Assignment-Extracting-Structure-solved

June 3, 2019

1 This tutorial is Largely based on the paper Context-aware Argumentative Relation Mining

Huy V. Nguyen Diane J. Litman

However the implementation is not the exact procedure stated in the paper, rather covers the overall intention. The main intention being somehow give the contextual information to the classifier model by extracting topics using LDA.

The implementation is divided into 6 parts :

```
Part 1 : Importing and structuring the Dataset
Part 2 : Window Context Extraction ( TODO )
Part 3 : LDA topic Extraction ( TODO )
Part 4 : Creating and Adding the features ( TODO )
Part 5 : Applying Classification Models ( TODO )
Part 6 : Hyperparameter tuning ( additional )
```

2 Part 1.

2.1 Importing and structuring the Dataset of 90 Persuasive ESSAYs

Importing the actual essays

```
In [104]: essay_dict['essay34']
```

```
Out[104]: ['Study at school or get a job?',
           'Many people believe that children should study at school to have more knowledge that
           'Others, however, think that these children may disrupt their school work and should
           'Personally, I tend to agree with the point of view that student have to be forced to
           'First of all, schools offer to students a good environment with experienced professo
           'It creates the best conditions for students education and can force them to focus on
           'Second of all, schools provide lots of academic knowledge to students.',
           'Students may learn professional skills, expand their understandings and gain experie
           'Therefore, they have more opprotunities to find a job and to be successful in the fu
           'For example, as we know, employer always prefer to hire an employee of high degree w
           'Nevertheless, it is not unreasonable that some people think that children should int
           'Whether children can learn a lot at school, there are many subjects that will be of
           'Furthermore, children can learn social skills when they have a job.',
```

```
 'They can get more experiences that can not be obtained at school.',
 "Working helps children be more independent and teach them to esteem and manage the m
 'Overall, I believe that students should study at school.',
 "Even though there are some advantages of leaving school to find a job, studying at s
 'There are many ways that can train children to learn independent and social skills i
```

Importing the Annotations

In [106]:
```
dataset.head()
```

Out[106]:

	src_id	src	src_type	src_strt	src_end	tgt_id	tgt	tgt_type	tgt_strt	tgt_end	relation	essay
0	T1	competition can effectively promote the develo...	Claim	78	140	T2	we should attach more importance to cooperation	MajorClaim	503	550	attacks	essay01
1	T1	competition can effectively promote the develo...	Claim	78	140	T3	In order to survive in the competition, compan...	Premise	142	283	no relation	essay01
2	T1	competition can effectively promote the develo...	Claim	78	140	T4	through cooperation, children can learn about ...	Claim	591	714	no relation	essay01
3	T1	competition can effectively promote the develo...	Claim	78	140	T5	What we acquired from team work is not only ho...	Premise	716	851	no relation	essay01
4	T1	competition can effectively promote the develo...	Claim	78	140	T6	During the process of cooperation, children ca...	Premise	853	1086	no relation	essay01

In [12]:
```
dataset[dataset['relation'] != 'no relation']['relation'].value_counts()
```

Out[12]:
```
supports    1312
attacks      161
Name: relation, dtype: int64
```

3 Part 3.

3.1 LDA Topic Extraction

Loading the extra Essay corpus

In [107]:
```
data[:5]
```

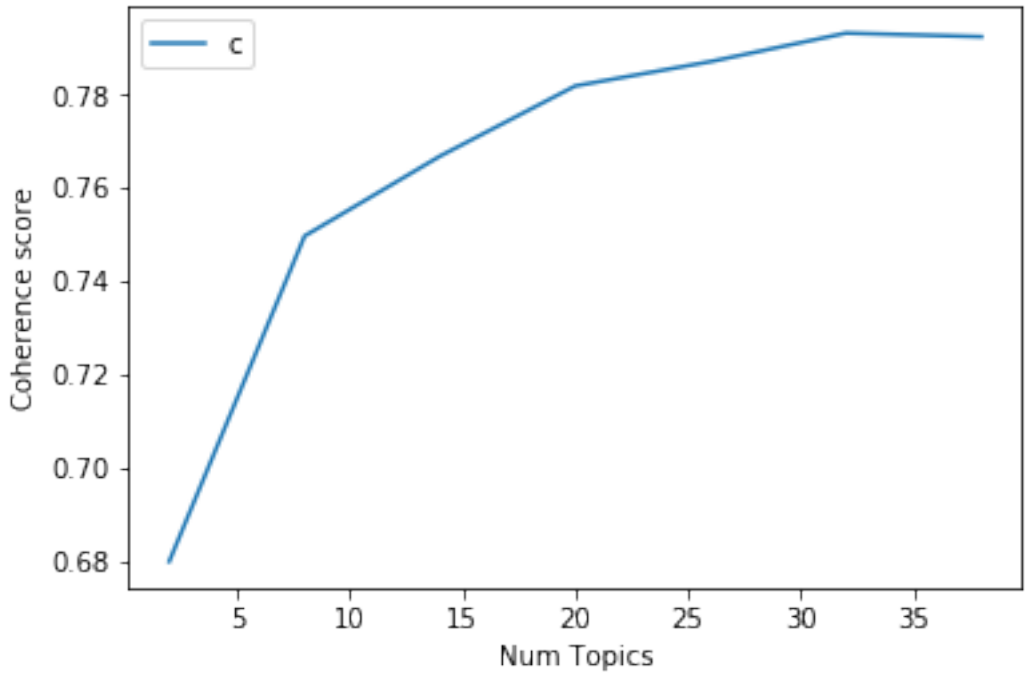
Out[107]:
```
['Should students be taught to compete or to cooperate? ',
 'More people are migrating to other countries than ever before ',
 'International tourism is now more common than ever before ',
 'International tourism is now more common than ever before ',
 'Living and studying overseas ']
```

```
In [23]:print(preprocess_pipeline(data)[:10])
```

```
[['student', 'teach', 'compete', 'cooperate'],
 ['people', 'migrate', 'country', 'ever'],
 ['international_tourism', 'common', 'ever'],
 ['international_tourism', 'common', 'ever'],
 ['live', 'study', 'overseas'],
 ['exercise'],
 ['newspaper', 'become', 'thing', 'past'],
 ['technology', 'cannot', 'solve', 'world', 'problem'],
 ['truth', 'cannabis'],
 ['single', 'international', 'language']]
```

```
In [26]: # Compute Coherence Score
         coherence_model_lda = CoherenceModel(model=lda_model, texts=texts, dictionary=id2word,
         coherence_lda = coherence_model_lda.get_coherence()
         print('\nCoherence Score: ', coherence_lda)
```

```
Coherence Score:  0.7923088842456681
```



4 Part 4.

4.1 Creating and adding Features

```
In [102]: topic_words[13][:5]
```

```
Out[102]: ['influence', 'artist', 'improve', 'rich', 'university']
```

```
In [95]: X.head()
```

```
Out[95]:    src_type_Claim  src_type_Premise  tgt_type_Claim  tgt_type_MajorClaim  \
         0               1                 0               0                    1
         1               0                 1               1                    0
         2               1                 0               0                    1
         3               0                 1               1                    0
         4               0                 1               1                    0

            tgt_type_Premise  abs_diff_strt  abs_diff_end  word_count_src  \
         0                 0            425           410              62
         1                 0             64           143             141
         2                 0             88           164             123
         3                 0            125           137             135
         4                 0            262           372             233

            word_count_tgt  src_next_sent1_comm      ...       topic_31_src  \
         0              47                    2      ...                  0
         1              62                    4      ...                  0
         2              47                    1      ...                  0
         3             123                    6      ...                  1
         4             123                    6      ...                  1

            topic_31_tgt  topic_32_src  topic_32_tgt  topic_33_src  topic_33_tgt  \
         0             0             0             0             1             0
         1             0             0             0             0             1
         2             0             0             0             0             0
         3             0             0             0             0             0
         4             0             2             0             1             0

            topic_34_src  topic_34_tgt  topic_35_src  topic_35_tgt
         0             0             0             0             0
         1             0             0             0             0
         2             0             0             0             0
         3             0             0             0             0
         4             1             0             1             0

         [5 rows x 97 columns]
```

5 Part 5.

5.1 Applying Classification models

Split X and Y into training and testing datasets using train_test_split.

```
In [88]:print(X_train.shape, Y_train.shape, X_test.shape, Y_test.shape)
```

```
(1178, 97) (1178,) (295, 97) (295,)
```

Calculating the precision macro, recall macro, f1 macro and accuracy of the model

```
In [90]: p_macro, r_macro, f_macro, support_macro = precision_recall_fscore_support(y_true=Y_tes
                                                                                    y_pred=Y_pre
                                                                                    labels=[0,1]
                                                                                    average='mac


         print('Accuracy:',round(accuracy_score(Y_test, Y_pred),2),
               '\nKappa:',round(cohen_kappa_score(Y_test,Y_pred),2),
               '\nMacro Precision:',round(p_macro,2),
               '\nMacro Recall:', round(r_macro,2),
               '\nMacro F1:',round(f_macro,2,),
               '\nF1:',round(f1_score(Y_test, Y_pred),2,)
               )
```

```
Accuracy: 0.87
Kappa: 0.24
Macro Precision: 0.68
Macro Recall: 0.59
Macro F1: 0.61
F1: 0.93
```