

Project 3: Using a Feature Grammar

Due date November 20, 2020

Goal: develop a feature grammar for sentiment analysis guided by a corpus

Data: you should be guided by examples in the NLTK data > `sentence_polarity` > `rt-polarity.neg`
You may simplify data samples to focus on the topic of the assignment:

1. *this may not have the dramatic gut-wrenching impact of other holocaust films , but it's a compelling story , mainly because of the way it's told by the people who were there .*
 - *it's a compelling story*
 - *it has low impact*
 - *it has low impact but it's a compelling story*
 - *it's a compelling story , but it has low impact*
 - *it has gut-wrenching impact and it is a compelling story*
 - *this does not have gut-wrenching impact but it's a compelling story*
 - *this compelling story with gut-wrenching impact*
2. *a perfect example of rancid , well-intentioned , but shamelessly manipulative movie making .*
 - *a perfect example*
 - *manipulative movie making*
 - *shamelessly manipulative movie making*
 - *well-intentioned movie making*
 - *rancid movie making*
 - *well-intentioned but manipulative movie making*
 - *a perfect example of well-intentioned bu manipulative movie making*

Note that these simplified snippets may have different sentiment values.

Description: You are to develop a feature grammar for NLTK's feature based Earley's Chart parser (parser class "FeatureEarleyChartParser").

You are trying to develop the best grammar you can: wider coverage, less acceptance of ungrammatical information, and you are trying to connect sentiment and discourse relation information.

You may use the NLTK feature grammars `feat0.fcfg`, `feat1.fcfg`, `simple-sem.fcfg` to get started, but you may choose to start from scratch. Use the Student Exchange Forum regularly, no matter how you proceed.

Required coverage

1. simple declarative sentence with a sentiment-bearing adjective (*He gave her an ugly sweater .*)
2. conjunction (*and, or, but*) of sentiment bearing adjectives or nouns (*I saw a dull and scary movie, It was a mess and a hazard*)
3. conjunction (*and, or, but*) of sentiment bearing sentences (*It was too long but entertaining*)

NLTK resources: NLTK Ch 9 explains how to write a feature grammar in detail. It is required reading for the last two projects.

In addition, you should use the feature grammars in NLTK, as well as the opinion lexicon and the sentence polarity dataset.

- **opinion_lexicon**

A list of positive and negative opinion words or sentiment words for English.

```
>>> from nltk.corpus import opinion_lexicon
>>> opinion_lexicon.words()[:4]
['2-faced', '2-faces', 'abnormal', 'abolish']
```

- **sentence_polarity**

The Sentence Polarity dataset contains 5331 positive and 5331 negative processed sentences.

```
>>> from nltk.corpus import sentence_polarity
>>> sentence_polarity.sents()
[['simplistic', ',', 'silly', 'and', 'tedious', '.'], ["it's", 'so', 'laddish',
'and', 'juvenile', ',', 'only', 'teenage', 'boys', 'could', 'possibly', 'find',
'it', 'funny', '.'], ...]
>>> sentence_polarity.categories()
['neg', 'pos']
>>> sentence_polarity.sents()[1]
["it's", 'so', 'laddish', 'and', 'juvenile', ',', 'only', 'teenage', 'boys',
'could', 'possibly', 'find', 'it', 'funny', '.']
```

Deliverables:

Create a file *Good* with your training sentences that your grammar parses and labels correctly.

Create a file *False* with your training sentences that your grammar does not parse or label correctly.

- 1 file: your well-annotated grammar (4pts, Grad Attr. 4,5,6)
- 1 file: *Good* with annotations (3pts, Grad Attr. 4,5,6)
- 1 file: *False* with annotations (1pt, Grad Attr. 4,5,6)

- 1 file: A report that includes:
 - one page justifying your design (Why did you make the grammar do what it does? Why did you make it not do what it doesn't?) (2pts, Grad Attr. 1, 6)
 - one page critiquing your design (What does your grammar not do that you think important?) (1pt, Grad Attr. 1, 6)
 - one page of ideas what kind of semantics you can do with your grammar (and how much additional work that would require) (1pts, Grad Attr. 1, 6)

Page limits are strict.

Negative starter data from NLTK data > sentence_polarity > rt-polarity.neg

*clayburgh and tambor are charming performers ; neither of them deserves eric schaeffer .
 a pale xerox of other , better crime movies .
 . . . a hokey piece of nonsense that tries too hard to be emotional .
 illiterate , often inert sci-fi action thriller .
 a perfect example of rancid , well-intentioned , but shamelessly manipulative movie making .
 the adventure doesn't contain half the excitement of balto , or quarter the fun of toy story 2 .
 essentially a collection of bits – and they're all naughty .
 a mess . the screenplay does too much meandering , norton has to recite bland police procedural
 details , fiennes wanders around in an attempt to seem weird and distanced , hopkins looks like
 a drag queen .*

Positive starter data from NLTK data > sentence_polarity > rt-polarity.pos

*a mostly intelligent , engrossing and psychologically resonant suspenser .
 it's this memory-as-identity obviation that gives secret life its intermittent unease , reaffirming
 that long-held illusions are indeed reality , and that erasing them recasts the self .
 hip-hop has a history , and it's a metaphor for this love story .
 in scope , ambition and accomplishment , children of the century . . . takes kurys' career to a
 whole new level .
 this may not have the dramatic gut-wrenching impact of other holocaust films , but it's a com-
 pelling story , mainly because of the way it's told by the people who were there .
 between the drama of cube ? s personal revelations regarding what the shop means in the big
 picture , iconic characters gambol fluidly through the story , with charming results .
 a gentle , compassionate drama about grief and healing .
 somewhere short of tremors on the modern b-scene : neither as funny nor as clever , though
 an agreeably unpretentious way to spend ninety minutes .
 digital-video documentary about stand-up comedians is a great glimpse into a very different
 world .
 unlike most teen flicks , swimming takes its time to tell its story , casts mostly little-known
 performers in key roles , and introduces some intriguing ambiguity .
 an enthralling , playful film that constantly frustrates our desire to know the 'truth' about this
 man , while deconstructing the very format of the biography in a manner that derrida would
 doubtless give his blessing to .*

Project 4: Extending a Feature Grammar

Due date December 4, 2020

Goal: extend your feature grammar for sentiment analysis to cover more complex cases. Compare resulting sentiment analysis to a baseline system.

Data: you should be guided by examples in the NLTK data > `sentence_polarity` > `rt-polarity.neg`. You may simplify data samples to focus on the topic of the assignment, as for Project 3, but for Project 4, you should attempt more complex, longer sentences and also paragraphs.

Description: You are to extend your feature grammar for NLTK's feature based Earley's Chart parser (parser class "FeatureEarleyChartParser").

You are trying to develop the best grammar you can: wider coverage, less acceptance of ungrammatical information, and you are trying to connect sentiment and discourse relation information.

Required coverage

1. complex declarative sentences with a sentiment-bearing words
2. conjunction (*and, or, but*) of sentiment bearing adjectives or nouns (*I saw a dull and scary movie, It was a mess and a hazard*)
3. conjunction (*and, or, but*) of sentiment bearing sentences (*It was too long but entertaining*)
4. negation (of text including sentiment bearing words)
5. small paragraphs of text that contain sentiment words

Data: you may continue to use the polarity dataset of Project 3 as inspiration, or you may consult the movie.reviews dataset in NLTK.

SSAP baseline: run the aFinn Simplest Sentiment Analysis in Python

- download the aFinn sentiment lexicon (https://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010)
- get the SSAP to work (<https://finnaarupnielsen.wordpress.com/2011/06/20/simplest-sentiment-analysis-in-python-with-af/>)
- use your sentences for Project 3 and 4 and record the performance of SSAP
- record the performance of SSAP as a baseline and compare the performance of your Project 4 in both tabular form and with error analysis on 10 examples of different complexity

Deliverables:

Create a file *Good* with your training sentences that your grammar parses and labels correctly.

Create a file *False* with your training sentences that your grammar does not parse or label correctly.

- 1 file: your well-annotated grammar (2pts, Grad Attr. 4,5,6)
- 1 file: *Good* with annotations (2pts, Grad Attr. 4,5,6)
- 1 file: *False* with annotations (2pts, Grad Attr. 4,5,6)
- 1 file: A report that includes:
 - one page explaining and critiquing your grammar design for complex declarative sentences (2pts, Grad Attr. 1, 6)
 - one page explaining and critiquing your grammar design for conjunction (*and*, *or*, *but*) of sentiment bearing adjectives or nouns (*I saw a dull and scary movie*, *It was a mess and a hazard*) with and without negation (2pts, Grad Attr. 1, 6)
 - one page explaining and critiquing your grammar design for conjunction (*and*, *or*, *but*) of sentiment bearing sentences (*It was too long but entertaining*) with and without negation (2pts, Grad Attr. 1, 6)
 - one page comparing SSAP and your grammar-based sentiment analysis of small paragraphs of text that contain sentiment words (2pt, Grad Attr. 1, 6)

Note on grading: if your Project 4 is better than your Project 3, it will count for both projects