

# 1: Sentiment Classification

Machine Learning and Real-world Data (MLRD)

Ann Copestake

(based on slides created by Simone Teufel)

Lent 2018

# This course: Machine Learning and Real-world Data (MLRD)

- Three Topics:
  - Classification: sentiment classification – thousands of movie reviews.
  - Sequence analysis: proteins – hundreds of amino acid sequences.
  - Network analysis: social networks — thousands of users and links between them.
- Different types of machine learning: straightforward approaches you can implement quickly.
- Emphasis on methodology: relevant for all approaches.
- Practical-based, each session starts with a short lecture introducing the main concepts.

# Computer Science as an empirical subject

- The style of solving tasks in this course is *empirical*.
- You will start from a hypothesis or an idea which you will test.
- Then you perform some manipulations on your data.
- You observe and record the results.
- You need a **lab book** to record your manipulations, observations and measurements.
  - physical book or electronic record

# Topic 1: Evaluative language and sentiment classification

- IMDb (= Internet Movie Data Base) has about 4.7 million titles (<http://www.imdb.com/pressroom/stats/>).
- Reviews: written in natural language by the general public.
- **Sentiment classification** — the task of automatically deciding whether a review is positive or negative, based on the text of the review.
- Standard task in **Natural Language Processing (NLP)**.



**Casablanca** (1942)

## User Reviews

[+ Add a Review](#)

1,152 Reviews

Ordered By: Helpfulness

★ 10/10

### A masterwork for all time...

[kdryan](#) 10 November 2004

There is a scene about halfway through the movie Casablanca that has become commonly known as 'The Battle of the Anthems' throughout the film's long history. A group of German soldiers has come into Rick's Café American and are drunkenly singing the German National Anthem at the top of their voice. Victor Lazlo, the leader of the French Resistance, cannot stand this act and while the rest of the club stares appalled at the Germans, Lazlo orders the band to play 'Le Marseilles (sic?)' the French National Anthem. With a nod from Rick, the band begins playing, with Victor singing at the top of HIS voice. This in turn, inspires the whole club to begin singing and the Germans are forced to surrender and sit down at their table, humbled by the crowd's dedication. This scene is a turning point in the movie, for reasons that I leave to you to discover.

As I watched this movie again tonight for what must be the 100th time, I noticed these

★ 10/10

### We'll Always Have Casablanca

[RWiggum](#) 29 March 2004

### Casablanca

#### Opinion

[Awards](#)

[FAQ](#)




[User Reviews](#)

[User Ratings](#)

[External Reviews](#)

[Metacritic Reviews](#)

[Explore More](#)

Share this page:   

[Like](#) 27K people like this. Be the first of your friends.

### User Lists

[Create a list »](#)

Related lists from IMDb users



#### movie i watched

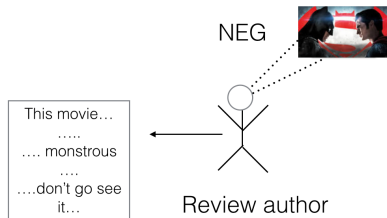
a list of 42 titles  
created 20 May 2012



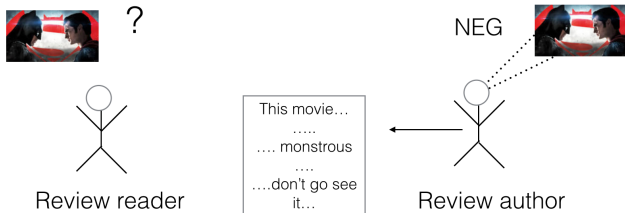
#### Twist Endings Movie

a list of 47 titles  
created 01 Nov 2013

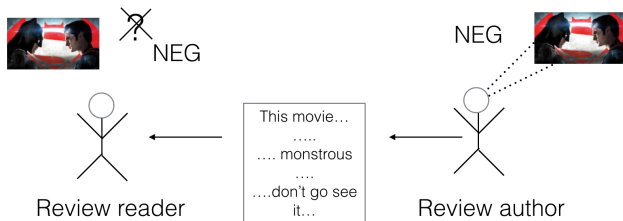
# Review sentiment



# Review sentiment

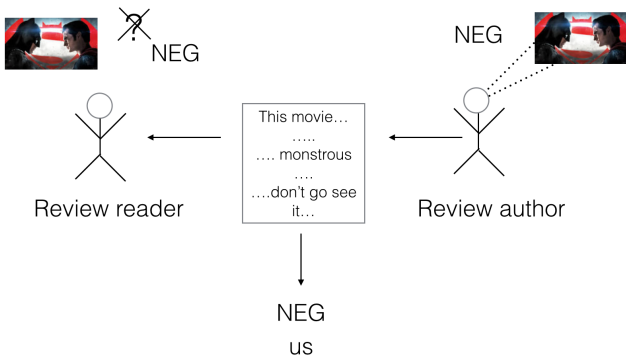


# Review sentiment





# Review sentiment



# From a good review

... He's incredible in fights. ... Also his relationship with Irons, who plays Alfred, is just wonderful in general. Irons was exceptional in the role.

# A bad review

This movie tries so hard... It completely fails on every single level. The movie is tedious and boring with characters that I just did not care about at all. ...

# Experiments with movie reviews

- Lots of possible NLP experiments . . .
- Today: use data about individual words to find sentiment.
  - Sentiment **lexicon** lists over 8000 words as positive or negative.
  - Hypothesis: a review that contains more positive than negative words is positive overall.

# Experiments with movie reviews

- Lots of possible NLP experiments ...
- Today: use data about individual words to find sentiment.
  - Sentiment **lexicon** lists over 8000 words as positive or negative.
  - Hypothesis: a review that contains more positive than negative words is positive overall.

```
word=foul intensity=weak polarity=negative
word=mirage intensity=strong polarity=negative
word=aggression intensity=strong polarity=negative
word=eligible intensity=weak polarity=positive
word=chatter intensity=strong polarity=negative
```

Note: a lexicon is a list of words with some associated information.

# Sentiment lexicon words in the good review

... He's incredible in fights. ... Also his relationship with Irons, who plays Alfred, is just wonderful in general. Irons was exceptional in the role.

- incredible **positive**
- wonderful **positive**
- exceptional **positive**

# Sentiment lexicon words in the bad review

This movie tries so hard... It completely fails on every single level. The movie is tedious and boring with characters that I just did not care about at all. ...

- try **negative**
- fail **negative**
- tedious **negative**
- boring **negative**
- care **positive**

## But it doesn't always work . . .

This movie tries so hard... **The ending should be exciting and fun and amazing.. and it just... wasn't.** It completely fails on every single level. The movie is tedious and boring with characters that I just did not care about at all. ...

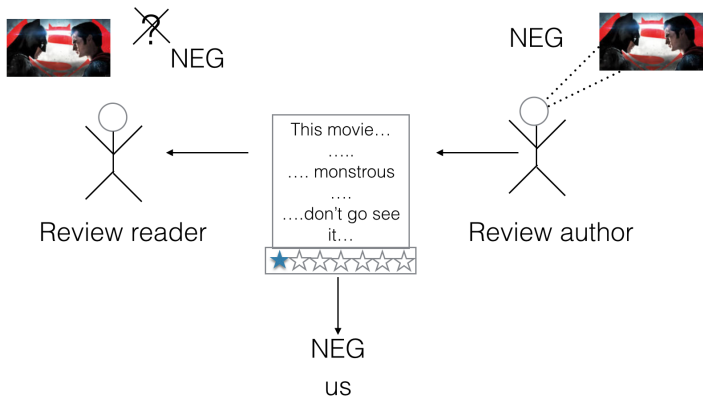
- try **negative**
- exciting **positive**
- fun **positive**
- amazing **positive**
- fail **negative**
- tedious **negative**
- boring **negative**
- care **positive**



# Evaluation

- No system predicts sentiment perfectly.
- How do we know the extent to which we've got it right?
- The author of the review told us the truth explicitly via a star rating (that's why NLP researchers like movie reviews).
- The rating has been extracted along with the review text.
- We will calculate a metric called  $A$  (accuracy).

# Star rating



# Accuracy

- The number of correct decisions  $c$  divided by total decisions (correct plus incorrect ( $i$ )):

$$A = \frac{c}{c + i}$$

- This metric is called  $A$  (accuracy).
- We know which decisions are “correct” because we can use the star rating as our definition of truth.

# Tokenisation: getting the words out

- Your code will look up words from your review document in the lexicon.
- So it needs to divide the text into words.
- Splitting on whitespace is not enough.
  - Words at the beginning of a sentence appear in upper case.
  - Words occurring before and after punctuation may be directly attached to the punctuation.
  - and many other things ...
- Your code will use a well-known basic **tokeniser** to split the text into individual words.
- Note: **type** vs **token** (see 'Further notes' in Session 2)

# Your tasks for today

## Task 1:

- explore the review data (1800 documents)
- explore the sentiment lexicon
- write a program that tests the sentiment lexicon approach
- write a program for using the star ratings to evaluate how well your program is doing
- and keep a record of what you do

# Example lab book page

## Simple Classifier: Virtual Programming Lab

16/1/18

Dataset: 900 positive, 900 negative reviews

### Simple Classifier

#### Method:

- Read a lexicon for positive & negative words
- Ignore neutral words
- loop through dataset, incrementing/decrementing a sentiment value
- Assign sentiment to review if sentiment value  $\geq$  threshold otherwise negative
- Set threshold to 0 as default

### Improving Classifier Options

- ① Change threshold to account for a natural bias, to use more +/-ive words
- ② Weighted lexicon relative to strong/weak subj.

### Results

Simple Classifier Accuracy: 63.5%

### Improved Classifier

Option 1:		Option 2 (Threshold=0)		Option 3 (Threshold=10)	
Threshold	Accuracy	Weight	Accuracy	Weight	Accuracy
0	63.5%	1	63.5%	1	68.0%
5	64.9%	2	64.3%	2	68.8%
10	68.0%	5	65.8%	5	68.6%

### Observation/Discussion

- Possible tendency to use more positive words in reviews  
for example, the negative review #652 starts with "Perhaps best remembered..."

# Practicalities

- 16 lectures (approx 25 minutes) [Mon, Fri]
- 16 demonstrated sessions in the Intel Lab: from immediately after lecture to 4:30pm [Mon, Fri]
- 12 tasks and 4 catch-up sessions
- 12 ticks: you should get them all
- Most tasks have automated tester: pass this first!
- Ticking during demonstrated sessions, queue on whiteboards.
- Lots more on Moodle ...