Joint Model

for Text and Aspect Ratings

(Titov & McDonald 2008, Zhao et al. 2010)

Review Mining

- User reviews: short, opinionated, and not necessarily thorough
    - Many times only cover a few aspects of a product
- Goal: thorough, crowd sourced reviews and ratings of a product, covering all aspects
- Solution: sentiment summarization

Aspects

- Aspect: a particular feature or component of a product or service:
  - Cameras: picture quality, battery life, size, weight, storage, etc.
  - Scanners: scanning bed, cover, scan quality, DPI, etc.
  - Hotels: location, service, room, decor, amenities, etc.

Aspects

- Opinion: subjective viewpoints
    - General opinion words:
        - Good, great, bad, terrible
    - Aspect specific opinion words:
        - Service: friendly, nice, mean, rude
        - Food: tasty, delicious, undercooked
        - Location: nearby, overlooking, ugly

Problem

Input:

Food: 5; Decor: 5; Service: 5; Value: 5

> The chicken was great. On top of that our service was excellent and the price was right. Can't wait to go back!

Food: 2; Decor: 1; Service: 3; Value: 2

> We went there for our anniversary. My soup was cold and expensive plus it felt like they hadn't painted since 1980.

Food: 3; Decor: 5; Service: 4; Value: 5

> The food is only mediocre, but well worth the cost. Wait staff was friendly. Lot's of fun decorations.

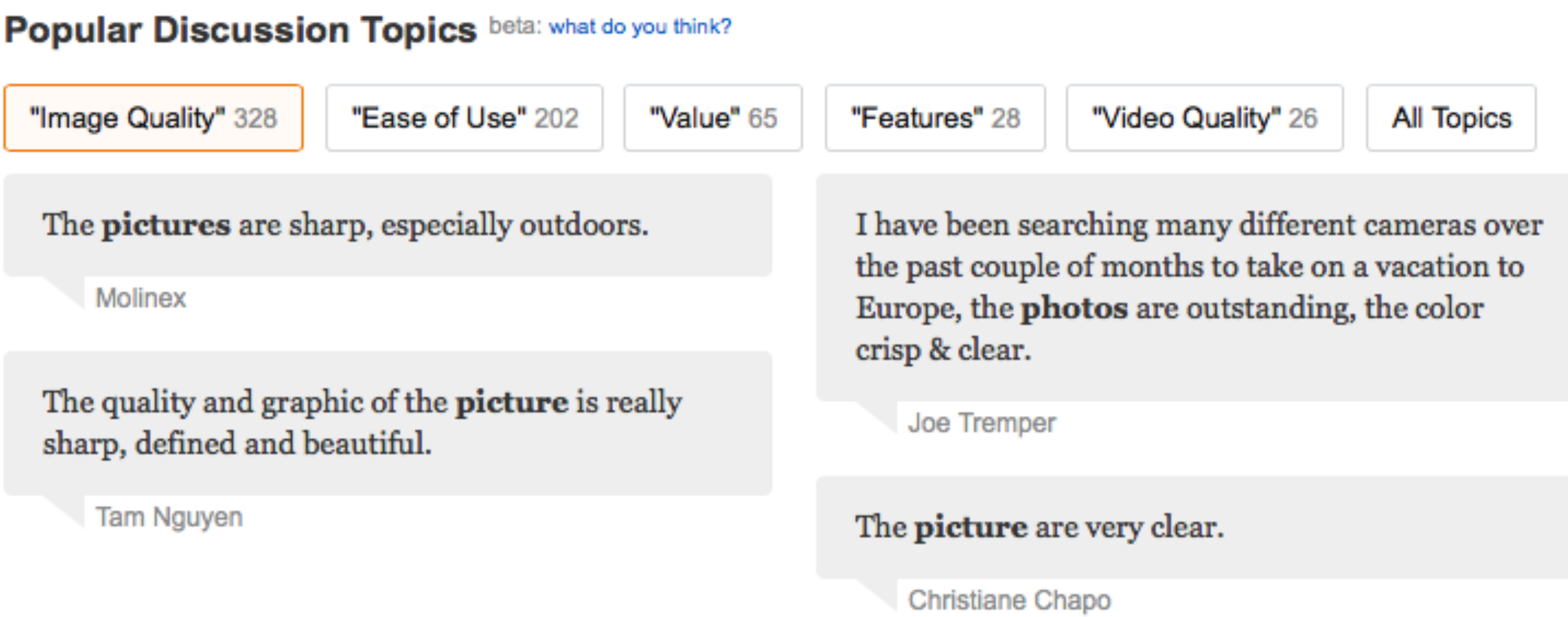
Problem

Output:

Food	"The chicken was great", "My soup was cold", "The food is only mediocre"
Decor	"it felt like they hadn't painted since 1980", "Lots of fun decorations"
Service	"service was excellent", "Wait staff was friendly"
Value	"the price was right", "My soup was cold and expensive", "well worth the cost"

Problem

Review mining in action (Amazon)



Problem

- Automatically recognize aspects for words/sentences
  - Without having to label large training data for word- or sentence-level aspects

- Find the sentences that support the aspect ratings

Multi-Aspect Sentiment Model (MAS)

- Words are *generated* from *topics*:
  - Topic 1 (*Service*): staff, friendly, concierge
  - Topic 2 (*Location*): walk, metro, minutes
  - Topic 3 (*Rooms*): shower, tv, bed
  - Topic 4: moscow, russia, petersburg
  - ...
- Aspects are “local topics” that are common to documents of various “global topics”
  - This multi-grain topic model correctly yields ratable aspects (unlike naive LDA)

Sentence Labeling; Model 1

- Label for each sentence is the highest count of labeled words:
  - wait/*service* staff/*service* was friendly/*service*
  - the tube station/*location* is about an 8 minute/*location* walk/*location*

- The result is comparable to a supervised classifier (but needs no training)

Sentence Labeling; Model 2

- One MaxEnt classifier per aspect, 10-fold cross-validation
  - unigram/bigram features

- Considered upper bound for unsupervised approach

Main Claim

- New model, Multi-Aspect Sentiment model
  - Text accompanying a review is predictive of the scores

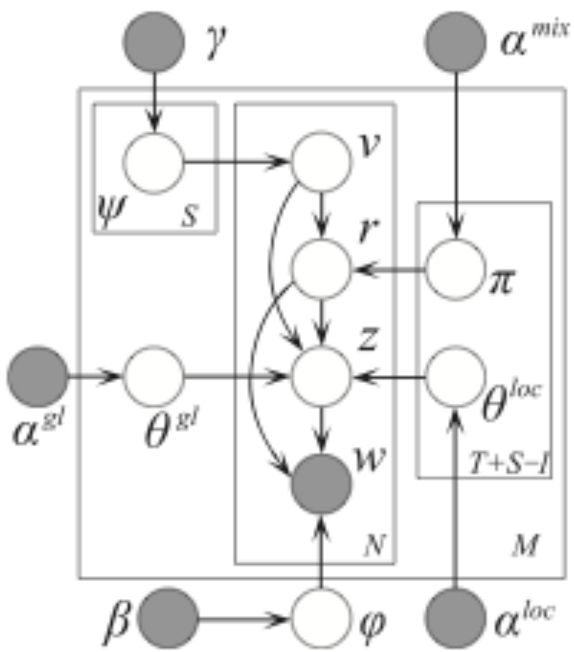
- Unsupervised method using MAS is comparable to supervised MaxEnt method

Data

- 10,000 reviews from tripadvisor.com
  - 109,024 sentences
- Each review marked for “service”, “location”, and “rooms”
- Tokenized and sentence split automatically
- 779 sentences labeled for aspect
  - 603 sentences marked for aspect

Joint model (MAS)

Generative story

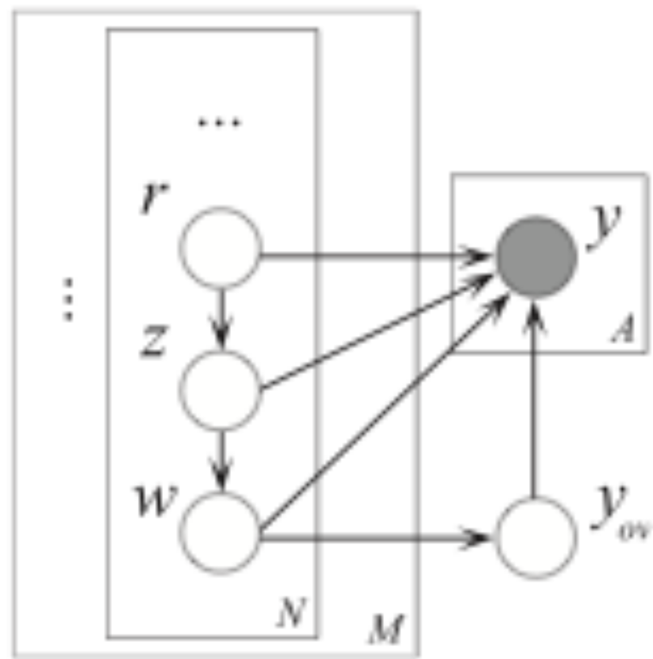


- Choose a distribution of global topics $\theta_d^{gl} \sim Dir(\alpha^{gl})$.
- For each sentence s choose a distribution over sliding windows $\psi_{d,s}(v) \sim Dir(\gamma)$.
- For each sliding window v
  - choose $\theta_{d,v}^{loc} \sim Dir(\alpha^{loc})$,
  - choose $\pi_{d,v} \sim Beta(\alpha^{mix})$.
- For each word i in sentence s of document d
  - choose window $v_{d,i} \sim \psi_{d,s}$,
  - choose $r_{d,i} \sim \pi_{d,v_{d,i}}$,
  - if $r_{d,i} = gl$ choose global topic $z_{d,i} \sim \theta_d^{gl}$,
  - if $r_{d,i}=loc$ choose local topic $z_{d,i} \sim \theta_{d,v_{d,i}}^{loc}$,
  - choose word $w_{d,i}$ from the word distribution $\varphi_{z_{d,i}}^{r_{d,i}}$.

Joint model (MAS)

User provided ratings are clues about how topics correspond to aspects

Inference: compute P(**r**, **z**|**w**, **y**) with approximate method (Gibbs sampling)

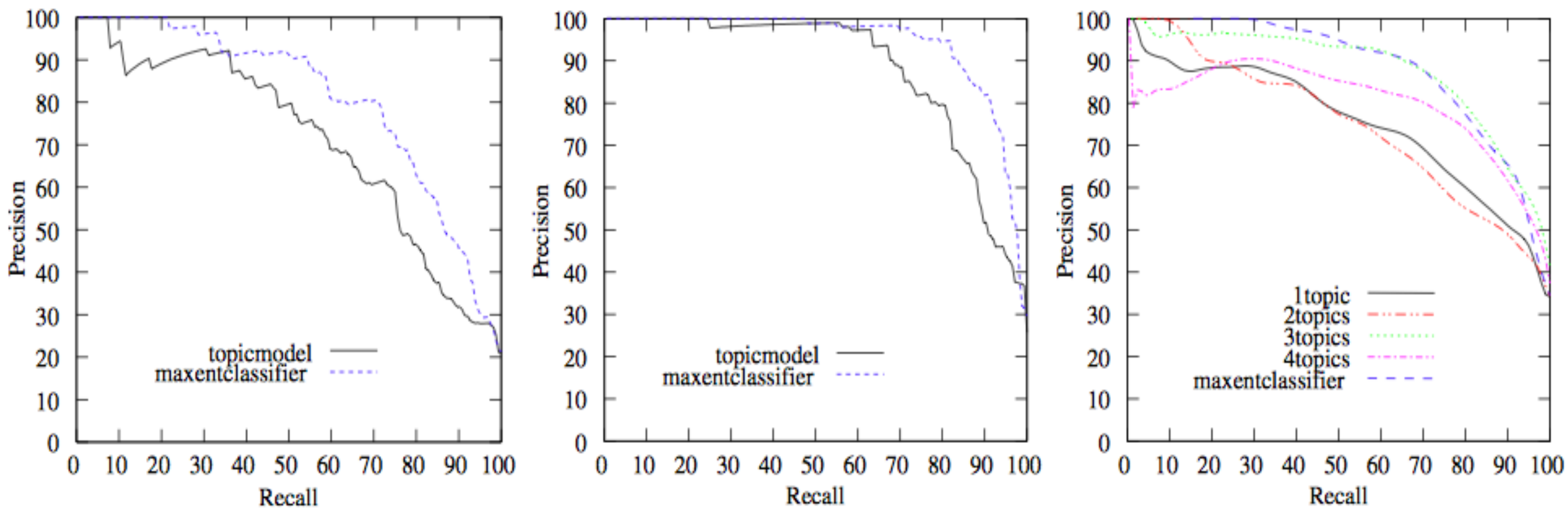


Evaluation

- *Qualitative* evaluation
  - Aspect words generated by the system “look good”
    - Topic 1 (*Service*): staff, friendly, concierge
    - Topic 2 (*Location*): walk, metro, minutes
    - Topic 3 (*Rooms*): shower, tv, bed
    - Topic 4: moscow, russia, petersburg

Evaluation

- Compared unsupervised method to supervised method to get upper bound
- Details light on scoring mechanism



MAS: Conclusion

Pros:

- Jointly models aspects and sentiment
- Flexible modeling by allowing multi-grain (global and local) topics
- Meets supervised baseline as an unsupervised model

MAS: Conclusion

Cons:

- Doesn’t discriminate between aspect words and opinion words
- Don’t discuss the weighting of particular sentences within window
- Some choices of meta-parameters are ad-hoc
- No standard test set for evaluation

Issues

- Why not WordNet?
- Popescu & Etzioni (2005) uses syntactic patterns, morphological cues, and WordNet to discover aspects.
- Group opinion but not aspect words

Issues

- Lots of issues dealt with in Zhao et al. (2010)
  - Different generative model: fixed topic per sentence
  - A Bernoulli RV (for each word) controls whether the word is aspect or opinion
    - The Bernoulli RV is drawn from a pre-trained MaxEnt model with POS features ⇒ still useful for domain adaptation
  - Standard test set from Ganu et al. (2009)

Issues

- Why is there no comparison of results on *sentiment analysis*?
    - Only mentioned as “future work”

Conclusions

- Unsupervised joint models can accurately capture relationships between aspect words and topics nearing the accuracy of supervised models
- More work needs to be done in evaluation techniques (or explaining them)
- Big improvements since Hu & Liu (2004)