

Evaluation scripts

Place your result file, the evaluation dataset, and the evaluation script in the same folder.

Note: you need to run an evaluation script for each result file separately.

task1_eval_script.py

`python task1_eval_script.py <path to your result csv file> <path to gold standard csv file>`

This is an example of how the outputs given by the script may look like:

```
The following similarity scores may need checking:
(absorb,learn) similarity score: 0.748, gold ranking: 5.48
(absorb,withdraw) similarity score: 0.997, gold ranking: 2.97
-----
(absorb,possess) similarity score: 0.5, gold ranking: 5.0
(absorb,withdraw) similarity score: 0.997, gold ranking: 2.97
-----
(acquire,obtain) similarity score: 0.157, gold ranking: 8.57
(acquire,find) similarity score: 0.638, gold ranking: 6.38
-----
(area,region) similarity score: 0.147, gold ranking: 9.47
(area,zone) similarity score: 0.833, gold ranking: 8.33
-----
(area,region) similarity score: 0.147, gold ranking: 9.47
(area,corner) similarity score: 0.207, gold ranking: 2.07
-----
```

task2_eval_script.py

`python task2_eval_script.py <path to your result csv file> <path to gold standard csv file>`

This is an example of how the outputs given by the script may look like:

```
Class level:
Class 1 precision: 1.0000 recall: 1.0000
Class 2 precision: 1.0000 recall: 1.0000
Class 3 precision: 1.0000 recall: 1.0000
Class 4 precision: 0.0000 recall: 0.0000
Class 5 precision: 0.3333 recall: 0.6667
Class 6 precision: 1.0000 recall: 0.5000
Class 7 precision: 0.8000 recall: 0.8000
Class 8 precision: 0.0000 recall: 0.0000
Class 9 precision: 1.0000 recall: 1.0000
Movie level:
Precision: 0.7800
Recall: 0.8167
```