Nueral Discourse Relation Classification Report

Wei Lu

November 4, 2018

In this project, I am trying to classify the discourse relation between two adjacent or non-adjacent discourse units using neural network models. I am very grateful to the TAs Yuchen Zhang and Che Yeol Chun, who helped me understand the course material more deeply and provided very helpful advice for my implementation. I particularly acknowledge the discussion with Che Yeol who made me realize the importance the connective words, which improved the model performance a lot.

1 Feature representation, neural network architecture, and hyper-parameters design

1.1 Feature representation

I used the embeddings of words to represent each sentence which is composed of connective (5 words), arg1 (30 words) and arg2 (30 words). So I truncated each sentence into a 65-word sentence with first 5 words as connective (0's are padding when the number of connective words is less than 5). I tried both random embeddings and pretrained google embeddings and found that the random embeddings works a little better. (This is elaborated in the experiments in the third part of the report)

1.2 Neural network architecture

I used a one layer fully connected neural network with 300 nodes and used cross entropy with L2 regularizers as the objective function. The optimizer is AdamOptimizer.

1.3 Hyper-parameters design

The batch size is 32, learning rate for AdamOptimizer is 0.0001. Regularizers coefficient is 0.01. Word embed size is 300.

2 Code structure

I defined the following functions:

```
def vocab_sense(rows):
```

which extracts the vocabulary and all the senses which appeared in the training set. A dictionary was used to store the vocabulary with word as key and its index as value. (When I used pretrained embedding, the corresponding value is the vector of the word).

```
def row2vec(row,vocab):
```

which featurizes a sentence composed of arg1, connective and arg2. It changes each word in the sentence into its corresponding index or 0 if it a word unseen.

```
def main(train_x_list, train_y_list, dev_x_list, dev_y_list, test_x_list,
 test_y_list, vocab_size):
```

which is the main part of the code. It is composed of two parts: first, it builds a feed-forward neural network architecture (symbolic programming). Secondly is the training part, which defines the cross entropy loss, optimizer (AdamOptimizer with L2 regularizers), and does the training.

You can run the code in the terminal with
"python relations_random_embedding_regul_300nodes_conn5_arg1_30_arg2_30.py", then a json file named "output_random_embedding_regul_300nodes_conn5_arg1_30_arg2_30.json" will be generated and can be used in evaluation by running the scorer.py programme.

3 Experiments

3.1 feature representation

I tried two methods for word embedding: self trained and pretrained google embeddings, and found that the self trained embeddng words a little better (although intuitively the pretrained embeddings should be more meaningful). For the length of each sentence I did two experiments: one is truncating each sentence using the first 30 words, another one is paying more attention to the connective words (after I discussed it with Che Yeol). Specifically, each sentence is represented by the first 30 words from arg1 and 5 words from connective, then 30 words from arg2. This is from the information of the training set: the average length of arg1, arg2 and connective is 16.6 15.6 and 1.1. More importantly it aligns all the sentences and shows that word31 to word35 are connective in each sentence. (Another way to do it is to move connective to the beginning of the sentence to make sure that it appears in the representation in case that the truncation is too short). When the connective are made to appear for sure, the performance improves a lot. Another experiment is to select the last 30 words from arg1 instead of the first 30 words since intuitively the last 30 words should play more role in discourse relation, however the results shown are similar. (The experimental results can be found at the end of the report, the last two are the best performance)

3.2 neural network architecture

Two architectures have been tried, one is a network with a single hidden layer with 50, 150 or 300 nodes; another one is with two hidden layers with nodes 256(first layer) -128(second layer), or 512(first layer) - 256(second layer). According to the experimental results, for the pretrained embeddings, two layers worded better than one layer and nodes 256 - 128 worked better than 512 - 256 due to the overfitting problem. For the self-trained embedding, the performance of one layer model and two layers model are similar, and 300 nodes worked better than 50 and 150 nodes.

3.3 regularizers

The L2 regularizers make the performance improve a lot by punishing the parameters to get rid of the overfitting problem.

3.4 experimental results

one layer network with 50 nodes and pretrained embedding without regularizers with sentence length 30:

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.2531

Temporal.Asynchronous.Succession    precision 0.33333      recall 0.015385     F1 0.029412
EntRel                              precision 0.11765      recall 0.03         F1 0.047809
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.27273      recall 0.054545     F1 0.090909
Expansion.Conjunction               precision 0.27495      recall 0.88218      F1 0.41924
Expansion.Restatement               precision 0.0          recall 0.0          F1 -
Contingency.Cause.Result            precision 0.0          recall 0.0          F1 -
Contingency.Cause.Reason            precision 0.16667      recall 0.01087      F1 0.020408
Comparison.Contrast                 precision 0.048387     recall 0.051724     F1 0.05
Expansion.Instantiation             precision 0.0          recall 0.0          F1 -
Contingency.Condition               precision 0.0          recall 0.0          F1 -
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.2531       recall 0.24113      F1 0.24697

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.34892

Temporal.Asynchronous.Succession    precision 0.33333      recall 0.015385     F1 0.029412
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.33333      recall 0.057692     F1 0.098361
Expansion.Conjunction               precision 0.39451      recall 0.86175      F1 0.54124
Contingency.Cause.Result            precision 0.0          recall 0.0          F1 -
Contingency.Cause.Reason            precision 0.33333      recall 0.02         F1 0.037736
Comparison.Contrast                 precision 0.064516     recall 0.068966     F1 0.066667
Expansion.Restatement               precision 0.0          recall 0.0          F1 -
Contingency.Condition               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -
Expansion.Instantiation             precision 0.0          recall 0.0          F1 -
Expansion.Alternative               precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.36742      recall 0.31908      F1 0.34155

============================================================
Evaluation for non-explicit discourse relations only

Accuracy: 0.17152

EntRel                              precision 0.26087      recall 0.03         F1 0.053812
Expansion.Restatement               precision 0.0          recall 0.0          F1 -
Contingency.Cause.Reason            precision 0.0          recall 0.0          F1 -
Expansion.Conjunction               precision 0.17857      recall 0.92105      F1 0.29915
Contingency.Cause.Result            precision 0.0          recall 0.0          F1 -
```

```
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Comparison.Contrast                 precision 0.032258     recall 0.034483     F1 0.033333
Expansion.Instantiation             precision 0.0          recall 0.0          F1 -
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.17152      recall 0.16944      F1 0.17047
```

two layers network with 256 - 128 nodes and pretrained embedding with L2 regularizers with sentence length 30:

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.31348

Temporal.Asynchronous.Succession    precision 0.5          recall 0.2          F1 0.28571
EntRel                              precision 0.32472      recall 0.44         F1 0.37367
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.14286      recall 0.018182     F1 0.032258
Expansion.Conjunction               precision 0.35839      recall 0.61934      F1 0.45404
Expansion.Restatement               precision 0.32143      recall 0.11842      F1 0.17308
Contingency.Cause.Result            precision 0.090909     recall 0.029412     F1 0.044444
Contingency.Cause.Reason            precision 0.23611      recall 0.18478      F1 0.20732
Comparison.Contrast                 precision 0.061224     recall 0.10345      F1 0.076923
Expansion.Instantiation             precision 0.45455      recall 0.11364      F1 0.18182
Contingency.Condition               precision 0.42308      recall 0.40741      F1 0.41509
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.27083      recall 0.26         F1 0.26531

Micro-Average                       precision 0.31348      recall 0.29866      F1 0.30589

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.36331

Temporal.Asynchronous.Succession    precision 0.8125       recall 0.2          F1 0.32099
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.16667      recall 0.019231     F1 0.034483
Expansion.Conjunction               precision 0.51079      recall 0.65438      F1 0.57374
Contingency.Cause.Result            precision 0.0          recall 0.0          F1 -
Contingency.Cause.Reason            precision 0.30435      recall 0.28         F1 0.29167
Comparison.Contrast                 precision 0.071429     recall 0.13793      F1 0.094118
Expansion.Restatement               precision 0.11765      recall 0.33333      F1 0.17391
Contingency.Condition               precision 0.5          recall 0.40741      F1 0.44898
Temporal.Asynchronous.Precedence    precision 0.39394      recall 0.325        F1 0.35616
Expansion.Instantiation             precision 1.0          recall 0.66667      F1 0.8
Expansion.Alternative               precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.41822      recall 0.33224      F1 0.3703

============================================================
```

```
Evaluation for non-explicit discourse relations only

Accuracy: 0.27106

EntRel                              precision 0.44444     recall 0.44         F1 0.44221
Expansion.Restatement               precision 0.41026     recall 0.10959      F1 0.17297
Contingency.Cause.Reason            precision 0.11538     recall 0.071429     F1 0.088235
Expansion.Conjunction               precision 0.21429     recall 0.55263      F1 0.30882
Contingency.Cause.Result            precision 0.13333     recall 0.05         F1 0.072727
Comparison.Concession               precision 0.0         recall 0.0          F1 -
Comparison.Contrast                 precision 0.047619    recall 0.068966     F1 0.056338
Expansion.Instantiation             precision 0.33333     recall 0.073171     F1 0.12
Expansion.Alternative               precision 0.0         recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0         recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0         recall 0.0          F1 -

Micro-Average                       precision 0.277       recall 0.26778      F1 0.27231
```

two layers network with 256 - 128 nodes and pretrained embedding with L2 regularizers with 30words_arg1 5words_connective 30words_arg2:

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.47725

Temporal.Asynchronous.Succession    precision 0.97917     recall 0.72308      F1 0.83186
EntRel                              precision 0.33179     recall 0.715        F1 0.45325
Comparison.Concession               precision 0.84615     recall 0.098214     F1 0.176
Temporal.Synchrony                  precision 0.58491     recall 0.56364      F1 0.57407
Expansion.Conjunction               precision 0.49559     recall 0.67976      F1 0.57325
Expansion.Restatement               precision 0.6         recall 0.039474     F1 0.074074
Contingency.Cause.Result            precision 0.71429     recall 0.14706      F1 0.2439
Contingency.Cause.Reason            precision 0.68293     recall 0.30435      F1 0.42105
Comparison.Contrast                 precision 0.18571     recall 0.22414      F1 0.20312
Expansion.Instantiation             precision 0.75        recall 0.068182     F1 0.125
Contingency.Condition               precision 0.88889     recall 0.88889      F1 0.88889
Expansion.Alternative               precision 0.0         recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.85714     recall 0.72         F1 0.78261

Micro-Average                       precision 0.47804     recall 0.45469      F1 0.46607

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.69065

Temporal.Asynchronous.Succession    precision 0.97917     recall 0.72308      F1 0.83186
Comparison.Concession               precision 0.84615     recall 0.1358       F1 0.23404
Temporal.Synchrony                  precision 0.59615     recall 0.59615      F1 0.59615
Expansion.Conjunction               precision 0.86957     recall 0.82949      F1 0.84906
Contingency.Cause.Result            precision 0.75        recall 0.32143      F1 0.45
Contingency.Cause.Reason            precision 0.68293     recall 0.56         F1 0.61538
Comparison.Contrast                 precision 0.19118     recall 0.44828      F1 0.26804
```

```
Expansion.Restatement                precision 0.5          recall 0.5          F1 0.5
Contingency.Condition                precision 0.88889      recall 0.88889      F1 0.88889
Temporal.Asynchronous.Precedence     precision 0.85714      recall 0.9          F1 0.87805
Expansion.Instantiation              precision 1.0          recall 0.66667      F1 0.8
Expansion.Alternative                precision 0.0          recall 0.0          F1 -

Micro-Average                        precision 0.74131      recall 0.63158      F1 0.68206

============================================================
Evaluation for non-explicit discourse relations only

Accuracy: 0.29556

EntRel                               precision 0.36203      recall 0.715        F1 0.48067
Expansion.Restatement                precision 0.75         recall 0.020548     F1 0.04
Contingency.Cause.Reason             precision 0.0          recall 0.0          F1 -
Expansion.Conjunction                precision 0.18219      recall 0.39474      F1 0.24931
Contingency.Cause.Result             precision 0.5          recall 0.025        F1 0.047619
Comparison.Concession                precision 0.0          recall 0.0          F1 -
Comparison.Contrast                  precision 0.0          recall 0.0          F1 -
Expansion.Instantiation              precision 0.5          recall 0.02439      F1 0.046512
Expansion.Alternative                precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence     precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                   precision 0.0          recall 0.0          F1 -

Micro-Average                        precision 0.29556      recall 0.29198      F1 0.29376
```

two layers network with 256 - 128 nodes and pretrained embedding with L2 regularizers with 5words_connective last 30words_arg1 30words_arg2:

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.47808

Temporal.Asynchronous.Succession     precision 1.0          recall 0.73846      F1 0.84956
EntRel                               precision 0.32697      recall 0.685        F1 0.44265
Comparison.Concession                precision 0.86667      recall 0.11607      F1 0.20472
Temporal.Synchrony                   precision 0.58621      recall 0.61818      F1 0.60177
Expansion.Conjunction                precision 0.49241      recall 0.6858       F1 0.57323
Expansion.Restatement                precision 0.4          recall 0.039474     F1 0.071856
Contingency.Cause.Result             precision 0.71429      recall 0.14706      F1 0.2439
Contingency.Cause.Reason             precision 0.77778      recall 0.30435      F1 0.4375
Comparison.Contrast                  precision 0.17391      recall 0.2069       F1 0.18898
Expansion.Instantiation              precision 0.75         recall 0.068182     F1 0.125
Contingency.Condition                precision 0.88889      recall 0.88889      F1 0.88889
Expansion.Alternative                precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence     precision 0.87805      recall 0.72         F1 0.79121

Micro-Average                        precision 0.47887      recall 0.45548      F1 0.46688

============================================================
Evaluation for explicit discourse relations only
```

```
Accuracy: 0.70144

Temporal.Asynchronous.Succession    precision 1.0         recall 0.73846      F1 0.84956
Comparison.Concession               precision 0.86667     recall 0.16049      F1 0.27083
Temporal.Synchrony                  precision 0.58621     recall 0.65385      F1 0.61818
Expansion.Conjunction               precision 0.87864     recall 0.8341       F1 0.85579
Contingency.Cause.Result            precision 0.75        recall 0.32143      F1 0.45
Contingency.Cause.Reason            precision 0.8         recall 0.56         F1 0.65882
Comparison.Contrast                 precision 0.17647     recall 0.41379      F1 0.24742
Expansion.Restatement               precision 0.42857     recall 0.5          F1 0.46154
Contingency.Condition               precision 0.88889     recall 0.88889      F1 0.88889
Temporal.Asynchronous.Precedence    precision 0.87805     recall 0.9          F1 0.88889
Expansion.Instantiation             precision 1.0         recall 0.66667      F1 0.8
Expansion.Alternative               precision 0.0         recall 0.0          F1 -

Micro-Average                       precision 0.75145     recall 0.64145      F1 0.6921

============================================================
Evaluation for non-explicit discourse relations only

Accuracy: 0.2879

EntRel                              precision 0.35677     recall 0.685        F1 0.46918
Expansion.Restatement               precision 0.375       recall 0.020548     F1 0.038961
Contingency.Cause.Reason            precision 0.0         recall 0.0          F1 -
Expansion.Conjunction               precision 0.18039     recall 0.40351      F1 0.24932
Contingency.Cause.Result            precision 0.5         recall 0.025        F1 0.047619
Comparison.Concession               precision 0.0         recall 0.0          F1 -
Comparison.Contrast                 precision 0.0         recall 0.0          F1 -
Expansion.Instantiation             precision 0.5         recall 0.02439      F1 0.046512
Expansion.Alternative               precision 0.0         recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0         recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0         recall 0.0          F1 -

Micro-Average                       precision 0.2879      recall 0.28442      F1 0.28615
```

two layers network with 256 - 128 nodes and self trained embedding with L2 regularizers with 30words_arg1 5words_connective 30words_arg2:

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.52605

Temporal.Asynchronous.Succession    precision 0.89286     recall 0.76923      F1 0.82645
EntRel                              precision 0.31898     recall 1.0          F1 0.48368
Comparison.Concession               precision 1.0         recall 0.071429     F1 0.13333
Temporal.Synchrony                  precision 0.52727     recall 0.52727      F1 0.52727
Expansion.Conjunction               precision 0.86905     recall 0.66163      F1 0.75129
Expansion.Restatement               precision 0.57143     recall 0.026316     F1 0.050314
Contingency.Cause.Result            precision 0.92857     recall 0.19118      F1 0.31707
Contingency.Cause.Reason            precision 0.80645     recall 0.27174      F1 0.4065
Comparison.Contrast                 precision 0.21333     recall 0.27586      F1 0.2406
Expansion.Instantiation             precision 0.83333     recall 0.11364      F1 0.2
```

```
Contingency.Condition               precision 0.88889      recall 0.88889      F1 0.88889
Expansion.Alternative               precision 1.0          recall 0.33333      F1 0.5
Temporal.Asynchronous.Precedence    precision 0.86364      recall 0.76         F1 0.80851

Micro-Average                       precision 0.52693      recall 0.50118      F1 0.51373

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.75899

Temporal.Asynchronous.Succession    precision 0.89286      recall 0.76923      F1 0.82645
Comparison.Concession               precision 1.0          recall 0.098765     F1 0.17978
Temporal.Synchrony                  precision 0.52727      recall 0.55769      F1 0.54206
Expansion.Conjunction               precision 0.90213      recall 0.97696      F1 0.93805
Contingency.Cause.Result            precision 0.90909      recall 0.35714      F1 0.51282
Contingency.Cause.Reason            precision 0.80645      recall 0.5          F1 0.61728
Comparison.Contrast                 precision 0.21333      recall 0.55172      F1 0.30769
Expansion.Restatement               precision 0.6          recall 0.5          F1 0.54545
Contingency.Condition               precision 0.88889      recall 0.88889      F1 0.88889
Temporal.Asynchronous.Precedence    precision 0.88372      recall 0.95         F1 0.91566
Expansion.Instantiation             precision 1.0          recall 0.66667      F1 0.8
Expansion.Alternative               precision 1.0          recall 0.5          F1 0.66667

Micro-Average                       precision 0.76311      recall 0.69408      F1 0.72696

============================================================
Evaluation for non-explicit discourse relations only

Accuracy: 0.32772

EntRel                              precision 0.31949      recall 1.0          F1 0.48426
Expansion.Restatement               precision 0.5          recall 0.0068493    F1 0.013514
Contingency.Cause.Reason            precision 0.0          recall 0.0          F1 -
Expansion.Conjunction               precision 0.41176      recall 0.061404     F1 0.10687
Contingency.Cause.Result            precision 1.0          recall 0.075        F1 0.13953
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Comparison.Contrast                 precision 0.0          recall 0.0          F1 -
Expansion.Instantiation             precision 0.75         recall 0.073171     F1 0.13333
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.32772      recall 0.32375      F1 0.32572
```

one layer network with 50 nodes and self trained embedding with L2 regularizers with 5words_connective 30words_arg1 30words_arg2:

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.5244

Temporal.Asynchronous.Succession    precision 0.97917      recall 0.72308      F1 0.83186
```

```
EntRel                              precision 0.31496      recall 1.0          F1 0.47904
Comparison.Concession               precision 0.73333      recall 0.098214     F1 0.17323
Temporal.Synchrony                  precision 0.5625       recall 0.49091      F1 0.52427
Expansion.Conjunction               precision 0.90213      recall 0.64048      F1 0.74912
Expansion.Restatement               precision 0.66667      recall 0.039474     F1 0.074534
Contingency.Cause.Result            precision 0.83333      recall 0.22059      F1 0.34884
Contingency.Cause.Reason            precision 0.58333      recall 0.30435      F1 0.4
Comparison.Contrast                 precision 0.2027       recall 0.25862      F1 0.22727
Expansion.Instantiation             precision 1.0          recall 0.090909     F1 0.16667
Contingency.Condition               precision 0.96296      recall 0.96296      F1 0.96296
Expansion.Alternative               precision 1.0          recall 0.33333      F1 0.5
Temporal.Asynchronous.Precedence    precision 0.95         recall 0.76         F1 0.84444

Micro-Average                       precision 0.5257       recall 0.49961      F1 0.51232

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.76259

Temporal.Asynchronous.Succession    precision 0.97917      recall 0.72308      F1 0.83186
Comparison.Concession               precision 0.78571      recall 0.1358       F1 0.23158
Temporal.Synchrony                  precision 0.5625       recall 0.51923      F1 0.54
Expansion.Conjunction               precision 0.90909      recall 0.96774      F1 0.9375
Contingency.Cause.Result            precision 0.91667      recall 0.39286      F1 0.55
Contingency.Cause.Reason            precision 0.59574      recall 0.56         F1 0.57732
Comparison.Contrast                 precision 0.2027       recall 0.51724      F1 0.29126
Expansion.Restatement               precision 0.8          recall 0.66667      F1 0.72727
Contingency.Condition               precision 0.96296      recall 0.96296      F1 0.96296
Temporal.Asynchronous.Precedence    precision 0.97436      recall 0.95         F1 0.96203
Expansion.Instantiation             precision 1.0          recall 0.66667      F1 0.8
Expansion.Alternative               precision 1.0          recall 0.5          F1 0.66667

Micro-Average                       precision 0.76812      recall 0.69737      F1 0.73103

============================================================
Evaluation for non-explicit discourse relations only

Accuracy: 0.32159

EntRel                              precision 0.31546      recall 1.0          F1 0.47962
Expansion.Restatement               precision 0.5          recall 0.013699     F1 0.026667
Contingency.Cause.Reason            precision 0.0          recall 0.0          F1 -
Expansion.Conjunction               precision 0.5          recall 0.017544     F1 0.033898
Contingency.Cause.Result            precision 0.66667      recall 0.1          F1 0.17391
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Comparison.Contrast                 precision 0.0          recall 0.0          F1 -
Expansion.Instantiation             precision 1.0          recall 0.04878      F1 0.093023
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.32159      recall 0.3177       F1 0.31963
```

one layer network with 150 nodes and self trained embedding with L2 regularizers with 5words_connective

30words_arg1 30words_arg2:

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.52936

Temporal.Asynchronous.Succession    precision 1.0         recall 0.70769      F1 0.82883
EntRel                              precision 0.31348     recall 1.0          F1 0.47733
Comparison.Concession               precision 0.9375      recall 0.13393      F1 0.23437
Temporal.Synchrony                  precision 0.55172     recall 0.58182      F1 0.56637
Expansion.Conjunction               precision 0.89873     recall 0.6435       F1 0.75
Expansion.Restatement               precision 0.63636     recall 0.046053     F1 0.08589
Contingency.Cause.Result            precision 0.8125      recall 0.19118      F1 0.30952
Contingency.Cause.Reason            precision 0.66667     recall 0.26087      F1 0.375
Comparison.Contrast                 precision 0.24324     recall 0.31034      F1 0.27273
Expansion.Instantiation             precision 1.0         recall 0.068182     F1 0.12766
Contingency.Condition               precision 0.92857     recall 0.96296      F1 0.94545
Expansion.Alternative               precision 1.0         recall 0.33333      F1 0.5
Temporal.Asynchronous.Precedence    precision 0.97436     recall 0.76         F1 0.85393

Micro-Average                       precision 0.53024     recall 0.50433      F1 0.51696

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.77878

Temporal.Asynchronous.Succession    precision 1.0         recall 0.70769      F1 0.82883
Comparison.Concession               precision 0.9375      recall 0.18519      F1 0.30928
Temporal.Synchrony                  precision 0.5614      recall 0.61538      F1 0.58716
Expansion.Conjunction               precision 0.90987     recall 0.97696      F1 0.94222
Contingency.Cause.Result            precision 0.91667     recall 0.39286      F1 0.55
Contingency.Cause.Reason            precision 0.66667     recall 0.48         F1 0.55814
Comparison.Contrast                 precision 0.24658     recall 0.62069      F1 0.35294
Expansion.Restatement               precision 0.8         recall 0.66667      F1 0.72727
Contingency.Condition               precision 0.92857     recall 0.96296      F1 0.94545
Temporal.Asynchronous.Precedence    precision 0.97436     recall 0.95         F1 0.96203
Expansion.Instantiation             precision 1.0         recall 0.66667      F1 0.8
Expansion.Alternative               precision 1.0         recall 0.5          F1 0.66667

Micro-Average                       precision 0.78442     recall 0.71217      F1 0.74655

============================================================
Evaluation for non-explicit discourse relations only

Accuracy: 0.317

EntRel                              precision 0.31447     recall 1.0          F1 0.47847
Expansion.Restatement               precision 0.5         recall 0.020548     F1 0.039474
Contingency.Cause.Reason            precision 0.0         recall 0.0          F1 -
Expansion.Conjunction               precision 0.25        recall 0.0087719    F1 0.016949
Contingency.Cause.Result            precision 0.5         recall 0.05         F1 0.090909
Comparison.Concession               precision 0.0         recall 0.0          F1 -
Comparison.Contrast                 precision 0.0         recall 0.0          F1 -
```

```
Expansion.Instantiation             precision 1.0          recall 0.02439      F1 0.047619
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.317        recall 0.31316      F1 0.31507
```

one layer network with 300 nodes and self trained embedding with L2 regularizers with 5words_connective 30words_arg1 30words_arg2: (one of the best performance)

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.53184

Temporal.Asynchronous.Succession    precision 1.0          recall 0.70769      F1 0.82883
EntRel                              precision 0.31397      recall 1.0          F1 0.4779
Comparison.Concession               precision 0.92308      recall 0.10714      F1 0.192
Temporal.Synchrony                  precision 0.6          recall 0.65455      F1 0.62609
Expansion.Conjunction               precision 0.89873      recall 0.6435       F1 0.75
Expansion.Restatement               precision 0.6          recall 0.039474     F1 0.074074
Contingency.Cause.Result            precision 0.875        recall 0.20588      F1 0.33333
Contingency.Cause.Reason            precision 0.65789      recall 0.27174      F1 0.38462
Comparison.Contrast                 precision 0.24658      recall 0.31034      F1 0.27481
Expansion.Instantiation             precision 1.0          recall 0.090909     F1 0.16667
Contingency.Condition               precision 0.92857      recall 0.96296      F1 0.94545
Expansion.Alternative               precision 1.0          recall 0.33333      F1 0.5
Temporal.Asynchronous.Precedence    precision 0.97436      recall 0.76         F1 0.85393

Micro-Average                       precision 0.53317      recall 0.5067       F1 0.5196

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.78058

Temporal.Asynchronous.Succession    precision 1.0          recall 0.70769      F1 0.82883
Comparison.Concession               precision 0.92308      recall 0.14815      F1 0.25532
Temporal.Synchrony                  precision 0.6          recall 0.69231      F1 0.64286
Expansion.Conjunction               precision 0.90987      recall 0.97696      F1 0.94222
Contingency.Cause.Result            precision 0.91667      recall 0.39286      F1 0.55
Contingency.Cause.Reason            precision 0.67568      recall 0.5          F1 0.57471
Comparison.Contrast                 precision 0.25         recall 0.62069      F1 0.35644
Expansion.Restatement               precision 0.75         recall 0.5          F1 0.6
Contingency.Condition               precision 0.92857      recall 0.96296      F1 0.94545
Temporal.Asynchronous.Precedence    precision 0.97436      recall 0.95         F1 0.96203
Expansion.Instantiation             precision 1.0          recall 0.66667      F1 0.8
Expansion.Alternative               precision 1.0          recall 0.5          F1 0.66667

Micro-Average                       precision 0.78766      recall 0.71382      F1 0.74892

============================================================
Evaluation for non-explicit discourse relations only
```

```
Accuracy: 0.32006

EntRel                              precision 0.31496      recall 1.0          F1 0.47904
Expansion.Restatement               precision 0.5          recall 0.020548     F1 0.039474
Contingency.Cause.Reason            precision 0.0          recall 0.0          F1 -
Expansion.Conjunction               precision 0.25         recall 0.0087719    F1 0.016949
Contingency.Cause.Result            precision 0.75         recall 0.075        F1 0.13636
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Comparison.Contrast                 precision 0.0          recall 0.0          F1 -
Expansion.Instantiation             precision 1.0          recall 0.04878      F1 0.093023
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.32006      recall 0.31619      F1 0.31811
```

one layer network with 300 nodes and self trained embedding with L2 regularizers with 5words_connective last 30words_arg1 30words_arg2: (one of the best performance)

```
============================================================
Evaluation for all discourse relations

Accuracy: 0.53267

Temporal.Asynchronous.Succession    precision 1.0          recall 0.72308      F1 0.83929
EntRel                              precision 0.31299      recall 1.0          F1 0.47676
Comparison.Concession               precision 0.82143      recall 0.20536      F1 0.32857
Temporal.Synchrony                  precision 0.57143      recall 0.50909      F1 0.53846
Expansion.Conjunction               precision 0.90987      recall 0.64048      F1 0.75177
Expansion.Restatement               precision 0.54545      recall 0.039474     F1 0.07362
Contingency.Cause.Result            precision 0.77778      recall 0.20588      F1 0.32558
Contingency.Cause.Reason            precision 0.58696      recall 0.29348      F1 0.3913
Comparison.Contrast                 precision 0.23729      recall 0.24138      F1 0.23932
Expansion.Instantiation             precision 1.0          recall 0.090909     F1 0.16667
Contingency.Condition               precision 0.92857      recall 0.96296      F1 0.94545
Expansion.Alternative               precision 1.0          recall 0.33333      F1 0.5
Temporal.Asynchronous.Precedence    precision 0.97436      recall 0.76         F1 0.85393

Micro-Average                       precision 0.534        recall 0.50749      F1 0.5204

============================================================
Evaluation for explicit discourse relations only

Accuracy: 0.78597

Temporal.Asynchronous.Succession    precision 1.0          recall 0.72308      F1 0.83929
Comparison.Concession               precision 0.82143      recall 0.28395      F1 0.42202
Temporal.Synchrony                  precision 0.58333      recall 0.53846      F1 0.56
Expansion.Conjunction               precision 0.90987      recall 0.97696      F1 0.94222
Contingency.Cause.Result            precision 0.91667      recall 0.39286      F1 0.55
Contingency.Cause.Reason            precision 0.6          recall 0.54         F1 0.56842
Comparison.Contrast                 precision 0.23729      recall 0.48276      F1 0.31818
Expansion.Restatement               precision 0.8          recall 0.66667      F1 0.72727
Contingency.Condition               precision 0.92857      recall 0.96296      F1 0.94545
```

```
Temporal.Asynchronous.Precedence    precision 0.97436      recall 0.95         F1 0.96203
Expansion.Instantiation             precision 1.0          recall 0.66667      F1 0.8
Expansion.Alternative               precision 1.0          recall 0.5          F1 0.66667

Micro-Average                       precision 0.7931       recall 0.71875      F1 0.7541

============================================================
Evaluation for non-explicit discourse relations only

Accuracy: 0.317

EntRel                              precision 0.31397      recall 1.0          F1 0.4779
Expansion.Restatement               precision 0.33333      recall 0.013699     F1 0.026316
Contingency.Cause.Reason            precision 0.0          recall 0.0          F1 -
Expansion.Conjunction               precision 0.0          recall 0.0          F1 -
Contingency.Cause.Result            precision 0.5          recall 0.075        F1 0.13043
Comparison.Concession               precision 0.0          recall 0.0          F1 -
Comparison.Contrast                 precision 0.0          recall 0.0          F1 -
Expansion.Instantiation             precision 1.0          recall 0.04878      F1 0.093023
Expansion.Alternative               precision 0.0          recall 0.0          F1 -
Temporal.Asynchronous.Precedence    precision 0.0          recall 0.0          F1 -
Temporal.Synchrony                  precision 0.0          recall 0.0          F1 -

Micro-Average                       precision 0.317        recall 0.31316      F1 0.31507
```