COMP 6751 Natural Language Analysis

Project 4 Report: Grammar Design and Baseline

Student: Yixuan Li   40079830

Table of Contents

Expectations of originality:

I, student 40079830, certify that this submission is my original work and meets the Faculty's Expectations of Originality.

Date: December 3 , 2020

I. Explanation on grammar design for complex sentences

As in project 3, a sentiment attribute called *SENTI* is attached to proper constituents and terminals, and in my grammar design, there are in total three different sentiments: *positive, neutral, negative*. As the parser parses a sentence or a small paragraph based on the grammar, the SENTI attribute is propagated up. To handle more complex declarative sentences, the effects of negation, discourse relations and subordinate clauses are considered.

1) Negation

Negation is to express the opposite meaning of a particular word. The most common negation word is "not", for example, *this movie is not dull*. "not" negates the sentiment of the word "dull", resulting in giving positive sentiment to the sentence.

The scope of negation is also considered in the feature grammar when discourse relation is introduced. In the example sentence, *this movie is not dull and scary.* The negation word "not" is applied on "dull and scary", and it is similar for the conjunction "or". However, in the example sentence, *this movie is not dull but entertaining.* The negation word "not" is only applied on "dull" instead.

In my grammar design, negation is applied on verb phrase (VP) and adjective phrase (ADJP)

2) Discourse relations

Discourse relations have to be considered to analyze the sentiment of more complex sentences. Connectives like "and", "but" are used to cohere with same constituents. However, not all connectives can go with every relation, and each can effect certain discourse relation types and text genre on the production[2].

In my grammar, conjunctions "and", "but" and "or" are considered on 4 different constituents: sentence (S), noun phrase (NP), adjective phrase (ADJP) and adverbial phrase (ADVP), and I will explain how they work in details in section II with concrete examples.

3) Subordinate clauses

Based on the definition of subordinate clauses: *a subordinate clause or dependent clause is a clause that provides a sentence element with additional information*[3], we know that it can affect the sentiment as well. For example, in the example sentence *He saw a movie which was fascinating.*, there is only one word "fascinating" with sentiment which is positive, so the clause "which was fascinating" affects the sentiment of the noun phrase, thus the entire sentence.

In my grammar, I marked subordinate clauses as *SBAR* and I considered two types of subordinate clause: relative (adjectival) clauses and content (noun) clauses.

- **noun clauses**[4]
  There are two main kinds of noun clauses: Declarative content clauses and interrogative content clauses. And the current grammar only supports declarative content clauses. For example: *He said **that he saw a scary movie**. (negative)
- **relative (adjectival) clauses**[5]
  Sentences like *He saw a scary movie **which was boring*** can be labeled as negative sentiment.

However, there are many other clause types such as adverbial clauses which are not considered in the grammar, and they are more complex.

- **adverbial clauses**[6]

An adverbial clause can modify a verb, an adjective or another adverb. It can express *time, condition, purpose, reason, concession, place, comparison, manner and results* according to the senses of their conjunctions[6], such as "in order to" for purpose, "as soon as" for time etc.

II. Explanation on grammar design for conjunction of sentiment bearing adjectives or nouns

1) Grammar design for sentiment bearing adjective phrase (ADJP)

Adjective phrases (ADJP) are commonly used in clusters with different combinations to modify NP, and there are certain limitations on the various combination of adjectives in ADJP with different conjunctions. Here's is the table of the mapping for **positive ADJP** with different conjunctions **without negation**, and it is similar to negative or neutral ADJP.

ADJP sentiment	1st part	Conjunction	2nd part	Example
positive	positive	and	positive	The movie is *interesting and entertaining* .
	positive		neutral	The movie is *interesting and long* .
	neutral		positive	The movie is *long and interesting* .
	positive	or	positive	He said that this movie was *interesting or entertaining* .
	negative	but	positive	He read a story which is *scary but fascinating* .
	neutral		positive	The car is *expensive but well-equipped* .

The corresponding grammar rules are:

```
ADJP[SENTI=?s] -> JJ[SENTI=?s] CC[+and] JJ[SENTI=?s] | JJ[SENTI=?s] CC[+and] ADJP[SENTI=?s]
ADJP[SENTI=positive] -> JJ[SENTI=positive] CC[+and] JJ[SENTI=neutral] | JJ[SENTI=neutral] CC[+and] JJ[SENTI=positive]
ADJP[SENTI=?s] -> JJ[SENTI=?s] CC[+or] JJ[SENTI=?s]
ADJP[SENTI=?s] -> JJ CC[+but] JJ[SENTI=?s]
```

And recursive grammar is used for ADJP when it has multiple adjectives:

```
ADJP[SENTI=?s] -> JJ[SENTI=?s] CC[+and] ADJP[SENTI=?s]
ADJP[SENTI=?s] -> JJ[SENTI=?s] COMMA ADJP[SENTI=?s]
ADJP[SENTI=?s] -> JJ COMMA CC[+but] ADJP[SENTI=?s]
```

With negation, if the negation word appears before the 1st part, then VP rules with negation will handle this. If the negation word appears in the 2nd part, then the corresponding grammar rules are:

```
ADJP[SENTI=?s] -> JJ[SENTI=?s] CC[+but] RB[+negation] JJ | ADJP[SENTI=?s] CC[+but] RB[+negation] JJ
ADJP[SENTI=?s] -> JJ[SENTI=?s] CC[+and] RB[+negation] JJ | ADJP[SENTI=?s] CC[+and] RB[+negation] JJ
ADJP[SENTI=?s] -> JJ[SENTI=?s] CC[+or] RB[+negation] JJ | ADJP[SENTI=?s] CC[+or] RB[+negation] JJ
```

2) Grammar design for sentiment bearing noun phrase (NP)

The sentiment of NP can be percolated from noun terminal (N), prepositional phrase (PP), adjective (JJ), subordinate clause (SBAR) and adjective phrase (ADJP).

NP sentiment	Grammar rules	Example
same as N	`NP[NUM=?n, PERSON=?p, SENTI=?s] -> N[NUM=?n, PERSON=?p, SENTI=?s]`	It is a **hazard**.
same as PP	`NP[NUM=?n, PERSON=?p, SENTI=?s] -> DT[NUM=?n] NP[NUM=?n, PERSON=?p] PP[SENTI=?s]`	an example **of perfect movie**
same as JJ	`NP[NUM=?n, PERSON=?p, SENTI=?s] -> JJ[SENTI=?s] NP[NUM=?n, PERSON=?p]`	this is a **perfect** example
same as SBAR	`NP[NUM=?n, PERSON=?p, SENTI=?s] -> NP[NUM=?n, PERSON=?p] SBAR[SENTI=?s]`	an example **which is perfect**
same as ADJP	`NP[NUM=?n, PERSON=?p, SENTI=?s] -> ADJP[SENTI=?s] NP[NUM=?n, PERSON=?p]`	It's a **shamelessly manipulative and boring movie making**

The negation on noun phrase is handled by the negation in SBAR or ADJP. I explained the negation on ADJP above, and here's the grammar rules for negation on SBAR.

```
SBAR[SENTI=positive] -> WP[+wh] V RB[+not] ADJP[SENTI=negative] | WP[+wh] V RB[+not] JJ[SENTI=negative]
SBAR[SENTI=negative] -> WP[+wh] V RB[+not] ADJP[SENTI=positive] | WP[+wh] V RB[+not] JJ[SENTI=positive]
SBAR[SENTI=neutral] -> WP[+wh] V RB[+not] ADJP[SENTI=?s] | WP[+wh] V RB[+not] JJ[SENTI=?s]
```

As in ADJP, if a noun phrase is consist of multiple noun phrases such as "a mess and a hazard"
  a. if conjunction is "and", the sentiment is determined by both parts or one of them if the other is neutral.
  b. if conjunction is "or", the sentiment is determined by one of the parts, however if two parts have opposite sentiments, such as "a good movie or a bad movie", then the NP sentiment is neutral.
  c. if conjunction is "but", the sentiment is always determined by the latter.

Due to the page limit, please check the corresponding grammar rules in *sentianalysis_grammar_s.fcfg*.

3) Limitations

As discussed in my project 3 report, there is an "of issue" in NP, and it causes the 1 false positive for feature-based grammar shown in section IV. The reason is that the sentiment of a NP which contains "of" can be determined by the part before "of" or by the part after "of". Therefore, the example sentence "a perfect example of boring movie" will get 1 positive vote and 1 negative vote.

III. Explanation on grammar design for conjunction of sentiment bearing sentences

1) Explanation on grammar design on the constituent S

The sentiment of a simple sentence is decided by its NP and VP, and the related grammar rule is

```
S[SENTI=?s, -INV] -> NP[NUM=?n, PERSON=?p, SENTI=?s] VP[TENSE=?t, NUM=?n, PERSON=?p, SENTI=?s]
```

When there are several compound sentences which are conjoined by conjunctions, the logic is almost same as the table shown in section II for positive ADJP. But I will explain more on negative and neutral cases this time.

S sentiment	1st S constituent	Comma (optional)	Conjunction	2nd S constituent	Grammar rules
negative	negative	,	and	negative	S[SENTI=negative, -INV] -> S[SENTI=negative, -INV] CC[+and] S[SENTI=negative, -INV] S[SENTI=negative, -INV] -> S[SENTI=negative, -INV] COMMA CC[+and] S[SENTI=negative, -INV]
	negative		and	neutral	S[SENTI=negative, -INV] -> S[SENTI=negative, -INV] CC[+and] S[SENTI=neutral, -INV] S[SENTI=negative, -INV] -> S[SENTI=negative, -INV] COMMA CC[+and] S[SENTI=neutral, -INV]
	neutral		and	negative	S[SENTI=negative, -INV] -> S[SENTI=neutral, -INV] CC[+and] S[SENTI=negative, -INV] S[SENTI=negative, -INV] -> S[SENTI=neutral, -INV] COMMA CC[+and] S[SENTI=negative, -INV]
	negative		or	negative	S[SENTI=?s, -INV] -> S[SENTI=?s, -INV] CC[+or] S[SENTI=?s, -INV]
	positive		but	negative	S[SENTI=negative, -INV] -> S[SENTI=positive, -INV] CC[+but] S[SENTI=negative, -INV] S[SENTI=negative, -INV] -> S[SENTI=positive, -INV] COMMA CC[+but] S[SENTI=negative, -INV]
	neutral		but	negative	S[SENTI=negative, -INV] -> S[SENTI=neutral, -INV] CC[+but] S[SENTI=negative, -INV] S[SENTI=negative, -INV] -> S[SENTI=neutral, -INV] COMMA CC[+but] S[SENTI=negative, -INV]
neutral	neutral	,	and	neutral	S[SENTI=neutral, -INV] -> S[SENTI=neutral, -INV] CC[+and] S[SENTI=neutral, -INV] S[SENTI=neutral, -INV] -> S[SENTI=neutral, -INV] COMMA CC[+and] S[SENTI=neutral, -INV]
	neutral		or	neutral	S[SENTI=?s, -INV] -> S[SENTI=?s, -INV] CC[+or] S[SENTI=?s, -INV] S[SENTI=?s, -INV] -> S[SENTI=?s, -INV] COMMA CC[+or] S[SENTI=?s, -INV]
	positive		or	negative	S[SENTI=neutral, -INV] -> S[SENTI=positive, -INV] CC[+or] S[SENTI=negative, -INV] | S[SENTI=negative, -INV] CC[+or] S[SENTI=positive, -INV]
	negative		or	positive	S[SENTI=neutral, -INV] -> S[SENTI=positive, -INV] COMMA CC[+or] S[SENTI=negative, -INV] | S[SENTI=negative, -INV] COMMA CC[+or] S[SENTI=positive, -INV]
	neutral		but	neutral	S[SENTI=neutral, -INV] -> S[SENTI=?s, -INV] CC[+but] S[SENTI=neutral, -INV] | S[SENTI=?s, -INV] COMMA CC[+but] S[SENTI=neutral, -INV]
	positive		but	neutral	
	negative		but	neutral	

The negation on S is solved by negation in NP rules and negation in VP rules, so there is no explicit grammar rules for S negation.

2) Limitations

The limitation on sentiment bearing S constituent is that the current grammar does not support to analyze the sentiment of questioning such as "Is this an interesting movie?". However, I added -INV attribute for detecting inversion on S constituent as an interface.

IV. Compare grammar-based sentiment analysis with SSAP and an academic paper

1) Test cases and results comparison

There are 5 ground-truth positive sentences, 5 ground-truth negative sentences and 3 neutral sentences in testing data in different complexity. Here's the results of feature grammar and SSAP baseline.

#	Ground-truth	Sentence	Complexity level	Grammar-based sentiment	SSAP polarity score	SSAP sentiment
1	positive	This is a well-intentioned movie making **and** it is funny **or** entertaining .	conj. : **and, or**	[positive]	1.604	positive
2		This is a well-intentioned movie making **and** it is funny **or** entertaining . It is a compelling **or** fascinating story , **and** it has gut-wrenching impact .	conj. : **and, or** small paragraph	[positive]	1.732	positive
3		This is a well-intentioned movie making **and** it is **not** boring **but** interesting **or** entertaining .	**negation** conj. : **and, or, but**	[positive]	0.243	positive
4		This may **not** have the dramatic gut-wrenching impact of other holocaust films , **but** it's a compelling **and** entertaining story , mainly because of the way it's told by the people *who were there* .	**negation** conj. : **and, but** *subordinate clause*	[positive]	0.333	positive
5		This is a perfect example of well-intentioned movie making . It's a compelling story with gut-wrenching impact , **and** it is entertaining **or** interesting **but not** too long. I like this movie very much .	**negation** conj. : **and, or, but** small paragraph	[positive]	1.5	positive
6	negative	I saw an entertaining **but** scary movie **and** it is a long story .	conj. : **and, but**	[negative]	0.0	neutral
7		I saw a dramatic **but** scary movie **and** it is a long story . It was a mess **and** a shamelessly manipulative **or** boring movie making .	conj. : **and, or, but** small paragraph	[negative]	-1.373	negative
8		This was a well-intentioned **but** manipulative movie making *which was **not** interesting **or** entertaining* .	**negation** conj. : **but, or** *subordinate clause*	[negative]	1.0	positive
9		It's a perfect example of rancid , well-intentioned , **but** shamelessly manipulative movie making *which is **not** entertaining **and** compelling* .	**negation** conj. : **and, but** *subordinate clause*	[positive, negative]*	1.091	positive
10		He was disappointed because he saw a movie *which was **not** interesting **or** entertaining*. He said *that it's a classic example of rancid , well-intentioned , **but** shamelessly manipulative movie making* .	**negation** conj. : **or, but** small paragraph *subordinate clause*	[negative]	0.354	positive
11	neutral	This is an neutral **or** general game .	conj. : **or**	[neutral]	0.0	neutral
12		This is a movie review, **and** it is **not** positive **or** negative . It is quite neutral .	**negation** conj. : **and, or** small paragraph	[neutral]	0.0	neutral
13		There are many cars *which have different colors* on the street , **but** the color of this car is black **and** it is **not** cheap **but** expensive .	negation conj. : **and, but** *subordinate clause*	[neutral]	0.0	neutral

[positive, negative]* means the sentence is labeled as positive or negative because they ties at the most number of parse trees with corresponding sentiment.

correctly labeled	wrongly labeled

2) Confusion matrix

SSAP baseline		Prediction		
		positive	negative	neutral
Ground-truth	positive	5	0	0
	negative	3	1	1
	neutral	0	0	3

feature-based grammar		Prediction		
		positive	negative	neutral
Ground-truth	positive	5	0	0
	negative	1	4	0
	neutral	0	0	3

3) More discussion on SSAP

The limitation of SSAP baseline is that it doesn't consider the negation and discourse relations, and it also doesn't take the part-of-speech into consideration. For example, in the example sentence "*He is pretty tall*.", "pretty" is an adverbial intensifier rather than an adjective with positive sentiment.

4

4) Compare the results with a modern academic paper

I compared the results with the results in paper: *How do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews*[7] where the authors used NLTK as the toolkit.

Approach	Precision	Recall	F-measure
Project 4 results*	5 / 6 = 0.833	5 / 5 = 1.0	0.908
Paper results (F_{NS} average)	0.601	0.506	0.549

* neutral class is temporarily omitted for calculating precision, recall and F1 measure.

Although the paper results seems to have lower performance mainly due to the small test set in this project, the model in the paper is trained on review text in various categories (e.g. games, productivity, social, etc.). The authors used *SentiStrength*[8] as the sentiment lexicon and used collocation finding algorithm in NLTK for extracting features from the review text.

The potential improvement for my model is to assign a sentiment score to each sentence, since the current feature grammar only gives the sentiment label. The way that paper authors used is taking the maximum and minimum scores given by SentiStrength among all the words in a sentence.

V. References

[1] References in Project 3 and pipeline and grammar productions in Project 3
[2] Discourse study on "and" and "but" (paper): https://journals.openedition.org/discours/10848
[3] Dependent clause (Wikipedia): https://en.wikipedia.org/wiki/Dependent_clause
[4] Content (noun) clause (Wikipedia): https://en.wikipedia.org/wiki/Content_clause
[5] Relative (adjectival) clause (Wikipedia): https://en.wikipedia.org/wiki/Relative_clause
[6] Adverbial clause (Wikipedia): https://en.wikipedia.org/wiki/Adverbial_clause
[7] *How do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews (paper)*: https://ieeexplore.ieee.org/document/6912257
[8] SentiStrength: http://sentistrength.wlv.ac.uk/

VI. Appendix

1. Annotations in Good.txt and False.txt
The sentences that the parser parses correctly and outputs the right sentiment label will be saved to Good.txt;
The sentences that the parser parses incorrectly or outputs the wrong sentiment label will be saved to False.txt.
The pattern in output files is shown below

```
Input Sentence
Ground-truth sentiment      |      [ Prediction(s) ]

parse tree(s)
```

An example:

```
This is an neutral or general game .
neutral       |      [neutral]

(S[-INV, SENTI='neutral']
  (NP[NUM='sg', PERSON=?p, SENTI=?s] (DT[NUM='sg'] This))
  (VP[NUM='sg', PERSON=3, SENTI='neutral', TENSE='pres']
    (V[+AUX, NUM='sg', PERSON=3, TENSE='pres', -passive] is)
    (NP[NUM='sg', PERSON=?p, SENTI='neutral']
      (DT[NUM='sg'] an)
      (NP[NUM='sg', PERSON=?p, SENTI='neutral']
        (ADJP[SENTI='neutral']
          (JJ[SENTI='neutral'] neutral)
          (CC[-and, -but, +or] or)
          (JJ[SENTI='neutral'] general))
        (NP[NUM='sg', PERSON=?p, SENTI='neutral']
          (N[NUM='sg', SENTI='neutral'] game))))))
```

5