Opinion Mining: Exploiting the Sentiment of the Crowd

Diana Maynard
Adam Funk
Kalina Bontcheva

University of Sheffield, UK

Burning questions you may have

- Which is more accurate: "Phone a Friend" or "Ask the Audience"?
- Could the "Fear Index" become reality?
- What does the Royal Wedding have to do with Pilates classes?
- Do people feel more miserable when stock prices fall?
- Can Twitter predict earthquakes?
- Can sentiment analysis find us the perfect husband or wife?

Aims of this tutorial

- Introduce the concepts of opinion mining and sentiment analysis from unstructured text
  - Why are they useful?
  - What tools and techniques are available?
- Introduce some general rule-based and machine learning techniques
- Take a look at what kind of problems are posed by opinion mining in general
- Take a look at some problems specific to processing social media

Tutorial Structure

09.00 – 10.00 Introduction to Opinion Mining

10.00 – 10.30 Machine Learning Applications

10.30 – 11.00 Coffee Break

11.00 – 12.00 Rule-based Applications

12.00 – 13.00 Opinion MIning and Social Media

Introduction to Opinion Mining

What is Opinion Mining?

- OM is a relatively recent discipline that studies the extraction of opinions using IR, AI and/or NLP techniques.
- More informally, it's about extracting the opinions or sentiments given in a piece of text
- Also referred to as Sentiment Analysis (though technically this is a more specific task)
- Web 2.0 nowadays provides a great medium for people to share things.
- This provides a great source of unstructured information (especially opinions) that may be useful to others (e.g. companies and their rivals, other consumers...)

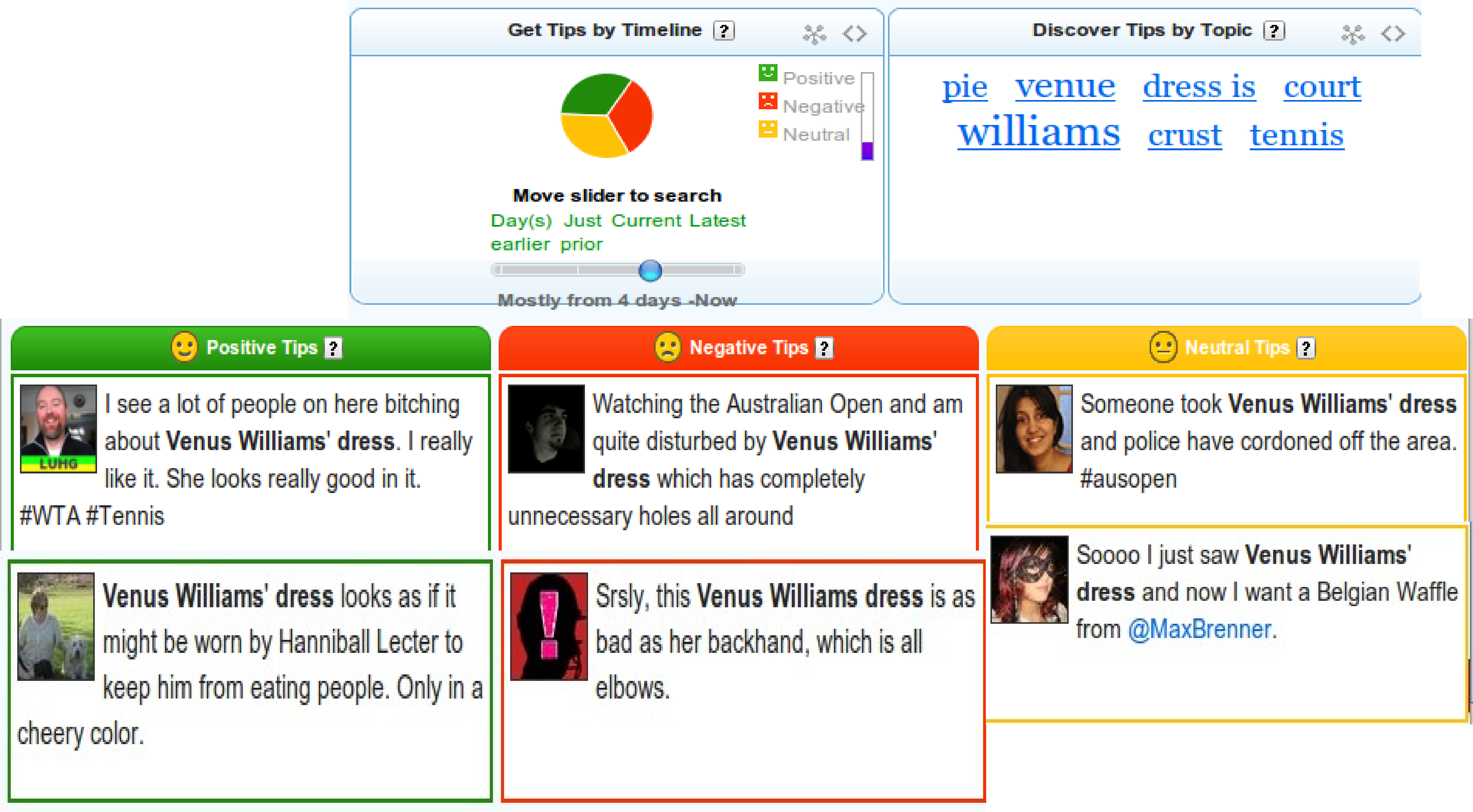
It's about finding out what people think...

Opinion Mining is Big Business

- Someone who wants to buy a camera
  - Looks for comments and reviews
- Someone who just bought a camera
  - Comments on it
  - Writes about their experience
- Camera Manufacturer
  - Gets feedback from customer
  - Improve their products
  - Adjust Marketing Strategies

Venus Williams causes controversy...

Opinion mining exposes these insights



I see a lot of people on here bitching about **Venus Williams' dress**. I really like it. She looks really good in it. #WTA #Tennis

Venus Williams' dress looks as if it might be worn by Hanniball Lecter to keep him from eating people. Only in a cheery color.

Watching the Australian Open and am quite disturbed by **Venus Williams' dress** which has completely unnecessary holes all around

Srsly, this **Venus Williams dress** is as bad as her backhand, which is all elbows.

Someone took **Venus Williams' dress** and police have cordoned off the area. #ausopen

Soooo I just saw **Venus Williams' dress** and now I want a Belgian Waffle from @MaxBrenner.

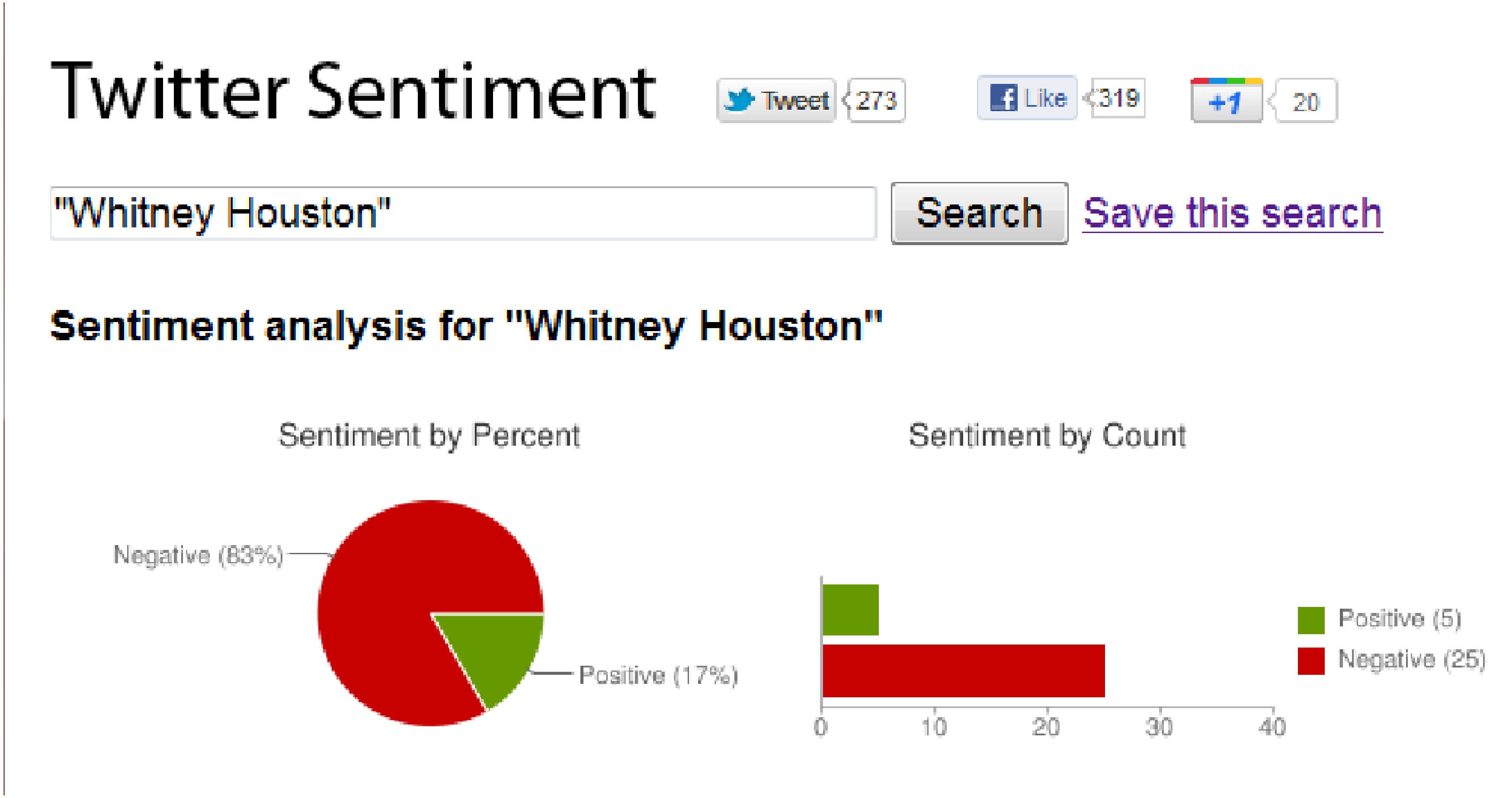
Online social media sentiment apps

- Try a search of your own on one of these:
  - Twitter sentiment http://twittersentiment.appspot.com/
  - Twends: http://twendz.waggeneredstrom.com/ http://twendz.waggeneredstrom.com/
  - Twittratr: http://twitrratr.com/
  - SocialMention: http://socialmention.com/
- Easy to search for opinions about famous people, brands and so on
- Hard to search for more abstract concepts, perform a non-keyword based string search
- e.g. to find opinions about Venus Williams, you can only search on “Venus Williams” to get hits

Why are these sites unsuccessful?

- They don't work well at more than a very basic level
- They mainly use dictionary lookup for positive and negative words
- They classify the tweets as positive or negative, but not with respect to the keyword you're searching for
- First, the keyword search just retrieves any tweet mentioning it, but not necessarily about it as a topic
- Second, there is no correlation between the keyword and the sentiment: the sentiment refers to the tweet as a whole
- Sometimes this is fine, but it can also go horribly wrong

Whitney Houston wasn't very popular...

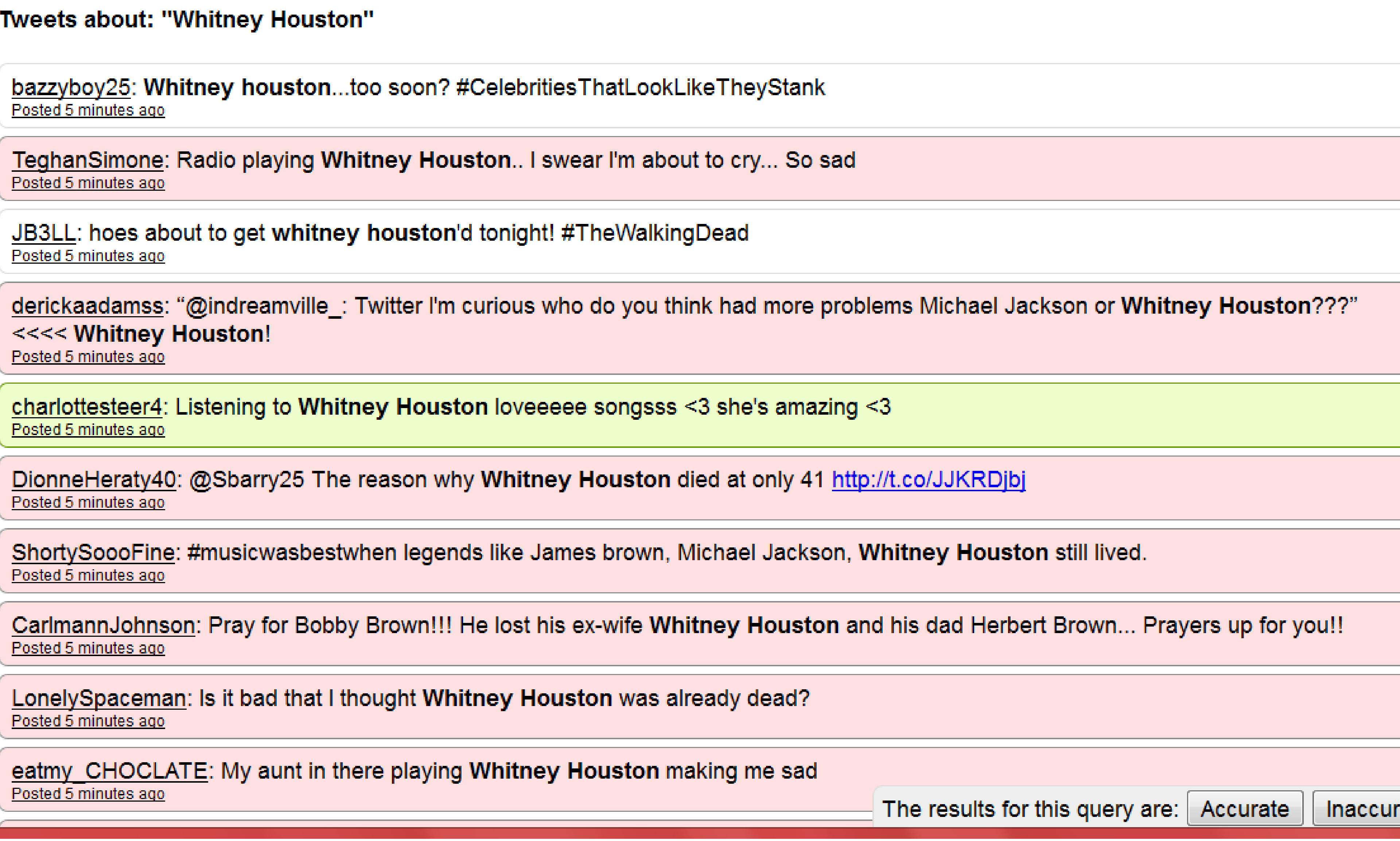


Twitter Sentiment

"Whitney Houston"

Sentiment analysis for "Whitney Houston"

Or was she?



Tweets about: "Whitney Houston"

bazzyboy25: **Whitney houston**...too soon? #CelebritiesThatLookLikeTheyStank
Posted 5 minutes ago

TeghanSimone: Radio playing **Whitney Houston**.. I swear I'm about to cry... So sad
Posted 5 minutes ago

JB3LL: hoes about to get **whitney houston**'d tonight! #TheWalkingDead
Posted 5 minutes ago

derickaadamss: “@indreamville_: Twitter I'm curious who do you think had more problems Michael Jackson or **Whitney Houston**???” <<<< **Whitney Houston**!
Posted 5 minutes ago

charlottesteer4: Listening to **Whitney Houston** loveeeee songsss <3 she's amazing <3
Posted 5 minutes ago

DionneHeraty40: @Sbarry25 The reason why **Whitney Houston** died at only 41 http://t.co/JJKRDjbj
Posted 5 minutes ago

ShortySoooFine: #musicwasbestwhen legends like James brown, Michael Jackson, **Whitney Houston** still lived.
Posted 5 minutes ago

CarlmannJohnson: Pray for Bobby Brown!!! He lost his ex-wife **Whitney Houston** and his dad Herbert Brown... Prayers up for you!!
Posted 5 minutes ago

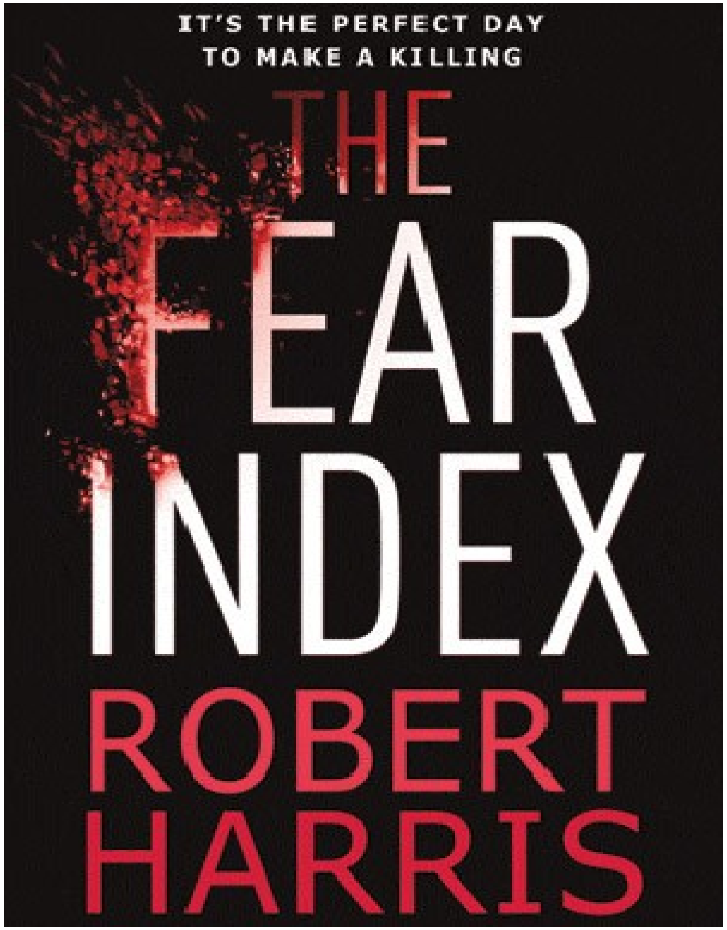
LonelySpaceman: Is it bad that I thought **Whitney Houston** was already dead?
Posted 5 minutes ago

eatmy_CHOCLATE: My aunt in there playing **Whitney Houston** making me sad
Posted 5 minutes ago

The results for this query are:

Opinion Mining for Stock Market Prediction

- It might be only fiction, but using opinion mining for stock market prediction has been already a reality for some years
- Research shows that opinion mining outperforms event-based classification for trend prediction [Bollen2011]
- At least one investment company currently offers a product based on opinion mining

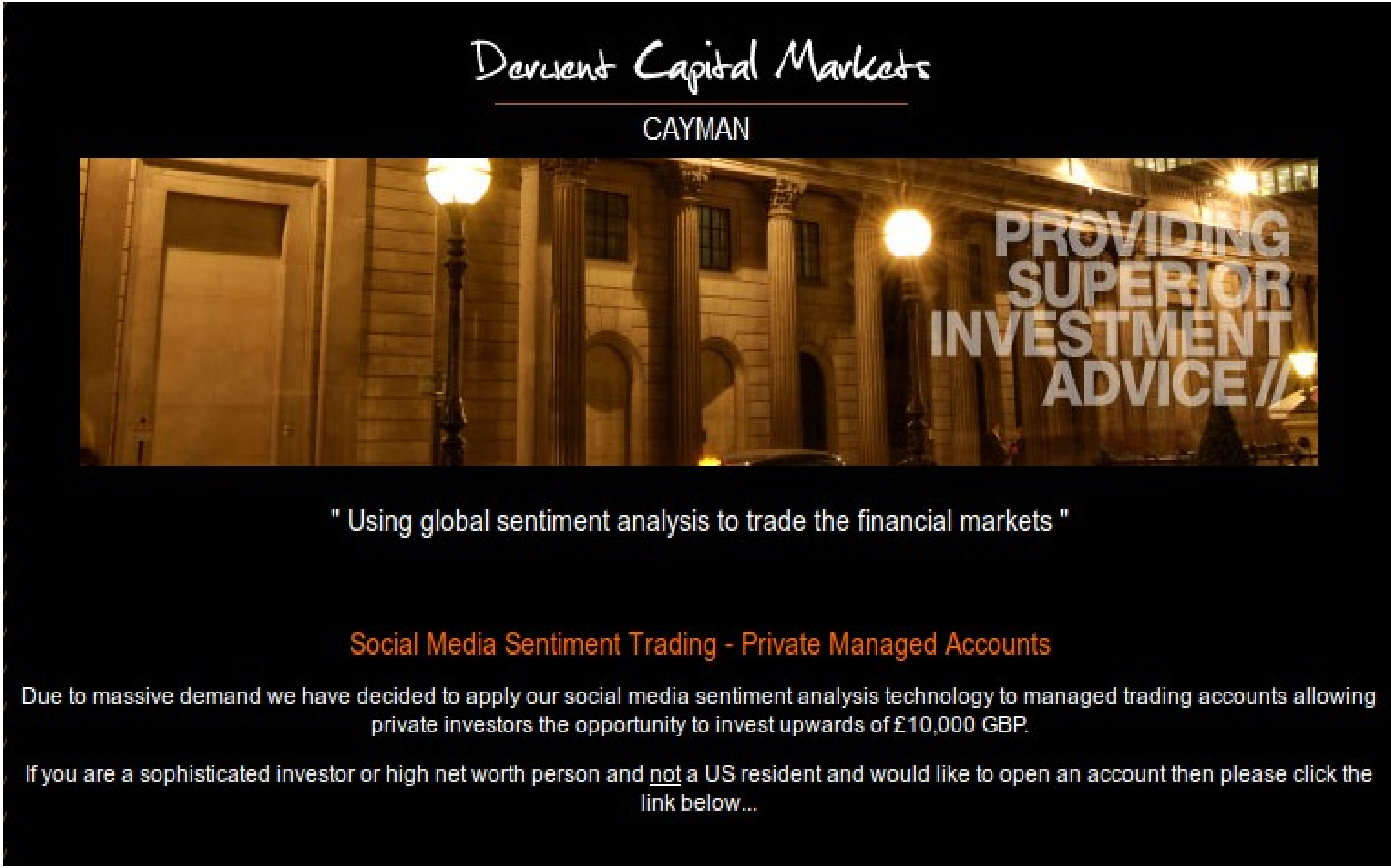


Using Twitter for Stock Market Prediction

“Hey Jon, Derek in Scunthorpe's having a bacon and egg, er, butty. Is that good for wheat futures?”

Derwent Capital Markets

- Derwent Capital Markets have launched a £25m fund that makes its investments by evaluating whether people are generally happy, sad, anxious or tired, because they believe it will predict whether the market will move up or down.
- Bollen told the Sunday Times: "We recorded the sentiment of the online community, but we couldn't prove if it was correct. So we looked at the Dow Jones to see if there was a correlation. We believed that if the markets fell, then the mood of people on Twitter would fall."
- "But we realised it was the other way round — that a drop in the mood or sentiment of the online community would precede a fall in the market."



Derwent Capital Markets

CAYMAN

" Using global sentiment analysis to trade the financial markets "

Social Media Sentiment Trading - Private Managed Accounts

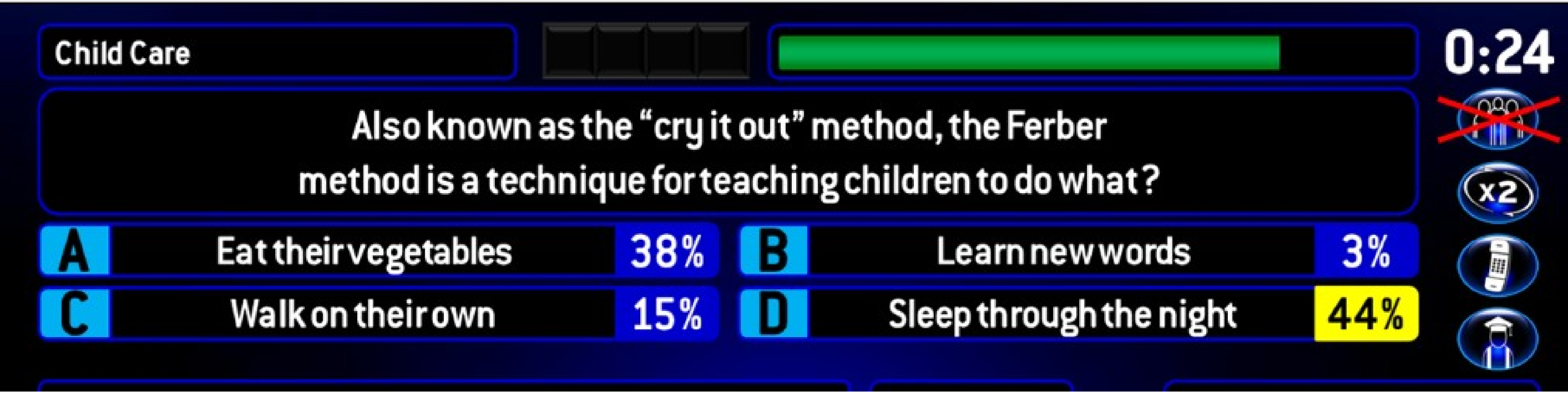
Due to massive demand we have decided to apply our social media sentiment analysis technology to managed trading accounts allowing private investors the opportunity to invest upwards of £10,000 GBP.

If you are a sophisticated investor or high net worth person and not a US resident and would like to open an account then please click the link below...

But don't believe all you read...

- It's been suggested recently that there are actually many flaws in Bollen's work, and that it's impossible to predict the stock market in this way
- If it were really possible, surely Bollen would be a millionaire by now and everyone would be using this technology?
- There's quite a lot of sloppiness in the reporting of methodology and results, so it's not clear what can really be trusted
- The advertised results are biased by selection (they picked the winners after the race and tried to show correlation)
- The accuracy claim is too general to be useful (you can't predict individual stock prices, only the general trend)
- http://sellthenews.tumblr.com/post/21067996377/noitdoesnot

Who Wants to be a Millionaire?



Child Care

0:24

Also known as the “cry it out” method, the Ferber method is a technique for teaching children to do what?

A Eat their vegetables 38%

B Learn new words 3%

C Walk on their own 15%

D Sleep through the night 44%

Ask the audience?

Or phone a friend?

Which do you think is better?

What's the capital of Spain?

A: Barcelona

B: Madrid

C: Valencia

D: Seville

What's the height of Mt Kilimanjaro?

A: 19,341 ft

B: 23,341 ft

C: 15,341 ft

D: 21,341 ft

Go for the majority or trust an expert?

- It depends what kind of question you're asking
- In Who Wants to Be a Millionaire, people tend to ask the audience fairly early on, because once the questions get hard, they can't rely on the audience getting it right

What's the height of Mt Kilimanjaro?

A: 19,341 ft
B: 23,341 ft
C: 15,341 ft
D: 21,341 ft

What's the capital of Spain?

A: Barcelona
B: Madrid
C: Valencia
D: Seville

Why bother with opinion mining?

- It depends what kind of information you want
- Don't use opinion mining tools to help you win money on quiz shows
- Recent research has shown that one knowledgeable analyst is better than gathering general public sentiment from lots of analysts and taking the majority opinion
- But only for some kinds of tasks

Whose opinion should you trust?

- Opinion mining gets difficult when the users are exposed to opinions from more than one analyst
- Intuitively, one would probably trust the opinion supported by the majority.
- But some research shows that the user is better off trusting the most credible analyst.
- Then the question becomes: who is the most credible analyst?
- Notions of trust, authority and influence are all related to opinion mining

All opinions are not equal

- Opinion Mining needs to take into account how much influence any single opinion is worth
- This could depend on a variety of factors, such as how much trust we have in a person's opinion, and even what sort of person they are
- Need to account for:
  - experts vs non-experts
  - spammers
  - frequent vs infrequent posters
  - “experts” in one area may not be expert in another
  - how frequently do other people agree?

Trust Recommenders

- Two types of trust:
  - relationship (local) trust
  - reputation (global) trust.
- **Relationship trust**: if you and I both rate the same things, and our opinions on them match closely, we have high relationship trust. This can be extended to a social networking group --> web of trust.
- **Reputation trust**: if you've recommended the same thing as other people, and usually your recommendation is close to what the majority of people think, then you're considered to be more of an expert and have high reputation trust.
- We can extend relationship trust to form clusters of interests and likes/dislikes
- We can narrow reputation trust to opinions about similar topics

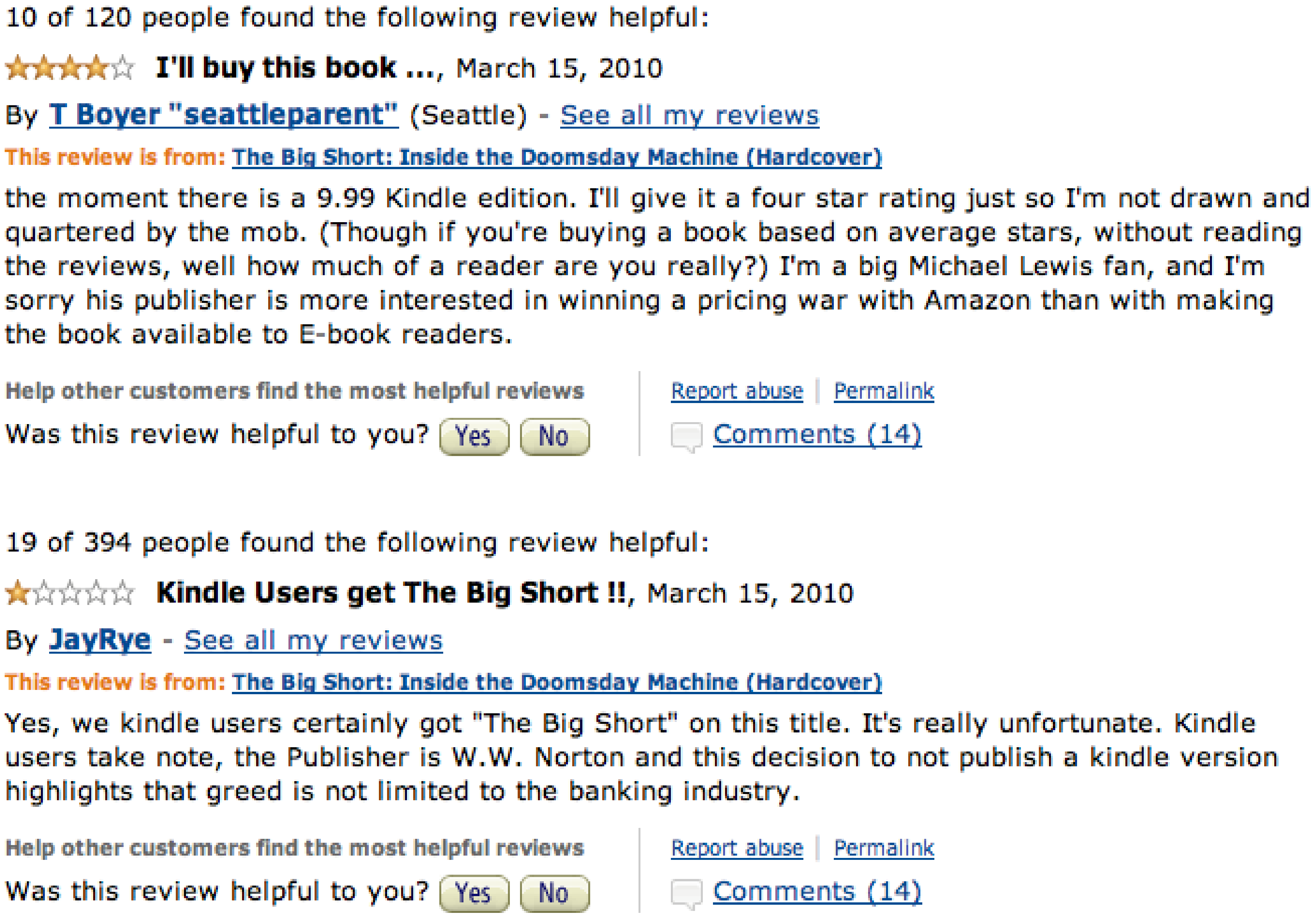
Opinion Mining Subtasks

- **Opinion extraction**: extract the piece of text which represents the opinion
  - I just bought a new camera yesterday. It was a bit expensive, but the battery life is very good.
- **Sentiment classification/orientation:** extract the polarity of the opinion (e.g. positive, negative, neutral, or classify on a numerical scale)
  - negative: expensive
  - positive: good battery life
- **Opinion summarisation**: summarise the overall opinion about something
  - price:negative, battery life: positive --> overall 7/10

Feature-opinion association

- **Feature-opinion association**: given a text with target features and opinions extracted, decide which opinions comment on which features.
  - “The battery life is good but not so keen on the picture quality”
- **Target identification**: which thing is the opinion referring to?
- **Source identification**: who is holding the opinion?
- There may be attachment and co-reference issues
  - “The camera comes with a free case but I don't like the colour much.”
  - Does this refer to the colour of the case or the camera?

Getting the target of the opinion right is crucial



10 of 120 people found the following review helpful:

★★★★☆ **I'll buy this book ...**, March 15, 2010

By T Boyer "seattleparent" (Seattle) - See all my reviews

This review is from: The Big Short: Inside the Doomsday Machine (Hardcover)

the moment there is a 9.99 Kindle edition. I'll give it a four star rating just so I'm not drawn and quartered by the mob. (Though if you're buying a book based on average stars, without reading the reviews, well how much of a reader are you really?) I'm a big Michael Lewis fan, and I'm sorry his publisher is more interested in winning a pricing war with Amazon than with making the book available to E-book readers.

Help other customers find the most helpful reviews

Was this review helpful to you? Yes No

Report abuse | Permalink

Comments (14)

19 of 394 people found the following review helpful:

★☆☆☆☆ **Kindle Users get The Big Short !!**, March 15, 2010

By JayRye - See all my reviews

This review is from: The Big Short: Inside the Doomsday Machine (Hardcover)

Yes, we kindle users certainly got "The Big Short" on this title. It's really unfortunate. Kindle users take note, the Publisher is W.W. Norton and this decision to not publish a kindle version highlights that greed is not limited to the banking industry.

Help other customers find the most helpful reviews

Was this review helpful to you? Yes No

Report abuse | Permalink

Comments (14)

Opinion spamming

Suppose we run a contest where people retweet our ad repeatedly, and the winner's whoever loses the most followers.

Spam opinion detection (fake reviews)

- Sometimes people get paid to post “spam” opinions supporting a product, organisation or even government
- An article in the New York Times discussed one such company who gave big discounts to post a 5-star review about the product on Amazon
- http://www.nytimes.com/2012/01/27/technology/for-2-a-star-a-retailer-
- Could be either positive or negative opinions
- Generally, negative opinions are more damaging than positive ones

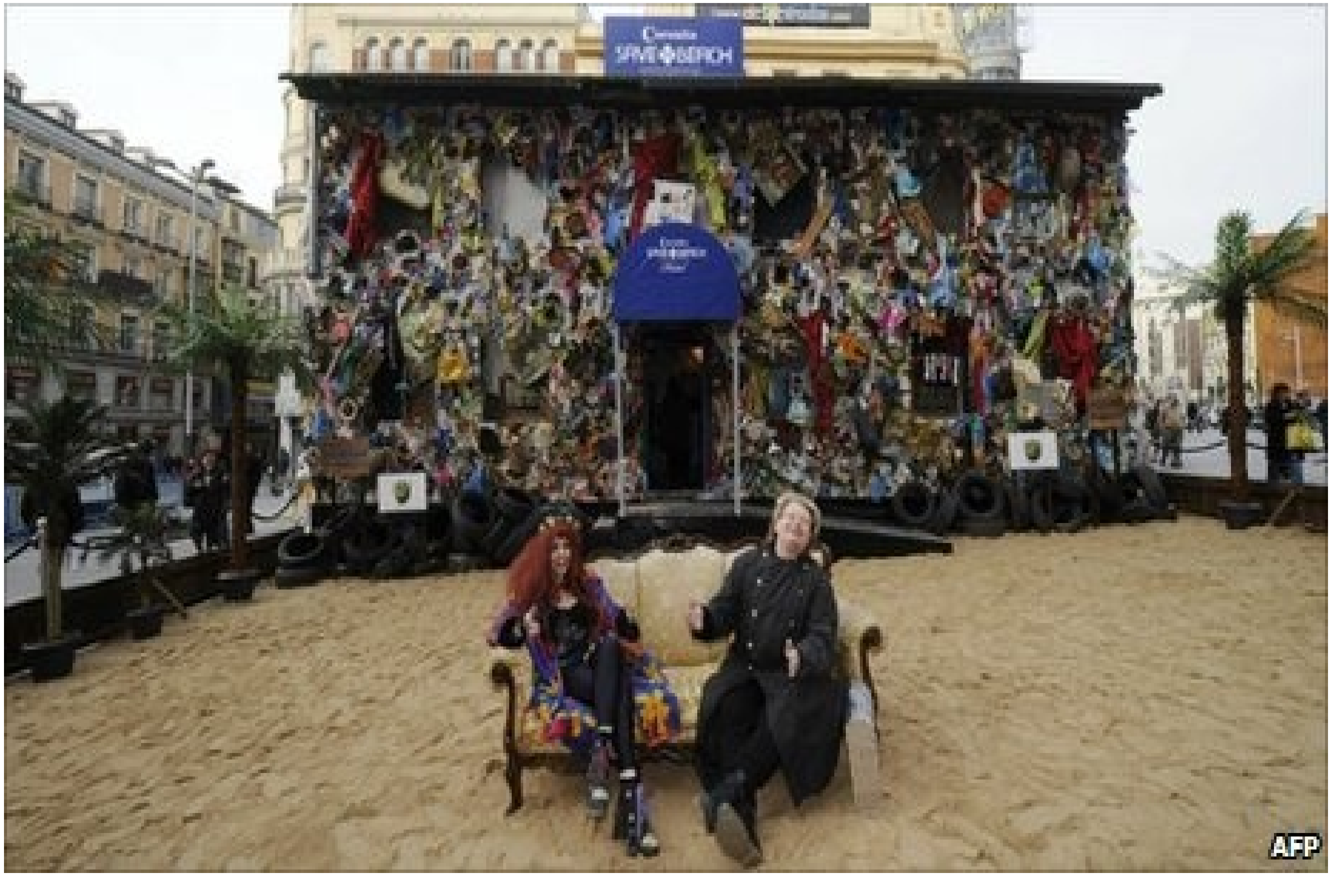
How to detect fake opinions?

- Review content: lexical features, content and style inconsistencies from the same user, or simlarities between different users
- Complex relationships between reviews, reviewers and products
- Publicly available information about posters (time posted, posting frequency etc)
- See anything wrong with these reviews? http://www.amazon.com/gp/pdp/profile/A3URRTIZEE8R7W

It's not just about cameras and dresses ...

- Film, theatre, books, fashion etc
  - impacts on the whole industry
  - predictions about changing society, trends etc.
- Monitoring political views
- Feedback/opinions about multimedia productions, e.g. documentaries, broadcasts etc.
- Feedback about events, e.g. conferences
- Scientific and technological monitoring, competitor surveillance etc.
- Monitoring public opinion
- Creating community memories

And it's not always as easy as it looks...



“Rubbish hotel in Madrid”

Opinion mining and social media

- Social media provides a wealth of information about a user's behaviour and interests:
  - *explicit*: John likes tennis, swimming and classical music
  - *implicit*: people who like skydiving tend to be big risk-takers
  - *associative*: people who buy Nike products also tend to buy Apple products
- While information about individuals isn't useful on its own, finding defined clusters of interests and opinions is

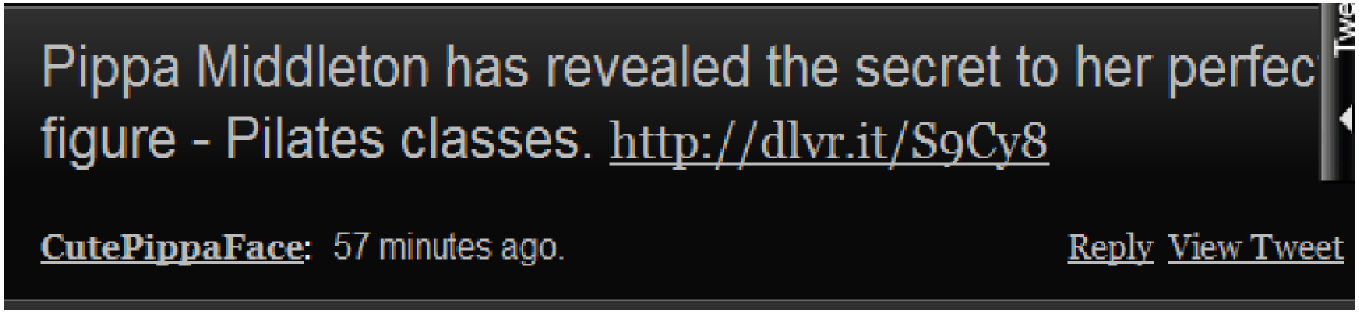
  If many people talk on social media sites about fears in airline security, life insurance companies might consider opportunities to sell a new service
- This kind of predictive analysis is all about understanding your potential audience at a much deeper level - this can lead to improved advertising techniques such as personalised ads to different groups

Analysing and preserving opinions

- Useful to collect, store and later retrieve public opinions about events and their changes or developments over time
- One of the difficulties lies in distinguishing what is important
- Opinion mining tools can help here
- Not only can online social networks provide a snapshot of such situations, but they can actually trigger a chain of reactions and events
- Ultimately these events might lead to societal, political or administrative changes

Pippa Middleton's assets

- One of the biggest Royal Wedding stories on Social Media sites
- Her bottom has its own twitter account, facebook page and website.
- Pilates classes have become incredibly popular since the Royal Wedding, solely as a result of all the social media



Pippa Middleton has revealed the secret to her perfec figure - Pilates classes. http://dlvr.it/S9Cy8

CutePippaFace: 57 minutes ago.

Accuracy of twitter sentiment apps

- Mine the social media sentiment apps and you'll find a huge difference of opinions about Pippa Middleton:
  - TweetFeel: 25% positive, 75% negative
  - Twendz: no results
  - TipTop: 42% positive, 11% negative
  - Twitter Sentiment: 62% positive, 38% negative
- Try searching for “Gaddafi” and you may be surprised at some of the results.

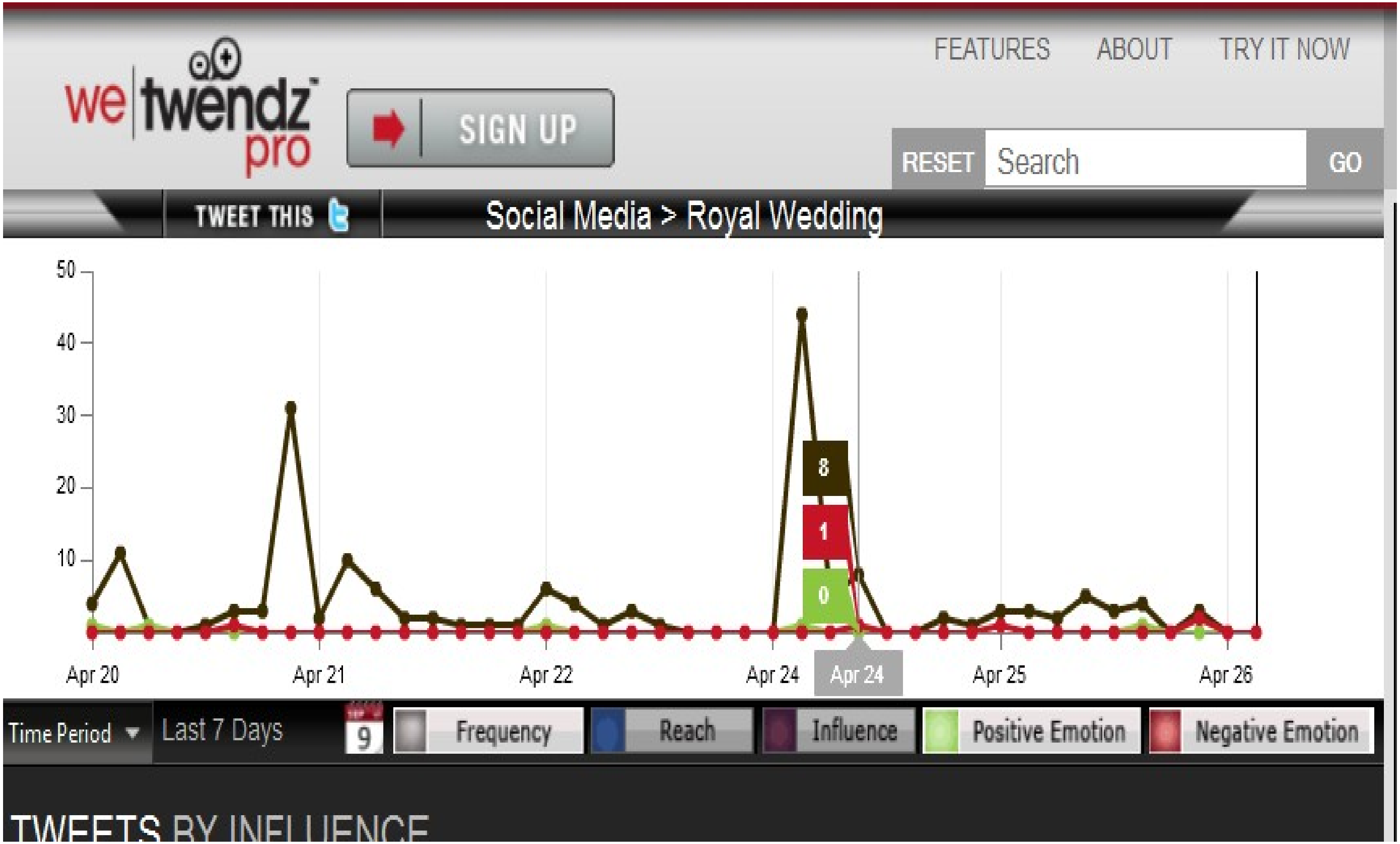
Twittrater's view of the Olympics

- A keyword search for Olympics shows exactly how existing systems fail to cut the mustard
- Lookup of sentiment words is not enough if
  - they're part of longer words
  - they're used in different contexts
  - the tweet itself isn't relevant
  - they're used in a negative or sarcastic sentence
  - they're ambiguous

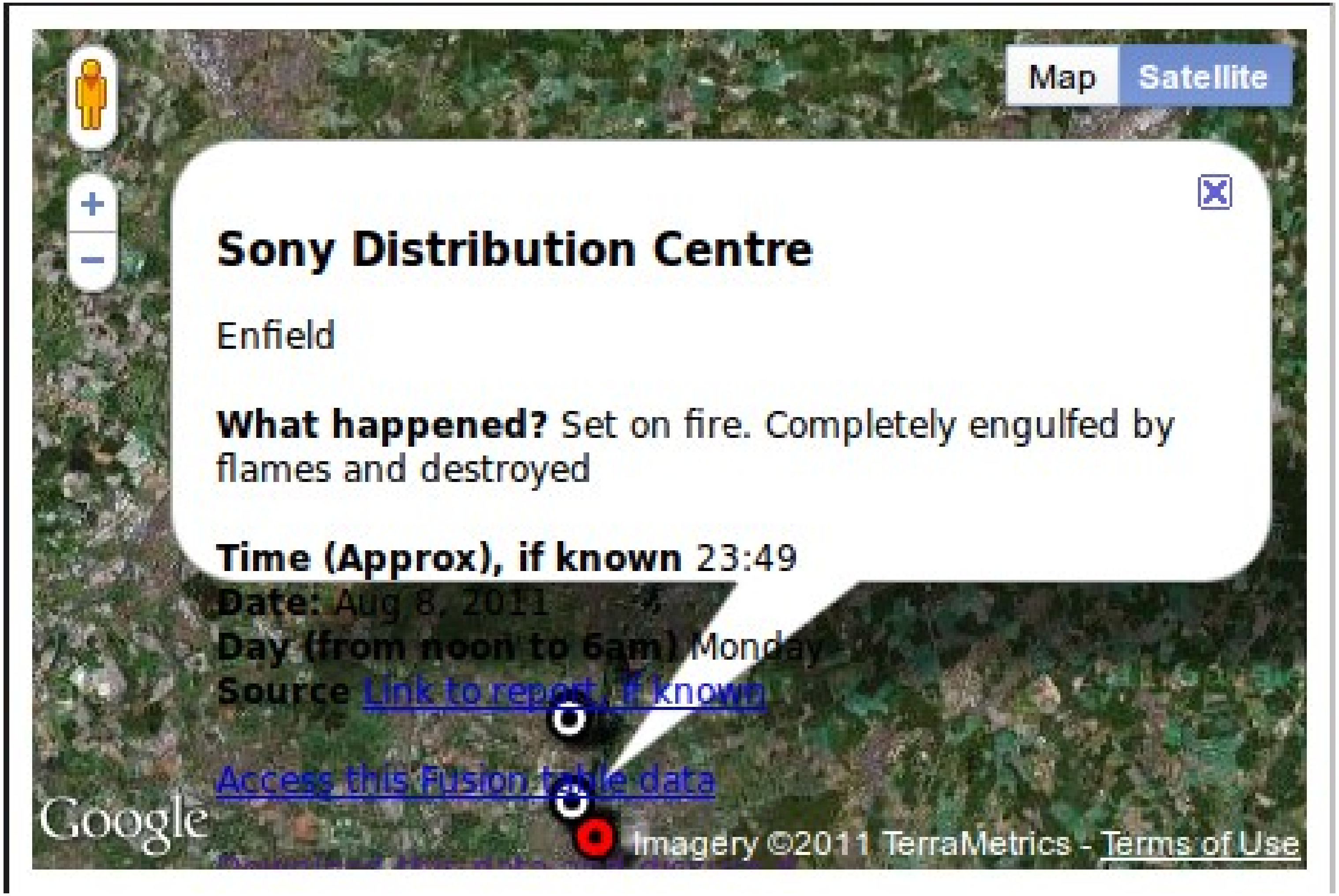
Tracking opinions over time

- Opinions can be extracted with a time stamp and/or a geo-location
- We can then analyse changes to opinions about the same entity/event over time, and other statistics
- We can also measure the impact of an entity or event on the overall sentiment about an entity or another event, over the course of time (e.g. in politics)
- Also possible to incorporate statistical (non-linguistic) techniques to investigate dynamics of opinions, e.g. find statistical correlations between interest in certain topics or entities/events and number/impact/influence of tweets etc.

Viewing opinion changes over time



Mapping dynamics from social media: UK riots demo



Sony Distribution Centre

Enfield

What happened? Set on fire. Completely engulfed by flames and destroyed

Time (Approx), if known 23:49
Date: Aug 8, 2011
Day (from noon to 6am) Monday
Source Link to report, if known

Access this Fusion table data

Predicting the future

Predicting other people's decisions

- It would be useful to predict what products people will buy, what films they want to see, or what political party they'll support

MY INVENTION CAN SCAN A PERSON'S BRAIN AND PREDICT HIS BUYING DECISIONS.
Dilbert.com DilbertCartoonist@gmail.com
IT SAYS YOU PLAN TO BUY. . . A BLUNT OBJECT SO YOU CAN KILL ME AND CLAIM CREDIT FOR MY INVENTION.
7-22-10 ©2010 Scott Adams, Inc./Dist. by UFS, Inc.
I ANTICIPATED THAT, WHICH IS WHY I INCLUDED A FEATURE TO MICROWAVE THE OFFENDING PART OF YOUR BRAIN.
FERT!

Predicting Presidential Candidates

- Michael Wu from Lithium did a study of sentiment data on various social web apps about presidential candidates in March 2012
- http://lithosphere.lithium.com/t5/Building-Community-the-Platform/E
- His analysis involved taking the positive sentiments minus the negative sentiments, over a 2 week period, and also including the neutral sentiments
- Neutral sentiments were weighted at 1/10 and added to the net sentiment
- He saw a close correlation between his analysis and the Gallup polls, but he warns us to be cautious...

Predictive Analysis Windows

- Predictive analytics is about trying to look into the future through the predictive window of your data.
- If you try to look outside this window, your future will look very blurry.
- It's like weather forecasting – the smaller the window, the more accurate you'll be
- The important question is not whether social media data can predict election outcome, but “how far ahead can it be predicted?”
- For something that changes very quickly like the financial market, the predictive window will be very short.
- For things that do not change as fast, the predictive window will be longer.
- For social media sentiment data, the window for election forecasting is about 1.5 to 2 weeks, (1 to be conservative).

Aggregate sentiment finding

- Aggregate sentiment finding (e.g. O'Connor et al 2010) uses shallow techniques based on sentiment word counting.
- Idea is that if you're only trying to find aggregates then such techniques are sufficient, even though they're far from perfect.
- Although the error rate can be high, with a fairly large number of measurements, these errors will cancel out relative to the quantity we are interested in estimating (aggregate public opinion).
- The claim is that using standard text analytics techniques on such data can actually be harmful, because they're designed to optimise per-document classification accuracy rather than assessing aggregate population proportions.
- Their method shows some correlation with public sentiment polls but they conclude that better opinion mining would be beneficial.

Social media and politics

- Twitter provides real-time feedback on debates that's much faster than traditional polling. Campaigns are paying close attention. That's because such chatter can gauge how a candidate's message is being received or even warn of a popularity dive.
- Campaigns that closely monitor the Twittersphere have a better feel of voter sentiment. That allows candidates to fine-tune their message for a particular state: “playing to your audience".
- However, applying complex algorithms to Twitter data, blogs, news sites and other media isn't yet perfect for predicting politics, e.g. you can't detect sarcasm reliably.
- Nevertheless, Twitter has played a role in intelligence gathering on uprisings around the world, showing accuracy at gauging political sentiment.
- http://www.usatoday.com/tech/news/story/2012-03-05/social-super-tuesday-prediction/53374536/1

Methods for Opinion Mining

- Machine learning methods
- Rule-based methods

GATE (General Architecture for Text Engineering)

- Our examples mostly use GATE – tool for LE in development in Sheffield since 2000.
- GATE includes:
  - **components** for language processing, e.g. parsers, machine learning tools, stemmers, IR tools, IE components for various languages...
  - tools for **visualising** and **manipulating** text, annotations, ontologies, parse trees, etc.
  - **various information extraction** tools
  - **evaluation** and **benchmarking** tools
- More info and freely available at http://gate.ac.uk

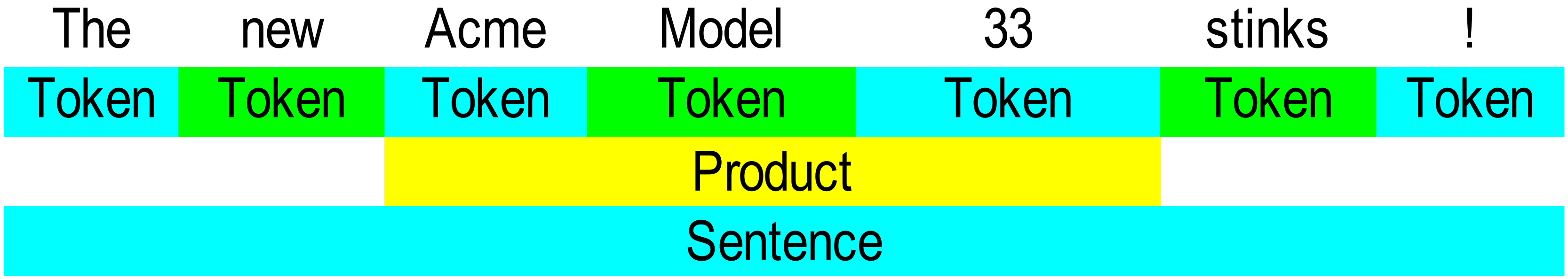
Machine learning

What is Machine learning?

- Automating the process of inferring new data from existing data
- In GATE, that means creating annotations or adding features to annotations by learning how they relate to other annotations

Learning a pattern

- For example, we have Token annotations with string features and Product annotations



The new Acme Model 33 stinks !

Token Token Token Token Token Token Token

Product

Sentence

- ML could learn that a Product close to the Token “stinks” expresses a negative sentiment, then add a polarity=“negative” feature to the Sentence.

How is that better than a rule-based approach?

- Not necessarily better, just different
- People are better at writing rules for some things, ML algorithms are better at finding some things
- With ML you don't have to create all the rules, **but** you have to manually annotate a training corpus—or get someone else to do it!
- Rule-based approaches (such as JAPE) and ML work well together; in GATE, JAPE is often used extensively to prepare data for ML.

Terminology: Instances

- Instances are cases that may be learned
- Every instance is a decision for the ML algorithm to make
- To which class does this instance belong?
  - “California” → Location
  - “This product stinks” → polarity=negative

Terminology: Attributes

- Attributes are pieces of information that we already know about instances (sometimes called “features” in machine learning literature).
- These can be GATE annotations, or annotation features that will be known before the ML algorithm is applied to new data
- Examples
  - Token.string == “stinks”
  - Token.kind == “punctuation”
  - Sentence contains Product

Terminology: Classes

- The class is what we want to learn
- Suppose we want to find opinions: for every Sentence instance, the question is “What kind of opinion does this express?” and the classes are *positive*, *negative*, *neutral*, and *none*.

ML Tasks

- GATE supports 3 types of ML tasks:
  - chunk recognition (named entity recognition, NP chunking)
  - text classification (sentiment classification, POS tagging)
  - relation annotation
- Most opinion mining tasks fall under text classification

Training

- Training involves presenting data to the ML algorithm from which it creates a model
- The training data consist of instances that have been annotated with correct classes as well as attributes
- Models are representations of decision-making processes that allow the ML algorithm to classify each instance based on its attributes

Application

- When the ML algorithm is applied, it creates new class annotations on data using the model
- The corpus it is applied to must contain the required attribute annotations
- The machine learner will work best if the application data is similar to the training data

Evaluation

- We want to know how good our machine learner is before we use it for a real task
- Therefore we apply it to some data for which we already have class annotations
  - the “right answers”, sometimes called “gold standard”
- If the machine learner creates the same annotations as the gold standard, then we know it is performing well
- GATE's ML PR has a built-in evaluation mode that splits the corpus into training and test sets and cross-validates them

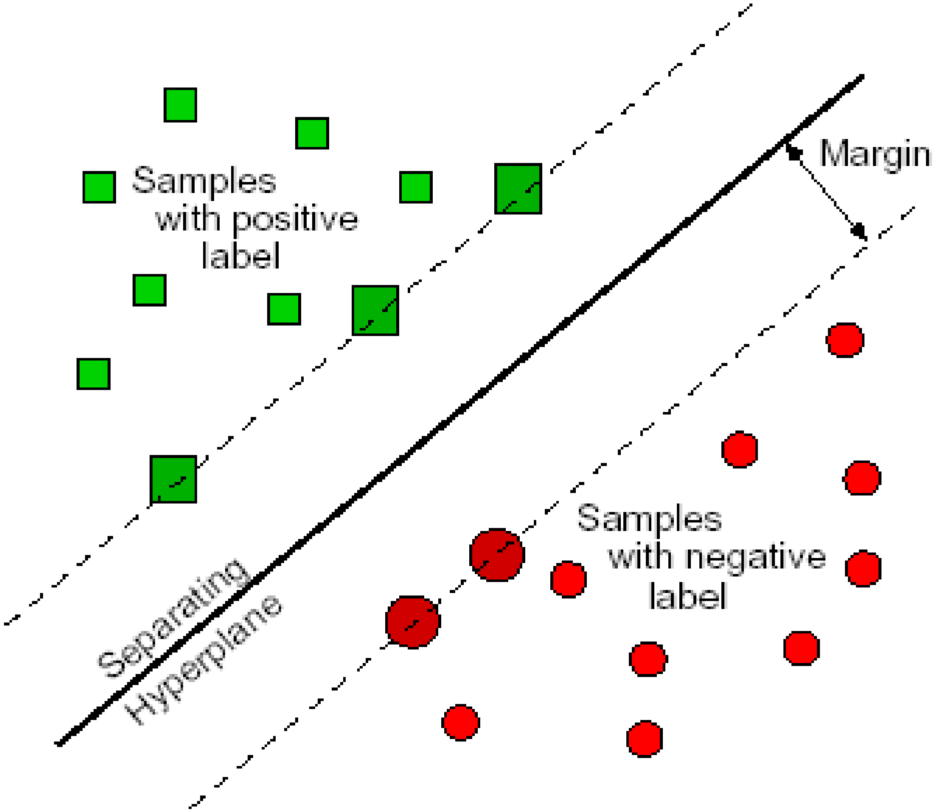
Perceptron and PAUM

- Perceptron is one of the oldest ML methods (invented in the 50s!)
- Like SVM (which will be covered later), it determines a hyperplane separator between the data points
- Theoretically SVM works a little better because it calculates the optimal separator, but in practice, however, there is usually little difference, and Perceptron is a lot faster!

Perceptron Algorithm with Uneven Margins (PAUM)

- Both Perceptron and SVM implement “uneven margins”
- This means that it doesn't position the separator centred between the points, but more towards one side

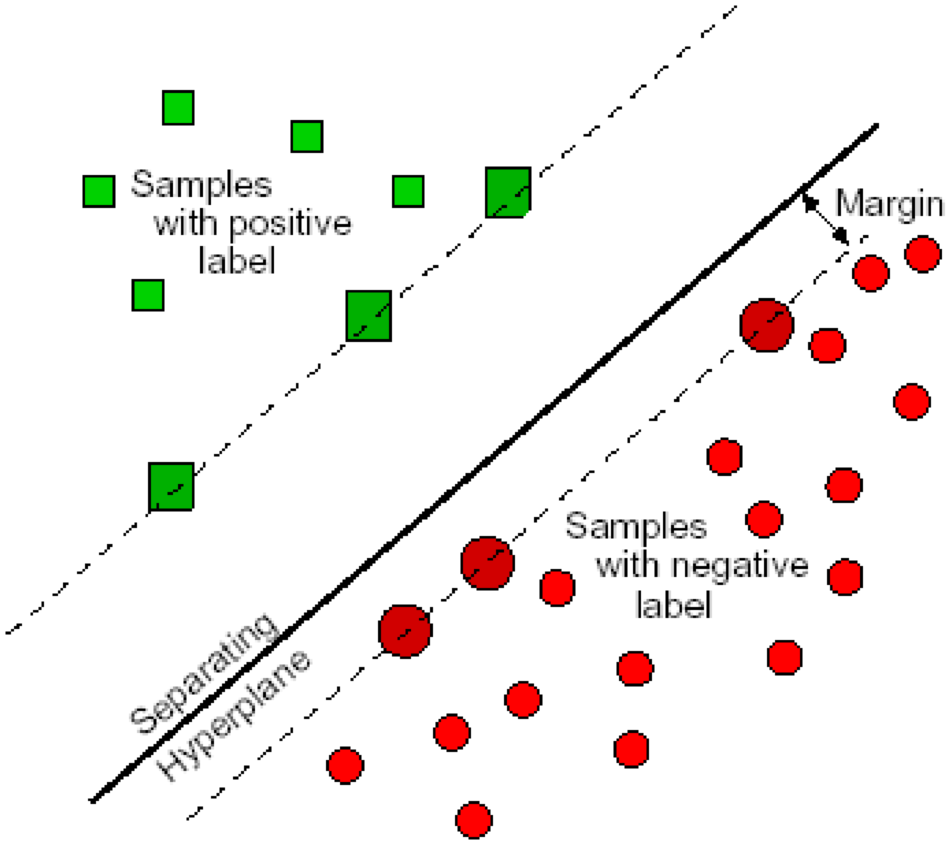
Even Margins



Why Uneven Margins?

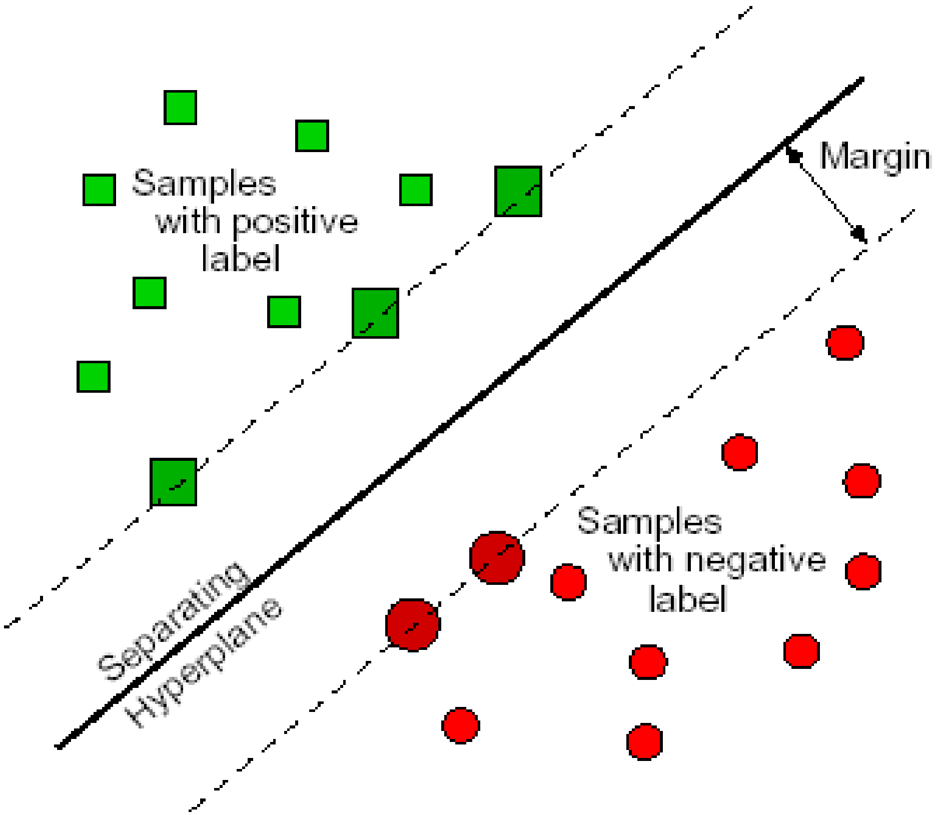
- In NLP the datasets are often very imbalanced.
- If you are tagging instances of “Person”, there are a few positive cases mixed with many words that are not persons.
- In opinion mining, you may have a few sentences with opinions but mostly sentences without them.
- So move the margin away from the smaller group of training examples.
- Y. Li, K. Bontcheva, and H. Cunningham. Using Uneven Margins SVM and Perceptron for Information Extraction. CoNLL-2005.

Uneven Margins



Support Vector Machines

- Like Perceptron, try to find a hyperplane that separates data
- But the goal here is to **maximize** the separation between the two classes
- Wider margin = greater generalisation



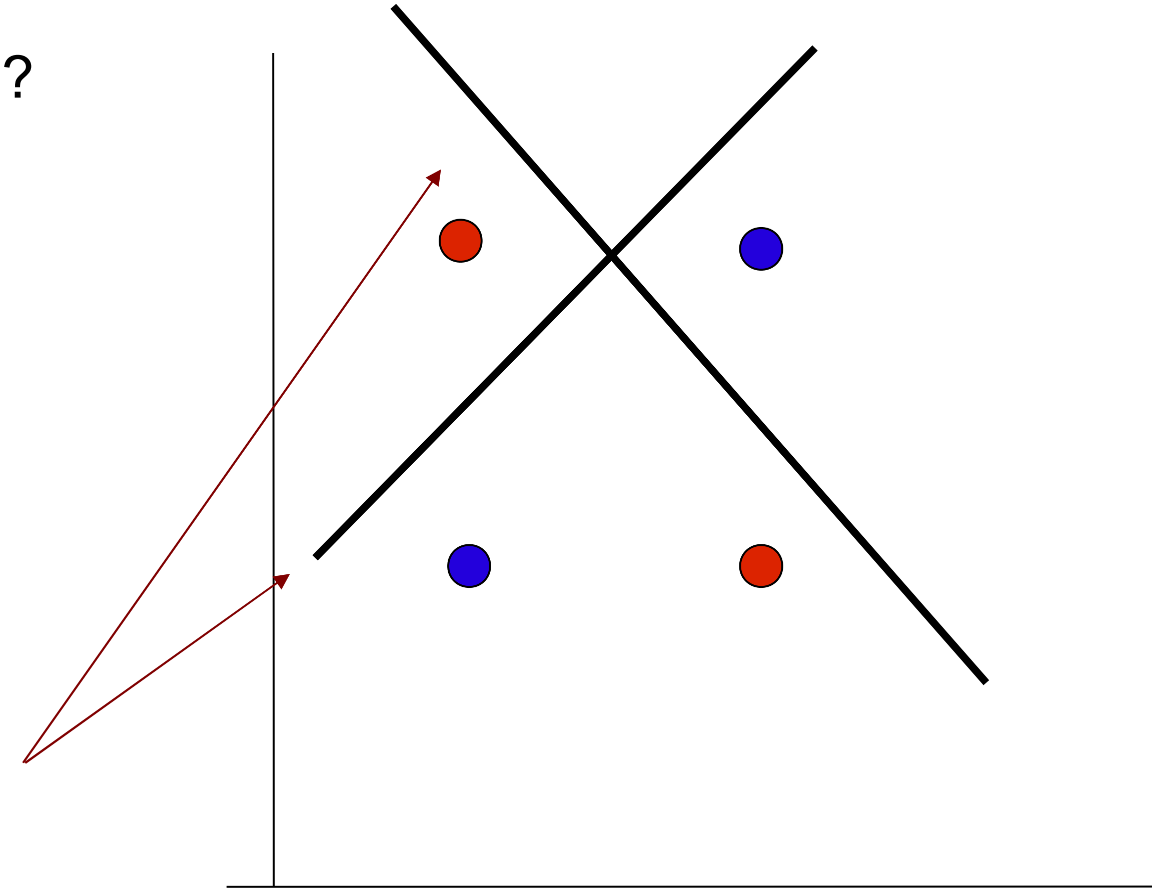
Support Vector Machines

- The points near the decision boundary are the “support vectors” (removing them would change boundary)
- The farther points are not important for decision-making
- What if you can't split the data neatly?
    - Soft boundary methods exist for imperfect solutions
    - However linear separator may be completely unsuitable

Support Vector Machines

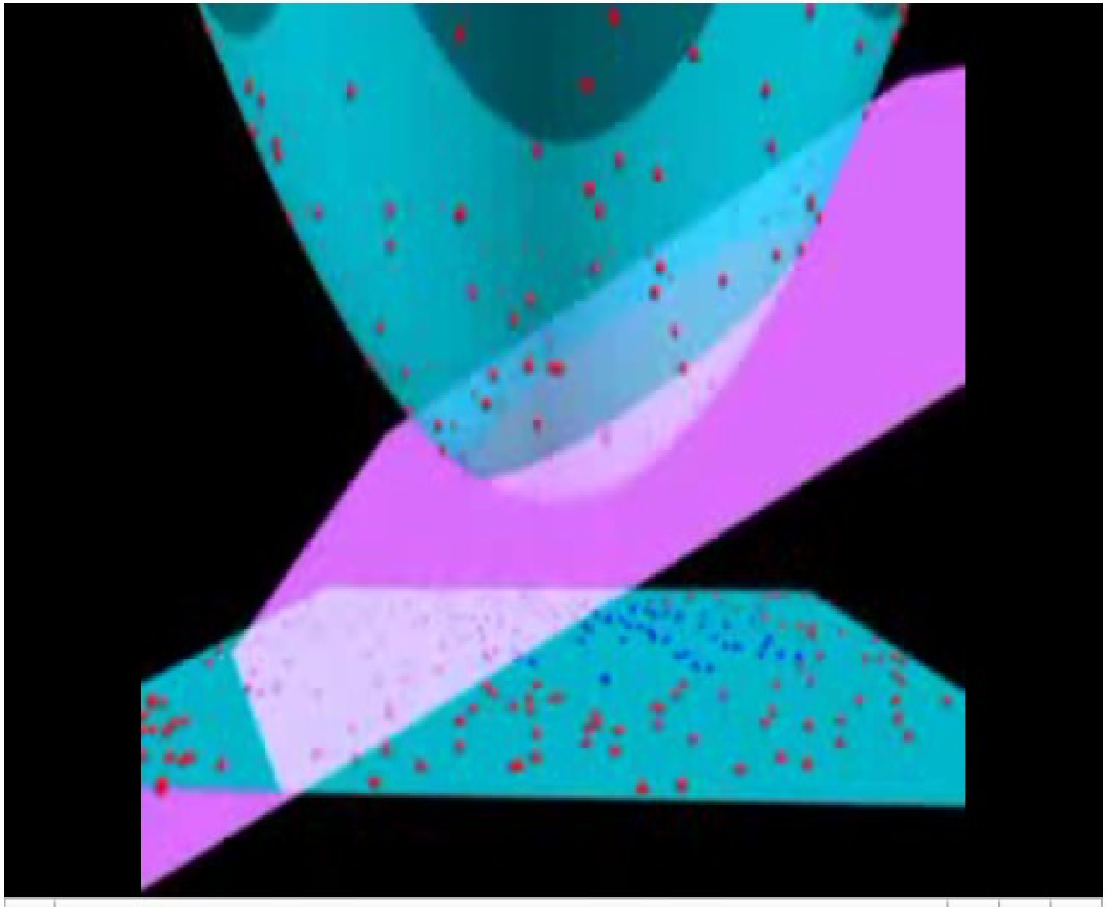
- What if there is no separating hyperplane?

They do not work!



Kernel Trick

- Map data into different dimensionality
- http://www.youtube.com
- As shown in the video, due to polynomial kernel elliptical separators can be created nevertheless.
- Now the points are separable!



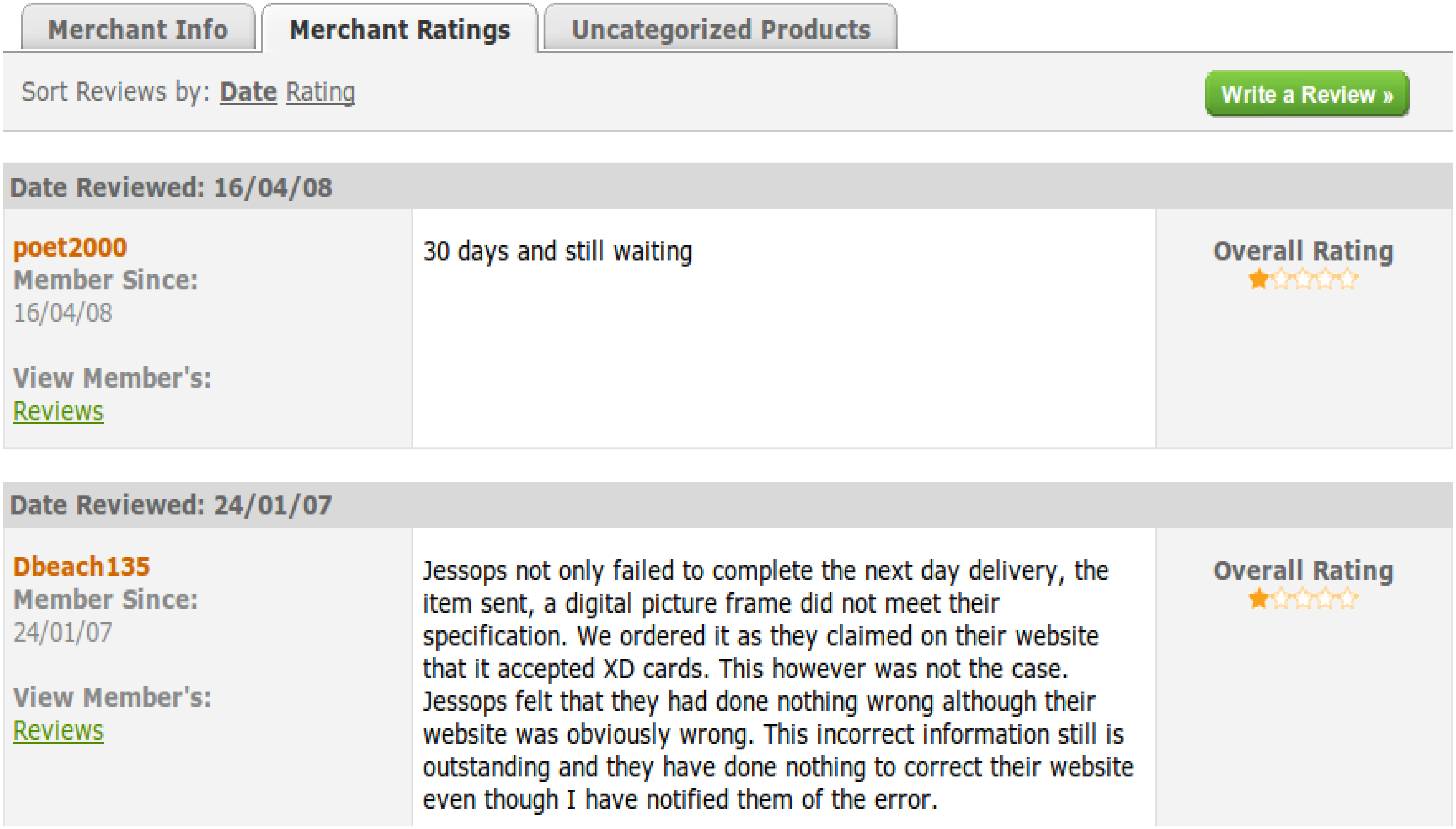
Kernel Trick in GATE and NLP

- Binomial kernel allows curved and elliptical separators to be created
- These are commonly used in language processing and are found to be successful
- In GATE, linear and polynomial kernels are implemented in Batch Learning PR's SVM engine

Machine Learning for Sentiment Analysis

- ML is an effective way to classify opinionated texts
- We want to train a classifier to categorize free text according to the training data.
- Good examples are consumers' reviews of films, products, and suppliers.
- Sites like www.pricegrabber.co.uk show reviews and an overall rating for companies: these make good training and testing data
- We train the ML system on a set of reviews so it can learn good and bad reviews, and then test it on a new set of reviews to see how well it distinguishes between them

Examples of consumer reviews



Date Reviewed: 16/04/08

poet2000
Member Since: 16/04/08
View Member's: Reviews

30 days and still waiting

Overall Rating ★☆☆☆☆

Date Reviewed: 24/01/07

Dbeach135
Member Since: 24/01/07
View Member's: Reviews

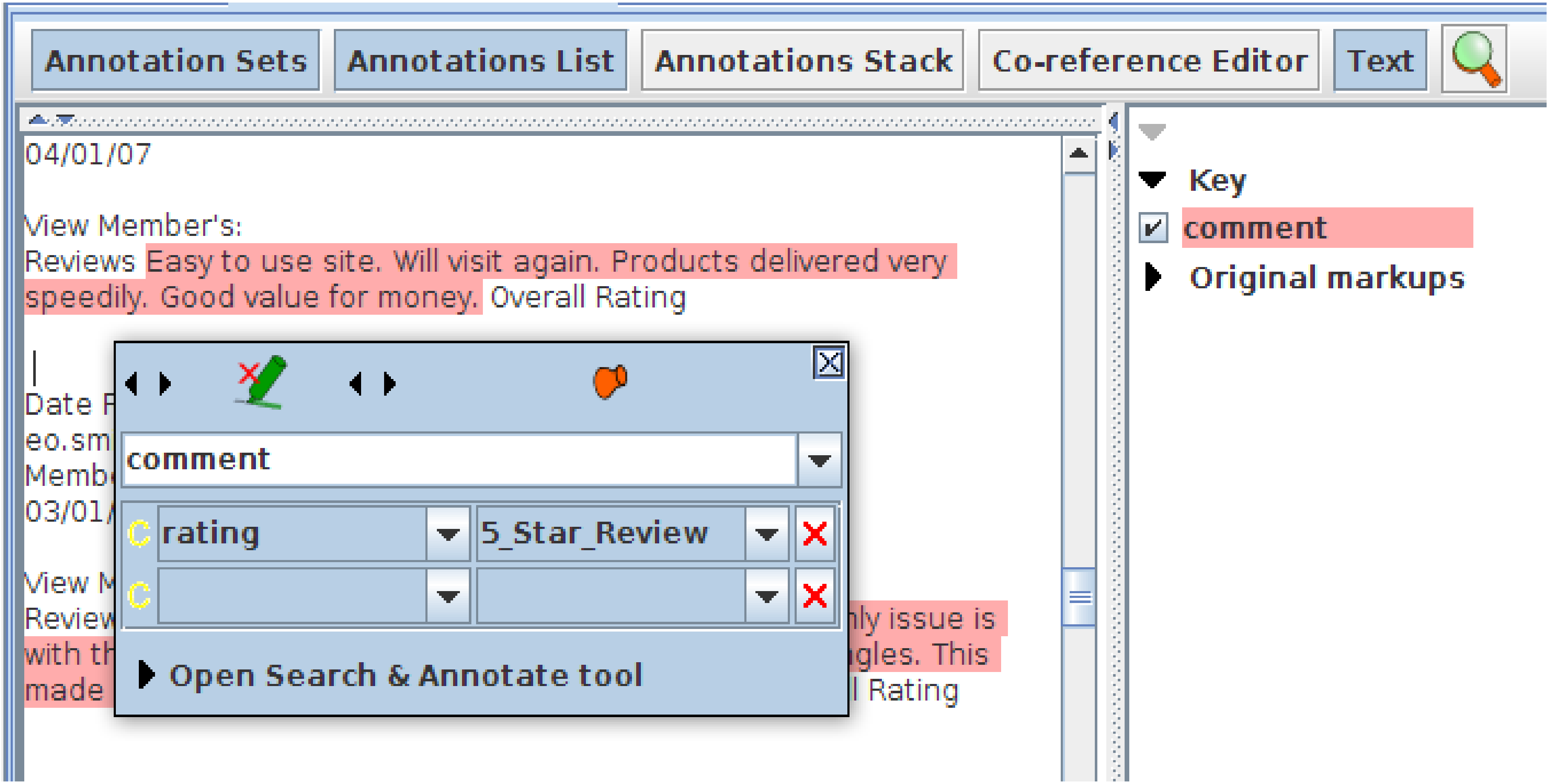
Jessops not only failed to complete the next day delivery, the item sent, a digital picture frame did not meet their specification. We ordered it as they claimed on their website that it accepted XD cards. This however was not the case. Jessops felt that they had done nothing wrong although their website was obviously wrong. This incorrect information still is outstanding and they have done nothing to correct their website even though I have notified them of the error.

Overall Rating ★☆☆☆☆

Preparing the corpus

- Corpus of 40 documents containing 552 company reviews.
- Each review has a 1- to 5-star rating.
- We pre-processed these in GATE to label each review with a comment annotation with a rating feature (free manual annotation!)
- In ML terms:
  - instance = *comment* annotation
  - class = *rating* feature on the *comment* annotation
  - attributes = NLP features of the underlying text
- We will keep the spans of the comment annotations and use ML to classify them with the *rating* feature
- We develop an application that runs a set of NLP components to provide ML instance attributes, and train the classifier

Annotated review



04/01/07

View Member's:
Reviews Easy to use site. Will visit again. Products delivered very speedily. Good value for money. Overall Rating

comment

rating: 5_Star_Review

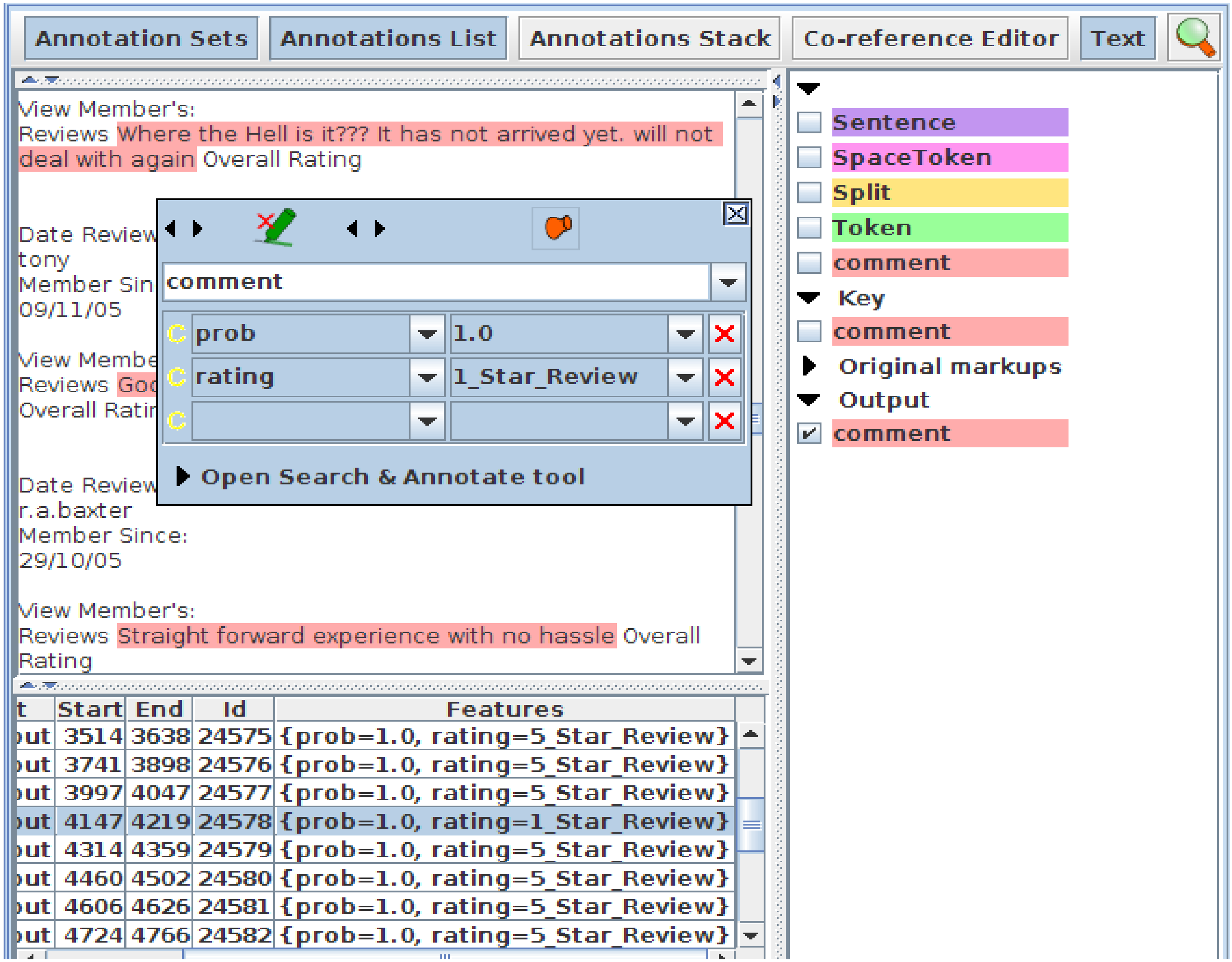
Musing ML configuration

- For this application, we used SVM (we would probably use PAUM now)
- Attributes: bag of lemmatised words (unigrams of lemmata) inside each comment annotation

Applying the training model

- To apply the classifier to our test corpus, we need to have comment annotations *without* rating features on the default AS
- These will give us the instances to classify
- A simple JAPE Transducer can do this
- When the pipeline is run, the classifier will get instances (*comment* annotations) and attributes from the default AS and put instances with classes (*rating* features) in the Output AS
  - Key set = user ratings
  - default set = instances with no classes
  - Output set = instances with ML classes

Annotation Results



View Member's:
Reviews Where the Hell is it??? It has not arrived yet. will not deal with again Overall Rating

Date Review
tony
Member Sin
09/11/05

View Membe
Reviews Goo
Overall Ratin

Date Review
r.a.baxter
Member Since:
29/10/05

View Member's:
Reviews Straight forward experience with no hassle Overall Rating

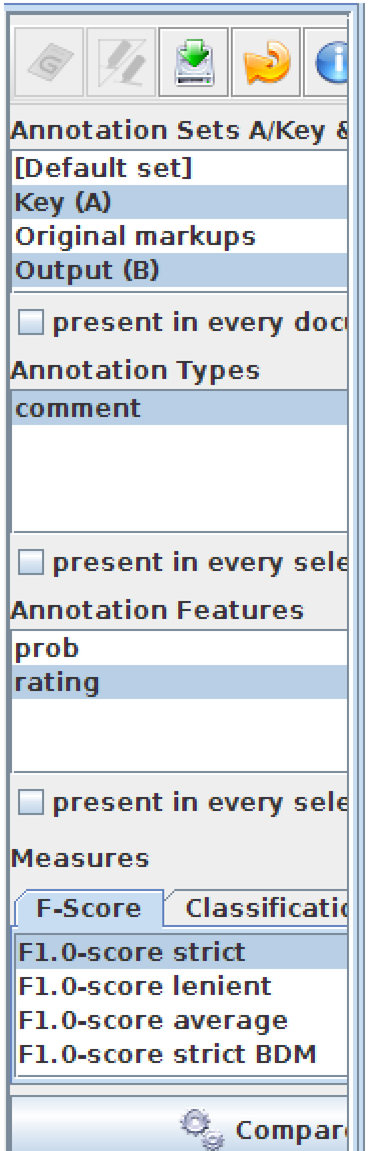
comment

prob	1.0
rating	1_Star_Review

- □ Sentence
- □ SpaceToken
- □ Split
- □ Token
- □ comment
- Key
  - □ comment
- Original markups
- Output
  - ☑ comment

t	Start	End	Id	Features
ut	3514	3638	24575	{prob=1.0, rating=5_Star_Review}
ut	3741	3898	24576	{prob=1.0, rating=5_Star_Review}
ut	3997	4047	24577	{prob=1.0, rating=5_Star_Review}
ut	4147	4219	24578	{prob=1.0, rating=1_Star_Review}
ut	4314	4359	24579	{prob=1.0, rating=5_Star_Review}
ut	4460	4502	24580	{prob=1.0, rating=5_Star_Review}
ut	4606	4626	24581	{prob=1.0, rating=5_Star_Review}
ut	4724	4766	24582	{prob=1.0, rating=5_Star_Review}

Evaluation: Corpus QA tool in GATE



Annotation Sets A/Key &

- [Default set]
- Key (A)
- Original markups
- Output (B)

□ present in every doc

Annotation Types

- comment

□ present in every sele

Annotation Features

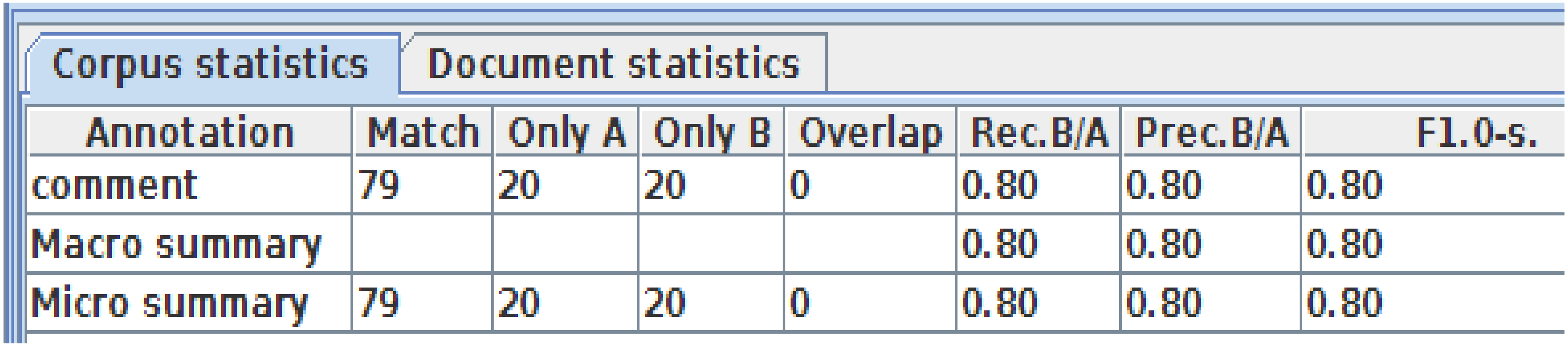
- prob
- rating

□ present in every sele

Measures

- F1.0-score strict
- F1.0-score lenient
- F1.0-score average
- F1.0-score strict BDM

Results



Annotation	Match	Only A	Only B	Overlap	Rec.B/A	Prec.B/A	F1.0-s.
comment	79	20	20	0	0.80	0.80	0.80
Macro summary					0.80	0.80	0.80
Micro summary	79	20	20	0	0.80	0.80	0.80

Cohen's Kappa and confusion matrices

- We can also use the Cohen's Kappa measure to show a confusion matrix
- The confusion matrix shows how many from each manually annotated class were automatically classified in each of the classes

	1	2	3	4	5
1	4	5	2	0	0
2	4	4	2	1	1
3	2	4	2	2	4
4	1	1	2	2	4
5	0	0	1	2	5

Cross-Validation

- Cross-validation is a standard way to “stretch” the validity of a manually annotated corpus, because it enables you to test on a larger number of documents
- Divide the corpus into 5 sub-corpora; train on ABCD and test on E; train on ABCE and test on D; etc.; average the results
- The 5-fold averaged result is more meaningful than the result obtained by training on 80% of the corpus and testing on the other 20% once.
- In GATE, you can't use the Corpus QA tool on the result, but you can get a detailed statistical report at the end, including P, R, & F1 for each class

Rule-based techniques

Rule-based techniques

- These rely primarily on sentiment dictionaries, plus some rules to do things like attach sentiments to targets, or modify the sentiment scores
- Examples include:
  - analysis of political tweets (Maynard and Funk, 2011)
  - analysis of opinions expressed about political events and rock festivals in social media (Maynard, Bontcheva and Rout, 2012)
  - SO-CAL (Taboada et al, 2011) for detecting positive and negative sentiment of ePinions reviews on the web.

Case study: Rule-based Opinion Mining from Political Tweets in GATE

Processing political tweets

- Application to associate people with their political leanings, based on pre-election tweets
  - e.g. “Had the pleasure of formally proposing Stuart King as Labour candidate for Putney”
- First stage is to find triple <Person, Opinion, Political Party>
  - e.g. John Smith is pro_Labour
- Usually, we will only get a single sentiment per tweet
- Later, we can collect all mentions of “John Smith” that refer to the same person, and collate the information
- John may be equally in favour of several different parties, not just Labour, but hates the Conservatives above all else

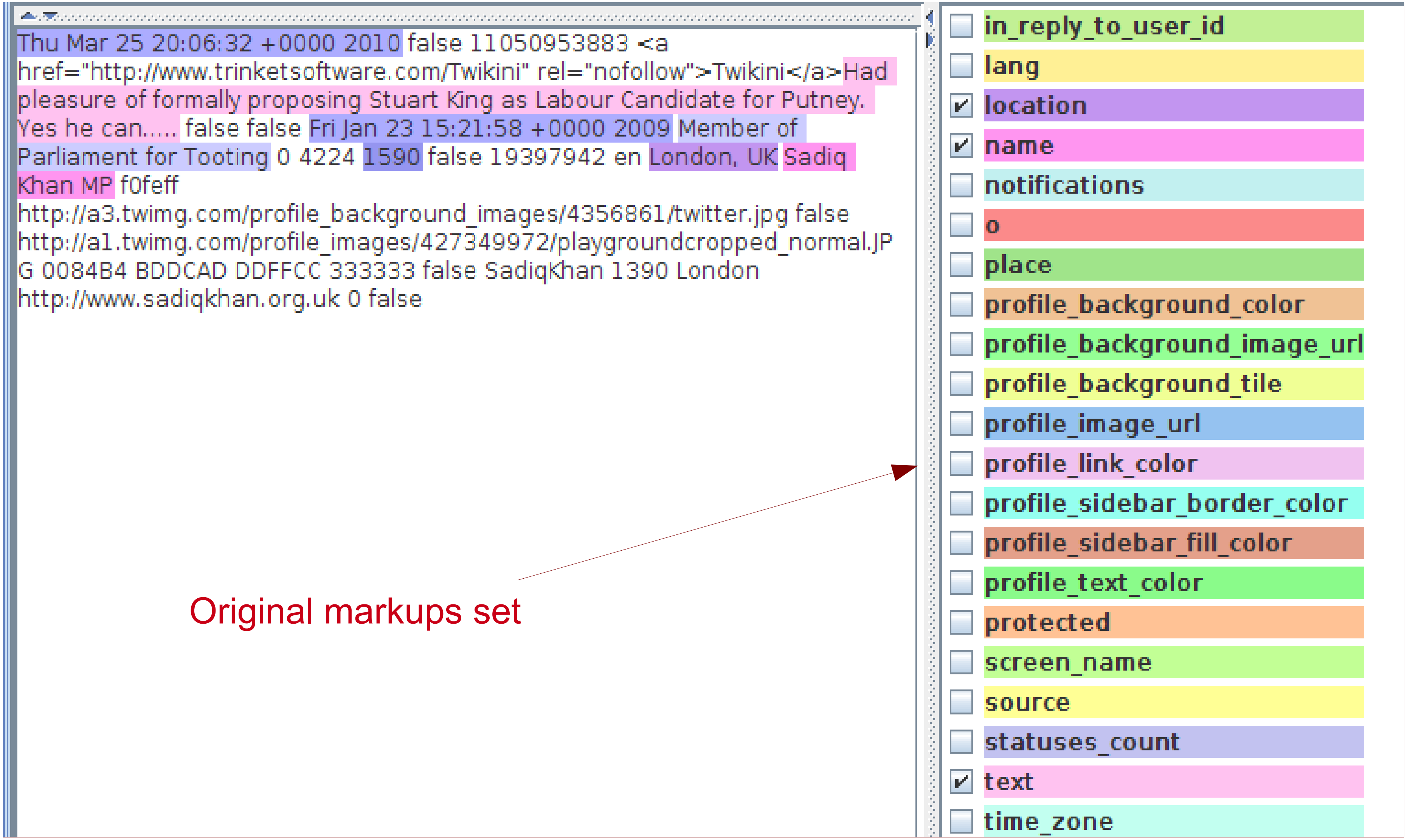
Creating a corpus

- First step is to create a corpus of tweets
- Used the Twitter Streaming API to suck up all the tweets over the pre-election period according to various criteria (e.g. use of certain hash tags, mention of various political parties etc.)
- Collected tweets in json format and then converted these to xml using JSON-:ib library
- This gives us lots of additional twitter metadata, such as the date and time of the tweet, the number of followers of the person tweeting, the location and other information about the person tweeting, and so on
- This information is useful for disambiguation and for collating the information later

Corpus Size

- Raw corpus contained around 5 million tweets
- Many were duplicates due to the way in which the tweets were collected
- Added a de-duplication step during the conversion of json to xml
- This reduced corpus size by 20% to around 4 million
- This still retains the retweets, however

Tweets with metadata

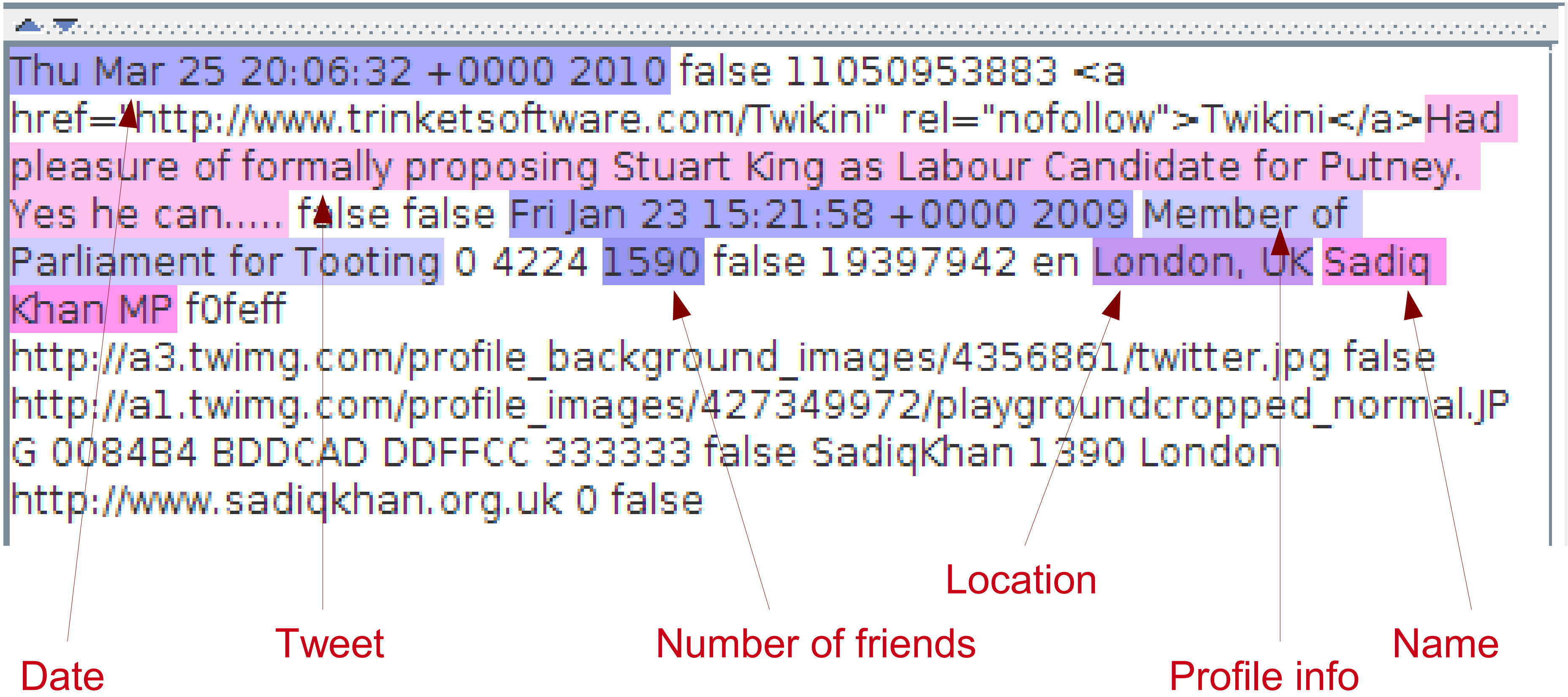


Thu Mar 25 20:06:32 +0000 2010 false 11050953883 <a href="http://www.trinketsoftware.com/Twikini" rel="nofollow">Twikini</a>Had pleasure of formally proposing Stuart King as Labour Candidate for Putney. Yes he can..... false false Fri Jan 23 15:21:58 +0000 2009 Member of Parliament for Tooting 0 4224 1590 false 19397942 en London, UK Sadiq Khan MP f0feff http://a3.twimg.com/profile_background_images/4356861/twitter.jpg false http://a1.twimg.com/profile_images/427349972/playgroundcropped_normal.JPG 0084B4 BDDCAD DDFFCC 333333 false SadiqKhan 1390 London http://www.sadiqkhan.org.uk 0 false

- ☐ in_reply_to_user_id
- ☐ lang
- ☑ location
- ☑ name
- ☐ notifications
- ☐ o
- ☐ place
- ☐ profile_background_color
- ☐ profile_background_image_url
- ☐ profile_background_tile
- ☐ profile_image_url
- ☐ profile_link_color
- ☐ profile_sidebar_border_color
- ☐ profile_sidebar_fill_color
- ☐ profile_text_color
- ☐ protected
- ☐ screen_name
- ☐ source
- ☐ statuses_count
- ☑ text
- ☐ time_zone

Original markups set

Metadata

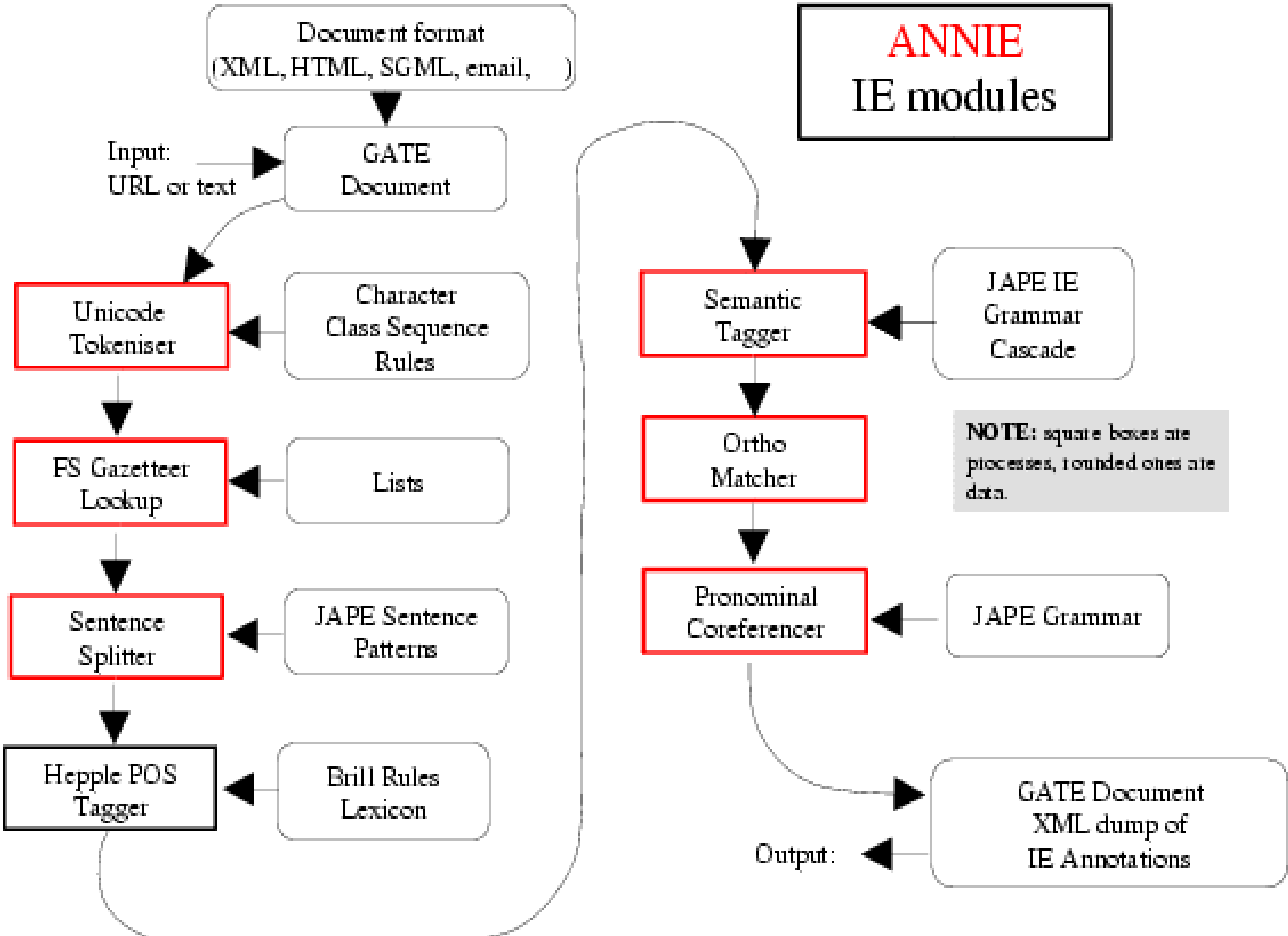


```
Thu Mar 25 20:06:32 +0000 2010 false 11050953883 <a
href="http://www.trinketsoftware.com/Twikini" rel="nofollow">Twikini</a>Had
pleasure of formally proposing Stuart King as Labour Candidate for Putney.
Yes he can..... false false Fri Jan 23 15:21:58 +0000 2009 Member of
Parliament for Tooting 0 4224 1590 false 19397942 en London, UK Sadiq
Khan MP f0feff
http://a3.twimg.com/profile_background_images/4356861/twitter.jpg false
http://a1.twimg.com/profile_images/427349972/playgroundcropped_normal.JP
G 0084B4 BDDCAD DDFFCC 333333 false SadiqKhan 1390 London
http://www.sadiqkhan.org.uk 0 false
```

Linguistic pre-processing

- Use standard set of pre-processing resources in GATE to identify tokens, sentences, POS tags etc., and also to perform NE recognition.
- Slightly adapted the standard ANNIE application (GATE's default IE application)

ANNIE
IE modules



Document format (XML, HTML, SGML, email,)

Input: URL or text

GATE Document

Unicode Tokeniser

Character Class Sequence Rules

FS Gazetteer Lookup

Lists

Sentence Splitter

JAPE Sentence Patterns

Hepple POS Tagger

Brill Rules Lexicon

Semantic Tagger

JAPE IE Grammar Cascade

Ortho Matcher

Pronominal Coreferencer

JAPE Grammar

GATE Document XML dump of IE Annotations

Output:

NOTE: square boxes are processes, rounded ones are data.

Gazetteers

- We create a flexible gazetteer to match certain useful keywords, in various morphological forms:
  - political parties, e.g. “Conservative”, “LibDem”
  - concepts about winning election, e.g. “win”, “landslide”
  - words for politicians, e.g. “candidate”, “MP”
  - words for voting and supporting a party/ person, e.g. “vote”
  - words indicating negation, e.g. “not”, “never”
- We create another gazetteer containing affect/emotion words from WordNet.
  - these have a feature denoting part of speech (category)
  - Keeping category information may be important, so we don't want a flexible gazetteer here

A negative sentiment list

Examples of phrases following the word “go”:

- down the pan
- down the drain
- to the dogs
- downhill
- pear-shaped

A positive sentiment list

- awesome category=adjective score=0.5
- beaming category=adjective score=0.5
- becharm category=verb score=0.5
- belonging category=noun score=0.5
- benefic category=adjective score=0.5
- benevolently category=adverb score=0.5
- caring category=noun score=0.5
- charitable category=adjective score=0.5
- charm category=verb score=0.5

Grammar rules: creating preliminary annotations

- Identify questions or doubtful statements as opposed to "factual" statements in tweets, e.g. look for question marks

Wont Unite's victory be beneficial to Labour?

- Create temporary Sentiment annotations if a Sentiment Lookup is found and if the category matches the POS tag on the Token (this ensures disambiguation of the different possible categories)

“Just watched video about awful days of Tory rule” vs “Ah good, the entertainment is here.”

“People like her should be shot.” vs “People like her.”

Question grammar

```
Phase:Preprocess
Input: Token
Options: control = appelt

Rule: Question
(
 {Token.string == "?"}
):tag
-->
:tag.Question = {rule = "Question"}
```

```
Phase:  Affect
Input: AffectLookup Token
Options: control = appelt

Rule: AffectAdjective
(
 {AffectLookup.category == adjective,Token.category == VBN}|
 {AffectLookup.category == adjective, Token.category == JJ}
):tag
-->
:tag.Affect = {kind = :tag.AffectLookup.kind,
      category = :tag.AffectLookup.category,
      rule = "AffectAdjective"}
```

Check category of both Lookup and Token are adjectives or past participles

copy category and kind values from Lookup to new Affect annotation

Grammar rules: finding triples

- We first create temporary annotations for Person, Organization, Vote, Party, Negatives etc. based on gazetteer lookup, NEs etc.
- We then have a set of rules to combine these into pairs or triples:
  - *<Person, Vote, Party> "Tory Phip admits he voted LibDem".*
  - *<Party, Affect> "When they get a Tory government they'll be sorry."*
- We create an annotation "Sentiment" which has the following features:
  - kind = "pro_Labour", "anti_LibDem", etc.
  - opinion_holder = "John Smith", "author" etc.

Identifying the Opinion Holder

- If the opinion holder in the pattern matched is a Person or Organization, we just get the string as the value of opinion_holder
- If the opinion holder in the pattern matched is a pronoun, we first find the value of the string of the antecedent and use this as the value of opinion_holder
- Currently we only match opinion holders within the same sentence.
- If no explicit opinion holder then we use "author" as the value of opinion_holder.
- Later we can grab the details of the twitterer instead of just using "author".

Grammar rules: finding antecedents

- Find the antecedents of pronouns within a sentence so that we can refer a sentiment back to the original opinion holder or object of the opinion.
- First run the pronominal coreference PR
- Then use a JAPE rule to find pronouns linked to a Person or Organization
- We can identify these because they will have the feature “ENTITY_MENTION_TYPE” (created by the coreferencer)
- The co-referring pronouns all have also an antecedent_offset feature pointing to the proper noun antecedent
- The matching proper noun antecedent is found and its string is added as a feature on the relevant pronoun annotation

Implicit Opinion Holders

- There may not always be an explicit opinion holder
- In many cases, the author of the tweet is the opinion holder

I'm also going to vote Tory. Hello new world.

  - Here we can co-refer "I" with the person tweeting (using the metadata)

- In other cases, there is no explicit opinion holder:

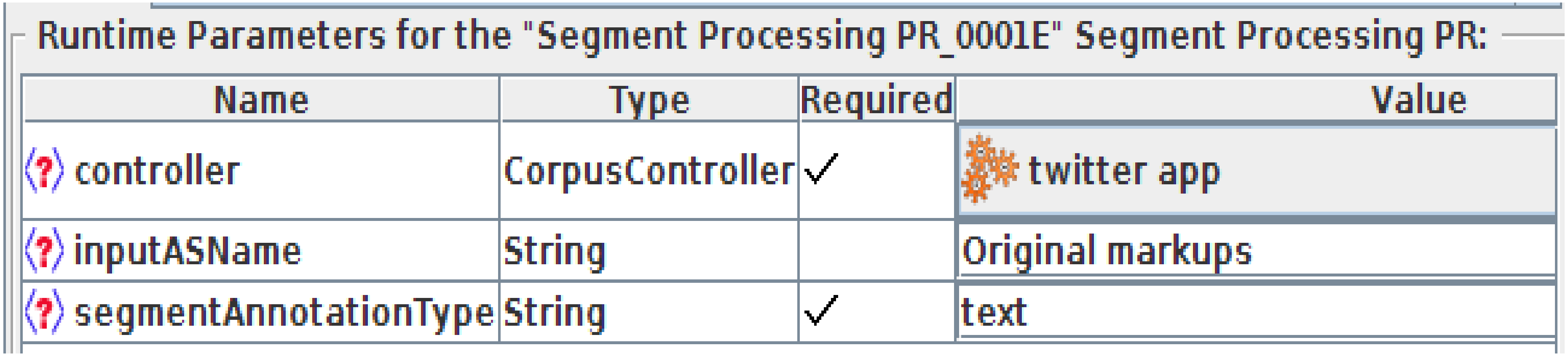
Vote for Labour. Harry Potter would.

  - However, we can infer by this instruction that the author of the tweet shares this opinion.

- In all these cases, we add the value "author" to the feature "opinion_holder"

Creating the Application

- We only want to process the actual text of the tweet, not all the other information
- To do this, we use a Segment Processing PR to run the sentiment app over just the "text" annotation in Original Markups set.
- So, we need two applications: one containing the Segment Processing PR and one containing the actual sentiment application



Runtime Parameters for the "Segment Processing PR_0001E" Segment Processing PR:

Name	Type	Required	Value
controller	CorpusController	✓	twitter app
inputASName	String		Original markups
segmentAnnotationType	String	✓	text

Corpus analysis tools

- Corpus analysis tools enable you to look at the results of processing and make sense of them manually
- In GATE, we have a tool called ANNIC which lets you analyse annotations in context.
- Like a KWIC index but works over annotations as well as just strings
- Enables you to search and analyse a whole corpus without knowing a priori what appears specifically in which document
- This is especially useful in a corpus of tweets where each document represents a single tweet

Pattern examples

- {Party}
- {Affect}
- {Lookup.majorType == negation} ({Token})*4 {Lookup.majorType == "vote"}{Lookup.majorType == "party"}
- {Token.string == "I"} ({Token})*4 {Lookup.majorType == "vote"} {Lookup.majorType == "party"}
- {Person} ({Token})*4 {Lookup.majorType == "vote"} {Lookup.majorType == "party"}
- {Affect} ({Token})*5 {Lookup.majorType == "candidate"}
- {Vote} ({Token})*5 {Lookup.majorType == "candidate"}

Opinion Finding in Arcomem

Arcomem project

- Arcomem is an EU project about storing community memories.
- Involves detection of entities, events, topics and opinions to guide the crawler
- Aims to answer questions such as:
  - What are the opinions on crucial social events and the key people involved?
  - How are these opinions distributed in relation to demographic user data?
  - How have these opinions evolved?
  - Who are the opinion leaders?
  - What is their impact and influence?

Arcomem Applications

- Developed a series of initial applications for opinion mining from social media using GATE
- Based on previous work identifying political opinions from tweets
- Extended to more generic analysis about any kind of entity or event, in 2 domains
  - Greek financial crisis
  - Rock am Ring (German rock festival)
- Uses a variety of social media including twitter, facebook and forum posts
- Based on entity and event extraction, and a rule-based approach

GATE Application

- Structural pre-processing, specific to social media types (such as separating the actual content of the tweet from the metadata)
- Linguistic pre-processing (including language detection), NE, term and event recognition
- Additional targeted gazetteer lookup
- JAPE grammars
- Aggregation of opinions
- Dynamics

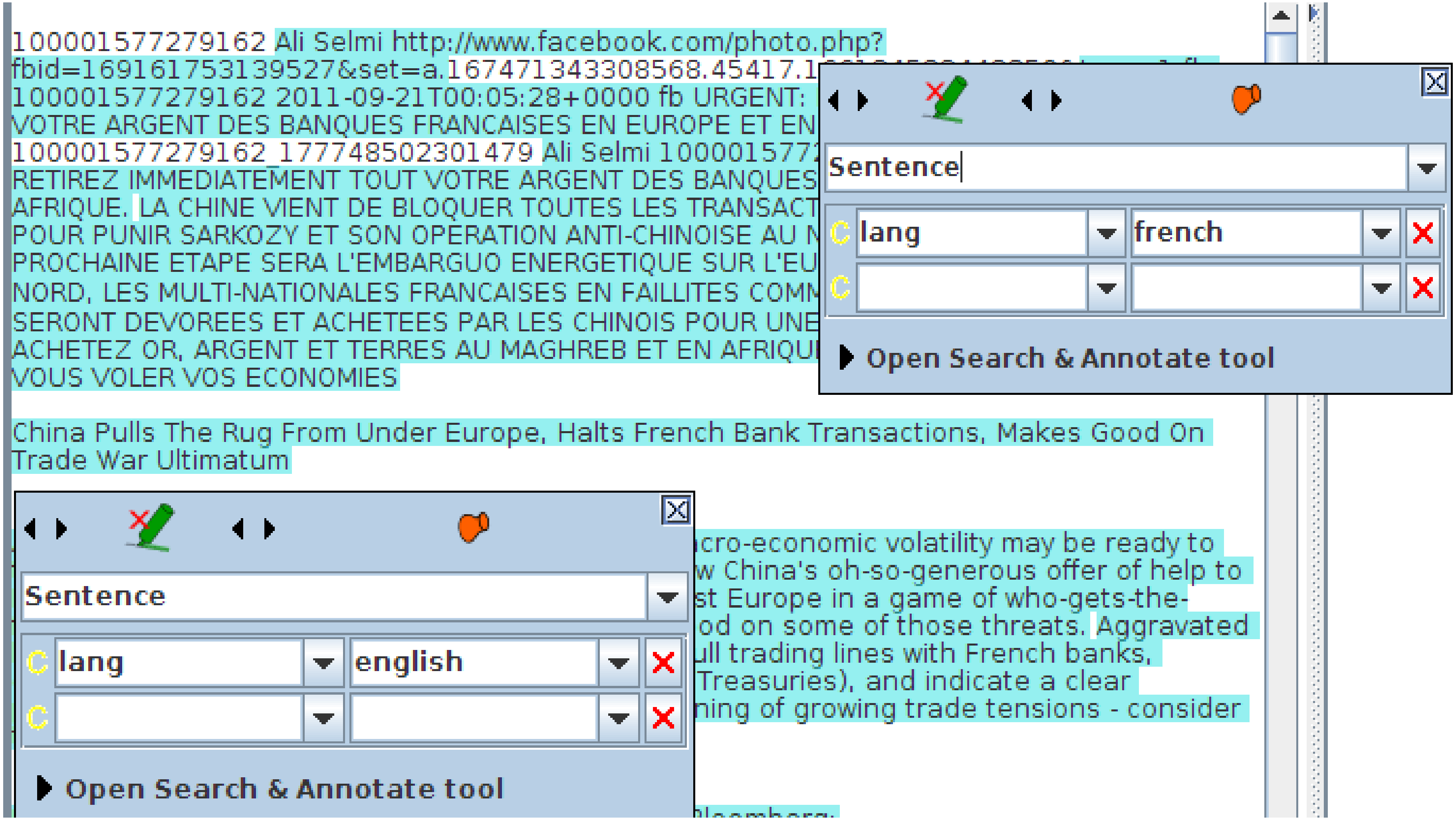
Why Rule-based?

- Although ML applications are typically used for Opinion Mining, this task involves documents from many different text types, genres, languages and domains
- This is problematic for ML because it requires many applications trained on the different datasets, and methods to deal with acquisition of training material
- Aim of using a rule-based system is that the bulk of it can be used across different kinds of texts, with only the pre-processing and some sentiment dictionaries which are domain and language-specific

Linguistic pre-processing

- Language identification (per sentence) using TextCat
- Standard tokenisation, POS tagging etc using GATE
- NE and Term recognition using modified versions of ANNIE and TermRaider
- Event recognition using specially developed GATE application (e.g. band performance, economic crisis, industrial strike)

Language ID with TextCat



100001577279162 Ali Selmi http://www.facebook.com/photo.php?
fbid=169161753139527&set=a.167471343308568.45417.1
100001577279162 2011-09-21T00:05:28+0000 fb URGENT:
VOTRE ARGENT DES BANQUES FRANCAISES EN EUROPE ET EN
100001577279162_177748502301479 Ali Selmi 10000157772
RETIREZ IMMEDIATEMENT TOUT VOTRE ARGENT DES BANQUES
AFRIQUE. LA CHINE VIENT DE BLOQUER TOUTES LES TRANSACT
POUR PUNIR SARKOZY ET SON OPERATION ANTI-CHINOISE AU N
PROCHAINE ETAPE SERA L'EMBARGUO ENERGETIQUE SUR L'EU
NORD, LES MULTI-NATIONALES FRANCAISES EN FAILLITES COMM
SERONT DEVOREES ET ACHETEES PAR LES CHINOIS POUR UNE
ACHETEZ OR, ARGENT ET TERRES AU MAGHREB ET EN AFRIQU
VOUS VOLER VOS ECONOMIES

Sentence

lang french

China Pulls The Rug From Under Europe, Halts French Bank Transactions, Makes Good On Trade War Ultimatum

icro-economic volatility may be ready to
w China's oh-so-generous offer of help to
st Europe in a game of who-gets-the-
od on some of those threats. Aggravated
ull trading lines with French banks,
Treasuries), and indicate a clear
ning of growing trade tensions - consider

Sentence

lang english

Basic approach for opinion finding

- Find sentiment-containing words in a linguistic relation with entities/events (opinion-target matching)
- Use a number of linguistic sub-components to deal with issues such as negatives, irony, swear words etc.
- Starting from basic sentiment lookup, we then adjust the scores and polarity of the opinions via these components

Sentiment finding components

- **Flexible Gazetteer Lookup**: matches lists of affect/emotion words against the text, in any morphological variant
- **Gazetteer Lookup:** matches lists of affect/emotion words against the text only in non-variant forms, i.e. exact string match (mainly the case for specific phrases, swear words, emoticons etc.)
- **Sentiment Grammars**: set of hand-crafted JAPE rules which annotate sentiments and link them with the relevant targets and opinion holders
- **RDF Generation**: create the relevant RDF-XML for the annotations according to the data model (so they can be used by other components)

Opinion scoring

- Sentiment gazetteers (developed from sentiment words in WordNet) have a starting “strength” score
- These get modified by context words, e.g. adverbs, swear words, negatives and so on
  - *The film was awesome* --> *The film was **** amazing*.
  - *The film was awful* --> *The film was **** awful..*
- Swear words on their own are classified as negative, however.
  - *Damed politicians and their lies*.
  - *RIP Fergie? It's SIR Alex Ferguson to you, Carlos, you runt.*

Evaluation

- Very hard to measure opinion polarity beyond positive / negative / neutral unless you have a product review corpus
- On a small corpus of 20 facebook posts, we identified sentiment-containing sentences with 55% Precision and 60% Recall. Of these, the polarity accuracy was 82%.
- Much better results for tweets, however.
- While this is not that high, not all the subcomponents are complete in the system, so we would expect better results with improved methods for negation and sarcasm detection
- NE recognition was high on these texts: 92% Precision and 69% Recall (compared with other NE evaluations on social media)

Comparison of Opinion Finding in Different Tasks

Corpus	Sentiment detection	Polarity detection	Target assignment
Political Tweets	78%	79%	97.9%
Financial Crisis Facebook	55%	81.8%	32.7%
Financial Crisis Tweets	90%	93.8%	66.7%

Using Machine Learning for the Arcomem task

- If we can collect enough manually annotated training data, we can also use an ML approach for this task
- Similar to that presented earlier for MUSING, but modified to take into account what we have subsequently learned and the differences in the data.
- Each MUSING product review had an opinion from 1 to 5 stars
- In Arcomem we classify sentences (the ML *instances)*, many of which do not contain opinions
- So the ML *classes* will be *positive*, *neutral*, *negative*, and *none* (contains no opinion, different from a *neutral* opinion)

Using Machine Learning for the Arcomem task

- We now know that PAUM is much faster than SVM but typically just as good for NLP tasks, so we will use PAUM instead
- We'll need to deal with the special issues of social media text (more on this later)
- For the ML *attributes*, we will use n-grams of tokens or lemmata
  - In MUSING, n-grams with n>2 did not improve accuracy but slowed the ML down
  - But it's worth trying 3-grams just in case they help with the smaller instances

Using Machine Learning for the Arcomem task

- Also worth trying other annotations such as named entities
- But these might exaggerate the effect of biased training data (this might not be a problem, but it's worth bearing in mind)
- For example, if most people who mention “Venus Williams” in the training data like her (or her dresses), we are training the ML model to expect positive opinions for that Person annotation; the real data might or might not match

Training on tweets

- You can use hashtags as a source of classes!
  - Example: collect a set of tweets with the **#angry** tag, and a set without it, and delete from the second set any tweets that look angry
  - Remove the **#angry** tag from the text in the first set (so you're not just training the ML to spot the tag)
  - You now have a corpus of manually annotated angry/non-angry data!
- This approach can work well, but if you have huge datasets, you may not be able to do the manual deletions
- Experimenting with **#sarcasm** is interesting (more on this later)

Challenges for opinion mining on social media

Linguistic issues

- What kinds of linguistic problems do we need to overcome?
  - Short sentences (problems for parsers etc)
  - Use of incorrect English
  - Negatives
  - Conditional statements
  - Use of slang/swear words
  - Use of irony/sarcasm
  - Ambiguity

Short sentences, e.g. tweets

- Social media, and especially tweets, can be problematic because sentences are very short and/or incomplete
- Typically, linguistic pre-processing tools such as POS taggers and parsers do badly on such texts
- Even basic tools like language identification can have problems
- The best solution is to try not to rely too heavily on these tools
  - Does it matter if we get the wrong language for a sentence?
  - Do we actually need full parsing?
  - Can we use other clues when POS tags may be incorrect?

Dealing with incorrect English

- Frequent problem in any NLP task involving social media
- Incorrect capitalisation, spelling, grammar, made-up words (eg swear words, infixes)
- Backoff strategies include
  - normalisation
  - using more flexible gazetteer matching
  - using case-insensitive resources (but be careful)
  - avoiding full parsing and using shallow techniques
  - using very general grammar rules
  - adding specialised gazetteer entries for common mis-spellings, or using co-reference techniques

Tokenisation

- Splitting a text into its constituent parts
- Plenty of “unusual”, but very important tokens in social media:
  - @Apple – mentions of company/brand/person names
  - #fail, #SteveJobs – hashtags expressing sentiment, person or company names
  - :-(, :-), :-P – emoticons (punctuation and optionally letters)
  - URLs
- Tokenisation key for entity recognition and opinion mining
- A study of 1.1 million tweets: 26% of English tweets have a URL, 16.6% - a hashtag, and 54.8% - a user name mention [Carter, 2013].

Example

#WiredBizCon #nike vp said when @Apple saw what http://nikeplus.com did, #SteveJobs was like wow I didn't expect this at all.

- Tokenising on white space doesn't work that well: Nike and Apple are company names, but if we have tokens such as #nike and @Apple, this will make the entity recognition harder, as it will need to look at sub-token level
- Tokenising on white space and punctuation characters doesn't work well either: URLs get separated (http, nikeplus), as are emoticons and email addresses

The GATE Twitter Tokeniser

- Treat RTs, emoticons, and URLs as 1 token each
- #nike is two tokens (# and nike) plus a separate annotation HashTag covering both. Same for @mentions
- Capitalisation is preserved, but an orthography feature is added: all caps, lowercase, mixCase
- Date and phone number normalisation, lowercasing, and other such cases are optionally done later in separate modules
- Consequently, tokenisation is faster and more generic

De-duplication and Spam Removal

- Approach from [Choudhury & Breslin, #MSM2011]:
- Remove as duplicates/spam:
  - Messages with only hashtags (and optional URL)
  - Similar content, different user names and with the same timestamp are considered to be a case of multiple accounts
  - Same account, identical content are considered to be duplicate tweets
  - Same account, same content at multiple times are considered as spam tweets

Language Detection

- There are many language detection systems readily available
- The main challenges on tweets/facebook status updates:
  - the short number of tokens (10 tokens/tweet on average)
  - the noisy nature of the words (abbreviations, misspellings).
- Due to the length of the text, we can make the assumption that one tweet is written in only one language
- Most language detection tools work by building n-gram language models for each language and then assigning the text to the most probable language from the trained model.

Language Detection for Social Media

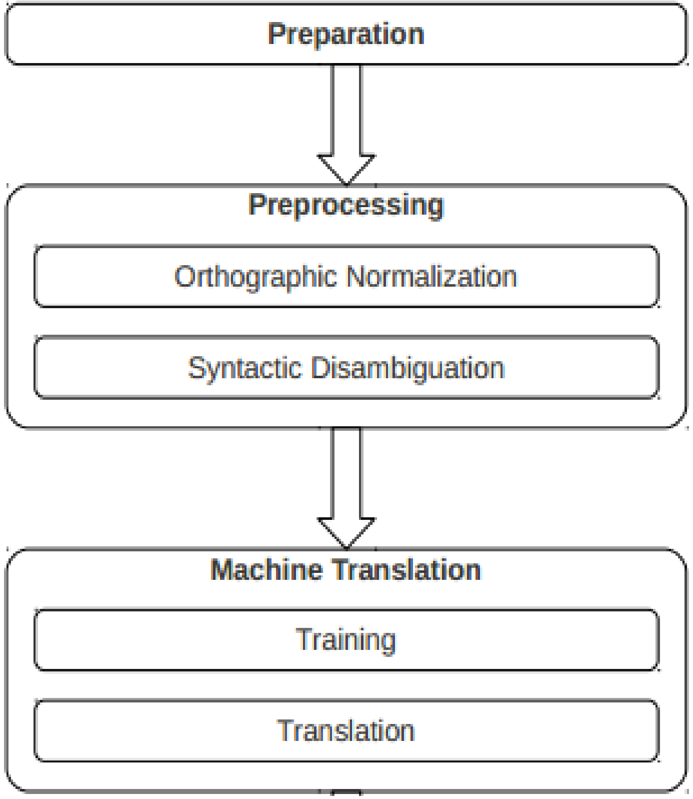
- Compare language detection methods [Lui and Baldwin, 2011]
- Best results with 1-nearest-neighbour (1NN) model
  - a test document is classified based on the language of the closest training document, as determined by the cosine similarity metric
  - Character bigrams or trigrams
- We have reimplemented their best method in Java, as part of TrendMiner
    - https://github.com/sinjax/trendminer-java/tree/master/text/nlp/src/main/java/org/openimaj/text/nlp
- Comes pre-trained on 97 languages and very fast

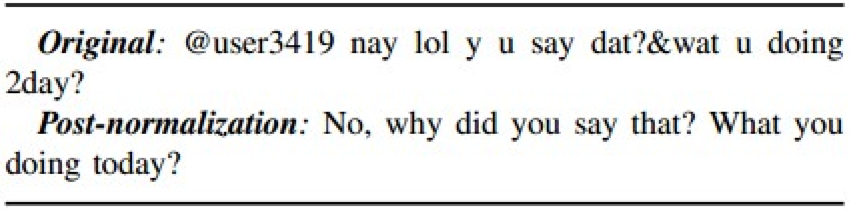
Normalisation

- “RT @Bthompson WRITEZ: @libbyabrego honored?! Everybody knows the libster is nice with it...lol...(thankkkks a bunch;))”
- OMG! I’m so guilty!!! Sprained biibii’s leg! ARGHHHHHH!!!!!!
- Similar to SMS normalisation
- For some later components to work well (POS tagger, parser), it is necessary to produce a normalised version of each token
- BUT uppercasing, and letter and exclamation mark repetition often convey strong sentiment
- Therefore some choose not to normalise, while others keep both versions of the tokens

Syntactic Normalisation [Kaufmann, 2010]

- Preparation: removing emoticons, tokenisation
- Orthographic mapping: 2moro, u
- Syntactic disambiguation
  - Determine when @mentions and #tags have syntactic value and should be kept in the sentence, vs replies, retweets and topic tagging
- Machine Translation: used MOSES
  - Trained on SMS and ANC corpora





Original: @user3419 nay lol y u say dat?&wat u doing 2day?

Post-normalization: No, why did you say that? What you doing today?

Stemming

- The Snowball stemmer is already integrated in GATE
- 11 European languages: Danish, Dutch, English, Finnish, French, German, Italian, Norwegian, Portuguese, Russian, Spanish and Swedish
- http://snowball.tartarus.org

NER in Tweets

- Performance of the Stanford NER drops to 48% [Liu et al, 2011] or even 29% on another tweet corpus [Ritter et al, 2011]
- Pre-processing used:
  - Stop words, user names, and links are removed
  - Specially adapted/trained POS tagger [Ritter et al, 2011]
  - NP Chunker adapted to tweets [Ritter et al, 2011]
  - Capitalisation information [Ritter et al, 2011]
  - Syntactic normalisation [Doerhmann, 2011]
  - Gazetteers derived from Freebase [Ritter et al, 2011]

NER for Tweets (2)

- Performance reported on 4 entity types (PER, LOC, ORG, PRODUCT): 80.2% f-score (81.6% P; 78.8% R) [Liu et al 2011]
- [Doerhmann, 2011] improved on Liu's results by normalising the tweets first
- Ritter's scores are lower but against more Freebase entity types: PERSON, GEO-LOCATION, COMPANY, PRODUCT, FACILITY, TV-SHOW, MOVIE, SPORTSTEAM, BAND, and OTHER

Other challenges of social media

- **Strongly temporal and dynamic**: temporal information (e.g. post timestamp) can be combined with opinion mining, to examine the volatility of attitudes towards topics over time (e.g. gay marriage).
- **Exploiting social context**: (Who is the user connected to? How frequently they interact). Derive automatically semantic models of social networks, measure user authority, cluster similar users into groups, as well as model trust and strength of connection
- **Implicit information about the user**: Research on recognising gender, location, and age of Twitter users. Helpful for generating opinion summaries by user demographics

More flexible matching techniques

- In GATE, as well as the standard gazetteers, we have options for modified versions which allow for more flexible matching
- BWP Gazetteer: uses Levenshein edit distance for approximate string matching
- Extended Gazetteer: has a number of parameters for matching prefixes, suffixes, initial capitalisation and so on

Extended Gazetteer

- Part of the StringAnnotation plugin in GATE
- Has the following additional characteristics:
  - Gives more control over which characters are considered to belong to words and non-word characters
  - Enables matching when an initial letter of a word is uppercase
  - matching of prefixes and suffixes
  - case-insensitive matching also deals with cases (such as German "ß" which maps to "SS")

Case-Insensitive matching

- This would seem the ideal solution, especially for gazetteer lookup, when people don't use case information as expected
- However, setting all PRs to be case-insensitive can have undesired consequences
  - POS tagging becomes unreliable (e.g. “May” vs “may”)
  - Back-off strategies may fail, e.g. unknown words beginning with a capital letter are normally assumed to be proper nouns
  - Gazetteer entries quickly become ambiguous (e.g. many place names and first names are ambiguous with common words)
- Solutions include selective use of case insensitivity, removal of ambiguous terms from lists, additional verification (e.g. use of coreference)

Finding Negatives

- What methods might we use for finding negatives?
  - List lookup
  - Verb analysis
  - Sarcasm

Find the hidden deer...

One of the trickiest tasks in opinion mining is spotting the hidden meaning in a piece of text.

Irony and sarcasm

- *The now abandoned HP TouchPad is officially the hottest piece of consumer electronics on Amazon.*
- *Life's too short, so be sure to read as many articles about celebrity breakups as possible.*
- *Loves being in this supah long line at the #DMV -- woo hoo*
- *I had never seen snow in Holland before but thanks to twitter and facebook I now know what it looks like. Thanks guys, awesome!*
- *On a bright note if downing gets injured we have Henderson to come in.*
- *Am glad 10 day forecast calling for lots of rain/cool temps. Was getting tired sun & dry conditions*

How do you know when someone is being sarcastic?

- Use of hashtags in tweets such as #sarcasm
- Large collections of tweets based on hashtags can be used to make a training set for machine learning
- But you still have to know which bit of the tweet is the sarcastic bit

To the hospital #fun #sarcasm

Man , I hate when I get those chain letters & I don't resend them , then I die the next day .. #Sarcasm

lol letting a baby goat walk on me probably wasn't the best idea. Those hooves felt great. #sarcasm

How else can you deal with it?

- Look for word combinations with opposite polarity, e.g. “rain” or “delay” plus “brilliant”

Going to the dentist on my weekend home. Great. I'm totally pumped. #sarcasm

- Inclusion of world knowledge / ontologies can help (e.g. knowing that people typically don't like going to the dentist, or that people typically like weekends better than weekdays.
- It's an incredibly hard problem and an area where we expect not to get it right that often

Ambiguity

- Social media can often pose ambiguities, for a number of reasons
- Misunderstandings: not much we can do

"I love Eminem" "I like Skittles better." "No, the rapper you idiot.." "You're the idiot! What's good about a M&M wrapper?!"

- Entity ambiguity: disambiguation techniques / linking to URI

I like how “RIP Fergie” is trending because of football and half the population of Twitter think that one of the Black Eyed Peas has died.

- But this is hard when there's no contextual reference...

Evaluation

- How can we evaluate opinion mining performance?
- What kind of results can we expect to get?
- What problems typically occur with evaluation?
- How can we compare existing tools and methods?

Comparing different opinion mining tools

- How do you compare different opinion mining tools, when there are so many out there and they all report different kinds of results?
- It is generally accepted that tools will be 50%-70% “accurate” out-of-the box.
- But what does this really mean?
- Seth Grimes has some pointers about this....

http://www.socialmediaexplorer.com/social-media-marketing/social-

1. Don't compare apples with oranges

- Not all tools do the same thing, even if they look the same
- Document-level vs topic-level sentiment
- One tool might be good at getting the overall sentiment of a tweet right, but rubbish at finding the sentiment about a particular entity
- e.g. the following tweet is classed as being negative about the Olympics:

skytrain seems to be having problems frequently lately. hope cause is upgraded and they work the kinks out before olympics.

- The tweet is (correctly) negative overall but not specifically about the Olympics

2. Use the same measurement scale

- Positive/negative/neutral vs scalar measurement (-5 to +5)
- Valence vs mood/orientation (e.g. happy, sad, angry, frustrated)
- Is reasonable emotion classification more useful to you than fantastic valence?
- How will you actually make use of the opinions generated to e.g. make decisions?

3. How is accuracy defined?

- NLP tools often use Precision, Recall and F-measure to determine accuracy
- But most opinion mining tools are only measured in terms of accuracy (Precision)
- How important is Recall?
- How important is the tradeoff between Precision and Recall?
- What about *contextual* relevance that incorporates timeliness, influence, activities, and lots of other still-fuzzy *social* notions?
- How trustworthy / important are the opinions? Sentiment from a valued customer may be more important than a one-time buyer

4. What's the impact of errors?

- Not all inaccuracies have the same impact
- If you're looking at aggregate statistics, a negative rating of a positive opinion has more impact than a neutral rating of a positive opinion
- How do neutral opinions affect aggregation? Are they considered? Should they be?
- In other cases, finding any kind of sentiment (whether with correct polarity or not) might be more important than wrongly detecting no sentiment and missing important information

Creating a gold standard

- Typically, we annotate a gold standard corpus manually and then compare the system results against that
- But have you ever tried doing manual annotation of tweets?
- It's harder than it looks...
- You have to be very clear what you want to annotate
- You have to understand what the author intended
- You need to decide how lenient you'll be
- You may need to decide if getting something right for the wrong reason is still OK

Positive or negative tweets?

RT @ssssab: Mariano: she used to be a very nice girl, before she discovered macdonalds

I'm tired after school today!

There was just a fire at work. Today is looking up.

Yesterday my son forgot his jacket at school. Today he remembered to bring home the jacket, but forgot his lunchbox.

Oh no. Ludo's got a new obsession with Dora the Explorer and now I find myself wondering around humming the theme tune.

I find myself sobbing at John Le Mesurier's beauty of soul. Documentary about him on BBC iPlayer

Opinionated or not?

The European sovereign debt crisis that’s spread from Greece to Italy and is roiling the region’s banks now has another potential victim: energy policy.

Labour got less this time than John Major did in 1997.

EUROPEAN LEADERSHIP - where is it?

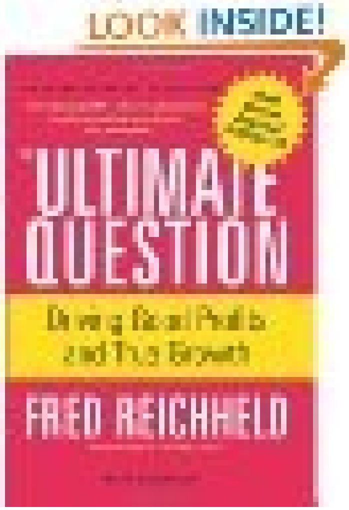
Looking into the future

- Typically, opinion mining looks at social media content to analyse people’s explicit opinions about a product or service
- This backwards-looking approach often aims primarily at dealing with problems, e.g. unflattering comments
- A forwards-looking approach aims at looking ahead to understanding potential new needs from consumers
- This is not just about looking at specific comments, e.g. “the product would be better if it had longer battery life”, but also about detecting non-specific sentiment
- This is achieved by understanding people's needs and interests in a more general way, e.g. drawing conclusions from their opinions about other products, services and interests.

Deep sentiment analysis

- The hardest thing about getting sentiment analysis right is uncovering exactly what is being meant
- Difference between a customer saying they merely like a brand and saying that they love it.
- Sentiment has many rich and nuanced dimensions that need to be teased apart to make it insightful.
- “An old lady told me that warm Dr. Pepper is delicious”
  - Is it only nice when warm? Does the author share the opinion of the old lady?
  - Could this be a new insight for the manufacturers/advertisers?
- “When I was a kid I loved Smarties”.
  - Should Smarties be targeted only at kids or do adults like them too?
- Classification of sentiment according to functional, insightful, emotional etc.

The Ultimate Question



- The book "The Ultimate Question" recently ranked #1 on the Wall Street Journal's Business Best-Sellers List and #1 on USA TODAY's Money Best-Sellers List.
- It's all about whether a consumer likes a brand enough to recommend it - this is the key to a company's performance.
- General sentiment detection isn't precise enough to answer this kind of question, because all kinds of “like” are treated equally
- Growing need for sentiment analysis that can get to very fine levels of detail, while keeping up with the enormous (and constantly increasing) volume of social media.

The problem of sparse data

- One of the difficulties of drawing conclusions from traditional opinion mining techniques is the sparse data issue
- Opinions tend to be based on a very specific product or service, e.g. a particular model of camera, but don't necessarily hold for every model of that brand of camera, or for every product sold by the company
- One solution is figuring out which statements can be generalised to other models/products and which are specific
- Another solution is to leverage sentiment analysis from more generic expressions of motivation, behaviour, emotions and so on, e.g. what type of person buys what kind of camera?

Approaches to Sentiment Analysis beyond opinion mining

- An interesting article from Seth Grimes about this:

  http://www.customerthink.com/article/mentions_to_meaning_analytics_journey

Summary

- Introduced the concept of Opinion Mining and Sentiment Analysis
- Simple examples of rule-based and ML methods for creating OM applications
- Dealing with social media
- Evaluation of opinion mining
- Looking ahead to the future

More information

- See the following paper for details and evaluation of a more complex version of the twitter application
- D. Maynard and A. Funk. Automatic detection of political opinions in tweets. In Proceedings of MSM 2011: Making Sense of Microposts. Workshop at 8th Extended Semantic Web Conference (ESWC 2011). Heraklion, Greece. June 2011 (download PDF)
- The EU-funded ARCOMEM and TrendMiner projects are dealing with lots of issues about opinion and trend mining from social media, and use GATE for this.
- http://www.arcomem.eu
- http://www.trendminer-project.eu/

References

- T. Baldwin and M. Lui. Language Identification: The Long and the Short of the Matter. In Proc. NAACL HLT ’10. http://www.aclweb.org/anthology/N10-1027.
- M. Kaufmann. Syntactic Normalization of Twitter Messages. http://www.cs.uccs.edu/~kalita/work/reu/REUFinalPapers2010/Kaufmann.pdf
- S. Choudhury and J. Breslin. Extracting Semantic Entities and Events from Sports Tweets. Proceedings of #MSM2011 Making Sense of Microposts. 2011.
- X. Liu, S. Zhang, F. Wei, M. Zhou. Recognizing Named Entities in Tweets. ACL'2011.
- A. Ritter, Mausam, Etzioni. Named entity recognition in tweets: an experimental study. EMNLP'2011.
- Doerhmann. Named Entity Extraction from the Colloquial Setting of Twitter. http://www.cs.uccs.edu/~kalita/work/reu/REU2011/FinalPapers/Doehermann.pdf
- S. Carter, W. Weerkamp, E. Tsagkias. Microblog Language Identification: Overcoming the Limitations of Short, Unedited and Idiomatic Text. Language Resources and Evaluation Journal. 2013 (Forthcoming)
- Johan Bollen, Huina Mao, Xiaojun Zeng, Twitter mood predicts the stock market, Journal of Computational Science, Volume 2, Issue 1, March 2011..

Some more demos to try

- http://sentiment.christopherpotts.net/lexicon/ Get sentiment scores for single words from a variety of sentiment lexicons
- http://sentiment.christopherpotts.net/textscores/ Show how a variety of lexicons score novel texts
- http://sentiment.christopherpotts.net/classify/ Classify tweets according to various probabilistic classifier models

Questions?