DiG: A Task-based Approach to Product Search

Scott Carter[1] Francine Chen[1] Aditi Muralidharan[2] Jeremy Pickens[1]

[1]FX Palo Alto Laboratory [2]UC Berkeley

DiG is a walk-up-and-use system that lets people make informed purchasing decisions about products.

DiG recommends products to users **based on how they plan to use the product.**

The interface uses information extracted automatically from sentences in online customer reviews. Statistical methods that cluster and find patterns are used to get detailed information about product features and uses.

Problem: fine-grained extraction

- Need to extract a list of a product's features and uses from customer reviews with high precision and low redundancy
- Cannot use existing web-scale algorithms (not enough data)
- Need to accurately evaluate customers' overall opinion of a product feature from customer reviews

Solution: Statistical and Linguistic Information

- Use human knowledge to:
  - Map product uses to features
  - Provide a few examples of product features
  - Annotate adjectives with sentiment
- Use grammatical structure to:
  - Filter out incoherent extractions
  - Characterize extracted items for grouping
  - Accurately identify opinion-giving adjectives
- Use statistical methods to:
  - Filter out infrequent and imprecise extractions
  - Cluster extracted items into groups
  - Predict the opinion polarities of adjectives

Data: Digital Cameras

- Standard Product Specifications
- Product Attributes
- Reviews
  - 168,638 reviews for 3700 cameras from Amazon
  - Over 1.1 million sentences

Processing Reviews

Product Feature Extraction

1 Learn extraction patterns

… digital camera features a < > and …
… the < > on this camera …
… the camera's < > is …
… camera has < >, which …

2 Pattern-match to get candidates

… digital camera features a <**macro setting**> and …
… the <**image stabilization**> on this camera …
… the camera's <**best selling point**> is …
… camera has <**grainy images at night**>, which …

3 Identify reliable product features

- SVM classifier
- Features based on web-PMI

4 Cluster product features

- Cluster noun sequences based on noun-adj distributions
- Filter with reliable product features

camera, body
photos, pics, pictures, shots
batter life, photo quality, quality, picture quality, image quality
zooms, zoom
screen, lcd, view screen, lcd screen, lcd display, display
lens, lenses
image, shot, picture
bang, deal, value, job
settings, setting
aa batteries, batteries

Opinion Mining

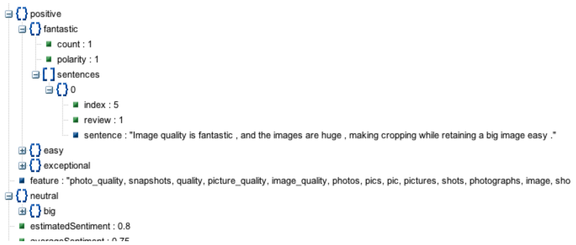
5 Classify adjectives: PMI features

Web PMI = log hits(a, b)/hits(a)hits(b)
+1: "good" :{'excellent':1.98 , 'poor': 0.32}, …
-1: "bad":{'excellent':0.27, 'poor':2.63}, …
?? "sharp":{'excellent':0.9, 'poor':0.21}

6 Evaluate opinion: use adjectives

image quality: grainy (-1), crisp (+1), clear (+1) …
battery life: long (+1), good (-1), excellent (+1) …
We use beta-binomial model to smooth estimates when there are only a few adjectives for a product feature

7 Extract summarizing sentences and terms



Product Use Extraction

8 Pattern-match to get candidates

- <picture term> <prep phrase> <noun phrase>

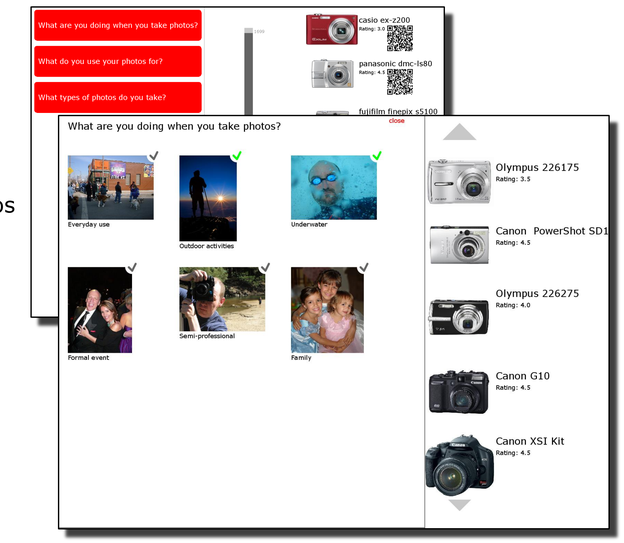
9 Filter candidates

light	in low light	in bright light
people	of people	with people
conditions	in low light conditions	in all conditions
time	at a time	at one time
kids	of the kids	of kids
family	of family	of the family
friends	of friends	of family and friends
computer	on the computer	on computer
price	for the price	at a great price
flowers	of flowers	of the flowers

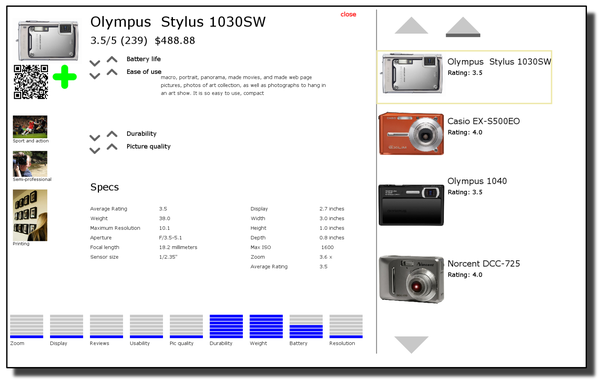
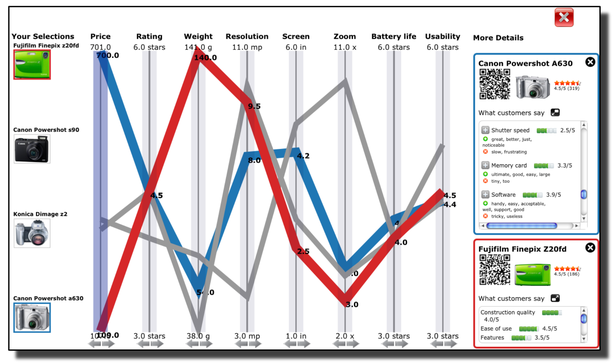
Integrating Analysis Results into an Interactive UI

- Allow users to describe how they will use a product
- Bridge between chosen uses and resultant features
- Compare a chosen set of cameras

Users can specify types of pictures, activities, and how they use their photos



Users can directly manipulate features



Parallel coordinates used to compare cameras