

...Report...

1. Learning of Polarity

Possible classes (for polarity) in readMe of MPQA3:

Possible values: positive, negative, both, neutral,
uncertain-positive, uncertain-negative, uncertain-both,
uncertain-neutral

Total number of objects of CSDS:

```
print(len(data['csds_objects']))  
  
4384
```

1. Percentage of the accuracy of pre-trained *bert-base-cased* model with 8-class training on dataset: more than 72%!

- Number of Epochs = 3
- Number of objects of CSDS that have polarity:

Number of data items can be used is: 4128

- How to select train set and validation set: 80 to 20 ratio with respect to the balance of classes (from each class to the mentioned ratio is selected)
- Number of data related to each class (train set):

```
uncertain-neutral : 11  
uncertain-negative : 24  
uncertain-both : 4  
negative : 1492  
uncertain-positive : 8  
both : 4  
neutral : 616  
positive : 1140
```

2. Percentage of the accuracy of pre-trained *bert-base-cased* model with 4-class training on dataset (aggregation of uncertainties with related certain, for example, aggregation of uncertain-pos with pos): more than 73%!

- Number of Epochs = 3
- Number of objects of CSDS that have polarity (with respect to aggregation of uncertainties):

Number of data items can be used is: 4128

- How to select train set and validation set: 80 to 20 ratio with respect to the balance of classes (from each class to the mentioned ratio is selected)
- Number of data related to each class (train set):

```
positive : 1148
both : 9
negative : 1517
neutral : 627
```

3. Percentage of the accuracy of pre-trained *bert-base-cased* model with 4-class training on dataset (ignoring uncertainties): more than 73%!

- Number of Epochs = 3
- Number of objects of CSDS that have polarity (with respect to ignore uncertainties):

Number of data items can be used is: 4067

- How to select train set and validation set: 80 to 20 ratio with respect to the balance of classes (from each class to the mentioned ratio is selected)
- Number of data related to each class (train set):

```
positive : 1140
both : 4
negative : 1492
neutral : 616
```

4. Percentage of the accuracy of pre-trained *bert-base-cased* model with 3-class training on dataset (ignoring uncertainties and both): more than 86%!

- Number of Epochs = 3
- Number of objects of CSDS that have polarity (with respect to ignore uncertainties and both):

Number of data items can be used is: 4061

- How to select train set and validation set: stratified 5-fold cross validation
- Number of data related to each class (dataset):

```
positive : 1425
negative : 1866
neutral  : 770
```

- accuracy, precision, recall, f1-score:

```
average_accuracy: 0.8604120844164106
average_precision: [0.83247633 0.86814572 0.89286082]
average_recall:    [0.82385965 0.88374934 0.87142857]
average_f1:        [0.82784705 0.87562486 0.88176023]
```

5.

2.

.....