

Natural Language Processing - CSE556

Output

MONSOON SEMESTER 2020

ANUNAY YADAV - 2018021
HAMZAH AKHTAR - 2018051

Extended Lexicon:

1. Word2Vec:

The complete extended L1 lexicon for the word2vec model is reported in the file ``L1_word2vec.csv``

- ('is', 'है', 'positive')
- ('on', 'पर', 'positive')
- ('many', 'कई', 'positive')
- ('its', 'इसकी', 'positive')
- ('from', 'से', 'positive')
- ('but', 'लेकिन', 'positive')
- ('from', 'से', 'negative')
- ('very', 'बहुत', 'positive')
- ('at', 'पर', 'negative')
- ('on', 'पर', 'negative')
- ('s', 'है', 'positive')
- ('girl', 'लड़की', 'positive')

Total addition for the word2vec model are 12

2. Glove:

The complete extended L1 lexicon for the word2vec model is reported in the file ``L1_glove.csv``

- ('option', 'विकल्प', 'positive')
- ('very', 'बहुत', 'positive')

Total addition for the glove model are 2

3. Total additions (Glove + Word2Vec) :

The complete extended L1 lexicon for the word2vec model is reported in the file ``L1_total.csv``

- ('is', 'है', 'positive')
- ('on', 'पर', 'positive')
- ('many', 'कई', 'positive')
- ('its', 'इसकी', 'positive')
- ('from', 'से', 'positive')
- ('but', 'लेकिन', 'positive')
- ('from', 'से', 'negative')
- ('very', 'बहुत', 'positive')
- ('at', 'पर', 'negative')
- ('on', 'पर', 'negative')
- ('"s", 'है', 'positive')
- ('girl', 'लड़की', 'positive')
- ('option', 'विकल्प', 'positive')
- ('very', 'बहुत', 'positive')

The total number of unique additions for both the model are 14

Naming Convention:

- L1.csv contains all english-hindi-polarity pairs.
- L1_glove.csv contains all English-Hindi-polarity pairs found by the glove model.
- L1_word2vec.csv contains all English-Hindi-polarity pairs found by word2vec model
- Lexicon.py is the base script which executes both the models.
- Model folder contains word2vec and Glove model
- Preprocessed folder contains lemmatized english.txt and hindi.txt
- Glove-master contains glove model.