Mining Reviews for Product Comparison and Recommendation

Dr. Minlie Huang

Dept. Computer Science, Tsinghua University
aihuang@tsinghua.edu.cn

Outline

- Background and Motivation
- System Architecture
- Product Comparison
  - Subjective perspective
  - Objective perspective
  - Overall perspective
- Product Recommendation
  - Evolution Tree
- Experiments
- System Demo

Background

- Web 2.0 times
  - Blogs, RSS, VSN, Forums, …
- Ecommerce
  - Amazon, Alibaba, …
- Express opinions, share opir

  - Epinions: www.epinions.com
  - IMDB: movie reviews
  - Many forums associate with specific products

Amazon

Customer Reviews

Canon Digital Rebel XSi 12.2 MP Digital SLR Camera with EF-S 18-

197 Reviews

5 star: (158)

Average Customer Review

(197 customer reviews)

re your thoughts with other customers

eate your own review

Epinions

Apple iPhone (8 GB) Smartphone

Overall Rating:

57 consumer reviews

Lowest Price: $549.00

at eBay SEE IT

Subscribe to reviews on this product

Compare Prices | Read Reviews (57) | View Details | Write a Review

Showing 1-15 of 57 reviews

Sort by Product Rating | Sort by Review Date

Product Rating:

Price Range: $549 - $700

COMPARE PRICES

Cohabitating With An Apple iPhone Addict

by Freak369 POPULAR AUTHOR - Top 10, Jul 01 '07

Pros: No-glare screen, internet browsing, built-in iPoc

Cons: Can take a while to get used to the touch scre

For months I have been seeing adverts for the Apple i iPhone owners. Two weeks ago the bedlam began; T

Read the full review

Product Rating:

Price Range: $549 - $700

COMPARE PRICES

Welcome to the i-Revolution !

by aashok TOP REVIEWER in Electronics, POPULAR

Pros: Museum quality design, large screen, virtual ke

Cons: Runs on dinky ATT wireless, no copy/cut/past

WHAT IS IT & WHY DID I BUY IT ? After all the medi heard the release date of 06/29. It is darn sure to cat

Read the full review

Product Rating:

Price Range: $549 - $700

COMPARE PRICES

The Regular Chick and *The Apple iPhone*.

by njchicaa POPULAR AUTHOR - Top 1000, Feb 06 '08

Pros: easy to use, does the job of 3 different devices

Cons: expensive, internet and typing could use some i

I am no technophile. I can use my laptop to surf the ne to make calls and send text messages. Don't ask me

Read the full review

IMDb

The Internet Movie Database

NOW PLAYING | MOVIE / TV NEWS | MY MOVIES | DVD & BLU-RAY

Home | Top Movies | Photos | Independe

search All

IMDb > Dead Poets Society (1989) > IMDb user comments

Watch It

Own the rights?

or

IMDb user comments for

Dead Poets Society (1989)

Filter: Best Hide Spoilers:

Page 1 of 37: [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11]

Index 363 comments in total

183 out of 199 people found the following comment useful :-

A powerful antidote to conformism, 22 August 2001

Author: Francisco Huerta (fjhuerta@hotmail.com) from Mexico

Motivation

- Product Reviews
  - Users may concern very specific features of a product , such as
    - The picture of a digital camera
    - The weight of a mobile phone
  - There is no comparison or recommer system (based on reviews) for produc

- Users have to read through all the product reviews to:
  - catch contents with mentions of user experience
  - seek product comparison or recommendation suggestions

Our System

- Present comparison and recommendation at the feature level
  - For example:
    - “Whose screen is better, Sony T200 or Canon G9?”

>

    - ”What products can be recommended whose screen is as good as Sony T200’s?”

Canon SD1000

Canon SD750

Fujifilm S5200

System Architecture

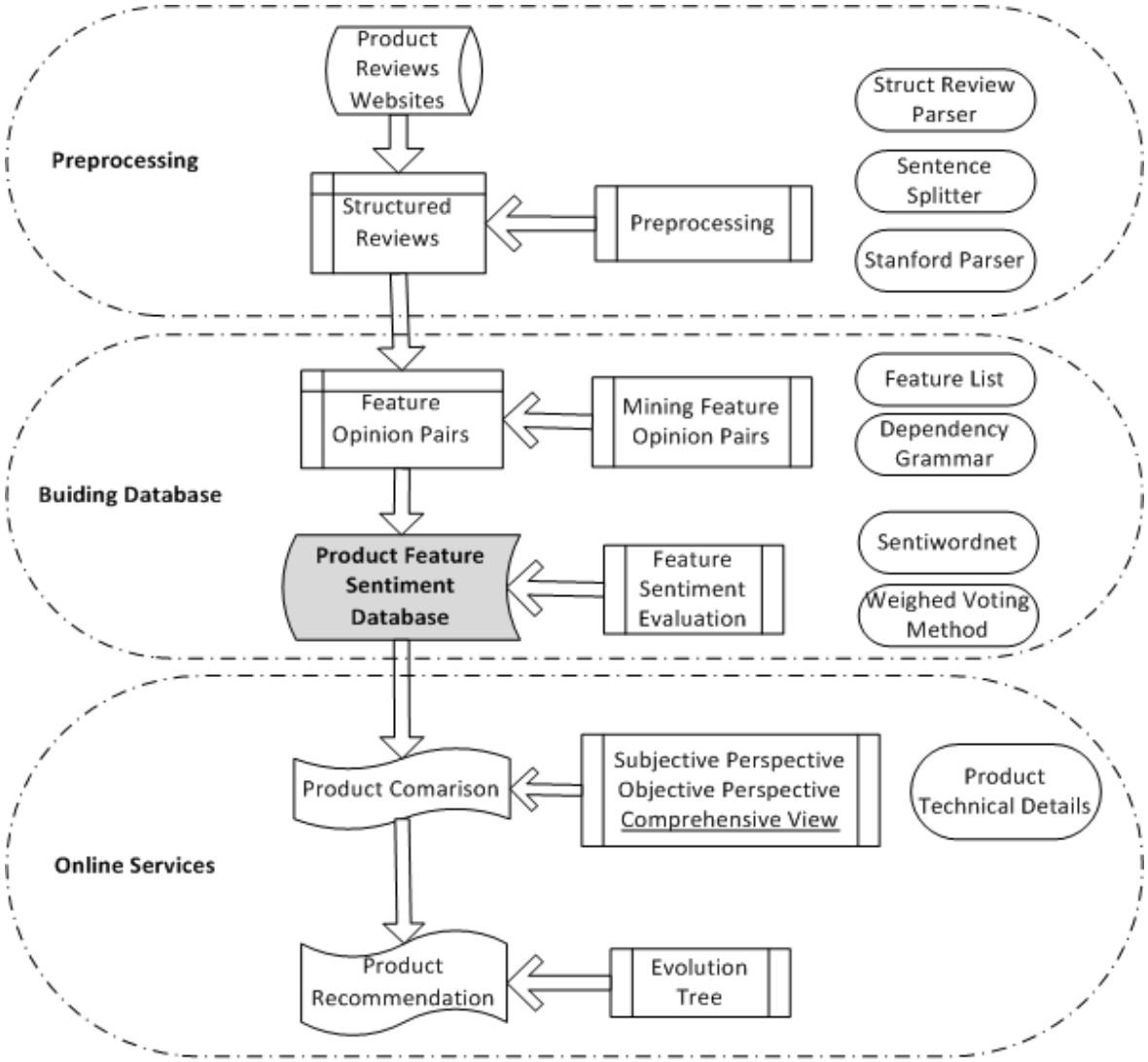


Figure: System Architecture

Feature Extraction

- N-Gram Feature (N<=3)
  - 1-Gram: “picture”, “screen”, “software”
  - 2-Gram: “memory card”, “picture quality”
  - 3-Gram: “red eye reduction”
- Feature Selection [2]
  - Frequently occurred N-Grams are selected as candidate feature words.
    - Noun phrase
    - The same NP-Chunker
    - Not containing stop-words
  - Various filtering methods are used on the candidates
    - Higher frequency
    - Higher Chi-square score between different product types

Product Feature for “Digital Camera”

[product]					
camera	camera	digit camera	canon	canon camera	canon powershot
size	size	pocket	shirt pocket		
color	color				
price	price	price rang	monei		
[quality]					
picture	pictur	pictur qualiti	photo	photo qualiti	imag
video	video	video qualiti	video clip	movi mode	video mode
[display]					
display	lcd screen	lcd	screen	view screen	lcd displai
view finder	view finder	optic viewfind	viewfind	viewfind	optic view find
[shot]					
zoom	lens	optic zoom	lens	wide angl	zoom len
shutter	shutter speed	shutter	shutter lag		
flash	flash				
focus	shot	focu	auto focu		
[special]					
……					
[parts]					
battery	batteri	batteri life	recharg batte	aa batteri	extra batteri
[service]					
customer ser	custom servic				

Mining Feature-Opinion Pairs

- Dependency Grammar Graph [3]
  - “Feature” keywords are used to retrieve sentences.
  - “Opinion” words have pos tags of “ADJ” “ADV” or “ADV ADJ”.
  - Finding dependency path from “Feature” to “Opinion” in the sentence.

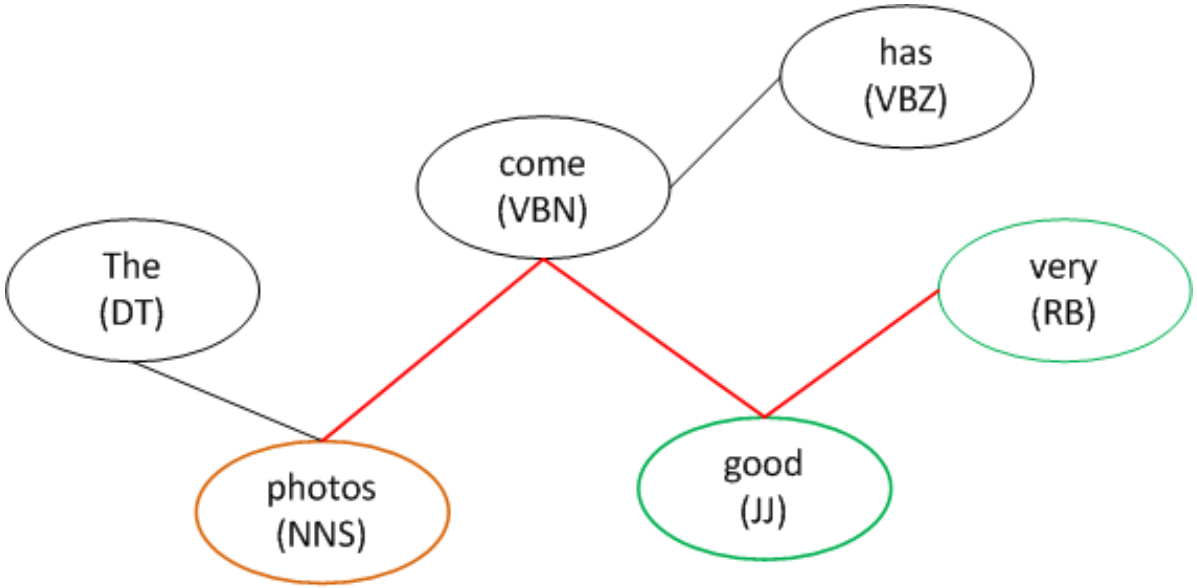


Figure: Dependency grammar graph from “Feature” to “Opinion”

Feature-level Product Assessment

- Using Sentiwordnet [4] to evaluate the sentimental strength of an "opinion" word.
  - Generated by WordNet [5].
  - Each word has a sentiment score (positive score, negative score)
- Evaluating the sentiment score of a "feature"
  - Integrating all the sentiment scores of "opinion" words related to "feature" by Feature-Opinion-Pairs(FOP).
  - Weighted voting method

$$feature_sentiment = \sum_{opinion_weight_k >= 2} opinion_weight_k \times opinion_sentiment_k$$

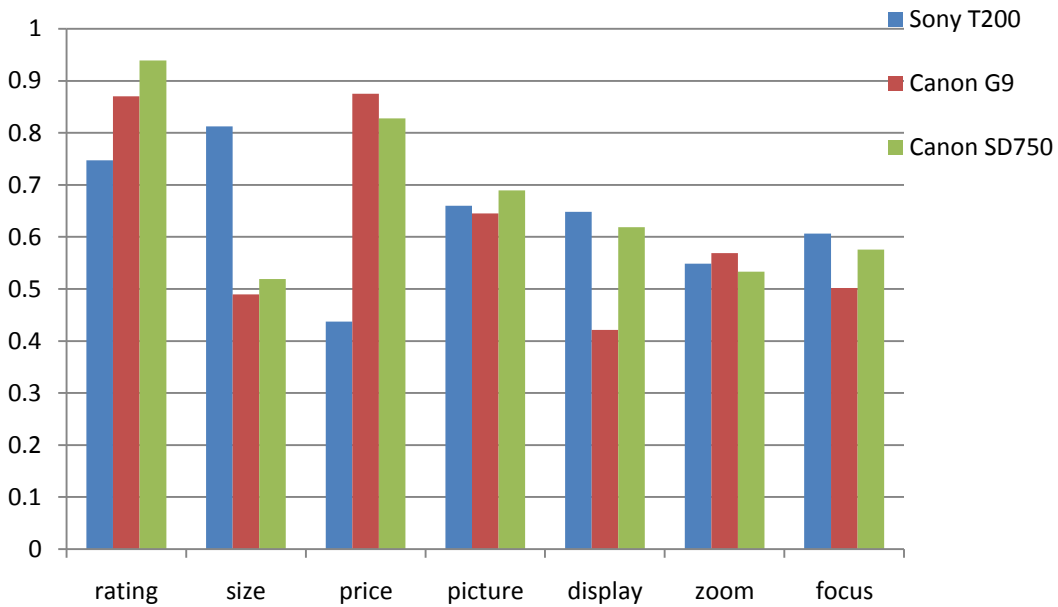
$$(opinion_weight \ \ represents \ \ the \ \ frequency \ \ this \ \ FOP \ \ ocurrs)$$

Product Comparison

- Subjective Perspective
  - Mining from reviews
  - Focusing on user experiences
- Objective Perspective
  - Incorporating technical details (physical parameters)
  - Reliable but not always comparable
- Overall Perspective
  - Rating products comprehensively according to review contents

Subjective Perspective

Compare various products based on "feature" sentiment.

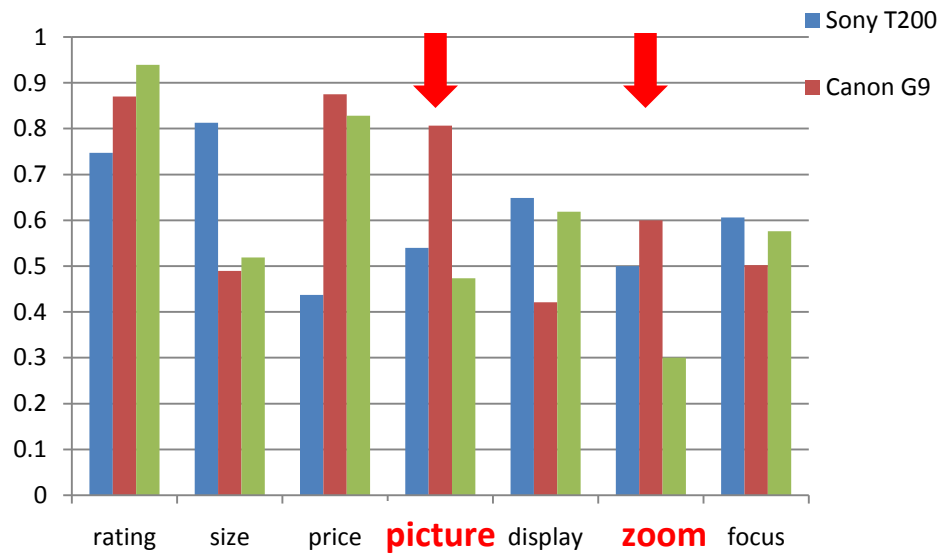


- **rating**: Sony T200 < Canon G9 < Canon SD750, based on Amazon [6] ratings.
- **size**: T200 and SD750 are with slim vertical design, while SD750 is not.
- **price**: SD750 ($184.49) is so cheap compared to its relatively high performance.
  G9($614) is a new product, with 12.1MP and 6x image-stabilized optical zoom, whose price is worth its performance.
  T200 ($599.44) is too expensive even if people don't care about the size.
- **picture, zoom, focus**: almost the same
- **display**: T200 and SD750 have bigger LCD screen, with clear display.

Objective Perspective

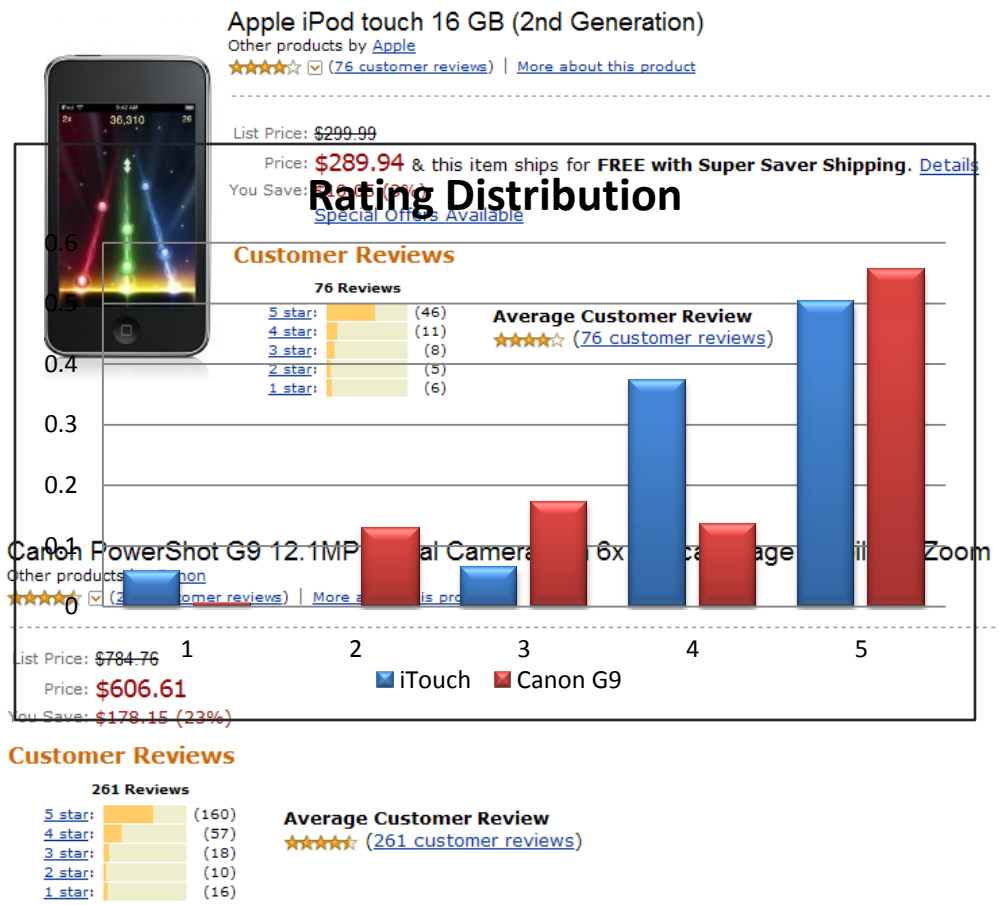
- Limitations of subjective perspective
  - Too subjective: sometimes make no sense.
  - Lack of user comments on some features.
- Incorporating technical details

	Sony T200	Canon G9	Canon SD500
Picture Revolution	8.1MP	12.1MP	7.1MP
Zoom	5	6	3



Overall perspective

- Sentiment polarities in reviews
  - traditional definition
    - positive vs. negative
    - 5 star scheme
  - our definition
    - rating distribution

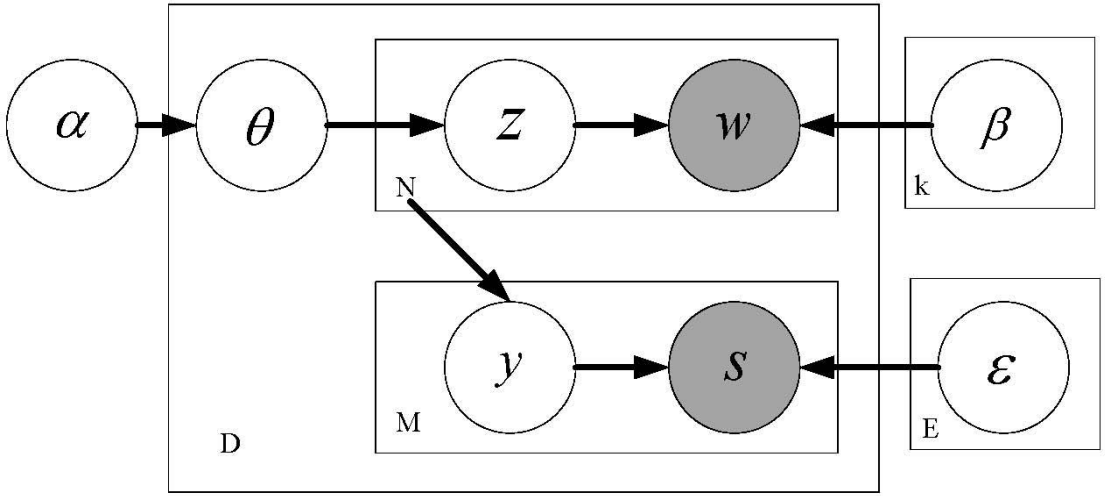


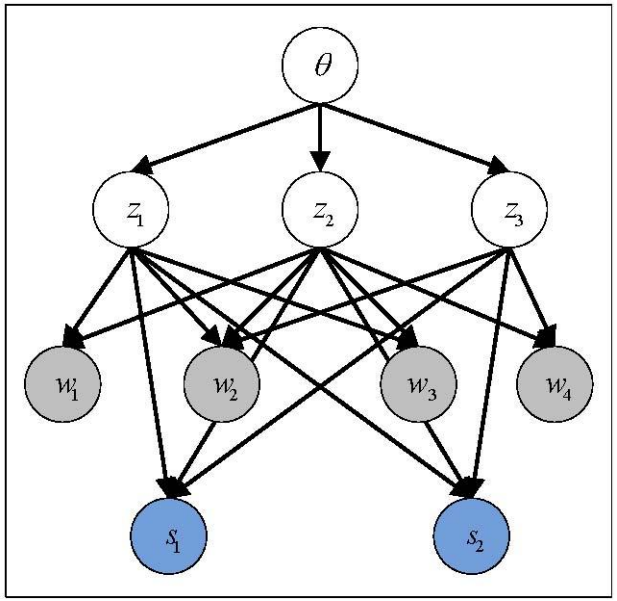
Analyze Content of Reviews

- Sentiment-oriention based approach
  - predefined “sentiment words”
  - disadvantages
    - ✗ Can not capture all expressions of subjectivity
- Our approach
  - all the content words
  - advantages
    - ✓ weight sentiment ratings according to all contents
    - ✓ capture general usage patterns

Word-Sentiment Model

- Generative Topic Model [7]





⊙: Dirichlet distribution
Z: topic variable
W: word
S: rating(1-5 star)

Generative Process

- Sample θ from Dirichlet distribution: $\theta \sim Dir(\alpha)$.
- For each word w_n, $n \in \{1, \cdots, N\}$
  - Sample topic factor z_n from multinomial distribution: $z_n \sim Mul(\theta)$
  - Sample w_n from multinomial distribution conditioned on z_n: $w_n \sim p(w_n \mid \beta, z_n)$
- For each sentiment rating s_m, $m \in \{1, \cdots, M\}$:
  - Sample indexing variable y_m from a uniform distribution: $y_m \sim Unif(1, \cdots, N)$
  - Sample s_m from multinomial distribution conditioned on z_{y_m}: $s_m \sim p(s_m \mid z_{y_m}, \beta)$.

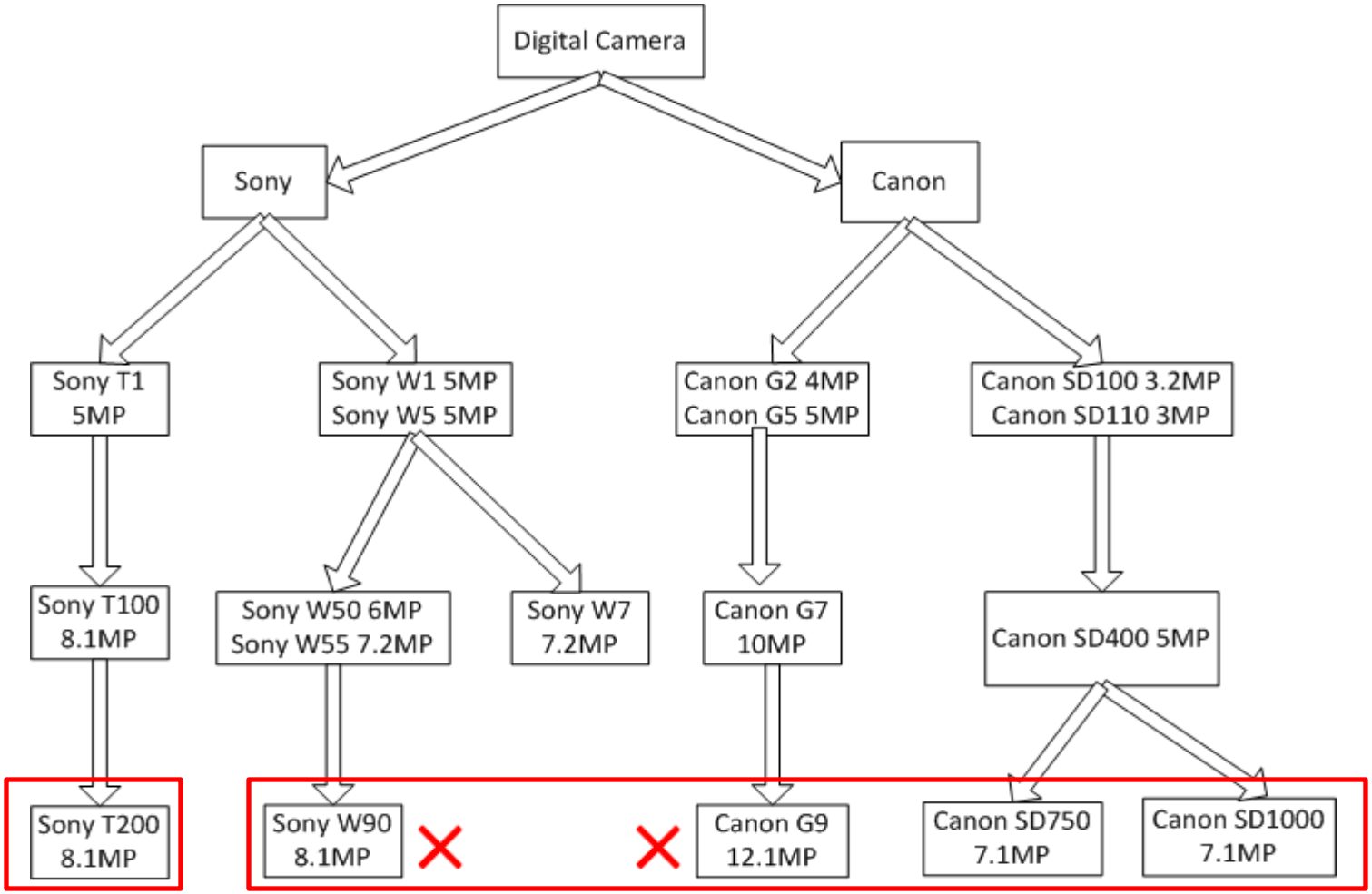
Evolution Tree

- When we propose recommendation, we need to suggest the products with the same generation.
  - Generation can be defined by various standards.
    - Time?
  - But must be significant in the product development.
    - Mobile phone: GSM → 3G
    - Memory card: SDRAM -> DDR -> DDR2 -> DDR3
- Evolution tree can provide a graphical representation
  - Informative
  - Intuitive
  - …

Product Recommendation

- We construct a simple evolution tree by
  - Product selling period extracted from reviews
  - Product full name (rule based)
  - Important technical details:
    - Digital Camera: picture resolution, zoom
- Recommendation procedure
  - Select all the products who have the same generation with the current product.
  - Keep the winners in comparison to the current product.
  - Rank the recommended products.

Evolution Tree: an Example

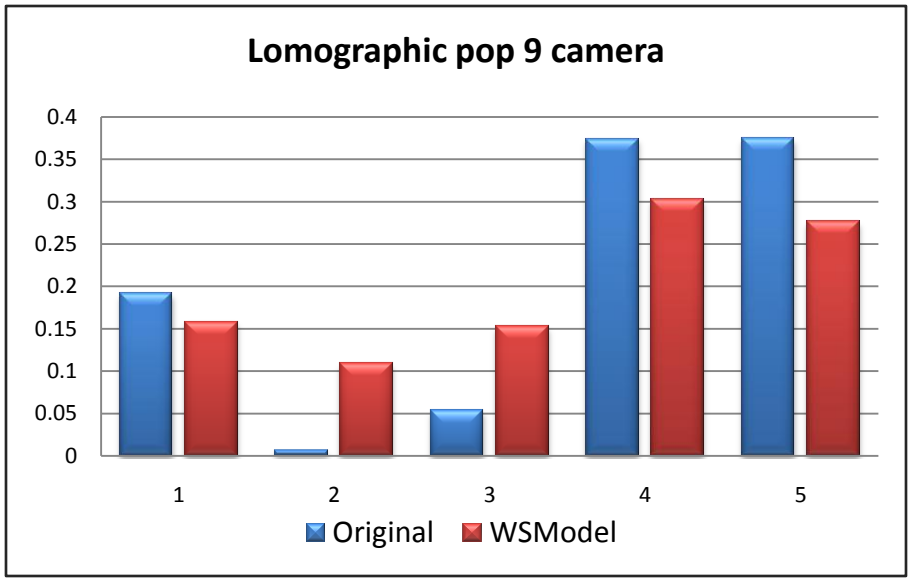


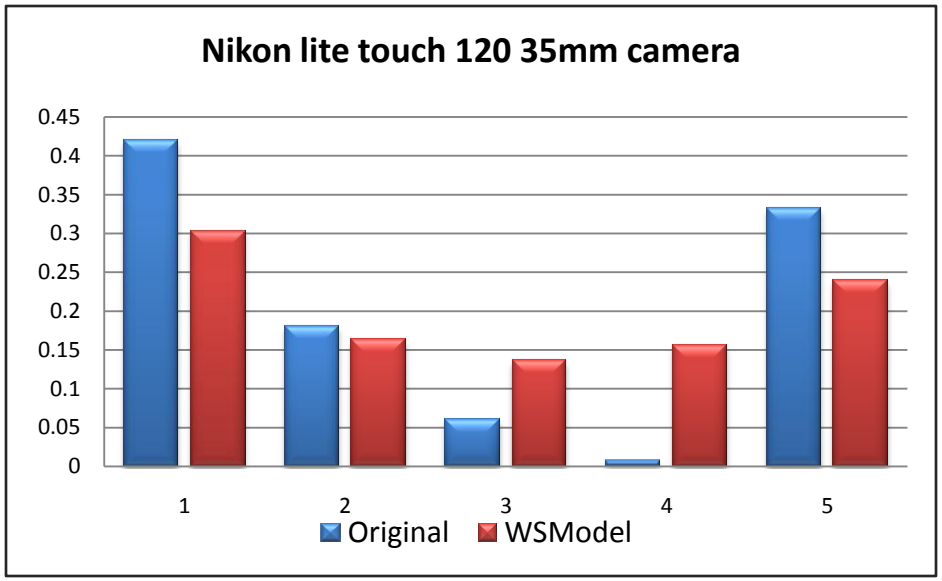
Screen evolution

Figure: Evolution tree of digital camera

Experiment Results

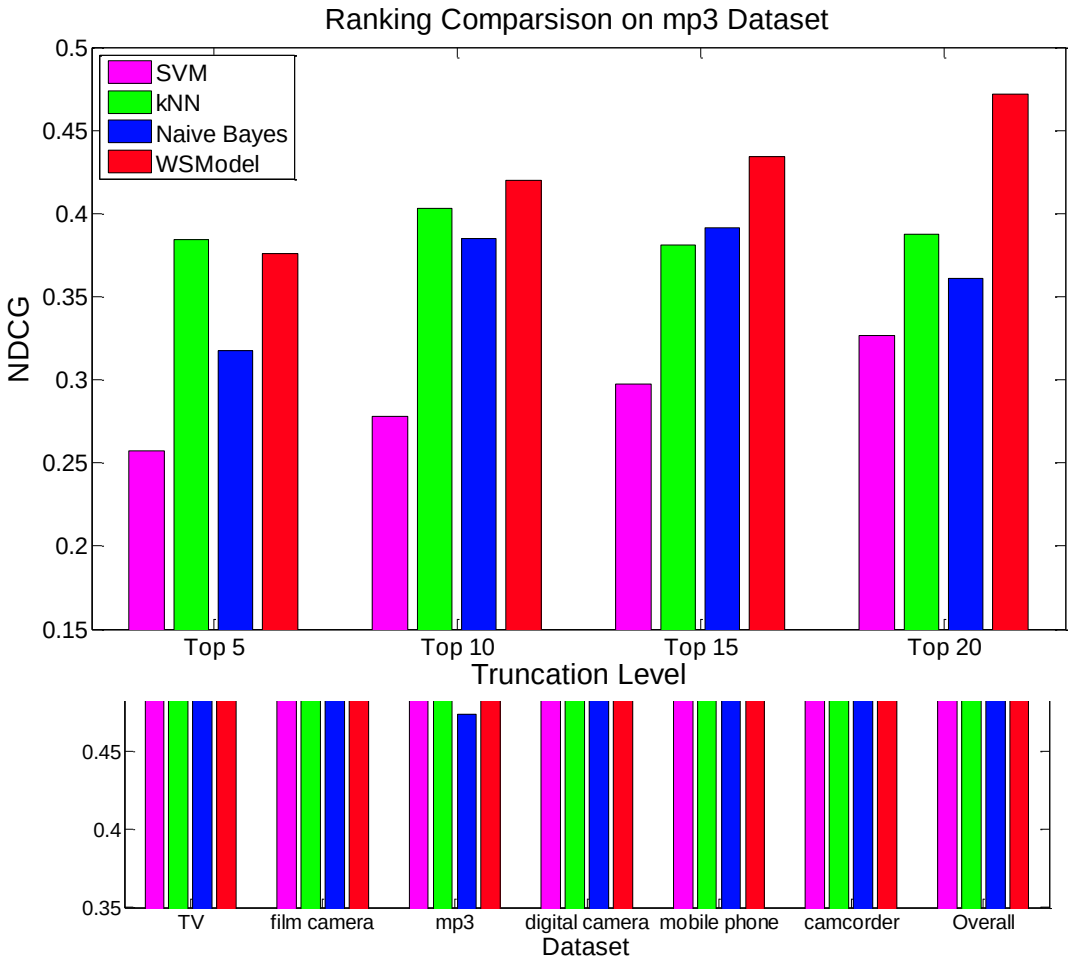
- Posterior Inference
  - $p(s \mid \mathbf{w}, \alpha, \beta, \eta) \approx \sum_{z_n} q(z_n \mid \phi_n) p(s \mid \varepsilon, z_n)$





Experiment Results

- Rating Classification
  - Accuracy
- Ranking
  - Normalized Discounted Cumulative Gain



Experiment Results

- Review Summary
  - By *WSModel*
  - $p(w|\alpha,\beta,\eta) \approx \sum_{z_n} q(z_n|\phi_n)\, p(w|\beta, z_n)$

Product	Description Words (TOP 10)
Minox ecx sub-miniature 8x 11mm camera	small, size, pocket, light, carry, design, hand, easily, choice, reasonable, image,
Olympus stylus epic 35mm camera	lens, great, shoot, zoom, flash, small, film, buy, good, stylus, best
Polaroid silver blue joycam instant camera	polaroid, fun, film, instant, buy, easy, product, photo, expensive, color, price

Summary

- Comparing product from subjective (opinion) and objective perspectives
- Simulating the rating system from the overall perspective (the topic generative system)
- Recommending products by evolution tree

Demo System

Home | Product Summary | Product Comparison | Product Recommended | Evolution Tree | About us

PCR System is a review mining system aiming at implementing product comparison and recommendation on the feature level.

The system has the following functions:

- Querying a product's review summary on each feature. The reviews are from Amazon and partly from other websites.
  "Please tell me some user's comments on Canon G9's feature."
- Comparing two product on the feature level, and telling customers such as:
  "Whose screen is better, Sony T200 or Canon G9."
- Recommending similar products to the current one when given the feature standard.
  "What products can be recommended whose screen is as good as Sony T200?"
- Generating products' evolution tree.

PCR System has employed many innovative methods and algorithms, including natural language processing, Feature-Opinion-Pairs Extraction, feature sentiment summmarization. For detailed descriptions of these algorithms, please contact us.

Website: http://166.111.138.87:2008

References

[1] Nitin Jindal and Bing Liu. Indentifying Comparative Sentences in Text Documents*. In Proceedings of SIGIR'06.*

[2] Minqing Hu and Bing Liu. *Mining and summarizing customer reviews. In Proceedings of ACM-KDD 2004,* pp.168-177.

[3] Li Zhuang, Feng Jing and Xiaoyan Zhu. Movie Review Mining and Summarization. *In Proceedings of CIKM'06.*

[4] Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of LREC, 417-422.*

[5] J. Kamps and M. Marx. 2002. Words with attitude. *In Proc. of the First International Conference on Global WordNet, pp.332-341.*

[6] *http://www.amazon.com*

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, Jan 2003.

Thank you!

Any questions?