2013 European Intelligence and Security Informatics Conference

The Effect of Personality Type on Deceptive Communication Style

Tommaso Fornaciari
CIMeC - University of Trento
Corso Bettini 31, Rovereto, Italy
Email: tommaso.fornaciari@gmail.com

Fabio Celli
CIMeC - University of Trento
Corso Bettini 31, Rovereto, Italy
Email: fabio.celli@unitn.it

Massimo Poesio
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, UK
Email: massimo.poesio@essex.ac.uk

***Abstract*—It has long been hypothesized that the ability to deceive depends on personality - some personality types are 'better' at deceiving in that their deception is harder to recognize. In this work, we evaluate how the pattern of personality of a speaker affects the effectiveness of machine learning models for deception detection in transcripts of oral speech. We trained models to classify as deceptive or not deceptive statements issued in Court by Italian speakers. We then used a system for automatic personality recognition to generate hypotheses about the personality of these speakers, and we clustered the subjects on the basis of their personality traits. It turned out that deception detection models perform differently depending on the patterns of personality traits which characterize the speakers. This suggests that speakers who show certain types of personality also have a communication style in which deception can be detected more, or less, easily.**

I. INTRODUCTION

Personality recognition and deception detection have been widely explored and studied using a number of different approaches [42], [30]. Nonetheless, it is only recently that these tasks have been approached using stylometric techniques, that is, making use of computational methods based on the stylistic features of written or spoken speech samples [32] [27]. In this perspective, while deception detection is the task of recognizing truth and deception in discourse and text, personality recognition from text is the task of classifying personality traits of authors, given fragments of text they wrote.

In the area of deception detection, Newman et al. 2003 [32] showed for the first time that machine learning techniques relying on sets of features automatically extracted from texts can be effective in the identification of deception. This outcome was confirmed in subsequent studies such as Strapparava & Mihalcea 2009 [41]. All of these studies however were based on data collected in laboratory, i.e., in settings where the psychological conditions of the subjects were consistently different from those in natural environment. This is because to study deception in high stakes scenarios is particularly difficult, for practical and ethical reasons [43] [15]. However, some recent studies [21] suggested that the same techniques can be effective even when applied to data collected on the field.

In psychology, personality is seen as an affect processing system that describes persistent human behavioural responses to broad classes of environmental stimuli [1], including communicative styles [7]. These differences appear to include the style used in deception: e.g., Enos et al. [14] found that several personality factors appear to correlate with the ability of a judge to detect deception in speech, and that humans perform worse than machines in the detection of deception [42].

In this work, we address the issue of the relationships between personality and the tendency towards lie and deception. Recent advances in automatic personality recognition from text [29], [7] allow us to extract personality types of subjects from the transcriptions of Court hearings, although the context and the interaction of the subject with others affect their communicative style. We exploited a system for automatic personality recognition from text available online[1] [7] in order to assess the personality of 31 Italian speakers, by the analysis of statements they issued in Court. These statements are part of DECOUR, a corpus of transcripts of hearings held in Italian Courts [20]. In particular, the hearings regarded criminal proceedings for calumny and false testimony, where the defendants were found guilty. Therefore, thanks to the information provided by the judgments of the Courts, every utterance issued by the speakers was annotated as true, false or uncertain, and DECOUR was employed to train models aimed to classify the utterances according to their degree of truthfulness [18] [19]. We carried out two text classification experiments using DECOUR, in which we evaluated the performance of the models considering each speaker and his personality traits.

The structure of the paper is as follows. Section II introduces some background regarding stylometry and personality recognition. Section III describes the dataset. In Section IV the results of the experiments are presented and discussed, followed by conclusions in section V.

II. STYLOMETRY AND PERSONALITY RECOGNITION

Stylometry is a discipline which analyses texts relying on their stylistic features only. Modern stylometry makes use of computational methods for automatic extraction of low-level linguistic cues from texts, and of machine learning techniques for their evaluation. These analyses have proven effective in several tasks, such as author profiling [13] [39], deducing age, sex and native language of authors of written texts [24], [34], author attribution [31], [28] and plagiarism analysis [40]. In these fields, researchers put a great deal of effort into reaching the best results. Also in deception detection, stylometric techniques were found reasonably successful [41], [15], [21].

Personality Recognition from Text is a computational linguistic task, partially connected to stylometry. It consists in the automatic classification of authors' personality traits using textual cues as features. Personality has been formalized in

[1] http://clic.cimec.unitn.it/fabio/pr2demo.php

978-0-7695-5062-6/13 $26.00 © 2013 IEEE
DOI 10.1109/EISIC.2013.8

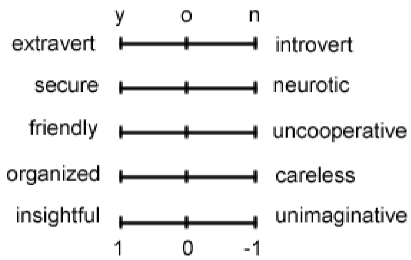

Fig. 1. Formalization of Personality.

various ways, and can be assessed by means of different questionnaires, such as the Myers-Briggs type indicators [4], that defines four personality types, the Interpersonal circumplex [26], that defines 8, and the Big5 [33], [10], [11], that defines five bipolar traits and has become a standard over the years.

The Big5 traits, introduced by Costa and MacCrae [10], are for instance: Extraversion, Emotional Stability/Neuroticism, Agreeableness, Conscientiousness and Openness to experience. Extraversion describes a person along the two opposite poles of sociability and shyness. Emotional stability, which is sometimes referred by its negative pole (neuroticism), describes the modality of impulse control along a scale that goes from control (a calm and stable person) to instability (an anxious and neurotic person). Agreeableness refers to the tendency to be sympathetic and cooperative towards others, rather than suspicious and antagonistic. Conscientiousness describes a person in terms of self-discipline versus disorganization. Openness to experience refers to the tendency to be creative and curious rather than unimaginative.

Most scholars working in the field of personality recognition from text [35] [29], [22], with some isolated exceptions, such as Luyckx and Daelemans [28], used the Big5 factor model. Mairesse et al 2007 [29] reported correlations between linguistic cues and personality traits, that can be exploited for automatic classification. The bipolar scales defined by the Big5 are suitable for computational processing, because they can be turned into continuous (-1, 0, 1) or nominal (y, o, n) variables, as shown in figure 1.

III. Data and Settings

A. Dataset

DeCour, widely described in [20], is a corpus constituted by hearings held in the Italian Courts of Bologna, Bolzano, Prato and Trento. The hearings are characterized by the fact that the speakers, who appeared in front of the judge as witnesses or as defendants for any criminal proceeding, issued statements which were suspected to be false and for this reason became body of evidence for a further criminal proceeding. In particular, DeCour collects 35 hearings where 31 subjects - 4 of them were heard twice - were found guilty of calumny or false testimony. In these cases, the judgments issued by Courts summarized the facts, pointing out precisely the lies told by the speakers. Therefore it was possible to annotate each utterance constituting DeCour as true, false or uncertain.

Table I shows the amount of utterances and tokens belonging to the different classes in DeCour. In order to evaluate

TABLE I. Labels of DeCour's utterances.

Label	Nr. utterances	Nr. tokens
True	1202	15456
Undecidable	868	10439
False	945	15924
Total	3015	41819

the reliability of these labels, an agreement study was carried out. Three annotators labeled about 20% of DeCour, and their agreement was measured using Kappa as metric [2]. The values of K were calculated in two conditions, that is considering the three mentioned classes and two classes only. In the last case, the true and uncertain utterances were collapsed in the class of not-false utterances. In these two conditions, the values of K were .57 and .64 respectively. The value for two classes indicates a moderate/substantial agreement [5], [25].

B. Text classification methods

In order to carry out text classification experiments, we followed the same approach more throughly described in [21]. Each utterance issued by the subjects heard in the hearings were represented as feature vector, whose values were constituted by the frequency of n-grams of lemmas and part-of-speech (POS) found into the utterance itself.

The n-grams which were considered as features were identified calculating the Information Gain of all the n-grams which appeared at least 5 times into the true or false utterances. The uncertain ones were not considered during the process of feature selection. Information Gain is a well known algorithm which "measures the decrease in entropy when the feature is given vs. absent" into the considered classes of instances [17]. In particular, the feature list was constituted by the n-grams (from uni-grams to epta-grams) with $IG \geq .01$. Since the experiments were carried out through n-fold cross-validation, in each fold the Information Gain was computed for the n-grams of the training set, in order to avoid the involvement of the test set in the feature selection.

Once the feature vectors were created, we trained models in order to carry out some text classification experiments. We used SVM as classifier [9], which performed particularly well on our dataset, under two different conditions:

- False vs. Not-False utterances. In this case, all the 3015 utterances of DeCour were classified, and true and uncertain utterances together constituted the class of not-false utterances.
- False vs. True utterances. In this experiment, only true and false utterances were employed, that is 2147 instances, while the uncertain ones were discarded.

Both the experiments were carried out through a 35-fold cross-validation, where each fold was represented by a single hearing of DeCour. This prevented over-fitting problems within the same hearing

Lastly, in order to evaluate the effectiveness of the models, we employed a heuristic baseline. This relies on the idea that, since in the hearings the speakers have to give answers regarding facts which have already been verified, they will be prone to lie denying them, and to be truthful confirming them. According to the heuristic baseline, a number of simulations were performed, where the utterances were considered as not-false (that is true or uncertain) if they began with the words 'Yes', 'I know' or 'I remember', and as false if they began with 'no', 'I do not know' or 'I do not remember'. The other utterances were classified randomly, according to the rate of the classes into the dataset. The performance of this algorithm was higher than that which could be found through completely random simulations. In particular, the threshold for the first experiment was 62.39% and for the second one 59.57%.

C. Personality Recognition System and Feature Set

For the annotation of DECOUR with personality labels, we exploited a system, available online, that performs instance-based classification of personality types in an unsupervised way, using language-independent features (see table II) [6], [7]. The system takes as input unlabeled text data with authors and an initial set of correlations between personality traits and linguistic or extralinguistic features. The output is one generalized hypothesis of personality for each author. Personality hypotheses are formalized as 5-characters strings, each one representing one trait of the Big5. Each character in the string can take 3 possible values: positive pole (y), negative pole (n) and omitted/balanced (o), which stands for classifier's abstention. For example "ynoon" stands for an extrovert neurotic and not open mindend person.

As initial feature set, the system exploits language-independent features extracted from LIWC [36] and MRC [12], whose correlations to personality are reported by Mairesse et al. 2007. The features are: punctuation (ap); question marks (qm); quotes (qt); exclamation marks (em); numbers (nb); parentheses (pa); repetition ratio (tt), word frequency (wf).

The pipeline of the personality recognition system has

TABLE II. FEATURE/CORRELATIONS SET, ADAPTED FROM MAIRESSE ET AL. 2007. * = p SMALLER THAN .05 (WEAK CORRELATION), ** = p SMALLER THAN .01 (STRONG CORRELATION).

feature	ext.	emo.	agr.	con.	ope.
ap	-.08**	-.04	-.01	-.04	-10**
em	-.00	-.05*	.06**	.00	-.03
mf	.05*	-.06**	.03	.06**	-.07**
nb	-.03	.05*	-.03	-.02	-.06**
pa	-.06**	.03	-.04*	-.01	.10**
qm	-.06**	-.05*	-.04	-.06**	.08**
qt	-.05*	-.02	-.01	-.03	.09**
tt	-.05**	.10**	-.04*	-.05*	.09**
wf	.05*	-.06**	.03*	.06**	.05**

three phases. In the preprocessing phase, the system samples 20% of the input unlabeled data, computing the average distribution of each feature of the correlation set, then assigns personality labels to the sampled data according to the correlations. In the processing phase, the system generates one personality hypothesis for each text in the dataset, mapping the features in the correlation set to specific personality trait poles, according to the correlations. Instances are compared to the average of the population sampled during the preprocessing phase and filtered accordingly. Only feature values above the average are mapped to personality traits. For example a text containing more punctuation than average will fire negative correlations with extraversion and openness to experience (see table II). The system keeps track of the firing rate of each single feature/correlation and computes personality scores for each trait, mapping positive scores into "y", negative scores into "n" and zeros into "o" labels. In this phase the system computes also per-trait confidence, defined as the coverage of the selected label over all the author's texts.

$$conf = \frac{m}{T}$$

where m is the count of instances of the selected pole of the personality trait, and T is the count of the author's texts.

In the evaluation phase the system compares all the hypotheses generated for each single text of each author and retrieves one generalized hypothesis per author by computing the majority class for each trait. In the evaluation phase the system computes average confidence and variability. Average Confidence is derived from per-trait confidence scores and gives a measure of the robustness of the personality hypothesis. Variability (var) gives information about how much one author tends to write expressing the same personality traits in all the texts. It is defined as

$$var = \frac{avg\ conf}{T}$$

where *avg conf* is the confidence averaged over the five traits and T is the count of all author's texts. The system can evaluate personality only for authors that have at least two texts, and all the defendants in DECOUR fit this requirement.

The system provides the following optional parameters: **Feature Score Normalization**. This option normalize the feature distribution computed in the preprocessing phase. **Automatic Feature Weighting**. When the parameter is activated, high feature firing rates decrease the personality score associated to that feature. This parameter boosts the infomation provided by low-frequency features. **Weak trait correction**. If the correction parameter is activated, the system generates random labels for the poles of the traits for which no features were detected in the preprocessing phase. **Variable hypothesis Average**. If this parameter is activated, the average distribution of each feature is recomputed on the fly during the processing phase. This allows to better fit the data at hand. **Hypothesis score normalization**. If this parameter is activated, the system normalizes hypotheses scores during the processing phase. If paired with Variable Hypothesis average, this parameter reduces consistently the amount of classifier's abstentions ("o" labels). **Automatic Pattern Extraction**. If this parameter is activated, the system automatically extracts new patterns from the data at hand exploiting the confidence scores, associates them to personality traits, and uses them as new correlations between patterns and personality traits.

The system has been tested on English and Italian (see [7]) obtaining f-measures between .63 and .68. To the best of our knowledge, it is the first system for personality recognition from text that has been tested on Italian.

IV. EXPERIMENTS

A. Utterance classification

Table III shows the performance of the models in the first experiment, where the classification task involved false and not-false utterances. The results of second experiment, in which uncertain utterances were removed, are presented in Table IV instead. In both cases, the accuracy is well above the threshold levels. Even though the performance of the second experiment is lower than that of the first one, the distance from the performance to the baseline is wider of about 1 point percent, due to the even lower value of the baseline. Therefore the models' performance can be considered as better when the uncertain utterances are discarded than when they are classified. This is not surprising. In fact, as discussed in [21] (where a wider error analysis is also carried out), the ground truth of the uncertain utterances is unknown and their class includes true and false statements merged together. Thence to remove them means to remove noisy data, and this enhances the models' performance.

TABLE III. False vs. Not-False utterances classification

	Correctly classified entities	Incorrectly classified entities
False	342	284
Not-False	1786	603
Total accuracy	70.58%	29.42%
Baseline	62.39%	

TABLE IV. False vs. True utterances classification

	Correctly classified entities	Incorrectly classified entities
False	511	234
True	968	434
Total accuracy	68.89%	31.11%
Baseline	59.57%	

B. Personality and deception classification

We extracted all the speakers from DeCour. We kept only the defendants and the witnesses, that were annotated with truth labels (true, false, uncertain), filtering out all the other speakers without a truth class label (that is judge, prosecutor, lawyer and few others). We will refer to this corpus as DECOURdef. We exploited the personality recognition system to annotate DECOURdef with personality types. As optional parameters we used feature score normalization and variable hypothesis average. We used personality labels as nominal features and truth labels (only true and false, because we substituted "uncertain" with a missing value) as target binary classes. The final DECOURdef corpus contains 31 defendants and about 1075 words for each one. Tables V and VI report a summary of the distribution of the personality labels among the defendants in the corpus and the rank of the personality types. As further experiment, we tested whether the information provided by the 5 personality labels can help the prediction of deception. We ran the experiment in Weka [44] using the five traits as features and truth labels as the target class, a 10-fold cross validation as evaluation setting, and four different algorithms. The results, reported in table VII, show that all the algorithms significantly outperform the majority baseline (zeroR), in particular for precision. This indicates that the information provided by personality labels is valuable for deception detection. In particular, decision trees [38] achieved the best performance, outperforming Naive Bayes [23], Ripper rule induction [8] and SMO Support Vector Machines [37]. It is very interesting to note that the only personality traits used by the decision tree for the classification are emotional stability/neuroticism and openness to experience. The tree, reported in figure 2, shows that secure and not open minded people tend to lie, while open minded people tend to tell the truth.

TABLE V. Personality labels in DECOURdef

label	x	e	a	c	o
y	8	10	15	14	24
o	0	0	1	1	0
n	27	25	19	20	11

TABLE VI. Personality type rank.

rank	personality type	freq
1	nnyyy	7
2	nynnn	7
3	ynyyy	7
4	nnnnn	4
5	nnnny	4
6	nynny	3
7	nnony	1
8	nnyoy	1
9	ynnny	1

TABLE VII. Deception classification via Personality.

algorithm	P	R	F
mbl (zeroR)	0.313	0.56	0.402
dt (J4.8)	0.579	0.586	0.55
nb (NaiveBayes)	0.548	0.562	0.538
svm (SMO)	0.582	0.585	0.533
ripper (JRip)	0.576	0.582	0.532

However these results are not directly comparable to the results reported in tables III and IV, because personality types are relative to each single defendant, while labels and features for utterance classification are relative to each single utterance. In other words, at hearing level the personality features are constant, therefore they have the same value in every utterance belonging to the same hearing. For this reason, so far to use personality types as features for utterances' classification did not improve the performance, thence we preferred to keep personality and utterance classification as two different tasks.

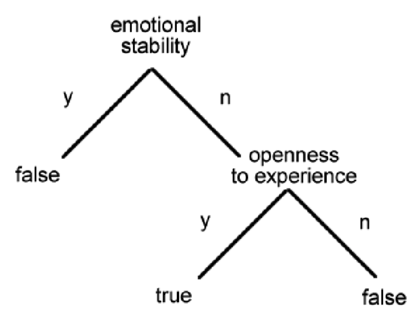

Fig. 2. Decision tree.

C. Clustering

In order to obtain a synthetic view of the distance between personality profiles of the speakers in DeCour, we transformed these profiles into a matrix of between-hearing distances and a Multi-Dimensional Scaling - MDS function has been applied to this matrix [3]. The results are shown in Figure 3 and 4. The entities in the charts represent the DECOUR's hearings, and their labels present the personality profile of the speaker and the models' accuracy in the classification task concerning the hearing itself. In Figure 3 the accuracies refer to the first experiment of utterances' classification, in Figure 4 to the second one. In both charts, the accuracies lower than the baseline (respectively 62.39% and 59.57%) are highlighted in red. As it can be seen in Figure 3, in the first experiment the not well classified hearings are dispersed through the whole chart.

In the second experiment, instead, all the hearings where the models' accuracy was lower than the baseline concentrate on the left part of the chart (Figure 4). By contrast, a group of well classified hearings lies on the right area of the chart. This suggested the presence of two different clusters of hearings, highlighted by the green and red ellipses.

In order to explore this hypothesis, two Student's t-test were carried out, concerning the accuracies of the two experiments showed in the Figures 3 and 4. In both cases, the comparison involved the group of 15 hearings identified by the green ellipsis and the group of 20 hearings included into the red ellipsis of Figure 4. The two groups turned out to belong to different populations, both regarding the accuracies of the first and of the second experiment. In particular, for the classification accuracies of false vs. not-false utterances, the significance was $p = .0412$, while for false vs. true utterances we obtained $p = .0485$ (paradoxically, the value was better in the first experiment, in spite of the presence, in that condition, of not well classified hearings in both groups).

V. CONCLUSIONS AND FUTURE WORK

In this study we combined deception detection and personality recognition techniques, in order to get some insight regarding the possible relation between deception and personality traits from the point of view of their linguistic expression. We have seen that personality traits, used as features, can be quite successful for deception detection, and that decision trees achieved the best performance in the prediction of true and false labels. Probably this result is due to the good performance of decision trees with nominal values.

However, to use personality traits to detect deception means to work at level of the whole narrative, rather than of the single statements, and this basically identifies the liar rather than the lie, as pointed out by [16]. Therefore we tried to examine how personality traits may correlate with deceptive communication style. Even though the relatively little amount of subjects allowed us to obtain only few types of personality, Figure 4 suggests that the machine learning models perform better with subjects showing certain kind of personality. This would mean that, in their communication style, deceptive statements are more easily recognizable; that is they differ more clearly from the truthful ones. The well classified personalities are "friendly", "organized" and "insightful", even though this last trait appears also in some hearings of the cluster with lower models" performance. Maybe not surprisingly, in the hearings difficult to classify, the subjects were "uncooperative". In most hearings, subjects showed to be "neurotic" rather than "secure": this evidence may be due to the particularly stressful conditions of the hearings in Court. Similarly, many subjects

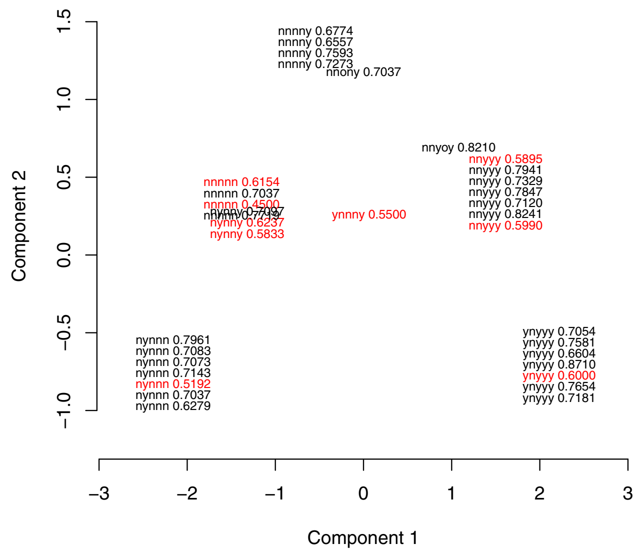

Fig. 3. Multi-Dimensional Scaling of personality profiles in DECOUR, with the models' performance in the False vs. Not-False utterances classification task. The entities highlighted in red are those in which the models' performance is lower than the threshold of 62.39%.

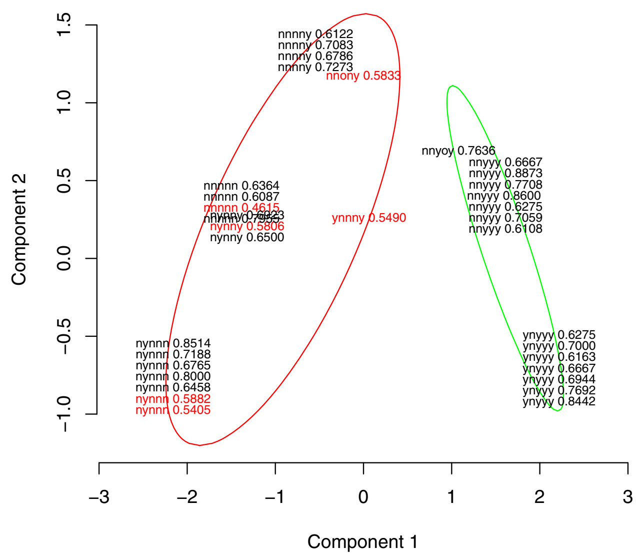

Fig. 4. Multi-Dimensional Scaling of personality profiles in DECOUR, with the models' performance in the False vs. True utterances classification task. The entities highlighted in red are those in which the models' performance is lower than the threshold of 59.57%.

turned out to be "introvert" rather than "extrovert". However, all the subjects which seemed to be "secure" were not easy to be classified for the models; by contrast, the "extrovert" ones belonged to the group where deception was detected more effectively.

These results cannot be considered entirely conclusive as many personality profiles were not taken into consideration in this study due to the limited number of subjects; but they are certainly suggestive, and expect to broaden the range of personalities analyzed in future work.

References

[1] Adelstein J.S., Shehzad Z., Mennes M., DeYoung C.G., Zuo X-N., Kelly C., Margulies D.S., Bloomfield A., Gray J.R., Castellanos X.F. and Milham M.P. Personality Is Reflected in the Brain's Intrinsic Functional Architecture. In *PLoS ONE* 6:(11), 1–12. 2011.

[2] Artstein, R. and Poesio, M. Inter-coder agreement for computational linguistics. *Comput. Linguist.*, 34(4):555-596. 2008.

[3] Baayen, R.H. Analyzing linguistic data: a practical introduction to statistics using R. *Cambridge University Press*. 2008.

[4] Briggs, I. and Myers, P.B. *Gifts differing: Understanding personality type*. Mountain View, CA: Davies-Black Publishing. 1980.

[5] Carletta, J. Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.* 22(2):249-254. 1996.

[6] Celli, F. Unsupervised Personality Recognition for Social Network Sites. In *Proc. of ICDS 2012 : The Sixth International Conference on Digital Society*, 59–62. 2012.

[7] Celli, F. *Adaptive Personality recognition from Text*. Lambert Academic Publishing. 2013.

[8] Cohen, W. W. Fast effective rule induction. In *Proc. of the 12th International Conference on Machine Learning*. 1995.

[9] Cortes C., Vapnik V. Support-vector networks.*Mach Learn*. 20(3):273–297. 1995.

[10] Costa P.T., Jr. and MacCrae, R.R. *The NEO Personality Inventory manual*. Psychological Assessment Resources. 1985.

[11] Costa, P. T. and MacCrae, R. R. Normal personality assessment in clinical practice: The neo personality inventory. Psychological assessment, 4(1):5. 1992.

[12] Coltheart, M. The MRC psycholinguistic database. In *Quarterly Journal of Experimental Psychology*. 33(A):497-505. 1981.

[13] Coulthard, M. Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*. 25(4):431-447. 2004.

[14] Enos, F. Benus, S. Cautin, R.L., Graciarena, M., Hirschberg, J., and Shriberg, E. Personality Factors in Human Deception Detection: Comparing Human to Machine Performance. In *Proc. of INTERSPEECH - ICSLP*. 813–816. 2006.

[15] Fitzpatrick, E. and Bachenko, J. Building a forensic corpus to test language-based indicators of deception. *Language and Computers*. 71(1):183-196. 2009.

[16] Fitzpatrick, E. and Bachenko, J. Building a data collection for deception research. *Ln Fitzpatrick, E., Bachenko, J., and Fornaciari, T., editors, Proc. of the EACL Workshop on Computational Approaches to Deception Detection*. 31-38. 2012.

[17] Forman G. An extensive empirical study of feature selection metrics for text classification. *J Mach Learn Res*. 3:1289–1305. 2003.

[18] Fornaciari, T. and Poesio, M. Lexical vs. surface features in deceptive language analysis. In *Proc. of the ICAIL2011, Workshop Applying Human Language Technology to the Law, in conj. to AHLTL2011*. 2-8. 2011a.

[19] Fornaciari, T. and Poesio, M. Sincere and deceptive statements in italian criminal proceedings. In *Proc. of the International Association of Forensic Linguists Tenth Biennial Conference, IAFL2011*. 126-138. 2011b.

[20] Fornaciari, T. and Poesio, M. DeCour: a corpus of DEceptive statements in Italian COURts. In Calzolari, N.C.C., Choukri, K., Declerck, T., Uur D., M., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., (editors), *Proc. of LREC12*. European Language Resources Association (ELRA). 1585–1590. 2012.

[21] Fornaciari, T. and Poesio, M. (2013). Automatic deception detection in italian court cases. *Artificial Intelligence and Law*. 1-38. 2013.

[22] Iacobelli, F., Gill, A.J., Nowson, S. and Oberlander, J. Large scale personality classification of bloggers. In *Lecture Notes in Computer Science (6975)*. 568–577. 2011.

[23] John, G. H. and Langley P. Estimating Continuous Distributions in Bayesian Classifiers. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. 1995.

[24] Koppel, M., Schler, J., Argamon, S., and Pennebaker, J. Effects of age and gender on blogging. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*. 1–7. 2006.

[25] Landis, R. J. and Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics*. 33(1):159-174. 1977.

[26] Locke, K.D. Circumplex Scales of Interpersonal Values: Reliability, validity, and applicability to interpersonal problems and personality disorders. *Journal of Personality Assessment*, 75: 249-267. 2000.

[27] Luyckx, K. and Daelemans, W. Authorship attribution and verification with many authors and limited data. In *Proc. of the 22nd International Conference on Computational Linguistics*. 1:513-520, 2008a

[28] Luyckx, K. and Daelemans, W. Personae: a corpus for author and personality prediction from text. In *Proc. of LREC2008, the Sixth International Language Resources and Evaluation Conference*, 2981–2987. 2008b.

[29] Mairesse, F., Walker, M.A., Mehl, M.R., and Moore, R.K. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial intelligence Research*, 30: 457–500. 2007.

[30] Mischel, W., Shoda, Y., and Ayduk, O. *Introduction to Personality: Toward an Integrative Science of the Person.* John Wiley & Sons. Hoboken, NJ. 2007.

[31] Mosteller, F. and Wallace, D. *Inference and Disputed Authorship: The Federalist*. Addison-Wesley, Boston, MA. 1964.

[32] Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M. Lying Words: Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*. 29(5):665-675. 2003.

[33] Norman, W., T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating. In *Journal of Abnormal and Social Psychology*, 66. 574–583. 1963.

[34] Nguyen, D. and Gravel, R. and Trieschnigg, D. and Meder, T. How Old Do You Think I Am?"; A Study of Language and Age in Twitter. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 2013

[35] Oberlander, J., and Nowson, S. Whose thumb is it anyway? classifying author personality from weblog text. In *Proc. of the 44th Annual Meeting of the Association for Computational Linguistics ACL*. 627–634. 2006.

[36] Pennebaker, J. W., Francis, M. E., Booth, R. J. *Inquiry and Word Count*. Lawrence Erlbaum, Mahwah, NJ. 2001.

[37] Platt, J. Machines using Sequential Minimal Optimization. In Schoelkopf, B., Burges, C., Smola, A. (ed), *Advances in Kernel Methods, Support Vector Learning*. 1998.

[38] Quinlan, R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers. 1993.

[39] Solan, L. M. and Tiersma, P. M. Author identification in american courts. *Applied Linguistics*. 25(4):448-465. 2004.

[40] Stein, B., Koppel, M., and Stamatatos, E. Plagiarism analysis, authorship identification, and near-duplicate detection pan07. *SIGIR Forum*. 41:68-71. 2007.

[41] Strapparava, C. and Mihalcea, R. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proc. ACLShort09 - Proc. of the ACL-IJCNLP 2009 Conference.* 1–4. 2009.

[42] Vrij, A. Detecting Lies and Deceit: Pitfalls and Opportunities. In *Wiley Series in Psychology of Crime, Policing and Law*. John Wiley & Sons, 2nd edition. 2008.

[43] Vrji, A. Criteria-based content analysis - A Qualitative Review of the First 37 Studies. *Psychology, Public Policy, and Law*. 11(1):3-41. 2005.

[44] Witten, I.H., and Frank, E. *Data Mining. Practical Machine Learning Tools and Techniques with Java implementations*. Morgan and Kaufman. 2005.