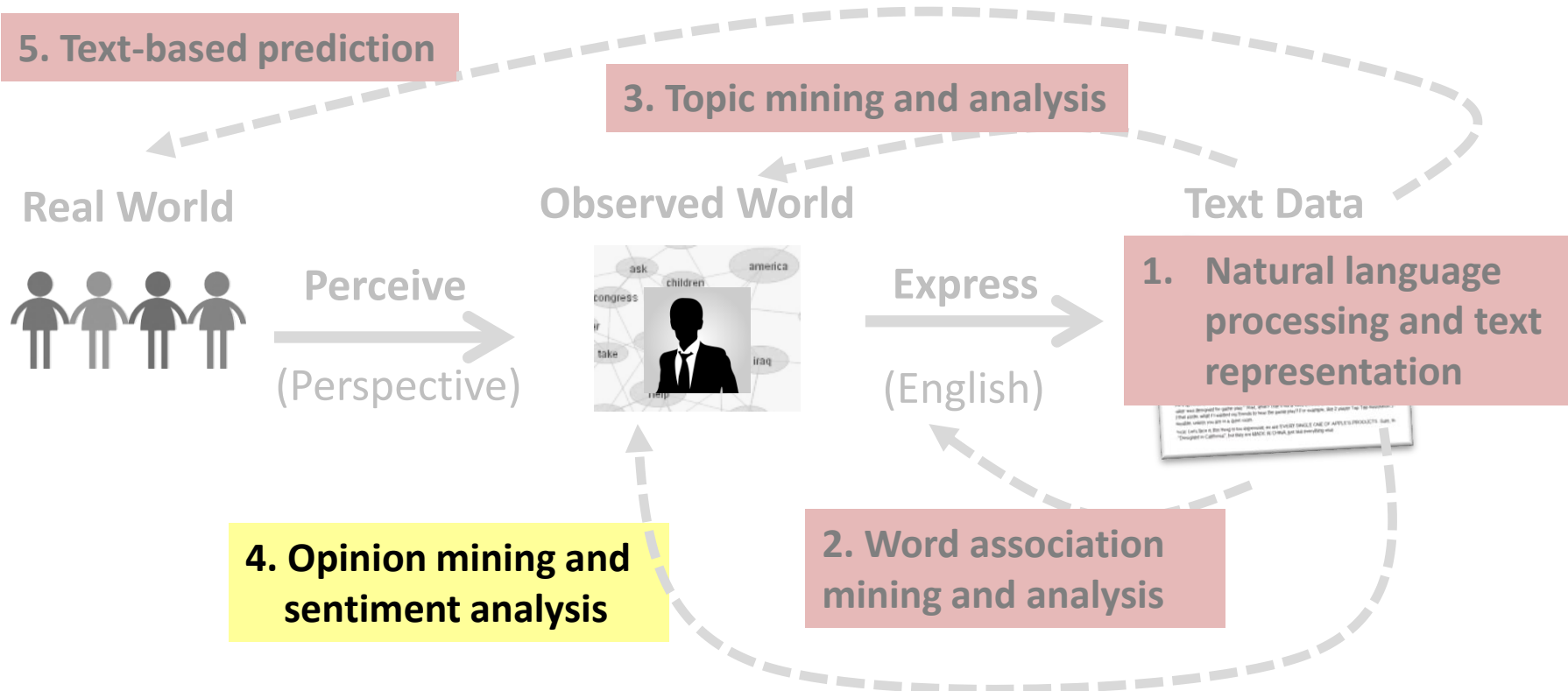


Opinion Mining and Sentiment Analysis: Latent Aspect Rating Analysis

Part 1

ChengXiang “Cheng” Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Opinion Mining and Sentiment Analysis: Latent Aspect Rating Analysis



Motivation

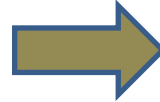
How to infer aspect ratings?

Hotel XYZ

Reviewer 1: ★★★★★

"Great location + spacious room = happy traveler"

Stayed for a weekend in July. Walked everywhere, enjoyed the comfy bed and quiet hallways....



★★★★★
★★★★★
★★★★★
★★★★★

Value
Rooms
Location
Service

Reviewer 2: ★★★★★

"Terrific service and gorgeous facility"

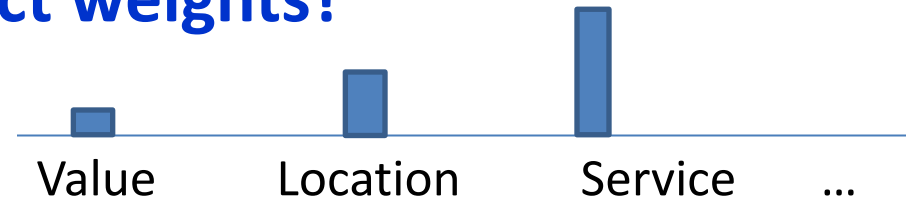
I stayed at the hotel with my young daughter for three nights June 17-20, 2010 and absolutely loved the hotel. The room was one of the nicest I've ever stayed in ...



★★★★★
★★★★★
★★★★★
★★★★★

Value
Rooms
Location
Service

How to infer aspect weights?



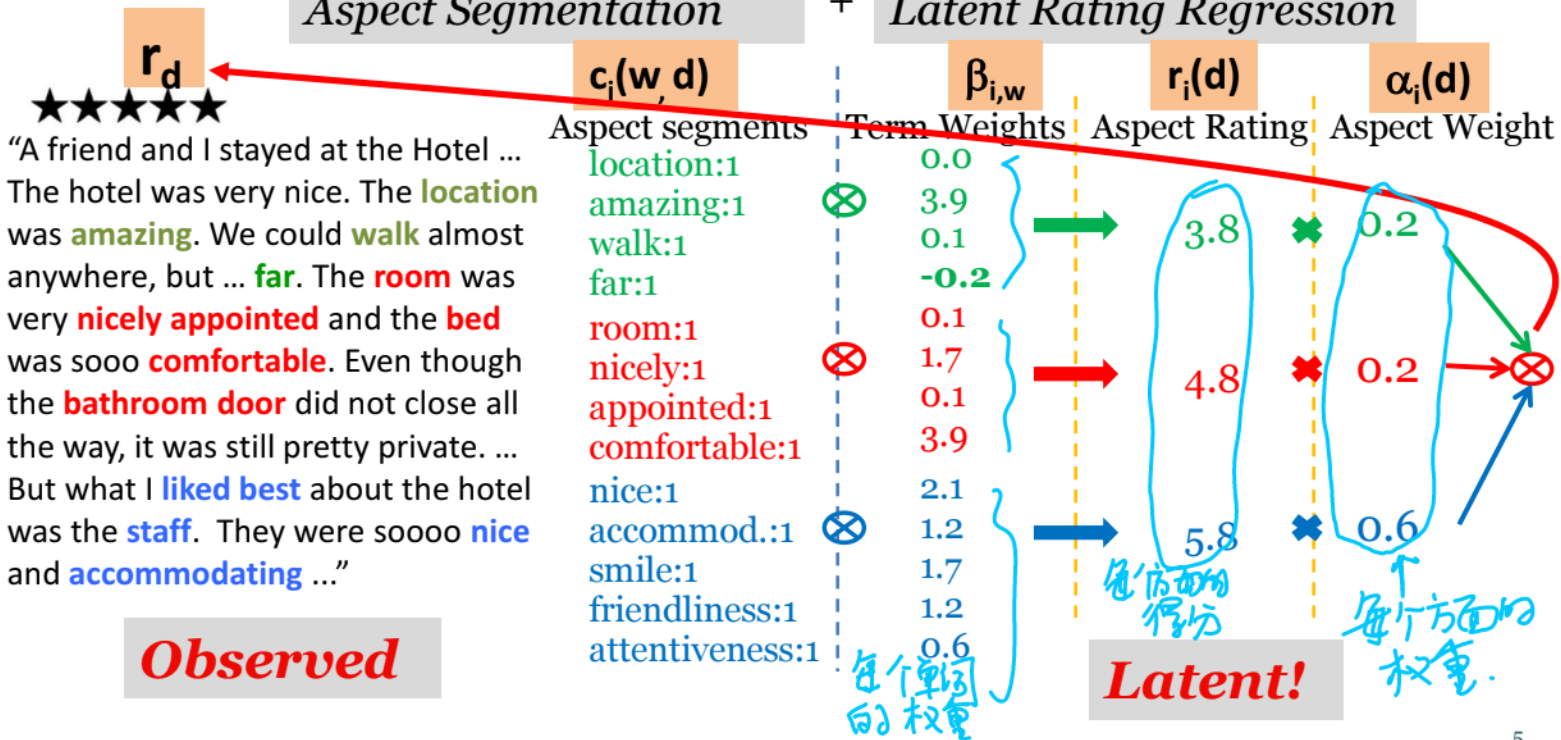
Latent Aspect Rating Analysis [Wang et al. 10]

- Given a set of review articles about a topic with overall ratings
- Output
 - Major aspects commented on in the reviews
 - Ratings on each aspect
 - Relative weights placed on different aspects by reviewers
- Many applications
 - Opinion-based entity ranking
 - Aspect-level opinion summarization
 - Reviewer preference analysis
 - Personalized recommendation of products
 - ...

Solving LARA in Two Stages

Aspect Segmentation

Latent Rating Regression



Latent Rating Regression [Wang et al. 10]

- Data: a set of review documents with overall ratings: $C=\{(d, r_d)\}$
 - d is pre-segmented into k aspect segments
 - $c_i(w, d)$ = count of word w in aspect segment i (zero if w didn't occur)
- Model: predict rating based on d : $p(r_d | d)$

Overall Rating = Weighted Average of Aspect Ratings

Multivariate Gaussian Prior

正态分布

$$r_d \sim N(\sum_{i=1}^k \alpha_i(d) r_i(d), \delta^2),$$

另一个正态分布

$$\bar{\alpha}(d) \sim N(\bar{\mu}, \Sigma)$$

$$r_i(d) = \sum_{w \in V} c_i(w, d) \beta_{i,w}$$

$$\beta_{i,w} \in \mathcal{R}$$

Aspect-Specific Sentiment of w

Aspect Rating = Sum of sentiment weights of words in the aspect

Latent Rating Regression (cont.)

- Maximum Likelihood Estimate

- Parameters: $\Lambda = (\{\beta_{i,w}\}, \bar{\mu}, \Sigma, \delta^2)$

- ML estimate: $\Lambda^* = \arg \max_{\Lambda} \prod_{d \in C} p(r_d | d, \Lambda)$

- Aspect Rating for aspect i

$$r_i(d) = \sum_{w \in V} c_i(w, d) \beta_{i,w}$$

$c_i(w, d) = 0$ for words not occurring in aspect segment i

- Aspect Weights: $\alpha_i(d)$ = weight on aspect i

$$\bar{\alpha}(d)^* = \arg \max_{\bar{\alpha}(d)} p(\bar{\alpha}(d) | \mu, \Sigma) p(r_d | d, \{\beta_{i,w}\}, \delta^2, \bar{\alpha}(d))$$

Maximum a Posteriori

Prior

Likelihood

1684

Suggested Reading

- [Wang et al. 10] Hongning Wang, Yue Lu, and ChengXiang Zhai, Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of ACM KDD 2010*, pp. 783-792, 2010. DOI=10.1145/1835804.1835903