

Information Retrieval

INFO 4300 / CS 4300

- Last class
 - Indexing
 - Index construction
 - Compression
 - Ranking model
- Today
 - (Fun) detour in honor of Fall Break
 - Text classification
 - Opinion spam detection
 - Marseille video: cs4740 class favorite

Classification

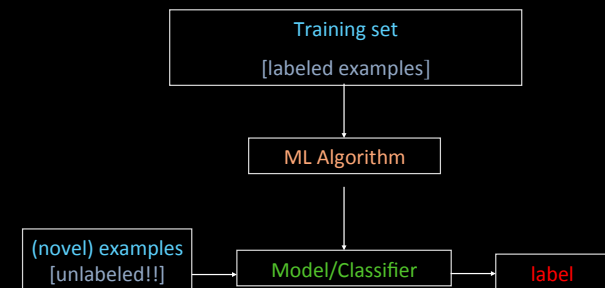
- Classification is the task of automatically applying labels to items
- Useful for many search-related tasks
 - Spam detection
 - Email categorization
 - Sentiment classification
 - Language identification
 - Online advertising

Classification

- Classification is a classical pattern recognition / machine learning problem
 - Asks “what class does this item belong to?”
 - **Supervised learning** task
- Items can be documents, queries, emails, entities, images, etc.

Classification

- Supervised learning methods

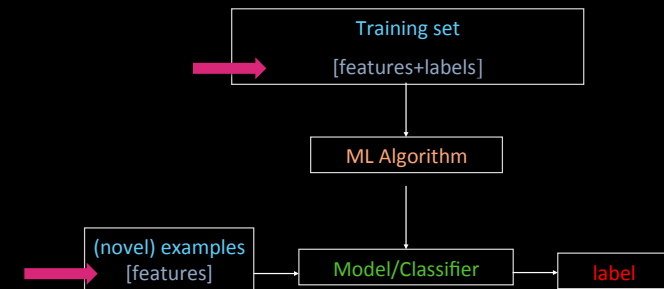


How to Classify?

- How do humans classify items?
- For example, suppose you had to classify the “healthiness” of a food
 - Identify set of **features** indicative of health
 - » fat, cholesterol, sugar, sodium, etc.
 - **Extract** features from foods
 - » Read nutritional facts, chemical analysis, etc.
 - **Combine evidence** from the features into a hypothesis
 - » Add health features together to get “healthiness factor”
 - Finally, **classify** the item based on the evidence
 - » If “healthiness factor” is above a certain value, then deem it healthy

Classification

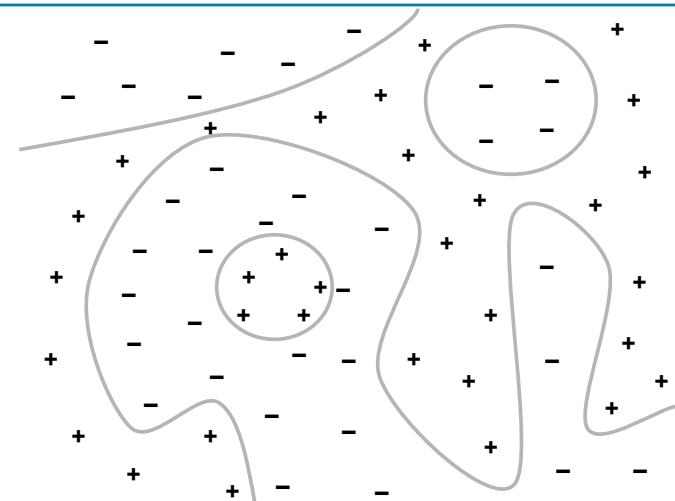
- Supervised learning methods



Ontologies

- Ontology is a labeling or categorization scheme
- Examples
 - Binary (spam, not spam)
 - Multi-valued (red, green, blue)
 - Hierarchical (news/local/sports)
- Different classification tasks require different ontologies

Nearest Neighbor Classification



Information Retrieval

INFO 4300 / CS 4300

- Last class
 - Indexing
 - » Index construction
 - » Compression
 - » Ranking model
- Today
 - (Fun) detour in honor of Fall Break
 - » Text classification
 - ➔ » Opinion spam detection
 - Marseille video: cs4740 class favorite

Online Reviews

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

The Statler Hotel at Cornell University

130 Statler Drive, Ithaca, NY 14853-6901

Hotel website | Hotel packages | (607) 257-2500 | Hotel amenities

Reviews you can trust

Traveler rating	See reviews for	Rating summary
Excellent 116	Families 64	Location @@@@
Very good 39	Couples 19	Sleep Quality @@@@
Average 11	Solo 22	Rooms @@@@
Poor 2	Business 50	Service @@@@
Terrible 2		Value @@@@
		Cleanliness @@@@

See which rooms travelers prefer - 43 traveler tips

170 reviews sorted by Data Rating (English first)

"Beautiful Campus Views, Excellent Hotel"
Reviewed May 3, 2013
Was this review helpful? Yes

Knowing that this hotel was connected with the world-leading hospitality management program at Cornell, I raised my expectations before arrival. The hotel did not disappoint. The staff provided excellent service and the room and facilities were also excellent. I had a wonderful view of the campus and rolling hills from my room. The TV was a nice 42" flatscreen that...

Ott, Cardie & Hancock [ACL 2011, WWW 2012, NAAACL 2013, ACL 2013]

Online Reviews

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

1 of 7 people found the following review helpful:

★★★★★ **Great Customer Service!!**, April 7, 2011
By **akaempf** - See all my reviews
Amazon Verified Purchase (What's this?)
This review is from: **Apple iPad 2 MC984LL/A Tablet (64GB, WiFi + AT&T 3G, White) NEWEST MODEL (Personal Computers)**

"WE SHIP TECH" is a great reliable company. I ordered the iPad2 late 3/30 @ 10:50pm and received the iPad2 4/1. When I wrote an email to them on the 3/31 they responded in about 20 min max. It's so hard to find great customer service and not get scammed these days that "We Ship Tech" is a breath of fresh air!!! I would surely use them again and highly recommend them to anyone who expects great products & service. Thank you We Ship Tech!!!!

Help other customers find the most helpful reviews | Report abuse | Permalink
Was this review helpful to you? Yes No | Comment

Online Reviews

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

Customer Reviews

Belkin F5U301 CableFree 4-Port USB 2.0 Hub with Dongle by Belkin

Average Customer Rating: 4.5 out of 5 (98 customer reviews)

5 star: (1) | 4 star: (2) | 3 star: (0) | 2 star: (2) | 1 star: (3)

8 of 119 people found the following review helpful:

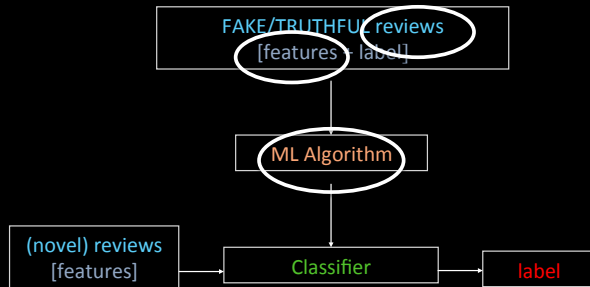
★★★★★ **Works Just as expected**, May 14, 2007
By **Laurie B. Cook** - See all my reviews

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.

Help other customers find the most helpful reviews | Report abuse | Permalink
Was this review helpful to you? Yes No | Comment (0)

A Classification Approach



Which of these two hotel reviews is *deceptive opinion spam*?

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 say 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.

This review is the subjective opinion of a TripAdvisor member and not of TripAdvisor LLC.

[View profile](#) | [Send message](#) | [Compliment reviewer](#)
[Report problem with review](#)

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.

This review is the subjective opinion of a TripAdvisor member and not of TripAdvisor LLC.

[View profile](#) | [Send message](#) | [Compliment reviewer](#)
[Report problem with review](#)

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.

This review is the subjective opinion of a TripAdvisor member and not of TripAdvisor LLC.

[View profile](#) | [Send message](#) | [Compliment reviewer](#)
[Report problem with review](#)

Previous Work

- Jindal & Liu (2008)
 - Opinion spam is different from e-mail or Web spam
 - No gold standard *deceptive* reviews
 - Identify *duplicate* vs. *non-duplicate* reviews
- Mihalcea & Strapparava (2009), Zhou et al. (2004, 2008)
 - N-gram-based features, small corpora
- Different deception tasks

ACL (2011), WWW (2012)

Overview

- Motivation and Background
- Gathering Data
- Human Performance
- Classifier Construction and Performance

Data: Deceptive Reviews

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

Data

- Mechanical Turk
 - Have: 20 chosen hotels
 - Want: 20 deceptive *positive* reviews / hotel
 - Offer: \$1 / review
 - Get: 400 reviews

Data

- Mechanical Turk
 - Have: 20 chosen hotels
 - Want: 20 deceptive **positive** reviews / hotel
 - Offer: \$1 / review
 - Get: 400 reviews

The screenshot shows a travel review page for 'James Chicago'. At the top, it lists the hotel class as five stars, the address as '55 East Ontario, Corner of Rush and Ontario, Chicago, IL 60611', and the phone number as '877.526.3755'. Below this, there are links for 'Hotel website' and 'E-mail hotel'. A section titled 'What travelers say about James Chicago' lists various categories of reviews with counts: 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), and Friendly and helpful (8). There is also a 'Reviews you can trust' section with a 'Filter traveler reviews' button and a 'Write a Review' button. The 'Filter traveler reviews' section has two columns: 'Trip type' and 'Traveler rating'. Under 'Trip type', there are radio buttons for 'All reviews (449)', 'Business reviews (94)', 'Couples reviews (194)', 'Family reviews (28)', 'Friends reviews (60)', and 'Solo travel reviews (62)'. Under 'Traveler rating', there are radio buttons for 'All (449)', 'Excellent (278)', 'Very good (116)', 'Average (23)', 'Poor (19)', and 'Terrible (13)'. At the bottom, it says '1-10 of 449 reviews' with navigation arrows.

Data

- Mechanical Turk
 - Have: 20 chosen hotels
 - Want: 20 deceptive **positive** reviews / hotel
 - Offer: \$1 / review
 - Get: 400 reviews



Instructions

Assume that you work for the hotel's marketing department, and pretend that your boss wants you to write a fake review (as if you were a customer) to be posted on a travel review website; additionally, the review needs to sound realistic and portray the hotel in a positive light.

Data

- Allow only a single submission per Turker
- Restrict our task to Turkers
 - Who are located in the United States
 - Who maintain an approval rating of at least 90%
- Check for plagiarism

Data: Truthful Reviews

- Mine all TripAdvisor.com reviews
 - From the 20 most-reviewed Chicago hotels (6,977)
 - Discard non-5-star reviews (3,130)
 - Exclude reviews written by first-time reviewers (1,607), under 150 characters, non-English. 2124 reviews left.
- Select 400 reviews such that the lengths are distributed similarly to the deceptive reviews

Validating the Deceptive Reviews

- Measure human performance
- Can also serve as a baseline

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

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

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

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

Performed at chance
(p-value = 0.1)

Performed at chance
(p-value = 0.5)

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

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

Classified fewer than 12% of opinions as deceptive!

Overview

- Motivation and Background
- Gathering Data
- Human Performance
- Classifier Construction and Performance

Classifier

- Linear SVM (Support Vector Machine)

Feature representation

- Three feature sets encode potentially complementary framings
 - Problem in [genre identification](#)
 - Instance of [psycholinguistic deception detection](#)
 - Standard [text categorization](#)

Features: genre identification

- 48 part-of-speech (PoS) features
- Expectations
 - Truth similar to informative writing
 - Deception similar to imaginative writing

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}

Analysis

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		Modal	-0.063	
	Superlative	-0.164		General	0.001	
PREPOSITIONS	General	0.064		COMPARISONS	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	0.017	
ADVERBS	Superlative	-0.094				

- POS feature analysis
 - superlatives

e.g., least often

e.g., fastest, nicest

Features: psycholinguistic

- Linguistic Inquiry 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

Features: psycholinguistic

- 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

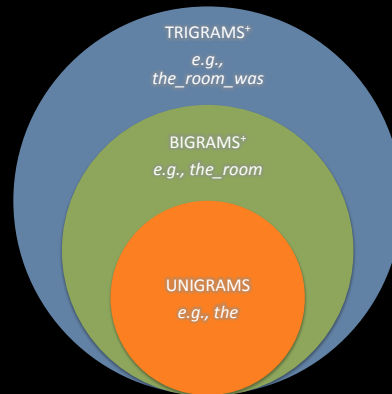
Performance

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

Features: text categorization

- Features
 - n-grams



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	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

Outperform all other methods

Analysis

- LIWC+BIGRAMS feature analysis
 - Spatial difficulties (Vrij et al., 2009)
 - Psychological distancing (Newman et al., 2003)

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

Analysis

- LIWC+BIGRAMS feature analysis
 - Spatial difficulties (Vrij et al., 2009)
 - Psychological distancing (Newman et al., 2003)

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

Analysis

- LIWC+BIGRAMS feature analysis
 - Spatial difficulties (Vrij et al., 2009)
 - Psychological distancing (Newman et al., 2003)

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

STRONG DECEPTIVE INDICATORS

A focus on who they were with
In this example, "My husband;" also words like "family."

Greater use of first-person singular
Fake reviews tend to use "I" and "me" more often.

Direct mention of where they stayed
Hotel and city names were less common in truthful reviews, which focus more on details about the hotel itself, like "small" or "bathroom."

"My husband and I stayed in the [hotel name] Chicago

and had a very nice stay! The rooms were large and comfortable. The view of Lake Michigan from our room was gorgeous. Room service was really good and quick, eating in the room looking at that view, awesome! The pool was really nice but we didn't get a chance to use it. Great location for all of the downtown Chicago attractions such as theaters and museums. Very friendly staff and knowledgeable, you can't go wrong staying here."

SLIGHT DECEPTIVE INDICATORS

High adverb use
"Very" and "really" are both used twice; "here" is used once.

High verb use
"Get", "go", "use", "can't", "didn't", "eating", "had", "looking", "stayed", "was" (three times), "were."

Use of "!" and positive emotion
Deceptive reviews tend to use exclamation points, while truthful reviews used more punctuation of other kinds, including "\$."

Conclusions

- People are **not good** at detecting fake on-line reviews
- Developed **automated classifier** capable of nearly **90% accuracy** when detecting (this one type of!) deceptive opinion spam

 Reviews you can trust

 reviews from our community

Follow-up and Ongoing Work

- Prevalence of opinions [WWW 2012]
- Negative opinions [NAACL 2013]
- Other domains (e.g., restaurants, doctors)
- Vary context of deception (e.g., domain experts vs. turkers)
- Countermeasures


www.reviewskeptic.com



Bates Motel, Psycho



"Clean room, nice shower."

 Reviewed August 30, 2011 new

Though the concierge was a little strange and I had some trouble with the shower. Otherwise, not so bad. The motel is family-owned and operated and you can tell that it's home to a lot of history. Affordable, quaint, mostly quiet, and one long block walk to Starbucks.

Marion Crane
Arizona
10 reviews
11 helpful votes

More ▾
Was this review helpful? Yes

Problem with this review?

Screen Junkies