Just the Way You Chat: Linking Personality, Style and Recognizability in Chats

Giorgio Roffo[1], Cinzia Giorgetta[2], Roberta Ferrario[2], and Marco Cristani[1]

[1] Università degli Studi di Verona, Strada Le Grazie 15, I-37134 Verona, Italy
[2] ISTC–CNR, via alla Cascata 56/C, I-38123 Povo (Trento), Italy

Abstract. Text chatting represents a hybrid type of communication, where textual information is delivered following turn-taking dynamics, which characterize spoken interactions. It is interesting to understand whether special interactional behavior can emerge in chats, similarly as it does in face-to-face exchanges. In this work, we focus on the writing style of individuals, analyzing how it can be recognized given a portion of chat, and how personality comes into play in this scenario. Two interesting facts do emerge: 1) some traits correlate significantly with some characteristics of people's chatting style, captured by stylometric features; 2) some of such features are very effective in recognizing a person among a gallery of diverse individuals. This seems to suggest that some personality traits could lead people to chat with a particular style, which turns out to be very recognizable. For example, motor impulsiveness gives a significative (negative) correlation with the use of the suspension points (...), that is also one of the most discriminative characteristics in chats. This and other relations emerge on a dataset on 45 subjects, monitored for 3 months, whose personality traits have been analyzed through self-administered questionnaires. What turns out is that chatting seems to be more than just typing.

Keywords: chat analysis, personality traits, authorship attribution.

1 Introduction

It is widely-known that written text (books, newspapers) conveys a great deal of information about the writer, relying on features that are not exhausted by those connected to semantic content; personality and identity traits can be inferred from text, through computational approaches, with a high accuracy [1–5]. But "written text" is not only books and newspapers: with the diffusion of the Internet, new kinds of media have arisen, such as online texts and electronic messages. Among these new media, text chats represent a social phenomenon: in 2009, 47 billion of instant messages have been delivered on a daily basis, with 1 billion of users worldwide. Under the perspective of pragmatics and social signal processing, chats are intriguing entities, representing crossbreeds of literary text and spoken conversations, due to the turn-taking dynamics with which text is delivered. In this respect, it is interesting to verify the presence of nonverbal cues

H.S. Park et al. (Eds.): HBU 2014, LNCS 8749, pp. 30–41, 2014.

in chats. Nonverbal signals enrich the spoken conversation by characterizing how sentences are uttered by a speaker [6], forging a unique style that characterizes the latter among many other subjects. In the same vein, they express personal beliefs, thoughts, emotions and personality. In this work we analyze how these insights may be applied to text chats, extracting features dubbed as "stylometric", since they codify the style of an author. Recently, specific features have been shown to finely recognize an author through his/her style. In this paper, we continue the work, checking whether style is effectively linked with the personality or other psychological or interactional traits of the speaker; this allows to close the loop, exploring how personality can make speakers recognizable by the uniqueness of their style.

To this sake, we consider a novel dataset of chats in italian language. The dataset contains data on 45 subjects, related to their chatting activity with single individuals for a time lapse of three months on average. In order to analyze the chatting style of a subject, we take into consideration 20 hybrid stylometric features recently proposed in [7]; to gather information on the personal characteristics of each individual, we use two well-known self-administered questionnaires, the former focused on 3 different impulsiveness factors (attentional, motor and non-planning impulsiveness), and the latter on 4 different psychological factors involved in human interactions (ability on taking others' perspective, empathic concern, fantasy scale and personal distress in interpersonal settings).

Two interesting results emerge: first, 5 psychological factors seem to correlate with 7 out of 20 stylometric features, in a statistically significant way ($p\text{-}value < 0.05$); for example, our data showed that subjects with higher lack of attention and cognitive instability were slower in providing their answers.

The second result follows by applying person recognition using stylometric features. The idea is to take a portion of a chat (10 turns of an individual) and to guess his/her identity, comparing against a gallery of 45 individuals. For this sake, we analyze each feature independently, capturing their appropriateness in distinguishing one subject from the other. Thus, we can single out highly expressive features (capable of recognizing a person with some accuracy) which are possibly correlated with a given personality trait expressed by a psychological factor. These findings seem to suggest that there are some personality traits that lead people to chat with a particular style, which turns out to be very recognizable. For example, the use of the suspension points (. . .) is a discriminative characteristic in chats, and shows a significant (negative) correlation with motor impulsiveness (the tendency to act on the spur of the moment).

The rest of the paper is organized as follows: in Section 2 we present the literature review, focusing on the computational approaches of chat analysis. In Section 3 we present our dataset, illustrating our analysis. In the same section our novel stylometric features are presented and the results of correlation between personality traits and stylometric features are reported in Section 3.3, while the results on identity recognition are presented in Section 3.4. Finally, conclusions and future perspectives are presented in Section 4.

2 Related Work

Although some computational approaches that try to infer personality traits from text are present in the literature, all these are focused on a semantic analysis of the content [2, 8]; on the contrary, in this work a semantic analysis of the text is, for privacy issues, absent. More importantly, none of the mentioned approaches have been applied to chats. Concerning the author recognition issue, the most related field of studies is that of Authorship Attribution (AA), that has as main aim to automatically recognize the author of a given text sample, based on the analysis of *stylometric* cues (see Table 1).

Typically, state-of-the-art approaches extract stylometric features from data and use discriminative classifiers to identify the author (each author corresponds to a class). The application of AA to chat conversations is recent (see [4] for a survey), with [1, 5, 9] the most cited works. In [5], a framework for authorship attribution of online messages has been developed to address the identity-tracing problem. Stylometric features were fed into SVM and neural networks on 20 subjects, validating the recognition accuracy on 30 random messages. PCA-like projection has been applied in [1] for authorship identification and similarity detection on 100 potential authors of e-mails, instant messages, feedback comments and program code. A unified data mining approach has been presented in [10] to address the challenges of authorship attribution in anonymous online textual communication (email, blog, IM) for the purpose of cybercrime investigation.

In the last ten years, authorship attribution and forensic analysis have extended their research to IM communication [11]. In [3], 4 authors of IM conversations have been identified based on the sentence structure and their use of special characters, emoticons, and abbreviations. Most recently, ad-hoc features for analyzing chats and perform author identification and verification have been presented in [7], dubbed as *turn-taking based* features (see Table 1).

In our work, we examine chats among pairs of people. These conversations can be considered as sequences of *turns*, where each *turn* is a set of symbols consecutively typed by one subject without being interrupted by the other person. Each conversation ends when no turn does occur for at least 30 minutes. In addition, each turn is composed by one or more *sentences*: a sentence is a stream of symbols which is ended by a "return" character. Each sentence is labeled by a temporal ID, reporting the time of delivery.

In the study each character constituting a word has been substituted by an X symbol, so that the content of the conversation is disregarded, but other features, such as length of words, punctuation, emoticons etc. are preserved and analyzed. Other elements that are left opaque are the relationship the speakers entertain with their interlocutors and the psychological character of the latter. Nonetheless, independently from this, we are able to re-identify the speaker(s) through their chat style.

Table 1. Synopsis of the 20 features used in our work. In squared parenthesis, the numeric range of values taken by each feature in this dataset.

Group	Subgroup	***Name*, description and [range]**	References
Lexical	Word level	***#Words (=W)***: number of words per turn - [0,253]; ***Avg. word length***: average word length in a turn - [0,443]	[1, 3–5, 10]
	Character level	***#Chars (=C)***: number of characters in a turn - [0,1624]; ***#Uppercase letters***: number of uppercase characters - [0,05]; ***#Uppercase/C***: number of uppercase characters divided by C - [0,1]	[1, 3–5]
Syntactic	Punctuation	***# ? and ! marks***: number of question and exclamation marks summed together, in a turn - [0,4]; ***#Three points (...)***: number of occurrences of three points (...) in a turn - [0,15]; ***#Marks (",.:*;)***: number of marks (",.:*;) in a turn - [0,119]	[1, 3–5, 10]
	Emoticons	***#Emoticons***: total number of Skype emoticons in a turn - [0,16]; ***#Emoticons/C***: number of emoticons divided by C - [0,1]; ***# emoticons/W***: number of emoticons divided by W - [0,4]; emoticons categories such as ***#Pos. emo.***, that counts the occurrences of happiness, love, intimacy, etc. icons (20 emot. types in total) - [0,4]; ***#Neg. emo.***: address fear, anger, etc. (19 emot. types in total) - [0,3]; and ***#Oth. emo.***, neutral emoticons portray actions, objects etc. (62 emot. types in total) - [0,16]	[3, 4, 7]
Turn-taking	Temporal (in seconds)	***Turn duration***: time spent to complete a turn - [0,1800]; ***Answer time***: time spent to answer a question expressed in the previous turn of the other interlocutor - [0, 1784]	[7]
	Tempo/lexical	***Char/Word writing speed***: number of typed characters or words per second - [0,350] and [0,67];	[7]
	Lexical	***Imitation rate/C, Imitation rate/W***: ratio between number of chars -or words- in current turn and number of chars -or words- in previous turn of the interlocutor - [0,1] and [0,1]	[7]

From each turn, all stylometric features are extracted, generating a number each; therefore, for each conversation, we obtain T feature values. The description of the employed features is presented in Tab. 1: for the sake of brevity, we did an initial feature selection, pruning away features producing less than 50% of nAUC (see next), coming out with 20 cues left. The features are nonverbal, that is, they ignore the content of the chat; in this sense, they are comparable to nonverbal signals in spoken conversation (nodding, speaking louder etc.).

3 Our Analysis

In this section we illustrate our analysis, discussing the results.

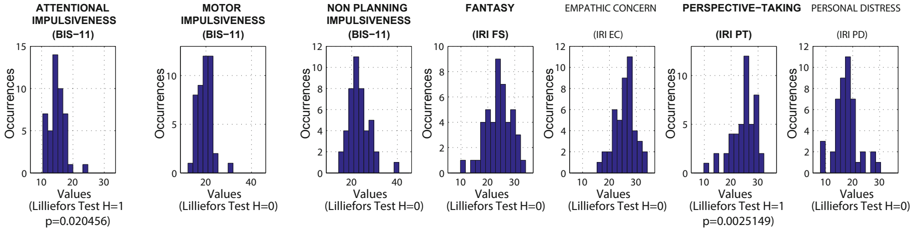

Fig. 1. Traits distribution. On the bottom, the output of the Lilliefors test for distribution normality. In bold, those personality traits which correlate with at least one stylometric feature.

3.1 The Dataset

Our dataset focuses on 45 subjects[1] (Master and PhD students, Post-doc associates and various professionals; 29 male, 16 female; average age: 30,25 years), involved in dyadic, spontaneous conversations; moreover, the chats have been performed before the beginning of the study, when the subjects were not aware that they would have been used for such purpose. This ensures that the behavior of the subjects is natural and no attempt has been made to modify the style in any sense. From the raw conversations, we extract the features discussed above. The number of turns per subject ranges between 20 and 100. Concerning the personality traits, participants were asked to fill the following two well-known self-administered questionnaires, aimed at evaluating psychological factors:

1. The Barratt Impulsiveness Scale, version 11 (Bis-11) [13], to measure the levels of impulsiveness, based on three different sub-scales: "attentional impulsiveness", indicating a lack of attention and cognitive instability, "motor impulsiveness", indicating a lack of control in motor behavior, and "non-planning impulsiveness", indicating a deficit in planning their own behavior. The total number of items is 30, answered on a 4-points scale, ranging from 'never' to 'always'.
2. Interpersonal Reactivity Index (IRI) [14], to test their ability to take others' perspective (PT) and empathic concerns (EC), to get caught up in fictional stories and imagine oneself in the same situations as fictional characters (FS), and finally to test their "self-oriented" feelings of personal anxiety in interpersonal settings (PD). This questionnaire presents 28 items answered on a 5-point Likert scale, ranging from 'does not describe me well' to 'describes me very well'.

In Fig. 1, we report the traits statistics.

[1] The number of participants in these experiments is relatively small, and this can be seen as a limitation, as it may partially restrict the generalization of the results. However, the present study is in line with the sample size of other studies in the psychological field, wherein, in within-subjects design, the number of participants is typically between 20 (e.g., for studies in neuroscience field) and 40 (e.g., for behavioral studies). Also, in other more related studies, such as [12], in which the role of personality traits in human-robot vs. human-human interaction has been investigated, the sample size is even lower ($N = 28$).

3.2 The Stylometric Features

Thanks to our framework, we were able to design a set of stylometric features which follow the idea of [7], that is, of using the turn length, and not simply the entire conversation, as fundamental entity for computing stylometric statistics.

We have paid much attention to privacy issues: the idea is to neglect the content of the conversation, accounting only for the way in which it is performed. Our features follow this guideline, avoiding any kind of natural language processing, while other features, as length of words, punctuation, emoticons etc. are preserved and analyzed.

For each person involved in a conversation, we analyze his/her stream of turns (suppose T), ignoring the semantic input from the other subject. This means that we assume that the chat style (as modeled by our features) is invariant through different interlocutors, with whom the subject entertain different kinds of relationships.

From each turn, a stylometric feature is extracted, generating a number; therefore, with T turns we obtain T feature values. Depending on the kind of feature and task to perform (measuring correlations or doing person identification), an histogram or the mean/median is computed, and the resulting measure becomes a part of the signature which characterizes a given individual. In the following, the list of the features together with their explanation is presented.

For the sake of clarity, our features will be presented following the taxonomy analyzed in the state of the art (see Tab. 1). For convenience, and whereas possible, we have kept the name of the features proposed in the literature.

How the features have been treated to calculate correlations with personality traits or similarities enabling recognition, will be discussed in the following. In the list below, the numbers in parenthesis indicate the feature ID.

Lexical Features

(1) **Number of words (#Words)**: number of words per turn. With "word" we intend a string of *characters* (see below);

(2) **Number of chars (#Chars)**: number of characters per turn. With "character", we intend every normal key on the XXX keyboard[2], ignoring special keys like the SPACE, CTRL, etc.;

(3) **Number of uppercase chars (#Uppercase Letters)**: number of uppercase characters in a turn;

(4) **Number of uppercase chars / number of chars (#Uppercase/#C)**: usually, entire words written in capital letters indicate a strong emotional message. This feature accounts for such communicative tendency;

(5) **Mean word length (Avg. word length)**: average length of the words in a turn;

(6-7) **1(2)-order length transitions (1oLT, 2oLT)**: these features resemble the n-grams of [1]; the strong difference here is in the fact that we consider

[2] The kind of keyboard has been removed for anonymity reasons.

solely the length of the words, and not their content. In practice, for a noLT of order $n = 1$ (1oLT), we build counting matrices that in the entry i, j, $1 \leq i, j \leq I$, exhibit the number of times we move from a word of length i to a word of length j. The count matrices are then normalized, so that they can be seen as transition matrices of a Markov chain. In our case, we set $I = 15$. 2oLT are modeled by 3D transition matrices of size $I \times I \times I$. We do not take into account superior orders, for sparsity issues (the resulting hypervolumes will have many entries at 0). These features are collected over whole conversations, rather than on single turns.

Syntactic Features

(8) **Number of ? and ! marks (#? & ! marks)**: we keep the "?" and the "!" marks in the same feature, since taken separately their relevance is very low. In practice this feature highlights the fact that a conversation is lively and conveys a great deal of (positive or negative) emotions. It also shows the will of the subject of generating a reaction in the interlocutor;

(9) **Number of suspension points (#Three points (...))**: this feature may show the uncertainty of the subject with respect to the sentence just written, or it may suggest to the interlocutor some unspoken consequent of what has been declared explicitly in the previous sentence;

(10) **Number of generic marks (#Marks (",.:*;))**: a high number of generic marks (",.:*;) usually indicates a more accurate writing style. In particular, they are mostly used to give structure to the sentences and articulate a discourse. This can either be a peculiar characteristic of the subject, or it may hint to the fact that the subject is very determined with respect to what he/she is saying and thus provides it with a structure to make explicit that it is the result of an elaborated thought;

(11,12,13) **Number of *positive*, *negative* and *uncategorized* emoticons (Pos. emo, Neg. emo, Oth. emo**, respectively**)**: features related to emoticons aim at individuating a particular mood expressed in a turn. In particular, 101 diverse emoticons have been divided in three classes, portraying positive emotions (happiness, love, intimacy, etc. – 20 emot.), negative emotions (fear, anger, etc. – 19 emot.) and other emoticons (portraying actions, objects etc. – 62 emot.);

(14) **Number of emoticons (#Emoticons)**: Number of emoticons in a turn, independently from their type;

(15-16) **Number of emoticons / number of words (chars) (#Emoticons/#W, #Emoticons/#C**, respectively**)**: it considers how often we inject pictorial symbols in a sentence considering the number of words (chars) typed; it measures how much the subject willingly shows his/her emotions, attitudes, intended humor etc.

Turn-taking Features

(17) **Turn duration**: We indicate with T the length of the period in which a turn is kept (before pressing the "return" key);

(18-19) **Word writing speed**: it measures the turn duration divided by the number $\#Words$ or $\#Chars$ of written words/chars in a turn;

(20-21) **Imitation per word, per char (Imitation rate / W, Imitation rate / C)**: ratio between number of chars -or words- in the current turn and number of chars -or words- in the previous turn of the interlocutor; this feature models the tendency of a subject to imitate the conversation style of the interlocutor (at least for what concerns the length of the turns). The imitation feature accounts for some interactional attitudes of the subjects;

(22) **Answer time**: this feature is the time spent to answer a question presented in the previous turn of the interlocutor. We assume the presence of a question whenever there is a question mark.

Since these features are collected for each turn (except the 1oLT and 2oLT features), and assuming there are T turns in a conversation, we end up with T numbers for each feature.

These features were subsequently used for two applications: correlation with psychological traits and subject recognition. Depending on the task at hand, the feature values have been treated differently, as discussed in the following.

3.3 Correlations between Traits and Features

In order to discover a connection between psychological traits and features, we calculate the Pearson correlation coefficient (whereas both the distribution of features values and traits were normal), the Spearman coefficient otherwise. The test for normality is the Lilliefors test. Results are shown in Table 2, showing 7 significant correlations (p-value< 0.05)[3].

Our findings seem to suggest that subjects who score higher in attentional impulsiveness, then with higher lack of attention and cognitive instability, are also slower in providing answers to their partners, whereas those higher in motor impulsiveness use a lower number of suspension points, indicating they are loosing less time (before making a question or providing an answer) to think about what they have to write. Also, subjects with deficit in planning their own behavior (non-planning impulsiveness) seem to use a lower number of emoticons expressing positive emotions. In the same direction, those higher in attentional impulsiveness seem to use a higher number of emoticons expressing negative emotions. Relatively to traits involved in interpersonal interactions, our data show that subjects higher in fantasy scale (FS) seem to spend less time in typing words and single letters, showing they are faster in imaging what they can

[3] Although the r values of these correlations are around .30, and thus indicating a not really strong correlation, these also show a p value which is lower than .05 and thus significant and worth to be taken into consideration.

Table 2. Correlations between psychological traits and stylometric features. In parenthesis, the nAUC score, witnessing how effective is the feature in distinguishing people (the higher the more effective, see Section 3.4). Numerical values in the table indicate statistically significant correlations (*p-value*< 0.05).

	Attentional impulsiveness	Motor impulsiveness	Non planning	Fantasy scale	Perspective taking
#Three points (...) (*57.5*)	/	-0.32	/	/	/
#Uppercase/C (*59.3*)	/	/	/	-0.32	-0.35
Word writing speed (*61.3*)	/	/	/	-0.33	/
Char writing speed (*63.3*)	/	/	/	-0.32	/
Answer time (*51.2*)	0.30	/	/	/	/
#Pos. Emo. (*62*)	/	/	-0.29	/	/
#Neg. Emo. (*54.5*)	0.33	/	/	/	/

write and then in doing that. Instead, subjects with lower scores in FS, as well as those with lower ability in taking others' perspective (PT), seem to use a higher number of upper cases, as if they need to outline what they are writing.

3.4 The Recognition Approach

At this point, we want to understand the appropriateness of stylometric features in the task of recognizing a person given a portion of his/her chat. Let us suppose to have collected the features related to the conversations for two subjects, A and B (one conversation each). We now have to exploit them for obtaining a single distance, describing the overall similarity between A and B. As a first step, we derive a plausible distance for each feature separately: since we want to extract as much information as possible from the conversation, given T values of a feature, we organize them into an 8-bin histogram, where the range of the quantization values has been fixed by considering all the samples of the gallery. To match the two conversations, we thus employ the Bhattacharyya distance. The turns of each subject are split into *probe* and *gallery* set, where the probe samples serve as test, and are given to the recognizer, which evaluates the matches with the gallery elements. Probe and gallery set include 10 turns each: in this way, any bias due to differences in the amount of available material should be avoided. When possible, we pick different turns selections (maintaining their chronological order) in order to generate different probe/gallery partitions. In this sense, for each feature we repeat the re-identification 30 times, varying probe and gallery partitions. A particular feature of a single subject is extracted from the probe set, and matched against the corresponding gallery of features of all subjects. This happens for all the N subjects, resulting in a $N \times N$ distance matrix. Ranking in ascending order the N distances for each probe element allows one to compute the *Cumulative Match Characteristic* (CMC) curve, i.e., the expectation of finding the correct match in the top n positions of the ranking. The CMC is an effective performance measure for authorship attribution approaches [15], and in our case is a valid measure for evaluating the task of *identity recognition*: given a test sample, we want to discover the identity among a set of N subjects. In particular, the value

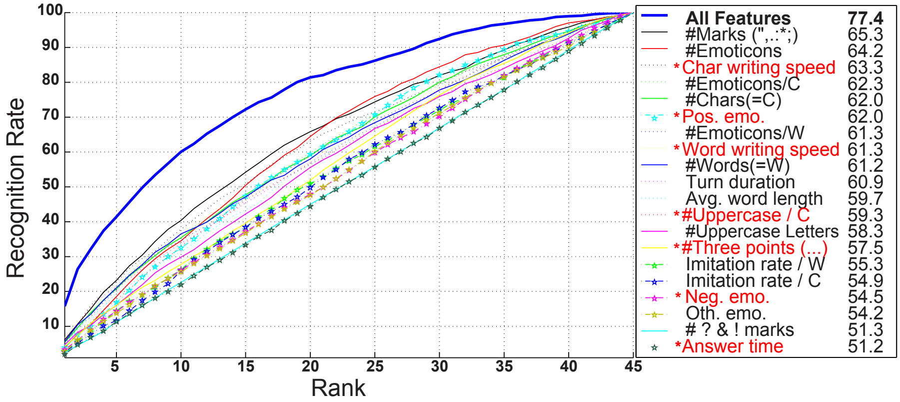

Fig. 2. CMC curves for each feature. After each feature, the value of the correspondent nAUC. With an asterisk, in red, we specify those features which correlate at least with a personality trait. The *All features* CMC indicates the performance of averaging the distance of all the features and calculating the ranking.

of the CMC curve at position 1 is the probability that the probe ID signature of a subject is closer to the gallery ID signature of the same subject than to any other gallery ID signature; the value of the CMC curve at position n is the probability of finding the correct match in the first n ranked positions.

Given the CMC curve for each feature (obtained by averaging on 30 trials), the normalized Area Under Curve (nAUC) is calculated as a measure of accuracy. The results are shown in Fig. 2, where the features are listed in decreasing order of accuracy. In red are portrayed those cues which correlate with at least one feature: for the sake of clarity, the nAUC value is reported also in Table 2. The Table seems to suggest an important fact: fantasy scale is the trait which is tightly related with the style of a person, making it very recognizable, since it affects the speed with which a person writes, and his/her usage of the uppercase letters. Once again, these data confirm that people with a higher ability in being caught up in fictional stories and in imaging themselves in the same situations of fictional characters are faster in imaging and then writing what they can say or answer to the person they are communicating with.

Another interesting observation can be assessed by looking at Fig. 2; all the CMC curves related to the different features are similar in expressivity, but not strongly effective; the probability of guessing the right people with only one attempt (corresponding to analyzing the performance of the CMC curve at rank 1) is below the 10%. Anyway, if we combine all features, mediating the related distances computed among the probe and the gallery subject, we obtain a much more informative curve (see Fig. 2, *All features*). In this case, getting the correct guess after the first attempt is above the 10%, and with the 60% of probability we can individuate the correct subject among the first 10 ranked subjects. This witnesses that the features model diverse and complementary stylistic aspects of a chat text; viceversa, except the #Uppercase/C stylometric feature, each feature seems to be connected with only one factor related with personality traits investigated here. These findings are very interesting as, congruently with the aim of this study, they clearly suggest that specific psychological factors related with impulsivity and involved in human interaction can be predictive of peculiar writing styles people use in the chat text. Finally, despite the recognition scores may appear low, one has to keep in mind that they are related to a soft biometric trait, that is, taken without the explicit cohoperation of the user [16]. In such a scenario, recognition performances are much lower than (hard) biometric traits (iris, fingerprint), and in line with the results presented.

4 Conclusions

The present study suggests that chatting via text is a very rich form of conversation; contrarily to what assumed until few years ago [17], chats allow to exchange more than simply verbal information: in particular, we have analyzed 7 personality traits, showing a significant correlation of 5 of them with 7 stylometric features. At the same time, these stylometric features turn out to be very helpful in discriminating one subject from the others, considering re-identification metrics. Putting together these two facts, what seems to emerge is that there are personality traits that could lead one to chat in a particular manner, which turns out to be very recognizable. This study constitutes an encouragement in pursuing this research direction, and should be considered a exploratory attempt, since many are the necessary improvements. First of all, the dataset should be enlarged for a more robust statistics support. Therefore, regression approaches will be needed for predicting the personality traits of an individual, given his/her chats. At the present moment, the data is not enough to support the training of a classifier, and the preliminary prediction results (done with Support Vector regression, under a Leave-One-Out cross-validation scheme) are not satisfactory. From a psychological perspective, a further step would be to investigate how these personality traits affect conversations in text chats focusing on more specific interactions, for example with co-workers, friends or unknown people. These specific interactions in particular can be decisive in determining whether it is possible to recognize the personality traits by written conversations, independently from the kind of relationships with the interlocutor and interactional setting. As for the application, this study paves the way for multimodal interfaces capable of recognizing the identity and/or the personality traits of a person, recommending for example specific typologies of interlocutors whom he/she would be more comfortable to talk with.

References

1. Abbasi, A., Chen, H., Nunamaker, J.F.: Stylometric identification in electronic markets: Scalability and robustness. Journal of Management Information Systems (JMIS) 25(1), 49–78 (2008)

2. Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research (JAIR) 30, 457–500 (2007)
3. Orebaugh, A., Allnutt, J.: Classification of instant messaging communications for forensics analysis. Social Networks, 22–28 (2009)
4. Stamatatos, E.: A survey of modern authorship attribution methods. Journal of the Association for Information Science and Technology (JASIST) 60(3), 538–556 (2009)
5. Zheng, R., Li, J., Chen, H., Huang, Z.: A framework for authorship identification of online messages: Writing-style features and classification techniques. Journal of the Association for Information Science and Technology (JASIST) 57(3), 378–393 (2006)
6. Knapp, M.L.: Nonverbal communication in human interaction, 8th edn., January 1. Cengage Learning (2013)
7. Roffo, G., Segalin, C., Vinciarelli, A., Murino, V., Cristani, M.: Reading between the turns: Statistical modeling for identity recognition and verification in chats. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2013 (2013)
8. Pennebaker, J.W., King, L.A.: Linguistic styles: language use as an individual difference. Journal of Personality and Social Psychology 77(6), 1296–1312 (1999)
9. Abbasi, A., Chen, H.: Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. ACM TOIS 26(2), 1–29 (2008)
10. Iqbal, F., Binsalleeh, H., Fung, B.C.M., Debbabi, M.: A unified data mining solution for authorship analysis in anonymous textual communications. Information Sciences (2011)
11. Gajadhar, J., Green, J.: Open Polytechnic of New Zealand, Open Polytechnic of New Zealand Staff. An Analysis of Nonverbal Communication in an Online Chat Group, Working papers, Open Polytechnic of New Zealand (2003)
12. Walters, M.L., Dautenhahn, K., Boekhorst, R., Koay, K.L., Kaouri, C., Woods, S., Nehaniv, C., Lee, D., Werry, I.: The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In: International Workshop on Robot and Human Interactive Communication, RO-MAN 2005, August 13-15, pp. 347–352 (2005)
13. Patton, J.H., Stanford, M.S., Barratt, E.S.: Factor structure of the barratt impulsiveness scale. Journal of Clinical Psychology 51, 768–774 (1995)
14. Davis, M.H.: A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology 10, 85–104 (1980)
15. Bolle, R., Connell, J., Pankanti, S., Ratha, N., Senior, A.: Guide to Biometrics. Springer (2003)
16. Dantcheva, A., Velardo, C., D'angelo, A., Dugelay, J.-L.: Bag of soft biometrics for person identification: New trends and challenges. Mutimedia Tools and Applications (October 2010)
17. Markey, P.M., Wells, S.M.: Interpersonal perception in internet chat rooms. Journal of Research in Personality 36(2), 134–146 (2002)