

A Framework for Aspect and Sentiment Extraction for Online Review

Prepared By:
Noor Rizvana Binti Ahamed Kabeer

Introduction



What others think has always been an important information for us in making decision.

“Which hotel should I stay?”



“How good is the product that I sell recently?”



“Whom to Ask for Opinion?”

Before Web



- Friends
- Reports
- Catalogues
- Experts

After Web



- Social networks (Facebook, Twitter)
- Blogs (Google blogs, Personal blogs)
- Review sites (TripAdvisor, CNET)
- E-commerce sites (Amazon, eBay)



“I Have a Lot of Reviews to Refer”

I have a lot of information on 1 matter/issue
(product / service/ person), now I can easily make decision.

TRUE

FALSE



“Not Really”

- ❑ Difficult to search reviews to find particular feature of a product / services.
- ❑ Eg: “Room in Hotel A vs room in hotel B”



“Not Really”

❑ Vast amount of information.

- Time consuming to read through each reviews.

- Reviews are written in different ways.

❑ Eg: “Good space in rooms but slightly old, but perfectly satisfactory style.”

❑ Eg: “Room is not too big and not too small either.”

❑ Eg: “Please don't waste your money on this hotel.”



“So How to Manage These Problems?”

“Have You Heard Of?”



Opinion
Mining

“Have You Heard OF?”



Aspect /
Feat ur es
and
Op i ni on/
Sent i ment
Ext r act i on

Aspect and Opinion



Hotel

Entity

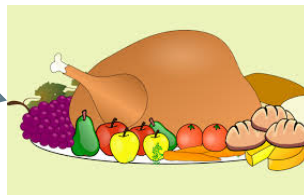
Room



Pool



Food

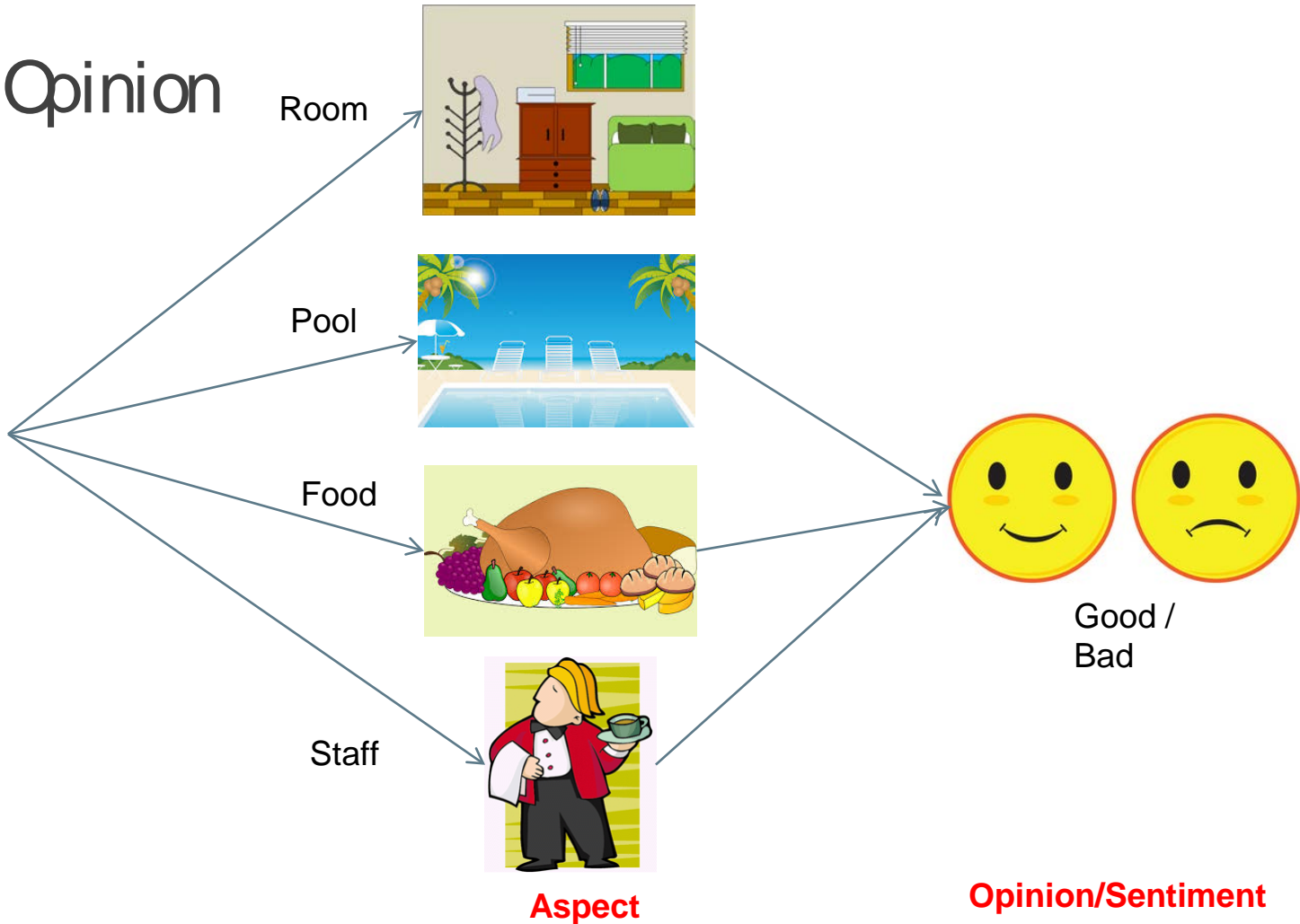


Staff



Good /
Bad

Opinion/Sentiment



Current Approaches

Sentence 1: Its a smoke free hotel.

POS Tags: Its_**PRP\$** a_**DT** smoke_**NN** free_**JJ** hotel_**NN** .**.**

Approach : Rule + pruning strategy + frequent aspect

Frequent noun

Nearest
adjectives

Smoke

Free

Hotel

Aspect

Opinion

Errors :

Wrong opinion were assigned to aspect

Should be :

Smoke



Free

Opinion

Hotel

Aspect

Hu, M, & Liu, B. (2004) &
Hu, M, Hu, M, Liu, B., & Liu, B. (2004)

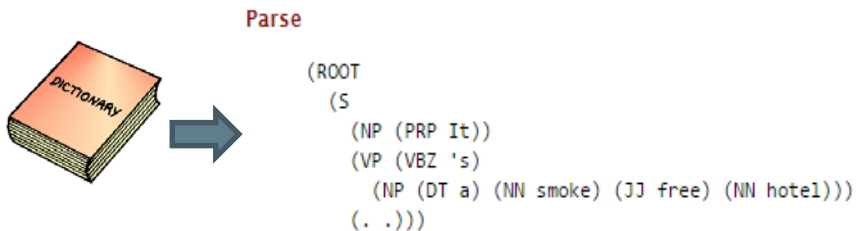
Problem Not only noun can aspect and adjective can be opinion but mixture of both can be aspect or opinion.

Relationship between aspect and opinion assigned wrongly.

Sentence 1: It a smoke free hotel.

POS Tags: Its_**PRP\$** a_**DT** smoke_**NN** free_**JJ** hotel_**NN** .**_.**

Approach : Lexicon + dependency parsing + rule



Aspect

Hotel

Smoke

Errors :

Lexicon based method inadequate to identify aspect that are new terminology

Opinion were ignored.

Should be :

Hotel

Aspect

Smoke

Free

Opinion

Veselovská, K., & Tamchyna, a. (2014)

Problem New terms like “App”, ‘iPhone’ and etc. needed to be added to lexicon consistantly for dictionary to be updated.

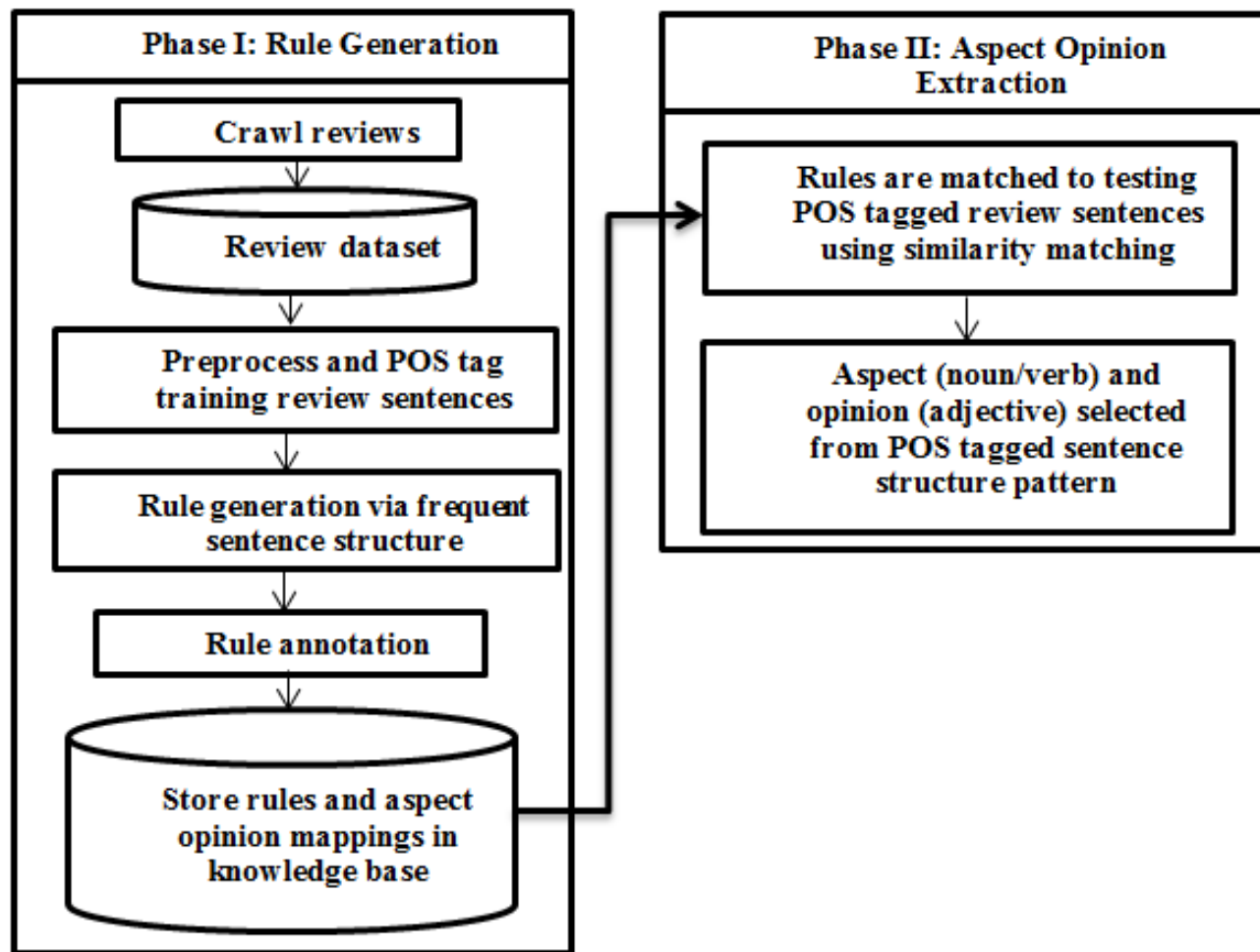
Some important opinion were ignored.

Objective

- To improve the extraction of aspects and its sentiments especially when they are written in a complex manner.
- To provide a domain independent solution by not relying to domain lexicon in its aspect and sentiments extraction.
- To identify the association between aspect and its corresponding opinion.



Rule-based Aspect Opinion Selection Framework



Phase 1: Rule Generation

a) Collect Reviews

- ❑ Collected from TripAdvisor
- ❑ English Language
- ❑ 9 hotels reviews in Bayan Lepas, Penang
- ❑ Total 761 reviews (1/01/2010 until 30/06/2015)

b) Preprocessing

- Let's take as sample review from U Hotel Penang for illustration.



Srinisri
Hyderabad

Level 5 Contributor



52 reviews



16 hotel reviews



76 helpful votes

“Clean rooms with smoke free”

★★★★☆ Reviewed 2 weeks ago

its a smoke free hotel. Nicely maintained. Rooms were very clean. Though the size of the room was bit small but great for the price paid. Hotel receptionists speak english. Hotel is located closer to good resturants



b) Preprocessing

- ❑ Sentences were tokenized from a review paragraph using punctuation (. , ! , ? ,) as end of sentence.
- ❑ Abbreviations excluded as end of sentence.
- S1: Its a smoke free hotel .
- S2: Rooms were very clean.

its a smoke free hotel.

Nicely maintained.

Rooms were very clean.

Though the size of the room was bit small but great for the price paid.

Hotel receptionists speak English.

Hotel is located closer to good restaurants

c) Part Of Speech (POS) Tagging

- Sentences were tagged using Stanford POS tagger.
- S1: Its **PRP\$** a **DT** smoke **NN** free **JJ** hotel **NN** . . .
- S2: Rooms **NNS** were **VBD** very **RB** clean **JJ** . . .

PRP\$	-Pronoun
DT	-Determiner
NN / NNS	-Noun
JJ	- Adjective
VBD	-Verb
RB	-Adverb

```
Its a smoke free hotel.
Its_PRPS a_DT smoke_NN free_JJ hotel_NN ...

Nicely maintained.
Nicely_RB maintained_VBN ...

Rooms were very clean.
Rooms_NNS were_VBD very_RB clean_JJ ...

Though the size of the room was bit small but great for the price paid.
Though_IN the_DT size_NN of_IN the_DT room_NN was_VBD bit_RB small_JJ but_CC great_JJ for_IN
the_DT price_NN paid_VBN ...

Hotel receptionists speak English.
Hotel_NNP receptionists_NNS speak_VBP English_NNP ...

Hotel is located closer to good restaurants
Hotel_NN is_VBZ located_II closer_RB to_TO good_JJ restaurants_NNS
```

d) Rules Generation via Frequent Sentence Structure and Rules Annotation

-
- Similar patterns were identified in sentence structures to determine the position of aspect and its opinion.
 - Simple Rule (1 aspect + 1 opinion)
 - Eg: **NN RB JJ**
 - Complex Rule (> 1 aspect + opinion)
 - Eg: **NN JJ NN**

1. [NN] [RB] [JJ]	-Simple Rule
2. [NN] [JJ]	-Simple Rule
3. [JJ] [NN] [IN] [NN]	-Complex Rule
4. [JJ] [NN] [NN]	-Complex Rule
5. [VBD] [NN]	-Simple Rule
6. [JJ] [NN]	-Simple Rule
7. [JJ] [NNS]	-Simple Rule
8. [NNS] [JJ]	-Simple Rule
9. [RB] [JJ]	-Simple Rule
10. [RB] [JJ] [NN]	-Simple Rule
11. [RB] [RB] [NN]	-Simple Rule
12. [RB] [RB] [JJ]	-Simple Rule
13. [VBN] [NN]	-Complex Rule
14. [VBD] [NN]	-Complex Rule
15. [VBN] [NNS]	-Complex Rule

e) Store Rules and Aspect Opinion Mappings inside Knowledge Base

Rules generated and its aspect opinion mappings were stored inside knowledge base.



Rules (R)

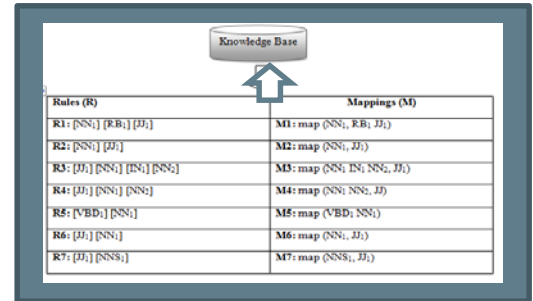
Mappings (M)

R1: [NN₁] [RB₁] [JJ₁]

M1: map (NN₁, RB₁ JJ₁)

R2: [NN₁] [JJ₁] [NN₁]

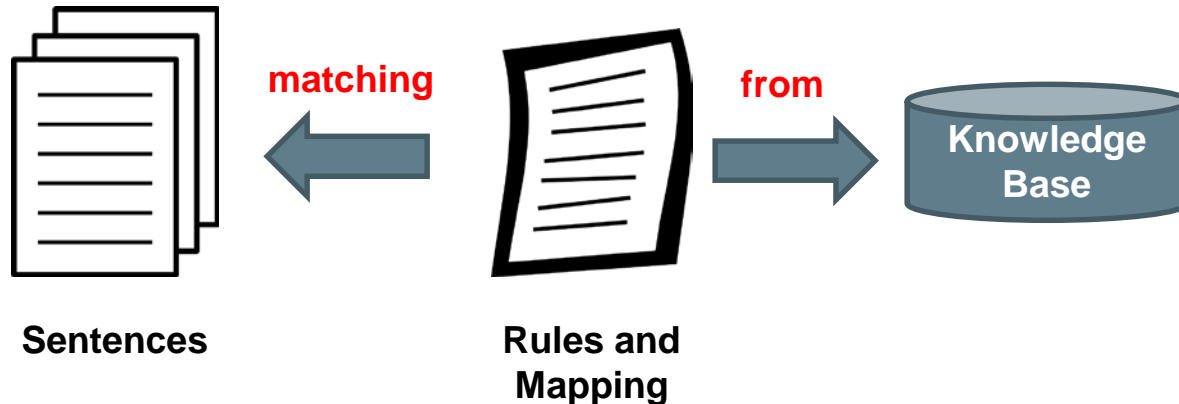
M2: map (NN₁ JJ₁, NN₂)



Phase 2 : Aspect Opinion Extraction

a) Similarity Matching

- For similarity matching, all the rules and sentences are matched to find the appropriate rule for that particular sentence.



a) Similarity Matching

R2 : NN JJ NN

S1 : PRP\$ DT NN JJ NN



Similarity = POS Matches / Total POS

Similarity = 3 / 8 = 0.38
= 38 %

a) Similarity Matching

R3 : NN JJ

S1 : PRP\$ DT NN JJ NN

Similarity = POS Matches / Total POS

Similarity = $2 / 7 = 0.29$
= 29 %

a) Similarity Matching

R1 : NN RB JJ NN

S1 : PRP\$ DT NN JJ NN

Similarity = POS Matches / Total POS

Similarity = $3 / 9 = 0.33$
= 33 %

a) Similarity Matching

The rule that will be assigned to **S1** is the one with highest similarity score:

R2 (38% similarity)

b) Aspect and Opinion Mapping

R2 : NN_1 JJ_1 NN_2

Extraction = *smoke, free, hotel*

Aspect (NN_2) = *hotel*

Opinion (NN_1, JJ_1) = *smoke, free*

M1: map (NN_1 JJ_1 , NN_2)

Opinion

Aspect

map (NN_1 JJ_1 , NN_2) = map (*smoke free, hotel*)

Mapping = *smoke free - hotel*

Conclusion

- ❑ **Framework proposed** - rule-based method that use POS tags and sentence structures for aspect and opinion extraction.
- ❑ **Can be used on simple and complex sentences** - in similar domain (product/service).
- ❑ **Application** - aspect and opinion extracted can be used to generate short summary (product/service).

Contact

Noor Rizvana

ana76_rez@yahoo.com



Thank You...