
Natural Language Engineering 2014

Assignment – Opinion Extraction Report

Candidate number: 118440

Date: 21th November, 2014

Word count: 2401 words

Table of Contents

1. Introduction	2
2. Design	2
3. Testing	4
4. Performance on DVD reviews	6
4.1 Sampling	6
4.2 Data	6
4.3 Analysis	15
5. Limitation.....	15
5.1 Deficiency in the opinion extractor algorithm	15
5.1.1 The Conjunction Issues.....	16
5.1.2 The “No” Issue	17
5.2 Errors made by the part-of-speech tagger.....	19
5.2.1 Syntactic ambiguity.....	19
5.2.3 Suffixes Issue.....	20
5.3 Errors made by the dependency parser	20
6. Summary	23
7. A Proposal for a DVD Summary Website	23
Background and Motivation	23
Proposed Content.....	23
Limitation	24

1. Introduction

An opinion extractor is a program that reads a certain amount of reviews and extracts opinions based on a specific aspect word. The design and methodology behind it will be introduced in this report. This report also aims at investigating into an opinion extractor, which will be tested and inspected on its performance. Analysis will then be made and some limitations of its functionalities will be discussed further.

2. Design

The algorithm of the opinion extractor involves 3 steps: initialize an empty list of opinion; find opinion under specified conditions; and return a list of outputs.

Provided below is a pseudo code of the algorithm.

Input:

aspect_token: the aspect that users are interested of extracting opinions in.

Output:

opinions: an array of tokens which constitute the extracted opinions of the aspect.

Function opinion_extractor:

```

opinions = []
adv = empty
adj = empty
conj = empty
head_of_aspect_token = parsed sentence

// CASE1: if HEAD of aspect token is ADJECTIVE
if the dependency relation of aspect_token = "nsubj" and
the PoS tag of head_of_aspect_token = "JJ", then:
    adj = form of head_of_aspect_token

    for each dependent in aspect_token's set of dependents:
        if dependency relation of dependent = "neg" Then:
            neg = True
        else, if dependency relation of dependent = "admod", then:
            adv = dependent's form
    if neg is true, then:
        append "not-" + adv + " " + adj to opinions list
    else:
        if adv = nothing then,
            append adj to opinions list
        else:
            append adv + "-" + adj to opinions list

// CASE2: if DEPENDANT of aspect token is ADJECTIVE
Else:
    for each dependent in aspect_token's set of dependents:
        if PoS tag of dependant == "JJ", then:
            adj = dependant's form

```

```

for each dependent2 in aspect_token's set of dependents:
    if dependency relation of dependent2 = "neg" Then:
        neg = True
    else, if dependency relation of dependent2 = "admod", then:
        adv = dependent2's form

for each dependent3 in dependent's set of dependents:
    if dependency relation of dependent3 = "neg" Then:
        neg = True
    else, if dependency relation of dependent3 = "admod", then:
        adv = dependent3's form

if neg is true, then:
    append "not-" + adv + " " + adj to opinions list
else:
    if adv = nothing then,
        append adj to opinions list
    else:
        append adv + "-" + adj to opinions list

// Find CONJUNCTIONS
for each dependent in head_of_aspect_token's set of dependents:
    if dependency relation of dependent = "conj" and
       dependant is not aspect_token then,
        conj = form of dependant
    for each dependent2 in aspect_token's set of dependents:
        if dependency relation of dependent2 = "neg" Then:
            neg = True
        else, if dependency relation of dependent2 = "admod", then:
            adv = dependent2's form

    if neg is true, then:
        append "not-" + adv + " " + conj to opinions list
    else:
        if adv = nothing then,
            append conj to opinions list
        else:
            append adv + "-" + conj to opinions list

return opinions

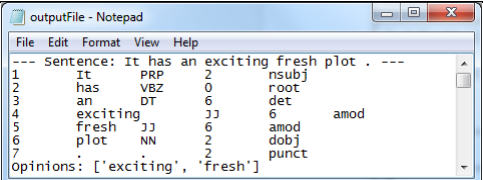
```

3. Testing

The first step after creating the opinion extractor would be testing the validity under 5 main categories: Adjectival modification, Adjectives linked by copulae, Adverbial modifiers, Negation and Conjunctions. I used 9 sentences from the example set to perform the testing, each one of them targeting one category of each extension. In table 3a, results of testing are shown, proving that the opinion extractor is valid under all 5 extensions.

Extension1

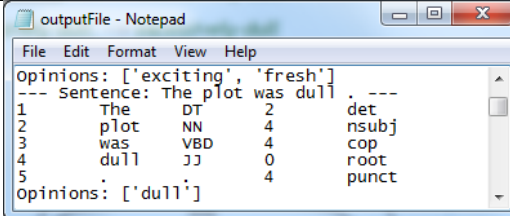
Table 3a – Testing results on example testing set

Extension	Input	Expectation	Output	Result
1: Adjectival modification	It has an exciting fresh plot.	fresh, exciting	 <pre> --- Sentence: It has an exciting fresh plot . --- 1 It PRP 2 nsubj 2 has VBZ 0 root 3 an DT 6 det 4 exciting JJ 6 amod 5 fresh JJ 6 amod 6 plot NN 2 dobj 7 . punct Opinions: ['exciting', 'fresh'] </pre>	✓

In Extension 1, the adjective of the aspect token is found by taking all dependents of the aspect word, finding those have an “amod” dependency relation with the aspect token and append it into the opinion list.

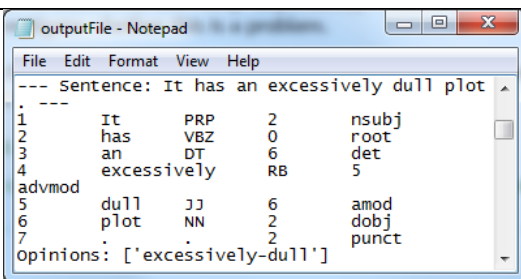
In the above example, “plot” is the aspect word where “exciting” and “fresh” are the Adjective. Correct output are shown.

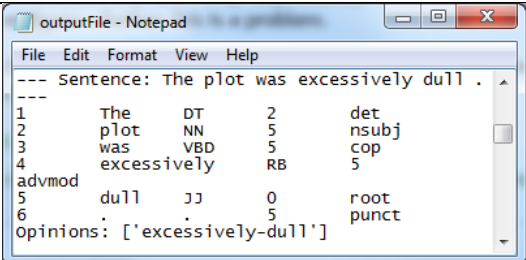
Extension2

Extension	Input	Expectation	Output	Result
2: Adjectives linked by copulae	The plot was dull.	dull	 <pre> Opinions: ['exciting', 'fresh'] --- Sentence: The plot was dull . --- 1 The DT 2 det 2 plot NN 4 nsubj 3 was VBD 4 cop 4 dull JJ 0 root 5 . punct Opinions: ['dull'] </pre>	✓

In extension 2, the adjective was found by checking if the aspect token is in an “nsubj” relationship with the head of the token(adjective), which in this case, “dull” is the adjective of “plot”.

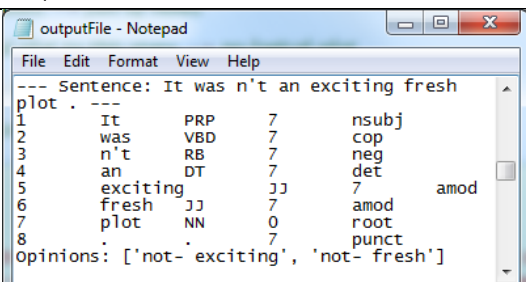
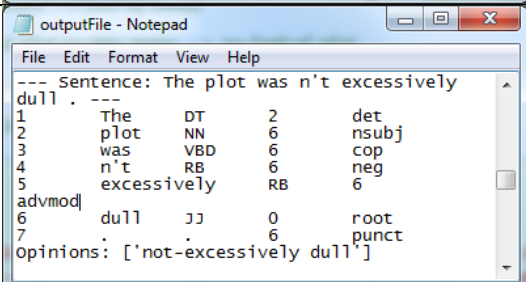
Extension3

Extension	Input	Expectation	Output	Result
3: Adverbial modifiers	It has an excessively dull plot.	excessively-dull	 <pre> --- Sentence: It has an excessively dull plot . --- 1 It PRP 2 nsubj 2 has VBZ 0 root 3 an DT 6 det 4 excessively RB 5 advmod 5 dull JJ 6 amod 6 plot NN 2 dobj 7 . punct Opinions: ['excessively-dull'] </pre>	✓

3: Adverbial modifiers	The plot was excessively dull.	excessively-dull		✓
------------------------	--------------------------------	------------------	--	---

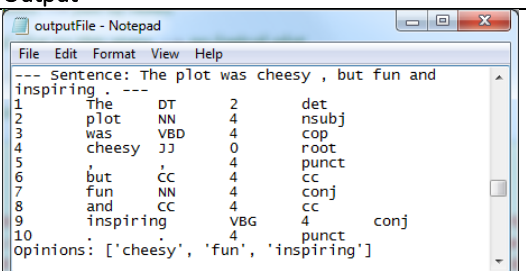
In extension3, adverb are found together with the adjective. In this case, “excessively” is an adverb of “dull”. To find “excessively”, all dependents of “dull”, which is the adjective, are checked, and the tokens with an “avmod” dependency relationship with the adjective will be added to the opinion list.

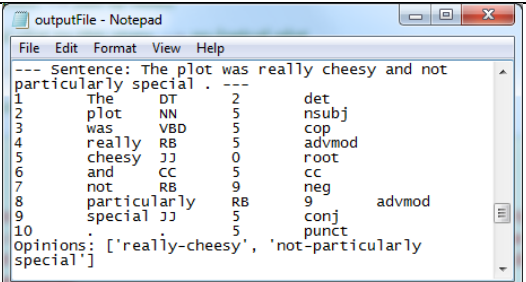
Extension4

Extension	Input	Expectation	Output	Result
4: Negation	It wasn't an exciting fresh plot.	not-exciting, not-fresh		✓
4: Negation	The plot wasn't excessively dull.	not-excessively-dull		✓

In extension4, negation are considered. All dependent of the adjective will be looked into, and if any token has a “neg” dependency relationship with the adjective token, “not-” will be added into the adjective as a prefix. In the testing sets, negation works on both adjective and those with an adverb.

Extension5

Extension	Input	Expectation	Output	Result
5: Conjunction	The plot was cheesy, but fun and inspiring.	cheesy, fun, inspiring		✓

5: Conjunction	The plot was really cheesy and not particularly special.	really-cheesy, not-particularly-special		✓
----------------	--	---	--	---

In extension5, conjunctions are found by scanning through the dependants of the first adjective. The conjunctions should be in a “conj” relationship with the adjective token, which will be added to the opinion list if founded. The two test examples shows that conjunctions are not only limited to adjectives only.

After testing, the opinion extractor will be applied to the DVD Amazon review to see if correct opinion will be generated.

4. Performance on DVD reviews

4.1 Sampling

In order to analyse the performance and further deduce limitations of the opinion extractor, 160 samples will be picked from the text file generated from 4 different Amazon DVD review aspects. Therefore, 40 samples will be picked from each of the 4 aspects, and then being examined manually one by one.

The next section will be showing 4 tables containing results of 4 inspected aspects.

4.2 Data

While examining the output opinions, errors have been spotted. Potential limitations are mainly due to the following errors:

- (i) Deficiency in the opinion extractor algorithm
- (ii) Errors made by the part-of-speech tagger
- (iii) Errors made by the dependency parser

The above errors will be labelled in the “result” column of the samples analysis result with the number (i), (ii) or (iii) on it.

Table 4.2a – 40 samples of analysis results on PLOT aspect of Amazon DVD reviews

Aspect	Input	Expectation	Output	Result / Error
--------	-------	-------------	--------	----------------

Plot	The plot was less coherent , but if you watch flicks like this for the plot , something is clinically wrong with you .	"Coherent"	'clinically-wrong'	X (i)
Plot	The movie plot is very much the same as the story of the first three books .	"Same"	"Same"	✓
Plot	Along the way , Thornhill meets a beautiful blond named Eve Kendall (Eva Marie Saint) who adds a lot of sex appeal along with several surprises to this wild and wonderful plot .	"Wild and wonderful"	'wonderful'	X (ii)
Plot	In summary , the music portion of this DVD rates a strong five stars , but the silly plot and the problems with the disk discussed above , bring this DVD down to a four star production	"Silly"	'discussed'	X (ii)
Plot	Although the plot is very thin , you just do n't care because it 's a fun movie .	"Very-thin"	'very-thin'	✓
plot	This movie does not take the easy way out by having a nice simple plot .	"Nice-simple"	'nice'	X (iii)
Plot	He IS adorable in the movie , but the plot is extremely corny , and I found myself rolling my eyes more than laughing out loud .	"extremely-corny"	'extremely-corny'	✓
Plot	They say that every show had the same plot .	"Same"	'same'	✓
Plot	The plot has already been summarized , but I will point out that I did n't find any of the characters overly-stereotypical , especially not in the derogatory sense .	No opinion found	'point'	X (i)
Plot	Set in WWII London , the dark hearted plot focuses on Navy Lieutenant Commander Charles Madison , an especially notorious personal assistant to the mentally unstable Admiral William Jessup .	"dark-hearted"	'dark'	X (ii)
Plot	There is not really any other plot .	no opinion found	"really-other"	X (i)
Plot	The plot is at the same time simple and deep , should you choose to look beyond the surface .	"simple", "deep"	"deep"	X (iii)
Plot	Poor plot , script , acting , everything was poor .	"poor"	"poor"	✓
Plot	The plot is great .	"great"	"great"	✓
Plot	The movie is sad and tragic with an amazing performance by Jackson and I shall always remember it for it 's unique plot and amazing delivery .	"unique"	"unique"	✓
Plot	The plot again is twisting and will keep you guessing until the very end , and , of course , the girls are fun to look at .	"twisting"	"keep"	X (ii)
Plot	The plot may be a bit serpentine and the dialogue stretched at the corners to cover maximum political ground , but Chad Allen 's nuanced and sexily hard-boiled performance easily compensates for these shortcomings .	"serpentine"	"stretched", "easily-compensates"	X (ii)
Plot	Anchored by a strong , fun plot , dazzling special effects and an instant classic performance by Johnny Depp as Captain Jack Sparrow , `` Pirates Of The Caribbean : The Curse Of The Black Pearl `` is a fantastic movie and the msot fun one can have while watching one .	"strong", "fun"	"strong"	X (ii)
Plot	While the plot is nearly identical , Alexander Payne (director) treats his characters much harsher than Perrotta .	"nearly-identical"	"nearly-identical"	✓
Plot	below average , the dialogues are way below average , while the plot is not bad at all .	"not-bad"	"not-at bad"	✓
Plot	The plot is slow yet intensely emotional .	"slow", "intensely emotional"	"slow"	X (iii)
Plot	The plot was bad and it did not have that many pokemon as I thought.And Mew does not fit in the story when it is about Lucario	"bad"	"bad", "not- have"	X (ii)
Plot	The plot is quite thin , but suffices in order to fill the time with what this movie is primarily concerned with - bringing five star entertainment to the screen in the form of musical numbers .	"quite-thin"	"quite-thin", "quite-fill"	X (iii)

Plot	Comparing it to Mary Poppins is silly , for Mary Poppins was a light-hearted comedy with no real plot , where Bedknobs and Broomsticks features a fantastic plot and clever dialogue .	“fantastic”	“fantastic”, “real”	X (i)
Plot	I can not rave about the content , plot or actors in any way that would be a fair comparison to a typical Hollywood production .	No opinion found	“actors”	X (i)
Plot	There is a solid plot , there is lush scenery and a time capsule feel for the late 60 's , depiction of social degradation amongst the wealthy class , exploitation of youth , betrayal , revenge , and murder , all rolled into a cleverly filmed psychedelic dreamlike piece of art-house cinema .	“solid”	“solid”	✓
Plot	The tenuous plot , which mainly serves to connect the brilliant dance numbers , concerns John `` Lucky `` Garnett (Astaire) , a gambler and professional dancer .	“ tenuous”	“ tenuous”	✓
Plot	In reality this is merely a bluff , a plot by the foreign spy organization to lure Thornhill into a false sense of security and eliminate him with the least amount of witnesses as possible .	“merely a bluff”	“eliminate”	X (i)
Plot	the plot was dry , like the other 2 of the series , but how exciting of a plot can you have with the amount of action and skilled driving packed into the movie .	“dry”	“dry”, “have”, “skilled”	X (ii)
Plot	I shall now name the similarities between these two silent classics : First of all , the films both have the same romantic plot : a clown falls in love with a fellow circus performer , while she has eyes for another .	“same”, “roman tic”	“same”, “romantic”	✓
Plot	The plot of `` Julie Johnson `` centers around two women who leave difficult marriages , move in together , develop romantic feelings for each other , and around the effects of their decisions upon each other , their friends , and their children .	No opinion found.	“develop”	X (i)
Plot	And when a good plot is stretched thin , too often do directors put 'unnecessary ' subplots throughout the movie .	“good”	“good”	✓
Plot	I 'm not a movie buff , but Deep Cover has all the elements you would expect of a classic : great acting , memorable lines , interesting (and logical) plot and a decent ending .	“interesting”, “logical”	'logical', 'lines', '(', 'decent'	X (iii)
Plot	Some important themes are well examined in the plot and though melodramatic the story is well told and holds the viewer 's interest throughout .	“well-examined”, “melodramatic”	“though”	X (i)
Plot	The plot is solid enough without being staid , the acting is good .	“solid enough”	'enough-solid'	✓
Plot	It had a really interesting plot , eye-catching visuals , and funny likeable characters .	“really interesting”	'really-interesting'	✓
Plot	An intricate plot .	'intricate'	'intricate'	✓
Plot	The main plot is that Aang wants to learn firebending from a reluctant teacher , Jong-Jong .	“main”	“main”	✓
Plot	It 's not a revolutionary plot , and a few scenes could have been better executed (the obligatory 'break-up ' before the happy ending is rather weak) , but on the whole the mushy side of me ate it up .	“not revolutionary”	'not-up revolutionary'	X (iii)

Table 4.2b– 40 samples of analysis results on CHARACTERS aspect of Amazon DVD

reviews

Aspect	Input	Expectation	Output	Result / Error
Characters	The main characters in Divided We Fall , Josef and Maria , are a married couple who end up harboring a Jewish acquaintance in their cellar for three years .	No opinion found.	“main”	X (i)

Characters	With a cast as large as this one , it 's only natural that short thrift be given to peripheral characters .	“peripheral”	“peripheral”	✓
Characters	Rocky & Bullwinkle are such dear characters , Fractured Fairytales , Aesop , Dudley Dornight and Mr. Peabody add to the wonderful time you 'll have watching this classic program .	“such dear”	“such dear”	✓
Characters	2nd , it offers some new twist in that the other characters do n't know at first who Jason is , and in one well drawn scene even we the audience do n't know .	No opinion found.	“other”	X (ii)
Characters	The first episode sets up the characters and begins to set up the plot , and does a really good job of doing so .	No opinion found.	'begins', 'does'	X (iii)
Characters	During my youth , someone had mentioned to me that Wordsworth the white-colored cat spoke in rhymes so I was quite curious about his dialogues as well as other characters ' .	No opinion found.	“other”	X(ii)
Characters	It 's almost got the sense of isolation of 'The Shining ' , but there are too many incidental characters .	“too many incidental”	'too-many', 'too-incidental'	✓
Characters	In reality , he is one of the only sympathetic characters in the movie , and the film 's climax is the only portion that packs a wallop : the script pits the only two characters in the film worth caring about against one another .	“sympathetic”	'only', 'sympathetic', 'only'	X (ii)
Characters	The actors are completely brilliant and if you accept their characters with all their quirky and often sad flaws , this film is more of a tragedy with a cast of eccentric (and often very funny) characters .	“very funny”	“very funny”	✓
Characters	Cohen came up with some wonderfully complex characters , put them in the situation , and then let the material speak for itself , which it does in an extremely funny way .	“wonderfully-complex”	“wonderfully-complex”	✓
Characters	He proceeds to make his way through London meeting up with various characters each more loathsome or desperate than the last .	“loathsome, “desperate”	'various', 'then-loathsome'	X (ii)
Characters	The citizens of Valentine Bluffs do as all horror film characters do and disregard the warning , finding themselves getting slaughtered one by one as Warden invokes his revenge .	No opinion found.	“disregard”	X (ii)
Characters	It 's also about actors - in “ Frantic ” French actors play French characters and American actors play American characters .	No opinion found.	'French', 'American'	X (ii)
Characters	The plot line does remind you of the old kung-fu movies but it has this ancient parallel universe to it where the whole world is comprised of Chinese-based characters and legends .	'Chinese-based'	'Chinese-based'	✓
Characters	No , you 're supposed to be uneasy , the lead characters do n't necessarily make it out okay & there should defiantly never eeeeeeeeeeeeeever be a happy ending .	No opinion found.	'not-eeeeeeeeeeeeever ending'	X (iii)
Characters	The acting of the main characters is superb , and their ability to deliver drama the way it is written and the way it should be is truly a gift to all fans of this show .	“superb”	“main”	X (iii)
Characters	There is a certain improvisational element to the two lead performances that contributes to this feeling and makes the characters easy to like despite upscale lifestyles that appear more appropriate in a Pottery Barn catalog .	“easy to like”	“easy”	X (i)
Characters	It does n't have many of the familiar characters but it is still great !	“no familiar”	“familiar”	X (iii)
Characters	The film is loaded with incredibly funny characters .	'incredibly-funny'	'incredibly-funny'	✓
Characters	It 's a wonderful film , but it could n't have made for an easy pitch : we 've got a group of disparate characters trying to save their run-down building from demolition by your basic big bad developer , and the miracle they all hope for comes in the form of these little alien gizmos - who not only go around fixing things , they also reproduce .	“Disparate”	“Disparate”	✓

Characters	It might make more sense if she was treated as some sort of pariah by those around her but all of the other characters seem to love her sowhat are they seeing that I 'm not ?	No opinion found	"other"	X (ii)
Characters	Most people see it because of Johnny Depp 's performance (as 2 characters) but that fact is very minor , although he does an excellent job .	"very minor"	"very minor"	✓
Characters	In my opinion one of the most underrated characters ever on `` ER , `` Jeannie has a very exciting year .	'most-underrated'	'most-underrated'	✓
Characters	`` The Mummy Returns `` abandons the characters , and the plot is utilized simply as a flimsy clothesline for action and F/X sequences .	No opinion found	'plot'	X (iii)
Characters	It 'll only serve to make your eyes bleed and I honestly do n't understand how people can give this movie five stars when the characters and story were so horribly portrayed and completely stripped of their essence .	"horribly portrayed", "completely-stripped"	"completely-stripped"	X (iii)
Characters	The cinematography is sweeping in it beauty (not unlike other films from its day such as the Sound of Music) and the actors were talented but the love story was so simplified that it lacked the emotional depth needed to give the characters arcs or even elicit sympathy .	No opinion found	'sympathy'	X (i)
Characters	I just had to deal with ignorant characters that were about to get five years in prison for their crimes , and nobody was worried .	'ignorant'	'ignorant'	✓
Characters	Kiss Of The Sabre - This is partially a fantasy episode where all the lead characters play other characters .	No opinion found	'other'	X (ii)
Characters	A lot of Indian movies have characters saying they 're in Goa and then show scenes for Switzerland .	No opinion found	'then-show'	X (i)
Characters	They create real characters !	'real'	'real'	✓
Characters	While the plot is nearly identical , Alexander Payne (director) treats his characters much harsher than Perrotta .	No opinion found	'Perrotta'	X (i)
Characters	The characters are well fleshed out and played quite well by their cast .	'well-fleshed-out' , 'well-played'	'well-played'	X (i)
Characters	There are also other characters who really add to the fun , and actors whose work you will appreciate .	'other'	'other'	✓
Characters	The characters change in between seasons a bit , but the premise remains sound , the characters true , and the writing top notch .	No opinion found	'notch'	X (i)
Characters	It contains maps of the battles and campaigns as well as a who 's who of them main characters and families , useful chapter notes and the authors appendix `` Some Words about War and Peace `` .	'main'	'main'	✓
Characters	Sidekicks are a little cheesy , but the main characters are handled well .	'main'	'main'	✓
Characters	Populated by a cast of real-life characters as colorful as any you 'll find (not least of all Ram Dass himself) , suffused with tenderness and sincere spiritual feeling , this film is a source of sadness , laughter , and somehow , tremendous comfort .	'real-life'	'real-life'	✓
Characters	To Hong Kong audiences , Li is known for playing various infallibly heroic characters , such as the folk heroes Wong Fei Hong and Fong Sai Yuk .	'infallibly-heroic'	'infallibly-heroic'	✓
characters	You really get involved with the main characters and we are anxious for the new season to begin	'main'	'main'	✓
characters	The reflexive anti-Americanism of Spain and Europe are integral to many of the interactions experienced by Taylor and Eigeman , the two main characters .	'main'	'main' , 'Eigeman'	X (iii)

*Table 4.2c – 40 samples of analysis results on CINEMATOGRAPHY aspect of
Amazon DVD reviews*

Aspect	Input	Expectation	Output	Result
--------	-------	-------------	--------	--------

cinematography	The plot is uninvolved and incoherent , the cinematography is cramped and completely lacking in style , the music is weak during a time that spawned so many memorable soundtracks and the direction lacks any focus whatsoever .	'completely-lacking', 'cramped'	'completely-lacking'	X (iii)
Cinematography	The sets are quite amazing , and the cinematography is absolutely beautiful .	'absolutely-beautiful'	'absolutely-beautiful'	✓
Cinematography	So , despite the beautiful cinematography , this movie is a hopeless muddle .	'beautiful'	'beautiful'	✓
Cinematography	Having said that , the cinematography was stunning and the music almost made the film worth watching .	'stunning'	'almost-made'	X (ii)
Cinematography	However , the awful cinematography , horrific editing , and terrible production value of this film make impossible for me to fathom why people did n't burn down the theaters when they realized this film was n't getting any better .	'awful'	'awful'	✓
Cinematography	There was good cinematography and music , and Kevin Bacon is always appealing , but that 's the best I can say .	'good'	'music-good', 'music-music', 'always-appealing', 'always-best'	X (ii)
Cinematography	Well , nice cinematography as set-designer work : nicely interiors and well dressed actors .	'nice'	'nice'	✓
Cinematography	Even the cinematography is utterly unimpressive .	'utterly-unimpressive'	'utterly-unimpressive'	✓
Cinematography	Although subtitled Grazia 's island (Grazia is the lead role , magnificently realized by Valeria Golino) , `` Respiro `` could have well been called `` Scenes from rural Sicilian life `` , as the scenography , cinematography and tableaux-like imagery seem as important to the director as her thin narrative line .	'important'	'seem'	X (i)
Cinematography	The film has some marvelous cinematography of the old Asbury Park area , and has an ending that while not totally satisfying is real .	'marvelous'	'marvelous', 'has'	X (i)
Cinematography	This film has it all , great performances , great direction , and great cinematography .	'great'	'great'	✓
Cinematography	The cinematography is absolutely beautiful .	'absolutely-beautiful'	'absolutely-beautiful'	✓
Cinematography	Reginald Rose 's original play reads like a powerhouse , even in a highschool english class , but given the direction of Sidney Lumet , the dynamite cast and the beautifully stark black and white cinematography it becomes a legend on film .	'beautifully-stark'	'beautifully-stark'	✓
Cinematography	Wonderful cinematography and a fascinating subject are undermined by uneven performances and some poor storytelling decisions , but there 's a lot to like .	'Wonderful'	'"s"	X (i)
Cinematography	In conclusion , this TV show has it all : good background music , great dialogue , interesting stories , good acting , and interesting cinematography .	'interesting'	'interesting'	✓
Cinematography	Not only are the musical numbers excellent ; the cinematography is superlative and the ending is a happy one !	'superlative'	'superlative', 'happy'	X (i)
Cinematography	I 've left out a ton of information as this film was close to three hours long , but watch the beautiful cinematography of Edward Yang , and his brilliant pacing .	'beautiful'	'beautiful', 'pacing'	X(i)
Cinematography	Ted Moore 's cinematography , especially the shot of bikini clad Ursula Andress emerging from the sea , is now a classic and also set a standard for future ` Bond Girls .	'classic'	'now-classic', 'also-set'	X(iii)
Cinematography	The soundtrack is awesome , the cinematography is great , the dialogue is realistic , and for once , a movie set in a foreign land that actually uses actors from that area speaking the language of that area !	'great'	'great'	✓

Cinematography	These three are classics , not to be confused with the fourth , which has less impact , since it tries too hard to be original , yet loses something along the way in terms of casting , acting , and cinematography , as the desert-themed atmosphere has n't worked that well .	No opinion found	'acting'	X (i)
Cinematography	Spike Lee 's cinematography is terrific .	'terrific'	'terrific'	✓
Cinematography	The beautiful cinematography , and the wonderful soundtrack featuring Elvis Presley , The Platters and many more contribute wonderfully to the film 's atmosphere of a simpler time .	'beautiful'	'beautiful'	✓
Cinematography	The cinematography by Joseph MacDonald is marvelous , the Edith Head gowns lavishly glamorous , and very important to the success of this film is the fabulous jazzy score , which is one of Elmer Bernstein 's finest .	'marvelous'	'marvelous'	✓
Cinematography	The fault in order of importance : poor screenplay , poor cinematography and poor performance by Pacino , who seems to be vying for the most over-the-top performance .	'poor'	'poor', 'performance'	X (i)
Cinematography	The soundtrack is awesome , the cinematography is great , the dialogue is realistic , and for once , a movie set in a foreign land that actually uses actors from that area speaking the language of that area !	'great'	'great'	✓
Cinematography	I was surprised by a decent story with suspense and action and some really beautiful cinematography .	'really-beautiful'	'really-beautiful', 'really-action'	X (iii)
Cinematography	Mr. Pellington falls back on some outdated `` tricks `` like good acting , excellent cinematography , and an absolutely stunning soundtrack .	'Excellent'	'soundtrack'	X(i)
Cinematography	All actors who play in this movie are the elite of Russian cinematography and hold a very special place in every Russian heart .	'Russian'	'Russian'	✓
Cinematography	The movie has a kind of organic feel to it and the cinematography is less-than-spectacular	'less-than-spectacular'	'less-than-spectacular'	✓
cinematography	The cinematography in this movie is amazing ; the camera disappears and we see real life transpiring on screen .	'amazing'	'see'	X (i)
Cinematography	But it does n't ; instead , director Andrei Zvyagintsev 's ambitions are realized through the believable acting , gorgeous cinematography , and intentionally open-ended storyline .	'gorgeous'	'gorgeous'	✓
Cinematography	No one blames you for that , and I really have no idea why you would buy this movie anyway , unless you simply love the beautiful cinematography .	'beautiful'	'beautiful'	✓
Cinematography	First , the black and white cinematography strikingly captures the mood and feel of a late 1940 's film .	'black and white'	'black'	X (i)
Cinematography	` The Joy Luck `` is a great film in all respects , the performances by the actresses , the general storyline , the cinematography and the music .	'great'	'music'	X (i)
Cinematography	The great director Hector Babenco (`` Kiss of the Spider Woman ``) also has a small role.It 's Bardem who carries this to greatness , with Schnabel 's sure hand ; cinematography was remarkably beautiful for an independent film .	'remarkably-beautiful'	'remarkably-beautiful'	✓
Cinematography	The cinematography is great , Molly Ringwald looks great , and it has a dark and peculiar charm all it 's own .	'great'	'great'	✓
Cinematography	A remake of an earlier film to be sure , but even this 1989 treatment looks and feels nostalgic with its amber-tinted cinematography , the sentimental presentation of the devoted fighter pilots , and the appearance of the ageless , magical Audrey Hepburn (sharp as a tack in her last film as a bright-eyed , no-nonsense angel) .	'amber-tinted'	'amber-tinted'	✓
Cinematography	The earthy cinematography in this film is comparable to the great murals and naturalistic portraits of Diego Rivera .	'comparable'	'comparable'	✓

Cinematography	Picture : Very good transfer captures Robert Richardson 's beautiful but gritty cinematography perfectly .	'gritty'	'perfectly-beautiful'	X (i)
Cinematography	The acting is the finest I have ever seen in a long time on film ; the cinematography is skillfully and artistically executed , propelling the ambiances , thoughts , and feelings of the various characters .	'Skillfully'	'.'	X(i)

Table 4.2d – 40 samples of analysis results on DIALOGUE aspect of Amazon DVD reviews

Aspect	Input	Expectation	Output	Result
Dialogue	Her dialogue is kept short and simple , and when she does have more than 3 lines at once , the complete lack of feeling , expression & emotion in her voice makes you consider turning off the movie and doing something worthwhile , like cleaning out the fridge or picking at your hang-nails .	'short', 'simple'	'makes'	X (iii)
Dialogue	The dialogue is at a much lower volume than the music and sound effects , making it impossible to view without constantly tinkering .	'lower volume'	'effects'	X (iii)
Dialogue	For the most part the bad dialogue and some of the terrible acting in parts were just too painful to sit through .	'bad'	'bad'	✓
Dialogue	All the characters were severely and tragically 2-dimensional with none of the depth you get from reading the graphic novel and very little attention was paid to developing better dialogue suitable for a big screen experience .	'suitable'	'suitable'	✓
Dialogue	Some of my favorite actors are wasted , either by being given cliched dialogue , over acting , or being put there merely for their brand name with no script to support their presence .	'cliched'	'cliched'	✓
Dialogue	It is a very dialogue-heavy film and the dialogue is not well-written enough to carry it off .	'not-enough well-written'	'not-enough well-written'	✓
Dialogue	Make sure all the bad guys bite off their dialogue and sound mean .	Mo opinion	'sound'	X (i)
Dialogue	The dialogue is difficult to follow especially since most of it is meaningless anyway , and the film never really follows it 's premise as much as it should : the rise of Factory .	'difficult-to-follow'	'difficult-to-follow'	X(i)
Dialogue	The production design and cinematography are sumptuos , but the script and dialogue are so banal , the fine cast delivering one empty performance after another .	'sumptuos'	'script'	X(ii)
Dialogue	While Baldwin and the rest of the cast did well in fleshing out Burke 's characters and have some tangy tough-guy/gal dialogue , they 're hampered by two things : 1 .	'tough-guy/gal'	'tough-guy/gal'	✓
Dialogue	Despite the sexy promises in the movie 's ads , none of the ladies share anything with Baldwin but dialogue and some kisses and/or embraces .	No opinion	'kisses'	X(i)
Dialogue	The English voice acting is embarassingly bad , especially the main character , but not as bad as the dialogue , internal and vocalized , that this character has been given to speak .	'bad'	'internal'	X (iii)
Dialogue	The sight gags are plentiful , the physical comedy abundant and often cruel , and the dialogue absurdly unrealistic .	'absurdly-unrealistic'	'absurdly-unrealistic'	✓
Dialogue	Anachronism , bad dialogue , bad plot and plain old bad acting plague this `` sequel `` .	'bad'	'bad'	✓
Dialogue	so I 'm supposed to believe that the horrible dialogue , bad acting , and ridiculous plotting is on purpose just because Mamet is a respected guy ?	'horrible'	'horrible'	✓
Dialogue	My real gripe about the movie was the dialogue and the acting .	'gripe'	'acting'	X (i)
Dialogue	It 's the way the story is told - its ham-fisted dialogue and plot , its forced obscenity at every turn - that reveals its teen sensibilities .	'ham-fisted'	'ham-fisted'	✓
Dialogue	The second set is strictly played for laughs -- there is little or no character development and the plots and dialogue run rather thin .	'thin'	'plots'	X (i)
Dialogue	The dialogue does n't fair any better , it 's also dull and uninspired .	'dull', 'uninspired'	'uninspired'	X(iii)

Dialogue	The plot is stupid , and the dialogue is so bad it 's over the top .	'so-bad'	'so-bad'	✓
Dialogue	There 's no joke too hackneyed , no visual too disgusting , and no dialogue too stupid to throw in the mix .	'too-stupid'	'too-stupid'	✓
Dialogue	The two dimensional nice old brother and sister are treated horribly , for no good reason with no motivation or payoff , the songs are unlistnable and the dialogue is hard to understand , this is not always a bad thing .	'hard to understand'	'hard'	X (i)
Dialogue	You do n't go to these movies to expect well-written dialogue , character depth or anything , you come for the kills and to tell you the truth , they 're slightly disappointing .	'disappointin g'	'well-written'	X (i)
Dialogue	The dialogue was sharp and you might have to watch twice to get all the subtle quips .	'sharp'	'sharp', 'have'	X (i)
Dialogue	The dialogue is wonderfully expressive and gives us the sense of authenticity and the kind of realistic effect seen only in the very best films .	'wonderfully -expressive'	'wonderfully -expressive', 'wonderfully -gives'	X (i)
Dialogue	Gould creates a sharp , amiable , chain-smoking , wise-cracking 70s detective whose 'quirky-alone ' lifestyle allow him the freedom to give his own Chandleresque , talking-to-himself dialogue (... 'it 's alright by me ... ') whilst interacting with a diverse range of characters within the narrative .	'talking-to-himself'	'talking-to-himself'	✓
Dialogue	The colors are dark , the dialogue between the characters is honest , real and simply wonderful .	'honest', 'real', 'simply-wonderful'	'real', 'simply-wonderful'	X (iii)
Dialogue	Having proven her ability to be a convincing Southerner in `` Gone With the Wind `` , Vivien Leigh expertly handles all the florid dialogue with her particular blend of defiance and vulnerability .	'florid'	'florid'	✓
Dialogue	this episode has a lot of funny dialogue , but that 's the only thing it has working for it .	'funny'	'funny'	✓
Dialogue	In conclusion , this TV show has it all : good background music , great dialogue , interesting stories , good acting , and interesting cinematography .	'great'	'great', 'stories'	X (i)
Dialogue	This is a wonderful film , with a silly but extremley funny plot , and some wonderful dialogue , particularly between Horton and Brady , who somehow manage to end up married to each other , much to their surprise .	'wonderful'	'wonderful', 'plot'	X (i)
Dialogue	The dialogue is sharp and witty , although there 's alot of swearing , that does n't mean the humour is just fowl mouthed , it 's hilarious , some of the lines go down in history as some of the funniest one liners ever .	'sharp', 'witty'	'sharp', 'witty'	✓
Dialogue	You 'll be sucked in , and surprised , within the first two minutes , and there 's very little dialogue so you need not worry about too much subtitle reading .	'very-little'	'very-little'	✓
Dialogue	It seems rare today to find such rich dialogue .	'such', 'rich'	'such', 'rich'	✓
Dialogue	The dialogue is n't hard-boiled except for a few lines delivered by Vicky .	'not- hard-boiled'	'not- hard-boiled'	✓
Dialogue	i appreciated the effects - nice touch - but i would have liked to see some real dialogue there .	'real'	'real'	✓
Dialogue	Relying on sharp dialogue and facial expressions , the actors manage to weave a gorgeous sensual dance together , both lighting up the screen and creating two of the most smolderingly complex characters ever seen in film .	'sharp'	'sharp'	✓
Dialogue	The dialogue is quick-witted , some the scenes have become classics (Allen sneezing on the cocaine) .	'quick-witted'	'quick-witted'	✓
Dialogue	The soundtrack is awesome , the cinematography is great , the dialogue is realistic , and for once , a movie set in a foreign land that actually uses actors from that area speaking the language of that area !	'realistic'	'for-realistic'	X (ii)
Dialogue	Some of my favorite actors are wasted , either by being given cliched dialogue , over acting , or being put there merely for their brand name with no script to support their presence .	'cliched'	'cliched'	✓

4.3 Analysis

From the 160 data samples above, an approximate idea of the performance of the opinion extractor towards Amazon DVD reviews could be observed, which is summarized in the following chart.

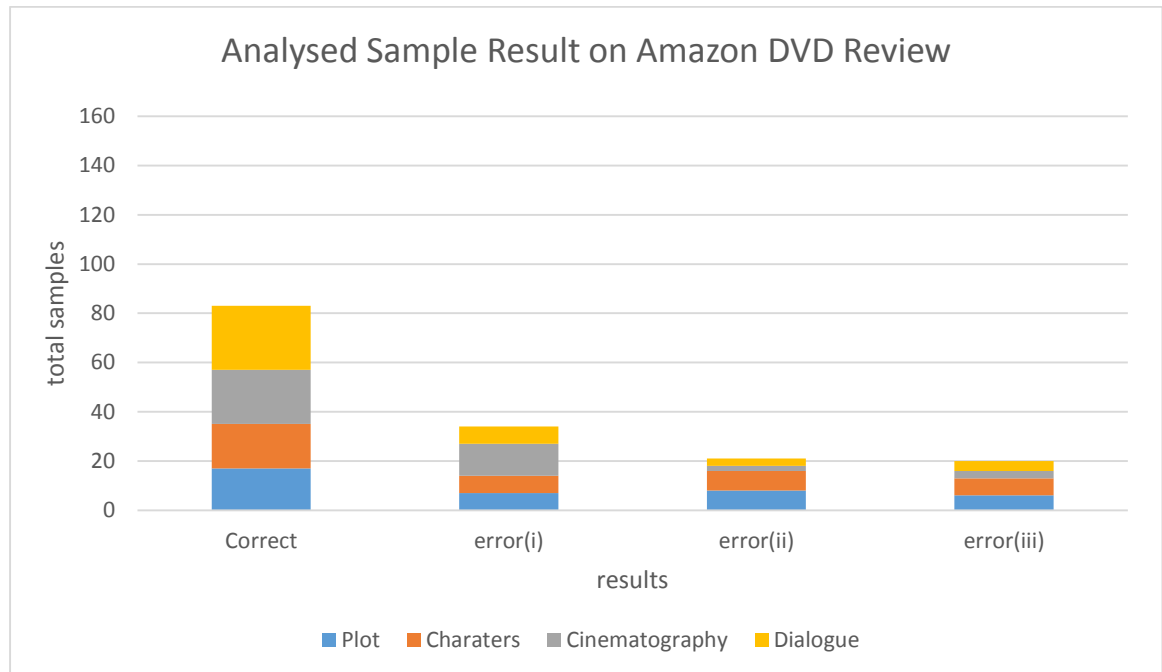


Figure 4.3a – a chart of Analysed Sample Result on Amazon DVD Review

As seen above, there are about 41.25% of opinion analysed from the extractor contains error, which will be justified and discussed in the next section.

5. Limitation

Although 5 extension have already been successfully implemented into the opinion extractor, the test set isn't detailed enough to test a sufficient amount of possible inputs, leading to some unreasonable output from the extractor.

There are 3 main reasons to explain these errors.

5.1 Deficiency in the opinion extractor algorithm

The first mistake that causes errors, is the deficiency of the extractor algorithm. Since some of the cases in 5 extensions are not being considered, the algorithm is not designed to deal with these unconsidered problems.

5.1.1 The Conjunction Issues

The conjunctions problem is one of the issues that haven't been well investigated before while designing the algorithm. Under observation during the last progress, most analysis errors occur under the problem of conjunction. An example is picked from the analysed data to explain this problem.

Example

```

--- Sentence: The film has some marvelous cinematography of the old Asbury Park area , and has an ending that while not totally satisfying is real . ---
1  The      DT      2      det
2  film     NN      3      nsubj
3  has      VBZ     0      root
4  some     DT      6      det
5  marvelous JJ      6      amod
6  cinematography NN  3      dobj
7  of       IN      6      prep
8  the      DT      12     det
9  old      JJ      12     amod
10 Asbury   NNP     11     nn
11 Park     NNP     12     nn
12 area     NN      7      pobj
13 ,        ,       3      punct
14 and      CC      3      cc
15 has      VBZ     3      conj
16 an       DT      15     dobj
17 ending   VBG     16     amod
18 that     IN      24     dep
19 while    IN      24     mark
20 not      RB      22     neg
21 totally  RB      22     advmod
22 satisfying VBG    24     csubj
23 is       VBZ     24     cop
24 real     JJ      16     dep
25 .        ,       3      punct
Opinions: ['marvelous', 'has']

```

Figure 5.1.1a - example on conjunction issue on cinematography aspect

In this example sentence in figure 5.1.1a, the token “has”, numbered line 15, was mistakenly appended into the opinion list since it was attached to the head token of the aspect token, with the relationship “conj”, which indicates a conjunction dependency relationship.

As this problem was not expected during the coding process of the opinion extractor, I will justify the reason of including all conjunctions to the opinion list.

Justification

```

--- Sentence: The plot was cheesy , but fun and inspiring . ---
1  The      DT      2      det
2  plot     NN      4      nsubj
3  was      VBD     4      cop
4  cheesy   JJ      0      root
5  ,        ,       4      punct
6  but      CC      4      cc
7  fun      NN      4      conj
8  and      CC      4      cc
9  inspiring VBG     4      conj
10 .        ,       4      punct
Opinions: ['cheesy', 'fun', 'inspiring']

```

Figure 5.1.1b - example from test set

Since not every conjuncts are adjective, therein lies the problem. As shown in the above parsed sentence, “cheesy”, “fun”, and “inspiring” are all opinions of plot, while “fun” and “inspiring” are not tagged as adjectives, but still in a conjunction dependency relationship with “cheesy”. Therefore if all conjuncts related to the aspect token are to be found by the opinion extractor, some of it might not be relevant.

5.1.2 The “No” Issue

The “no” issue have been neglected during the algorithm designing process. Here is an example of how this issue could lead to error.

Example

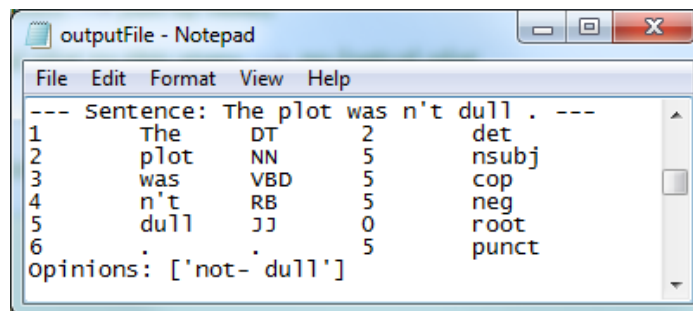
```

--- Sentence: Comparing it to Mary Poppins is silly, for Mary Poppins was a light-hearted
comedy with no real plot, where Bedknobs and Broomsticks features a fantastic plot and
clever dialogue. ---
1 Comparing VBG 7 csubj
2 it PRP 1 dobj
3 to TO 1 prep
4 Mary NNP 5 nn
5 Poppins NNP 3 pobj
6 is VBZ 7 cop
7 silly RB 0 root
8 , punct 7
9 for IN 15 mark
10 Mary NNP 11 nn
11 Poppins NNPS 15 nsubj
12 was VBD 15 cop
13 a DT 15 det
14 light-hearted JJ 15 amod
15 comedy NN 7 dep
16 with IN 15 prep
17 no DT 19 det
18 real JJ 19 amod
19 plot NN 16 pobj
20 , punct 19
21 where WRB 25 advmod
22 Bedknobs NNP 25 nsubj
23 and CC 22 cc
24 Broomsticks NNP 22 conj
25 features NNS 19 rcmmod
26 a DT 28 det
27 fantastic JJ 28 amod
28 plot NN 25 dobj
29 and CC 28 cc
30 clever NN 31 nn
31 dialogue NN 28 conj
32 : punct 7
Opinions: ['real', 'fantastic']

```

Figure5.1.2a - example on “no” issue on plot aspect

In this case, the sentence in figure5.1.2a “for Mary Poppins was a light-hearted comedy with no real plot...” should be considered “the plot is not real” for the opinion of the sentence. However, one of the output is “real” instead of “no-real-plot”, which has a totally different meaning.

Justification*Figure 5.1.2b - example from test set*

From this example sentence from the test set, “not” is considered as the “n’t” in “wasn’t” is in a negation dependency relationship with “was”, while “no” is in another dependency relationship with “plot”. As a result, the integrated extensions only considered negation, but “no” was not taken into account.

5.2 Errors made by the part-of-speech tagger

Various mistake was made by the part-of-speech (PoS) tagger leading to unwanted results in the opinion extractor, some due to lexical ambiguity, some due to suffixes problem.

Part-of-Speech (PoS) tagging is a process to decide the part of speech of a token. It is fairly important as it decides the property of a token so the extractor knows where the opinion words are.

While a token could have different meaning in different parts of a sentence, confusion could occur by the part-of speech tagger, which will be demonstrated as the followings.

5.2.1 Syntactic ambiguity

Example

```

--- Sentence: The plot is at the same time simple and deep , should you
choose to look beyond the surface . ---
1      The      DT      2      det
2      plot     NN      3      nsubj
3      is       VBZ     0      root
4      at       IN      3      prep
5      the      DT      7      det
6      same     JJ      7      amod
7      time     NN      4      pobj
8      simple   NN      7      dep
9      and      CC      3      cc
10     deep     VB      3      conj
11     ,        ,      3      punct
12     should   MD      14     aux
13     you      PRP     14     nsubj
14     choose   VB      3      ccomp
15     to       TO      16     aux
16     look     VB      14     xcomp
17     beyond   IN      16     prep
18     the      DT      19     det
19     surface  NN      17     pobj
20     Opinions: ['deep']

```

Figure5.2.1a - example on syntactic ambiguity

In this sample parsed sentence, “The plot is at the same time simple and deep...” , “simple” and “deep” should both be the adjective of “plot”. However, only “deep” in found.

Justification

This explanation of this problem could be syntactic ambiguity, which means the confusion due to alternate syntactic sentence structure. Considering the example sentence, “plot” itself could have different meanings which could leads to two different meaning of the whole sentence: (1) the plot is simple and deep as the message behind it ; (2) the plot as a district has two names “simple” and “deep” at the same time. This could explain the reason of why the PoS tagger have tagged the token “simple” as a noun.

Syntactic ambiguity problem would be hard to solve as there are infinite ways of expressing a sentence.

5.2.3 Suffixes Issue

Example

```

--- Sentence: The colors are dark , the dialogue between the characters is honest , real and simply
wonderful . ---
1      The      DT      2      det
2      colors   NNS     4      nsubj
3      are      VBP     4      cop
4      dark     JJ      0      root
5      ,        ,        4      punct
6      the      DT      7      det
7      dialogue NN      12     nsubj
8      between IN     7      prep
9      the      DT     10     det
10     characters NNS   8      pobj
11     is       VBZ    12     cop
12     honest   JJS     4      ccomp
13     ,        ,        12     punct
14     real     JJ      12     conj
15     and      CC      12     cc
16     simply   RB      17     advmod
17     wonderful JJ     12     conj
18     .        .        4      punct
opinions: ['real', 'simply-wonderful']

```

Figure 5.2.3a - example on suffixes issue on Dialogue aspect

In this sentence, “...the dialogue between the characters is honest”, “dialogue” is in a “nsubj” dependency relationship with “dialogue”. However, “honest” is tagged as JJS and is not found by the opinion extractor.

Justification

In this case, “honest” is tagged as JJS, which means superlative adjective (eg. Biggest, tallest etc.) instead of Adjective. The reason behind this could be that since “honest” contains a suffix “est”, the PoS tagger could have interpreted wrongly as “hon-est”, resulting a mistaken PoS tagged.

5.3 Errors made by the dependency parser

After a token have gone through the PoS tagging process, it will be pared by the dependency parser to create relationship between different tokens in a sentences. While there is infinite ways to interpret a sentence, errors might occur depending on how the parser is trained to recognise different sentence structures.

In the followings, examples will be found on how dependency parser have made errors and causing incorrect analysis.

```

--- Sentence: I 'm not a movie buff , but Deep Cover has all the elements you would expect of a
classic : great acting , memorable lines , interesting ( and logical ) plot and a decent ending
.
1      I      PRP      6      nsubj
2      'm     VBP      6      cop
3      not    RB       6      neg
4      a      DT       6      det
5      movie  NN        6      nn
6      buff   NN        0      root
7           ,      ,      6      punct
8      but    CC       6      cc
9      Deep   NNP      10     nn
10     Cover  NNP      11     nsubj
11     has    VBZ      6      conj
12     all    DT       14     predet
13     the    DT       14     det
14     elements NNS     11     dobj
15     you    PRP      17     nsubj
16     would  MD       17     aux
17     expect VB       14     rcmmod
18     of     IN       17     prep
19     a      DT       20     det
20     classic JJ      18     pobj
21     :      :        6      punct
22     great  JJ       23     amod
23     acting NN       0      root
24           ,      ,      23     punct
25     memorable JJ     26     amod
26     lines  NNS     23     conj
27           ,      ,      23     punct
28     interesting JJ    29     amod
29     (      NN       23     conj
30     and    CC       23     cc
31     logical JJ     33     amod
32     )      NN       33     nn
33     plot   NN       23     conj
34     and    CC       23     cc
35     a      DT       36     det
36     decent NN      23     conj
37     ending VBG     36     partmod
38     :      :        23     parataxis
Opinions: ['logical', 'lines', '(', 'decent_']

```

Figure5.3a - example on plot aspect

In Figure5.3a, from the sentence "...interesting (and logical) plot...", it was normally interpreted as "the plot is interesting and logical, where "interesting" and "logical" should be attached to "plot" since "plot" is the noun subject. However, only "logical" have been attached to "plot", and "interesting" is mistakenly attached to "(" which should not have happened.

```

--- Sentence: Aside from the stirring performances of Dustin Hoffman , Lawrence Olivier and William Devane
the cinematography and editing reminded me that they used to make movies for adults . ---
1  Aside      NNP      30      nsubj
2  from       IN       1       prep
3  the        DT       5       det
4  stirring   NN       5       nn
5  performances NNS     2       pobj
6  of         IN       5       prep
7  Dustin     NNP      8       nn
8  Hoffman    NNP      6       pobj
9  ,          ,       8       punct
10 Lawrence'   NNP      11      nn
11 Olivier    NNP      8       conj
12 and        CC       8       cc
13 William    NNP      14      nn
14 Devane     NNP      8       conj
15 ,          ,       8       punct
16 the        DT       17      det
17 cinematography NN     8       appos
18 and        CC       1       cc
19 editing    VBG     20      auxpass
20 reminded   VBN     1       conj
21 me         PRP     20      dobj
22 that       IN       24      complm
23 they       PRP     24      nsubj
24 used       VBD     20      ccomp
25 to         TO       26      aux
26 make       VB       24      xcomp
27 movies     NNS     26      dobj
28 for        IN       26      prep
29 adults     NNS     28      pobj
30 .          .       0       root
Opinions: ['Olivier', 'Devane']

```

Figure5.3b - example on cinematography aspect

In Figure5.3b, the sentence “Aside from the stirring performance of Dustin Hoffman, Lawrence Olivier and William Devane...”, two of the names have been incorrectly attached by the dependency parser. When looking at line 11 of the parsed sentences, “Olivier” in “Lawrence Olivier” was attached to “Hoffman” which is another irrelevant name. This would lead to error since “Olivier” should be attached to “Lawrence” with the noun relationship since it is the first and last name.

6. Summary

This report provides an overall introduction to an opinion extractor. A number of samples are picked from the Amazon DVD review to investigate the performance of the extractor. Several limitations of the extractor have been found, resulting from the deficiency of the algorithm as well as errors made by the part-of-speech tagger and dependency parser. These limitations have been further justified and discussed.

Due to time and work load limitations, only an approximate analysis have been made for the performance of the extractor. In the future, a more precise investigation and improvements based on the limitations mentioned in the above sessions are recommended and hopefully will lead to a more detailed understanding towards opinion extracting and natural language processing.

7. A Proposal for a DVD Summary Website

Background and Motivation

This proposal aims at proposing to build a website that automatically generates a summary of reviews towards a DVD.

With the non-stopping expansion of the Internet, reviews of DVDs are ubiquitous all over social platforms and e-shopping websites, such as Twitter and Amazon. Together with the development on language processing in the IT industry, it would be pleasant to use these technologies to extract opinions out of these reviews and classify them to produce summaries on varies DVD on the market. To implement this, a website would be one of the most convenient way for clients to easily access it.

Proposed Content

The website targets on two kinds of potential clients in two ways:

(1) General Public

A general summary of DVDs under different aspects will be shown on the website for general public how wants to know a rough opinion of a certain DVD from other public audience. This includes a top-ten most mentioned opinion as well as a rating for the DVD, based on the number of negative and positive opinions.

(2) Business Companies

This will be a paid version for DVD summaries. Clients paid will be able to see the ranking of all opinion in both positive and negative side, as well as the trend of opinion throughout the last 3 years. All statistics will be generated into charts and tables for presentation.

Based on the above, the following are a summary on functionalities behind the DVD summary website:

- Build a database to store reviews from Amazon and Twitter
- Automatic program to generate DVD review opinions in different aspects, using opinion extractor and word classifier
- Automatically generated ranking, rating, and the trend within a time period, of opinions of a certain DVD review summary.

Limitation

Accuracy would be one of the biggest problem for the DVD review summary website. Currently, the accuracy of an opinion extractor in Amazon DVD reviews is about 50%. Although the accuracy of the opinion extractor is not very high, it can still be improved further by word filtering and enhancement. The idea behind this website is still achievable under further developments of the software.