

Topic Presentation on Review Mining

1. M. Hu and B. Liu. “Mining and Summarizing Customer Reviews”.
2. M. Ganapathibhotla and B. Liu. “Mining Opinions in Comparative Sentences”.

Papers

- M. Hu and B. Liu. “Mining and Summarizing Customer Reviews,” In Proceedings of KDD, 2004. (Primary reference)
- M. Ganapathibhotla and B. Liu. “Mining Opinions in Comparative Sentences,” In Proceedings of COLING, 2008. (Supplemental reference)

Summary of [Hu 2004], “Mining and Summarizing Customer Reviews”

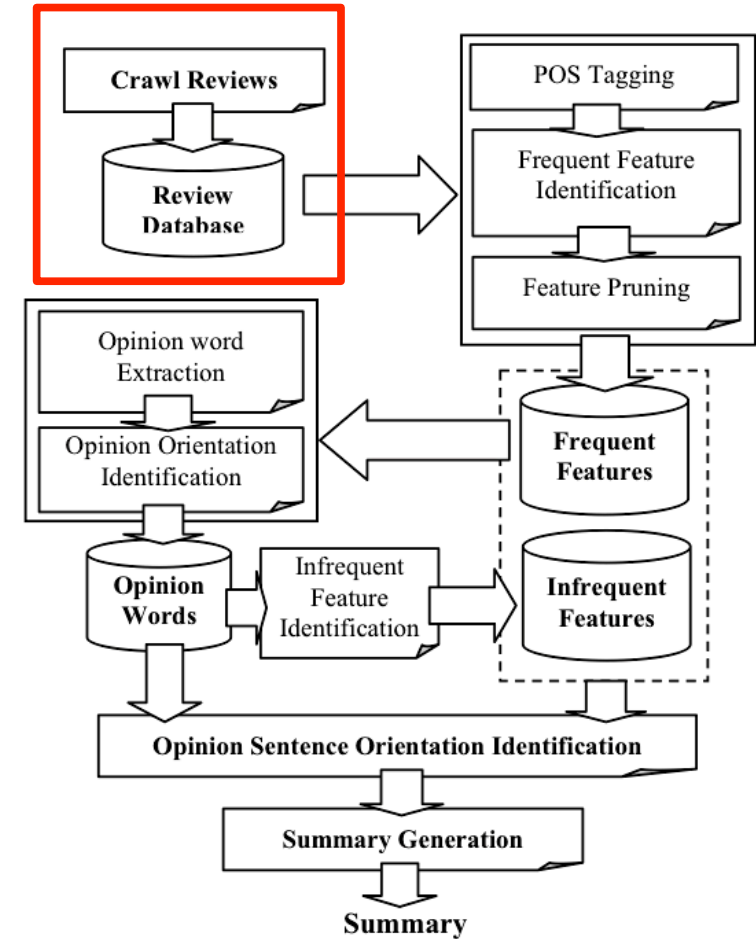
- Problem:
 - Products are sold online (e.g. digital cameras, MP3 player, DVD player)
 - Customers write reviews stating their opinion
 - How can we generate a summary of these customer reviews?
- Approach:
 - Identify product features/attributes/aspects
 - Identify opinion sentences (that contain a product feature and an opinion word)
 - Determine positive or negative orientation of the opinion sentences

```
Digital_camera_1:  
Feature: picture quality  
Positive: 253  
          <individual review sentences>  
Negative: 6  
          <individual review sentences>  
  
Feature: size  
Positive: 134  
          <individual review sentences>  
Negative: 10  
          <individual review sentences>  
  
...
```

Figure 1: An example summary

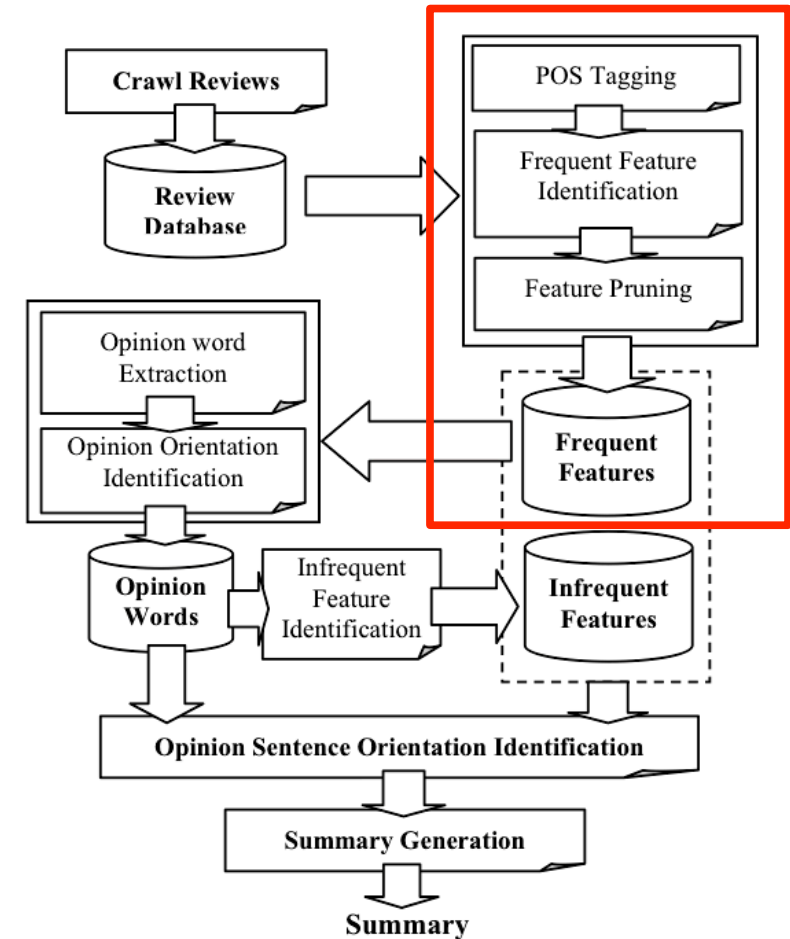
Overview of workflow (1/7)

- Retrieve data
- Experiments used reviews from Amazon.com and Cnet.com



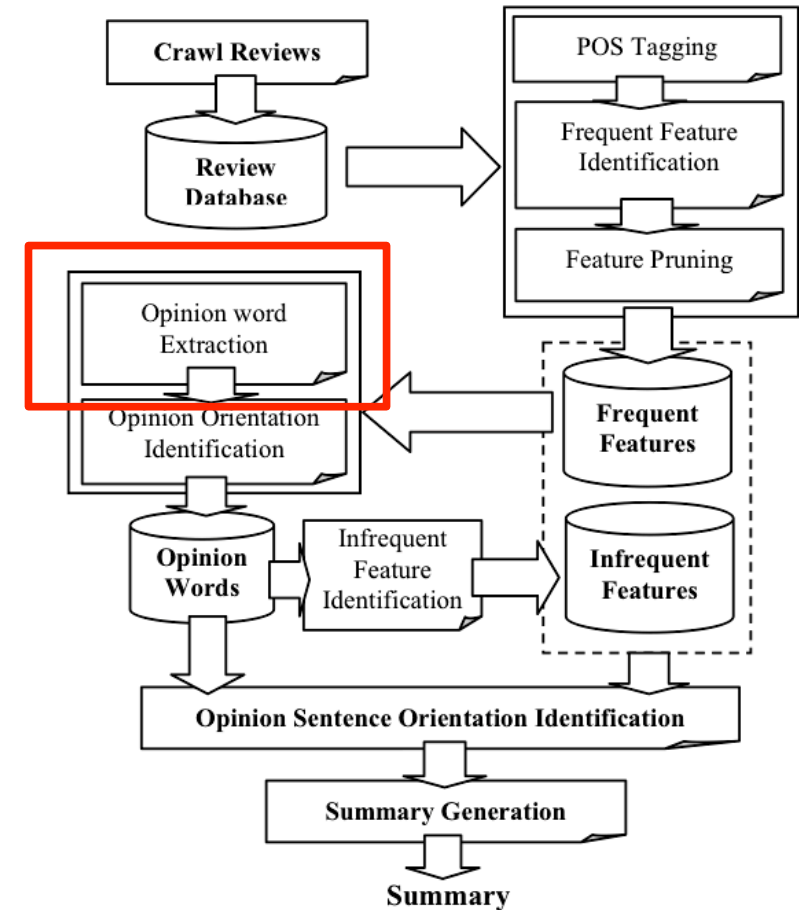
Overview of workflow (2/7)

- POS tagging
 - Find N and NP to identify product features
 - Remove stopwords, apply stemming
- Finding product features in sentences
 - Only get explicitly-mentioned N/NPs
 - *The **pictures** are clear.*
 - *While light, it won't easily fit in pockets.*
 - Frequently occurring N/NPs are likely to be product features
 - Used Apriori algorithm
 - Kept N/NPs that occurred in >1% of review sentences
- Pruning
 - Keep only NPs that occur in some order
 - Remove single N that are also part of larger NP
 - Improves precision



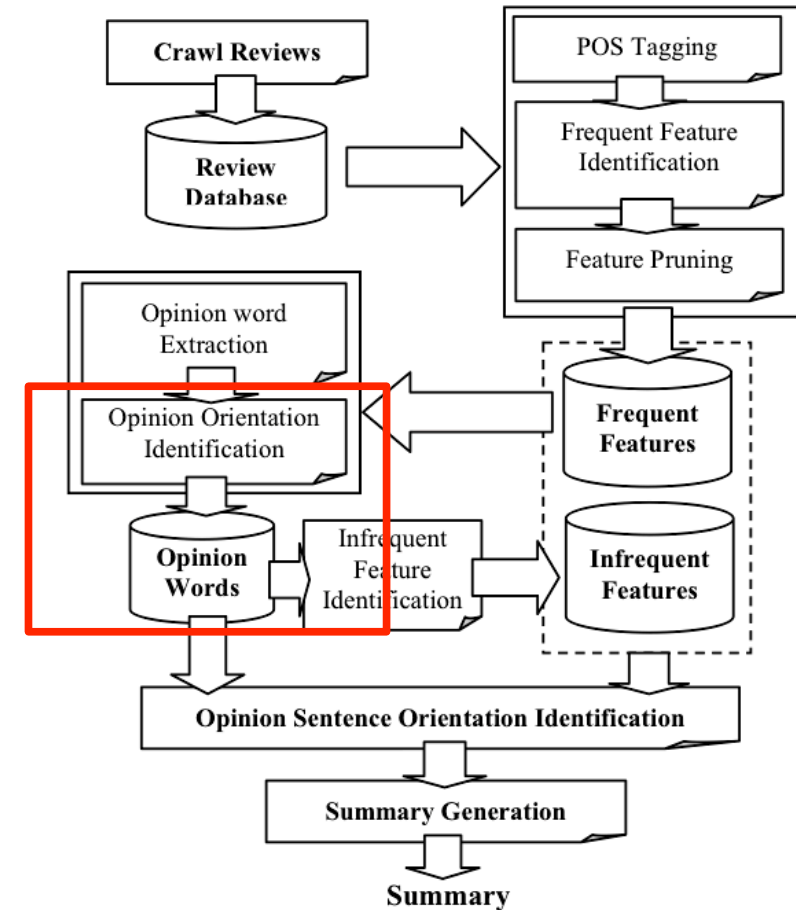
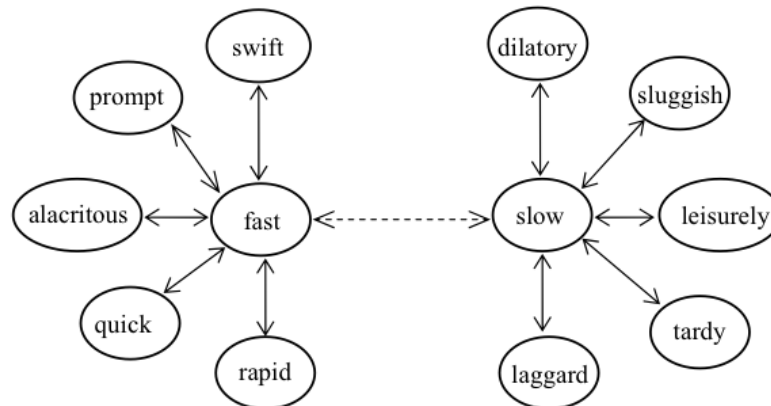
Overview of workflow (3/7)

- Extract opinion words from each opinion sentence
 - An opinion word is the adjective:
 - closest to a product feature (section 3.6)
 - and has a positive or negative orientation
 - An opinion sentence has 1+ product feature and 1+ opinion word
 - Example: *The **strap** is **horrible** and gets in the way of parts of the camera you need access to.*



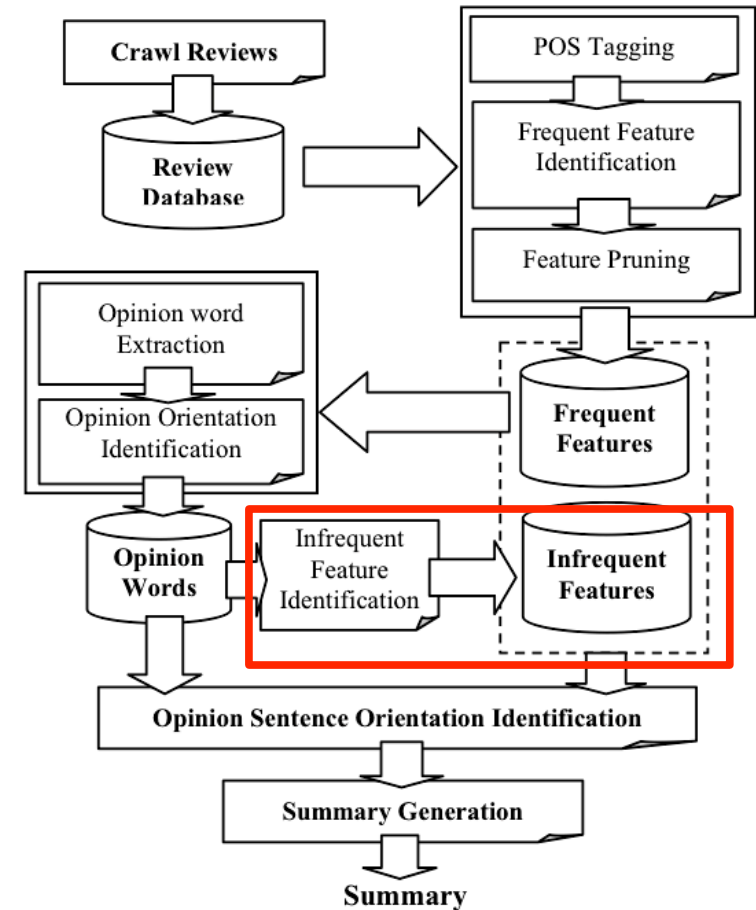
Overview of workflow (4/7)

- Identify the positive or negative orientation of opinion words (adjectives)
 - Applied bootstrapping to grow list of oriented words
 - Started with 30 seed words that are clearly positive or negative (e.g. great, fantastic, bad, dull)
 - For a given opinion word:
 - use Wordnet to find a related opinionated adjective
 - assign same orientation
 - add to growing list of oriented words



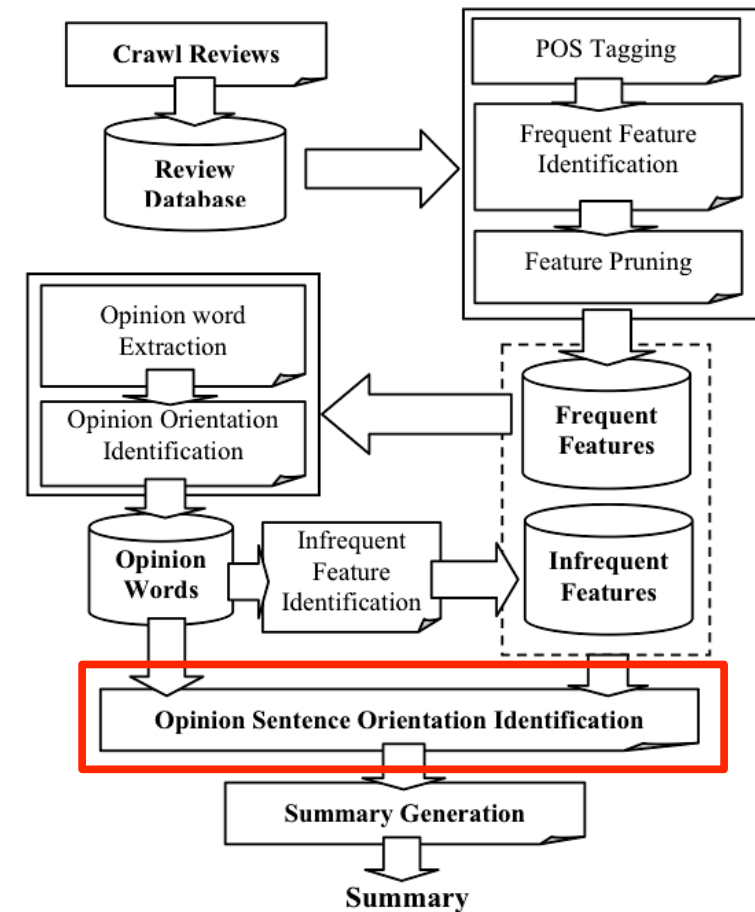
Overview of workflow (5/7)

- Infrequent features
 - Find features that did not pass Apriori's minimum support (occurred in less than 1% of review sentences)
 - For a sentence with known opinion words (adjectives) but no known frequent features, pick the closest N/NP closest to the opinion word
 - Improves recall



Overview of workflow (6/7)

- Identify the positive or negative orientation of opinion sentences
 - Count the number of positive and negative opinion words
 - The orientation with the higher count wins and is assigned to the sentence
 - Tie breaking rules, such as:
 - Use average orientation for feature(s)
 - Use orientation of previous sentence
 - Rules for handling *but/however* clause
 - Negate orientation of opinion word if negation word occurs within 5 words



Overview of workflow (7/7)

- Generate summary

12 opinion sentences related to **picture** with positive orientation

Feature: picture

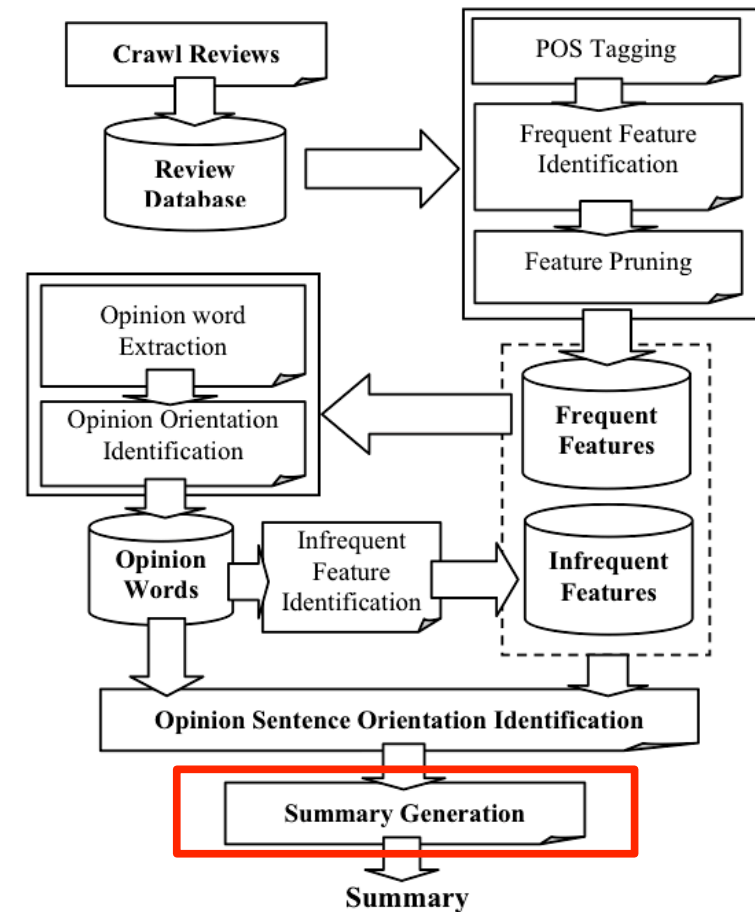
Positive: 12

- Overall this is a good camera with a really good picture clarity.
- The pictures are absolutely amazing - the camera captures the minutest of details.
- After nearly 800 pictures I have found that this camera takes incredible pictures.

...

Negative: 2

- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.



Experimental results

Table 1: Recall and precision at each step of feature generation

Product name	No. of manual features	Frequent features (association mining)		Compactness pruning		P-support pruning		Infrequent feature identification	
		Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Digital camera1	79	0.671	0.552	0.658	0.634	0.658	0.825	0.822	0.747
Digital camera2	96	0.594	0.594	0.594	0.679	0.594	0.781	0.792	0.710
Cellular phone	67	0.731	0.563	0.716	0.676	0.716	0.828	0.761	0.718
Mp3 player	57	0.652	0.573	0.652	0.683	0.652	0.754	0.818	0.692
DVD player	49	0.754	0.531	0.754	0.634	0.754	0.765	0.797	0.743
Average	69	0.68	0.56	0.67	0.66	0.67	0.79	0.80	0.72

- For each product, authors created list of features by manual inspection
- Recall and precision for features found by algorithm at each stage

Experimental results

Table 3: Results of opinion sentence extraction and sentence orientation prediction

Product name	Opinion sentence extraction		Sentence orientation accuracy
	Recall	Precision	
Digital camera1	0.719	0.643	0.927
Digital camera2	0.634	0.554	0.946
Cellular phone	0.675	0.815	0.764
Mp3 player	0.784	0.589	0.842
DVD player	0.653	0.607	0.730
Average	0.693	0.642	0.842

- Authors manually identified opinion sentences and determined orientation
- Recall and precision of determining if a sentence is an opinion sentence
- Accuracy of orientation

Critiques

- Hard to critique
 - Paper published in 2004
 - 1992 citations on Google Scholar
 - Almost 200 citations per year

Critiques

- The output is a summary of individual features from individual sentences, not of the overall opinion
 - Do customers like the product or not?
- No context
 - *The laptop had a fast hard drive.*
 - *The fast hard drive nonetheless produced a lot of noise.*
- Context is discussed further in [Ganapathibhotla 2008]

```
Digital_camera_1:  
Feature: picture quality  
Positive: 253  
          <individual review sentences>  
Negative: 6  
          <individual review sentences>  
  
Feature: size  
Positive: 134  
          <individual review sentences>  
Negative: 10  
          <individual review sentences>  
...
```

Figure 1: An example summary

Critiques

- No reference resolution – losing a lot of information
 - *The lens turns quickly. It is easy to use.*
 - *The former was great, but the latter was horrible.*
- How to handle mixed products in the review?
 - Example: *I liked using CameraX. I previously owned CameraY. The **lens** was **horrible** on that one.*

```
for each sentence in the review database
  if (it contains a frequent feature, extract all the adjective
      words as opinion words)
    for each feature in the sentence
      the nearby adjective is recorded as its effective
        opinion. /* A nearby adjective refers to the adjacent
        adjective that modifies the noun/noun phrase that is a
        frequent feature. */
```

Figure 3: Opinion word extraction

Critiques

- Lots of heuristics, section 3.6
- Regarding negation:
 - *“By ‘closely’ we mean that the word distance between a negation word and the opinion word should not exceed a threshold (in our experiment, we set it to 5).”*
- Regarding subordinate clauses:
 - *“For a sentence that contains a ‘but’ clause ... we first use the effective opinion in the clause to decide the orientation of the features. If no opinion appears in the clause, the opposite orientation of the sentence will be used.”*
- Regarding adjectives and nouns:
 - *“Effective opinion is the closest opinion word for a feature in an opinion sentence.”*

Critiques

- Assumed reviews are legitimate
 - Promotional spam / astroturfing
 - Malicious trolling
 - Sarcasm

Things I learned

- Identifying product features by (1) finding N/NP and (2) applying a minimum threshold
 - Seems to work well in practice
 - A. Popsecu and O. Etzioni, “Extracting product features and opinions from reviews.” In Proceedings Empirical Methods and in NLP, 2005.
 - R. Feldman, “Techniques and Applications for Sentiment Analysis,” CACM, April 2013.
- Adjectives as sentiment-bearing tokens
 - As discussed in 4/15 lecture
 - Growing a list of opinionated adjectives by searching Wordnet for synonyms and antonyms is useful
 - Some opinionated adjectives are unambiguous – or are they?

Summary of [Ganapathibhotla 2008], “Mining Opinions in Comparative Sentences”

- Problem:
 - Customers write reviews stating their opinion, often with sentences comparing two products
 - How can we identify which product the user prefers?
- Approach:
 - Clearly opinionated comparatives are easy
 - *CameraX is **better** than CameraY.*
 - *CameraX is **more** beautiful than CameraY.* (where *beautiful* is a known positive adjective from Hu’s paper)
 - Context-dependent comparatives are hard
 - *CameraX has a **higher** build quality than CameraY.*
 - *CameraX has a **higher** failure rate than CameraY.*
 - Check external source (e.g. epinion.com’s Pros and Cons lists) to see if *higher build quality* and *higher failure rate* are positive or negative

Applications

The screenshot displays the Amazon.com interface for a PlayStation 4 Console. At the top, the Amazon Prime logo is on the left, and navigation links for 'Thomas's Amazon.com', 'Today's Deals', 'Gift Cards', 'Sell', and 'Help' are in the center. On the right, there is a 'Spring Cleaning' banner with a flower icon and a 'Shop now' link, and a circular icon. Below the header, a search bar contains 'Video Games' and a 'Go' button. User account information shows 'Hello, Thomas', 'Your Account', 'Your Prime', a shopping cart with '0' items, and a 'Wish List'. A horizontal menu lists various categories: Video Games, Fire TV, Xbox One, Xbox 360, PS4, PS3, Wii U, Wii, 3DS, PS Vita, Digital Games, Kindle Fire Games, Deals, and Best Sellers.

The main product area features a large image of the PlayStation 4 console and its DualShock 4 controller. To the left of the main image is a vertical strip of smaller thumbnail images. The product title is 'PlayStation 4 Console' by Sony. Below the title, it specifies the platform as 'PlayStation 4' and is rated 'Kids to Adults'. Customer feedback is shown as a 4.5-star rating from 6,872 reviews and 27 answered questions. Pricing information includes a list price of \$399.99, a current price of \$399.00 with a Prime logo, and a 'You Save' of \$0.99. The status is 'In Stock' and it is shipped and sold by Amazon.com. A 'Gift-wrap available' option is also present.

Below the main product image, there are four bundle options: 'Battlefield 4 and PS Plus Launch Bundle', 'DS4 Launch Bundle', 'Killzone and PS Plus Launch Bundle', and 'Knack Launch Bundle'. The 'PlayStation 4' bundle option is highlighted with an orange border.

On the right side of the product page, there are several action buttons: 'Add to Cart' with a shopping cart icon, 'or 1-Click Checkout', 'Buy now with 1-Click' with a mouse cursor icon, and 'Order within 5hr 5min'. Delivery options show 'Get it:' with 'Wed +7.99' and 'Thu Free' buttons. The shipping location is set to 'Thomas Phan - San Jose' with a dropdown arrow, and there is a checkbox for 'This is a gift'. Below these are 'Add to Wish List' and 'Sell Us Your Item' sections. The 'Sell Us Your Item' section offers up to a \$219.48 gift card and includes a 'Trade in' button and a 'Learn more' link. At the bottom right, a 'More Buying Choices' section shows '370 used & new' items starting from \$339.00.

Applications

If I were an algorithm, I would look for this

amazon Prime
Thomas's Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search Video Games

Video Games Fire TV Xbox One Xbox 360 PS4 PS3 Wii

Customer Reviews

★★★★☆ (6,872)
4.1 out of 5 stars

5 star	4,637
4 star	621
3 star	232
2 star	125
1 star	1,257

[See all 6,872 customer reviews](#)

Most Helpful Customer Reviews

695 of 820 people found the following review helpful

★★★★★ **PS4 launch day console review! So far I'm Impressed.**
November 22, 2013
By tech
Edition: PlayStation 4 | **Verified Purchase**

I purchased the launch version at 12:15 Friday night of the 15th. I'm glad I did. Got up the next morning and looked online and they were already sold out.

“ Been playing Killzone and Assassins Creed 4 plus the free games for Playstation Plus like Resogun.”
J. M. Castaneda | 1,298 reviewers made a similar statement

“ My PS4 is working fine but people should know by now that if you buy one of the first systems there is a good chance there will be some sort of issues.”
Amber Brown | 430 reviewers made a similar statement

“ The system is very fast and responsive with a clean interface/OS that is easy to use.”
Johnny | 618 reviewers made a similar statement

Applications

Product features

The screenshot shows an Amazon product page for a Canon PowerShot camera. The page includes a search bar with the text "canon powershot", a star rating of 4.2 out of 5 stars based on 271 reviews, and a "Popular Discussion Topics" section. A red box highlights the "Image Quality" topic, which has 161 reviews. Below this, several customer reviews are displayed, many of which mention "picture quality" or "pictures".

Customer Reviews
4.2 out of 5 stars (271)

Popular Discussion Topics beta: what do you think?
"Image Quality" 161 | "Ease of Use" 59 | "Value" 23 | "Zoom" 22 | "Features" 21 | All Topics

Most Helpful Customer Reviews
423 of 428 people found the following review helpful
★★★★★ **As an Amateur Photographer, I Love It**
By [Laura](#) on March 14, 2013
Color Name: Gray
I'll note right off the bat – I've been interested in photography for many years and have been honing my skills, but I am not a professional photographer. As a result, I can only speak to the qualities of

Thank you for staying awake