Identity Deception Detection

Verónica Pérez-Rosas[1], **Quincy Davenport**[1], **Anna Mengdan Dai**[1],
Mohamed Abouleinen[2] and **Rada Mihalcea**[1]
[1]Computer Science and Electrical Engineering. University of Michigan
[2]Computer and Information Science. University of Michigan, Dearborn
{vrncapr,quincyd,annadai,zmohamed,mihalcea}@umich.edu

Abstract

This paper addresses the task of detecting identity deception in language. Using a novel identity deception dataset, consisting of real and portrayed identities from 600 individuals, we show that we can build accurate identity detectors targeting both age and gender, with accuracies of up to 88%. We also perform an analysis of the linguistic patterns used in identity deception, which lead to interesting insights into identity portrayers.

1 Introduction

With the ever growing usage of social media and other online interactions, cyber-crimes such as identity thief, fraud, and sexual predation have become increasingly common. The availability of a wide variety of social platforms and apps further facilitates these kind of crimes, which are often characterized by the ease of deception and concealment of one's real identity. Moreover, the increased sense of security due to the spatial and temporal "distanciation" involved with online communications leads to a growing number of occurrences of identity deception.

Identity deception is defined as "pretending to be someone you are not" (Rowe, 2009). The intentions of identity deceivers differ between achieving monetary benefits, committing fraud, sexual predating, appearing more attractive in online dating, and so on. The risk is further increased with the massive number of children and teens using social media as well as the elderly who get involved in online interactions that they assume to be trustworthy by default due to their lack of experience. Multiple stories of teenagers who became victims of these activities, as well as elderly who lost hundreds of thousands of dollars are now often encountered in the news. For instance, a recent study reported that 92% of teens go online on daily basis including 24% who are online "almost constantly". [1] The study additionally reported that females tend to use social media more than males.

Hence, due to the financial, social, physical, and psychological damages associated with cyber-crimes, there is a growing need to learn more about the patterns and behaviors of identity deceivers, in order to develop computational approaches that can aid in preventing such crimes.

In this work, we seek to identify linguistic differences in individuals' self-presentation when engaging in identity deception behaviors. We target the two most common behaviors related with online identity deception i.e., individuals portraying themselves as either younger and older, and individuals lying about their gender –also known as gender switching (Macwan and inz. Grzegorz Filcek, 2017; Herring and Martinson, 2004) .

We perform a set of experiments to explore three main research questions. First, given a deceptive corpus consisting of written samples of gender switching, can we build fake identity detectors that predict gender deception? Second, given a deceptive corpus consisting of written samples of age deception, can we build fake identity detectors that predict age deception? Lastly, are there linguistic differences associated with individuals' gender and age in identity deception?

2 Related Work

Several approaches have been developed to detect and prevent identity deception (Tsikerdekis and Zeadally, 2015). Integration of latent textual features with spatial and temporal features has been suggested using probabilistic generative

[1] http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/

Proceedings of the The 8th International Joint Conference on Natural Language Processing, pages 885–894,
Taipei, Taiwan, November 27 – December 1, 2017 ©2017 AFNLP

modelling to detect identity thieves (Wang et al., 2017). Another study analyzed the connection between time traces of stolen accounts and compared them to that of the original users using support vector machines (Villar-Rodrguez et al., 2016).

As linguistic patterns represent the main method for detecting identity theft, previous work used text mining techniques by extracting semi-structured information from online news stories and reports on the topic of identity theft, and identified behavioral and temporal patterns and resources used by identity thieves (Zaeem et al., 2017). An analysis of thieves' behavior in the Massive Multiuser Online Role Playing Game suggested a detection model based on specific sequences including item production, item sales, and acquisition of game money (Kim et al., 2015).

Focusing specifically on the patterns associated with impersonating the opposite gender, a research study using the Turing Game identified stereotypical content performed by the deceivers as well as stylistic cues to their real gender (Herring and Martinson, 2004). The study additionally suggested that linguistic cues referring to gender seem to be unconsciously generated.

One of the common practices of identity thieves is online sexual predation. To identify such behavior, word patterns and search engine query detection were suggested as means for detecting pedophilic activity (Macwan and inz. Grzegorz Filcek, 2017). Examples include the detection of predators using lexical and behavioral features and calculating the predator-hood score as a function of features weights (Dhouioui and Akaichi, 2016). A method was developed to identify sexual predation using phrase-matching and rule-based systems and reported the usefulness of statement lengths in chat lines for improving the identification process (Mcghee et al., 2011). A study (Bogdanova et al., 2012) analyzing a corpus of chats for detecting cyberpedophilia found that character n-grams are capable of discriminating pedophiles' chats. However, higher-level features that modeled behavior and emotion were required to detect conversations with cyberpedophiles from cybersex chat logs (Bogdanova et al., 2014).

An analysis of twenty chat rooms to detect sex offenders indicated eight recurrent themes, including 'implicit and explicit content of discourse', 'on-line solicitation', 'fixated discourse', 'use of colloquialisms', 'conscience', 'acknowledging illegal/immoral behavior', 'risk minimization', and 'preparing to meet offline' (Egan et al., 2011). Other work used automatic text categorization techniques with a support vector machine and distance weighted k-NN classifiers to distinguish between a sexual predator and a pseudo victim using data collected from volunteers posing as underage victims (Pendar, 2007). A language processing module was developed in order to successfully differentiate between porn titles and titles of encyclopedia articles using support vector machine and linear regression (Panchenko et al., 2012).

More recently, a conversational agent was developed by mimicking the behavior of a teenager and inferring its conversational rules from a real dialogue corpus in order to detect cyberpedophilia (Callejas-Rodríguez, 2016). The study reported a successful approach using a combination of the least frequent word and the most frequent bigram.

In summary, most of the previous work has focused on specific topics and in particular sexual predation and gaming, and relied primarily on limited text mining approaches. Instead, our approach uses a novel dataset where the participants are not restricted from discussing any topics, and hence we can identify general patterns of identity deception related to both gender and age. Moreover, an advantage of our dataset is the knowledge of the real identity and demographic information of the participants. Furthermore, we conduct multiple experiments using a wide variety of linguistic features in order to explore and analyze the textual clues that identify identity deceivers.

3 Data Collection

We seek to examine written samples of individuals presenting themselves with their real identity as well as a fake identity. To achieve this, we collect a corpus of writings from several participants, including responses to open-ended questions about their real identity, as well as about a pre-assigned fake identity. We target four fake identities: 1) 18-year-old female, 2) 18-year-old male, 3) 65-year-old female, 4) 65-year-old male. The choice of the fake identities was based on two practical considerations. First, the most prominent form of online identity deception is gender switching, thus we seek to collect samples of individuals portraying themselves as being from the opposite sex. Second, 18 and 65 years are extreme values on the age

FAKE IDENTITY	REAL IDENTITY
I'm Ashley, and I'm currently enrolled in a freshman in the nursing program at my university. My family is very important to me. I have two sisters, both of whom are younger than me. I do volunteer work in my free time through my church I enjoy hanging out with my friends, traveling (I get a lot of chances to do this through mission trips I go on). I want a husband and family, but my career comes first. I'm excited to see what new opportunities will appear on life's path.	I am a guy in his mid-30's. I have a cat and live with two roommates. I am self employed, selling various items on eBay as well as doing my Mechanical Turk work. I love music (alt and indie rock, electronica, and 80's especially), concerts, karaoke, video games, DC Comics, technology, reading, streaming TV and movies, camping and hiking, exploring the city, and good food and coffee. I'm currently involved in various projects to improve my life.
I am a male, aged 65 years old. I have a wife who is 59 years old and two grown children ages 30 and 32, both boys. I worked for Boeing making airplanes for thirty years before I was injured and needed shoulder surgery. I retired just recently and teach an after school class about airplanes and how they are built. I check the mail at 3 pm each day and go to the grocery store four times a week, my wife picks me up a 6 pack of Ice House nightly ...	I am a 35 year old female with bleached blonde hair and hazel eyes. I am a graduate in Public Admin and I enjoy bodybuilding. I have four children and a husband and we live in Seattle. I like to take hikes outside in the mountains but not when it is too hot outside. I am from Maine originally but lived in florida for ten years and hated it. The weather is too hot and there are too many snakes and lizards crawling all over your yard.

Table 1: Sample responses from our dataset. Top row: 35-year-old male acting as an 18-year-old female; bottom row: 35-year-old female acting as a 65-year-old male

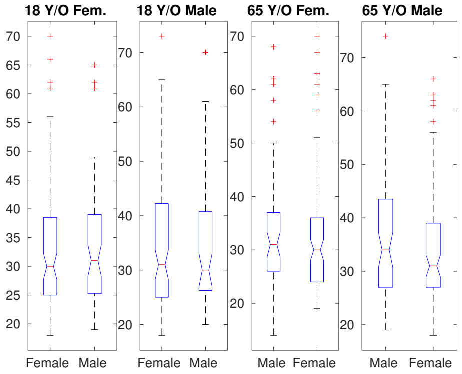

Figure 1: Participants' age clustered by gender when assuming four fake identities

distribution of Internet users, and this can allow us to explore deceptive behaviors in individuals portraying themselves as being younger or older than they actually are.

We designed four surveys using the Qualtrics survey software and distributed them via Mechanical Turk. In each survey, participants were asked to respond to the following prompt:

1. *Describe yourself.*
2. *Discuss any topic that comes to your mind.*
3. *Imagine you are trying to convince someone that you are a* 18 | 65 *year old male* | *female. Briefly describe yourself, pretending you are a* 18 | 65 *year old male* | *female.*
4. *Imagine you are trying to convince someone that you are a* 18 | 65 *year old male* | *female. Discuss any topic that comes to your mind as if pretending you are a* 18 | 65 *year old male* | *female.*

Note that the participants are asked to respond to two open-ended questions when presenting themselves with their real and fake identity: one asking for a self-description, and another one asking for an essay on a topic of choice. Our motivation for this setup is to avoid stereotypical responses for each target identity, i.e., descriptions focusing on physical appearance and activities frequently associated to the target gender and age. Two sample responses (self-description only) from individuals posing as female and male are shown in Table 1.

We also collect demographic data from participants, including their gender, age, education, primary language, ethnicity, country of origin, and country of residence. During the data collection, participants self-reported their gender, using the male/female options. In the future, with larger-scale data collections, we will be able to work with an extended set of gender options. Also, note that individual's true identity is self-reported so we assume workers to provide accurate information – previous studies analyzing Mechanical Turk as a participant's pool show that this is often the case (Paolacci and Chandler, 2014).

In order to obtain responses with a reasonable amount of text, we constrain the response's length to be at least 400 characters. Additionally, we

Fake Identity	Respondents	Male	Female	Age (M,SD)
18 Y/O Fem.	151	41.7%	58.3%	(33.24, 10.82)
18 Y/O Male	154	35.7%	64.3%	(34.24,11.62)
65 Y/O Fem.	150	38.0%	62.0%	(33.42, 12.03)
65 Y/O Male	149	42.3%	57.7%	(35.07,11.96)

Table 2: Gender distribution and age statistics for four fake identities.

manually check for coherence and relevance to the prompt. We reject contributions that failed to follow the provided guidelines and did not pass the manual verification.

After this filtering, we obtained a total of 604 completed surveys, each of them containing descriptions of participants' real and fake identities, as well as their corresponding demographic data. The data statistics are shown in Table 2.

Figure 1 shows the participants real age, clustered by gender, when assuming different fake identities. We observe that the average age of the respondents ranges between 33-35 year-old, thus suggesting a reasonable distance between their actual and fake age. In addition, the graph suggests differences in how deceivers portray themselves given their actual gender, which we further explore in section 5.

4 Features

In this section, we describe the sets of features extracted, which are used to build our classifiers.

Unigrams We extract unigrams and bigrams derived from the bag of words representation of each identity response.

POS These features consist of part-of-speech (POS) tags obtained with the Stanford Parser (Chen and Manning, 2014).

Semantic LIWC These features include the 74 semantic classes present in the LIWC lexicon 2015 (Pennebaker et al., 2015). Each feature represents the number of words in a response belonging to a specific semantic class, normalized with respect to the length of the response.

Semantic Word2Vec To obtain these features, we use the word2vec (Mikolov et al., 2013) implementation available in the Gensim toolkit (Řehůřek and Sojka, 2010) to obtain word vectors with dimension 300 for each word in the responses. The final identity vector is calculated by adding all the word vectors in the response.

Lexical diversity This set includes four lexical diversity metrics, including type/token ratio (McCarthy and Jarvis, 2010), mean word frequency, and the Yule's I and K indexes (Oakes, 2000).

Readability We also extract features that indicate text understandability. These include readability metrics such as the Flesch-Kincaid, Flesch Reading Ease, Gunning Fog, and the Automatic Readability Index (ARI) (Kincaid et al., 1975; Senter and Smith, 1967)

5 Experimental Results

We conduct several experiments to answer the research questions formulated at the beginning of this paper. During our experiments, we perform the evaluations at individual level, by merging the two open-ended questions asked during the survey (i.e., the self-description and topic-of-choice essay), for both the real and the fake identities. Also, given the contributors' age distribution shown in Figure 1, we opted to cluster the participants age into into two groups: young ($\leq$ 30 years) and old ($>$30 years).

The classifiers are built using the SVM algorithm[2] and the different sets of features described in section 4. We perform leave-one-out cross-validation in all our experiments. In all cases, we use the majority class baseline as a reference value.

5.1 Classification of Real and Fake Identities

Before focusing on our main research questions, we seek to evaluate whether deception detection can be conducted using our fake identity dataset. Thus, we focus on two main classification tasks. First, using our entire dataset, we explore whether we can discriminate between the portrayed identities and the real identities. Second, we once again attempt to discriminate between real and fake identities, but this time filtering by either age or gender.

[2] As implemented in the LIBLINEAR library, using (L2) SVM classification.

Features	All Identities	Gender		Age	
		Female	Male	Young	Old
Baseline	50.12	49.93	50.21	49.90	50.14
Ngrams	86.67	86.59	84.25	85.35	84.31
POS	66.61	68.39	68.93	68.97	67.30
LIWC	63.28	64.70	64.89	60.69	66.56
Word2Vec	77.51	77.70	72.34	77.26	75.51
LexDiv	49.45	50.47	46.38	46.62	48.09
Readability	54.55	52.25	51.16	51.55	54.45
All features	85.76	87.50	85.10	85.74	85.48

Table 3: Classification results for deception detection regardless of gender or age (All Identities); for people with a certain gender, according to their real identity (Gender); and people with a certain age according to their real identity (Age).

		Identity		
		Fake	Real	Total
Gender	Male	234	236	470
	Female	355	366	721
Age	Old	340	342	682
	Young	259	260	519
All identities		599	602	1201

Table 4: Class distribution for all identities, filtered by gender, and filtered by age.

To perform these experiments, we start by creating five subsets from all the data. The first one consists of all the real and fake identities that we collected; the second and third one consist of responses filtered by the individuals' actual gender, i.e., male and female; and the last two consist of responses filtered by the individuals' actual age, i.e., old and young. During this process, we discard those instances where the fake identity overlaps with the actual identity. The class distribution of the resulting subsets is shown in Table 4.

We then build classification models that attempt to discriminate between the real and fake identities under the following scenarios: a) having no previous knowledge of the actual individual's age or gender, i.e., all data; b) knowing that the individual's actual gender is female; c) knowing that the individual's actual gender is male; d) knowing that the individual's actual age is 30 years and under; and e) knowing that the individual's actual age is over 30 years.

Classification results and the corresponding majority class baselines are shown in Table 3. For most of the prediction tasks, except for the first one, the best performing feature set is the combination of all features (*All features*), followed by *Ngrams*. The remaining sets of features achieve accuracy values ranging from 63% to 77%, which still represent a noticeable improvement over the majority class baseline (the only exception is the lexical diversity feature set).

Real Identity	Fake Identity	
	Male	Female
Male	185	181
Female	116	120
Total	301	301

Table 5: Class distribution for gender deceivers

While the overall results are largely similar across the five experiments, note that the size of the datasets used in the gender and age-filtered experiments is much smaller than the one used in the All Identities experiment. This suggests that the gender (or age) of the writer plays an important role in this classification, and the gender (or age) characteristics can make up for the smaller data size.

5.2 Classification of Gender Deceivers

Motivated by our previous findings, we investigate our first research question. Can we build fake identity detectors that predict gender deception? This question focuses on the scenario in which we are interested in knowing if a text whose author claims to be of a certain gender is indeed authored by that particular gender. E.g., if the author claims to be a female, is the writer indeed a female or a male? This can be useful in the verification of user profiles in dating websites, where gender switching is a common form of deception.

For this classification task, we use only the data from the portrayed identities. That is, we use the responses from the 16 Y/O male, 65 Y/O male, 16 Y/O female, and 65 Y/O female identities. While we could have potentially also used the real identities, we chose to focus only on the portrayed ones so that we obtain a more consistent dataset. The distribution for this subsets is shown in Table 5.

For each portrayed gender, the classifier aims to

Features	Fake Identity		
	Female+Male	Female	Male
Baseline	50.00	60.00	61.00
Ngrams	86.20	82.72	87.04
LIWC	70.40	66.11	73.08
Word2Vec	75.90	63.78	65.11
POS	67.40	64.45	66.77
LexDiv	54.30	57.14	54.15
Readability	61.29	54.48	71.42
All features	86.04	83.05	87.70

Table 6: Classification results for gender deceivers, overall (Female+Male column) and broken down by portrayed gender.

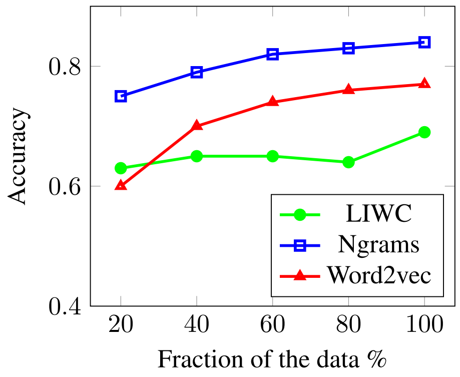

Figure 2: Learning curves on gender deceivers classification using three feature sets

predict the real gender. Classification accuracies for the resulting models are shown in Table 6. This table also shows the results of the models built using the portrayed male and female responses separately. In these experiments, the *Ngrams* features outperform all the other feature sets. The second best performing features are the ones based on the LIWC lexicon, followed by the part-of-speech features *POS*. Overall, the results suggest that it is easier to identify females portraying themselves as males (*Male* identity column) than males portraying themselves as females (*Female* identity column).

As an additional experiment, we investigate whether larger amounts of training data can improve the identification of gender deceivers. We plot the learning curves of the best performing sets of features using incremental amounts of data as shown in Figure 2. In this graph, the learning trends for the *LIWC* and *Word2Vec* features suggest that larger amounts of training data could improve the classification performance.

Real Identity	Fake Identity	
	Young	Old
Young	168	174
Old	137	123
Total	305	297

Table 7: Class distribution for age deceivers

Features	Fake Identity		
	Young+Old	Young	Old
Baseline	51.00	55.00	58.00
Ngrams	83.38	86.22	83.83
LIWC	71.26	73.11	67.67
Word2Vec	77.24	73.77	65.99
POS	69.76	63.27	64.30
LexDiv	49.66	50.16	56.90
Readability	60.32	60.54	53.87
All features	82.72	87.21	81.81

Table 8: Classification results for age deceivers overall, and broken down by portrayed age.

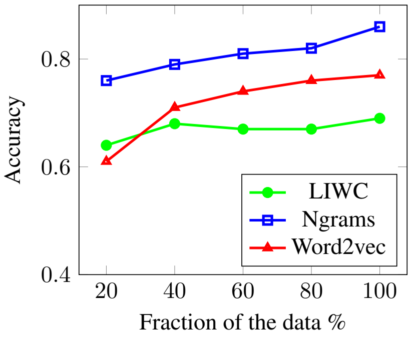

Figure 3: Learning curves on age deceivers classification using three feature sets

5.3 Classification of Age Deceivers

Next, we focus our attention on identifying age deceivers to answer our second research question: can we build fake identity detectors that predict age deception? This time we focus on the scenario where the author of a text claims to be either young or old (using our earlier definition of young/old), and we want to determine if the real age of the writer is indeed corresponding to their claims. This can be particularly useful in the identification of sexual predators who target younger people, or scammers who target elderly people.

For this classification task, we once again use only the data from the portrayed identities, i.e., we use only the responses corresponding to the portrayed 16 Y/O male, 65 Y/O male, 16 Y/O female, and 65 Y/O female identities. The distribution for this dataset is shown in Table 7.

For each portrayed age range (young or old), the classifier aims to predict the real age range. Clas-

Female as a Male		Real Female		Male as a Female		Real Male	
SWEAR	2.35	SEXUAL	2.03	FILLER	4.07	ACHIEV	1.57
FILLER	2.35	ANX	1.48	FEMALE	3.16	MONEY	1.51
ASSENT	2.15	INGEST	1.44	SWEAR	2.29	WORK	1.50
FRIEND	1.96	BIO	1.36	FAMILY	1.96	SEXUAL	1.36
INFORMAL	1.81	RISK	1.31	BODY	1.94	DEATH	1.34
YOU	1.75	HEALTH	1.28	NETSPEAK	1.91	RISK	1.29
DEATH	1.70	NUMBER	1.20	FRIEND	1.81	RELIG	1.27
NETSPEAK	1.61	RELIG	1.19	YOU	1.67	CAUSE	1.24
MOTION	1.50	INSIGHT	1.18	MALE	1.63	INSIGHT	1.23
FOCUSFUTURE	1.43	ADJ	1.17	INFORMAL	1.60	IPRON	1.18

Table 9: Top ranked semantic classes from the LIWC lexicon associated to gender impersonators and actual gender.

Younger as Older		Real Young		Older as Younger		Real Old	
YOU	2.02	ASSENT	1.76	NETSPEAK	5.43	RELIG	2.71
FAMILY	2.00	ANX	1.63	FILLER	4.48	INGEST	1.92
DEATH	1.86	ACHIEV	1.56	SWEAR	4.05	HEALTH	1.84
FILLER	1.79	NETSPEAK	1.47	FRIEND	3.05	SEXUAL	1.75
FEMALE	1.72	WORK	1.34	INFORMAL	2.82	RISK	1.73
THEY	1.66	HEAR	1.34	ASSENT	2.32	NUMBER	1.59
FOCUSPAST	1.63	LEISURE	1.25	FOCUSFUTURE	1.84	DEATH	1.56
TIME	1.42	ANGER	1.25	YOU	1.62	SAD	1.48
SEXUAL	1.39	NEGEMO	1.23	MOTION	1.51	BIO	1.48
HEALTH	1.33	POSEMO	1.18	HEAR	1.42	ANX	1.26

Table 10: Top ranked semantic classes from the LIWC lexicon associated to age impersonators and actual age.

sification accuracies for the resulting models are shown in Table 8. This table also shows the results of models built using the portrayed young or portrayed old responses separately. In these experiments, the *Ngrams* features also outperform all the other feature sets. The second best performing features are the ones based on the semantic vector obtained with *Word2vec*, followed by *LIWC* and part-of-speech *POS* features. Overall, the results suggest that it is easier to identify older individuals portraying themselves as being younger (*Young* identity column) than younger individuals portraying themselves as being older (*Old* identity column).

To explore whether more training data would be beneficial to improve classifiers performance, we plot the learning curves of the bests sets of features using incremental amounts of data as shown in Figure 3. As observed, all feature sets show a positive learning trend suggesting that more training data might improve the performance in the age deception task.

5.4 Analysis of Linguistic Differences Associated to Gender and Age Deceivers

Seeking to answer our third research question: are there linguistic differences associated to individuals' gender and age in identity deception? we analyze differences in word usage that might reveal the real identity of age and gender impersonators.

From the gender-based analysis, we use responses from actual females writing as males and responses from actual males writing as females as well as their truthful self-descriptions. Similarly, in the age-based analysis, we consider the responses from younger respondents writing as older and older respondents writing as younger as well as their truthful self-descriptions.

Our analyses are based on the semantic word classes from the LIWC lexicon and the semantic word-class scoring by (Mihalcea and Pulman, 2009). Tables 9 and 10 show the top classes for each deception group and their real identities.

The analyses reveal interesting word usage patterns among gender impersonators. On the one hand, when females pose as males they use more 'swear', 'fillers', and 'informal talk' words. On the other hand, males that impersonate females use more 'fillers', 'female', and 'family' words. When looking at the word associations for actual genders, it would seem that there is no clear relation with how males and females portray each other when faking their gender. We believe that this could be attributed to gender-related stereotypes and biases. In the age deceiver case, the younger individuals who portray themselves as older use

	Real Female	Male as Female	Real Male	Female as Male
Real Female	1			
Male as Female	0.670**	1		
Real Male	0.663**	0.994**	1	
Female as Male	0.668**	0.999**	0.995**	1

Table 11: Correlation of LIWC classes across real and fake gender identities. ** Correlation is significant at 0.001 level (2-tailed)

	Real Young	Old as Young	Real Old	Young as Old
Real Young	1			
Old as Young	0.973**	1		
Real Old	0.975**	0.996**	1	
Young as Old	0.976**	0.997**	0.994**	1

Table 12: Correlation of LIWC clases across real and fake age identities. ** Correlation is significant at 0.001 level (2-tailed)

more 'you', 'family', 'death' and 'filler' words. In contrast, older individuals who portray themselves as younger use more internet slang 'netspeak', 'fillers', and 'swear' words. Similar to the gender findings, age deception seems to be affected by stereotypes.

For further insights into the linguistic differences between gender and age impersonators, we also calculate the Pearson correlation of the LIWC class counts across actual and corresponding fake identities. We use word class counts normalized by the number of words in the sentence.

The correlation between *Real Female and Real Male* shown in Table 5.3 present an estimate of how similar the actual male and female writings are. We observe a positive mid-strength correlation as compared with the other correlation pairs. The *Real Male vs Female as Male* correlation suggests that females are good at emulating males' writing. In contrast, the analysis shows that males are not as good emulating female language (see correlation of *Real Female vs Male*).

Surprisingly, the analysis of age deceivers in Table 5.3 shows a strong correlation trend between the different age-based identities. In particular, the correlation between *Real Old and Young* suggests high language similarity between the two groups. Furthermore, the correlation between the *Real Old vs Old as Young* and *Real Young vs Young as Old* indicates that in general, older people are good at imitating younger people and vice versa.

6 Conclusions

In this paper, we addressed the task of identity deception detection. We collected a novel identity deception dataset, consisting of individuals portraying themselves with four fake identities, targeting different ages and genders.

Through several experiments, we showed that we can build accurate identity detectors. Specifically, we focused on the prediction of gender and age impersonators. We were able to identify identity deceivers with accuracies up to 88%. Our main findings showed that it is easier to identity females posing as males and similarly, it is easier to identify older individuals posing as younger individuals.

Furthermore, we presented a statistical analysis of linguistic patterns that differentiate between fake and real identities based on age and gender.

The datasets introduced in this paper are publicly available under http://lit.eecs.umich.edu/downloads.html.

Acknowledgments

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

References

Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. 2012. On the impact of sentiment and emotion based features in detecting online sexual predators. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*. Association for Computational Linguistics, Stroudsburg, PA, USA, WASSA '12, pages 110–118.

Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. 2014. Exploring high-level features for detecting cyberpedophilia. *Computer Speech & Language* 28(1):108 – 120.

Ángeland Villatoro-Tello Esaúand Meza Ivanand Ramírez-de-la-Rosa Gabriela Callejas-Rodríguez. 2016. *From Dialogue Corpora to Dialogue Systems: Generating a Chatbot with Teenager Personality for Preventing Cyber-Pedophilia*, Springer International Publishing, Cham, pages 531–539.

Danqi Chen and Christopher D Manning. 2014. A fast and accurate dependency parser using neural networks. In *Emnlp*. pages 740–750.

Zeineb Dhouioui and Jalel Akaichi. 2016. Privacy protection protocol in social networks based on sexual predators detection. In *Proceedings of the International Conference on Internet of Things and Cloud Computing*. ACM, New York, NY, USA, ICC '16, pages 63:1–63:6. https://doi.org/10.1145/2896387.2896448.

V. Egan, J. Hoskinson, and D. Shewan. 2011. *Perverted justice: a content analysis of the language used by offenders detected attempting to solicit children for sex*, Nova Science Publishers, Inc, pages 273–297.

Susan C. Herring and Anna Martinson. 2004. Assessing gender authenticity in computer-mediated language use. *Journal of Language and Social Psychology* 23(4):424–446. https://doi.org/10.1177/0261927X04269586.

Hana Kim, Byung Il Kwak, and Huy Kang Kim. 2015. A study on the identity theft detection model in mmorpgs. *Journal of the Korea Institute of Information Security and Cryptology* 25(3):627–637.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch.

Shweta Macwan and Dr. inz. Grzegorz Filcek. 2017. Web mining against pedophilia. *International Education and Research Journal* 3(5).

Philip M McCarthy and Scott Jarvis. 2010. Mtld, vocd-d, and hd-d: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods* 42(2):381–392.

India Mcghee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra Mcbride, and Emma Jakubowski. 2011. Learning to identify internet sexual predation. *International Journal on Electronic Commerce* 15(3):103–122. https://doi.org/10.2753/JEC1086-4415150305.

Rada Mihalcea and Stephen Pulman. 2009. Linguistic ethnography: Identifying dominant word classes in text. In *Computational Linguistics and Intelligent Text Processing*, Springer, pages 594–602.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* .

Michael P. Oakes. 2000. Statistics for corpus linguistics. *International Journal of Applied Linguistics* 10(2):269–274.

Alexander Panchenko, Richard Beaufort, and Cédrick Fairon. 2012. Detection of Child Sexual Abuse Media on P2P Networks: Normalization and Classification of Associated Filenames. In Zygmunt Vetulani and Edouard Geoffrois, editors, *Proceedings of the Language Resources for Public Security Workshop 2012 (LRPS 2012) at LREC 2012*. Instanbul, Turkey, pages 27–31.

Gabriele Paolacci and Jesse Chandler. 2014. Inside the turk. *Current Directions in Psychological Science* 23(3):184–188. https://doi.org/10.1177/0963721414531598.

Nick Pendar. 2007. Toward spotting the pedophile telling victim from predator in text chats. In *Proceedings of the International Conference on Semantic Computing*. IEEE Computer Society, Washington, DC, USA, ICSC '07, pages 235–241. https://doi.org/10.1109/ICSC.2007.102.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, pages 45–50.

Neil C Rowe. 2009. *The Ethics of Deception in Cyberspace*, IGI Global, pages 529–541.

RJ Senter and Edgar A Smith. 1967. Automated readability index. Technical report, CINCINNATI UNIV OH.

Michael Tsikerdekis and Sherali Zeadally. 2015. Detecting and preventing online identity deception in social networking services. *IEEE Internet Computing* 19(3):41–49. https://doi.org/10.1109/MIC.2015.21.

Esther Villar-Rodrguez, Javier Del Ser, Ana I. Torre-Bastida, Miren N. Bilbao, and Sancho Salcedo-Sanz. 2016. A novel machine learning approach to the detection of identity theft in social networks based on emulated attack instances and support vector machines. *Concurrency and Computation: Practice and Experience* 28(4):1385–1395. Cpe.3633. https://doi.org/10.1002/cpe.3633.

Chen Wang, Bo Yang, and Jing Luo. 2017. Identity theft detection in mobile social networks using behavioral semantics. In *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*. pages 1–3.

Razieh Nokhbeh Zaeem, Monisha Manoharan, Yongpeng Yang, and K. Suzanne Barber. 2017. Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security* 65:50–63.