Opinionated Recommendation

Michael O'Mahony

Overview

Motivations & Preliminaries

Algorithms & Techniques

From Reviews to Recommendations, Case Studies…

Part 1 – Preliminaries

How would you start to build a recommender system for a hotel booking site?

Hotels, homes and more

- Hotels 168
- Apartments 114
- Motels 41
- Hostels 18
- Holiday homes 16
- Inns 8
- Bed and breakfasts 5

Review score

- Superb: 9+ 7
- Very good: 8+ 89
- Good: 7+ 170
- Pleasant: 6+ 211
- No rating 13

Facility

- WiFi 328
- Parking 269
- Airport shuttle 72
- Fitness centre 104
- Non-smoking rooms 267
- Indoor pool 7
- Spa and wellness centre 15
- Family rooms 208
- Outdoor pool 31
- Pets allowed 92
- Facilities for disabled guests 125

Hilton San Francisco Union Square ★★★★

Union Square, San Francisco – Metro access

deal 2928

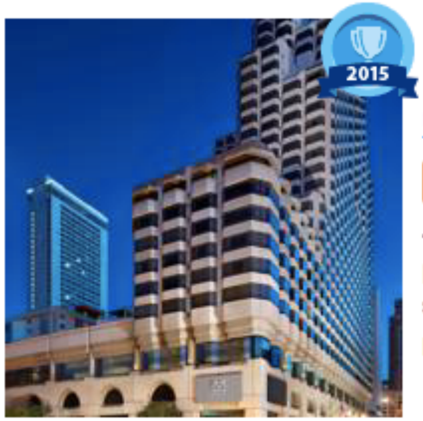
Located in the heart of downtown San Francisco and just a 5-minute walk from the Powell Street subway station, this Hilton features an on-site Starbucks and a beautiful courtyard pool.

Booked 25 times today

Good 7.6

3,220 reviews

Show prices



Parc 55 San Francisco - a Hilton Hotel ★★★★

Union Square, San Francisco – Metro access

deal 3243

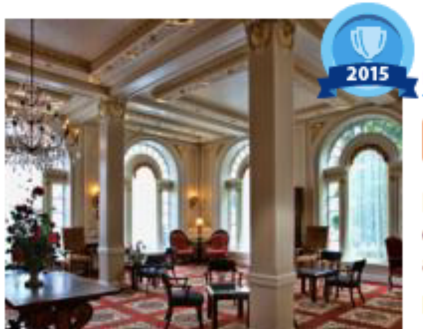
This downtown San Francisco hotel is 1 block from Union Square and the Hallidie Plaza, which features a cable car stop. The hotel offers massage services and a tour desk.

Booked 22 times today

Good 7.9

2,777 reviews

Show prices



Beresford Arms ★★★

Union Square, San Francisco

deal 2525

Named on the National Register of Historic Places, this hotel offers Victorian charm combined with modern amenities and is located a short stroll from area attractions, including cable car lines.

Booked 17 times today

Good 7.8

3,424 reviews

Show prices

Sources of Recommendation Knowledge

User ratings (scores, stars etc) ⇒ Collaborative Filtering

If we have information about which users stayed in which hotels or (even better) how they rated these hotels then we can use this ratings data to build an CF style recommender system (user-based, item-based, MF).

Metadata (facilities, price, etc) ⇒ Content-Based Rec.

If we have information about facilities (leisure centre, business centre, restaurant, bar, wifi etc) then we can represent hotels and users using these features to produce a content-based recommender.

Hybrid approach - e.g. combined CF and content-based approach

Any other sources of useful information?

Hotels, homes and more

- Hotels 168
- Apartments 114
- Motels 41
- Hostels 18
- Holiday homes 16
- Inns 8
- Bed and breakfasts 5

Review score

- Superb: 9+ 7
- Very good: 8+ 89
- Good: 7+ 170
- Pleasant: 6+ 211
- No rating 13

Facility

- WiFi 328
- Parking 269
- Airport shuttle 72
- Fitness centre 104
- Non-smoking rooms 267
- Indoor pool 7
- Spa and wellness centre 15
- Family rooms 208
- Outdoor pool 31
- Pets allowed 92
- Facilities for disabled guests 125

2015

Hilton San Francisco Union Square ★★★★

Union Square, San Francisco – Metro access

deal 2928

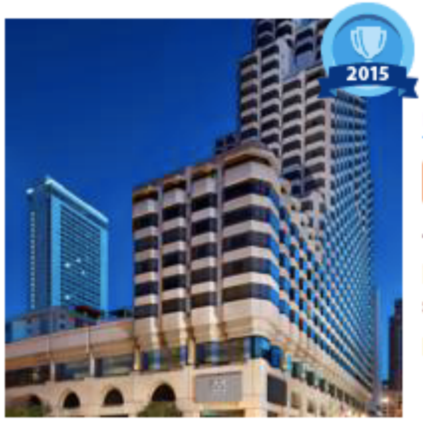
Located in the heart of downtown San Francisco and just a 5-minute walk from the Powell Street subway station, this Hilton features an on-site Starbucks and a beautiful courtyard pool.

Booked 25 times today

Good 7.6

3,220 reviews

Show prices



2015

Parc 55 San Francisco - a Hilton Hotel ★★★★

Union Square, San Francisco – Metro access

deal 3243

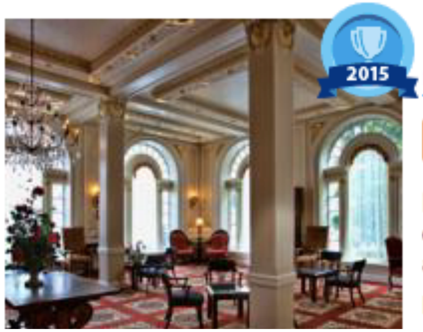
This downtown San Francisco hotel is 1 block from Union Square and the Hallidie Plaza, which features a cable car stop. The hotel offers massage services and a tour desk.

Booked 22 times today

Good 7.9

2,777 reviews

Show prices



2015

Beresford Arms ★★★

Union Square, San Francisco

deal 2525

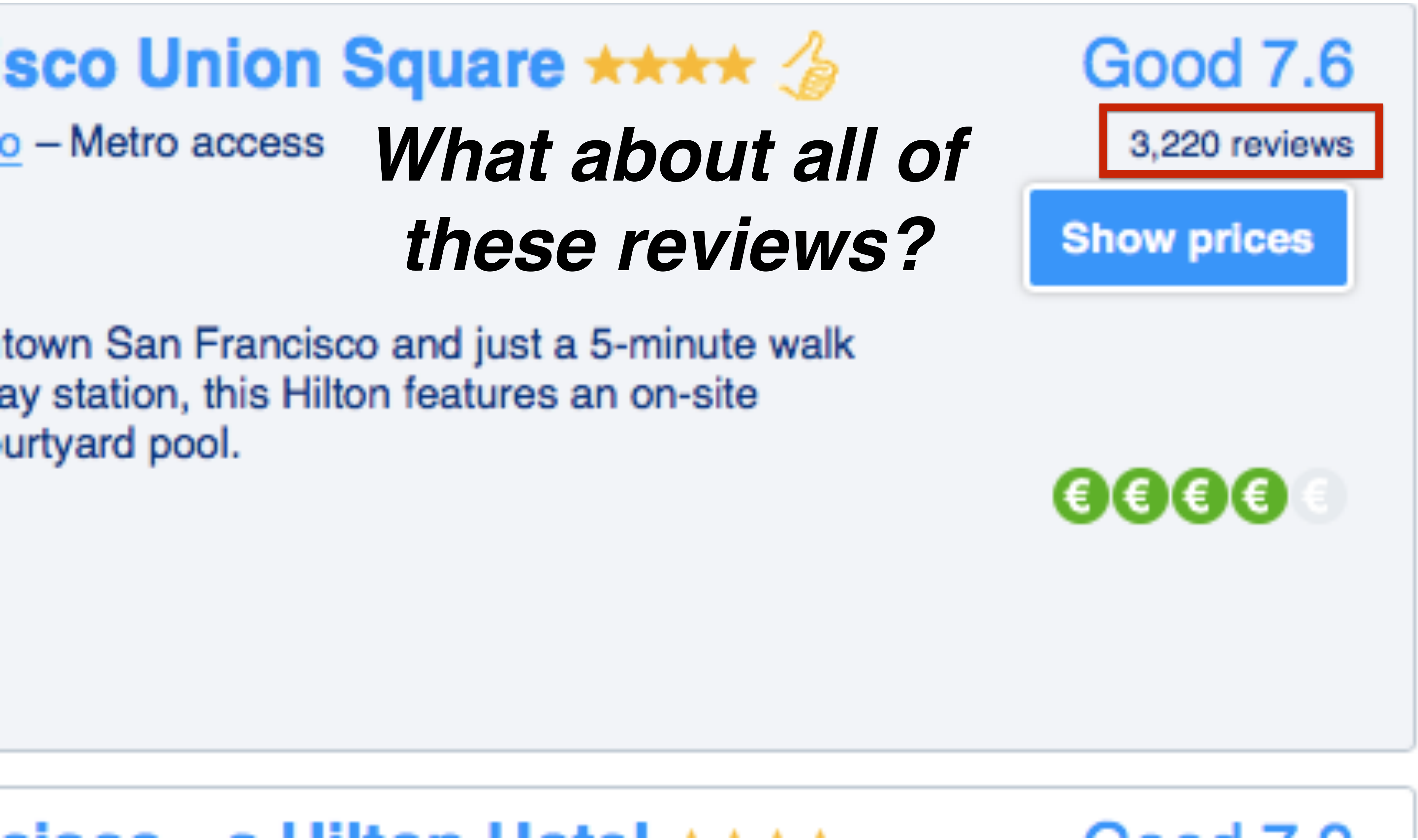
Named on the National Register of Historic Places, this hotel offers Victorian charm combined with modern amenities and is located a short stroll from area attractions, including cable car lines.

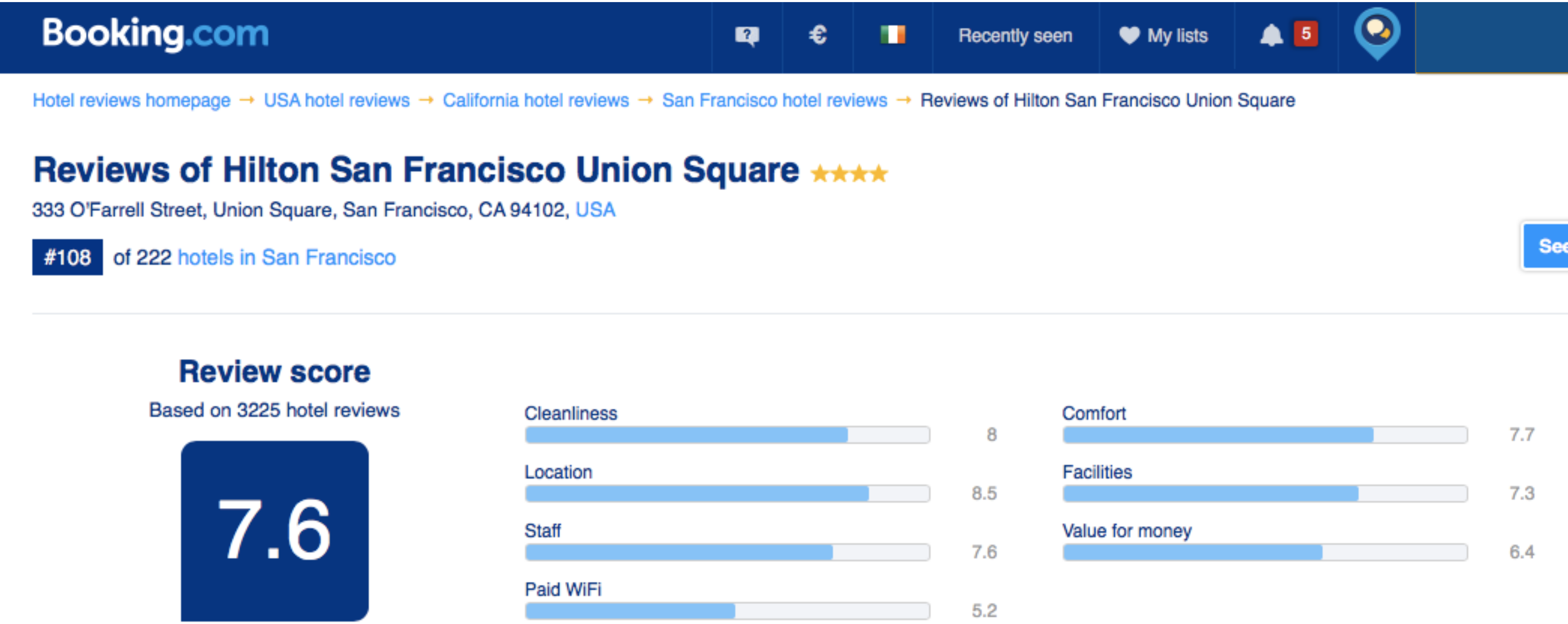
Booked 17 times today

Good 7.8

3,424 reviews

Show prices


sco Union Square
Good 7.6
3,220 reviews
o – Metro access
What about all of these reviews?
Show prices
town San Francisco and just a 5-minute walk
ay station, this Hilton features an on-site
urtyard pool.


Booking.com
Recently seen
My lists
5
Hotel reviews homepage → USA hotel reviews → California hotel reviews → San Francisco hotel reviews → Reviews of Hilton San Francisco Union Square
Reviews of Hilton San Francisco Union Square
333 O'Farrell Street, Union Square, San Francisco, CA 94102, USA
#108 of 222 hotels in San Francisco
Review score
Based on 3225 hotel reviews
7.6
Cleanliness 8
Comfort 7.7
Location 8.5
Facilities 7.3
Staff 7.6
Value for money 6.4
Paid WiFi 5.2

David

USA

1 review

7.5 **"We liked our stay! Good weekend!"**

- Leisure trip
- Couple
- Double Room with Two Double Beds
- Stayed 1 night
- Submitted via mobile

NEW

Guest stayed February 2016

The cost was more than we usually like to pay. Maybe a breakfast should be included in the price if the cost and parking is going to be so high. We really enjoyed our stay though.

Atmosphere was fun!

Worldtraveller

Australia

1 review

5.8 **"Booking in advance didn't help..."**

- Leisure trip
- Couple
- 2 rooms
- Stayed 4 nights

Guest stayed January 2016

Despite our best efforts - a request one year in advance!! - and a huge hotel - we could not get my mother's room together with our room on the same floor. My mother's room had no refrigerator and was small. We had a fridge and had to share it with her. The wi fi was complicated, but eventually sorted out by reception. The hotel is expensive, but we booked it hoping the brand would be reliable.

Opinions, Reviews, and Recommendations

Opinionated Recommendation

What if we could harness the opinions expressed in reviews?

Could we produce new product/service/item descriptions based on the features discussed in reviews?

Would this change the way we approached recommendation? Similarity vs sentiment?

Could we use these opinions to help explain or justify recommendations to user?

Overview

Sources of Recommendation Knowledge

User-generated reviews as a novel and plentiful source of recommendation knowledge.

Opinion Mining Primer

Shallow NLP for opinion mining from text.

Case-Studies in using Opinion Mining for Recommendation

A variety of related, interconnected case-studies looking at the use of opinion mining in various aspects of recommendation and related tasks.

Active Research

The work described in these lectures is the subject of ongoing research in the Recommender Systems Group in Insight UCD.

It is based on work contained in about a dozen published scientific papers (see next slide).

The work continues to be developed …

Ruihai Dong, Kevin McCarthy, Michael P. O'Mahony, Markus Schaal, Barry Smyth: Towards an intelligent reviewer's assistant: recommending topics to help users to write better product reviews. IUI 2012: 159-168

Ruihai Dong, Markus Schaal, Michael P. O'Mahony, Kevin McCarthy, Barry Smyth: Harnessing the Experience Web to Support User-Generated Product Reviews. ICCBR 2012: 62-76

Ruihai Dong, Michael P. O'Mahony, Markus Schaal, Kevin McCarthy, Barry Smyth: Sentimental product recommendation. RecSys 2013: 411-414

Ruihai Dong, Markus Schaal, Michael P. O'Mahony, Barry Smyth: Topic Extraction from Online Reviews for Classification and Recommendation. IJCAI 2013

Ruihai Dong, Markus Schaal, Michael P. O'Mahony, Kevin McCarthy, Barry Smyth: Opinionated Product Recommendation. ICCBR 2013: 44-58

Ruihai Dong, Michael P. O'Mahony, Barry Smyth: Further Experiments in Opinionated Product Recommendation. ICCBR 2014: 110-124

Khalil Muhammad, Aonghus Lawlor, Rachael Rafter, Barry Smyth: Great Explanations: Opinionated Explanations for Recommendations. ICCBR 2015: 244-258

Khalil Muhammad, Aonghus Lawlor, Barry Smyth: A Live-User Study of Opinionated Explanations for Recommender Systems, IUI 2016

Khalil Muhammad, Aonghus Lawlor, Barry Smyth: On the Use of Opinionated Explanations to Rank and Justify Recommendations , FLAIRS 2016

Khalil Muhammad, Aonghus Lawlor, Barry Smyth: When You’re Explaining, You’re ... Ranking: Explanation and Ranking in Opinionated Recommender Systems , IJCAI 2016

Sources of Recommendation Knowledge

Transactional & Behavioural Data

Clicks, purchases, likes, rating, actions (save to wish-list)

Content & Meta Data

Features, tags, and terms, structured and unstructured.

Experiential Data

User-generated opinions. Based on real subjective experiences vs product-owner catalog metadata.

Why are Reviews Useful?

Reviews are ubiquitous and abundant.
Most large sites now host millions of user-generated reviews.

Reviews are independent (usually) and (often) insightful.
Real user reviews help us to understand the reality of a product or service, its pros and cons, rather than simply highlighting the manufacturer's marketing promises.

Reviews help us to make better decisions.
Reviews matter. Research shows that reviews help users to make better decisions. They increase conversion rates and improve satisfaction.

Consumers Trust Reviews

88% have read reviews to determine the quality of a local business (regularly or occasionally).

39% read reviews on a regular basis …

… and only 12% of people don't pay attention to reviews.

BrighLocal, Consumer Review Survey

Reviews vs Recommendation

88% of consumers say they trust online reviews as much as personal recommendations (vs. 79% in 2013)

Only 13% said they do not trust reviews as much as personal recommendations (vs. 21% in 2013)

BrighLocal, Consumer Review Survey

Reviews & Influence

61% of customers read online reviews before making a purchase decision, **50** or more reviews per product can mean a **4.6%** increase in conversion rates and can lead to an **18%** increase in sales revenue.

63% of users more likely to purchase from sites that have user generated reviews.

Visitors who interact with reviews and customer Q&A are **105%** more likely to purchase, and spend **11%** more than others.

Consumer reviews are **12 times** more trusted than manufacturer descriptions.

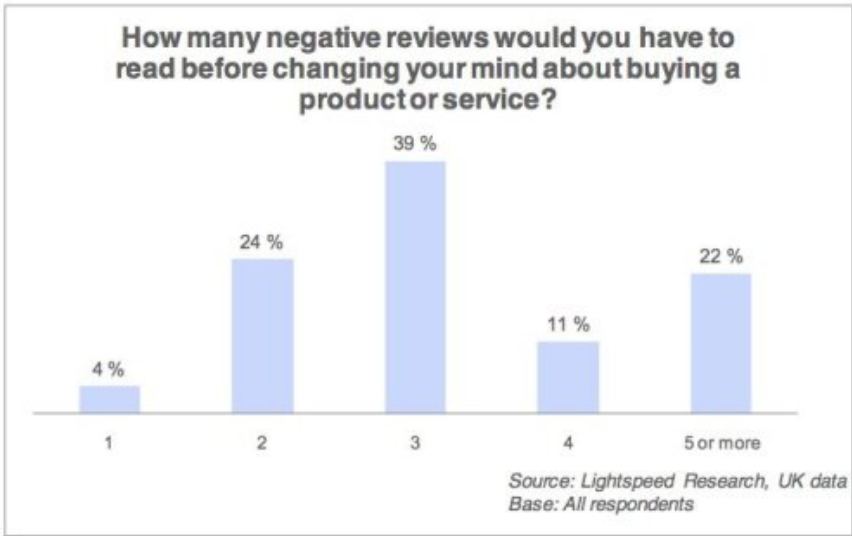
Negative Reviews

Negative reviews are valuable too...

A mix of positive and negative reviews helps to improve consumer trust in the opinions they read. Recent stats from Reevoo suggest that the presence of bad reviews actually improves conversions by 67%.

Too many bad reviews aren't good …

The benefits of bad reviews depends on the proportion of good to bad. The negative reviews make the positive ones more believable, but there is a point at which they ring alarm bells for consumers.



The Dark Side of Reviews

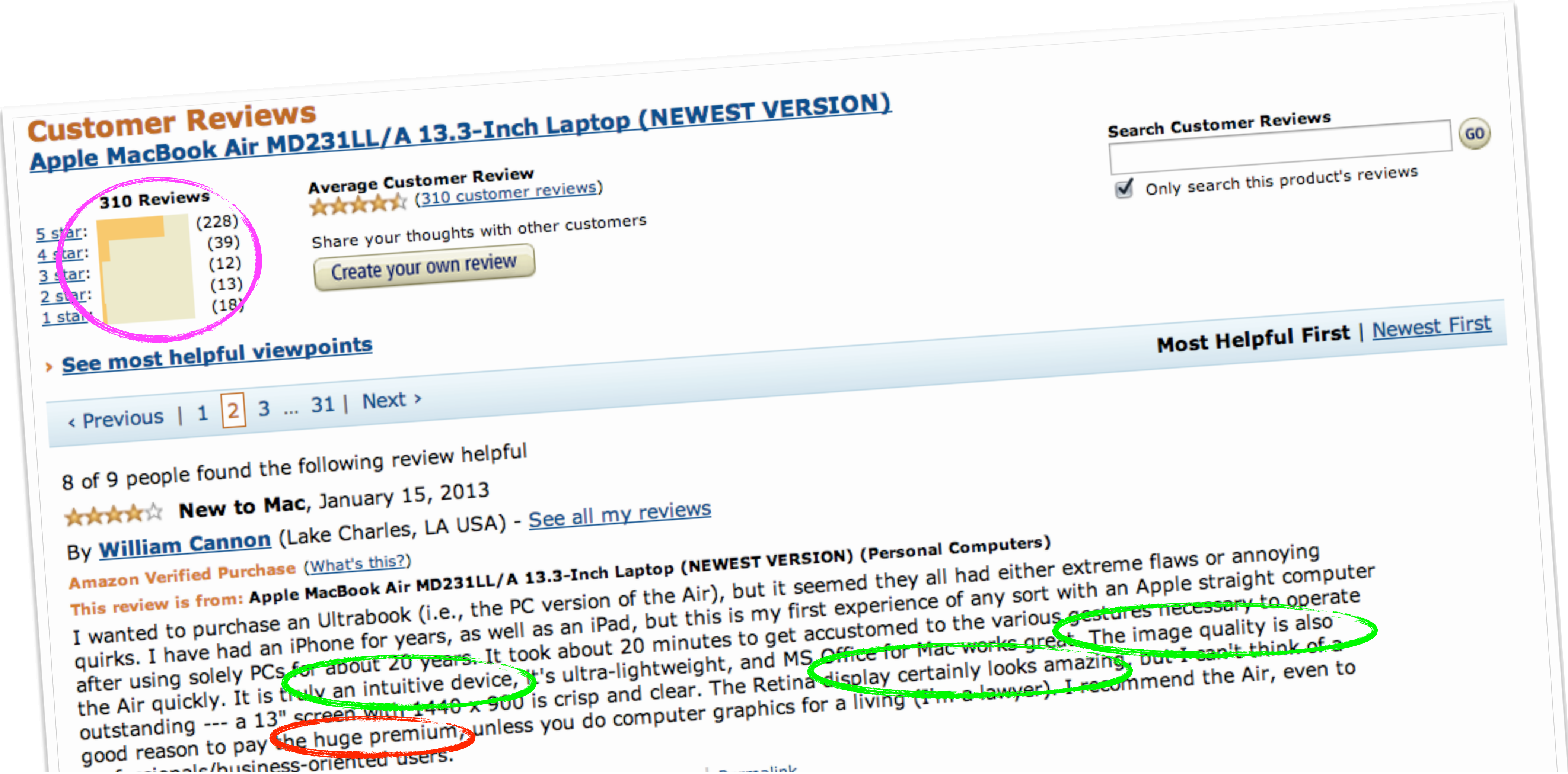
Fake reviews are a problem! Especially for recommender systems.

In 2015 Amazon filed a lawsuit against more than 1000 people it claimed were writing fake reviews.

Researchers actively looking for ways to automatically detect fake reviews…

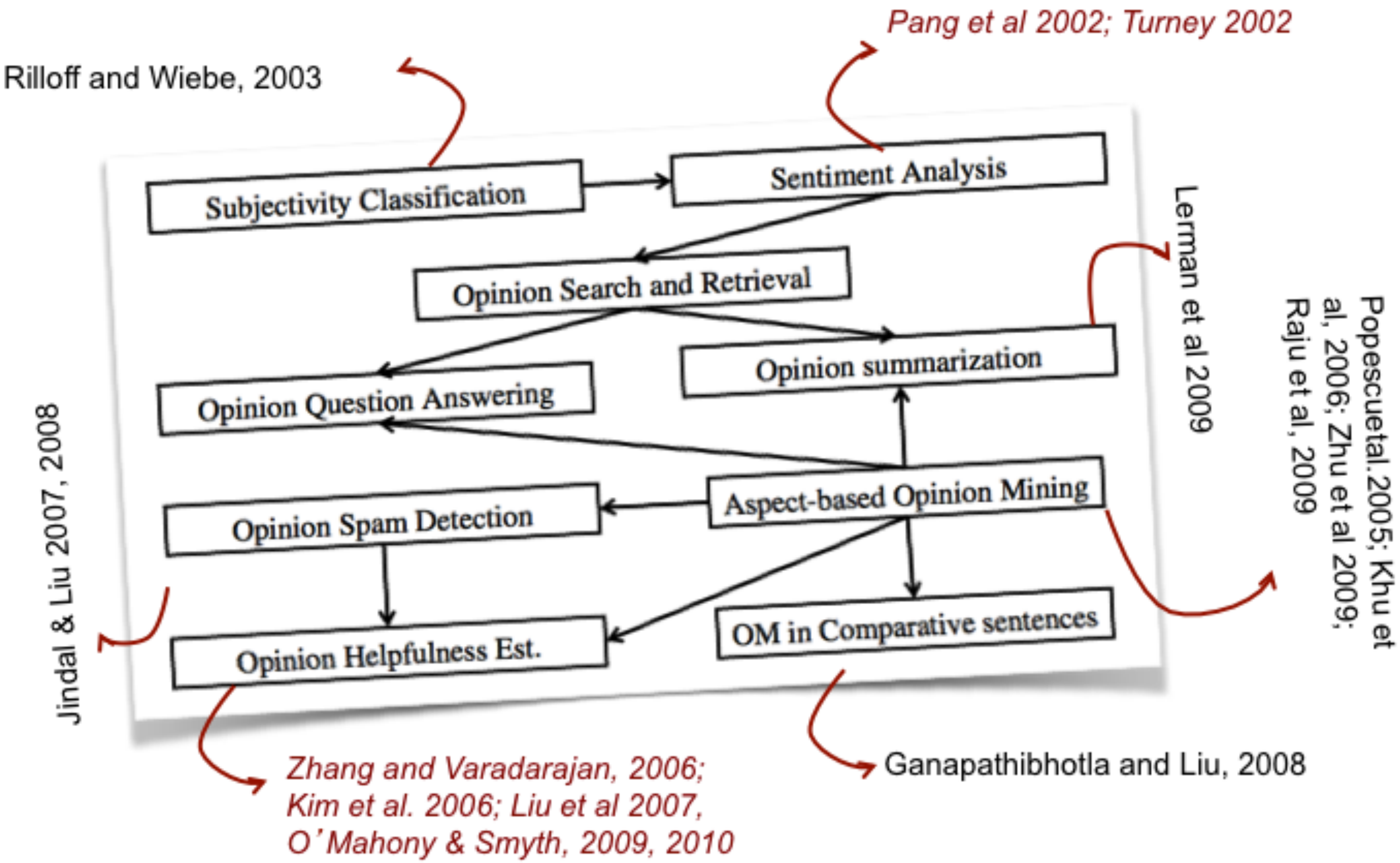
An Opinion Mining Primer

Opinion Mining Example



The Opinion Mining Landscape

source: Ester WWW (2013)



Of interest to us …

Sentiment Analysis
Automatically understanding the *polarity* of an opinion; is what has been said positive or negative or neutral?

Opinion Helpfulness
Automatically understanding the value of the opinion in terms of how helpful it might be to others.

Aspect-Oriented Opinion Mining
Automatically identifying the aspects or features discussed in an opinion; e.g. what product features is the reviewer discussing?

Opinion Summarisation
The ability to automatically summarise a potentially large space of opinions; e.g. all of the reviews for a hotel.

But first …

Opinion Mining Primer

Given a set of reviews how can we begin to extract opinions, identify features, and estimate sentiment?

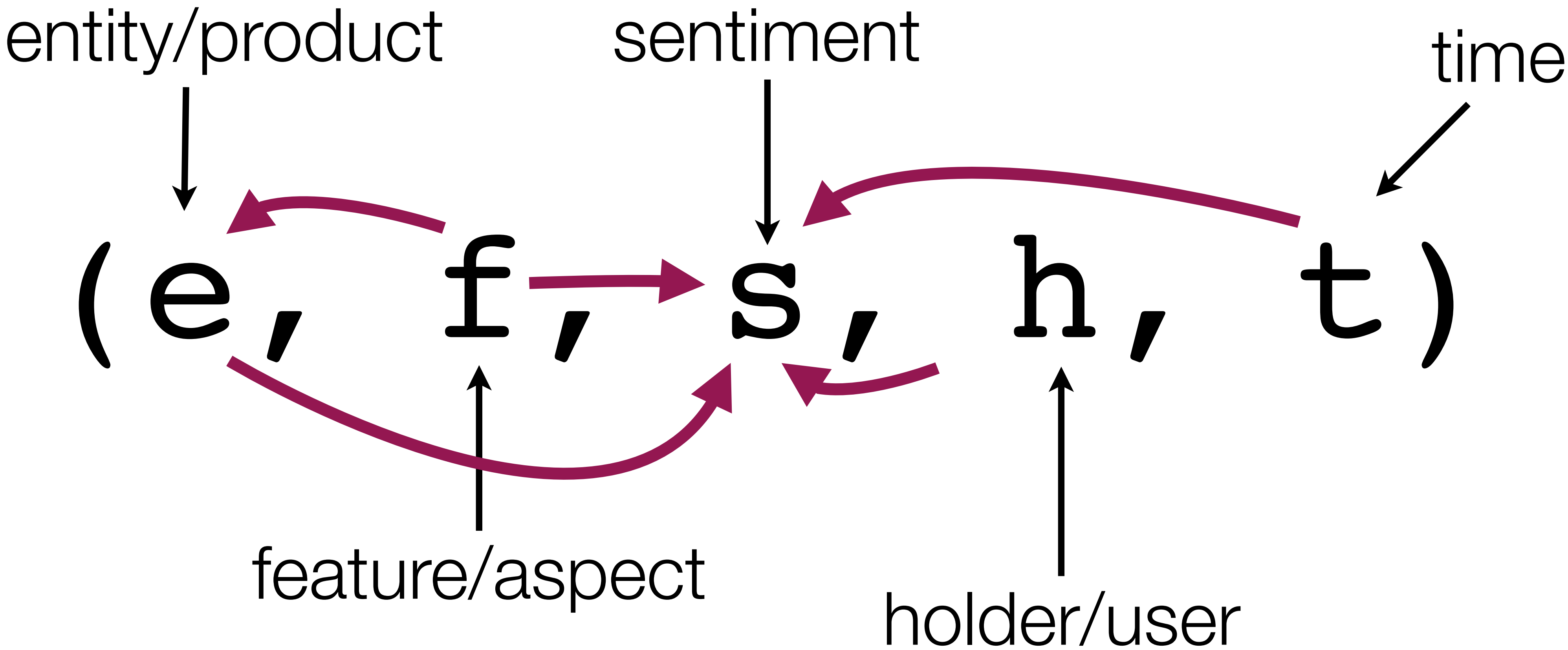
In what follows we will describe the use of shallow natural language processing techniques to do this.

NLP is a huge field in itself and we will gloss over it.
Use a standard NLP toolkit to do the heavy lifting and (for this course) it is more important for you to know how to do this.

The Anatomy of an Opinion

`(e, f, s, h, t)`

The Anatomy of an Opinion



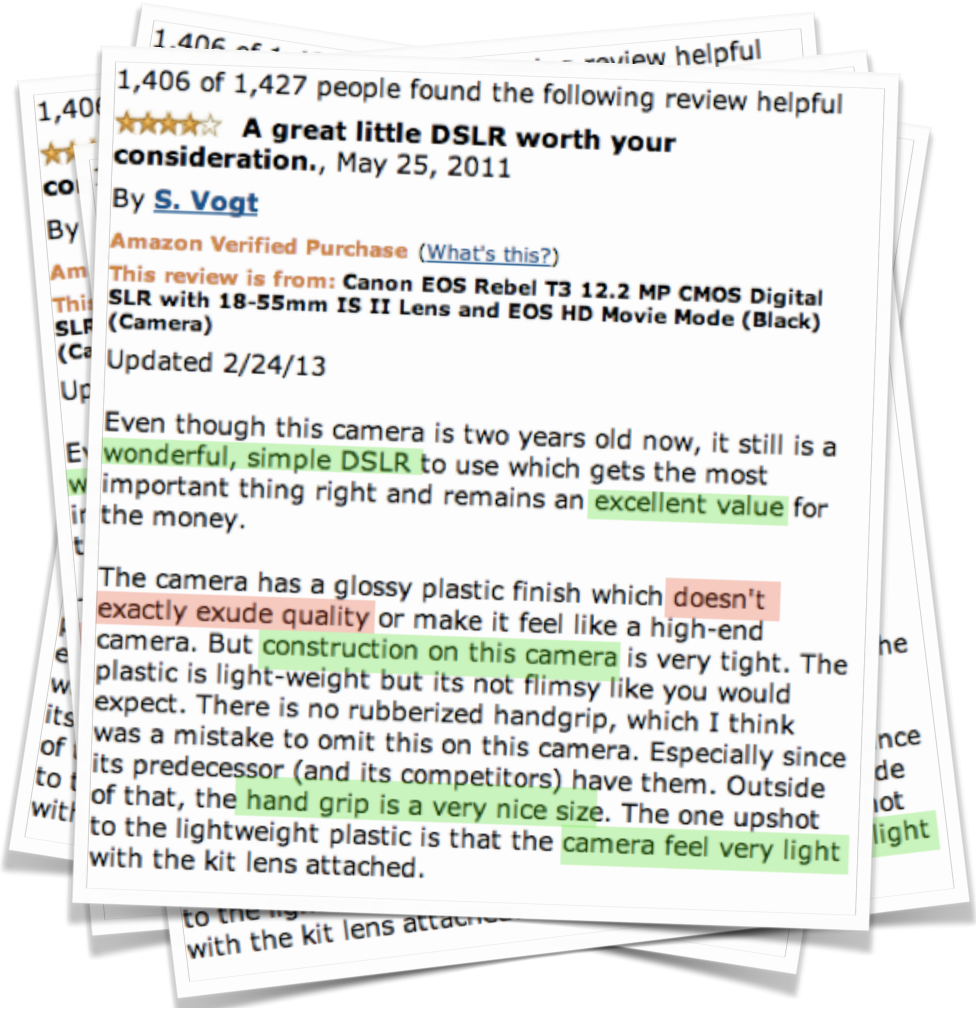
The Anatomy of an Opinion

... focus on extracting only *features* and *sentiment* from a collection of reviews for a product ...

`(e, f, s, h, t)`

... to produce aggregate product descriptions for each product.

Aspect-based Opinion Mining



DSLR:	++
Value:	++
Build Quality:	-
Weight:	+
Grip:	+
Image Quality:	+
Resolution:	-
Price:	+
Battery Life:	---

Nikon D90 Review (Amazon)

Raw Review Text

It provides amazing picture quality and is perfect for amateur photographers. The battery life is truly outstanding, I am not even going to buy a spare battery. It's weight is heavier than the D60, but that's not a problem. The stereo sound and autofocus are limited, however. The viewfinder is also really poor.

Aspects/Features

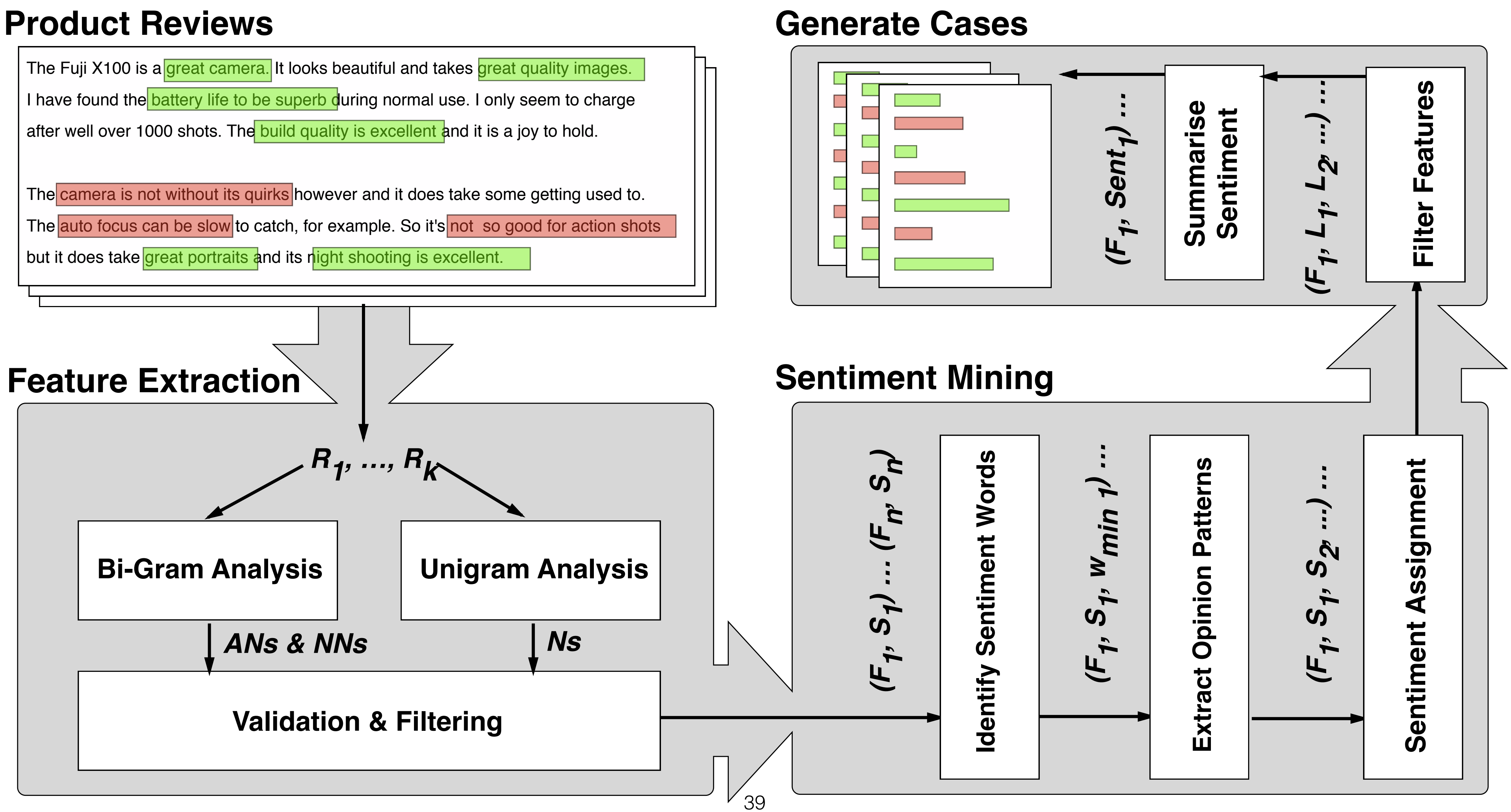
It provides amazing **picture quality** and is perfect for amateur photographers. The **battery life** is truly outstanding, I am not even going to buy a spare **battery**. It's **weight** is heavier than the D60, but that's not a problem. The stereo **sound** and **autofocus** are limited, however. The **viewfinder** is also really poor.

Sentiment Analysis

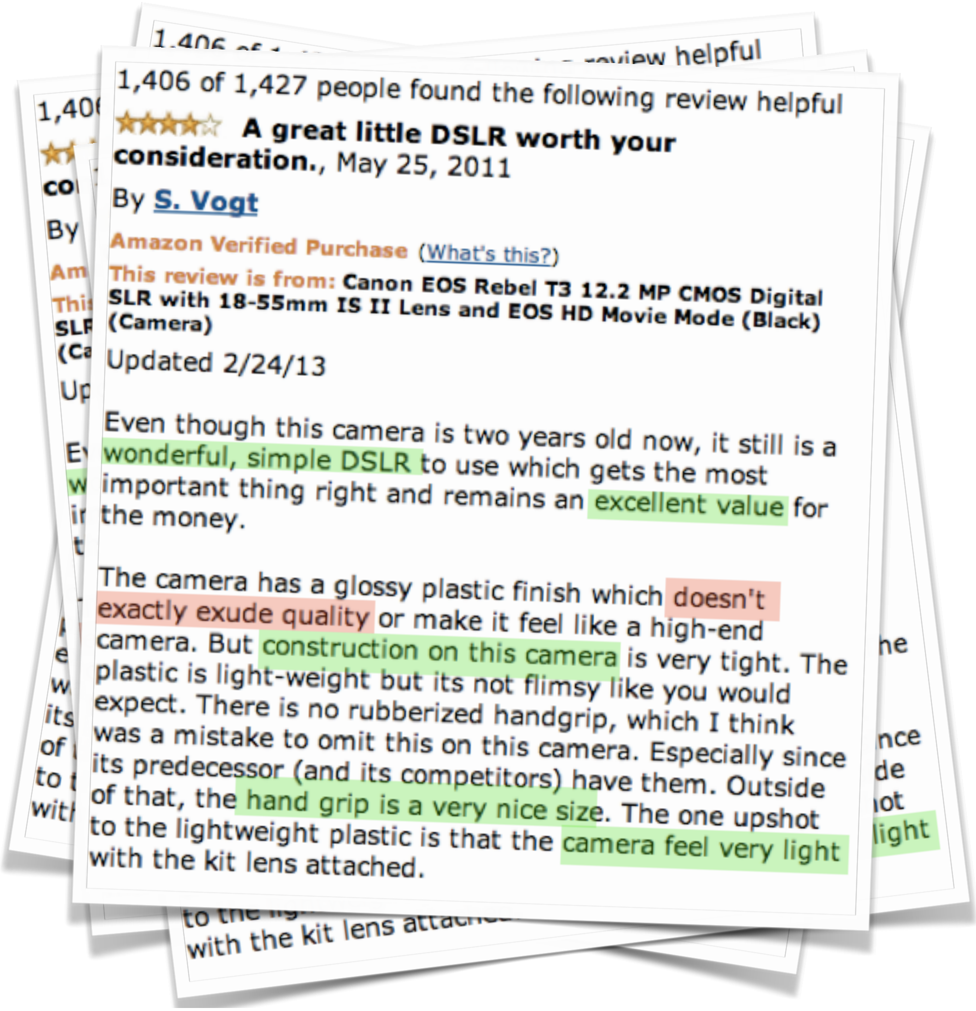
It provides *amazing* **picture quality** and is perfect for amateur photographers. The **battery life** is truly *outstanding*, I am not even going to buy a spare **battery**. It's **weight** is heavier than the D60, but that's *not* a *problem*. The stereo **sound** and **autofocus** are *limited*, however. The **viewfinder** is also really *poor*.

Part 2 – Algorithms

(Dong et al, ICCBR 2013)



Experiential Case



1,406 of 1,427 people found the following review helpful

A great little DSLR worth your consideration., May 25, 2011

By S. Vogt

Amazon Verified Purchase (What's this?)

This review is from: **Canon EOS Rebel T3 12.2 MP CMOS Digital SLR with 18-55mm IS II Lens and EOS HD Movie Mode (Black) (Camera)**

Updated 2/24/13

Even though this camera is two years old now, it still is a wonderful, simple DSLR to use which gets the most important thing right and remains an excellent value for the money.

The camera has a glossy plastic finish which doesn't exactly exude quality or make it feel like a high-end camera. But construction on this camera is very tight. The plastic is light-weight but its not flimsy like you would expect. There is no rubberized handgrip, which I think was a mistake to omit this on this camera. Especially since its predecessor (and its competitors) have them. Outside of that, the hand grip is a very nice size. The one upshot to the lightweight plastic is that the camera feel very light with the kit lens attached.

DSLR:	++
Value:	++
Build Quality:	-
Weight:	+
Grip:	+
Image Quality:	+
Resolution:	-
Price:	+
Battery Life:	---

An Algorithm Sketch

Identify ***candidate features*** in review text.

Filter out ***unlikely/weak candidates*** to preserve high quality features.

Assign ***sentiment scores*** for features.

Handle ***negation terms,*** which may invert sentiment polarity (e.g. "This camera is not bad!").

Produce ***experiential cases*** by aggregating features, frequency, sentiment information at item level.

Preparing the Text

The NLP Bit

Sentence Segmentation

Split raw review text into distinct sentences; simple text splitting.

Text Chunking

Dividing sentences into syntactically correlated chunks

Part-of-Speech Tagging

Marking word tokens with corresponding POS/word-type tags

Apache OpenNLP

Easy to use, open NLP processing library … for all your NLP needs!

Text Chunking

Text chunking consists of dividing a text in syntactically related sets of words; eg. noun groups and verb groups.

Consider "He reckons the current account deficit will narrow to only 1.8 billion in September." … which is divided as follows:

[NP He] [VP reckons] [NP the current account deficit] [VP will narrow] [PP to] [NP only 1.8 billion] [PP in] [NP September]

Why Text Chunking?

“The battery life is truly outstanding.”

vs

“It has a great battery, life is good!”

With the first, the feature is *battery life* but in the second the feature is *battery*…

Why Text Chunking?

“The battery life is truly outstanding.”

vs

“It has a great battery, life is good!”

… by chunking the sentences we can focus on the appropriate sentence fragment.

Text Chunking

Token	Chunk Tag
It	B-NP
's	B-VP
weight	B-NP
is	B-VP
heavier	B-ADJP
than	B-PP
the	B-NP
D60	I-NP
,	O
but	O
that	B-NP
's	B-VP
not	O
a	B-NP
problem	I-NP
.	O

“It's weight is heavier than the D60, but that's not a problem.”

Phrase Level Tags:

ADJP - Adjective Phrase
NP - Noun Phrase
VP - Verb Phrase

B-CHUNK = 1st word of chunk & I-CHUNK for each other word in chunk; the O tag is used for tokens which are not part of any chunk

see <u>http://www.clips.ua.ac.be/pages/mbsp-tags</u>

POS Tagging

Token	Chunk Tag	POS Tag
It	B-NP	PRP
's	B-VP	VBZ
weight	B-NP	NN
is	B-VP	VBZ
heavier	B-ADJP	JJR
than	B-PP	IN
the	B-NP	DT
D60	I-NP	NNP
,	O	,
but	O	CC
that	B-NP	DT
's	B-VP	VBZ
not	O	RB
a	B-NP	DT
problem	I-NP	NN
	O	.

Word Level Tags:

IN - Preposition or subordinating conjunction
JJ - Adjective
JJR - Adjective, comparative
NN - Noun, singular
NNS - Noun, plural
NNP - Proper noun, singular
NNPS - Proper noun, plural

For full list, see:
http://www.clips.ua.ac.be/pages/mbsp-tags

Summary

Sentence segmentation, text chunking, and POS tagging allow us to structure review sentences into meaningful sequences of categorised words.

From here we can start to look for patterns of words and word types that are likely to correspond to expressions of opinions.

And from there we can identify features and evaluate sentiment polarity.

Identifying Features

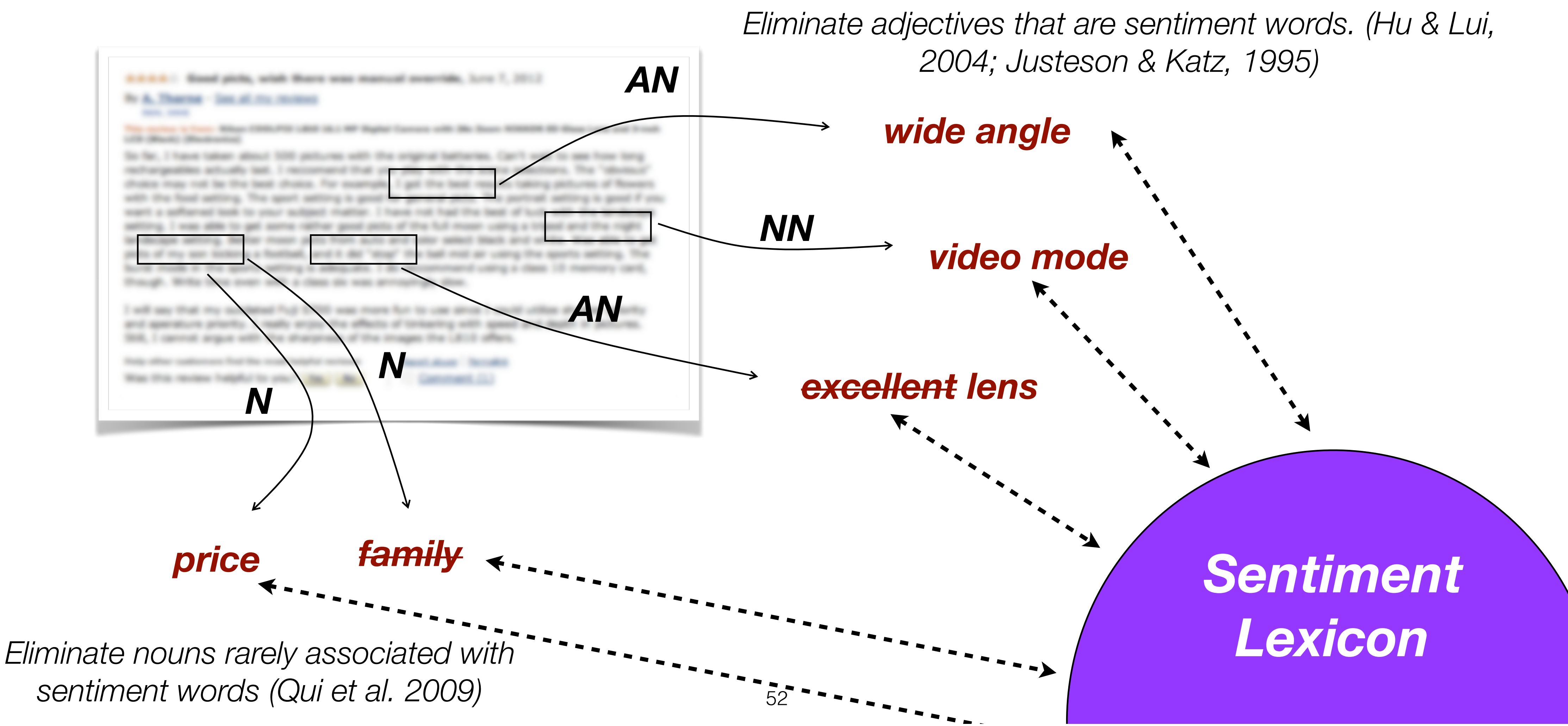
Feature Extraction

For our purposes we will focus on two basic types of review features…

Single Noun Features - single nouns that can be identified in a review.

Bigram Features - pairs of words; focus on adjectives followed by a noun (AN) and nouns followed by nouns (NN)

Extracting Features: Unigrams & Bigrams



Sentiment Lexicons

Use sentiment lexicons …

… which are essentially lists of words marked as either positive or negative.

There are a number of different lexicons available…

Extracting Bi-Gram Features

Look for occurrences of AN and NN’s in text chunks.

E.g. “wide angle” (AN) and “video mode” (NN)

But … exclude AN who's adjective is a sentiment word.
For example, “… great flash …” is an AN but it clearly relates to a single noun feature (“flash”) with “great” acting as the sentiment modifier.

That is, exclude bi-grams whose adjective is in the sentiment lexicon.

Extracting Single Noun Features

Extract candidate set of (non stop-word) nouns from reviews.

For each of these candidates calculate how often it appears close to a sentiment word (same chunk) based on all occurrences in reviews.

Eliminate those candidates that appear with a sentiment word less frequently than some threshold (e.g. <70%).

This will eliminate nouns that are unlikely to be product features (e.g. “family”, “day”, “vacation”).

Filtering Weak Features

Post-hoc Filtering

Now we have a set of potential features, $\boldsymbol{f_1}$***, …,*** $\boldsymbol{f_n}$.

Some/many of these may be very unusual aspects that are rarely discussed in reviews.

On the grounds that this may be because they are not valid features, the final step is to eliminate those that occur in fewer than ***k*** out of ***n*** reviews**.**

Typically we use a different ***k*** for single-noun and bi-gram features, determined by experiment.

Assigning Sentiment

Evaluating Opinion Sentiment

Evaluate the sentiment of f_i, in sentence s_j from review r_k.

First locate any sentiment words in s_j.

Find words w_1, w_2,… that are in s_j and that are also in the sentiment lexicon.

If there are no such words, f_i is marked as ***neutral.***

… and you can move on to the next feature.

Evaluating Feature Sentiment

If there are sentiment words then the feature may be a valid sentiment-laden aspect.

However, not every feature that is close to a sentiment word makes for a valid/good aspect.

In what follows we describe an approach to ***opinion pattern mining*** that attempts to identify good features/aspects which conform to common speech patterns …

Opinion Patterns

If there are sentiment words …

Determine the one that is closest to $\boldsymbol{f_i}$ in $\boldsymbol{s_j}$ (checking either side of $\boldsymbol{f_i}$). We call this closest word $\boldsymbol{w_{min}}$

$$\text{“The } \underbrace{\text{battery life}}_{f_i} \text{ is truly } \underbrace{\text{outstanding}}_{w_{min}}\text{.”}$$

Opinion Patterns

Next, get the POS tags for the words between $\boldsymbol{f_i}$ and $\boldsymbol{w_{min}}$.

This POS sequence is an ***opinion pattern***, which in this case is ***{FEATURE, VBZ, RB, JJ}***.

$$\text{“The } \underbrace{\text{battery life}}_{\textit{FEATURE},\ \boldsymbol{f_i}}\ \overbrace{\text{is}}^{\textit{VBZ}}\ \overbrace{\text{truly}}^{\textit{RB}}\ \underbrace{\text{outstanding}}_{\textit{JJ},\ \boldsymbol{w_{min}}}\text{.”}$$

Opinion Patterns

We produce these opinion patterns for all of the reviews in a given product set (typically a given product category).

For each pattern, e.g. ***{FEATURE, VBZ, RB, JJ}***, we count how often it occurs and deem those patterns that occur ***>t*** times to be ***valid***.

If a pattern is deemed not to be valid it's corresponding feature is assigned a ***neutral*** sentiment.

Assigning Sentiment

For our valid patterns we now need to assign sentiment to their corresponding features.

If sentence s_j containing f_i contains a ***negation term*** *then the sentiment of* f_i *is opposite polarity to the sentiment of* w_{min}.

That is, if there is a negation term and pos(w_{min}*) then sentiment(*f_i*) = neg.*

If s_j *does not contain any negation terms then the sentiment of* f_i *is the polarity of* w_{min}.

That is, if there is no negation term and pos(w_{min}*) then sentiment(*f_i*) = pos.*

Dealing with Negation

Negation Terms

Valid negation terms are “***not***” and “…***n’t***” subject to the following rules:

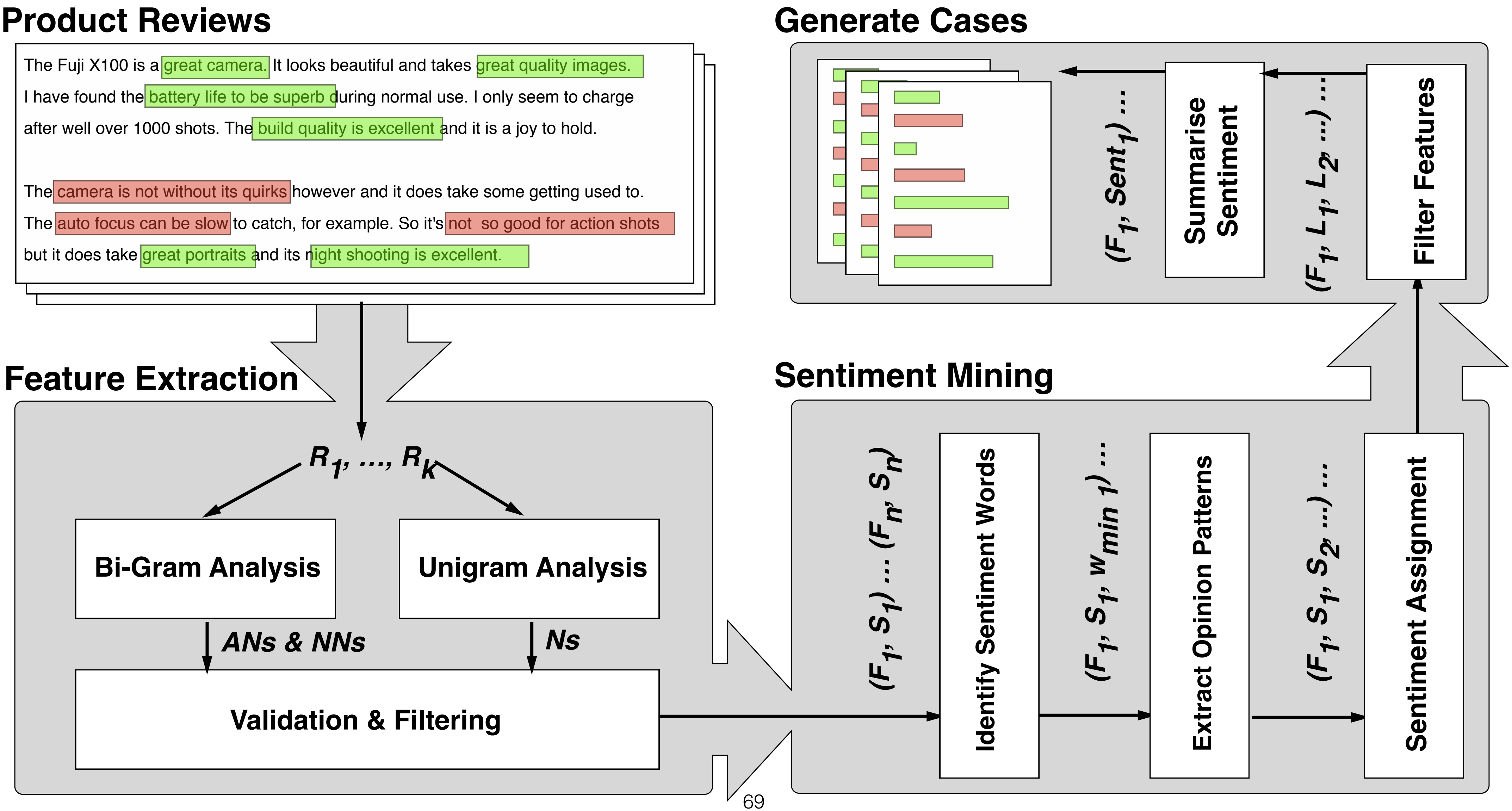
1. The negation term must be an adverb (RB); otherwise ignore it.

2. If then negation term is “***not***” and it is followed by “***only***” then it is ignored.

It provides *amazing* **picture quality** and is perfect for
amateur photographers. The **battery life** is truly
outstanding, I am not even going to buy a spare **battery**. It's
weight is heavier than the D60, but that's *not* a *problem*.
The **stereo sound** and **autofocus** are *limited*, however.
The **viewfinder** is also really *poor*.

Experiential Cases

(Dong et al, ICCBR 2013)



Quick Recap

Starting with some product ***P*** and set of reviews ***R*** …

We can analyse each of the reviews $\mathbf{r_i}$ in **R** to extract a set of features, $\boldsymbol{f_1}$, $\boldsymbol{f_2}$, **…**

For each feature we can calculate its frequency of occurrence in ***R*** and for each occurrence we can assign a sentiment label from ***{pos, neg, neut}***.

We can aggregate these features, frequency/popularity scores, and sentiment scores to produce a representation for ***P*.**

Case Generation

Experience (product) cases are then assembled from the features extracted from each items's reviews $\boldsymbol{r_1}$, $\boldsymbol{r_2}$, *…*

$\mathbf{f_1} \Rightarrow \{\text{pos, pos, neg, ...}\}$

$\mathbf{f_2} \Rightarrow \{\text{neut, pos, neg, ...}\}$

:

:

$\mathbf{f_n} \Rightarrow \{\text{neg, neg, neut, ...}\}$

Generating Experiential Cases

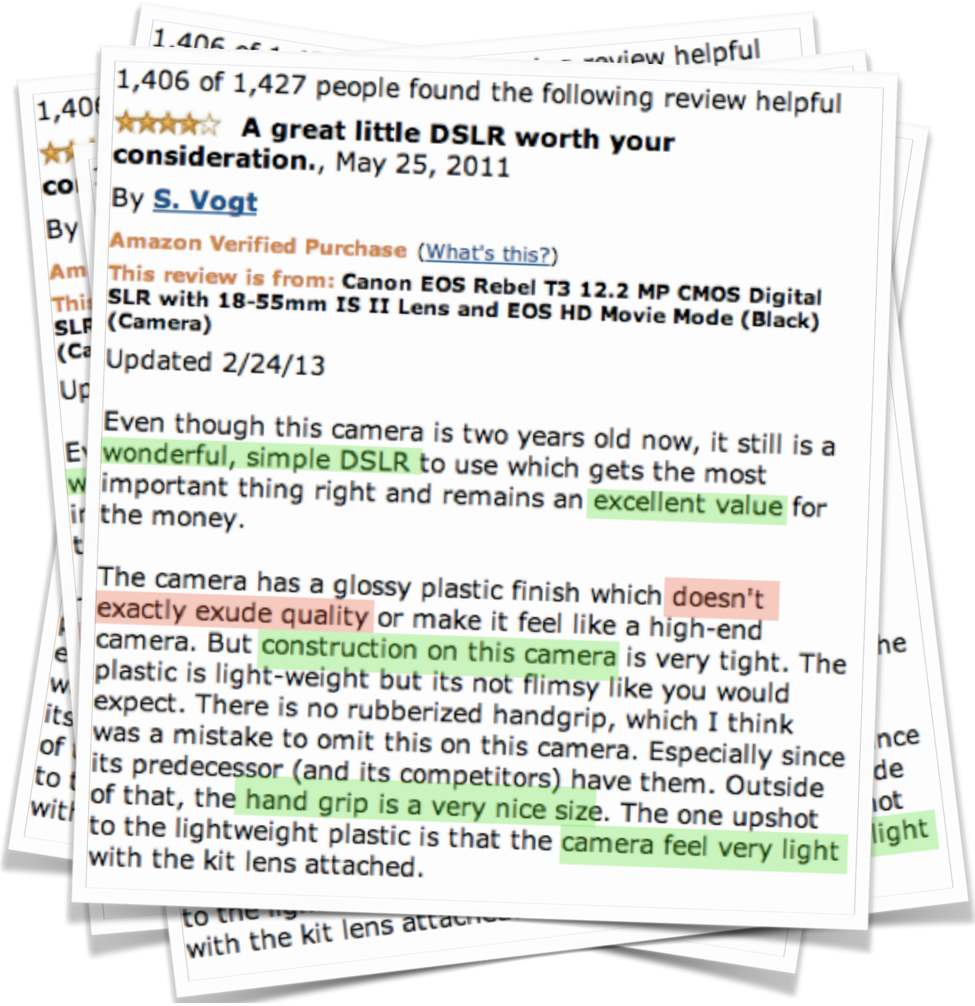
$$Case(P) = \{[F_i, Sent(F_i, P), Pop(F_i, P)] : F_i \in F(P)\}$$

$$Sent(F_i, P) = \frac{Pos(F_i, P) - Neg(F_i, P)}{Pos(F_i, P) + Neg(F_i, P) + Neut(F_i, P)}$$

$$Pop(F_i, P) = \frac{|\{R_k \in Reviews(P) : F_i \in R_k\}|}{|Reviews(P)|}$$

$Pos(Fi, P)$ (or $Neg(Fi, P)$, $Neut(Fi, P)$) denotes the number of positive (or negative, neutral) sentiment labels for feature Fi

From Reviews to Cases



1,406 of 1,427 people found the following review helpful

A great little DSLR worth your consideration., May 25, 2011

By S. Vogt

Amazon Verified Purchase (What's this?)

This review is from: **Canon EOS Rebel T3 12.2 MP CMOS Digital SLR with 18-55mm IS II Lens and EOS HD Movie Mode (Black) (Camera)**

Updated 2/24/13

Even though this camera is two years old now, it still is a wonderful, simple DSLR to use which gets the most important thing right and remains an excellent value for the money.

The camera has a glossy plastic finish which doesn't exactly exude quality or make it feel like a high-end camera. But construction on this camera is very tight. The plastic is light-weight but its not flimsy like you would expect. There is no rubberized handgrip, which I think was a mistake to omit this on this camera. Especially since its predecessor (and its competitors) have them. Outside of that, the hand grip is a very nice size. The one upshot to the lightweight plastic is that the camera feel very light with the kit lens attached.

DSLR: ++
Value: ++
Build Quality: -
Weight: +
Grip: +
Image Quality: +
Resolution: -
Price: +
Battery Life: ---

In other words …

We can generate a new structured, content-based representation for products in which features can be associated with sentiment.

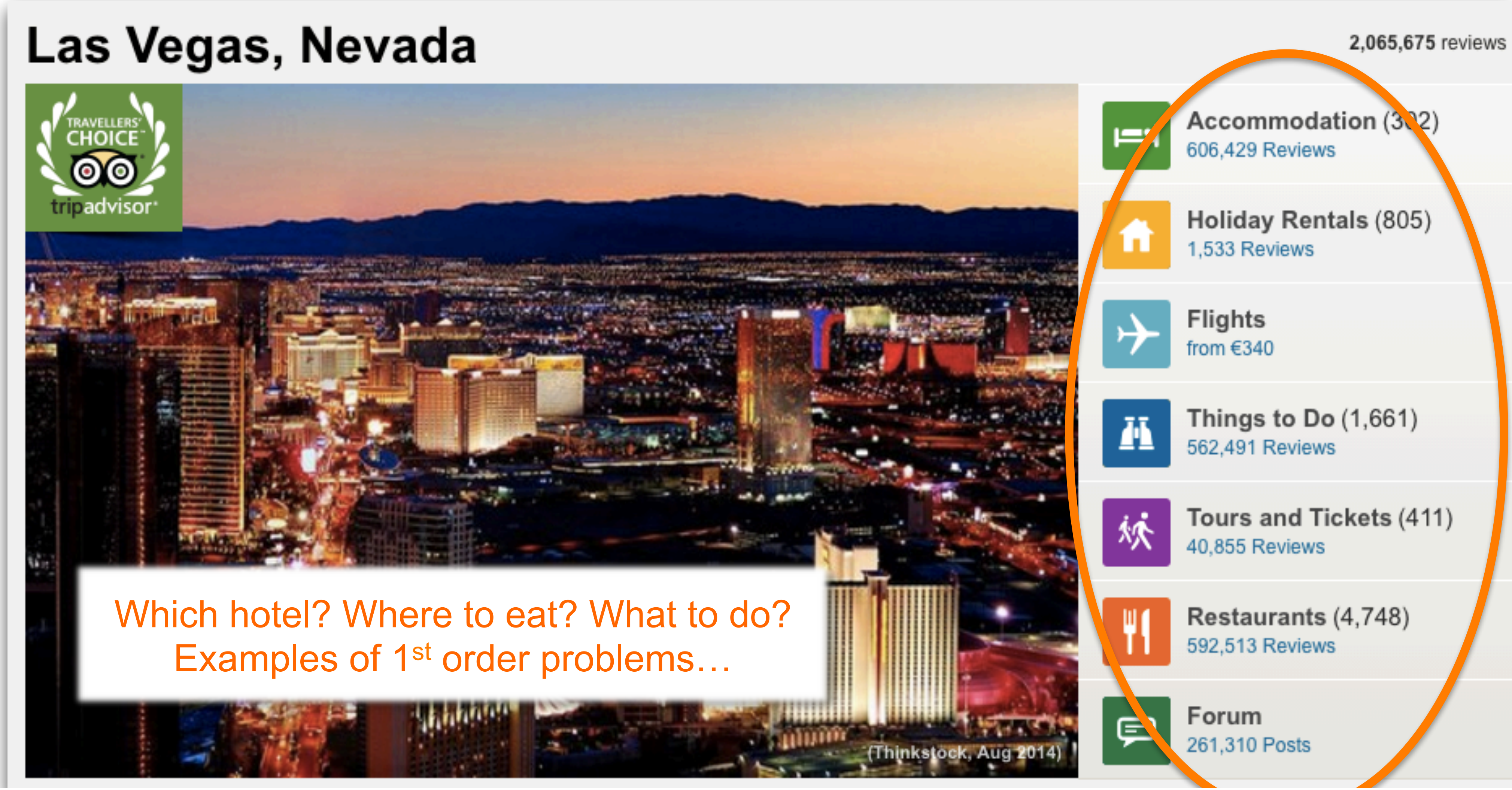
Feasible, as long as there are reviews, and even in the absence of formal metadata.

The availability of sentiment information hints at an alternative approach to computing product “similarity” during recommendation.

Part 3 – Case Studies

From Reviews to Classification

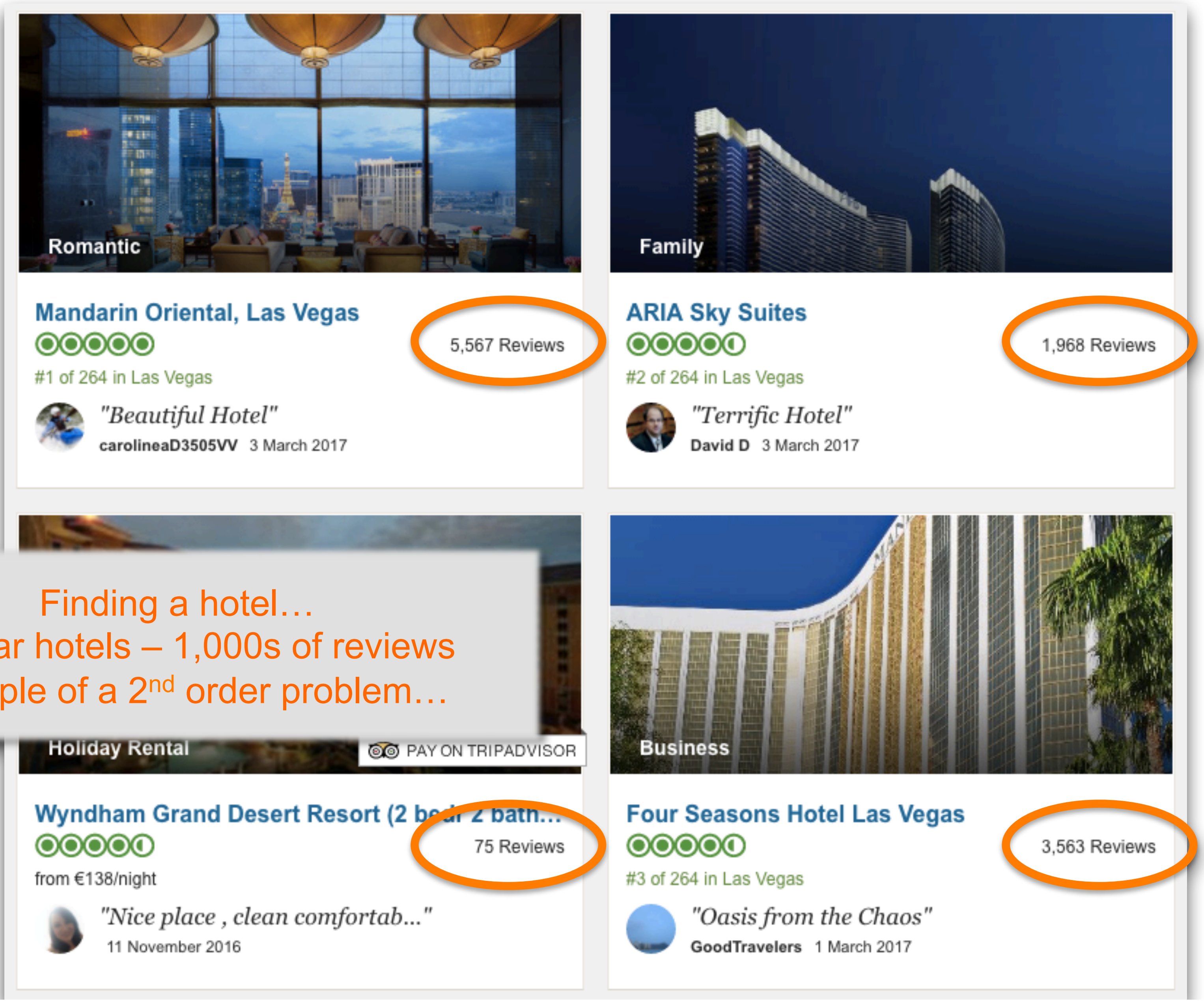
Planning a Trip to Las Vegas…



Finding a hotel…

Popular hotels – 1,000s of reviews

Example of a 2nd order problem…



A Further Challenge…

Popular products attract large numbers of reviews… an additional form of *information overload*

No guarantee that reviews are *informative*, *comprehensive* or free from *bias*

How best to present reviews to consumers?

Ranking Reviews

Reviews (3,626)

Traveller rating

- Excellent 2,882
- Very good 375
- Average 100
- Poor 35
- Terrible 22

Traveller type

- Families
- Couples
- Solo
- Business
- Friends

Time of year

- Mar-May
- Jun-Aug
- Sep-Nov
- Dec-Feb

Language

- All languages
- English (3,414)
- Chinese (Sim.) (79)
- Chinese (Trad.) (79)

More languages

Show reviews that mention

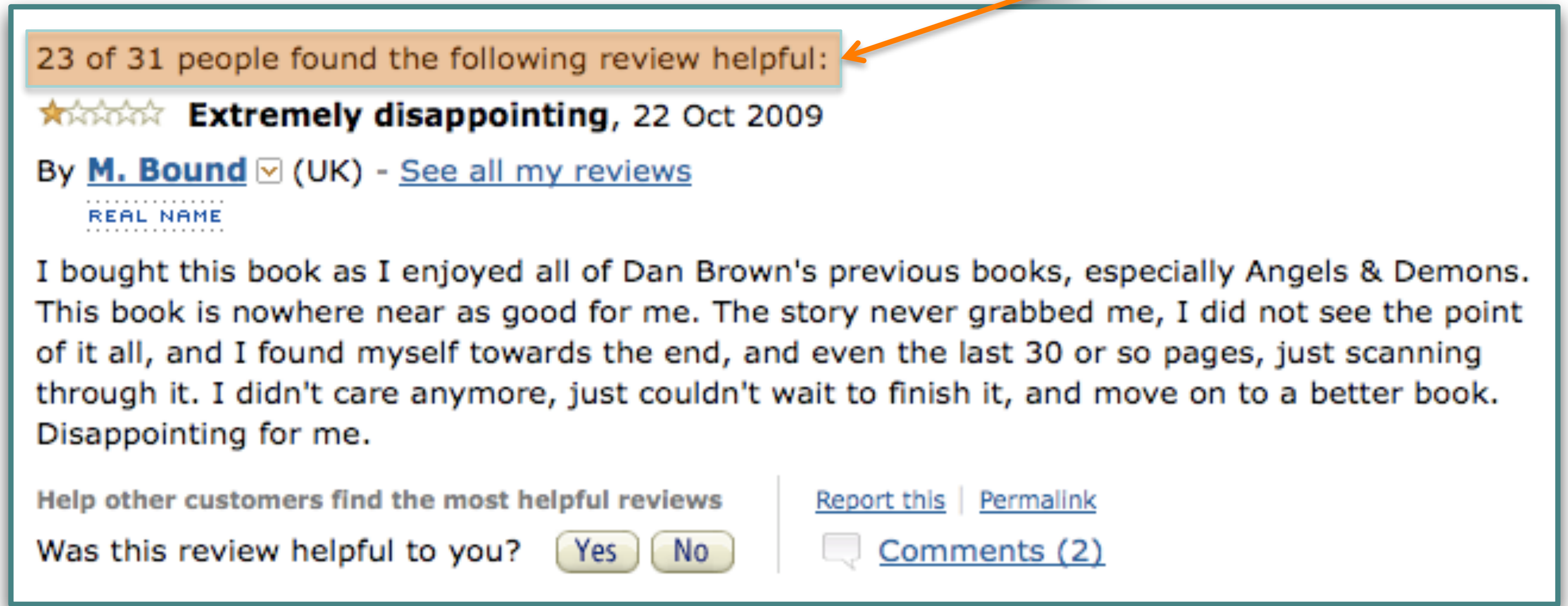
Search reviews

All reviews | cellar restaurant | garden wing | art collection | the main house | art tea | patrick guilbaud | highly recommend the merrion | government buildings | drawing room | public rooms | beautiful hotel | national gallery | turn down service | five star hotel | grafton street | trinity college | meriton

Ranking Reviews

Some sites facilitate *feedback* on *review helpfulness*…

helpfulness = 23/31



Ranking Reviews

Rank reviews based on helpfulness - consider Amazon.com:

Customer Reviews
Flight

137 Reviews

5 star: (47)
4 star: (47)
3 star: (19)
2 star: (13)
1 star: (11)

Average Customer Review
(137 customer reviews)

Share your thoughts with other customers

Create your own review

Search Customer Reviews

GO

Only search this product's reviews

The most helpful favorable review	**The most helpful critical review**
128 of 139 people found the following review helpful	42 of 53 people found the following review helpful
A Contemplative Adult Drama About Addiction Featuring One Of Denzel Washington's Most Complex Performances	**Washington Soars But This Flight Has More Than One Bumpy Landing**
I know that some viewers might feel like there is a bit of a bait-and-switch when sitting down to watch the very adult drama "Flight." I was primed to see a movie about a harrowing plane crash. And indeed, this initiating sequence is staged with tremendous suspense and gut wrenching detail. It may be one of the more memorable scenes you are likely to see this year...	That Denzel Washington is a movie star of the first rank is proven once again in Robert Zemeckis' 2012 crisis-driven character study, a film that starts out very strong but then turns into something of a morality play about addiction and personal accountability. Despite the slowdown in the narrative momentum, the actor lends a chilling honesty and his considerable...
Read the full review ›	**Read the full review ›**
Published 2 months ago by K. Harris	Published 2 months ago by Ed Uyeshima
› See more 5 star, 4 star reviews	› See more 3 star, 2 star, 1 star reviews

Vs.

Review Feedback

Collecting helpfulness ratings to rank and filter reviews can be useful but it takes time to collect the data and it tends to favour older reviews.

Some TripAdvisor statistics:

- Based on 225K reviews by 45K users for 70K hotels
- Some 20% of reviews received no helpfulness scores
- Only 35% of reviews received helpfulness scores on at least 5 occasions

In order to rank reviews by their helpfulness, we first need a method to learn the helpfulness of all reviews for each product…

Can we automatically infer how helpful a review is likely to be by analysing its content?

Related Work

Kim et al (2006) considered features related to structural (review length, #sentences, mean sentence length…), syntactical (#nouns, verbs, adjectives, adverbs in reviews), and semantic properties (#product features & sentiment words in reviews) of reviews…
… to conclude that review length and average review rating were the most predictive of overall helpfulness.

Lui et at (2008) found reviewer expertise to be a useful predictor of helpfulness.

O'Mahony & Smyth (2010) found *readability* features to be good predictors.

Opinion Mining for Review Classification

Can we use features/opinions mined from reviews to improve review classification?

Focus on bi-gram/single-noun features with sentiment as previously described.

Encode reviews as training instances based on 7 different categories of features.

Review Instances

$$\mathit{Instance}(\mathit{Review}_i)=\{\mathit{AGE},\mathit{RAT},\mathit{SIZE},\mathit{READ},\mathit{TOP},\mathit{SENT},\mathit{CN}\}$$

AGE : days since posting
RAT : normalized product rating
SIZE : sentences and word counts
READ : standard readability metrics

TOP : topical coverage (breadth, depth, etc)
SENT : # pos, neg, neutral topics, etc.
CNT: Vector of the top 50 most popular review topics, indicating whether it is present in the review or not

Type	Feature	#	Description
AGE	Age	1	The number of days since the review was posted.
RAT	$NormUserRating$	1	A normalised rating score obtained by scaling the user's rating into the interval $[0, 1]$.
SIZE	$NumSentences$	1	The number of sentences in the review text.
	$NumWords$	1	The total number of words in the review text.
TOP	$Breadth$	1	The total number of topics mined from the review.
	$Depth$	1	The average number of words per sentence containing a mined topic.
	$Redundancy$	1	The total word-count of sentences that are not associated with any mined topic.
	$TopicRank$	1	The sum of the reciprocal popularity ranks for the mined topics present; popularity ranks are calculated across the target product.
SENT	$NumPos\ (Neg, Neutral)$	3	The number of positive, negative, and neutral topics, respectively.
	$Density$	1	The percentage of review topics associated with non-neutral sentiment.
	$NumUPos\ (Neg, Neutral)$	3	The number of *unique* topics with positive/negative/neutral sentiment.
	$WPos\ (Neg, Neutral)$	3	The number of positive, negative, and neutral topics, weighted by their reciprocal popularity rank.
	$RelUPos\ (Neg, Neutral)$	3	The relative proportion of unique positive/negative/neutral topics.
	$SignedRatingDiff$	1	The value of $RelUPos$ minus $NormUserRating$
	$UnsignedRatingDiff$	1	The absolute value of $RelUPos$ minus $NormUserRating$
READ	$NumComplex$	1	The number of 'complex' words (3 or more syllables) in the review text.
	$SyllablesPerWord$	1	The average number of syllables per word
	$WordsPerSen$	1	The average number of words per sentence
	$GunningFogIndex$	1	The number of years of formal education required to understand the review.
	$FleschReadingEase$	1	A standard readability score on a scale from 1 (30 - very difficult) to 100 (70 - easy).
	$KincaidGradeLevel$	1	Translates FleschReadingEase into KincaidGradeLevel required (U.S. grade level).
	$SMOG$	1	Simple Measure of Gobbledygood (SMOG) estimates the years of education required, see [DuBay, 2004].
CNT		50	The top 50 most frequent topics that occur in a particular product's reviews.

Evaluation

51,837 Amazon Reviews, 1,384 products, 4 Product Categories
Digital Cameras, GPS, Laptops and Tablets

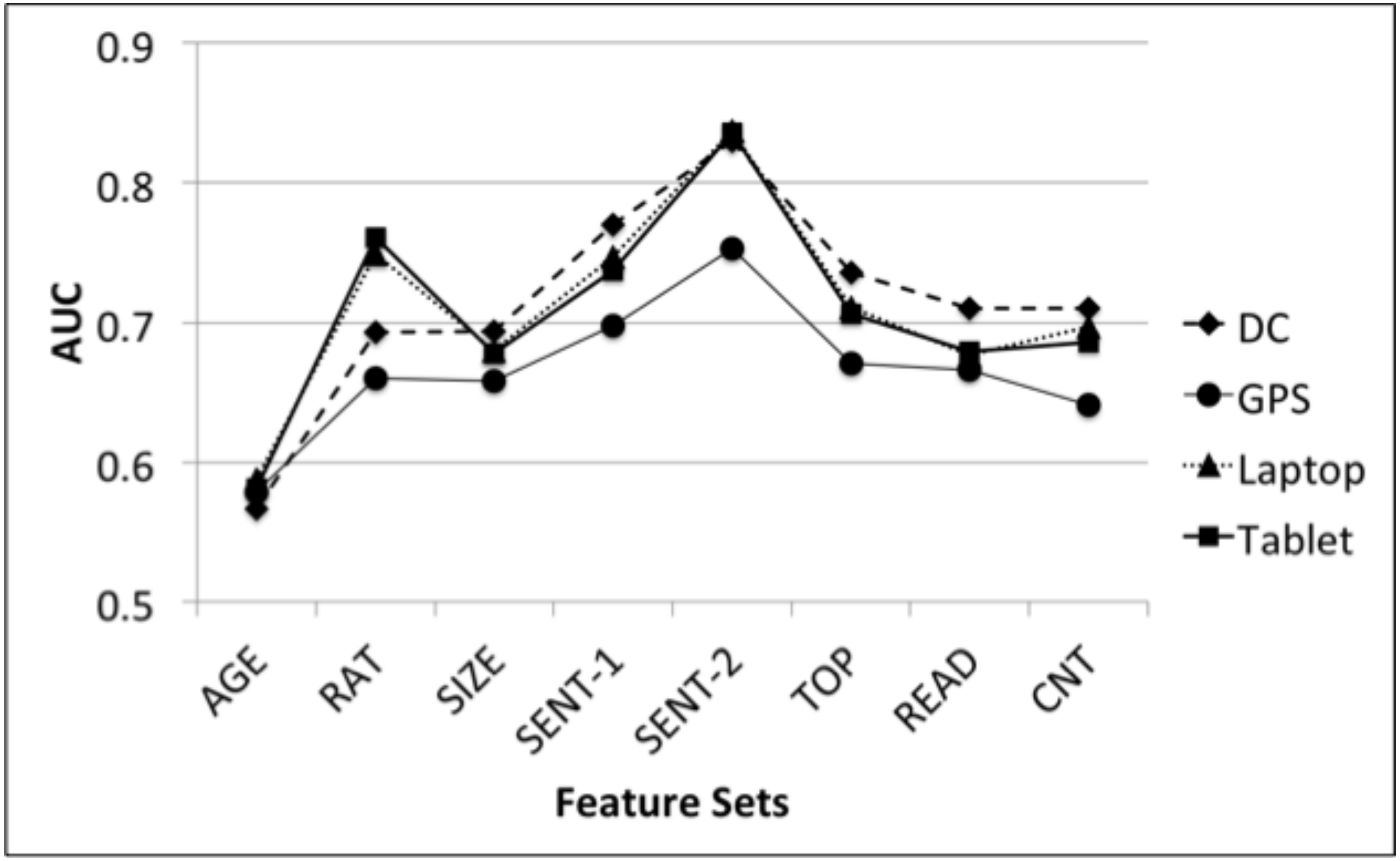
Reviews labeled as helpful/unhelpful based on available ground truth
Review labeled as helpful if >75% of helpfulness scores are positive.
All reviews have >4 helpfulness scores and sampling used to produce balanced training set.

Category	#Reviews	#Prod.	Avg. Helpfulness		
			Help.	Unhelp.	All
DC	3180	113	0.93	0.40	0.66
GPS Devices	2058	151	0.93	0.46	0.69
Laptops	4172	592	0.93	0.40	0.67
Tablets	6652	241	0.92	0.39	0.65

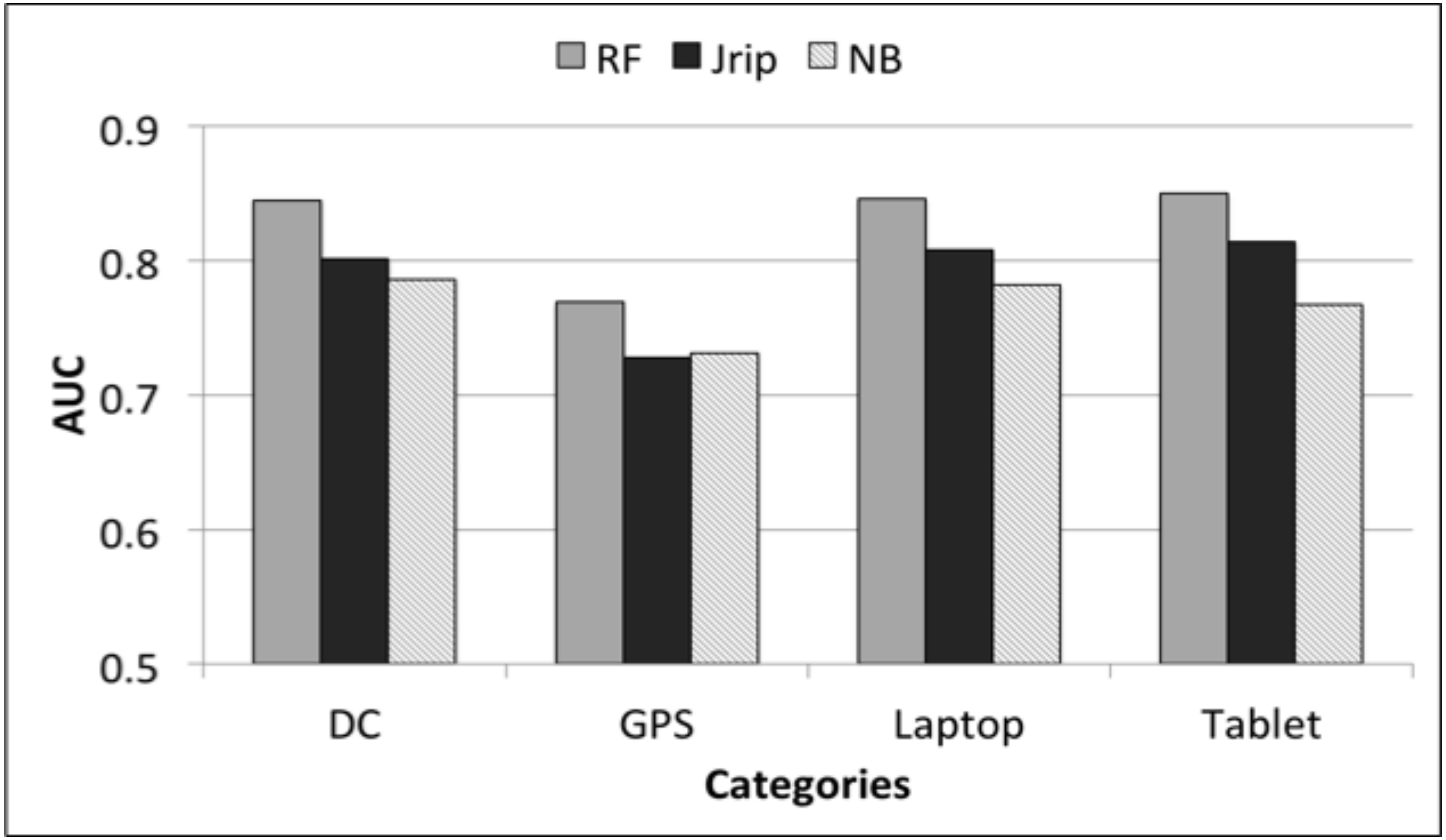
Each review turned into feature based training instance using above features.
Note: SENT-1 = all sentiment features except ratings difference features; SENT-2 = all sentiment features

Compare performance of Random Forest (RF) classifier on different subsets of features using 10-fold cross validation.

Random Forest



All Features



Discussion

AUC < 0.7 (> 0.9) considered poor (excellent) from a classification viewpoint.

Overall RF tends to produce better classification performance across the various feature sets.

Traditional feature types (AGE, SIZE, READ, CNT) perform less well compared to opinion-based groups.

Overall, SENT-2 features provided best performance, followed by SENT-1, RAT/TOP feature sets.

Recommending Helpful Reviews

As per previous example, Amazon presents the most helpful positive and most helpful critical review from a review collection.

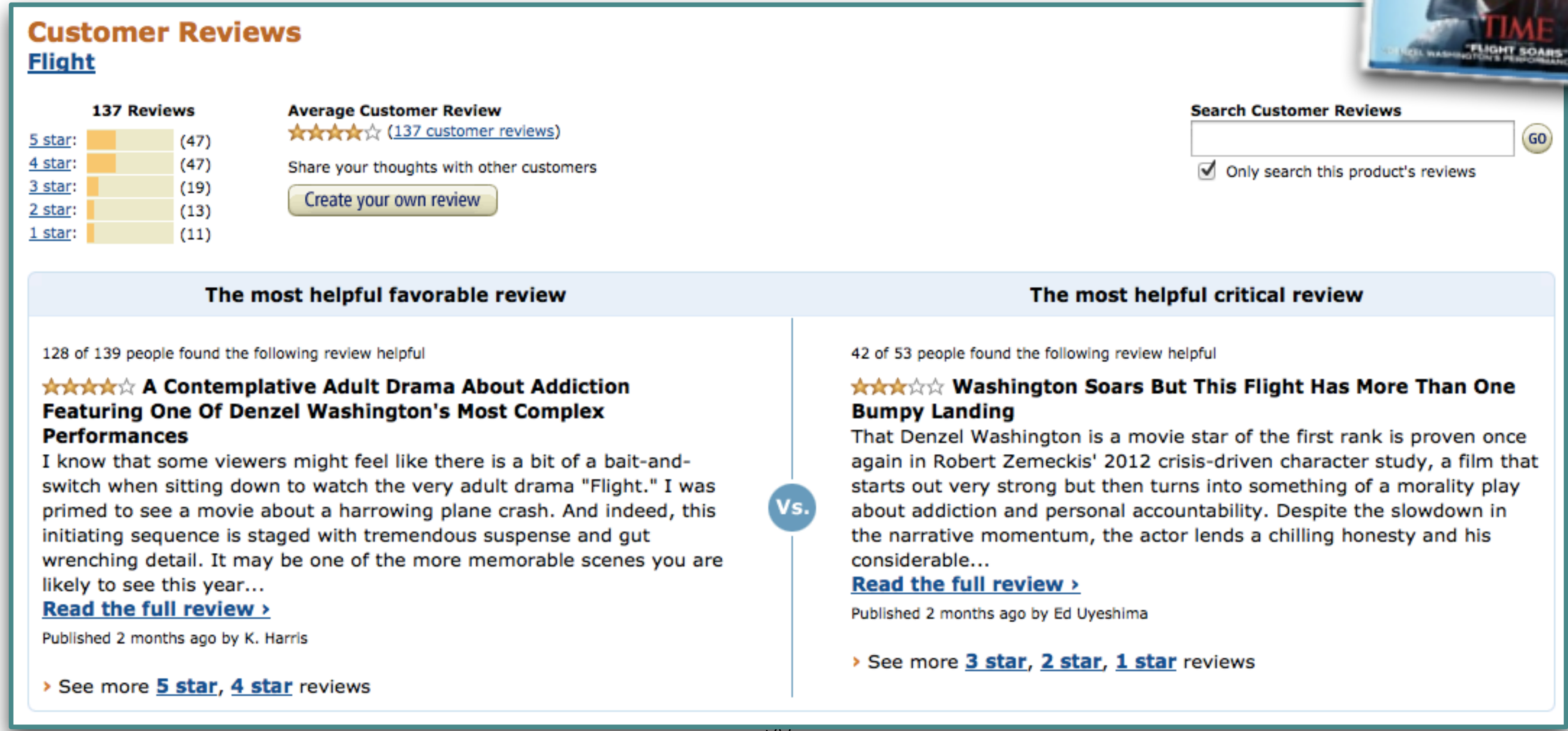
To evaluate the ability of the classification approach to make review recommendations, we use *classification confidence* to rank-order helpful reviews and select the top-ranked review for recommendation to the user.

In this experiment, the single most confident helpful review for each individual product is selected; refer to this strategy as ***Pred***.

Remember this recommendation is made without the presence of actual helpfulness scores and relies only on the ability to predict whether a review will be helpful.

Ranking Reviews

Rank reviews based on helpfulness - consider Amazon.com:



Customer Reviews
Flight

137 Reviews

5 star: (47)
4 star: (47)
3 star: (19)
2 star: (13)
1 star: (11)

Average Customer Review
(137 customer reviews)

Share your thoughts with other customers

Create your own review

Search Customer Reviews

GO

Only search this product's reviews

The most helpful favorable review		The most helpful critical review
128 of 139 people found the following review helpful **A Contemplative Adult Drama About Addiction Featuring One Of Denzel Washington's Most Complex Performances** I know that some viewers might feel like there is a bit of a bait-and-switch when sitting down to watch the very adult drama "Flight." I was primed to see a movie about a harrowing plane crash. And indeed, this initiating sequence is staged with tremendous suspense and gut wrenching detail. It may be one of the more memorable scenes you are likely to see this year... **Read the full review ›** Published 2 months ago by K. Harris › See more 5 star, 4 star reviews	Vs.	42 of 53 people found the following review helpful **Washington Soars But This Flight Has More Than One Bumpy Landing** That Denzel Washington is a movie star of the first rank is proven once again in Robert Zemeckis' 2012 crisis-driven character study, a film that starts out very strong but then turns into something of a morality play about addiction and personal accountability. Despite the slowdown in the narrative momentum, the actor lends a chilling honesty and his considerable... **Read the full review ›** Published 2 months ago by Ed Uyeshima › See more 3 star, 2 star, 1 star reviews

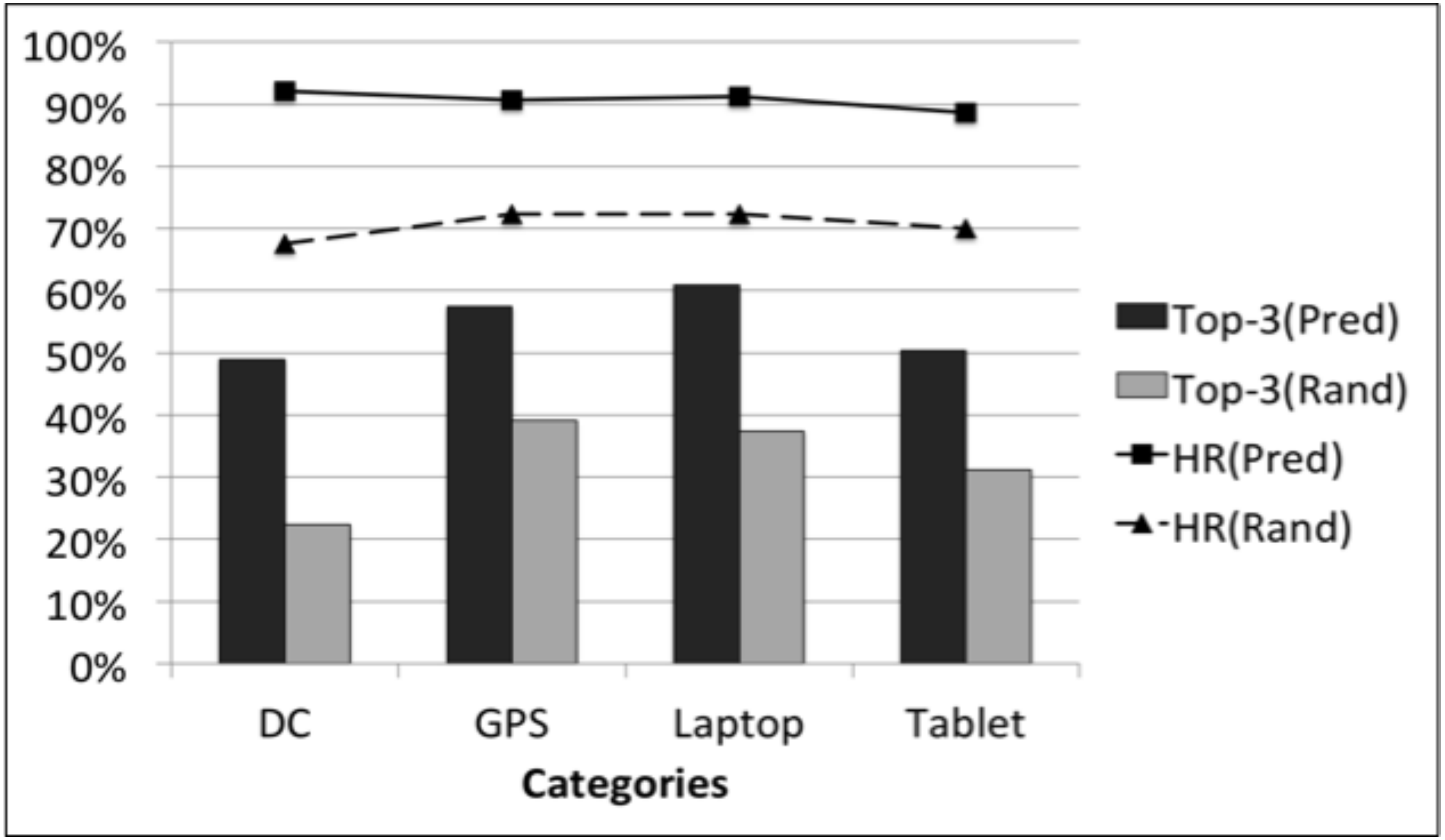
Recommending Helpful Reviews

As a baseline recommendation strategy, a review is also selected at random; this strategy is referred to as ***Rand***.

Compare these strategies to the actual helpfulness score of the most helpful review for a product (optimal helpfulness score)
Compute a helpfulness ratio (HR) – actual helpfulness score of the recommended review to the actual helpfulness score of the most helpful review.

Also compare based on how often one of the (actual) top-3 most helpful reviews are recommended by ***Pred*** and ***Rand***.

Recommendation Results



Discussion

Pred outperforms ***Rand*** wrt helpfulness ratio (approx. 0.9 compared to 0.7).
This means that ***Pred*** is, on average, capable of recommending a review that has a helpfulness score equal to 90% that of the actual most helpful review.

Pred recommends a top-3 review between 1.5 and 2 times as frequently as ***Rand***.

Therefore, the helpfulness classifier can be used to recommend helpful reviews, without the need for explicit helpfulness information…
and that these recommendations are almost as good as the optimal helpful reviews that could be chosen if perfect helpfulness information was available.

This approach has wider applicability: Facebook messages, blog posts…