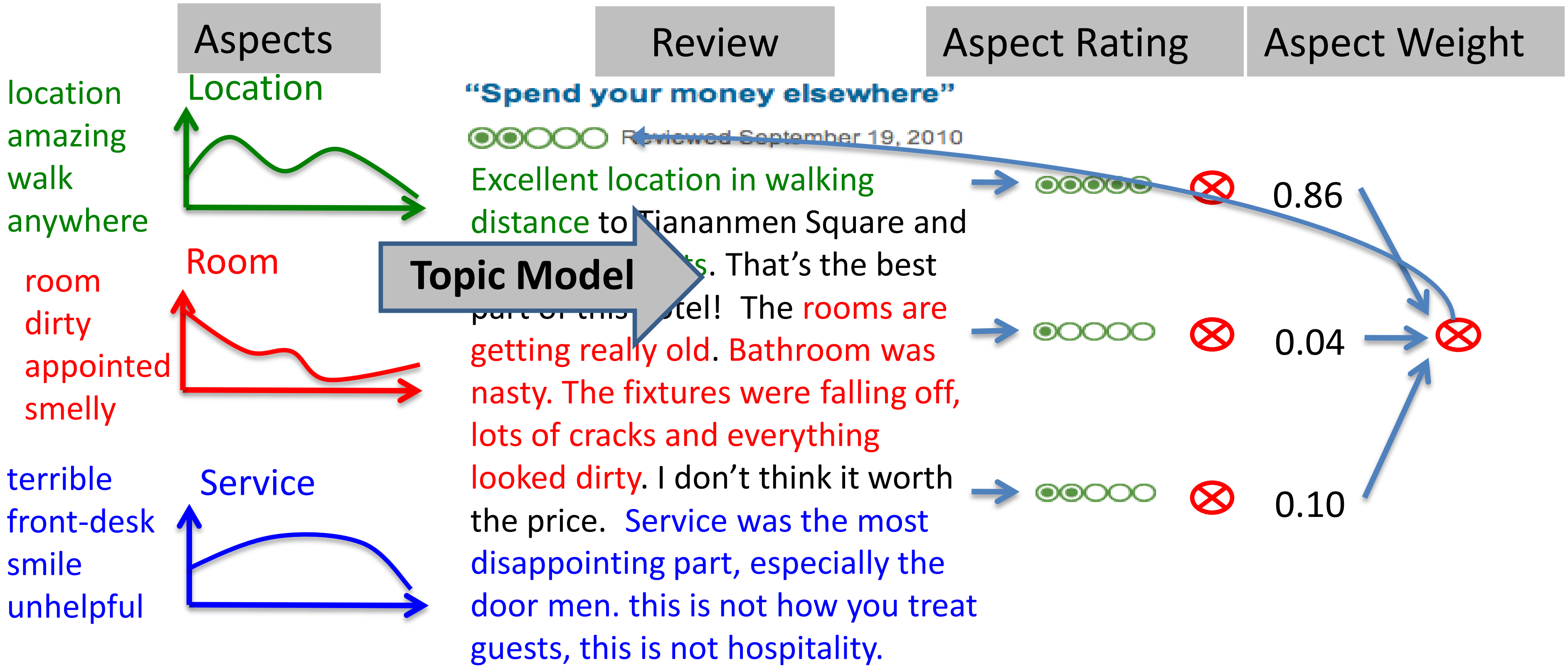
Opinion Mining and Sentiment Analysis: Latent Aspect Rating Analysis

Part 2

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

A Unified Generative Model for LARA [Wang et al. 11]

Any Entity



Sample Result 1: Rating Decomposition [Wang et al. 10]

- Hotels with the same overall rating but different aspect ratings

(All 5 Stars hotels, ground-truth in parenthesis)

总分相同，各个aspect都不同。人的需求不一样

Hotel	*Value*	*Room*	*Location*	*Cleanliness*
HOTEL 1	4.2(4.7)	3.8(3.1)	4.0(4.2)	4.1(4.2)
HOTEL 2	4.3(4.0)	3.9(3.3)	3.7(3.1)	4.2(4.7)
HOTEL 3	3.7(3.8)	4.4(3.8)	4.1(4.9)	4.5(4.8)

- Reveal detailed opinions at the aspect level

Sample Result 2: Comparison of Reviewers [Wang et al. 10]

- Per-Reviewer Analysis
  - Different reviewers' ratings on the same hotel

Reviewer	*Value*	*Room*	*Location*	*Cleanliness*
Reviewer 1	3.7(4.0)	3.5(4.0)	3.7(4.0)	5.8(5.0)
Reviewer 2	5.0(5.0)	3.0(3.0)	5.0(4.0)	3.5(4.0)

  - Reveal differences in opinions of different reviewers

Sample Result 3: Aspect-Specific Sentiment Lexicon [Wang et al. 10]

Value	*Rooms*	*Location*	*Cleanliness*
resort 22.80	view 28.05	restaurant 24.47	clean 55.35
value 19.64	comfortable 23.15	walk 18.89	smell 14.38
excellent 19.54	modern 15.82	bus 14.32	linen 14.25
worth 19.20	quiet 15.37	beach 14.11	maintain 13.51
bad -24.09	*carpet -9.88*	*wall -11.70*	*smelly -0.53*
money -11.02	*smell -8.83*	*bad -5.40*	*urine -0.43*
terrible -10.01	*dirty -7.85*	*road -2.90*	*filthy -0.42*
overprice -9.06	*stain -5.85*	*website -1.67*	*dingy -0.38*

Learn sentimental information directly from the data.

Sample Result 4: Validating Preference Weights [Wang et al. 10]

Top-10: Reviewers with the highest Val/X ratio (emphasize "value")

Bot-10: Reviewers with the lowest Val/X ratio (emphasize a non-value aspect)

City	*Avg. Price*	*Group*	*Val/Loc*	*Val/Rm*	*Val/Ser*
Amsterdam	241.6	top-10	190.7	214.9	221.1
		bot-10	**270.8**	**333.9**	**236.2**
San Francisco	261.3	top-10	214.5	249.0	225.3
		bot-10	**321.1**	**311.1**	**311.4**
Florence	272.1	top-10	269.4	248.9	220.3
		bot-10	**298.9**	**293.4**	**292.6**

Higher!

Application 1: Rated Aspect Summarization

Aspect	*Summary*	*Rating*
Value	Truly unique character and a great location at a reasonable price Hotel Max was an excellent choice for our recent three night stay in Seattle.	**3.1**
	Overall not a negative experience; however, considering that the hotel industry is very much in the impressing business, there was a lot of room for improvement.	**1.7**
Location	The location, a short walk to downtown and Pike Place market, made the hotel a good choice.	**3.7**
	When you visit a big metropolitan city, be prepared to hear a little traffic outside!	**1.2**
Business Service	You can pay for wireless by the day or use the complimentary Internet in the business center behind the lobby, though.	**2.7**
	My only complaint is the daily charge for Internet access when you can pretty much connect to wireless on the streets anymore.	**0.9**

Application 2: Discover Consumer Preferences [Wang et al. 2011]

- Amazon reviews: No guidance

Table 2: Topical Aspects Learned on MP3 Reviews

Low Overall Ratings			High Overall Ratings		
unit	jack	service	files	player	vision
usb	headphone	charge	format	music	video
battery	warranty	problem	included	download	player
charger	replacement	support	easy	headphones	quality
reset	problem	hours	convert	button	great
time	player	months	mp3	set	product
hours	back	weeks	videos	hours	sound
work	months	back	file	buds	radio
thing	buy	customer	wall	volume	accessory
wall	amazon	time	hours	ear	fm
battery life	accessory	service	file format	volume	video

Application 3: User Rating Behavior Analysis [Wang et al. 10]

	Expensive Hotel		*Cheap Hotel*	
	5 Stars	*3 Stars*	*5 Stars*	*1 Star*
Value	0.134	0.148	**0.171**	0.093
Room	0.098	0.162	0.126	0.121
Location	0.171	0.074	0.161	0.082
Cleanliness	0.081	**0.163**	0.116	**0.294**
Service	**0.251**	0.101	0.101	0.049

People like expensive hotels because of good service.

People like cheap hotels because of good value.

Application 4: Personalized Ranking of Entities [Wang et al. 10]

Query: 0.9 value
0.1 others

Non-personalized

Personalized

(Query-specific)

	Hotel	Overall Rating	Price	Location
Approach 1	Majestic Colonial	5.0	339	Punta Cana
	Agua Resort	5.0	753	Punta Cana
	Majestic Elegance	5.0	537	Punta Cana
	Grand Palladium	5.0	277	Punta Cana
	Iberostar	5.0	157	Punta Cana
Approach 2	Elan Hotel Modern	5.0	216	Los Angeles
	Marriott San Juan Resort	4.0	354	San Juan
	Punta Cana Club	5.0	409	Punta Cana
	Comfort Inn	5.0	155	Boston
	Hotel Commonwealth	4.5	313	Boston

Summary of Opinion Mining

- Very important with a lot of applications!
- Sentiment analysis can be done using text categorization techniques
  - With enriched feature representation
  - With consideration of ordering of the categories
- Generative models are powerful for mining latent user preferences
- Most approaches were proposed for product reviews
- Opinion mining from news and social media remains challenging

Suggested Reading

- Bing Liu, *Sentiment analysis and opinion mining*, Morgan & Claypool Publishers, 2012.
- Bo Pang and Lillian Lee, Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval* 2(1-2), pp. 1–135, 2008.
- 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
- Hongning Wang, Yue Lu, and ChengXiang Zhai. 2011. Latent aspect rating analysis without aspect keyword supervision. In *Proceedings of ACM KDD 2011*, pp. 618-626. DOI=10.1145/2020408.2020505