

# Finding Deceptive Opinion Spam by Any Stretch of the Imagination

Myle Ott,<sup>1</sup> Yejin Choi,<sup>1</sup> Claire Cardie,<sup>1</sup> and Jeff Hancock<sup>2</sup>  
 Dept. of Computer Science,<sup>1</sup> Communication<sup>2</sup>  
 Cornell University, Ithaca, NY

## Motivation

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



Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Motivation

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



Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Motivation

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



Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Motivation

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

★★★★★ **Works Just as expected**, May 14, 2007  
By [Laurie B. Cook](#)  - [See all my reviews](#)  
REAL NAME  
**This review is from: Belkin F5U301 CableFree 4-Port USB 2.0 Hub with Dongle (Electronics)**  
Supplies good range and does provide true wireless USB. Software worked right out of the box. I have been recommending this nifty little device to all my friends. Very useful device.

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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.

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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.

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Overview

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Gathering Data

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

Home → United States → Illinois (IL) → Chicago → Chicago Hotels → James Chicago

**James Chicago**

Hotel class ★★★★★  
55 East Ontario, Corner of Rush and Ontario, Chicago, IL 60611  
877.526.3755 [Hotel website](#) [E-mail hotel](#)

**What travelers say about James Chicago**

- Great location (33)
- Room service (20)
- Very nice (18)
- Trader joe (16)
- Boutique hotel (15)
- Magnificent mile (14)
- Very good (13)
- Michigan avenue (13)
- Comfortable bed (10)
- Friendly and helpful (8)

**Reviews you can trust**

Filter traveler reviews [Write a Review](#)

Trip type	Traveler rating
<input checked="" type="radio"/> All reviews (449)	<input checked="" type="radio"/> All (449)
<input type="radio"/> Business reviews (94)	<input type="radio"/> Excellent (278)
<input type="radio"/> Couples reviews (194)	<input type="radio"/> Very good (116)
<input type="radio"/> Family reviews (28)	<input type="radio"/> Average (23)
<input type="radio"/> Friends reviews (60)	<input type="radio"/> Poor (19)
<input type="radio"/> Solo travel reviews (62)	<input type="radio"/> Terrible (13)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Gathering Data

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

1-10 of 449 reviews

Sort by [Date ▼] [Rating] English first

**"Amazing Hotel!"**

Date of review: Apr 25, 2011 - New

emmbake...  
Farnborough, UK  
2 contributions

Stayed at this hotel in May 2010. Came on business from the UK with my husband for the Snack and Candy Expo at McCormick Place and decided that this place was an easy taxi ride away but within walking distance for our spare time. Wow, the hotel was amazing, one of the best we've stayed in. Our room wasn't ready...

[more ▼](#)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Gathering Data

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



Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Gathering Data

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Overview

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Human Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Human Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Human Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Human Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Human Performance

(p-value = 0.1)		Accuracy	TRUTHFUL			DECEPTIVE		
			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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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
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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Human Performance

		Accuracy	TRUTHFUL			DECEPTIVE		
			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
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

No more truth bias!

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Overview

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination



## Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

Approach	Features	Accuracy	TRUTHFUL			DECEPTIVE		
			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})

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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	<b>0.041</b>
	Proper, singular	<b>-0.041</b>		Present participle	-0.089
	Proper, plural	0.091		Singular, present	-0.031
ADJECTIVES	General	0.002		Third person singular, present	<b>0.026</b>
	Comparative	0.058		Modal	-0.063
	Superlative	<b>-0.164</b>	ADVERBS	General	<b>0.001</b>
PREPOSITIONS	General	0.064		Comparative	-0.035
DETERMINERS	General	0.009	PRONOUNS	Personal	-0.098
COORD. CONJ.	General	0.094		Possessive	-0.303
VERBS	Past participle	0.053	PRE-DETERMINERS	General	<b>0.017</b>
ADVERBS	Superlative	<b>-0.094</b>			

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# 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	<b>0.041</b>
	Proper, singular	<b>-0.041</b>		Present participle	-0.089
	Proper, plural	0.091		Singular, present	-0.031
ADJECTIVES	General	0.002		Third person singular, present	<b>0.026</b>
	Comparative	0.058		Modal	-0.063
	Superlative	★ <b>-0.164</b>	ADVERBS	General	<b>0.001</b>
PREPOSITIONS	General	0.064		Comparative	-0.035
DETERMINERS	General	0.009	PRONOUNS	Personal	-0.098
COORD. CONJ.	General	0.094		Possessive	-0.303
VERBS	Past participle	0.053	PRE-DETERMINERS	General	<b>0.017</b>
ADVERBS	Superlative	★ <b>-0.094</b>			

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Classifier Performance

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Classifier Performance

- 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Classifier Performance

- 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Classifier Performance

- 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

- 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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

Approach	Features	Accuracy	TRUTHFUL			DECEPTIVE		
			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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

Approach	Features	Accuracy	TRUTHFUL			DECEPTIVE		
			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)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

- Text categorization (n-grams)
  - Unigrams
  - Bigrams<sup>+</sup>
    - Includes unigrams
  - Trigrams<sup>+</sup>
    - Includes unigrams and bigrams

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

Approach	Features	Accuracy	TRUTHFUL			DECEPTIVE		
			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	<b>90.3</b>	88.6
	BIGRAMS	89.6%	<b>90.1</b>	89.0	89.6	89.1	<b>90.3</b>	89.7
	LIWC+BIGRAMS	<b>89.8%</b>	89.8	<b>89.8</b>	<b>89.8</b>	<b>89.8</b>	89.8	<b>89.8</b>
	TRIGRAMS	89.0%	89.0	89.0	89.0	89.0	89.0	89.0

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
on	hotel
location	,_and
)	luxury
allpunctLIWC	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)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

Approach	Features	Accuracy	TRUTHFUL			DECEPTIVE		
			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	<b>90.3</b>	88.6
	BIGRAMS	89.6%	<b>90.1</b>	89.0	89.6	89.1	<b>90.3</b>	89.7
	LIWC+BIGRAMS	<b>89.8%</b>	89.8	<b>89.8</b>	<b>89.8</b>	<b>89.8</b>	89.8	<b>89.8</b>
	TRIGRAMS	89.0%	89.0	89.0	89.0	89.0	89.0	89.0

Outperforms all other methods!

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
★ on	hotel
★ location	,_and
)	luxury
allpunctLIWC	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)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
on	hotel
location	,_and
)	luxury
allpunct <sub>LIWC</sub>	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)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	my
on	hotel
location	,_and
)	luxury
allpunct <sub>LIWC</sub>	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)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Classifier Performance

LIWC+BIGRAMS	
TRUTHFUL	DECEPTIVE
-	chicago
...	★ my
on	hotel
location	,_and
)	luxury
allpunct <sub>LIWC</sub>	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)

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

# Overview

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

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Conclusion

- First large-scale gold-standard deception dataset
  - [http://www.cs.cornell.edu/~myleott/op\\_spam](http://www.cs.cornell.edu/~myleott/op_spam)
- Evaluated human deception detection performance
- Developed automated classifiers capable of nearly 90% accuracy
  - Relationship between deceptive and imaginative text
  - Importance of moving beyond universal deception cues

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Conclusion

- First large-scale gold-standard deception dataset
  - [http://www.cs.cornell.edu/~myleott/op\\_spam](http://www.cs.cornell.edu/~myleott/op_spam)
- Evaluated human deception detection performance
- Developed automated classifiers capable of nearly 90% accuracy
  - Relationship between deceptive and imaginative text
  - Importance of moving beyond universal deception cues

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Conclusion

- First large-scale gold-standard deception dataset
  - [http://www.cs.cornell.edu/~myleott/op\\_spam](http://www.cs.cornell.edu/~myleott/op_spam)
- Evaluated human deception detection performance
- Developed automated classifiers capable of nearly 90% accuracy
  - Relationship between deceptive and imaginative text
  - Importance of moving beyond universal deception cues

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Conclusion

- First large-scale gold-standard deception dataset
  - [http://www.cs.cornell.edu/~myleott/op\\_spam](http://www.cs.cornell.edu/~myleott/op_spam)
- Evaluated human deception detection performance
- Developed automated classifiers capable of nearly 90% accuracy
  - Relationship between deceptive and imaginative text
  - Importance of moving beyond universal deception cues

Finding Deceptive Opinion Spam by Any Stretch of the Imagination



## Conclusion

- First large-scale gold-standard deception dataset
  - [http://www.cs.cornell.edu/~myleott/op\\_spam](http://www.cs.cornell.edu/~myleott/op_spam)
- Evaluated human deception detection performance
- Developed automated classifiers capable of nearly 90% accuracy
  - Relationship between deceptive and imaginative text
  - Importance of moving beyond universal deception cues

Finding Deceptive Opinion Spam by Any Stretch of the Imagination

## Thank you. Questions?

- First large-scale gold-standard deception dataset
  - [http://www.cs.cornell.edu/~myleott/op\\_spam](http://www.cs.cornell.edu/~myleott/op_spam)
- Evaluated human deception detection performance
- Developed automated classifiers capable of nearly 90% accuracy
  - Relationship between deceptive and imaginative text
  - Importance of moving beyond universal deception cues

Finding Deceptive Opinion Spam by Any Stretch of the Imagination