

# Mining and Summarizing Customer Reviews

By Minqing Hu and  
Bing Liu

Presented by Peng Zang

# Motivation

- Too many reviews to read
- Need good sampling of reviews
- Traditional summarization techniques inadequate

# Example Result:

## Digital\_camera\_1:

Feature: *picture quality*

Positive: 253

“Wonderful picture quality”

“Love the quality of the pictures”

...

Negative: 6

“Picture quality not up to par, especially with the price.”

...

Feature: *size*

Positive: 134

“Amazing, it can fit in the palm of your hand”

...

Negative: 10

“Sized too small to be easily usable”

...

...

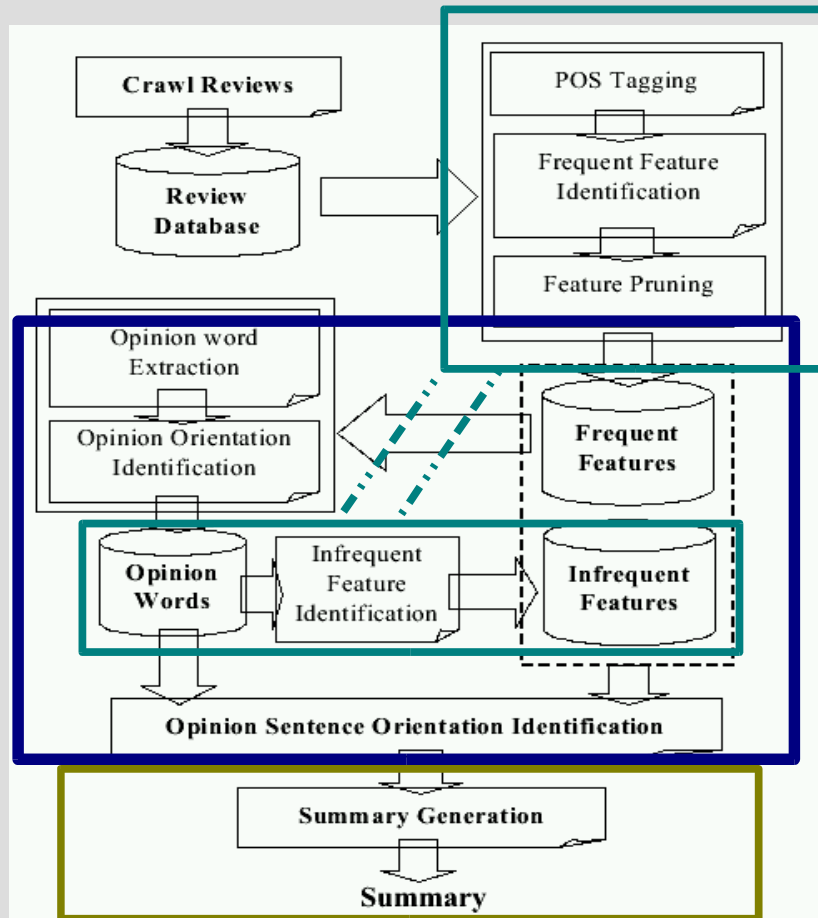
# New Technique – Basic Idea

- Identify product features
- Summarize each feature
- Produce overall summary

# Three Simple Steps

- Identify product features
- Categorize opinion sentences into POS/NEG
- Summarize result

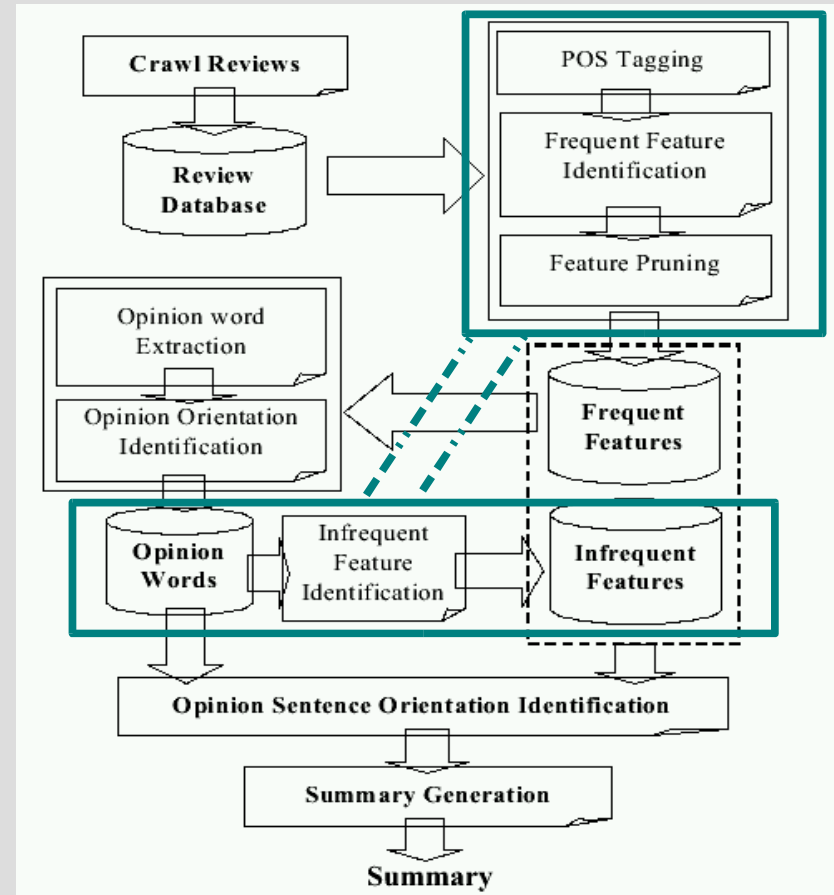
# System Architecture



- 1) Identify product features
- 2) Categorize opinion sentences into **POS/NEG**
- 3) Summarize result

# Feature Identification, key insight

“customer review[s] contain many things that are not directly related to product features... However, when they comment on product features, the words that they use converge”

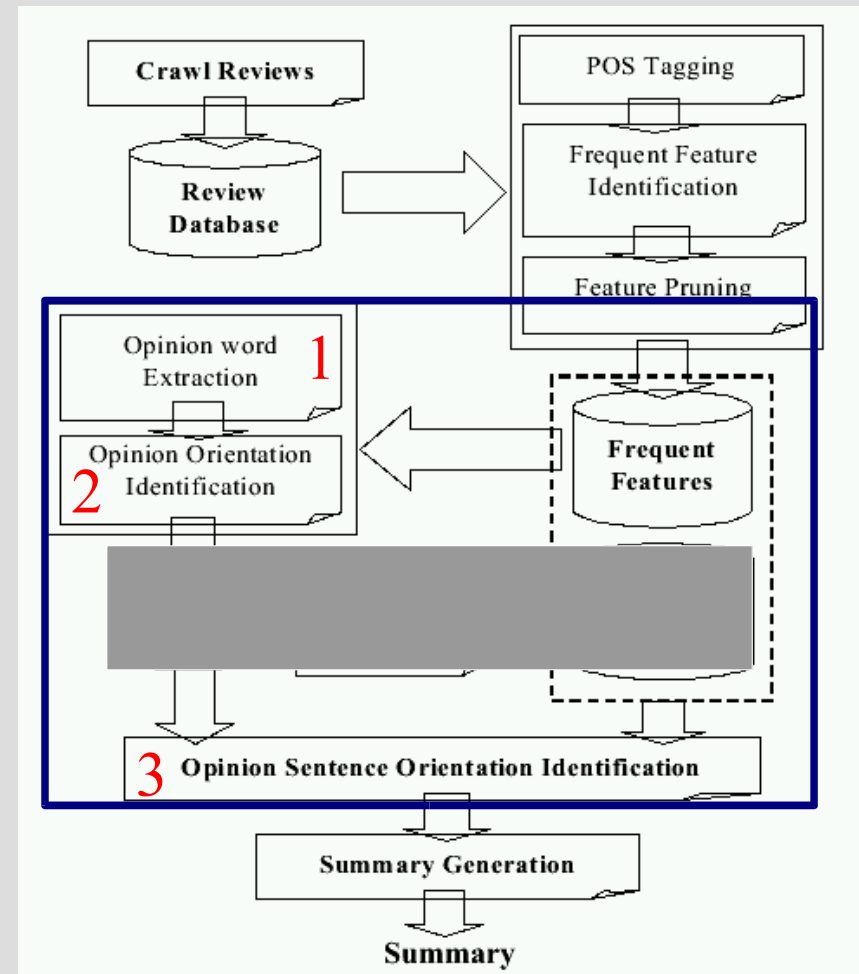


# Feature Identification (cont.)

- Association mining to find frequent itemsets
- Compactness pruning
- Redundancy pruning
- Misses infrequent features... so later, identify those via opinion words and add to feature list.

# Categorize Opinion Sentences

- 1) Adjectives from sentences containing features.
- 2) Identification derived from synonym /antonym.
  - 1) Initial seeds manually identified.
  - 2) Expanded with WordNet.
- 3) Sentence orientation based on sum of word orientations.



# Finally, Summary Generation

## Digital\_camera\_1:

Feature: *picture quality*

Positive: 253

“Wonderful picture quality”

“Love the quality of the pictures”

...

Negative: 6

“Picture quality not up to par, especially with the price.”

...

Feature: *size*

Positive: 134

“Amazing, it can fit in the palm of your hand”

...

Negative: 10

“Sized too small to be easily usable”

...

...

# Experimental Results

- Feature extraction: (R , P)
  - Association mining (0.68, 0.56)
  - Compactness pruning (0.67, 0.66)
  - P-support pruning (0.67, 0.79)
  - Infrequent feature addition (0.80, 0.72)
  - Far better than FASTR: (0.17, 0.03)
- Opinion sentence extraction (0.69, 0.64)
- Sentence accuracy: 84.2%

# Caveats

- Limited features
  - Only nouns allowed
  - must be explicitly mentioned, eg. can't process “While light, it will not easily fit in pockets.”
- Minimum support and p-support setting
- Synonym searching can be misleading
  - eg. Cool is a positive word, but a synonym might be aloof or unfriendly.
- Can't deal with sentences like:  
“Picture is good, but lens too big.”

# Questions

- Seems like good system, but any broad impacts?
- What's the big picture?
- What can we take away?