Gender Differences in Deceivers Writing Style

Verónica Pérez-Rosas and Rada Mihalcea

University of North Texas, University of Michigan
vrncapr@umich.edu, mihalcea@umich.edu

Abstract. The widespread use of deception in written content has motivated the need for methods to automatically profile and identify deceivers. Particularly, the identification of deception based on demographic data such as gender, age, and religion, has become of importance due to ethical and security concerns. Previous work on deception detection has studied the role of gender using statistical approaches and domain-specific data. This work explores gender detection in open domain truths and lies using a machine learning approach. First, we collect a deception dataset consisting of truths and lies from male and female participants. Second, we extract a large feature set consisting of n-grams, shallow and deep syntactic features, semantic features derived from a psycholinguistics lexicon, and features derived from readability metrics. Third, we build deception classifiers able to predict participant's gender with classification accuracies ranging from 60-70%. In addition, we present an analysis of differences in the linguistic style used by deceivers given their reported gender.

Keywords: deception, linguistics, machine learning

1 Introduction

The increasing presence of deceit in written content has motivated the need for automatic methods able to identify deceptive behavior. Particularly, the identification of deception based on demographic data such as gender, age, education, and religion among others, has become of importance due to ethical and security concerns. Online date websites, forums, and social media, have reported multiple cases of strategic misrepresentation, with people lying mainly about their gender, age, and physical attributes such as height and weight [15, 17, 6]. Among these aspects, we focus on the identification of gender in deception, which can also be associated with gender imitation or gender misrepresentation.

We start by collecting a deception dataset consisting of truths and lies from male and female participants. Unlike other studies, where authors established a specific domain, the domain of our dataset is not pre-determined as we hypothesize that when lying in an open domain setting deceivers will show natural bias towards specific topics related to gender.

Using this dataset, we extract a large feature set consisting of n-grams, shallow and deep syntactic features, semantic features derived from a psycholinguistics lexicon, and features derived from readability metrics. Most of these features have been previously found to be effective for the prediction of deceptive behavior.

We perform a set of experiments to explore three research questions. First, can we build deception classifiers using short open domain truths and lies? Second, given a deceptive corpus from female and male deceivers, can we build deception classifiers able to predict deceiver's gender? Third, what are the topics more frequently discussed by male and female deceivers? Finally, we discuss our main findings and future work directions.

2 Related work

Several efforts have been presented to approach the automatic identification of deceivers in written sources using computational linguistic approaches. Lie detection has been explored in different domains such as e-mail communication, dating websites, blogs, forums, chats, and social network websites.

Research in this area has shown the effectiveness of features derived from text analysis, including n-grams, sentence counts, and sentence length. More recently, features derived from syntactic Context Free Grammar (CFG) parse trees, and part-of-speech (POS) tags have also been used to aid the deceit detection [4, 18]. Syntactic complexity has been also found to be correlated with deception [19] as related research suggests that deceivers might create less complex sentences in an effort to conceal the truth and being able to recall their lies more easily [2].

A widely used resource for incorporating semantic information is the Linguistic Inquiry and Word Count (LIWC) dictionary [13]. LIWC is a lexicon of words grouped into semantic categories relevant to psychological processes. Several research works have relied on the LIWC lexicon to build deception models using machine learning approaches [10, 1] and showed that the use of semantic information is helpful for the automatic identification of deceit.

Deception detection has usually been applied to discriminate true-tellers from liars. For instance, Ott et al. [12] identified spam producers by analyzing deceptive reviews. Also, Fornaciari and Poesio [5] analyzed transcripts of court cases to identify deceptive testimonies.

Despite the fact that gender imitation and misrepresentation has been reported as one of the main forms of deception in online sources [7], very little attention has been paid to address the identification of deception based on demographic data using computational approaches. It is however worth mentioning important efforts in the field of psychology to analyze demographics influence during the deception process. Studies have revealed interesting findings regarding the role of gender during deception. For instance, according to Kaina et al. [8], females are more easily detectable when lying than their male counterparts. On the other hand Tilley et al. [14] reported that females are more successful in deception detection than male receivers. Furthermore, gender perception has an important effect on the receiver and it can lead to important implications. For instance, gender perception can have an impact on trustworthiness, as females are perceived as more cooperative and less dominant than males [3].

Finally, it is important to point out that the scarcity of resources for this task so far made it difficult to approach the problem using machine-learning techniques. We are aware of only one other resource for deception detection where demographic data is available [16]. The lack of standard datasets for this task motivated us to build our own dataset, which is publicly available at http://lit.eecs.umich.edu and represents an additional contribution of this work.

3 Dataset

In order to collect a deception dataset, we set up a task on Amazon Mechanical Turk where we asked workers to provide seven lies and seven truths, each consisting of one sentence, on topics of their choice. Participants were asked to provide plausible lies and avoid non-commonsensical statements such as "A cat can bark." We also collected demographic data for the contributors, such as gender, age, and education level. The final dataset consists of 3584 truths and lies provided by 512 contributors. The dataset distributions for gender, truths, and lies are presented in Table 2. Sample one-liners containing truths and lies are presented in Table 2.

Table 1. Dataset distribution

Gender	Lies	Truths	Total
Female	2086	2086	4172
Male	1498	1498	2996
Total	3584	3584	7168

Table 2. Sample open domain lies and truths provided by a male and a female participant

.

Female	
Lie	Truth
I'm allergic to alcohol	Giraffes are taller than zebras.
I am missing a toe on my left foot.	Humans are not able to fly.
My shoes cost me over a hundred dollars.	The meat industry is cruel to animals.
Male	
Lie	Truth
I own two Ferraris, one red and one black	I love to play soccer with my friends
I wake up at 11 o clock every day	I wake up at 6 am because I have to work at 7 am
I have a jumping bed in my backyard	I own a 2003 white lancer and a 2008 silver Toyota 4runner

4 Features

We extract a large number of features, consisting of several features that have been previously found to correlate with deception cues.

Unigrams: We extract unigrams derived from the bag of words representation of the one-liners present in our dataset. Our unigrams features are encoded as term frequency inverse document frequency (tf-idf) values.

Shallow and deep syntax: Following Feng et al. [4] we extract a set of features derived from POS tags and production rules based on CFG trees. We use the Berkeley parser to obtain both POS and CFG features. Our POS features are encoded as the tf-idf values of each POS tag occurring in the dataset. The CFG derived features consist of all lexicalized production rules combined with their grandparent node and are also encoded as tf-idf values.

LIWC derived features: We use features derived from the LIWC lexicon, except for the paralinguistic classes. These features consist of word counts for each of the 80 semantic classes present in the LIWC lexicon.

Syntactic complexity and readability score features: To extract these features we use a tool provided by Lu et al. [9], which generates fourteen indexes representing sentence syntactic complexity including: mean length of sentence (MLS), mean length of T-unit (MLT), mean length of clause (MLC), clauses per sentence (C/S), verb phrases per T-unit (VP/T), clauses per T-unit (C/T), dependent clauses per clause (DC/C), dependent clauses per T-unit (DC/T), T-units per sentence (T/S), complex T-unit ratio (CT/T), coordinate phrases per T-unit (CP/T), coordinate phrases per clause (CP/C), complex nominals per T-unit (CN/T), and complex nominals per clause (CP/C). In addition, we also incorporate standard readability metrics such as Flesch-Kincaid and Gunning Fog.

5 Experiments

We perform several experiments to answer the research questions formulated at the beginning of this paper.

5.1 Is it possible to build accurate deception classifiers for short open domain truths and lies?

To answer this question, we build deception classifiers using each set of features described in section 4, and also using a combination of unigrams and each of the remaining feature sets. All classifiers were created using the Support Vector Machine (SVM) algorithm as implemented in the Weka toolkit with the default parameter configuration. Results are obtained using a ten-fold cross-validation.

First, we evaluate the deception detection task. We build classifiers using all truths and lies present in the dataset, regardless of gender. The purpose of this experiment is to evaluate the deception detection task when using short deceptive and truthful statements and also to explore which feature set is more suitable for these data sets. Figure 1 presents the overall and per-class accuracies obtained by each classifier.

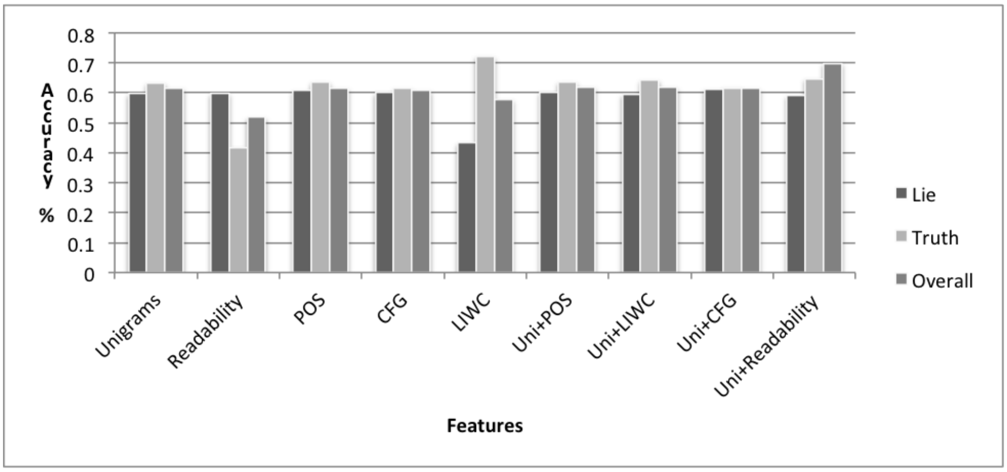

Fig. 1. Deception classification results in terms of accuracy percentages using different feature sets.

As the graph shows, the individual use of unigrams and features derived from POS and CFG leads to similar classification accuracies. Also, the classifier built with features representing syntactic complexity and readability scores is the worst performing classifier. However, when combined with unigrams, this turns out to lead to the best classifier with the highest overall accuracy values. In general, from this graph we can observe that given the various classifiers, the deceptive class is always more difficult to predict than the truthful one. Interestingly, the LIWC-based classifier shows the best performance for the truthful class.

Second, we explore the deception detection within gender. We split our dataset based on reported gender and obtain two datasets. One dataset consists of truths and lies from males (male dataset), while the other consists of truths and lies from females (female dataset). As before, we build deception classifiers for each dataset using the different feature sets. Classification results are reported in Table 2 and 3 respectively.

From these figures we notice that combining unigrams with readability and syntactic complexity features seem to help the most while predicting deception. We can observe also interesting differences in the classification accuracy per class. For instance, the classification accuracies obtained for the truth and lie classes on the male dataset are very similar, thus suggesting that truths and lies are equally difficult to predict. However, for the female dataset, the detection of truths seems easier than the detection of lies.

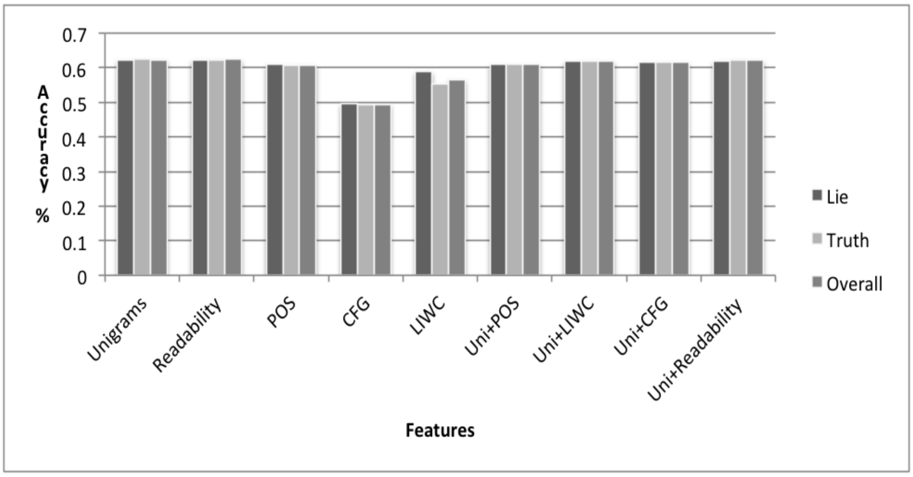

Fig. 2. Deception classification of female truths and lies for several feature sets.

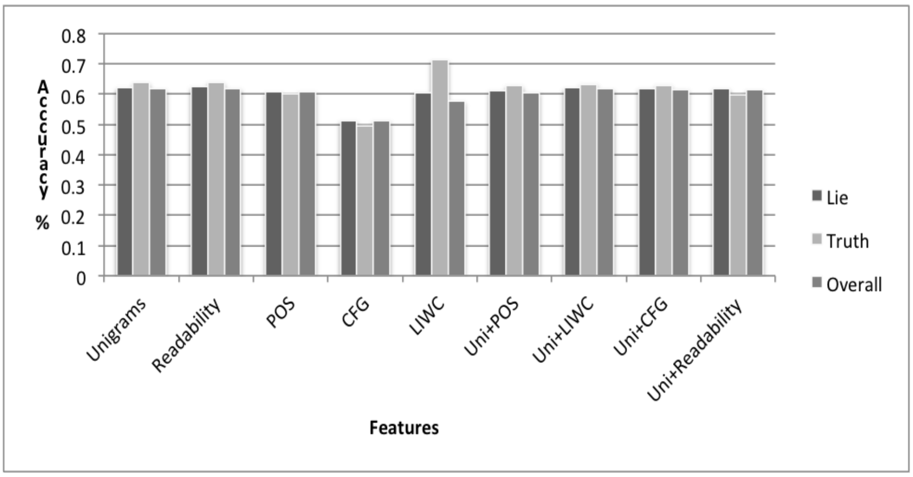

Fig. 3. Deception classification of male truths and lies for several feature sets.

5.2 Can we build deception classifiers able to predict deceiver's gender?

To answer this question we focus on predicting the gender of an author of a deceptive sentence. Thus, in these experiments, we use a subset from our deception dataset consisting of only lies.

Using this deception corpus, we build several deception classifiers using the each of feature sets described in section 4. As before, our experiments are performed using the SVM algorithm and using 10-fold cross-validation. Note that the class distribution for this deceptive corpus is unbalanced as we have 2086 female and 1498 male instances. Thus, the baseline, corresponding to the majority class is 58.2%.

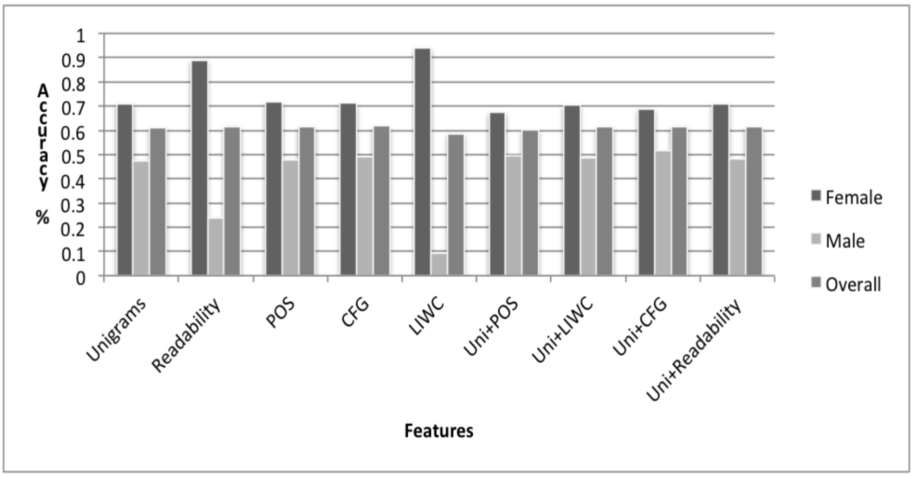

Fig. 4. Gender classification results in terms of accuracy percentage using different feature sets

Figure 4 shows the accuracy results for female and male classes, and overall accuracy. From this figure, we can observe that the female class is more easily predicted than the male class. While this can be in part attributed to the dataset imbalance, these results are also in line with the findings reported in Ho et al. [7], where female deceivers were more easily identifiable than the males ones. Among the different classifiers presented in this graph, we can observe that overall accuracy values are very similar to each other and the male class is always more difficult to predict. This time the combination of unigrams and CFG features seems to provide the best performance.

5.3 What are the topics more frequently discussed by male and female deceivers?

To answer this question, we initially attempted to automatically identify topics from lies generated by male and female deceivers. We further split our deception corpus into male and female sets and applied Latent Dirichlet Allocation (LDA) topic modeling to identify which topics are associated to each gender. However, this approach generated a very large number of topics and we were unable to identity specific topics that allowed us to make such comparisons.

Table 3. Results from LIWC word class analysis. Top ranked semantic classes associated to truths and lies generated by both female and males (Male+Female), male only (Male), and female only (Female)

Lies					
Male+Female		Male		Female	
Class	Score	Class	Score	Class	Score
Certain	1.94	Other	2.22	**Certain**	1.87
Negate	1.79	**Negate**	2.08	**Negate**	1.63
You	1.68	**Certain**	2.06	**You**	1.59
Anger	1.64	Death	2.04	Motion	1.47
Down	1.42	**Anger**	2.03	**Down**	1.45
Motion	1.41	**You**	1.77	**Money**	1.35
Money	1.38	**Friends**	1.71	**Anger**	1.28
Friends	1.37	**Othref**	1.67	Future	1.20
Othref	1.35	**Down**	1.47	**Othref**	1.19
Death	1.28	**Money**	1.44	Sports	1.15
Other	1.26	Sleep	1.41	**Eating**	1.15
Eating	1.22	**Eating**	1.36	**Friends**	1.14
Truths					
Male+Female		Male		Female	
Incl	0.87	Leisure	0.82	Number	0.85
Number	0.86	Posemo	0.82	Music	0.85
Discrep	0.85	Sports	0.79	Tentat	0.83
Posemo	0.85	Occup	0.75	We	0.83
School	0.83	Job	0.73	Tv	0.76
Family	0.82	**Posfeel**	0.72	Metaph	0.75
Sexual	0.81	**Relig**	0.70	**Posfeel**	0.75
Tv	0.79	Sexual	0.70	Anx	0.74
See	0.75	School	0.64	Discrep	0.72
Music	0.74	Music	0.60	See	0.67
Posfeel	0.73	Groom	0.59	**Relig**	0.62
Relig	0.65	Family	0.59	Sleep	0.61

In order to provide some insight about word usage differences, we opted instead for applying the method proposed by Mihalcea et al. [11] and obtained the most dominant semantic word classes, extracted using the LIWC lexicon, associated to each gender. Table 3 shows the most dominant words classes used by both female and male deceivers as well as only female and only male. To provide a more comprehensive analysis, this table also shows the most dominant classes used by true-tellers. To facilitate the comparisons based on gender, we show in bold the overlapping classes taken from the top twelve ranking classes for both deceptive and truthful categories.

From this table, we can observe that, regardless of their gender, deceivers make use of negation, negative emotions, and references to others. On the other hand, when telling the truth, more positive emotion (Posemo) and positive feeling (Posfeel) words are used. Also, some word classes suggest topics associated with truths, such as religion, family, and school.

Regarding the differences between genders, we first observe that the overlap of semantic word classes associated with deception is greater than the overlap of classes associated with true-telling. One possible explanation for this is that lies are told about similar topics while truths seem to be more diverse. Tables 4 and 5 show sample lies from male and female participants for four overlapping and non-overlapping semantic classes. As observed, both male and female share some commonalities on the words used when deceiving. For instance, the use of negation and anger words, and words referring to friends and eating. On the other hand, in Table 5 we can see that females lie more about sports and future actions, while males lie about topics such as death and sleep.

Table 4. Sample lies from male and female participants for overlapping semantic classes

.

Class	Male	Female
Negate	I don't have a steady job.	I do not lie.
	I do not work out at gym.	Sports are not dangerous.
	I don't own a cellphone.	I do not love shopping very much.
Anger	I hate polygamy	I hate dogs.
	I killed someone last night.	I have never told a lie.
	I hate my little brother and think he is annoying.	Blue eyed people are dangerous
Friends	I like the neighbors.	I have three boyfriends.
	Ricky martin has a lovely girlfriend.	The neighbors across the street have twelve children.
	I don't care about my friends.	I have never made friends over the Internet.
Eating	I woke up this morning and had breakfast in bed.	You will lose weight if you buy my diet chocolate pies.
	Eating a mushroom every morning in empty stomach helps to reduce weight	I am going to eat at a restaurant that pays me to eat there.
	You can lose weight without exercising or changing your diet.	People eat breakfast at night.

Table 5. Sample lies from male and female participants for non-overlapping semantic classes

Class	Sample lies
	Male
Dead	My girlfriend committed suicide yesterday.
	Drinking cyanide doesn't kill humans.
Sleep	The kids go to bed without any trouble
	I will go to sleep early today.
Other	Superman has the letter "m" on his chest.
.	My friend can smell with the help of his fingers.
	Female
Future	I am excited about going to work tomorrow.
	Obamacare may be obtained after extensive genetic testing.
Sports	I love to watch football, i have never missed a game.
	Shaq o'neill is a famous tennis player.
Motion	I learned driving and used to drive at the age of 5.
	I am able to fly a plane.

To further analyze the word usage by gender, we also obtain the most dominant words per gender using the same method we used before for the semantic classes, but this time applied to the most frequent words used by deceivers. Results from this analysis are reported in Table 6.

From this table, we can observe significant differences in the word usage by each gender. Interestingly, the word safe is present in lies being told by both males and females. From a closer look at the deceptive corpus, we can frequently observe lies such as: "Drinking and driving is a winning and safe combination," "The internet is a totally safe way for children to spend their day, "East los Angeles is a safe place," or "Your privacy is safe on the internet."

Table 6. Top dominant words used per gender in the deception corpus

Male		Female	
Word	Score	Word	Score
Woman	12.39	Dollars	9.93
Really	10.48	Government	6.46
Night	5.40	Times	5.46
Safe	5.24	Never	4.56
he	5.08	Everyone	3.97
Think	4.76	Know	3.97
Always	4.44	Won	3.72
Drinking	4.13	Safe	3.31
His	4.13	Million	2.9
Never	4.08	Always	2.98
They	3.40	Car	2.88

6 Conclusions

In this paper, we presented a set of experiments where we explored the gender detection task in deceptive content. We collected a deception dataset consisting of one-liners truths and lies. Through several experiments, we showed that this data can be used to build deception classifiers able to discriminate between truths and lies. We also explored the gender detection on a fraction of the data consisting of only lies. Our results showed that the female deceivers are more easily detected than males and that classifiers based on unigrams show robust performance. We provided also an analysis of the differences in topics and words used by deceivers from each gender. Our results showed that is more difficult to identify lies than truths. Also, when it comes to gender, lies being told by females are more easily identifiable than lies being told my males. In the future, we are planning to conduct a more detailed analysis where we will study differences related to age and gender perception.

Acknowledgments

This material is based in part upon work supported by National Science Foundation awards #1344257 and #1355633, by grant #48503 from the John Templeton Foundation, and by DARPA-BAA-12-47 DEFT grant #12475008. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the John Templeton Foundation, or the Defense Advanced Research Projects Agency.

References

1. Almela, A., Valencia-García, R., Cantos, P.: Seeing through deception: A computational approach to deceit detection in written communication. In: Proceedings of the Workshop on Computational Approaches to Deception Detection. pp. 15–22. Association for Computational Linguistics, Avignon, France (April 2012), http://www.aclweb.org/anthology/W12-0403
2. DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological Bulletin 129(1) (2003)
3. Dreber, A., Johannesson, M.: Gender differences in deception. Economics Letters 99(1), 197–199 (2008)
4. Feng, S., Banerjee, R., Choi, Y.: Syntactic stylometry for deception detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2. pp. 171–175. ACL '12, Association for Computational Linguistics, Stroudsburg, PA, USA (2012), http://dl.acm.org/citation.cfm?id=2390665.2390708
5. Fornaciari, T., Poesio, M.: Automatic deception detection in italian court cases. Artificial Intelligence and Law 21(3), 303–340 (2013), http://dx.doi.org/10.1007/s10506-013-9140-4

6. Guadagno, R.E., Okdie, B.M., Kruse, S.A.: Dating deception: Gender, online dating, and exaggerated self-presentation. Comput. Hum. Behav. 28(2), 642–647 (Mar 2012), http://dx.doi.org/10.1016/j.chb.2011.11.010
7. Ho, S.M., Hollister, J.M.: Guess who? an empirical study of gender deception and detection in computer-mediated communication. Proceedings of the American Society for Information Science and Technology 50(1), 1–4 (2013)
8. Kaina, J., Ceruti, M.G., Liu, K., McGirr, S.C., Law, J.B.: Deception detection in multicultural coalitions: Foundations for a cognitive model. Tech. rep., DTIC Document (2011)
9. Lu, X.: Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics 15(4), 474–496 (2010)
10. Mihalcea, R., Strapparava, C.: The lie detector: Explorations in the automatic recognition of deceptive language. In: Proceedings of the Association for Computational Linguistics (ACL 2009). Singapore (2009)
11. Mihalcea, R., Pulman, S.: Linguistic ethnography: Identifying dominant word classes in text. In: Computational Linguistics and Intelligent Text Processing, pp. 594–602. Springer (2009)
12. Ott, M., Choi, Y., Cardie, C., Hancock, J.: Finding deceptive opinion spam by any stretch of the imagination. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1. pp. 309–319. HLT '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011), http://dl.acm.org/citation.cfm?id=2002472.2002512
13. Pennebaker, J., Francis, M.: Linguistic inquiry and word count: LIWC (1999), erlbaum Publishers
14. Tilley, P., George, J.F., Marett, K.: Gender differences in deception and its detection under varying electronic media conditions. In: Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 1 - Volume 01. pp. 24.2–. HICSS '05, IEEE Computer Society, Washington, DC, USA (2005), http://dx.doi.org/10.1109/HICSS.2005.284
15. Toma, C., Hancock, J., Ellison, N.: Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles. Personality and Social Psychology Bulletin 34(8), 1023–1036 (2008), http://psp.sagepub.com/content/34/8/1023.abstract
16. Verhoeven, B., Daelemans, W.: Clips stylometry investigation (csi) corpus: A dutch corpus for the detection of age, gender, personality, sentiment and deception in text. In: Chair), N.C.C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland (may 2014)
17. Warkentin, D., Woodworth, M., Hancock, J.T., Cormier, N.: Warrants and deception in computer mediated communication. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work. pp. 9–12. ACM (2010)
18. Xu, Q., Zhao, H.: Using deep linguistic features for finding deceptive opinion spam. In: Proceedings of COLING 2012: Posters. pp. 1341–1350. The COLING 2012 Organizing Committee, Mumbai, India (December 2012), http://www.aclweb.org/anthology/C12-2131
19. Yancheva, M., Rudzicz, F.: Automatic detection of deception in child-produced speech using syntactic complexity features. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 944–953. Association for Computational Linguistics, Sofia, Bulgaria (August 2013), http://www.aclweb.org/anthology/P13-1093