Text Summarization of Review Sentiments

Eric Jensen

Summize, Inc.

S U M M I Z E

BETA

Home   Trends   Sign In

Search

We've summized **44,336,409** opinions across the web

Find reviews and sentiments on millions of items

Enter a band, actor, author, game, camera, printer... | Summize

Discover what people **like** and **dislike**

Overall Opinion
Total sentiments across our entire site

44,336,409 sentiments
on 20,878,187 products
by 4,345,674 people

What is Summize?

Summize connects you with the **opinions** of millions of people on zillions of products.

We scour the web for the latest user reviews and blogger discussions, surmise their **sentiments**, and summarize them here. **Learn more**.
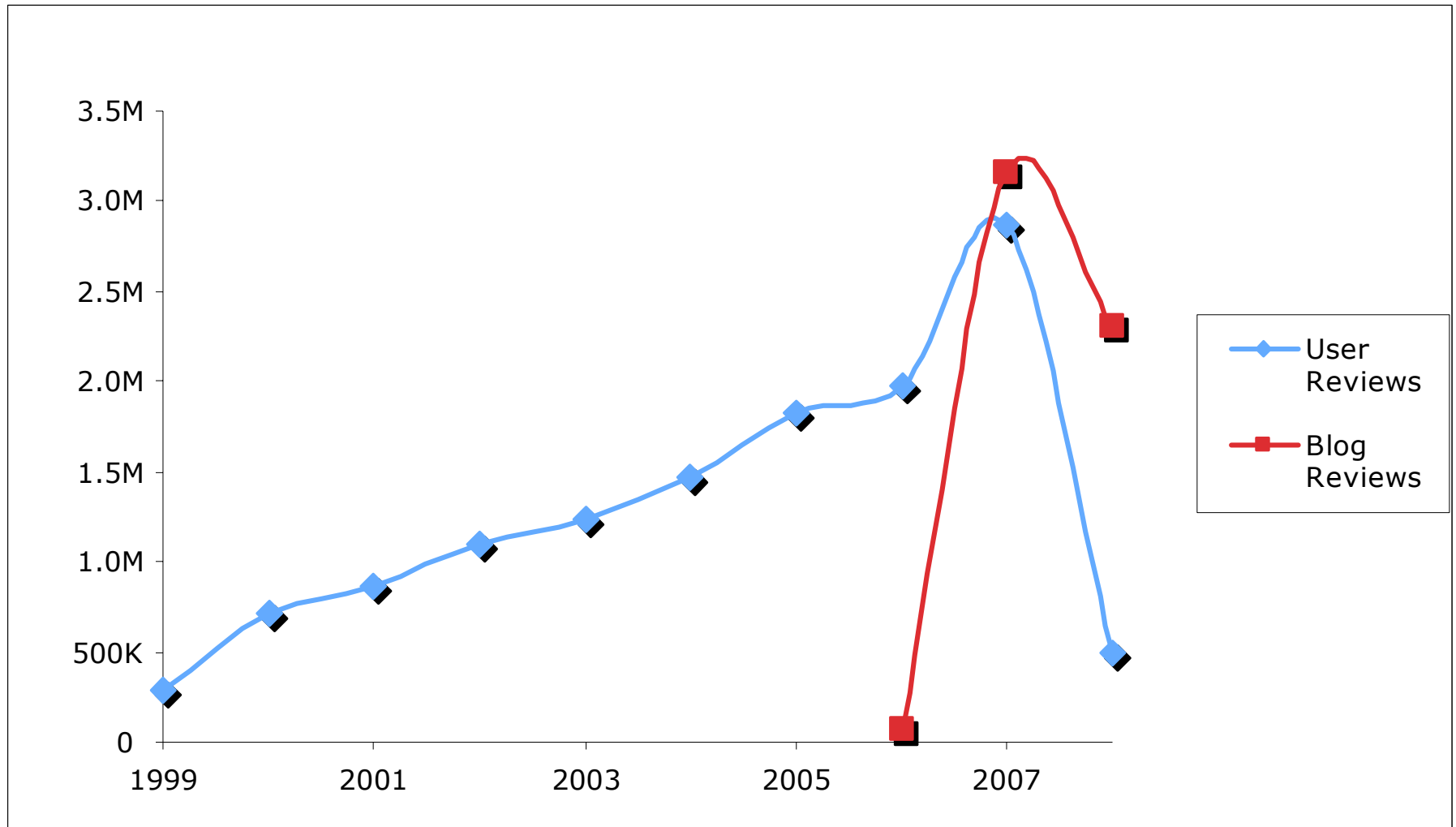
Now on the iPhone!

Get reviews on the go with our new **Summize** webapp.

Trending books by...

Andrew Peterson
Larry Smith
Dante
Michael Masterson
Jack Kerouac
Joseph E. Stiglitz
Stefanie Wilder-Taylor
Joshilyn Jackson
Michael Connelly
Sarah Addison Allen

Trending music by...

Counting Crows
The B-52's
Destroyers
Lindsey Buckingham
Muses
Panic at the Disco
Ghostface Killah
Toby Keith
Great Northern
Adam Green

Trending movies by...

Megumi Okina
Thomas Jane
Will Arnett
Tyler Perry
Jerry O'Connell
Leslie Mann
Stephen Sommers
Tommy Lee Jones
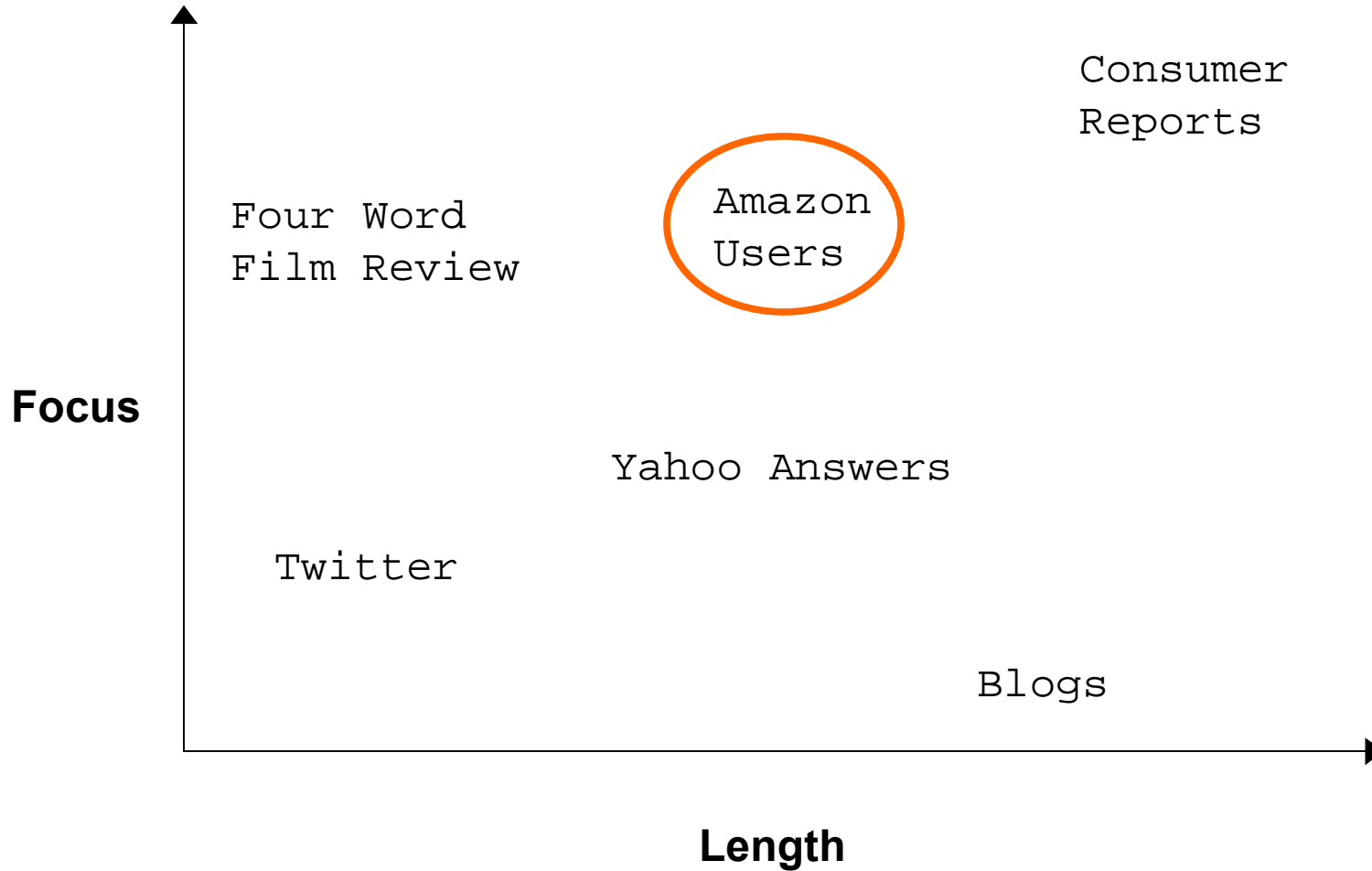Michael Chiklis
Pierce Brosnan

Outline

- Opinions on the web
- Opinion mining
- Text summarization
  - The problem
  - Proposed algorithm
  - Results
- Conclusions

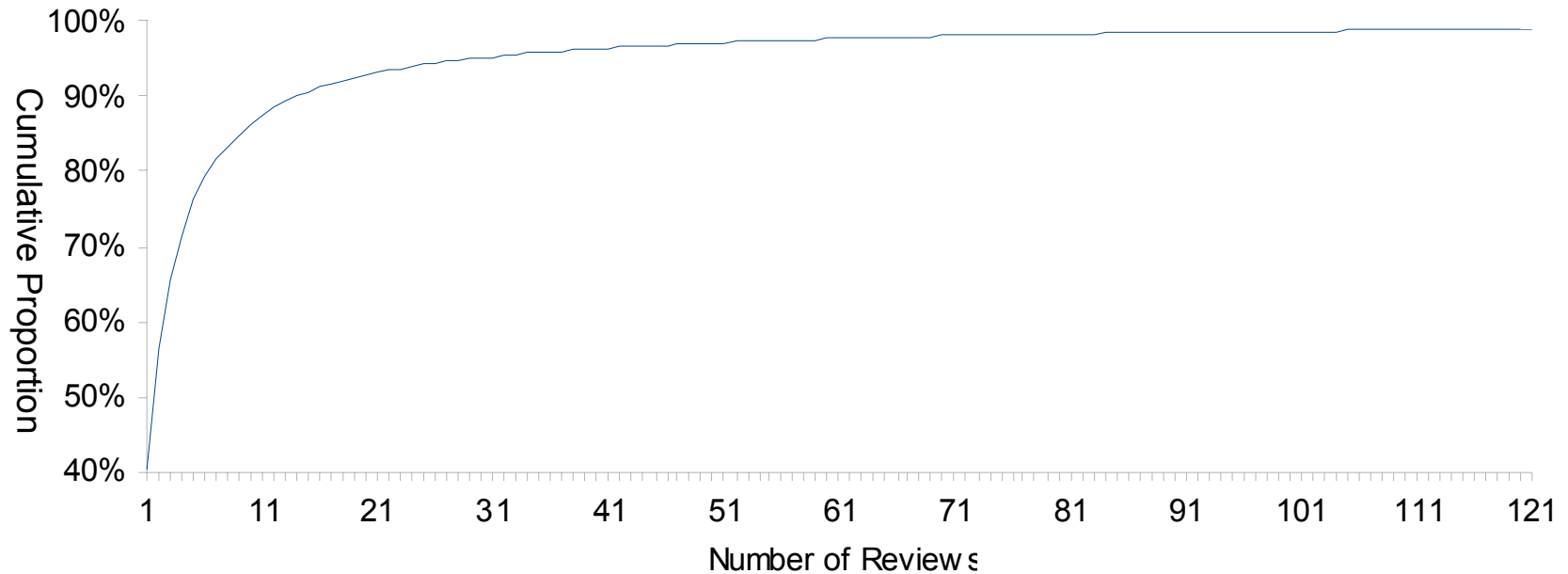# Growth of Amazon, IMDB, and Blogs

Opinions on the web

Focus

Consumer
Reports

Four Word
Film Review

Amazon
Users

Yahoo Answers

Twitter

Blogs

Length

Support (or lack of?)

How many are you willing to read?

Opinion mining

- Sentiment analysis
- Facet mining
- **Text summarization**

Sentiment analysis! (Pang EMNLP 2002, Dave, et. al WWW 2003)

I Am Legend

944 sentiments
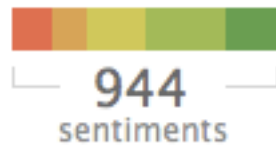
"I won't review the movie because this has already been done. What I will rate is the 2-disc 'Special Edition' of this movie…Overall, I feel this 2-disc edition is not worth the extra money it costs."

Facet mining (Hu and Liu KDD 2004, Popescu and Etzioni EMNLP 2005, Titov and McDonald WWW 2008)

- **Digital camera**
  - Resolution
  - Zoom
  - User interface
- **I Am Legend**
  - Acting
  - Special effects
  - *2-disc special edition?*

Text summarization

The problem: understand the prevailing sentiments as quickly as possible

- Leverage the ratings users provide to produce *more* meaningful summaries
- Don't restrict to fixed categories/facets
- *Why* did the users rate it this way

Example

I Am Legend

- **riveting** **movie**
- **hollywood** **ending**
- **amazing** **story** • **excellent** **character** • **riveting** **performance**
- **dark sci-fi** • **grotesque film**

Experimentation
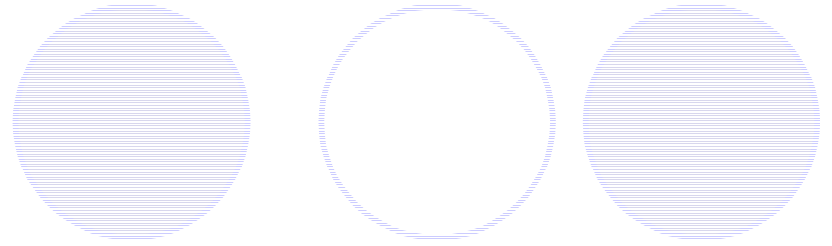
- Dataset
- Evaluation
- Baseline
- Results
- Consensus Building

Experimentation: Dataset

- Amazon and IMDB
- 10 million user reviews
- 3.6 million products
- Books, movies, music, and others

Evaluation

- Sampled 30 products
  - Stratified by category
  - Minimum of 10 reviews each
- Task: ideal 10-word summary of the prevailing sentiments about that product
  - Mix positive and negative in appropriate ratio
  - Arbitrary length phrases
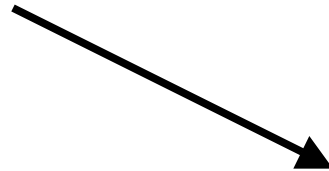- E.g. **vacuum cleaner**: high suction, heavy, do not buy

Evaluation: Metrics

- Text Analysis Conference (formerly DUC)

- Overlap of reference summaries highly correlated with manual evaluation (Lin & Hovy HLT-NAACL 2003)

$$ROUGE - N = \frac{\sum\limits_{gram_n \in reference} Count_{match}(gram_n)}{\sum\limits_{gram_n \in reference} Count(gram_n)}$$

Framework

Input

944 sentiments
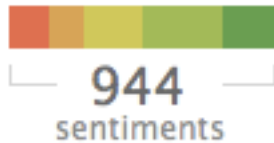
Output

- **riveting** **movie**
- **hollywood** **ending**
- **amazing** **story** • **excellent** **character** • **riveting** **performance**
- **dark sci-fi** • **grotesque film**

Baseline: Adapted facet-oriented mining (Hu and Liu KDD 2004)

1. Identify noun phrases and treat adjacent adjectives as opinion words

2. Rank noun phrases by TFxIDF

3. *Choose top opinion word by frequency*

4. *Choose top summary phrases by frequency*

   - 3 & 4 our adaptation

Proposed algorithm
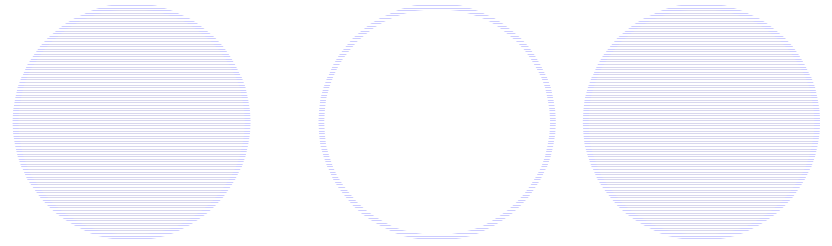
1. Identify each opinion word and treat the following word as a "facet" word

2. Rank facet words by frequency

3. Choose top opinion word by frequency

4. Choose top phrases by frequency

Results

Method / Metric	Precision	Recall	$F_{0.5}$
Facets ROUGE-1	0.329	0.189	0.215
Summize ROUGE-1	0.293	0.263	0.273
			+26.81%
Facets ROUGE-2	0.105	0.025	0.033
Summize ROUGE-2	0.050	0.044	0.045
			+36.25%
Facets ROUGE-SU4	0.161	0.054	0.059
Summize ROUGE-SU4	0.107	0.088	0.091
			+55.03%

Consensus Building
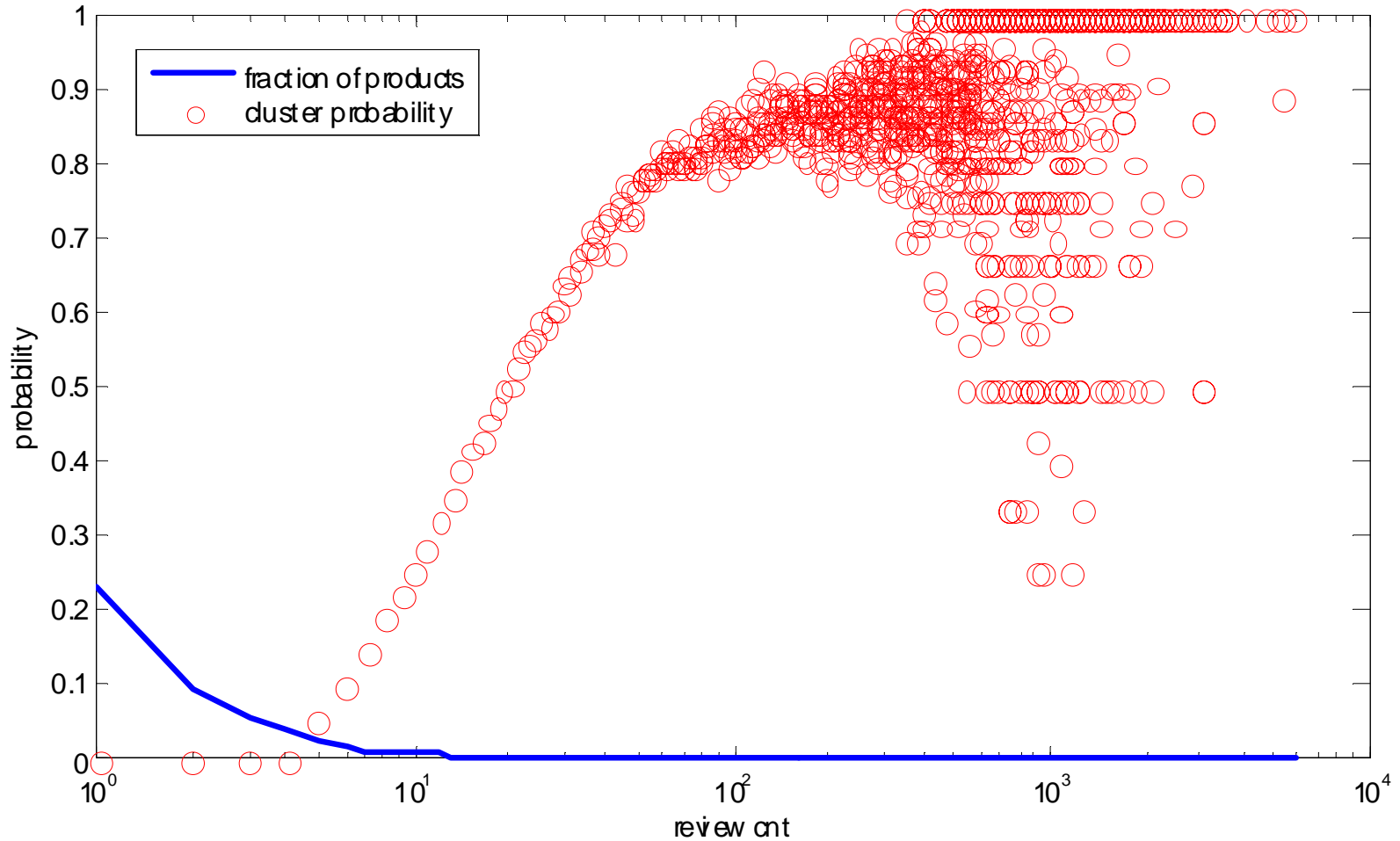
Conclusions

- Number of opinions on the web are growing faster than anyone wants to read
- Text summarization reveals the *why* behind the ratings
- Facets do not capture the ideal summaries (sentiment-oriented ones are 26% closer)
- **Scaling is both a problem and an opportunity**

Future Directions

- Scale to more and more reviews
- Analyze opinions from unstructured sources (blogs, twitters, etc.)

Plugging my other work

- Semi-automatic evaluation (ACM TOIS '07)

- Query classification (ACM TOIS '07)

- Query log analysis (SIGIR '04)