren Yenikent

Symanto Research  
Nuremberg, Germany

seren.yenikent@symanto.com

## Abstract

Automatic detection of the Myers-Briggs Type Indicator (MBTI) from short posts attracted noticeable attention in the last few years. Recent studies showed that this is quite a difficult task, especially on commonly used Twitter data. Obtaining MBTI labels is also difficult, as human annotation requires trained psychologists, and automatic way of obtaining them is through long questionnaires of questionable usability for the task. In this paper, we present a method for collecting reliable MBTI labels via only four carefully selected questions that can be applied to any type of textual data.

## 1 Introduction

The Myers-Briggs Type Indicator (MBTI) model (Briggs-Myers and Myers, 1995) is one of the most widely used non-clinical psychometric models (Štajner and Yenikent, 2020). It classifies people into two groups across four dimensions: extraversion/introversion (E/I), sensing/intuition (S/N), thinking/feeling (T/F), and judgement/perception (J/P). This leads to a total of 16 personality types. The first three dimensions are based on the theoretical work of Carl Jung (1921), while the fourth dimension was added later by Myers and Briggs-Myers (1995). The MBTI personality framework has already been used for decades in educational and industry settings, e.g. for finding jobs that best resonate with the person's preferences for information processing (S/N and T/F dimensions), finding work organization types that best resonate with the person's preferred judgement processes (J/P dimension) thus leading to better job satisfaction, and for better matching work environments with the person's preferences (E/I dimension) to lower employee turnover (Briggs-Myers and Myers, 1995).

The original MBTI questionnaire contains 93 questions and is not freely available.<sup>1</sup> Due to the

popularity of MBTI framework (it is estimated that more than 2 million US adults complete the inventory every year),<sup>2</sup> there is a number of freely available alternative MBTI questionnaires on the internet, with the *16personalities test*<sup>3</sup> being one of the most popular ones. According to the Myers-Briggs Foundation<sup>4</sup> and the 16personality test website,<sup>5</sup> both questionnaires satisfy the accepted standards for test validity and reliability. Nevertheless, the MBTI questionnaires have received a noticeable criticism from the academic community (Pittenger, 1993; Boyle, 1995) for not relying on a scientifically proven (i.e. data-driven) background, but rather on qualitative measures such as observation and introspection. The other common criticism is the binary nature of the questionnaire as it is known that the majority of people usually lies somewhere in the middle of the scales (Pittenger, 1993).

The questionnaire-based personality detection has several weaknesses: it requires trained human assessors; it is prone to social desirability bias (Krumpal, 2011) and reference-group effect (Heine et al., 2002); it is questionable if answering questionnaires is a natural way of showing ones personality (as opposed to free writing or behaviour “when nobody watches”). To detect MBTI typologies in a more natural way and without necessity for trained human assessors, many studies have attempted at building systems for automatic detection of MBTI personality types from text in the last several years. Attempts have been made for automatic detection of MBTI personality types from: tweets written in English (Plank and Hovy, 2015), six other Western European languages (Ver-

professional/versions-of-the-mbti-questionnaire/

<sup>2</sup><https://www.verywellmind.com/the-myers-briggs-type-indicator-2795583#the-mbti-today>

<sup>3</sup><https://www.16personalities.com/free-personality-test>

<sup>4</sup>[myersbriggs.org](https://www.myersbriggs.org)

<sup>5</sup><https://www.16personalities.com/articles/reliability-and-validity>

<sup>1</sup><https://www.myersbriggs.org/using-type-as-a->

hoeven et al., 2016), and Japanese (Yamada et al., 2019); English posts collected from Personality Cafe forum<sup>6</sup> available in Kaggle;<sup>7</sup> and English Reddit comments (Gjurković and Šnajder, 2018; Gjurković et al., 2020). Despite being trained on large amounts of textual data (over one million), and modelled as four binary classification tasks, the best systems performed only slightly better than the random and majority-class baselines, regardless of the architecture used.

Some studies suggested that tweets might not contain sufficient amounts of MBTI signals (even after concatenating up to 150-200 tweets per user) due to the nature of Twitter posts (Celli and Lepri, 2018; Štajner and Yenikent, 2020, 2021). Another issue with all those studies and obtained results might be that the systems are supervised and were trained with gold labels obtained via MBTI questionnaires that suffer from all earlier mentioned weaknesses. In our recent study (Štajner and Yenikent, 2021), we found a low association between the MBTI types obtained via questionnaires and the MBTI signals found in the short texts written by participants (tweets and free texts on carefully chosen topics). At the same time, the inter-annotator agreement of two expert annotators assigning MBTI types based on those free texts was quite high (Štajner and Yenikent, 2021).

**Contributions.** To avoid all previously mentioned problems in automatic MBTI detection from texts, in this study, we propose a carefully designed set of four questions with answers on a 1-5 scale (Section 3) that aim to capture the main MBTI characteristics without taking much time from participants, and can be administered together with any open-end questions without need for trained human assessors. The validity of our questionnaire has been assessed via expert human annotation following previously proposed annotation methodology (Štajner and Yenikent, 2021). The agreement between the answers to the newly proposed questions and the expert human annotations was found to be similar as between two trained annotators (Section 5.2). Another advantage of the proposed method is that it goes beyond binary typology, by offering a 5-point scale for each MBTI dimension. This creates a possibility for filtering out those instances written by people who exhibit similar amount of signals from both polarities. As

it is known that many people have characteristics of both polarities across MBTI dimensions (Pittenger, 1993), such filtering of training datasets might lead to better performances of automatic systems for MBTI detection from texts by removing noise.

## 2 Related Work

Plank and Hovy (2015) were the first to explore the use of Twitter data for obtaining a large-scale dataset for open-vocabulary automatic detection of MBTI personality traits. They collected a corpus of 1.2M English tweets automatically labelled for gender and MBTI type. To identify the users for whom an MBTI type can be automatically assigned, the authors relied on mentions of any of the 16 MBTI types plus the word “*Briggs*”. Additionally, each user was labelled as *female* or *male* whenever it was discernible; those users for whom the gender was not discernible were excluded from the study. For each selected Twitter user, the authors collected up to 2000 most recent tweets (to be included, each user had to have at least 100 tweets). Plank and Hovy (2015) found that the distribution of MBTI types across the selected Twitter users significantly differs from the distribution of MBTI types across the general US population. The authors further trained binary classification models (for each MBTI dimension separately) using various features and model architectures. The best systems outperformed majority-class baselines only for I/E and T/F dimensions.

Verhoeven et al. (2016) used a similar strategy for obtaining large-scale MBTI datasets for six other languages: German, Italian, Dutch, French, Portuguese, and Spanish. As opposed to the work of Plank and Hovy (2015), the triggers for identifying users whose MBTI types can be automatically assigned were mentions of one of the 16 personality types and the word “*personality*” or pronouns and verb forms such as “*I am*” or “*I have*”, for each of the six languages. All retrieved contexts were manually checked for whether or not they describe the personality of the writer of the post. For all users whose posts passed this check, the gender was annotated based on the user’s name, handle, description, and profile picture (Verhoeven et al., 2016). Distributions of MBTI types across Twitter users of the six languages were found to be similar, with only a few exceptions (Verhoeven et al., 2016). The authors also trained binary classifiers using the dataset with 200 concatenated tweets for each user

<sup>6</sup><https://www.personalitycafe.com/>

<sup>7</sup><https://www.kaggle.com/datasnaek/mbti-type>

and LinearSVC classifier with binary word and character  $n$ -gram features. Similar as for English (Plank and Hovy, 2015), in most of the languages, the best classifiers outperformed the majority-class baselines only for E/I and T/F dimensions.

Gjurković and Šnajder (2018) compiled a large-scale MBTI dataset from English Reddit comments by relying on *flairs*—short introductions of users on various subreddits—which, in the case of the MBTI-related subreddits, usually contain the users’ MBTI results. In the subsequent study (Gjurković et al., 2020), dataset was further enriched with demographic information about the users (age, gender, location, and language), and the labels for two other personality models. The distribution of MBTI types in this dataset also significantly deviated from the general US population (see Figure 3 in Section 6 for comparison of MBTI type distribution among different populations/datasets).

Automatic assignment of MBTI type to each user in all above-mentioned studies is based on automatic extraction of contexts in which a certain MBTI type is mentioned. Without manual inspection of each such mention—which was only reported for the study by Verhoeven et al. (2016)—the assigned labels might not be reliable, as they may refer to someone else mentioned in the tweet and not the writer of the tweet, or they might be a part of a larger phrase, e.g. “I think/believe I am an INTP” or “I expect to get ESFJ as the result if I do personality assessment”.

To the best of our knowledge, the only study in which MBTI labels were obtained by explicitly asking participants to report their MBTI type, if they had done an MBTI personality test in the past, is our recent study (Štajner and Yenikent, 2021). The Amazon Mechanical Turk workers were also asked to describe their favourite type of vacations and preferred hobbies in minimum 300 characters each. We found that this type of texts (responses to carefully selected open-end questions) contain more MBTI signals than tweets (even if concatenated together for each user). We further proposed detailed guidelines for MBTI personality annotation from textual data, and showed that expert human annotators have a high level of agreement among themselves on obtained textual answers when following provided guidelines. At the same time, we found that the annotators have a low level of agreement with the MBTI types reported by participants (based on their previous MBTI personality testing

The image shows a 'Demographics' form with the following sections:

- Gender \***: A dropdown menu with 'Female' selected.
- Age \***: A dropdown menu with '18-25' selected.
- English is: \*** (check all that apply):
  - My native language
  - The language I use in my everyday work/studies
  - None of the above
- The highest degree I have obtained: \***:
  - High school diploma
  - Bachelor's degree or equivalent
  - Master's degree or equivalent
  - Doctoral degree
  - Other: \_\_\_\_\_

Figure 1: Demographic questions.

via popular questionnaires), which might be an indication that MBTI results obtained via questionnaires do not resonate well with the MBTI signals found in more natural textual forms.

The current study aims to overcome previously reported issues by proposing four questions with the answers on a 1–5 scale to obtain MBTI labels that better resonate with the expert human MBTI annotations on short texts.

### 3 Questionnaire

The whole questionnaire consisted of one optional question “*You might have obtained your MBTI type in the past via questionnaires. If you know your MBTI type, please type it here*”, four compulsory demographic questions, four compulsory questions with answers on a 1–5 scale that aimed to capture the participants MBTI type, and two compulsory open-end questions. Demographic questions encompassed gender, age, whether or not English is their native language, and the highest level of education obtained (Figure 1). The gender question had four possible answers: *female*, *male*, *other*, *prefer not to specify*. Five age groups were offered to choose from: *18–25*, *26–35*, *36–45*, *46–55*, and *over 55*.

After answering demographic questions, partici-

For my free time, I ideally prefer to \*

1 2 3 4 5

Attend social activities and events where I can be surrounded by people, meet and talk to new people      Be alone and enjoy some quiet time by myself

---

I love hobbies that are \*

1 2 3 4 5

Technical and hands-on      Artistic and imaginative

---

My favorite hobbies are \*

1 2 3 4 5

Those that allow me to learn new things      Those that give me an escape from reality and let me get connected with my artistic side

---

For my holidays, I like \*

1 2 3 4 5

To be completely spontaneous without any pre-booked activities      To have everything organized and booked in advance

Figure 2: Questions for obtaining MBTI labels.

pants were provided with four questions that aimed to capture their MBTI type, and were asked to provide an answer on a 1–5 points scale. Those four questions are the central contribution of this study. By following the idea that aspects of leisure time represent the most natural version of personality, as it is directed by high degrees of intrinsic motivation (Štajner and Yenikent, 2021), the questions are focussed on typical leisure time activities—hobbies and vacations. This also gave us the opportunity to utilize the previously proposed open-end questions (Štajner and Yenikent, 2021) in the validation process (Section 5). In deciding the content of the questions for each individual dimension, we followed the main definitions provided by Briggs-Myers and Myers (1995). Although each MBTI dimension corresponds to multiple practical and behavioral characteristics, the core theoretical focus for every dimension is consistent.

The first question (for the E/I dimension) was designed with the idea of capturing whether the person prefers to be surrounded by people and social interactions, on one end of the scale (1 = extraverted), or to spend quiet and calm time by themselves, on the other end of the scale (5 = introverted). The second question (for the S/N dimension) aims to capture the characteristics of the tasks people would prefer to process, concrete or

intuitive, by asking whether they prefer technical and hands-on hobbies (1 = sensing) or abstract and imaginative (5 = intuitive). The third MBTI dimension (T/F) is fundamentally about how people make their decisions, whether based on rational or emotional motives. As people do not engage with strict decision-making processes during their free time, which is ultimately based on their personal interests, the question measured the preference for rational (1 = thinking) or emotional (5 = feeling) reasoning for liking a certain hobby. The fourth question aimed to capture the preference for spontaneous and flexible (1 = perceiving), or a well-planned (5 = judging) schedule at vacations.

We initially prepared two questions per each MBTI dimension and performed a pilot study with 30 participants to choose those questions (Figure 2) that better correspond to the MBTI types provided by the participants, and the MBTI annotations by two annotators.

Finally, participants were asked to answer to two open-end questions, which we previously proposed (Štajner and Yenikent, 2021) as the optimal questions for annotating MBTI types from texts:

- Describe which kind of vacations you typically enjoy and why.
- Describe what type of hobbies you enjoy and why.

The two questions were preceded by the following instructions: “*The following questions aim to understand your life style preferences. While answering, please write down the first things that come to your mind without much contemplation.*” To be accepted, each answer needed to contain a minimum of 300 characters.

#### 4 Challenges in Data Collection

Data was collected via Amazon Mechanical Turk (AMT) platform. We prepared the questionnaire as Google Forms and provided the link to it in the HIT of the AMT platform. We experimented with various setups in the platform: different values for monetary compensations, allowing only those participants with high scores on previous tasks, different times for validation of the answers and payment. The only variable that noticeably influenced the time needed for obtaining completed HITs was whether or not we restrict the participants according to their performance on the previous HITs. Without any restrictions, we were

obtaining approximately 50 completed HITs per hour. The main bottleneck in the whole procedure was the need for manually checking “honesty” of the answers to the open-end questions.

We manually checked all answers by particularly focusing on checking whether or not: (1) all personality questions contained the same answer (although theoretically possible that a person has the same answer to all four questions, it is very likely that this behaviour instead signifies that the AMT worker just wanted to finish the task as soon as possible and get the monetary compensation); (2) the answers to the open-end questions make sense, i.e. are not just a random sequence of 300 characters; (3) the answers to the open-end questions are copied from the internet; (4) the same worker has already completed the task, and if they did, we checked whether the answers to both multiple-choice and open-end questions were similar in both completed HITs. We found that approximately one third of the completed HITs contained answers that were copied from the internet. Those HITs were disregarded and those workers did not obtain monetary compensation to prevent them from doing it again. In those cases where more than one HIT was completed by the same worker, if the answers were similar, we paid the monetary compensation for all of them and maintained only one randomly chosen HIT for our dataset. In those cases where the answers in the HITs of the same worker had many significant differences, we paid the monetary compensation, but excluded all HITs from that worker from the dataset.

## 5 Validation

We performed two types of validation of our questionnaire. First, we calculated the agreement of the answers to our personality questions with the MBTI labels provided by the participants, for those cases where the MBTI label was provided (Section 5.1). Given that the MBTI labels obtained by using popular questionnaires might not be reliable for MBTI type detection from textual utterances (Štajner and Yenikent, 2021), the results obtained through this validation method should be taken with the grain of salt. Second, we validated our questionnaire via manual annotation of MBTI types on the answers to the open-end questions (Section 5.2).

### 5.1 Agreement with the MBTI Types

Given the wide popularity of the MBTI framework in non-research communities, we expected that a substantial number of AMT workers had taken the MBTI or 16Personalities tests before. In total, 340 participants responded to the optional MBTI question, with the following distribution of the MBTI traits: 87 extraversion / 253 introversion, 249 sensing / 91 intuition, 148 thinking / 192 feeling, and 206 judging / 134 perceiving.

We compared the MBTI types provided by AMT workers and the respective answers to our personality questions. To be able to compare them, we converted the 5-point scores into binary ones to match with the MBTI binary typology. For every dimension, the scores of 1 and 2 were merged as one polarity (e.g. extravert), and 4 and 5 were merged as the other polarity (e.g. introvert). The remaining 3s were considered as middle scores and were treated differently in two setups: (1) always counted as correct, regardless of the MBTI type provided by the AMT workers; (2) excluded from the analysis. In the second case, the total number of excluded cases per each dimension was: 79 (23.2%) for E/I; 76 (22.4%) for S/N; 67 (19.7%) for T/F; and 54 (15.9%) for J/P.

The middle scores either represent the cases in which participant equally exhibits characteristics of both polarities, or they indicate indecisiveness. Both types of participants could score as either of the two polarities in the MBTI questionnaires as those only offer two options. Therefore, the results of the first setup could be seen as an upper bound, and the results of the second setup a lower bound of the validity score. The exact result of the validation procedure cannot be calculated due to the limitation of the MBTI labels to capture middle cases which are common (Pittenger, 1993).

The percentage of cases in which the collected answers to our personality questions correspond to the MBTI type provided by the AMT workers is given in Table 1.

### 5.2 Agreement with Human Annotations

We asked two paid expert annotators, well-versed in MBTI framework from psychology perspective, to read 30 answers to the Vacations question and 30 answers to the Hobbies questions, and annotate each of them with one of the polarities (e.g. E or I), or the label MIDDLE if they see equal amount of signs for both polarities. The annotators

Scenario	E/I	S/N	T/F	J/P
Entire dataset (upper bound)	80.00%	64.71%	63.53%	69.91%
Reduced dataset (lower bound)	73.95%	54.55%	54.58%	63.99%

Table 1: Agreement with the provided MBTI types.

Instance	E/I	S/N	T/F	J/P
I like going on beach vacations to relax and get tan. I like to have drinks on the beach and eat a lot of yummy good food. I like vacations where my hardest choice is whether or not I want to spend my day either by the pool or on the beach. Then I go shower for the night and go to dinner in cute summer outfits.	I(4)	NA	NA	P(4)
ocean or sea side, somewhere chill so i can stare for hours into the calm waters while having a coffee in the morning or an alcoholic beverage at night. that place needs to be close to a big city where I can explore the local cuisine, but also explore the nature. Something like Hawaii or Miami where you have those secluded hotels on the ocean but then you get into a city and it's super lively.	Mid(3)	NA	NA	Mid(3)
Like to go for a trip to theme parks where i can make myself engaged and actively enjoy with the friends by playing various games and rides rather than sight seeing where we need to be enjoy alone or we can only interact with the nature. Also like to go to a vacation to new places and explore many places and their interest and values.	E(1)	NA	NA	J(1)
I enjoy playing video games because it gives me an outlet to forget about the world around me. when i can lock into the story and progress through making my own decisions i will forget about all the stress from work and just enjoy the fantasy world for hours on end. I also enjoy writing music for the same reasons. The escape from reality for a small amount of time is nice	NA	N(5)	F(5)	NA
My hobbies are a plethora of sorts. I like to read books that are either nonfiction or fiction. NF books allow me to learn something new that I don't learn in my typical coursework, and fiction books allow me to delve into a unique story. I also enjoy playing video games to compete and take on challenges.	NA	Mid(3)	Mid(3)	NA
I like to run, swim, travel, watch movies and series, because they are several activities where I develop and train at the same time, and I can enjoy life doing sports and studying to work not only the body, but also our minds. I love to study, run, travel and be on the side of the people we love and share those moments.	NA	S(1)	T(2)	NA

Table 2: Examples of agreements between the human annotators and the answers to our questions.

were provided with the previously proposed annotation guidelines (Štajner and Yenikent, 2021) and instructed to annotate the answers to the Vacations question only for the E/I and J/P dimensions, and the answers to the Hobbies questions only for the S/N and T/F dimensions.

The 30 answers per each open-end question were randomly selected with the constraint that 10 or those answers are from workers who chose 1 or 2 as the answers to the two respective personality questions, 10 answers are from workers who chose 3 as the answers to the two respective personality questions, and 10 answers are from workers who chose 4 or 5 as the answer to the two respective

personality questions.

We compared the labels obtained by human annotation (in a ternary classification task) with the labels obtained by transforming the answers to the personality questions into three classes: 1 and 2 to one polarity; 4 and 5 to the other polarity; and 3 to MIDDLE. Human annotations corresponded to the automatically obtained labels in 66.7-90.0% of the cases, depending on the annotator and the dimension in question. Several examples from this experiment are given in Table 2. Given the complexity of the task and the results of the agreement between the two trained annotators on a similar task (Štajner and Yenikent, 2021, Table 7), we find



Figure 3: Distributions of MBTI dimensions across different populations/datasets (presented in percentages of all respective users): general US population (Briggs Myers et al., 1998), Twitter users (Plank and Hovy, 2015), Reddit users (Gjurković et al., 2020), Amazon Mechanical Turk users who completed our HIT (MTURK-MBTI corresponds to those users who reported their MBTI type and is based on the MBTI type they provided, while MTURK-Q corresponds to all users and is based on the answers they chose on the scale 1–5).

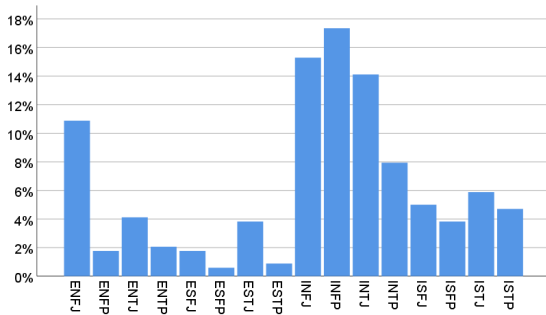


Figure 4: MBTI distribution (340 users).

these results satisfactory.

## 6 Dataset Statistics

The total number of users whose answers we collected via Amazon Mechanical Turk platform, after the manual quality check (Section 4) was 1038.

Distributions of labels for each dimension separately, for the whole dataset (MTURK-Q) and the portion that contained MBTI types entered by AMT workers (MTURK-MBTI) is provided in Figure 3

together with the distribution of labels in different populations/datasets mentioned in Section 2. As can be observed, the distributions of labels for the S/N dimension vary significantly across different datasets, with our dataset being the only one that follows a similar distribution as the one found in the general US population. For the E/I dimension, distribution of labels in all datasets deviate from the distribution in the general US population. As already mentioned in some of the previous studies (Plank and Hovy, 2015; Verhoeven et al., 2016; Štajner and Yenikent, 2020), this is not surprising, as it is known that introverts prefer online communication (Goby, 2006). For the T/F and J/P dimensions, all datasets roughly follow the distribution in the general US population.

The distribution of MBTI types across 340 AMT workers that provided their previously obtained MBTI type is presented in Figure 4. In our dataset, the three most frequent MBTI types are INFP, INFJ, and INTJ. It is interesting that those three MBTI types were also reported as the most frequent ones in the MBTI Twitter dataset (Plank

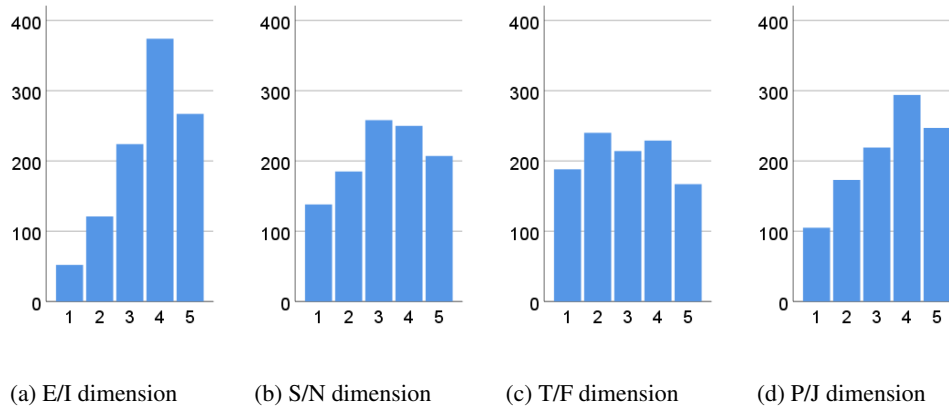


Figure 5: Distributions of answers on the 1–5 scale (y-axis represents the number of instances).

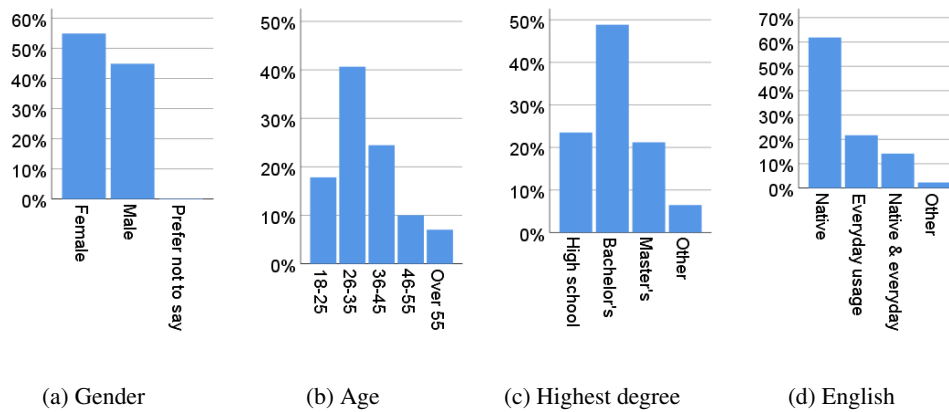


Figure 6: Distributions of the answers to the demographics questions.

and Hovy, 2015), while being reported among the least frequent ones in the general US population (Briggs Myers et al., 1998).<sup>8</sup>

Distributions of the answers to our personality questions and demographics questions are presented in Figures 5 and 6.

## 7 Conclusions

In this study, we proposed a set of four questions for quickly obtaining MBTI labels that better correspond to the expert human annotations of MBTI traits from short texts than the commonly used MBTI labels obtained via lengthy questionnaires. Apart from being faster to administer and allowing for obtaining large quantities of short texts on various topics, the proposed method also offers a more fine-grained MBTI typology, overcoming thus the common objections about the binary nature of MBTI questionnaires. This is particularly important for advancing research efforts in automatic MBTI personality detection from texts, as

<sup>8</sup>The statistics from the MBTI manual are also available at: <https://www.careerplanner.com/MB2/TypeInPopulation.cfm>.

those instances that come from people that exhibit equal preferences to both polarities could be filtered out from the training data, thus lowering the noise in the models. Finally, it seems that proposed way of compiling MBTI dataset via Amazon Mechanical Turk platform leads to obtaining labels with distributions closer to those of the general US population.

## Acknowledgements

We would like to thank Yassine Benajiba and Francisco Rangel for supporting this work.

## References

- Gregory J. Boyle. 1995. Myers-Briggs type indicator (MBTI): Some psychometric limitations. *Australian Psychologist*, 30(1).
- Isabel Briggs Myers, Mary H. McCaulley, Naomi L. Quenk, and Allen L. Hammer. 1998. *MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*. Consulting Psychologists Press.



- Isabel Briggs-Myers and Peter B. Myers. 1995. *Gifts differing: Understanding personality type*. Davies-Black Publishing.
- Fabio Celli and Bruno Lepri. 2018. [Is Big Five Better than MBTI? A Personality Computing Challenge Using Twitter Data](#). In *CLiC-it*.
- Matej Gjurković, Mladen Karan, Iva Vukojević, Mihaela Bošnjak, and Jan Šnajder. 2020. [PAN-DORA talks: Personality and demographics on reddit](#). *CoRR*, abs/2004.04460.
- Matej Gjurković and Jan Šnajder. 2018. [Reddit: A gold mine for personality prediction](#). In *Proceedings of the Second Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media*, pages 87–97, New Orleans, Louisiana, USA. Association for Computational Linguistics.
- Valerie Priscilla Goby. 2006. Personality and online/offline choices: MBTI profiles and favored communication modes in a Singapore study. *CyberPsychology Behavior*, 9:5–13.
- Steven J. Heine, Darrin. R. Lehman, Kaiping Peng, and Joe Greenholtz. 2002. [What’s wrong with cross-cultural comparisons of subjective likert scales?: The reference-group effect](#). *Journal of Personality and Social Psychology*, 82(6):903—918.
- Carl G. Jung. 1921. *Psychological Types: Volume 6*. Routledge.
- Ivar Krumpal. 2011. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality Quantity*, 47(4).
- David J. Pittenger. 1993. Measuring the MBTI... and coming up short. *Journal of Career Planning and Employment*, 54:48–52.
- Barbara Plank and Dirk Hovy. 2015. 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*, pages 92–98, Lisbon, Portugal. Association for Computational Linguistics.
- Ben Verhoeven, Walter Daelemans, and Barbara Plank. 2016. [Twisty: a multilingual twitter stylometry corpus for gender and personality profiling](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1632–1637, Portoroz, Slovenia. European Language Resources Association (ELRA).
- Sanja Štajner and Seren Yenikent. 2020. [A survey of automatic personality detection from texts](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6284–6295, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Sanja Štajner and Seren Yenikent. 2021. [Why is MBTI personality detection from texts a difficult task?](#) In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3580–3589, Online. Association for Computational Linguistics.
- Kosuke Yamada, Ryohei Sasano, and Koichi Takeda. 2019. [Incorporating textual information on user behavior for personality prediction](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 177–182, Florence, Italy. Association for Computational Linguistics.