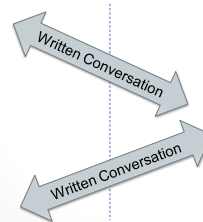


Web Intelligence

Quratulain Rajput
Faculty of Computer Science, IBA
Spring2013

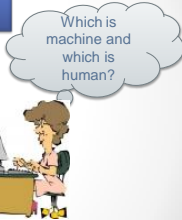
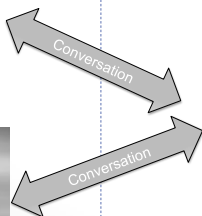
What is Intelligence?

Allan Turing Test



What is Intelligence?

Allan Turing Test



Turing Claim

If the machine successfully foolish the human judge in to believing that it is not a machine but a human that machine should be considered as Intelligent.

The Reverse Turing test

Which is machine and which is human?

Conversation

Conversation

• Quratulain Rajput

Reverse Turing Test

- Examples

Enter the characters you see

D3-N=PPe

overlooks inquiry

Type the last words:

• Quratulain Rajput

The Reverse Turing test

Which is man and which is woman?

Written Conversation

Written Conversation

Sentiment analysis

- Likes/dislikes
- Shopper/surfer
- Rich/poor ...

• Quratulain Rajput

AI web Applications

- Language translation

Personally, I believe the biggest feature of artificial intelligence is the expectation set by science fiction for what AI is and what it can be. That isn't able to be built fast enough by scientists, thinkers, and entrepreneurs. In other words, there seems to be a wide gap between research/exploration and commercialization, mostly because expectations for what AI can do for us is massively overemphasized. It prevents us from

- Image recognition:

Joggles

Facebook, Picasa

• Quratulain Rajput

Sentiment Analysis

- What is sentiment analysis (SA)?
- Why is it worth doing?
- How is it done?
- What are the challenges?

• Quratulain Rajput

• 9

What is sentiment?

- Sentiment = feelings
 - Attitudes
 - Emotions
 - Opinions
- Generally, a binary values is assumed
 - For/against, like/dislike, good/bad, etc.
- Subjective impressions, not facts

• Quratulain Rajput

• 10

Semantic Analysis

- Using NLP, statistics, or machine learning methods to
 - extract, identify, or otherwise characterize the sentiment content of a text unit

• Quratulain Rajput

• 11

Why sentiment analysis

- Is this product review positive or negative?
- Is this customer email satisfied or dissatisfied?
- Based on a sample of tweets, how are people responding to this ad campaign/product release/news item?
- How have bloggers' attitudes about the president changed since the election?

• Quratulain Rajput

• 12

Application in business intelligence

- Question: Why aren't consumers buying our laptop?
- Question: How to promote product?
- Question: What are the reviews about a book?

Thus,

- search the web for opinions and reviews of this and competing laptops.
- Blogs, opinions, amazon, tweets, etc.

• Quratulain Rajput

• 13

What to classify

- There are many possibilities for what we might want to classify:
 - Users
 - Texts
 - Sentences (paragraphs, chunks of text)
 - Tweets/updates

• Quratulain Rajput

• 14

Sentiment Analysis

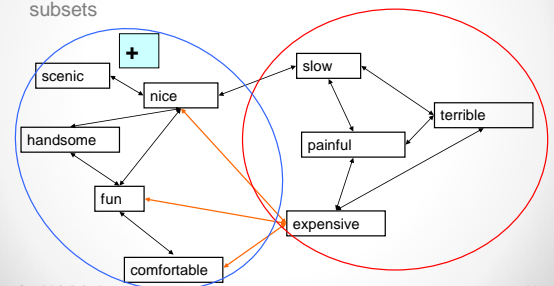
- We can inquire about sentiment at various linguistic levels:
 - Words – objective, **positive**, **negative**, **neutral**
- For example, consider adjectives
 - positive: **honest** **important** **mature** **large** **patient**
 - negative: **harmful** **hypocritical** **inefficient**
 - subjective (but not positive or negative): **curious**, **peculiar**, **odd**, **likely**, **probable**

• Quratulain Rajput

• 15

Sentiment Analysis

- A **clustering algorithm** partitions the adjectives into two subsets



• Quratulain Rajput

• 16

How to do it (using machine learning)

count		Sentiment
2000	I really like this course and am learning a lot	positive
800	I really hate this course and think it is a waste of time	negative
200	The course is really too simple and quite a bore	negative
3000	The course is simple, fun and very easy to follow	positive
1000	I'm enjoying this course a lot and learning something too	positive
400	I would enjoy myself a lot if I did not have to be in this course	negative
600	I did not enjoy this course enough	negative

$p(+)$ = 6000/8000 = .75; $p(-)$ = 2000/8000 = .25 Aprior probability

$p(\text{like}|+) = 2000/6000 = .33$; $p(\text{enjoy}|+) = .16$; ... $p(\text{hate}|+) = 1/6000 = .0002$...
 $p(\text{hate}|-) = 800/2000 = .4$; $p(\text{bore}|-) = .1$; $p(\text{like}|-) = 1/2000 = .0001$;
 also ... $p(\text{enjoy}|-) = 1000/2000 = .5$! and while $p(\text{lot}|+) = .5$, $p(\text{lot}|-) = .4$!

• Quratulain Rajput

• 17

Bayesian sentiment analysis

positive likelihoods	negative likelihoods
$p(\text{like} +) = .33$	$p(\text{like} -) = .0001$
$p(\text{lot} +) = .5$	$p(\text{lot} -) = .4$
$p(\text{hate} +) = .0002$	$p(\text{hate} -) = .4$
$p(\text{waste} -) = .0002$	$p(\text{waste} -) = .4$
$p(\text{simple} +) = .5$	$p(\text{simple} -) = .1$
$p(\text{easy} +) = .5$	$p(\text{easy} -) = .0001$
$p(\text{enjoy} +) = .16$	$p(\text{enjoy} -) = .1$

now faced with a new tweet:

I really like this simple course a lot

$$L = \frac{p(\text{like}|+)p(\text{lot}|+)[1 - p(\text{hate}|+)] [1 - p(\text{waste}|+)] p(\text{simple}|+) [1 - p(\text{easy}|+)] [1 - p(\text{enjoy}|+)] p(+)}{p(\text{like}|-)p(\text{lot}|-)[1 - p(\text{hate}|-)] [1 - p(\text{waste}|-)] p(\text{simple}|-) [1 - p(\text{easy}|-)] [1 - p(\text{enjoy}|-)] p(-)}$$

we get $L = \frac{.026}{.00005} \gg 1$ so the system labels this tweet as 'positive'

• Quratulain Rajput

• 18

Challenges

Language is ambiguous, consider the following text:

"The watch isn't water resistant" - In a product review this could be negative.

"The canon camera is better than the Fisher Price one" - comparisons are hard to classify.

"Wall's the ice cream is luuuuuuuuuvely" - slang and the way we communicate in general needs to be processed.

• Quratulain Rajput

• 19

Conclusion

- Sentiment analysis is a difficult task
- The difficulty increases with the unclear meaning and complexity of opinions expressed
- Product reviews, etc. are relatively easy
- Books, movies, art, music are more difficult
- Policy discussions, indirect expressions of opinion more difficult still
- Non-binary sentiment (political inclinations etc.) is extremely difficult

• Quratulain Rajput

• 20

