Detecting Deception

Lecture by: Myle Ott[1]
Incl. joint work with: Claire Cardie,[1,2] Yejin Choi,[1] Jeff Hancock[2,3]
Depts of C.S.,[1] I.S.,[2] Comm.[3]
Cornell University, Ithaca, New York

Background

- Language use varies:
  - By location
    - soda vs. pop vs. coke
    - "koo" vs. "coo" (Eisenstein et al., 2010; 2011)
    - Also Johnstone (2010), Mei et al. (2006; 2007), Labov et al. (2006), Tagliamonte (2006), …

Background

- Language use varies:
  - By genre
    - British National Corpus: Koppel et al. (2002), Rayson et al. (2001), Biber et al. (1999), …
    - Web: Mehler et al. (2010), Rehm et al. (2008), …
    - Twitter: Westman and Freund (2010), …

Background

- Language use varies:
  - By the author's gender
    - British National Corpus: Koppel et al. (2002), …
    - Blogs: Mukherjee and Liu (2010), …
    - Twitter: Burger et al. (2011), …
    - Cross-topic/domain: Sarawgi et al. (2011)

Background

- Language use varies:
  - By the author's beliefs, feelings, opinions
    - Opinion mining and sentiment analysis: Pang and Lee (2008), …
    - Belief annotation and tagging: Prabhakaran et al. (2010), Diab et al. (2009), …
    - Detecting hedges: CoNLL 2010 Shared Task, …

Background

- Language use varies:
  - By whether the author is being truthful or deceptive
  - Studies have considered deception involving:
    - Emotional states: Ekman and Friesen (1969), …
    - Views on social issues, e.g., death penalty: Newman et al. (2003), Mihalcea and Strapparava (2009), …
    - Online dating profiles: Hancock et al. (2007), …
    - Online product reviews: Ott et al. (2011; 2012), …
    - …

Outline

- Briefly go over a few important studies and meta-analyses of deception:
  - Bond and DePaulo (2006)
  - Newman et al. (2003)
  - Vrij (2008)
- Case study on detecting deceptive online reviews of hotels: Ott et al. (2011)

Bond and DePaulo (2006)

- Meta-analysis of over 200 studies of deception
- Finds that human judges are relatively bad at detecting deception, with an average accuracy of just 54%
- Poor performance due in part to truth-bias
  – Human judges are more likely to erroneously judge something as truthful than erroneous judge something as deceptive

Newman et al. (2003)

- Hundreds of true and false verbal and written samples from undergraduates across three subjects: stance on abortion, feelings about friends, and a mock crime
- Language analyzed using the Linguistic Inquiry and Word Count (LIWC) software, developed by James Pennebaker (a co-author of the study)

Newman et al. (2003)

- LIWC
  - Counts instances of ~4,500 keywords
    - Regular expressions, actually
  - Keywords are divided into 80 psycholinguistically-motivated dimensions across 4 broad groups
  - Reports means and standard deviations

Newman et al. (2003)

- LIWC
  - Linguistic processes
    - e.g., average number of words per sentence
  - Psychological processes
    - e.g., talk, happy, know, feeling, eat
  - Personal concerns
    - e.g., job, cook, family
  - Spoken categories
    - e.g., yes, umm, blah

Newman et al. (2003)

- LIWC
  - Linguistic processes
    - e.g., average number of words per sentence
  - Psychological processes
    - e.g., talk, happy, know, feeling, eat
  - Personal concerns
    - e.g., job, cook, family
  - Spoken categories
    - e.g., yes, umm, blah

Newman et al. (2003)

- LIWC
  - Linguistic processes
    - e.g., average number of words per sentence
  - Psychological processes
    - e.g., talk, happy, know, feeling, eat
  - Personal concerns
    - e.g., job, cook, family
  - Spoken categories
    - e.g., yes, umm, blah

Newman et al. (2003)

- LIWC
  - Linguistic processes
    - e.g., average number of words per sentence
  - Psychological processes
    - e.g., talk, happy, know, feeling, eat
  - Personal concerns
    - e.g., job, cook, family
  - Spoken categories
    - e.g., yes, umm, blah

Newman et al. (2003)

- Results showed that deceptive samples have:
  - Reduced first-person singular (psychological distancing)
    - Liars avoid taking ownership of their lies, either to "dissociate" or due to a lack of personal experience
  - Increased negative emotion words
    - Possibly due to discomfort and guilt about lying
  - Reduced complexity and less exclusive language
    - Possibly due to increased cognitive load

Vrij (2008)

- Comprehensive review of the current state of deception detection research
- In addition to the previous findings:
  - Meta-analysis of 30 studies shows that deceivers have difficulty encoding spatial and temporal information into their deceptions

Outline

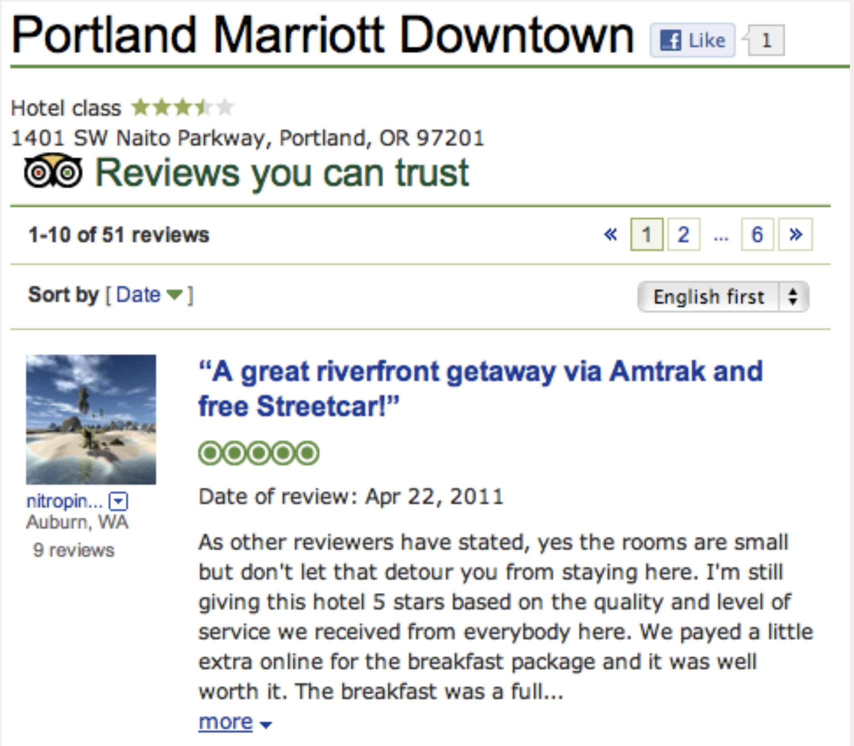
- Briefly go over a few important studies and meta-analyses of deception:
  - Bond and DePaulo (2006)
  - Newman et al. (2003)
  - Vrij (2008)
- Case study on detecting deceptive online reviews of hotels: Ott et al. (2011)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

Myle Ott,[1] Yejin Choi,[1] Claire Cardie,[1] and Jeff Hancock[2]
Dept. of Computer Science,[1] Communication[2]
Cornell University, Ithaca, NY

Motivation

- Consumers increasingly rate, review and research products online
- Potential for opinion spam
  - Disruptive opinion spam
  - Deceptive opinion spam



Motivation

- Consumers increasingly rate, review and research products online
- Potential for opinion spam
  - Disruptive opinion spam
  - Deceptive opinion spam

Motivation

- Consumers increasingly rate, review and research products online
- Potential for opinion spam
  - Disruptive opinion spam
  - Deceptive opinion spam

Motivation

- Consumers increasingly rate, review and research products online
- Potential for opinion spam
  - Disruptive opinion spam
  - Deceptive opinion spam

Motivation

Which of these two hotel reviews is deceptive opinion spam?

Date of review: Jun 9, 2006

4 people found this review helpful

I have stayed at many hotels traveling for both business and pleasure and I can honestly stay that The James is tops. The service at the hotel is first class. The rooms are modern and very comfortable. The location is perfect within walking distance to all of the great sights and restaurants. Highly recommend to both business travellers and couples.

Date of review: Jun 9, 2006

4 people found this review helpful

My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful!! The area of the hotel is great, since I love to shop I couldn't ask for more!! We will definatly be back to Chicago and we will for sure be back to the James Chicago.

Motivation

Which of these two hotel reviews is deceptive opinion spam?

Answer:

Date of review: Jun 9, 2006

4 people found this review helpful

My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful!! The area of the hotel is great, since I love to shop I couldn't ask for more!! We will definatly be back to Chicago and we will for sure be back to the James Chicago.

Overview

- Motivation
- Gathering Data
- Human Performance
- Classifier Performance
- Conclusion

Gathering Data

- Label existing reviews
  - Can’t manually do this
  - Duplicate detection (Jindal and Liu, 2008)
- Create new reviews
  - Mechanical Turk

Gathering Data

- Label existing reviews
  - Can't manually do this
  - Duplicate detection (Jindal and Liu, 2008)
- Create new reviews
  - Mechanical Turk

Gathering Data

- Label existing reviews
  - Can’t manually do this
  - Duplicate detection (Jindal and Liu, 2008)
- Create new reviews
  - Mechanical Turk

Gathering Data

- Label existing reviews
  - Can't manually do this
  - Duplicate detection (Jindal and Liu, 2008)
- Create new reviews
  - Mechanical Turk

Gathering Data

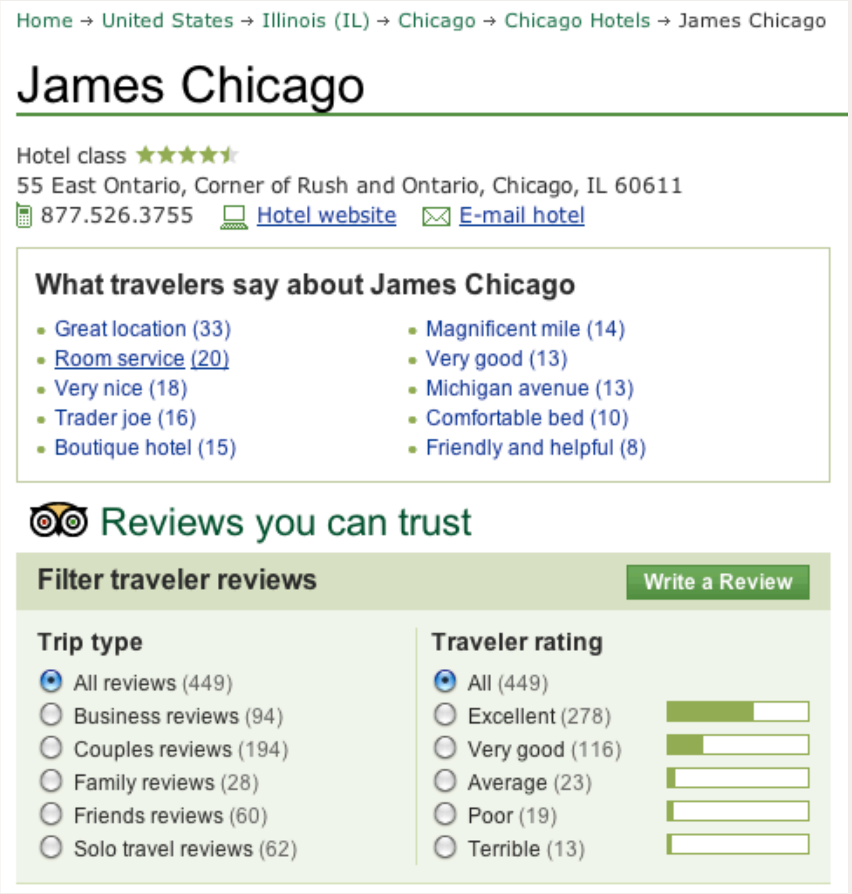
- Label existing reviews
  - Can't manually do this
  - Duplicate detection (Jindal and Liu, 2008)
- Create new reviews
  - Mechanical Turk

Gathering Data

- Mechanical Turk
  - 20 hotels
  - 20 reviews / hotel
  - Offer $1 / review
  - 400 reviews

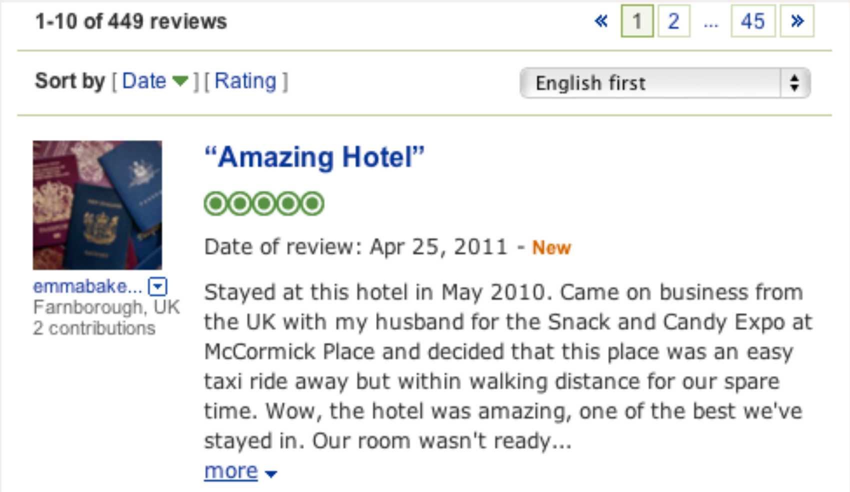
Gathering Data

- Mechanical Turk
  - 20 hotels
  - 20 reviews / hotel
  - Offer $1 / review
  - 400 reviews



Gathering Data

- Mechanical Turk
  - 20 hotels
  - 20 reviews / hotel
  - Offer $1 / review
  - 400 reviews



Gathering Data

- Mechanical Turk
  - 20 hotels
  - 20 reviews / hotel
  - Offer $1 / review
  - 400 reviews

Gathering Data

- Mechanical Turk
  - 20 hotels
  - 20 reviews / hotel
  - Offer $1 / review
  - 400 reviews
- Average time spent: > 8 minutes
- Average length: > 115 words

Gathering Data

- 400 truthful reviews
  - TripAdvisor.com
  - Lengths distributed similarly to deceptive reviews

Overview

- Motivation
- Gathering Data
- Human Performance
- Classifier Performance
- Conclusion

Human Performance

- Why bother?
  - Validates deceptive opinions
  - Baseline to compare other approaches

Human Performance

- Why bother?
  - Validates deceptive opinions
  - Baseline to compare other approaches

Human Performance

- Why bother?
  - Validates deceptive opinions
  - Baseline to compare other approaches

Human Performance

- 80 truthful and 80 deceptive reviews
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Human Performance

		Accuracy	TRUTHFUL P	TRUTHFUL R	TRUTHFUL F	DECEPTIVE P	DECEPTIVE R	DECEPTIVE F
HUMAN	JUDGE 1	**61.9%**	57.9	87.5	**69.7**	74.4	36.3	48.7
	JUDGE 2	56.9%	53.9	**95.0**	68.8	**78.9**	18.8	30.3
	JUDGE 3	53.1%	52.3	70.0	59.9	54.7	36.3	43.6

- 80 truthful and 80 deceptive reviews
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Human Performance

Performed at chance (p-value = 0.1)

			TRUTHFUL			DECEPTIVE		
		Accuracy	**P**	**R**	**F**	**P**	**R**	**F**
HUMAN	JUDGE 1	**61.9%**	57.9	87.5	**69.7**	74.4	36.3	48.7
	JUDGE 2	56.9%	53.9	**95.0**	68.8	**78.9**	18.8	30.3
	JUDGE 3	53.1%	52.3	70.0	59.9	54.7	36.3	43.6

Performed at chance (p-value = 0.5)

- 80 truthful and … eviews
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Human Performance

		Accuracy	TRUTHFUL P	TRUTHFUL R	TRUTHFUL F	DECEPTIVE P	DECEPTIVE R	DECEPTIVE F
HUMAN	JUDGE 1	**61.9%**	57.9	87.5	**69.7**	74.4	36.3	48.7
	JUDGE 2	56.9%	53.9	**95.0**	68.8	**78.9**	18.8	30.3
	JUDGE 3	53.1%	52.3	70.0	59.9	54.7	36.3	43.6

- 80 truthful and 80 deceptive reviews
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Human Performance

		Accuracy	TRUTHFUL P	TRUTHFUL R	TRUTHFUL F	DECEPTIVE P	DECEPTIVE R	DECEPTIVE F
HUMAN	JUDGE 1	**61.9%**	57.9	87.5	**69.7**	74.4	36.3	48.7
	JUDGE 2	56.9%	53.9	**95.0**	68.8	**78.9**	18.8	30.3
	JUDGE 3	53.1%	52.3	70.0	59.9	54[illegible]	36.3	43.6

Classified fewer than 12% of opinions as deceptive!

- 80 truthful and 80 deceptive reviews
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Human Performance

			TRUTHFUL			DECEPTIVE		
		Accuracy	**P**	**R**	**F**	**P**	**R**	**F**
HUMAN	JUDGE 1	**61.9%**	57.9	87.5	**69.7**	74.4	36.3	48.7
	JUDGE 2	56.9%	53.9	**95.0**	68.8	**78.9**	18.8	30.3
	JUDGE 3	53.1%	52.3	70.0	59.9	54.7	36.3	43.6

- 80 truthful and 80 deceptive reviews
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Human Performance

		Accuracy	TRUTHFUL P	TRUTHFUL R	TRUTHFUL F	DECEPTIVE P	DECEPTIVE R	DECEPTIVE F
HUMAN	JUDGE 1	**61.9%**	57.9	87.5	**69.7**	74.4	36.3	48.7
	JUDGE 2	56.9%	53.9	**95.0**	68.8	**78.9**	18.8	30.3
	JUDGE 3	53.1%	52.3	70.0	59.9	54.7	36.3	43.6
META	MAJORITY	58.1%	54.8	92.5	68.8	76.0	23.8	36.2
	SKEPTIC	60.6%	**60.8**	60.0	60.4	60.5	**61.3**	**60.9**

- 80 truthful and 80 deceptive reviews
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Human Performance

		Accuracy	TRUTHFUL P	TRUTHFUL R	TRUTHFUL F	DECEPTIVE P	DECEPTIVE R	DECEPTIVE F
HUMAN	JUDGE 1	**61.9%**	57.9	87.5	**69.7**	74.4	36.3	48.7
	JUDGE 2	56.9%	53.9	**95.0**	68.8	**78.9**	18.8	30.3
	JUDGE 3	53.1%	52.3	70.0	59.9	54.7	36.3	43.6
META	MAJORITY	58.1%	54.8	92.5	68.8	76.0	23.8	36.2
	SKEPTIC	60.6%	**60.8**	60.0	60.4	60.5	**61.3**	**60.9**

No more truth bias!

- 80 truthful and 80 deceptive
- 3 undergraduate judges
  - Truth bias
- 2 meta-judges

Overview

- Motivation
- Gathering Data
- Human Performance
- Classifier Performance
- Conclusion

Classifier Performance

- Three feature sets
  - Genre identification
  - Psycholinguistic deception detection
  - Text categorization
- Linear SVM

Classifier Performance

- Three feature sets
  - Genre identification
  - Psycholinguistic deception detection
  - Text categorization
- Linear SVM

Classifier Performance

- Genre identification
  - 48 part-of-speech (PoS) features
  - Baseline automated approach
- Expectations
  - Truth similar to informative writing
  - Deception similar to imaginative writing

Classifier Performance

- Genre identification
  - 48 part-of-speech (PoS) features
  - Baseline automated approach
- Expectations
  - Truth similar to informative writing
  - Deception similar to imaginative writing

Classifier Performance

- Genre identification
  - 48 part-of-speech (PoS) features
  - Baseline automated approach
- Expectations
  - Truth similar to informative writing
  - Deception similar to imaginative writing

Classifier Performance

- Genre identification
  - 48 part-of-speech (PoS) features
  - Baseline automated approach
- Expectations
  - Truth similar to informative writing
  - Deception similar to imaginative writing

Classifier Performance

			TRUTHFUL			DECEPTIVE		
Approach	**Features**	**Accuracy**	**P**	**R**	**F**	**P**	**R**	**F**
GENRE IDENTIFICATION	POS	73.0%	75.3	68.5	71.7	71.1	77.5	74.2

Classifier Performance

			TRUTHFUL			DECEPTIVE		
Approach	**Features**	**Accuracy**	**P**	**R**	**F**	**P**	**R**	**F**
GENRE IDENTIFICATION	POS	73.0%	75.3	68.5	71.7	71.1	77.5	74.2

Outperforms human judges!
(p-values = {0.06, 0.01, 0.001})

Classifier Performance

TRUTHFUL/INFORMATIVE			DECEPTIVE/IMAGINATIVE		
Category	**Variant**	**Weight**	**Category**	**Variant**	**Weight**
NOUNS	Singular	0.008	VERBS	Base	-0.057
	Plural	0.002		Past tense	**0.041**
	Proper, singular	**-0.041**		Present participle	-0.089
	Proper, plural	0.091		Singular, present	-0.031
ADJECTIVES	General	0.002		Third person singular, present	**0.026**
	Comparative	0.058			
	Superlative	**-0.164**		Modal	-0.063
PREPOSITIONS	General	0.064	ADVERBS	General	**0.001**
DETERMINERS	General	0.009		Comparative	-0.035
COORD. CONJ.	General	0.094	PRONOUNS	Personal	-0.098
VERBS	Past participle	0.053		Possessive	-0.303
ADVERBS	Superlative	**-0.094**	PRE-DETERMINERS	General	**0.017**

- Rayson et. al. (2001)
  - Informative on left, imaginative on right

Classifier Performance

TRUTHFUL/INFORMATIVE			DECEPTIVE/IMAGINATIVE		
Category	**Variant**	**Weight**	**Category**	**Variant**	**Weight**
NOUNS	Singular	0.008	VERBS	Base	-0.057
	Plural	0.002		Past tense	**0.041**
	Proper, singular	**-0.041**		Present participle	-0.089
	Proper, plural	0.091		Singular, present	-0.031
ADJECTIVES	General	0.002		Third person singular, present	**0.026**
	Comparative	0.058			
	Superlative ★	**-0.164**		Modal	-0.063
PREPOSITIONS	General	0.064	ADVERBS	General	**0.001**
DETERMINERS	General	0.009		Comparative	-0.035
COORD. CONJ.	General	0.094	PRONOUNS	Personal	-0.098
VERBS	Past participle	0.053		Possessive	-0.303
ADVERBS	Superlative ★	**-0.094**	PRE-DETERMINERS	General	**0.017**

e.g., best, finest

e.g., most

- Rayson et. al. (2001)
  - Informative on left, imaginative on right

Classifier Performance

- Linguistic Inquire and Word Count (Pennebaker et al., 2001; 2007)
  - Counts instances of ~4,500 keywords
    - Regular expressions, actually
  - Keywords are divided into 80 dimensions across 4 broad groups

Classifier Performance

			TRUTHFUL			DECEPTIVE		
Approach	**Features**	**Accuracy**	**P**	**R**	**F**	**P**	**R**	**F**
GENRE IDENTIFICATION	POS	73.0%	75.3	68.5	71.7	71.1	77.5	74.2
PSYCHOLINGUISTIC DECEPTION DETECTION	LIWC	76.8%	77.2	76.0	76.6	76.4	77.5	76.9

Classifier Performance

			TRUTHFUL			DECEPTIVE		
Approach	**Features**	**Accuracy**	**P**	**R**	**F**	**P**	**R**	**F**
GENRE IDENTIFICATION	POS	73.0%	75.3	68.5	71.7	71.1	77.5	74.2
PSYCHOLINGUISTIC DECEPTION DETECTION	LIWC	76.8%	77.2	76.0	76.6	76.4	77.5	76.9

Outperforms PoS!
(p-value = 0.02)

Classifier Performance

- Text categorization (n-grams)
  - Unigrams
  - Bigrams^{+}
    - Includes unigrams
  - Trigrams^{+}
    - Includes unigrams and bigrams

Classifier Performance

			TRUTHFUL			DECEPTIVE		
Approach	**Features**	**Accuracy**	**P**	**R**	**F**	**P**	**R**	**F**
GENRE IDENTIFICATION	POS	73.0%	75.3	68.5	71.7	71.1	77.5	74.2
PSYCHOLINGUISTIC DECEPTION DETECTION	LIWC	76.8%	77.2	76.0	76.6	76.4	77.5	76.9
TEXT CATEGORIZATION	UNIGRAMS	88.4%	89.9	86.5	88.2	87.0	**90.3**	88.6
	BIGRAMS	89.6%	**90.1**	89.0	89.6	89.1	**90.3**	89.7
	LIWC+BIGRAMS	**89.8**%	89.8	**89.8**	**89.8**	**89.8**	89.8	**89.8**
	TRIGRAMS	89.0%	89.0	89.0	89.0	89.0	89.0	89.0

Classifier Performance

Approach	Features	Accuracy	TRUTHFUL P	TRUTHFUL R	TRUTHFUL F	DECEPTIVE P	DECEPTIVE R	DECEPTIVE F
GENRE IDENTIFICATION	POS	73.0%	75.3	68.5	71.7	71.1	77.5	74.2
PSYCHOLINGUISTIC DECEPTION DETECTION	LIWC	76.8%	77.2	76.0	76.6	76.4	77.5	76.9
TEXT CATEGORIZATION	UNIGRAMS	88.4%	89.9	86.5	88.2	87.0	**90.3**	88.6
	BIGRAMS	89.6%	**90.1**	89.0	89.6	89.1	**90.3**	89.7
	LIWC+BIGRAMS	**89.8**%	89.8	**89.8**	**89.8**	**89.8**	89.8	**89.8**
	TRIGRAMS	89.0%	89.0	89.0	89.0	89.0	89.0	89.0

Outperforms all other methods!

Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
on	hotel
location	,_and
)	luxury
allpunct$_{LIWC}$	experience
floor	hilton
(	business
the_hotel	vacation
bathroom	i
small	spa
helpful	looking
$	while
hotel_.	husband
other	my_husband

- Spatial difficulties (Vrij et al., 2009)
- Psychological distancing (Newman et al., 2003)

Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
★ on	hotel
★ location	,_and
)	luxury
allpunct$_{LIWC}$	experience
★ floor	hilton
(	business
the_hotel	vacation
★ bathroom	i
★ small	spa
helpful	looking
$	while
hotel_.	husband
other	my_husband

- Spatial difficulties (Vrij et al., 2009)
- Psychological distancing (Newman et al., 2003)

Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
on	hotel
location	,_and
)	luxury
allpunct$_{LIWC}$	experience
floor	hilton
(	★ business
the_hotel	★ vacation
bathroom	i
small	spa
helpful	looking
$	while
hotel_.	★ husband
other	★ my_husband

- Spatial difficulties (Vrij et al., 2009)
- Psychological distancing (Newman et al., 2003)

Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
on	hotel
location	,_and
)	luxury
$\text{allpunct}_{\text{LIWC}}$	experience
floor	hilton
(	business
the_hotel	vacation
bathroom	i
small	spa
helpful	looking
$	while
hotel_.	husband
other	my_husband

- Spatial difficulties (Vrij et al., 2009)
- Psychological distancing (Newman et al., 2003)

Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	★ my
on	hotel
location	,_and
)	luxury
$\text{allpunct}_{\text{LIWC}}$	experience
floor	hilton
(	business
the_hotel	vacation
bathroom	★ i
small	spa
helpful	looking
$	while
hotel_.	husband
other	my_husband

- Spatial difficulties (Vrij et al., 2009)
- Psychological distancing (Newman et al., 2003)

Overview

- Motivation
- Gathering Data
- Human Performance
- Classifier Performance
- Conclusion

Conclusion

- Language use varies depending on features of the text and the author
- It seems likely that whether the author is being truthful or deceptive influences their language use
- Research into detecting deception has interesting real-life applications, e.g., detecting fake reviews
- Standard n-gram text categorization can outperform human performance on this task

- Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. 2010. A latent variable model for geographic lexical variation. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 1277-1287.
- Jacob Eisenstein, Noah A. Smith, and Eric P. Xing. 2011. Discovering sociolinguistic associations with structured sparsity. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 1365-1374.
- B. Johnstone. 2010. Language and place. In R. Mesthrie and W. Wolfram, editors, Cambridge Handbook of Sociolinguistics. Cambridge University Press.
- Qiaozhu Mei, Chao Liu, Hang Su, and ChengXiang Zhai. 2006. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In Proceedings of the 15th international conference on World Wide Web (WWW '06). ACM, New York, NY, USA, 533-542.
- Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In Proceedings of the 16th international conference on World Wide Web (WWW '07). ACM, New York, NY, USA, 171-180.
- Labov, W., Ash, S. & Boberg, C. (2006). The atlas of North American English: phonetics, phonology, and sound change: a multimedia reference tool. Mouton de Gruyter
- Tagliamonte, S. (2006). Analysing sociolinguistic variation. Cambridge Univ Press.
- Koppel, M., Argamon, S., Shimoni, A. R.. 2002. Automatically Categorizing Written Text by Author Gender. Literary and Linguistic Computing.
- P. Rayson, A. Wilson, and G. Leech. 2001. Grammatical word class variation within the British National Corpus sampler. Language and Computers, 36(1):295–306.
- D. Biber, S. Johansson, G. Leech, S. Conrad, E. Finegan, and R. Quirk. 1999. Longman grammar of spoken and written English, volume 2. MIT Press.
- Mehler, S. Sharoff and M. Santini. 2010. Genres on the Web: Computational Models and Empirical Studies. TEXT, SPEECH AND LANGUAGE TECHNOLOGY
- Rehm, Georg; Santini, Marina; Mehler, Alexander; Braslavski, Pavel; Gleim, R¨udiger; Stubbe, Andrea; Symonenko, Svetlana; Tavosanis, Mirko and Vidulin, Vedrana (2008): “Towards a Reference Corpus of Web Genres for the Evaluation of Genre Identification Systems”. In: Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008). Marrakech, Morocco.
- S. Westman and L. Freund. Information interaction in 140 characters or less: genres on twitter. In IIiX '10, pages 323{328, 2010.
- Arjun Mukherjee and Bing Liu. 2010. Improving gender classification of blog authors. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, October. Association for Computational Linguistics.
- John D. Burger, John Henderson, George Kim, and Guido Zarrella. 2011. Discriminating gender on Twitter. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11). Association for Computational Linguistics, Stroudsburg, PA, USA, 1301-1309.

- Ruchita Sarawgi, Kailash Gajulapalli, and Yejin Choi. 2011. Gender attribution: tracing stylometric evidence beyond topic and genre. In Proceedings of the Fifteenth Conference on Computational Natural Language Learning, CoNLL '11, pages 78–86, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2, 1-135.
- Vinodkumar Prabhakaran, Owen Rambow, and Mona Diab. 2010. Automatic committed belief tagging. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 1014-1022.
- Mona T. Diab, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, and Weiwei Guo. 2009. Committed belief annotation and tagging. In Proceedings of the Third Linguistic Annotation Workshop (ACL-IJCNLP '09). Association for Computational Linguistics, Stroudsburg, PA, USA, 68-73.
- Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica, 1, 49-98.
- M.L. Newman, J.W. Pennebaker, D.S. Berry, and J.M. Richards. 2003. Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin, 29(5):665.
- R. Mihalcea and C. Strapparava. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 309–312. Association for Computational Linguistics
- Jeffrey T. Hancock, Catalina Toma, and Nicole Ellison. 2007. The truth about lying in online dating profiles. In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07). ACM, New York, NY, USA, 449-452. DOI=10.1145/1240624.1240697 http://doi.acm.org/10.1145/1240624.1240697
- Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. 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 (HLT '11), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 309-319.
- Myle Ott, Claire Cardie, and Jeff Hancock. 2012. Estimating the prevalence of deception in online review communities. In Proceedings of the 21st international conference on World Wide Web (WWW '12). ACM, New York, NY, USA, 201-210. DOI=10.1145/2187836.2187864 http://doi.acm.org/10.1145/2187836.2187864
- C.F. Bond and B.M. DePaulo. 2006. Accuracy of deception judgments. Personality and Social Psychology Review, 10(3):214.
- A. Vrij. 2008. Detecting lies and deceit: Pitfalls and opportunities. Wiley-Interscience.
- N. Jindal and B. Liu. 2008. Opinion spam and analysis. In Proceedings of the international conference on Web search and web data mining, pages 219–230. ACM.
- A. Vrij, S. Leal, P.A. Granhag, S. Mann, R.P. Fisher, J. Hillman, and K. Sperry. 2009. Outsmarting the liars: The benefit of asking unanticipated questions. Law and human behavior, 33(2):159–166.