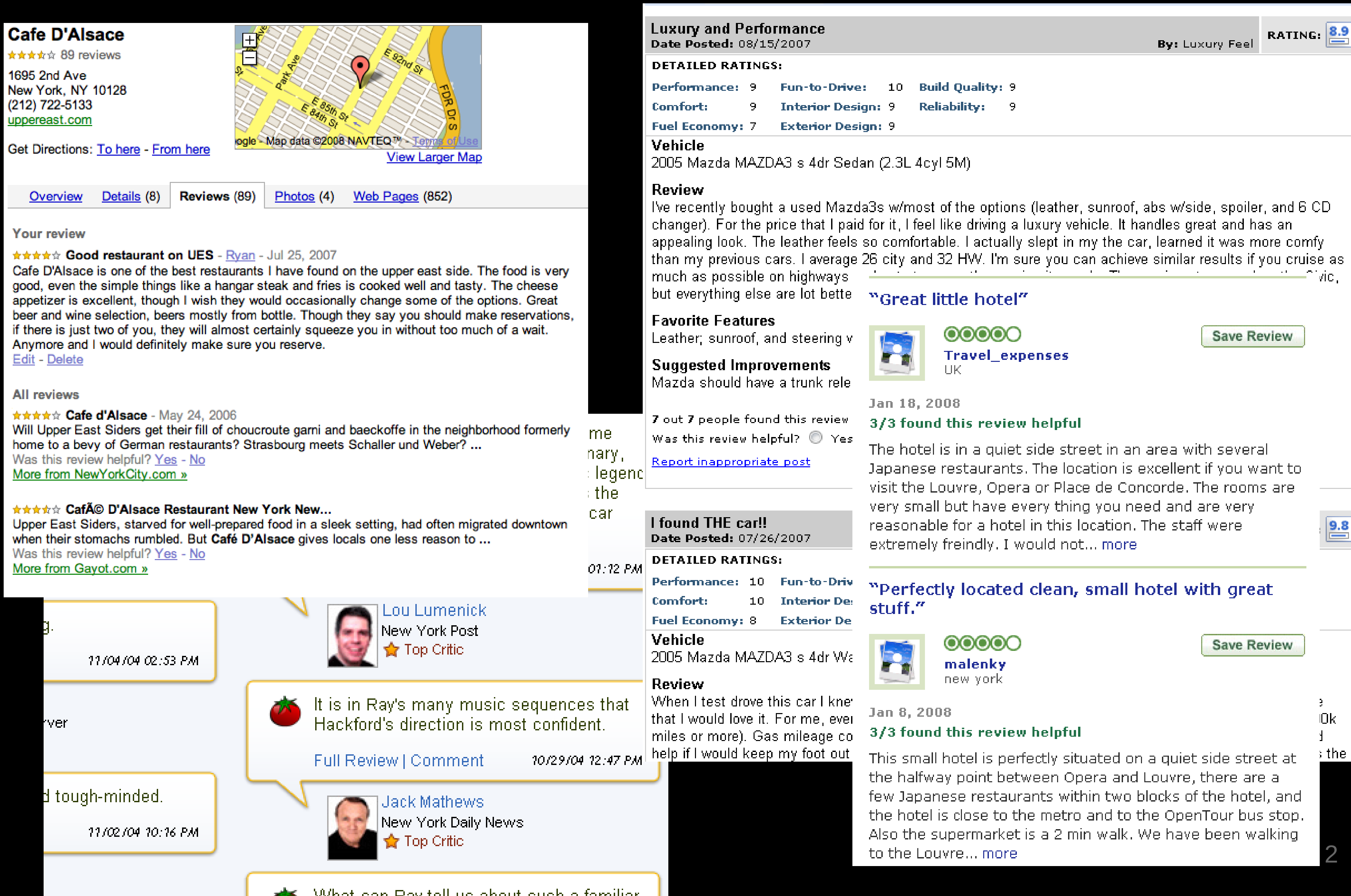
Modeling Online Reviews with Multi-grain Topic Models

Ivan Titov[1] and Ryan McDonald[2]

[1]University of Illinois at Urbana-Champaign
[2]Google Inc.

Online User Reviews



Sentiment Classification

- Sentiment and Opinion Classification (Wiebe, 00), (Pang et al, 02), (Turney, 02)
  - However, sentiment is often provided by the user
    - Isn’t this just a contrived task?
  - Not always:
    - Train sentiment classifiers on reviews use it 4 blogs
    - Train review ratings, apply on phrases or sentences
    - Train on one blog, apply to unannotated blogs

Sentiment Summarization

- Take a set of reviews for an entity and summarize them
- Aspect-based summarization (Hu & Liu 2004)
  - Summarize along key aspects

Nikos' Fine Dining

Food	4/5	"Best fish in the city", "Excellent appetizers"
Decor	3/5	"Cozy with an old world feel", "Too dark"
Service	1/5	"Our waitress was rude", "Awful service"
Value	5/5	"Good Greek food for the $", "Great price!"

- Many real world manual examples, e.g., Zagat.com

Text Segmentation

- Focus on models for segmenting text

The chicken was great. On top of that our service was excellent and the price was right. Can't wait to go back!
We went there for our anniversary. My soup was cold and expensive plus it felt like they hadn't painted since 1980.
The food is only mediocre, but well worth the cost. Wait staff was friendly. Lot's of fun decorations.

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"

Previous Work

- Hu and Liu ’04
  - Aspect-based summarization
  - String-based aspects + lexicon sentiment
- Popescu and Etzioni ’05: Opine system
- Gamon et al. ’05
  - Aspect clusters: use most frequent word label
- Carenini ’06
  - String-based + ontologies
- Mei et al. ’07
  - Generative topic-sentiment models (at document level)

Three Tasks

- Identify Aspects
  - Often we know this (pros-cons, tech specs, ontologies)
- Extract Mentions
  - We always have to do this
- Aggregate Sentiment
  - Again, we often know this (star ratings, eg, TripAdvisor)

Nikos' Fine Dining

Food	4/5	"Best fish in the city", "Excellent appetizers"
Decor	3/5	"Cozy with an old world feel", "Too dark"
Service	1/5	"Our waitress was rude", "Awful service"
Value	5/5	"Good Greek food for the $", "Great price!"

Aspect Identification and Extraction

Nikos' Fine Dining

Food	4/5	"Best fish in the city", "Excellent appetizers"
Decor	3/5	"Cozy with an old world feel", "Too dark"
Service	1/5	"Our waitress was rude", "Awful service"
Value	5/5	"Good Greek food for the $", "Great price!"

- Common method: String-based extraction
  - Find frequently occurring nouns that are modified by opinion words
  - Take top K as relevant aspects
  - Extract all sentences / phrases that match
  - Problem: Get a long list of aspects w/ no clustering

Aspect Identification and Extraction

- String-based example: restaurants
  - Is list really summarization?
  - How far down to get “cozy”, “fish”, “$”, “waitress”, “dark”?
- We really want to cluster these

General
Food
Ambiance
Service
Value

food
place
pizza
service
restaurant
atmosphere
time
wine
meal
prices
value
sauce
hour
price
dim
selection
experience
crust
dining
ingredients

Nikos’ Fine Dining

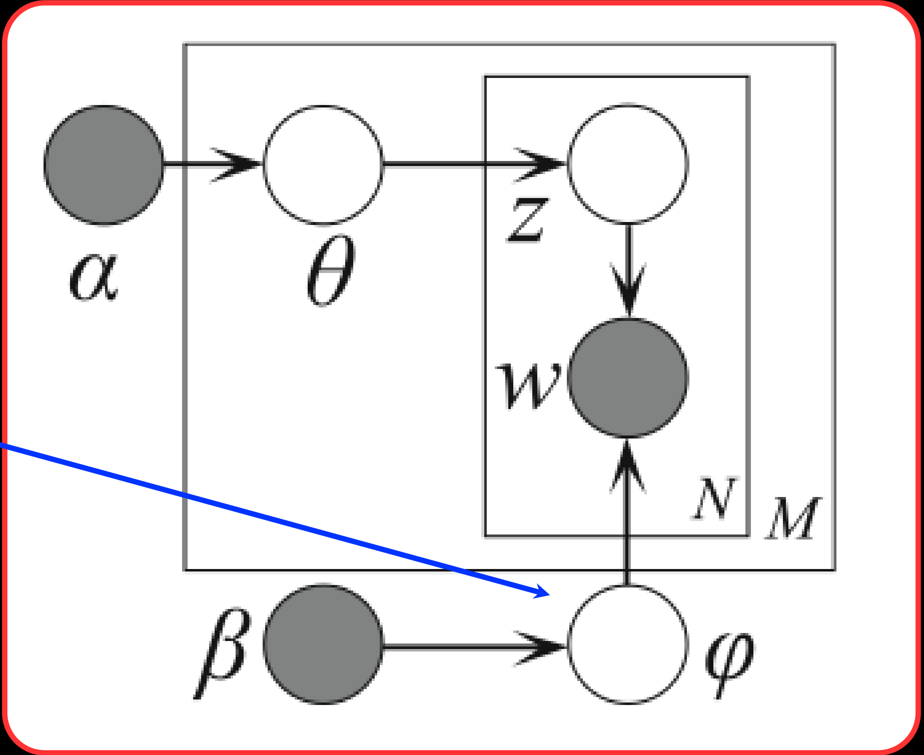
Food	4/5	“Best fish in the city”, “Excellent appetizers”
Decor	3/5	“Cozy with an old world feel”, “Too dark”
Service	1/5	“Our waitress was rude”, “Awful service”
Value	5/5	“Good Greek food for the $”, “Great price!”

Topic Models

- Studied in ML and Data Mining
  - LSA, PLSA, LDA, Pachinko Allocation, ...
- Build semantic “topics” of data collections
  - e.g., newsgroups into “religion”, “politics”, “science”, ...
- Simple hypothesis
  - Topics in reviews correspond to clustered aspects
- We will focus on LDA type models (Blei et al. ’03)
  - Others produce similar observations

LDA

- Generative model of text
- Sample multinomial word distributions for each topic
- The for each document d:



- choose distribution of topics $\theta_d \sim Dir(\alpha)$
- for each word i in document d
  - choose topic $z_{d,i} \sim \theta_d$,
  - choose word $w_{d,i} \sim \varphi_{z_{d,i}}$.

Side Note: Inference

- All methods use collapsed Gibbs (Griffiths & Steyvers ’04)
- A sample from the chain used to approx:
    - Distribution of words in topics
    - Distribution of topics in text fragments
- We tried variational techniques, but they didn’t work for our models
- More details in the paper

LDA

- Problem with LDA (and most other topic models)
  - Co-occurrences modeled at document level
  - Topics are about instances not aspects
    - e.g., iPod versus Creative Labs
  - Often clusters are meaningless

(Service??) Topic 0: product player did support bought work unit problem $
(Creative Labs) Topic 1: gigabeat deleted waiting jukebox creative playback
(iPod) Topic 11: ipod apple mac firewire dock itunes x display aac

Most topics are incoherent. Only 4 out of first 40 can be viewed as aspects.

LDA

- Simple solutions: LDA over sentences
  - Co-occurrence counts too sparse
  - Can use sliding window, but results look like LDA
  - Still can't distinguish aspect topics from the rest
- Another solution: Multi-grain topic models
  - Model local topics (aspects) and global topics (types)
  - Creates a bottleneck for local topics
  - Words generated from sliding windows

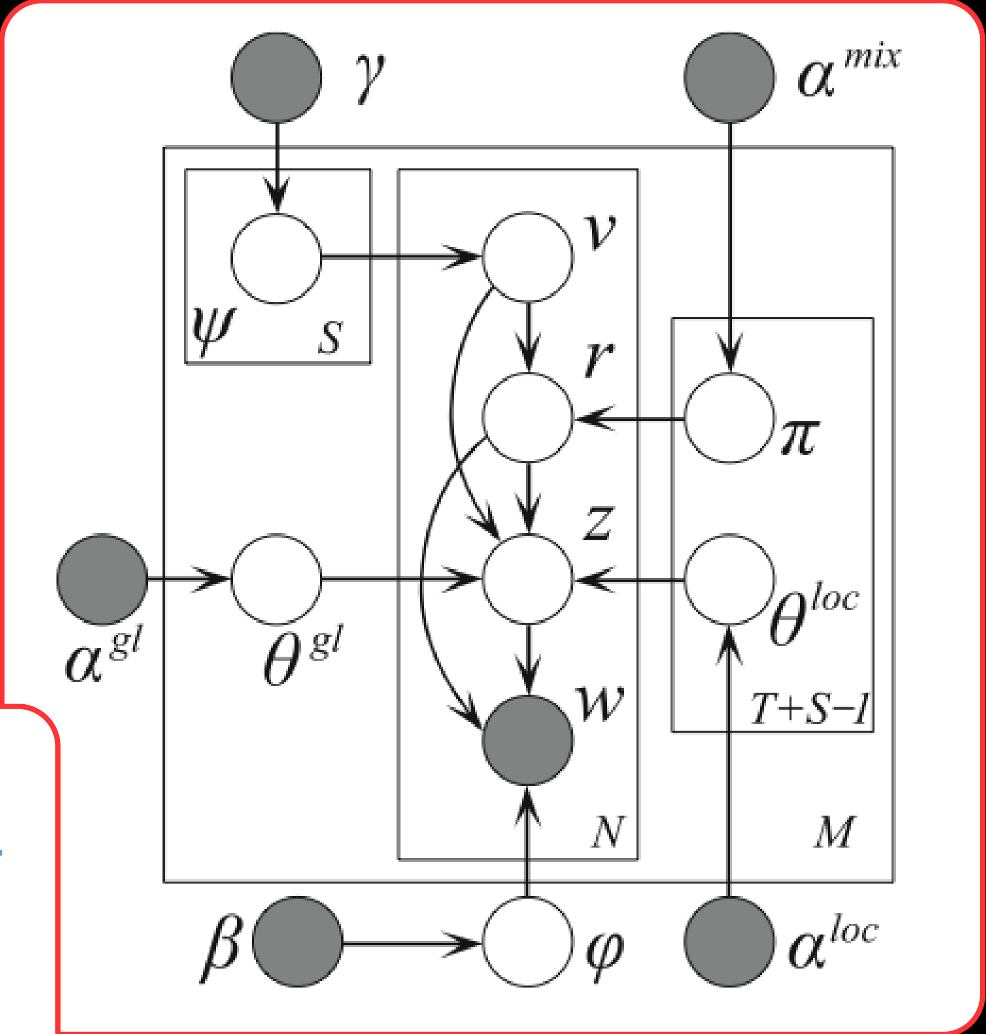
Varying Granularity

"... public transport in London is straightforward, the tube station is about an 8 minute walk … or you can get a bus for £ 1.50 ....
We had a stunning view (from the floor to ceiling window) of the Tower and the Thames."

- Global topic London: *London, tube, £ , Tower, Thames*
  - Global topic dist is assigned to the document
- Local topics:
  - Location: *transport, station, walk, bus, minute*
  - View: *view, window*
  - Local topic dist is assigned to current sliding window

MG-LDA

- Draw global topic word dist.
- Draw local topic word dist.
- For each document d:



- Choose a distribution of global topics $\theta_d^{gl} \sim Dir(\alpha^{gl})$.
- For each sentence s choose a distribution $\psi_{d,s}(v) \sim Dir(\gamma)$.
- 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 $\varphi_{z_{d,i}}^{r_{d,i}}$.

Multi-Grain models

- For local topics instead of sliding windows we could use topical ngrams (Wang and McCallum 05, Wallach 06)
- But topical ngrams are computationally expensive
- Efficiency of MG-LDA is comparable with that of LDA
- Crucial property is that the topic distributions are associated with different scopes in a text

Evaluation

- Qualitative comparison of LDA and MG-LDA topics
- Quantitative comparison: improving multi-aspect ranking by using information about topic distribution

Qualitative Evaluation

- Data: 3,872 reviews of MP3 players, 32,861 reviews of hotels and 32,563 reviews of restaurants (from Google Product and Local Search)
- Stop words and punctuation are removed

MG-LDA Topics (Mp3)

First 8 Local Topics!!

Sound Quality	Features	PC Connection	Tech Problems	Looks	Controls	Battery	Accessories
sound	games	usb	reset	case	button	battery	usb
quality	features	pc	noise	pocket	play	hours	cable
headphones	clock	windows	backlight	silver	track	life	headphones
volume	contacts	port	slow	screen	menu	batteries	adapter
bass	calendar	transfer	freeze	plastic	song	charge	remote
earphones	alarm	computer	turn	clip	buttons	aaa	plug
ear	notes	mac	remove	easily	volume	rechargeable	power
rock	game	software	playing	small	album	time	charger
settings	quiz	cable	hot	blue	tracks	power	included

First 4 Global Topics

iPod	Creative Zen	Sony Walkman	Video Players
ipod	zen	sony	video
music	creative Zen	walkman	screen
apple	micro	memory	videos
songs	touch	stick	device
use	xtra	sonicstage	photos
mini	pad	players	tv
very	nomad	atrac3	archos
just	waiting	mb	pictures
itunes	labs	atrac	camera

LDA
- 40 topics
- Only 4 aspect topics
- A couple other coherent topics
- Good topics in no order
- Mostly junk topics

MG-LDA: first 8 Local Topics!!

MG-LDA vs. LDA (Hotels)

Amenities	Food/Drink	Noise/AC	Bathroom	Breakfast	Spa	Parking	Staff
coffee	food	air	shower	breakfast	pool	parking	staff
microwave	restaurant	noise	water	coffee	area	car	friendly
fridge	bar	door	bathroom	continental	hot	park	helpful
tv	good	room	hot	morning	tub	lot	very
ice	dinner	hear	towels	fruit	indoor	valet	desk
room	service	open	toilet	fresh	nice	garage	extremely
refrigerator	breakfast	night	tub	buffet	swimming	free	help
machine	ate	conditioning	bath	included	outdoor	street	directions
kitchen	eat	loud	sink	free	fitness	parked	courteous

LDA: examples of LDA topics (out of 45), only 9 aspect topics

Beach resorts	Las Vegas	?	Smells/ stains?	Getting there	Breakfast	Front desk	Opinion
beach	vegas	motel	room	airport	breakfast	room	hotel
great	strip	rooms	did	hotel	coffee	hotel	best
pool	great	nice	smoking	shuttle	fruit	told	stay
very	casino	hotel	bed	bus	room	desk	hotels
place	$	like	night	very	juice	did	stayed
ocean	good	place	went	minutes	fresh	manager	reviews
stay	hotel	stay	like	flight	eggs	asked	service
view	food	parking	desk	hour	continental	said	great
just	las	price	smoke	free	very	service	time

Local LDA

- LDA applied on individual sentences ('Local' LDA) on MP3 reviews:
  - Infers a number of valid aspects
  - Still many are related to brands of MP3 players

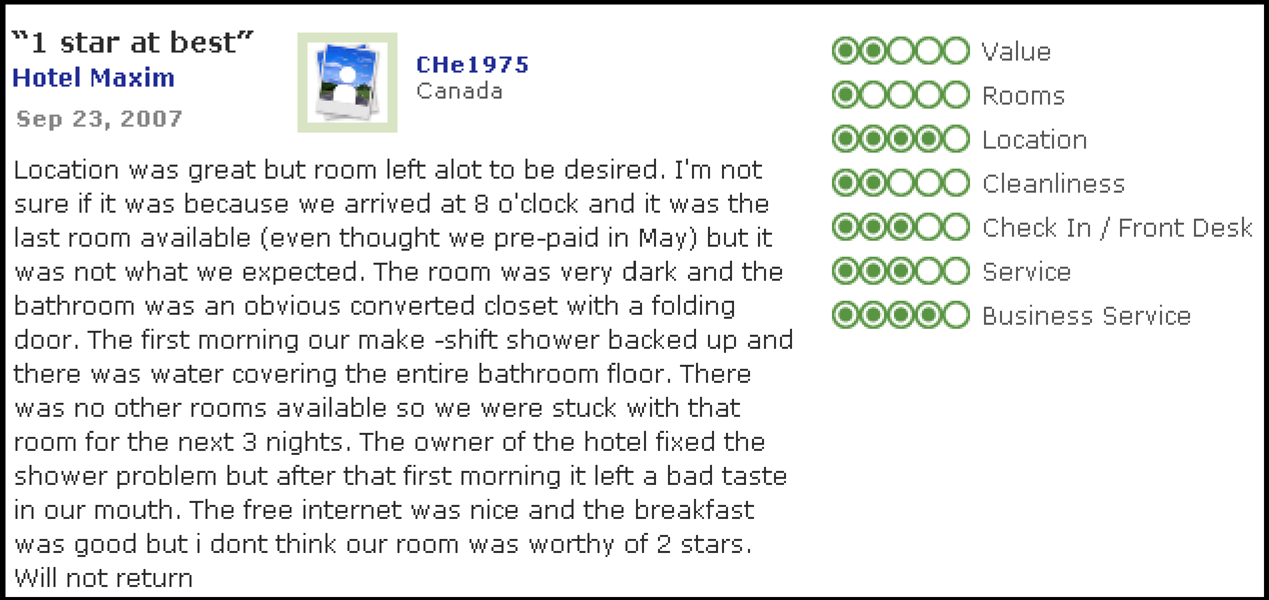
    20 top words for a half of topics contained brand names ('ipod', 'sony', 'yepp',...)
- Simultaneous modeling of both local and global topics is important for discovery of coherent aspects

Restaurant Reviews

- Restaurant reviews are challenging, only the following aspects are inferred:
  - MG-LDA : *service*, *location*, *atmosphere* and *decor*
  - LDA: *waiting time* and *location*
- Challenge: majority of aspects are specific for a type of restaurants: *pizza* and *pasta* for a Italian rest., *sushi* and *noodles* for Japanese
- Good results with MG-LDA when applied to a specific restaurant type (9 aspects are inferred for Italian).
- Hierarchical topic modeling could address this problem

Quantitative Evaluation

- We wish to evaluate how well the learned topics correspond to aspects that users typically rate
- Multi-aspect opinion rating (Snyder and Barzilay,07):
  - predicting a rating for multiple aspects of an object
- Data: 27,564 reviews labeled reviews from TripAdvisor



Using a topic model

- Information about a topic of a sentence should help a classifier:
  - “*The X was great*”,
    - X – is “*duck*”, “*steak*”, “*soup*” - food rating is high
    - X – is “*music*”, “*light*” – atmosphere rating is high
- We extended standard ngram features of text x by adding top topics and their bucketed probabilities:
  - *x contains “great” & topic = 3 & bucket = 0.4-0.5*

Results

- Ranking loss: $\sum_n \frac{|\text{actual\_rating}_n - \text{predicted\_rating}_n|}{N}$

  (lower – better)

- PRank (Crammer and Singer, 02) – perceptron for ranking

Unigram features only

Model	Overall	Check-in	Service	Value	Location	Rooms	Cleanliness
Baseline	1.118	1.126	1.208	1.272	0.742	1.356	1.002
PRank	0.774	0.831	0.799	0.793	0.707	0.798	0.715
PRank + LDA	0.735	0.786	0.762	0.749	0.677	0.746	0.690
PRank + MG-LDA	**0.706**	**0.748**	**0.731**	**0.725**	**0.635**	**0.719**	**0.676**

Unigram, bigram and trigram features

Model	Overall	Check-in	Service	Value	Location	Rooms	Cleanliness
PRank	0.689	0.735	0.725	0.710	0.627	0.700	0.637
PRank + LDA	0.682	0.728	0.717	0.705	0.620	0.684	0.637
PRank + MG-LDA	**0.669**	**0.717**	**0.700**	**0.696**	**0.607**	**0.672**	0.636

Summary

- We demonstrated that:
  - topic models are appropriate for discovery of aspects in user reviews
  - simultaneous modeling of both local and global topics is important for discovery of coherent ratable aspects

Future/Recent work

- Integrating user annotation for topic discovery:
  - (Titov and McDonald, ACL 2008)
- Hierarchical topic modeling
  - E.g, infer a hierarchy representing cuisines and with each cuisine extract cuisine specific aspects
- Combining aspect extraction with sentiment classification:
  - Examining joint models of topics and sentiment (Mei et al, 07)