



Airbnb Review Summarizer

By: Clark Li

Business Case

Motivation: The inconvenience of reading through large number of reviews calls for a simpler way to highlight key aspects only relevant to user needs

Business Case

Offered Solutions:

Goal #1:

- Provide a high level summary of pros/cons around the most frequently appearing aspects

Goal #2:

- Allow users to filter the aspects and focus only on what they most care about

Methodology

- Aspect Detection
 - Tokenization, Part-of-Speech Tagging, Associative Rule mining, Dependency Parsing
- Opinion Mining
 - Sentiment analysis with polarity scores for each aspect



Use Case 1: Top Aspects

Private Room in Youth Hostel. # of Reviews: 66. Rating: 69

→ Room

- ◆ **Positive:** 20 -> decent, very simple clean, safe enough, accurate, more spacious, etc
- ◆ **Negative:** 1 -> small

→ Host

- ◆ **Positive:** 13 -> very attentive, very communicative, even available late, quick, friendly, etc
- ◆ **Negative:** 0

→ Place

- ◆ **Positive:** 10 -> safe, nice, good, great, clean, cheap
- ◆ **Negative:** 0

→ Neighborhood

- ◆ **Positive:** 2 -> comfortable, best
- ◆ **Negative:** 8 -> sketchy, quite creepy, dangerous, kinda rough, not very clean, tough, scary

→ Bathroom

- ◆ **Positive:** 6 -> enough clean, adequate, Shared spartan, extremely clean, very private clean
- ◆ **Negative:** 2 -> dirty, very dirty

Use Case 2: Aspect of Interest

SF cottage-easy parking & transport. # of Reviews: 286; Rating: 97

Aspects	Positive	Negative	Total
host	131	4	135
cottage	96	13	109
home	88	4	92
view	48	0	48
stay	40	0	40
full	31	1	32
garden	27	2	29
parking	20	6	26
time	20	3	21
location	20	0	20
neighborhood	15	2	17

Parking

Positive: 20 -> easy, free, very convenient, always adequate, readily available, easy quick, etc

Negative: 6 -> hard, difficult, how difficult, etc











Source Reviews:

It is located close enough to the city area and parking is very convenient

At night, spots were full, so parking could be difficult to find.

Validation Web App

Aspect Mining Labeling

Listing_id	Review_id	Review_sentence	Aspect #1	Sentiment Orientation #1	Aspect #2	Sentiment Orientation #2
311259	960520	This cottage is a gem!	<input type="text" value="cottage"/>	Positive 	<input type="text"/>	Neutral 
	960520	It was everything I needed for my short stay in San Francisco for a conference.	<input type="text"/>	Neutral 	<input type="text"/>	Neutral 
	960520	I made a last minute booking – the host was really nice, allowing me to check in early.	<input type="text" value="host"/>	Positive 	<input type="text"/>	Neutral 
	960520	It was easy to find, and I found parking immediately outside the place.	<input type="text" value="parking"/>	Positive 	<input type="text"/>	Neutral 
	960520	I got in using a code for the lock that the host provided.	<input type="text"/>	Neutral 	<input type="text"/>	Neutral 

Validation

Approach:

- Manually labeled 100 reviews consisting of 281 aspects given the constraint of time

Aspect Detection scores:

accuracy: 0.43

precision: 0.60

recall: 0.60

Sentiment scores:

accuracy: 0.91

Future Work

- Expand validation dataset by manually labeling through the validation pipeline
- Identify hidden aspects that are not explicitly defined in the text
- Inspect more complex sentence structures for sentiment analysis

Thank You

Questions?