Assignment-Discussion Vector Based WSD

200050006 – Ameeya Ranjan Sethy

200050053 – Jasani Parth Alpeshbhai

200050063 – Kumari Pragati Gupta

27th November, 2022

First Baseline

- Most Frequent Sense (MFS)

- Precision : 0.60

- Recall : 0.70

- F2-score : 066

Vector Based Extended Lesk

- PS : For given word sequence of words produce synset IDs of word which POS tag is NOUN.

- Data : semcor

- Algorithm1 : Extended Lesk

- Algorithm2 : Page-Rank

Performance report of extended lesk

Metrics	value
Precision	0.7172896749077547
Recall	0.5928444733857615
F1-score	0.6267663941848407
Accuracy	0.420788521427609

Performance report for Rank Based WSD

Metrics	value
Precision	0.6994952164393669
Recall	0.5876234768882604
F1-score	0.621535092767554
Accuracy	0.379824406667255

Confusion Cases (for Extended Lesk)

- Sentence : The efforts of the various interest groups to control or influence governmental decision, particularly when taken in conjunction with the impact of industrialization, led to a concentration of attention on the legislative power and the means whereby policy could be formulated and enforced as law through bureaucratic institutions.
  - Word : power
  - Actual : power.n.05 : one possessing or exercising power or influence or authority
  - Predicted : power.n.05 : supply the force or power for the functioning of

- Sentence : Mr Skyros did not at all like the look on angelo's regular featured, almost girlishly good-looking face- or indeed anything about Angelo.
  - Word : face
  - Actual : face.n.01 : the front of the human head from the forehead to the chin and ear to ear
  - Predicted : confront.v.03 : present somebody with something, usually to accuse or criticize
- Sentence : The bone was found under the ash tree.
  - Word : ash
  - Actual : ash.n.03 : strong elastic wood of any of various ash trees; used for furniture and tool handles and sporting good such as basketball bats.
  - Predicted : ash.n.01 : The residue that remains when something is burned.

- Sentence : The cops arrested the person.
  - Word : cops
  - Actual : bull.n.05 : uncomplimentary terms for a policeman
  - Predicted : collar.v.01 : take into custody
- Sentence : Now you don't want to go talking that way
  - Word : way
  - Actual : manner.n.01 : how something is done or how it happens
  - Predicted : way.n.01 : doing as one pleases or chooses

Confusion Cases (for Page-Rank)

- Sentence : There was a fellow named blatz over smithtown way.
  - Word : way
  - Actual : direction.n.01 : a line leading to a place to point
  - Predicted : manner.n.01 : how something is done or how it happens
- Sentence : Up to date, however, his garden was still more or less of a mess, he hadn't even started his workshop and if there was a meadow pond in the neighbourhood he hadn't found it.
  - Word : mess
  - Actual : mess.n.01 : a state of confusion and disorderliness
  - Predicted : batch.n.02 : (often followed by `of) a large number or amount or extent

- Sentence : Phil started to take his clothes off and frankie sat down on the bench again.
  - Word : bench
  - Actual : bench.v.01 : a long seat for more than one person
  - Predicted : bench.v.01 : take out of a game; of players
- Sentence : what the hell right did eddie have saying a thing like that?
  - Word : right
  - Actual : right.n.01 : an abstract idea of that which is due to a person or governmental body by law or tradition or nature
  - Predicted : right.n.04 : those who support political or social or economic conservatism; those who believe that things are better left unchanged
- Sentence : phil shut the door behind him.
  - Word : door
  - Actual : door.n.01 : a swinging or sliding barrier that will close the entrance to a room or building or vehicle
  - Predicted : dorrway.n.01 : the entrance through which you enter or leave a room or building; the space that a door can close

Interpretation of confusion (error analysis : Extended lesk)

- For some words like 'face' and 'cops', POS in which it is used is sometimes predicted wrongly.

- Some function words like 'were' predicted as abbreviations due to the context in which they were used like 'in' was thought to be 'inch.n.01' and 'or' was thought to be operation room.

- Senses having very closely related definitions are often mispredicted.

- Sometime context of a word may not have maximum similarity with correct sense. Ex : The bone was found under ash tree.

- Dictionary entries rarely take into account the distributional constraints of different word senses. Ex : cigarette and ash never co-occur in a dictionary.

Interpretation of confusion (error analysis : Page Rank)

- Words like 'own' and 'have', 'yield' and 'give', 'receive' and 'get' are used interchangeably by humans which leads to confusion.

- To calculate similarity between two nodes of a graph, we are using only the definition of the sense and not any other relation like hyponyms or hypernyms which can lead to lower accuracy.

- The use of POS tags can also reduce the search space and give better accuracy because sometimes the POS disambiguation itself is wrong which gives rise to wrong sense.

- The sense like 'door.n.01' and 'doorway.n.01' have very similar words in their definition which leads to similar score with both senses.

Data Processing info

- **<u>Obtaining word-vectors</u>**
  - We used word2vec model trained on google news dataset and for extracting word vectors from that we used genism module

- We created dictionary of word vector of semcor dataset and for calculating cos-similarity of sentence we average of all word vector of word in that sentence. And for any sentence word vector is not present for any word then we return zeros of same shape.

GUI – Extended Lesk

```python
test_sentence = "She likes to play cricket"
tokens = word_tokenize(test_sentence)
for word in tokens:
  pred_sense = predict_synset_extended(word, tokens)
  if(pred_sense!='notag'):
    print("Target Word : " + word + "\nPredicted Sense : " + pred_sense + " : " + wn.synset(pred_sense).definition() + "\n")
  else:
    print("Target Word : " + word + "\nPredicted Sense : " + pred_sense + "\n")
```

```
Target Word : She
Predicted Sense : notag

Target Word : likes
Predicted Sense : like.n.02 : a kind of person

Target Word : to
Predicted Sense : notag

Target Word : play
Predicted Sense : playing_period.n.01 : (in games or plays or other performances) the time during which play proceeds

Target Word : cricket
Predicted Sense : cricket.n.02 : a game played with a ball and bat by two teams of 11 players; teams take turns trying to score runs
```

GUI – Page Rank

```python
test_sentence = "She is a government servant"
tokens = word_tokenize(test_sentence)
pred_dict = prediction(tokens)
for x,y in pred_dict.items():
  if(y!='notag'):
    print("Target Word : " + x + "\nPredicted Sense : " + y + " : " + wn.synset(y).definition() + "\n")
  else:
    print("Target Word : " + x + "\nPredicted Sense : " + y + "\n")
```

```
Target Word : She
Predicted Sense : notag

Target Word : is
Predicted Sense : be.v.12 : to remain unmolested, undisturbed, or uninterrupted -- used only in infinitive form

Target Word : a
Predicted Sense : deoxyadenosine_monophosphate.n.01 : one of the four nucleotides used in building DNA; all four nucleotides

Target Word : government
Predicted Sense : government.n.03 : (government) the system or form by which a community or other political unit is governed

Target Word : servant
Predicted Sense : servant.n.01 : a person working in the service of another (especially in the household)
```