Overcoming catastrophic forgetting

Scielo dataset

The Scielo dataset contains monolingual and parallel texts from scientific publications in English, Spanish, Portuguese and French.

This databset was used in WMT16 for Biomedical Translation Task, with tthe aims to evaluate systems on the translation of scientific publications for the biological and health domains. more (https://www.statmt.org/wmt16/biomedical-translation-task.html)

Preprocessing

Parallel corpus

This dataset is available the BioC XML format (http://bioc.sourceforge.net/) and require some preprocessing and aligment in order to end up with a usable paralell corpora.

Although we have preprocessed and aligned this dataset from scratch, we have end up using the aligned corpora provided by the WMT in order to reuse the test sets from the WMT15 challenge.

Working format

To ease the work with this dataset, we have convert it to a CSV file (UFT-8),

Cleaning

We have cleaned every corpus using the following steps:

1. Since Remove repeated whitespaces
2. Convert text to NFD to avoid encoding problems
3. Strip whitespace
4. Removing all pairs where one of the language didn't have a translation

Since both, the health and biological domain can make a heavy use of Chemical formulas and drugs names, we have consider that lowering the text was not convenient.

Also, we do not differenciate between texts from titles or abstracts.

In addition to this, we have prepare two version of the dataset:

- **Unconstrained:** Datasets that use all the available data
- **Constrained:** Datasets limited to a maximum number of pairs. This limit correspond to the number of pairs of the smallest dataset

Cleaning output (UNCONSTRAINED)

Train (tr+val):

```
/home/salvacarrion/anaconda3/envs/mltests/bin/python /home/salva
carrion/Documents/Programming/Python/mltests/paper1/code/preproc
ess/clean.py
Processing training files...
Reading file... (es-en-gma-biological.csv)
100%|██████████| 125828/125828 [00:10<00:00, 11960.31it/s]
Stats for: es-en-gma-biological.csv **************************
    - Documents: 17672
    - Sentences: 123597
        - Removed: 2231 (1.77%)
    - Titles/Abstracts: 8867/114730 (7.73%)
File saved!

Reading file... (es-en-gma-health.csv)
100%|██████████| 587299/587299 [00:48<00:00, 12119.58it/s]
Stats for: es-en-gma-health.csv **************************
    - Documents: 75856
    - Sentences: 580482
        - Removed: 6817 (1.16%)
    - Titles/Abstracts: 44508/535974 (8.30%)
File saved!

Reading file... (fr-en-gma-health.csv)
100%|██████████| 9127/9127 [00:00<00:00, 11951.10it/s]
Stats for: fr-en-gma-health.csv **************************
    - Documents: 1135
    - Sentences: 9040
        - Removed: 87 (0.95%)
    - Titles/Abstracts: 740/8300 (8.92%)
File saved!

Reading file... (pt-en-gma-biological.csv)
100%|██████████| 121874/121874 [00:10<00:00, 12057.78it/s]
Stats for: pt-en-gma-biological.csv **************************
    - Documents: 18180
    - Sentences: 120301
        - Removed: 1573 (1.29%)
    - Titles/Abstracts: 5766/114535 (5.03%)
File saved!

Reading file... (pt-en-gma-health.csv)
100%|██████████| 512564/512564 [00:42<00:00, 12043.31it/s]
Stats for: pt-en-gma-health.csv **************************
    - Documents: 65659
    - Sentences: 507987
        - Removed: 4577 (0.89%)
    - Titles/Abstracts: 48986/459001 (10.67%)
File saved!
```

```
Done!

Process finished with exit code 0
```

Test:

```
/home/salvacarrion/anaconda3/envs/mltests/bin/python /home/salva
carrion/Documents/Programming/Python/mltests/paper1/code/preproc
ess/1_clean.py
Processing test files...
Reading file... (test-gma-en2es-biological.csv)
100%|██████████| 4003/4003 [00:00<00:00, 12253.74it/s]
Stats for: test-gma-en2es-biological.csv ***********************
***
    - Documents: 500
    - Sentences: 3933
        - Removed: 70 (1.75%)
    - Titles/Abstracts: 513/3420 (15.00%)
File saved!

Reading file... (test-gma-en2es-health.csv)
100%|██████████| 4854/4854 [00:00<00:00, 12497.96it/s]
Stats for: test-gma-en2es-health.csv ***************************
    - Documents: 500
    - Sentences: 4816
        - Removed: 38 (0.78%)
    - Titles/Abstracts: 506/4310 (11.74%)
File saved!

Reading file... (test-gma-en2fr-health.csv)
100%|██████████| 5268/5268 [00:00<00:00, 11930.75it/s]
Stats for: test-gma-en2fr-health.csv ***************************
    - Documents: 500
    - Sentences: 5023
        - Removed: 245 (4.65%)
    - Titles/Abstracts: 510/4513 (11.30%)
File saved!

Reading file... (test-gma-en2pt-biological.csv)
100%|██████████| 4100/4100 [00:00<00:00, 12347.91it/s]
Stats for: test-gma-en2pt-biological.csv ***********************
***
    - Documents: 500
    - Sentences: 4055
        - Removed: 45 (1.10%)
    - Titles/Abstracts: 502/3553 (14.13%)
File saved!

Reading file... (test-gma-en2pt-health.csv)
100%|██████████| 3694/3694 [00:00<00:00, 12371.66it/s]
Stats for: test-gma-en2pt-health.csv ***************************
    - Documents: 500
    - Sentences: 3676
        - Removed: 18 (0.49%)
    - Titles/Abstracts: 504/3172 (15.89%)
```

```
File saved!

Reading file... (test-gma-es2en-biological.csv)
100%|██████████| 4028/4028 [00:00<00:00, 12156.61it/s]
Stats for: test-gma-es2en-biological.csv ************************
***
    - Documents: 500
    - Sentences: 3982
        - Removed: 46 (1.14%)
    - Titles/Abstracts: 511/3471 (14.72%)
File saved!

Reading file... (test-gma-es2en-health.csv)
100%|██████████| 5027/5027 [00:00<00:00, 11226.09it/s]
Stats for: test-gma-es2en-health.csv ***************************
    - Documents: 500
    - Sentences: 4961
        - Removed: 66 (1.31%)
    - Titles/Abstracts: 508/4453 (11.41%)
File saved!

Reading file... (test-gma-fr2en-health.csv)
100%|██████████| 5356/5356 [00:00<00:00, 11859.21it/s]
Stats for: test-gma-fr2en-health.csv ***************************
    - Documents: 500
    - Sentences: 5072
        - Removed: 284 (5.30%)
    - Titles/Abstracts: 506/4566 (11.08%)
File saved!

Reading file... (test-gma-pt2en-biological.csv)
100%|██████████| 3914/3914 [00:00<00:00, 10782.80it/s]
Stats for: test-gma-pt2en-biological.csv ************************
***
    - Documents: 500
    - Sentences: 3871
        - Removed: 43 (1.10%)
    - Titles/Abstracts: 506/3365 (15.04%)
File saved!

Reading file... (test-gma-pt2en-health.csv)
100%|██████████| 3757/3757 [00:00<00:00, 12287.98it/s]
Stats for: test-gma-pt2en-health.csv ***************************
    - Documents: 500
    - Sentences: 3738
        - Removed: 19 (0.51%)
    - Titles/Abstracts: 506/3232 (15.66%)
File saved!
```

```
Done!

Process finished with exit code 0
```

Cleaning output (CONSTRAINED)

Train (tr+val):

```
/home/salvacarrion/anaconda3/envs/nmt/bin/python /snap/pycharm-p
rofessional/230/plugins/python/helpers/pydev/pydevd.py --multipr
oc --qt-support=auto --client 127.0.0.1 --port 33541 --file /hom
e/salvacarrion/Documents/Programming/Python/mltests/paper1/code/
preprocess/1_clean.py
Connected to pydev debugger (build 203.7148.72)
Processing training files...
Reading file... (es-en-gma-biological.csv)
100%|██████████| 125828/125828 [00:27<00:00, 4642.78it/s]
Stats for: es-en-gma-biological.csv **************************
    - Documents: 17672
    - Sentences: 123597
        - Removed: 2231 (1.77%)
    - Titles/Abstracts: 8867/114730 (7.73%)
File saved!

Reading file... (es-en-gma-health.csv)
100%|██████████| 587299/587299 [02:05<00:00, 4671.29it/s]
Limiting size to 123597
Stats for: es-en-gma-health.csv **************************
    - Documents: 59237
    - Sentences: 123597
        - Removed: 463702 (78.96%)
    - Titles/Abstracts: 9548/114049 (8.37%)
File saved!

Reading file... (fr-en-gma-health.csv)
100%|██████████| 9127/9127 [00:01<00:00, 4570.30it/s]
Stats for: fr-en-gma-health.csv **************************
    - Documents: 1135
    - Sentences: 9040
        - Removed: 87 (0.95%)
    - Titles/Abstracts: 740/8300 (8.92%)
File saved!

Reading file... (pt-en-gma-biological.csv)
100%|██████████| 121874/121874 [00:26<00:00, 4654.42it/s]
Stats for: pt-en-gma-biological.csv **************************
    - Documents: 18180
    - Sentences: 120301
        - Removed: 1573 (1.29%)
    - Titles/Abstracts: 5766/114535 (5.03%)
File saved!

Reading file... (pt-en-gma-health.csv)
100%|██████████| 512564/512564 [01:49<00:00, 4699.29it/s]
Limiting size to 120301
Stats for: pt-en-gma-health.csv **************************
    - Documents: 54294
```

```
    - Sentences: 120301
        - Removed: 392263 (76.53%)
    - Titles/Abstracts: 11662/108639 (10.73%)
File saved!

Done!

Process finished with exit code 0
```

Test:

These files are the same as the ones used in unconstrained dataset

Splits

Since no validation set was provided, we have create our owns (from the training set), using the following formula: `min(5000, max(3000, TRAININGSET_LEN*0.03))` . This formula is arbitrary and was used to have a validation set similar in size to the test set but not too large.

Split sizes: (UNCONSTRAINED)

```
#Command: wc -l $(ls)
(Values indicate the number of sentences)
```

	Dataset	Split	Size
0	biological-es-en	train	119890
1	biological-es-en	val	3707
2	biological-es-en	test	3982
3	biological-pt-en	train	116692
4	biological-pt-en	val	3609
5	biological-pt-en	test	3871
6	health-es-en	train	575482
7	health-es-en	val	5000
8	health-es-en	test	4961
9	health-pt-en	train	502987
10	health-pt-en	val	5000
11	health-pt-en	test	3738
12	merged-es-en	train	699079
13	merged-es-en	val	5000
14	merged-es-en	test	8943
15	merged-pt-en	train	623288
16	merged-pt-en	val	5000
17	merged-pt-en	test	7609

 sizes

Split sizes: (CONSTRAINED)

```
#Command: wc -l $(ls)
(Values indicate the number of sentences)

0   biological-es-en  train  119890
1   biological-es-en    val    3707
2   biological-es-en   test    3982
3   biological-pt-en  train  116692
4   biological-pt-en    val    3609
5   biological-pt-en   test    3871
6       health-es-en  train  119890
7       health-es-en    val    3707
8       health-es-en   test    4961
9       health-pt-en  train  116692
10      health-pt-en    val    3609
11      health-pt-en   test    3738
12      merged-es-en  train  242194
13      merged-es-en    val    5000
14      merged-es-en   test    8943
15      merged-pt-en  train  235602
16      merged-pt-en    val    5000
17      merged-pt-en   test    7609
Figure saved!
```

 sizes

Tokenization and BPE

Finally, we have tokenized the dataset using Moses, and fastBPE to learn and apply the subword tokenization with a maximum dictionary size of 32,000 tokens.

Examples

Health

Translation pair #1:

- DE@@ TEC@@ CIÓN R@@ Á@@ P@@ ID@@ A DEL VI@@ R@@ US DE LA G@@ AS@@ TR@@ O@@ EN@@ T@@ ERI@@ TI@@ S TR@@ AN@@ SM@@ I@@ SI@@ B@@ LE EN C@@ ER@@ DO A TR@@ AV@@ É@@ S DE LA RE@@ AC@@ CIÓN EN C@@ AD@@ ENA DE LA PO@@ LI@@ MER@@ ASA
- R@@ AP@@ ID DE@@ TEC@@ TION OF SW@@ INE TR@@ AN@@ SM@@ IS@@ SI@@ B@@ LE G@@ AS@@ TR@@ O@@ EN@@ T@@ ERI@@ TI@@ S VI@@ R@@ US BY N@@ EST@@ ED P@@ OL@@ Y@@ MER@@ AS@@ E CH@@ A@@ IN RE@@ AC@@ TION

Translation pair #2:

- Una ingestión excesiva de calorías , combinada con una vida sedent@@ aria , promueve la expresión fenotípica de este síndrome en individuos con predisposición genética .
- Ex@@ ces@@ sive intake of calories together with lack of exercising leads to phen@@ otyp@@ ical expression of this syndrome in individuals with genetic predisposition .

Translation pair #3:

- Pacientes : catorce pacientes con una edad media de 60 años diagnosticados de cirrosis hepática en estadios A y B que se les practicó colecistectomía laparoscópica .
- Patients : we studied 14 patients (mean age 60 y@@ rs) with Child 's Class A and Class B hepatic cirrhosis who underwent laparoscopic cholecystectomy .

Biological

Translation pair #1:

- La caracterización de los extractos puri@@ ficados se realizó por barrido espectrofoto@@ métrico , pruebas bioquímicas de electroforesis SDS @-@ PAGE , ensayo Pur@@ pal@@ d y la prueba cromo@@ génica de L@@ AL .
- The characterization of the purified extracts was performed by spectrophotometric scanning , SDS @-@ PAGE biochemical tests , Pur@@ pal@@ d assay and chromo@@ genic L@@ AL test .

Translation pair #2:

- Se utilizó un diseño experimental completamente al azar , con arreglo factorial , con cinco repeticiones .
- A completely randomized experimental design with a factorial arrangement and five replications was used .

Translation pair #3:

- Y su relación con el desarrollo de Bemisia tabaci Gen@@ n .
- And their relationship to the development of Bemisia tabaci Gen@@ n .

Merge

Translation pair #1:

- Clínicamente muestran aspecto sol@@ ap@@ ables con las forma ex@@ trat@@ es@@ ticulares .
- Clin@@ ically they show features overl@@ apping the ex@@ trat@@ esticular types .

Translation pair #2:

- Estrategias como la coordinación e investigación pueden optim@@ iz@@ arse para alcanzar sus objetivos .
- Strategies such as coordination and research could be optim@@ ised so that their objectives can be reached .

Translation pair #3:

- Los resultados obtenidos fueron muy satisfactorios .

- The results obtained were satisfactory .

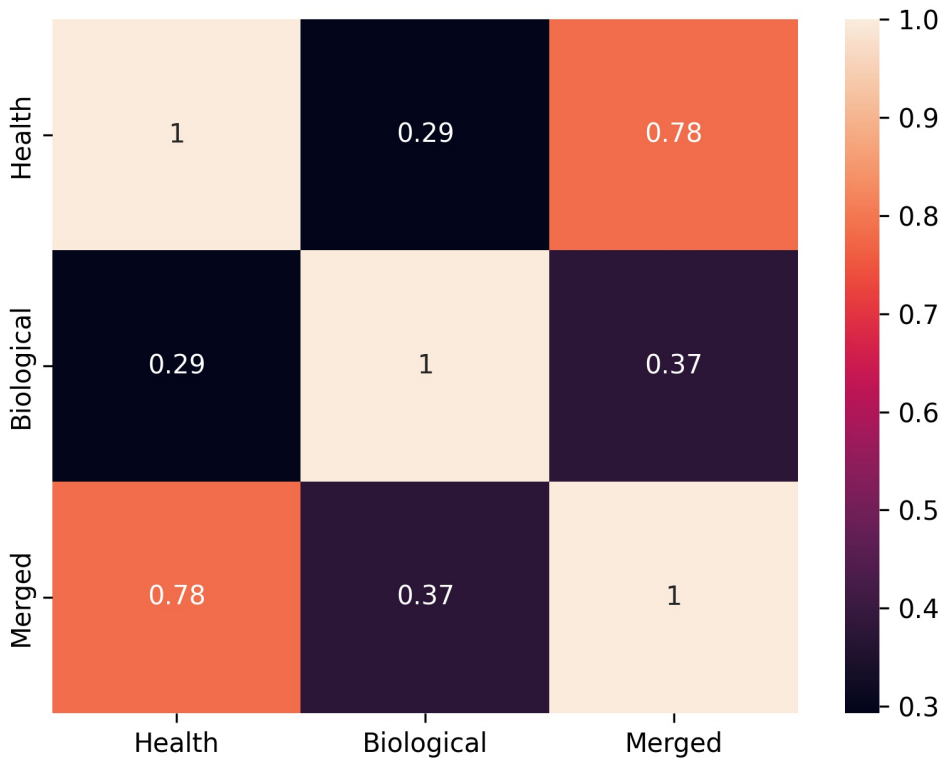
Dataset overlapping

With the objective to visualize the vocabulary overlap among domains, we have plot two matrices: one with IoU values, and another with the overlapping between domain1 and domain2:

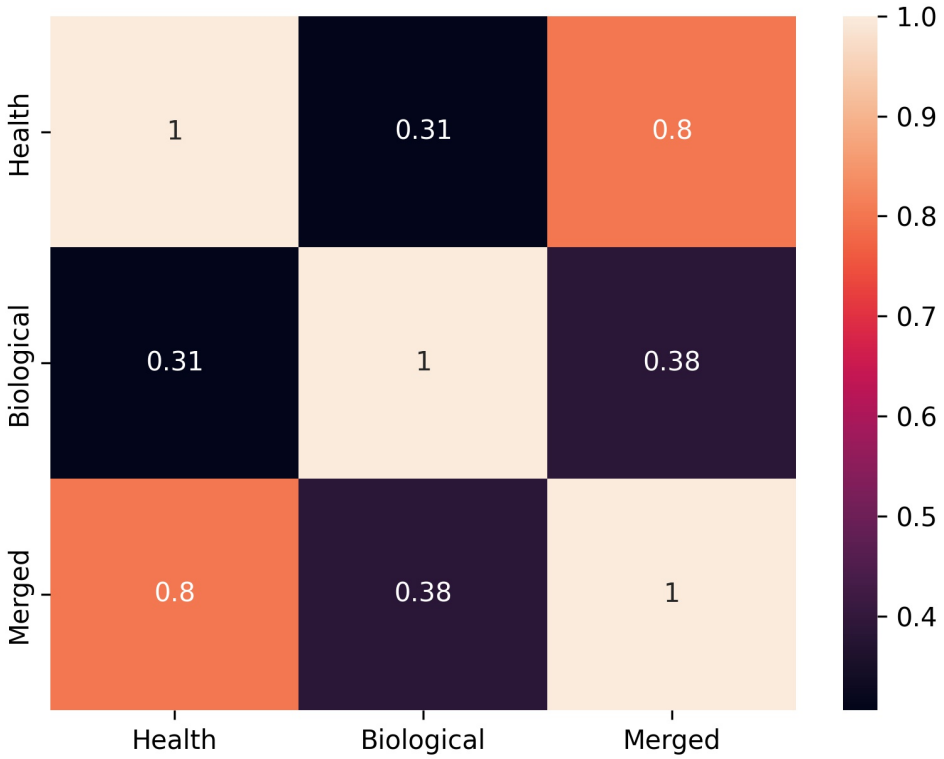
Intersection Over Union (unconstrained)

> Formula:
> `iou=vocab1.intersection(vocab2)/len(vocab1.union(vocab2))`

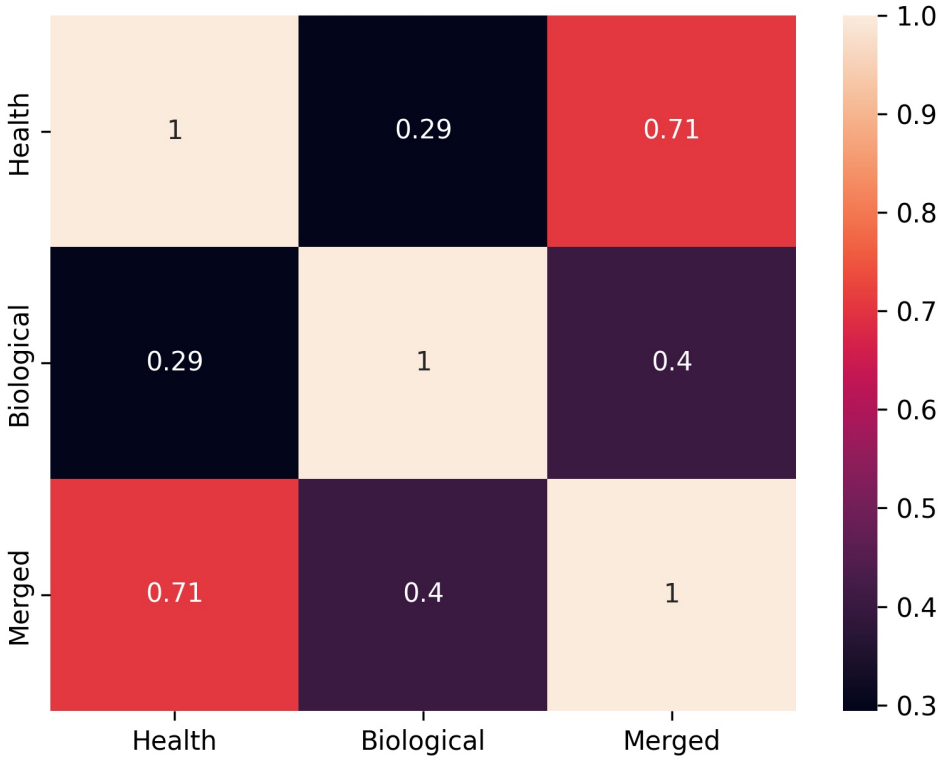
es-en | data: es



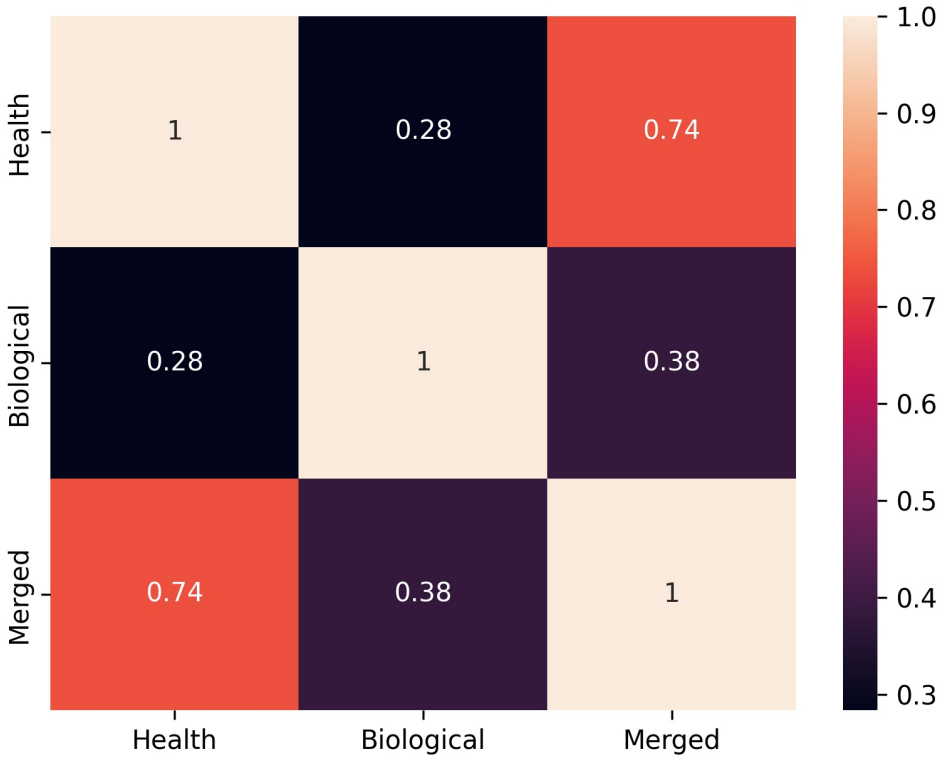
es-en | data: en


Health
Biological
Merged
1
0.31
0.8
0.31
1
0.38
0.8
0.38
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5
0.4

pt-en | data: pt


Health
Biological
Merged
1
0.29
0.71
0.29
1
0.4
0.71
0.4
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3

pt-en | data: en

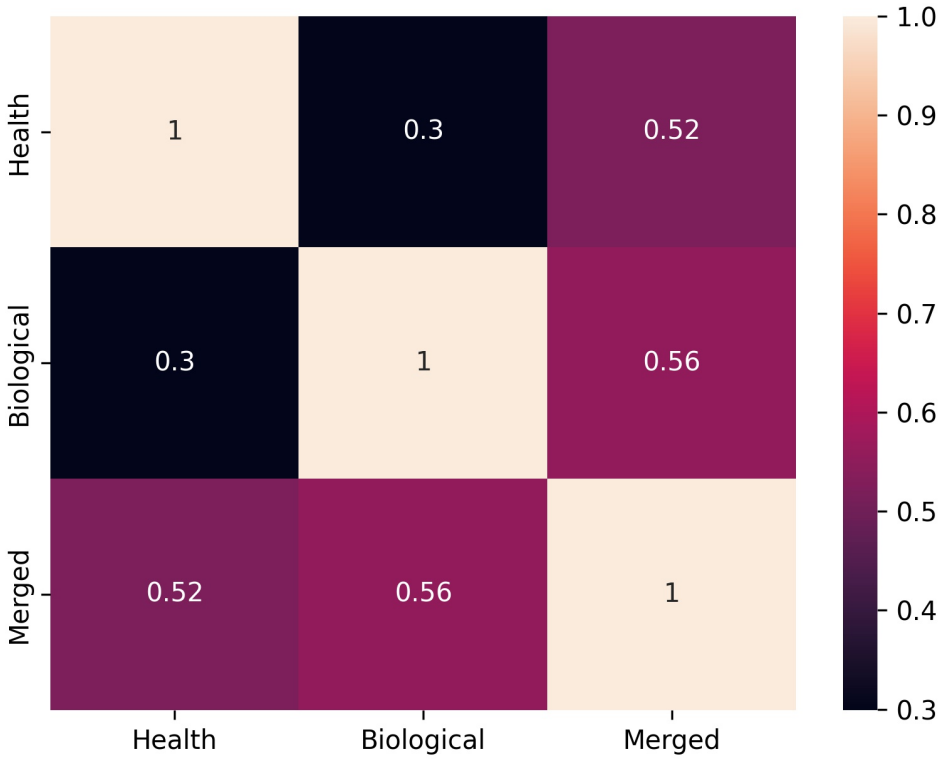


Intersection Over Union (constrained)

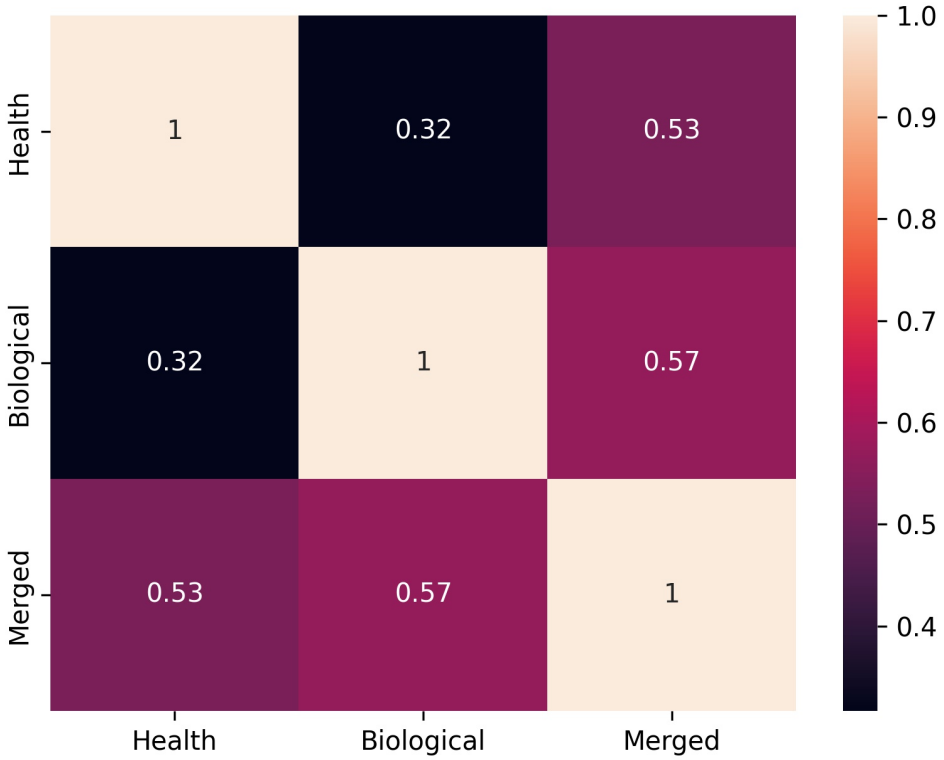
Formula:

```
iou=vocab1.intersection(vocab2)/len(vocab1.union(vocab2))
```

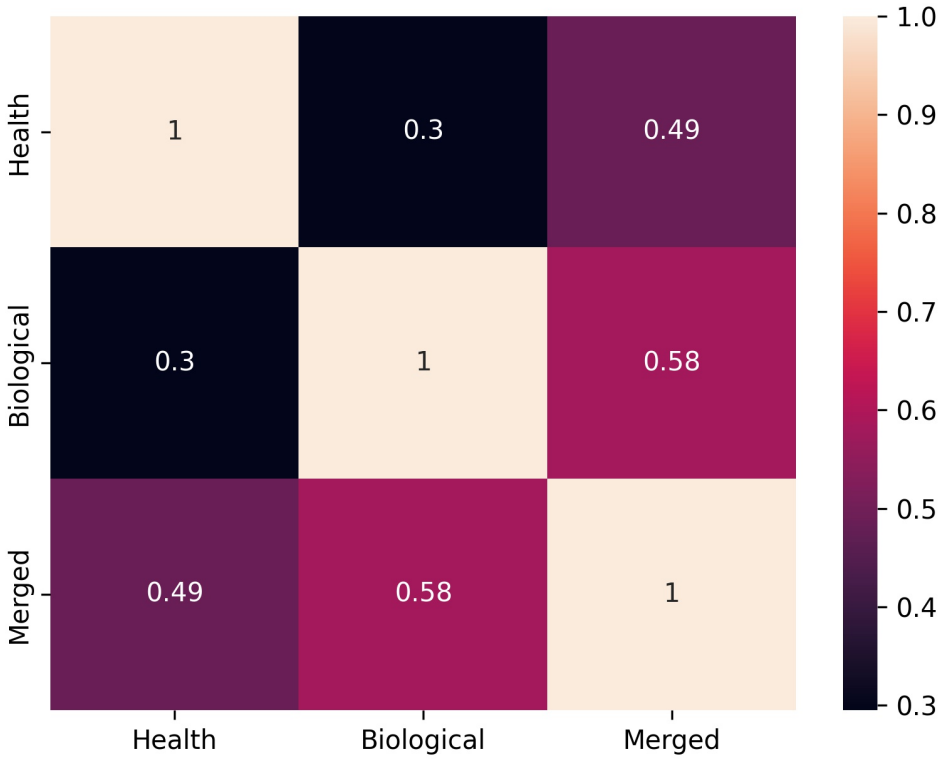
es-en | data: es


Health
Biological
Merged
1
0.3
0.52
0.3
1
0.56
0.52
0.56
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3

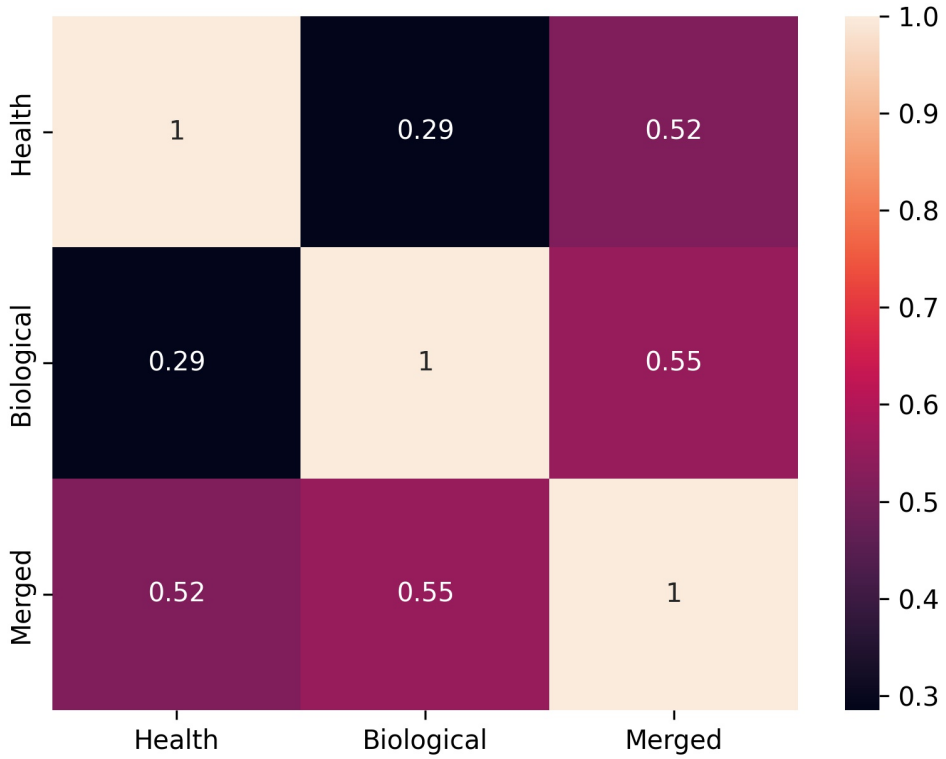
es-en | data: en


Health
Biological
Merged
1
0.32
0.53
0.32
1
0.57
0.53
0.57
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5
0.4

pt-en | data: pt



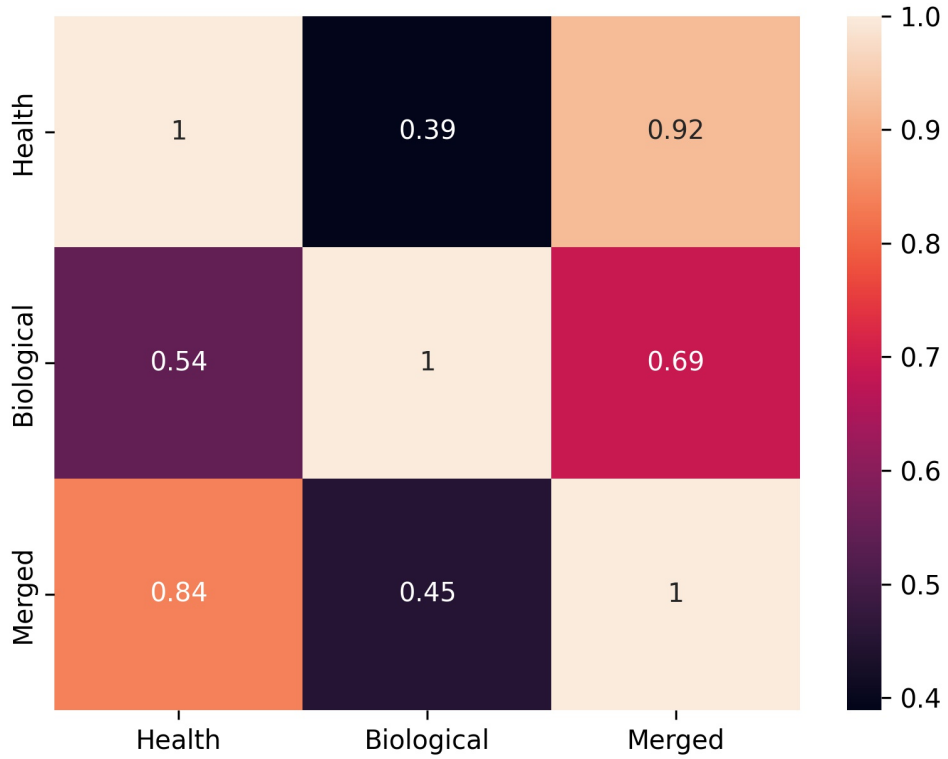
pt-en | data: en



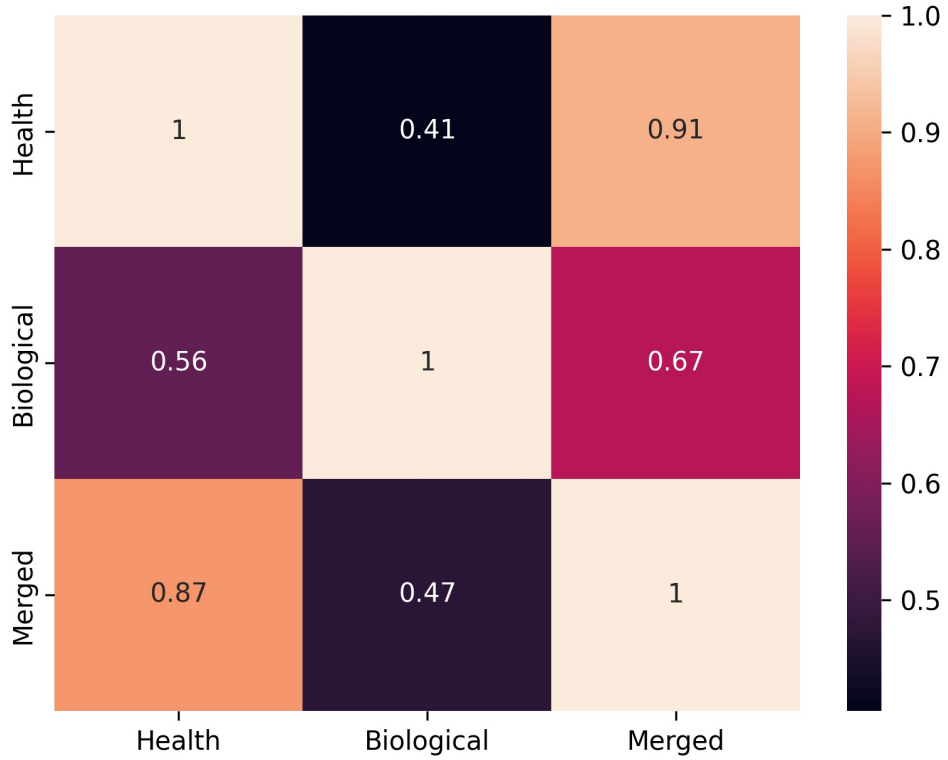
Overlapping (domain1 - domain2) (unconstrained)

Formula: `overlap=vocab1.intersection(vocab2)/len(vocab1)`

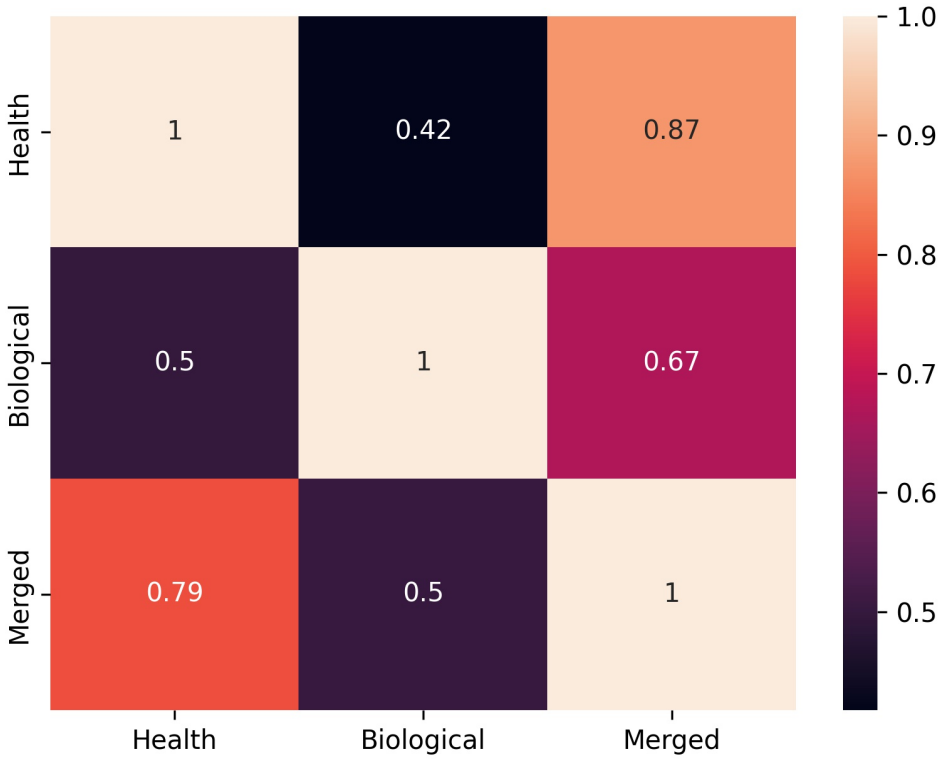
es-en | data: es


Health
Biological
Merged
1
0.39
0.92
0.54
1
0.69
0.84
0.45
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5
0.4

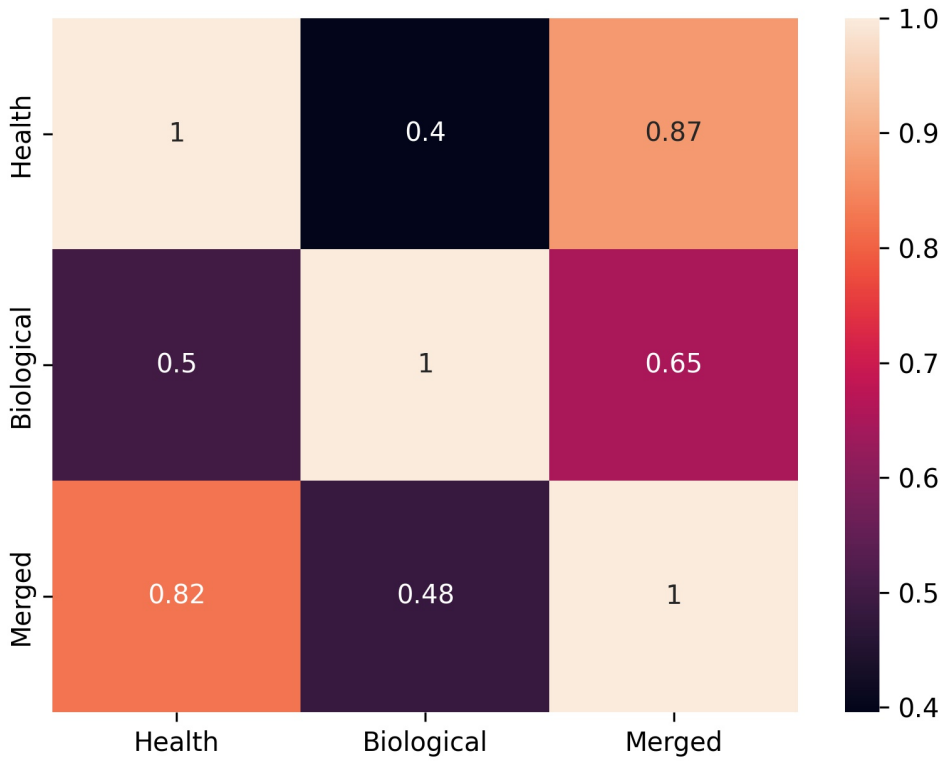
es-en | data: en


Health
Biological
Merged
1
0.41
0.91
0.56
1
0.67
0.87
0.47
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5

pt-en | data: pt



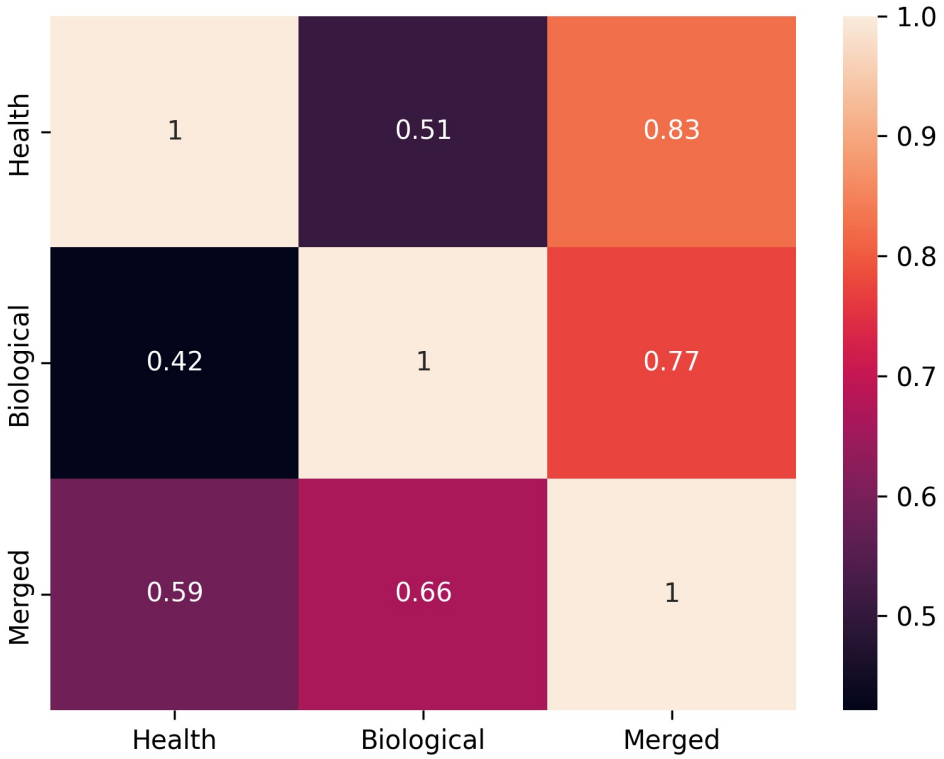
pt-en | data: en



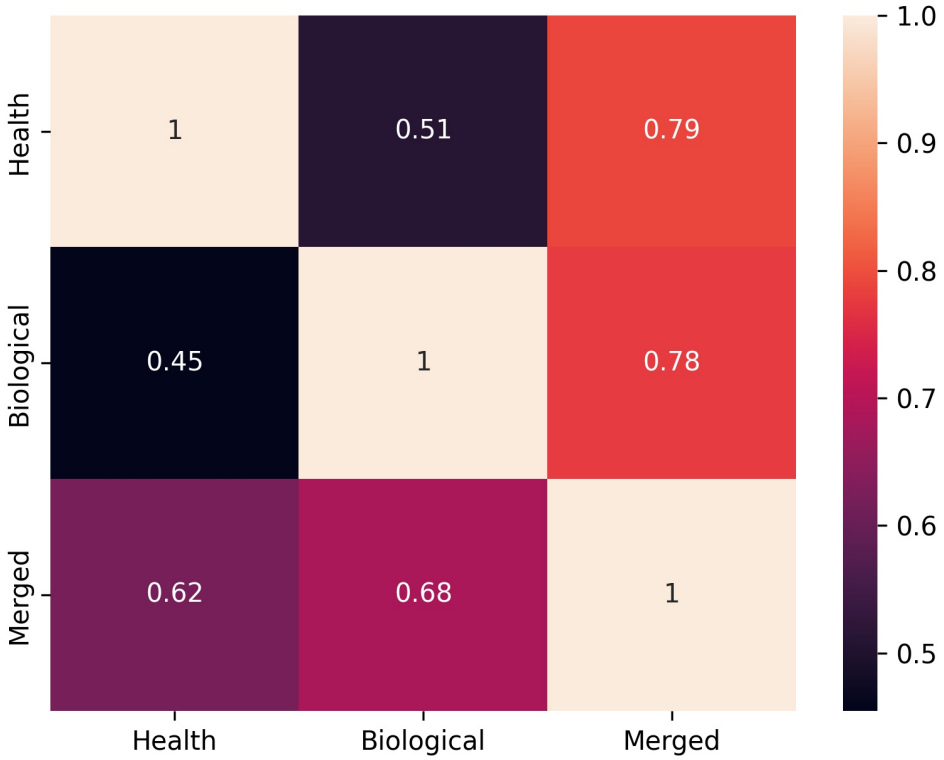
Overlapping (domain1 - domain2) (constrained)

Formula: `overlap=vocab1.intersection(vocab2)/len(vocab1)`

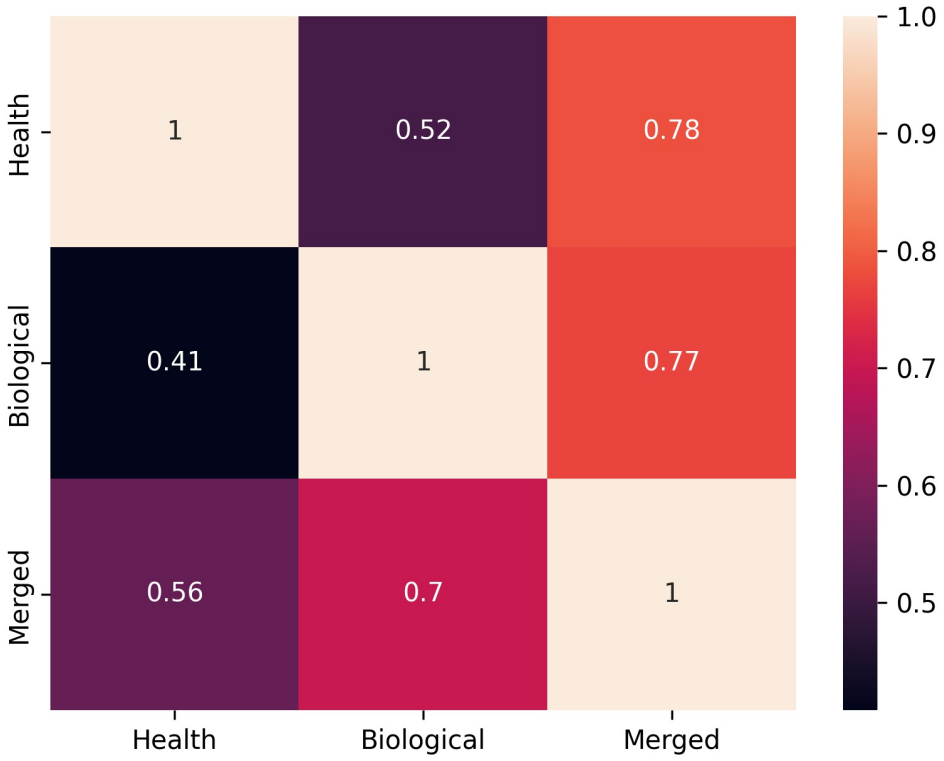
es-en | data: es


Health
Biological
Merged
1
0.51
0.83
0.42
1
0.77
0.59
0.66
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5

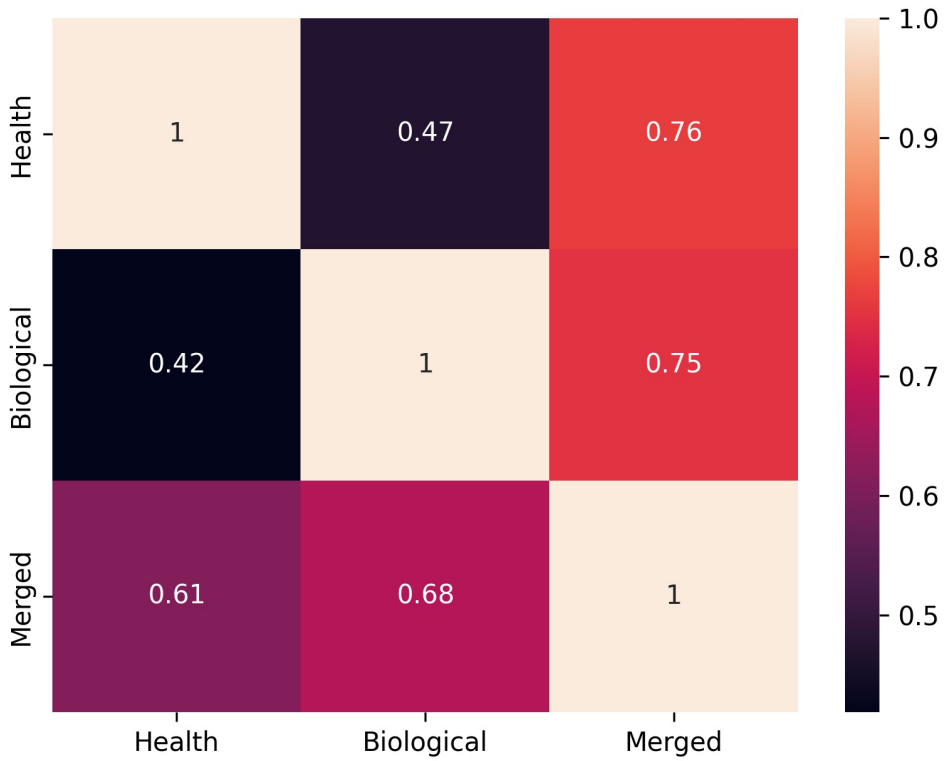
es-en | data: en


Health
Biological
Merged
1
0.51
0.79
0.45
1
0.78
0.62
0.68
1
Health
Biological
Merged
1.0
0.9
0.8
0.7
0.6
0.5

pt-en | data: pt



pt-en | data: en



Methodology

Baseline #1: Naive training

- Train one model per language and domain (2 language pairs x 3 domains).
  - *Example: (es-en) and (pt-en) for (health, biological and health+biological)*

Baseline #2: Sequencial training

- Evidence of the catastrophic forguetting problem

Baseline #3: Approaches for mitigating catastrohpic forguetting

- Fixed/Bayesian Interpolation
- Elastic Weight Consolidation

Proposal: Freeze architecture, extend, and train extension

- Fine tuning
- Model surgery
- Reinforcement learning

Training

We have trained 6 models* using the same base architecture for all of them, and the same configuration. *(Toolkit: fairseq 0.10.2)*

Trained models:

- **es-en:** health, biological, health+biological, health->biological
- **pt-en:** health, biological, health+biological, health->biological

Main parameters:

```
    --arch transformer \
    --optimizer adam \
    --criterion label_smoothed_cross_entropy --label-smoothing
 0.1 \
    --max-tokens 4096 \
    --num-workers    $(nproc) \
    --max-epoch    50 \
    --seed 1 \
    --save-dir $BASE_PATH/checkpoints \
    --log-format simple \
    --no-epoch-checkpoints \
    --tensorboard-logdir $BASE_PATH/logdir \
    --update-freq 8 \
    --eval-bleu \
    --eval-bleu-args '{"beam": 5, "max_len_a": 1.2, "max_len_b":
10}' \
    --eval-bleu-detok moses \
    --eval-bleu-remove-bpe \
    --eval-bleu-print-samples \
    --best-checkpoint-metric bleu --maximize-best-checkpoint-met
ric \
    --clip-norm 1.0 \
    --lr 1e-3  \
    --lr-scheduler reduce_lr_on_plateau  \
    --warmup-updates 4000 \
    --dropout 0.1 --weight-decay 0.0001 \
    --patience 5 \
    --wandb-project "mltests"
```

Progress (unconstrained)

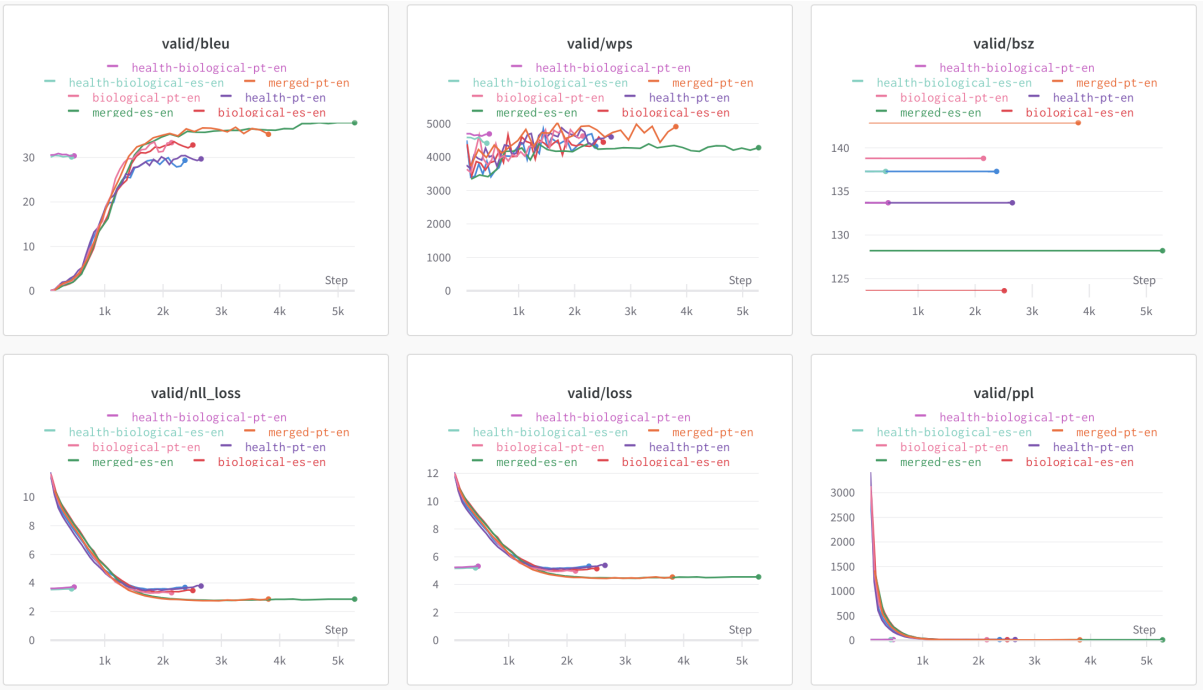
Training:

Validation:

Progress (constrained)

Training:

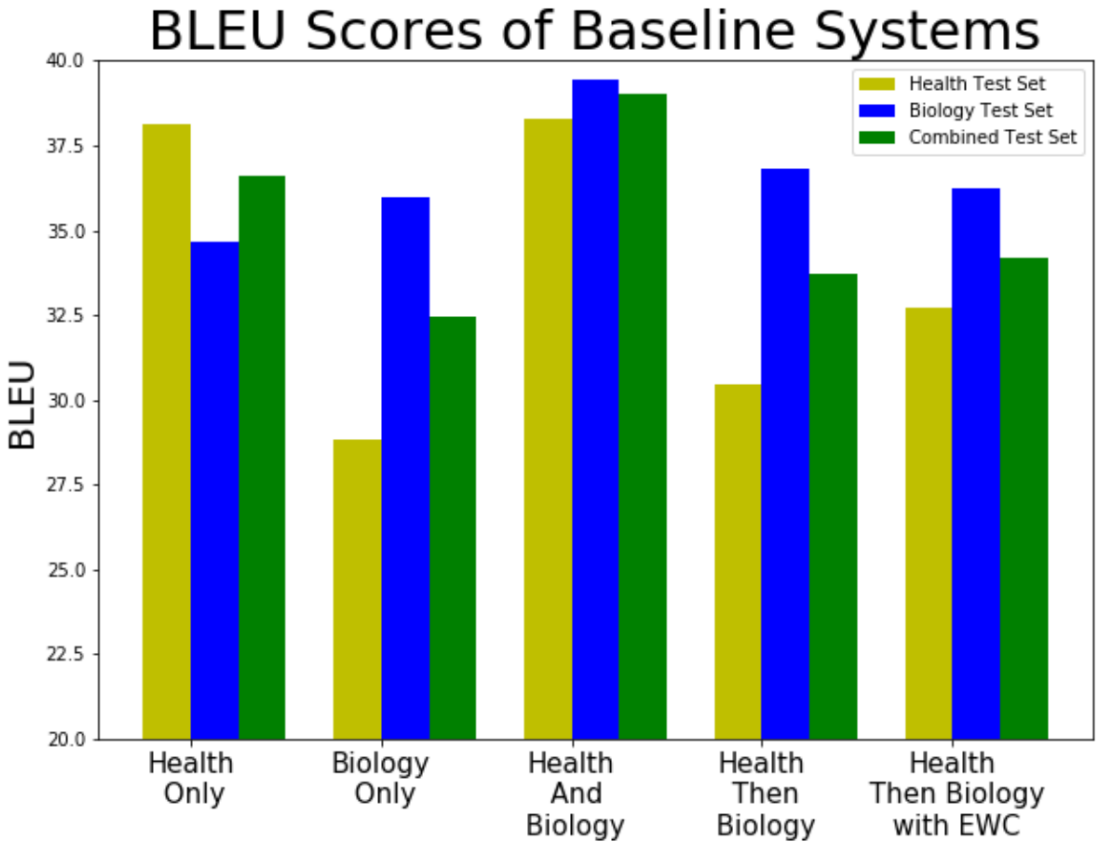
Validation:



Evaluation

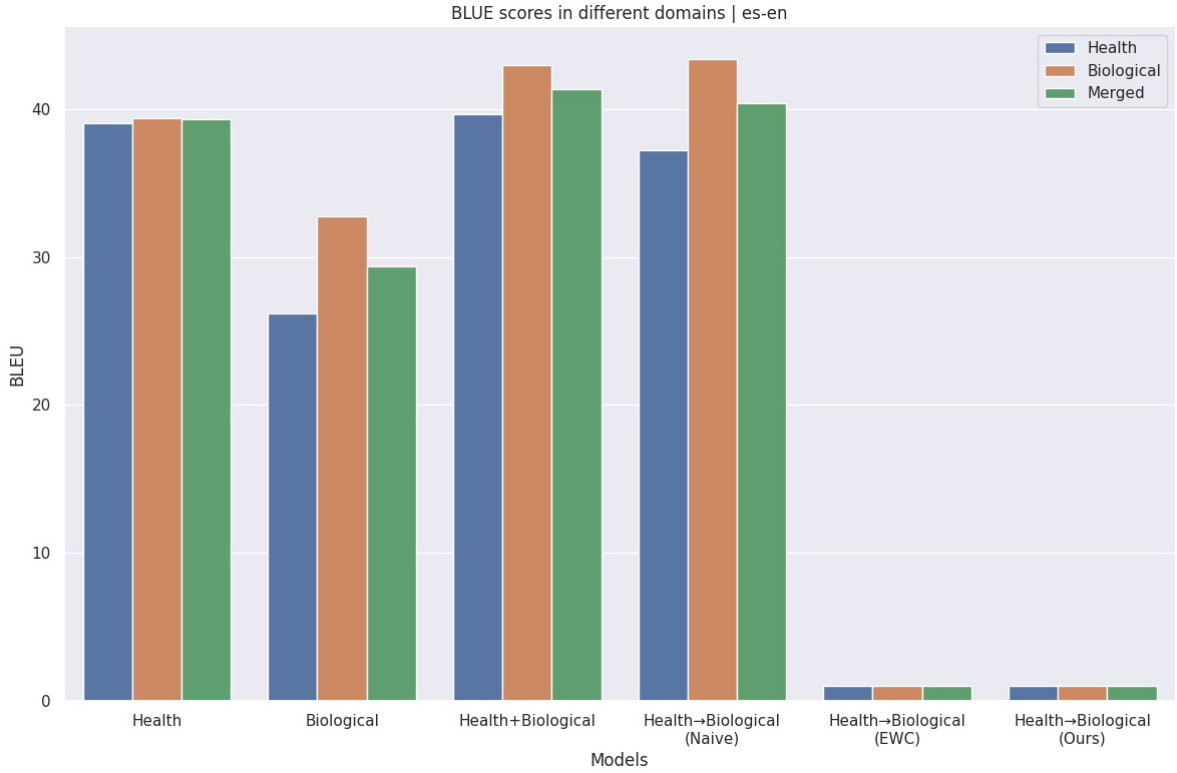
Each one of the 6 trained models has been evaluated on each domain, using the BLEU score from Fairseq (used for the WMT).

Reference (kell master thesis)

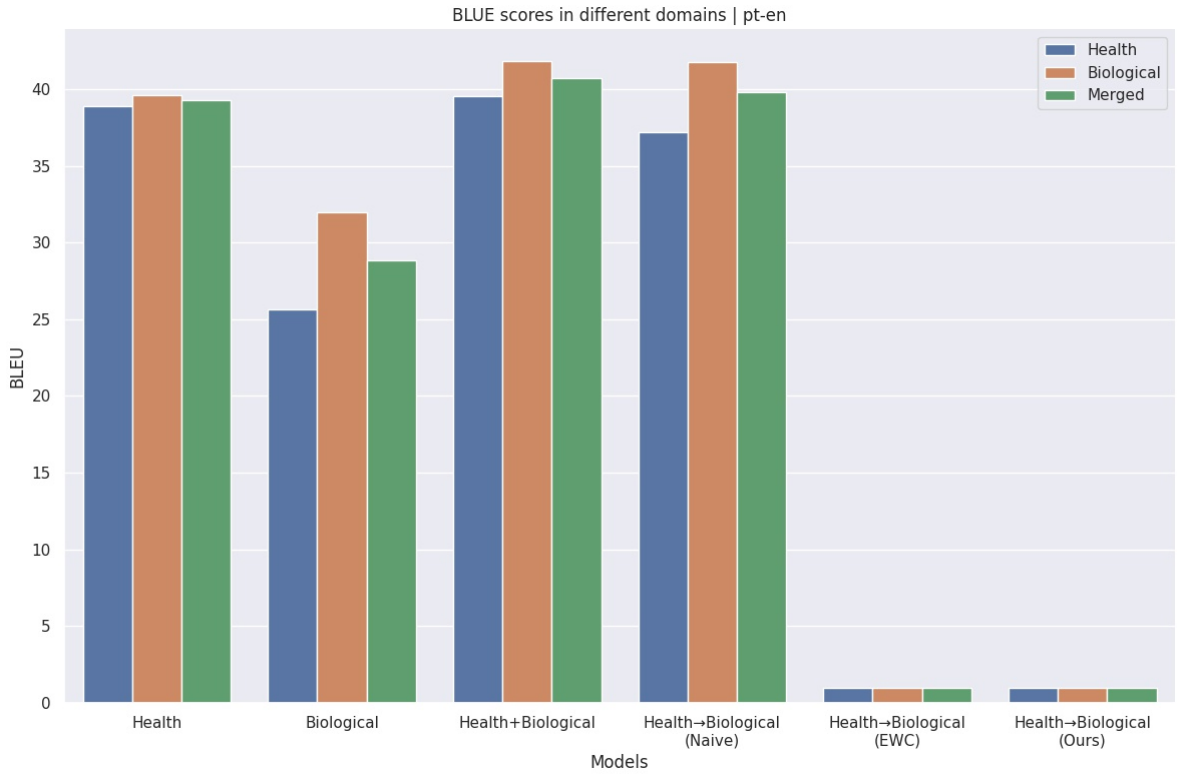


Unconstrained

Spanish-English:



Portuguese-English:



Comments:

- The results from es-en and pt-en are pretty similar across all tested domains. Therefore, the comments below can apply to both of them.
- **Health model:** This model is trained on the health domain exclusively. However, this case is interesting since we expect the health model to perform very well on its domain (health), then to perform reasonably well on the merged domain (as there is more data), and finally, to perform worse on the biology domain. However, the health model performs very well in all domains with no significant differences among them. We argue that this could be related to the amount of data available for each domain. The training set from the health domain contains around ~550k sentences, while the biological domain has only 100k sentences. This imbalance in terms of data could lead to inadvertively train more performant models when using the health domain. To test this hyphotesis, we have trained a new model but using the same amount of data as it is available for the biological domain (later).
- **Biological model:** This model is trained on the biological domain exclusively. It performs well in its domain, a little worse in the merged domain and finally, the health domain is the one with the worst results as expected.
- **Health+Biology:** This model is trained on the health and biological domain at once. Here, we observed a consistent increase in performance for all domains. As the amount of data in the biology domain is significantly smaller than the health domain, the improvements are only slightly better than the health domain.
- **Health->Biology:** This model is trained sequentially. First on the health domain, and then fine-tuned on the biological domain. As a result of this sequential training, the performance in the fine-tuned domain increased but at expense of worsen its performance on a previously learned domain (health). This phenomenon is known as "catastrophic forgetting" and it is the one we will try to mitigate.

Table:

In [4]:
```
import pandas as pd
df = pd.read_csv("../data/unconstrained/test_data.csv")
print(df)
```

```
    Unnamed: 0                        Model Test domain   lang   BLEU
0            0                       Health      Health  es-en  39.02
1            1                       Health  Biological  es-en  39.38
2            2                       Health      Merged  es-en  39.34
3            3                   Biological      Health  es-en  26.19
4            4                   Biological  Biological  es-en  32.76
5            5                   Biological      Merged  es-en  29.36
6            6           Health+Biological      Health  es-en  39.63
7            7           Health+Biological  Biological  es-en  42.95
8            8           Health+Biological      Merged  es-en  41.32
9            9  Health→Biological\n(Naive)      Health  es-en  37.20
10          10  Health→Biological\n(Naive)  Biological  es-en  43.35
11          11  Health→Biological\n(Naive)      Merged  es-en  40.38
12          12    Health→Biological\n(EWC)      Health  es-en   1.00
13          13    Health→Biological\n(EWC)  Biological  es-en   1.00
14          14    Health→Biological\n(EWC)      Merged  es-en   1.00
15          15   Health→Biological\n(Ours)      Health  es-en   1.00
16          16   Health→Biological\n(Ours)  Biological  es-en   1.00
17          17   Health→Biological\n(Ours)      Merged  es-en   1.00
18          18                       Health      Health  pt-en  38.86
19          19                       Health  Biological  pt-en  39.61
20          20                       Health      Merged  pt-en  39.26
21          21                   Biological      Health  pt-en  25.67
22          22                   Biological  Biological  pt-en  31.94
23          23                   Biological      Merged  pt-en  28.86
24          24           Health+Biological      Health  pt-en  39.54
25          25           Health+Biological  Biological  pt-en  41.80
26          26           Health+Biological      Merged  pt-en  40.74
27          27  Health→Biological\n(Naive)      Health  pt-en  37.16
28          28  Health→Biological\n(Naive)  Biological  pt-en  41.78
29          29  Health→Biological\n(Naive)      Merged  pt-en  39.77
30          30    Health→Biological\n(EWC)      Health  pt-en   1.00
31          31    Health→Biological\n(EWC)  Biological  pt-en   1.00
32          32    Health→Biological\n(EWC)      Merged  pt-en   1.00
33          33   Health→Biological\n(Ours)      Health  pt-en   1.00
34          34   Health→Biological\n(Ours)  Biological  pt-en   1.00
35          35   Health→Biological\n(Ours)      Merged  pt-en   1.00
```

Translated examples

Leyend: Source (S), Target (T), Hypothesis (H), Detokenized (D), Probabilities (P)

Translations not cherry picked

Trained on: Health, Tested on: Health (es-en)

```
S-4762    Algunos b @-@ bloque@@ adores son , también , vasodila
t@@ adores .
T-4762    Some b @-@ blockers are also vasodi@@ l@@ ators .
H-4762    -0.2545783817768097    Some b @-@ blockers are also va
sodi@@ l@@ ators .
D-4762    -0.2545783817768097    Some b @-@ blockers are also va
sodi@@ l@@ ators .
P-4762    -0.2415 -0.0850 -0.1610 -0.1649 -0.3808 -0.3433 -0.750
6 -0.0631 -0.3405 -0.1388 -0.1310
S-858    Dis@@ función miccional post @-@ bra@@ quit@@ erapia en
el cáncer prostático
T-858    Objectives :
H-858    -0.656700849533081    Post @-@ brachy@@ therapy voiding
dysfunction in prostate cancer
D-858    -0.656700849533081    Post @-@ brachy@@ therapy voiding
dysfunction in prostate cancer
P-858    -4.2436 -0.1390 -0.2959 -0.0589 -0.4146 -0.2406 -0.1419
-0.6925 -0.1862 -0.1538
S-2323    La tomografía computadorizada mostró imágenes con aspe
cto de embol@@ ia pulmonar .
T-2323    CT images suggested pulmonary embolism .
H-2323    -0.556445837020874    Computed tomography showed image
s with aspect of pulmonary embolism .
D-2323    -0.556445837020874    Computed tomography showed image
s with aspect of pulmonary embolism .
P-2323    -1.0355 -0.0882 -0.9586 -0.1854 -0.5684 -2.6379 -0.134
8 -0.1575 -0.0706 -0.1530 -0.1310
S-3812    Consumo de alcohol en la consulta de una sala de emerg
encia .
T-3812    Alcohol consumption in the emergency room .
H-3812    -0.5013466477394104    Alcohol consumption in an emerg
ency room .
D-3812    -0.5013466477394104    Alcohol consumption in an emerg
ency room .
P-3812    -0.1431 -0.2048 -0.5730 -1.4506 -0.2033 -0.1445 -1.148
6 -0.1427
```

Trained on: Health, Tested on: Biological (es-en)

```
S-3944    Las técnicas de fec@@ und@@ ación artificial siguen de
sarroll@@ ándose día a día .
T-3944    Arti@@ ficial reproductive technology ( ART ) is impro
ving constantly .
H-3944    -0.7757904529571533    Arti@@ ficial fertilization tec
hniques continue to develop a day to day .
D-3944    -0.7757904529571533    Arti@@ ficial fertilization tec
hniques continue to develop a day to day .
P-3944    -0.8396 -0.0449 -0.8764 -0.3654 -1.3506 -0.7778 -1.550
5 -0.9983 -0.2864 -2.3688 -0.3376 -0.1536 -0.1354

S-1367    El intervalo de tiempo promedio entre aportes de presa
fue de 7@@ 98 s.
T-1367    The mean time interval between pre@@ y deliveries was
 7@@ 98 s .
H-1367    -0.6061871647834778    The average time interval betwe
en pre@@ y contributions was 7@@ 98 .
D-1367    -0.6061871647834778    The average time interval betwe
en pre@@ y contributions was 7@@ 98 .
P-1367    -0.4470 -0.7220 -0.2297 -0.5246 -0.3596 -3.0708 -0.970
8 -0.1361 -0.1207 -0.0876 -0.0675 -0.9636 -0.1805

S-727    El Her@@ b@@ ario US@@ J de Costa Rica : trayectoria y
 contribuciones .
T-727    US@@ J Her@@ bari@@ um of Costa Rica : history and cont
ributions .
H-727    -0.3842181861400604    The Her@@ bari@@ um US@@ J of Co
sta Rica : trajectory and contributions .
D-727    -0.3842181861400604    The Her@@ bari@@ um US@@ J of Co
sta Rica : trajectory and contributions .
P-727    -0.3028 -1.3042 -0.4261 -0.3853 -0.3091 -0.1093 -1.2289
-0.0751 -0.0860 -0.1367 -0.8913 -0.1663 -0.0718 -0.1423 -0.1280
```

Trained on: Biological, Tested on: Health (es-en)

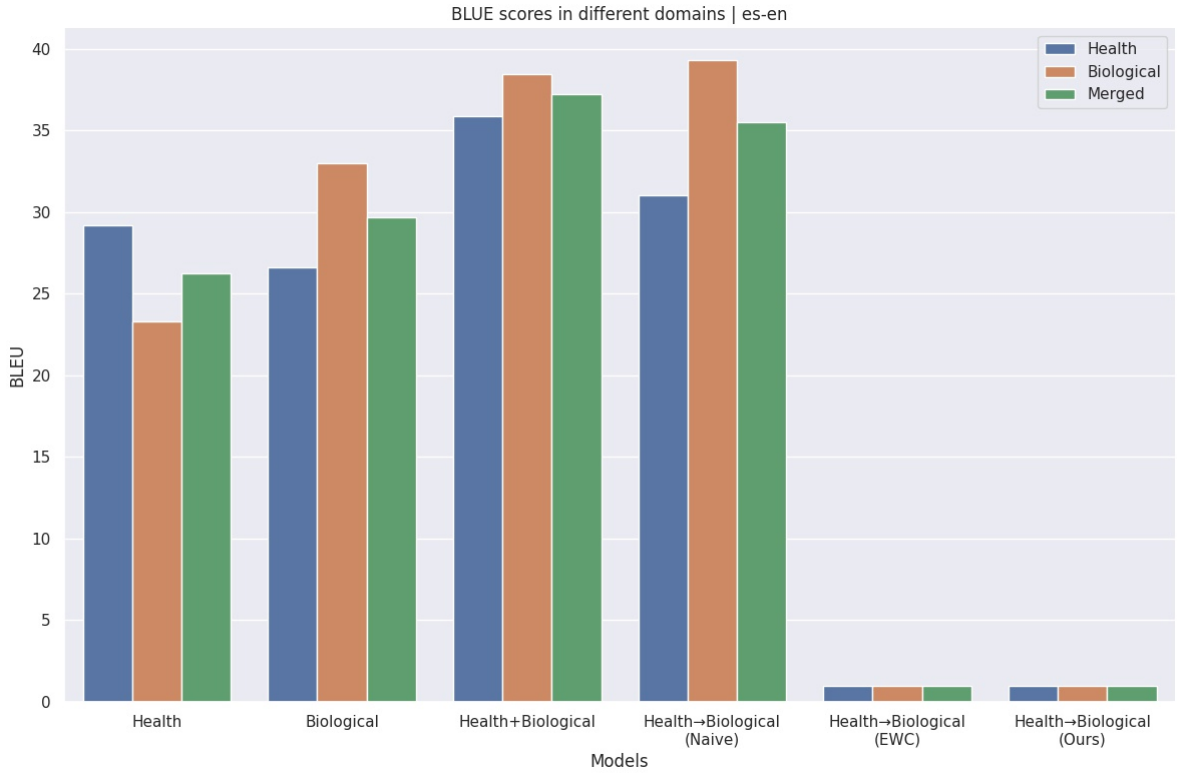
```
S-3629    Un paciente HIV positivo fal@@ leci@@ ó debido a epide
r@@ mo@@ lisis .
T-3629    One HIV positive patient died of epider@@ mol@@ y@@ si
s .
H-3629    -1.3282438516616821    A HIV infection caused HIV infe
ction due to epider@@ mo@@ ist lysis .
D-3629    -1.3282438516616821    A HIV infection caused HIV infe
ction due to epider@@ mo@@ ist lysis .
P-3629    -0.5730 -1.2475 -1.7452 -2.2903 -2.1050 -1.4641 -0.736
7 -0.1582 -3.7445 -0.0702 -1.6936 -2.5032 -0.0915 -0.1726
S-916    In@@ dic@@ aciones médicas y quirúrgicas de los potenci
ales ev@@ oc@@ ados .
T-916    Medical and surgical indications of ev@@ oked potential
s .
H-916    -1.0327579975128174    medical and surgical approach of
the potential ev@@ oked potentials .
D-916    -1.0327579975128174    medical and surgical approach of
the potential ev@@ oked potentials .
P-916    -2.9271 -0.3403 -0.2766 -3.6245 -0.2752 -2.0038 -0.3977
-0.1251 -1.9874 -0.1362 -0.1323 -0.1669
S-3544    Este estudio tiene por objetivo presentar una revisión
sobre el asun@@ to .
T-3544    This study aimed at revie@@ wing this subject .
H-3544    -0.6708385348320007    This study aims to present a re
view on the matter of the matter .
D-3544    -0.6708385348320007    This study aims to present a re
view on the matter of the matter .
P-3544    -0.3408 -0.2574 -2.1074 -0.2635 -1.2550 -0.7099 -0.019
0 -0.8654 -0.2694 -0.4990 -0.3849 -1.1015 -1.6698 -0.1248 -0.194
9
S-4839    Los valores plasmáticos del sodio y del potasio no se
 alter@@ aron .
T-4839    Plas@@ ma sodium and potassium were not changed .
H-4839    -0.45692428946495056    Plas@@ ma and potassium values
were not altered .
D-4839    -0.45692428946495056    Plas@@ ma and potassium values
were not altered .
P-4839    -2.8977 -0.1235 -0.3627 -0.0258 -0.2182 -0.1091 -0.170
9 -0.4268 -0.0846 -0.1500
```

Trained on: Biological, Tested on: Biological (es-en)

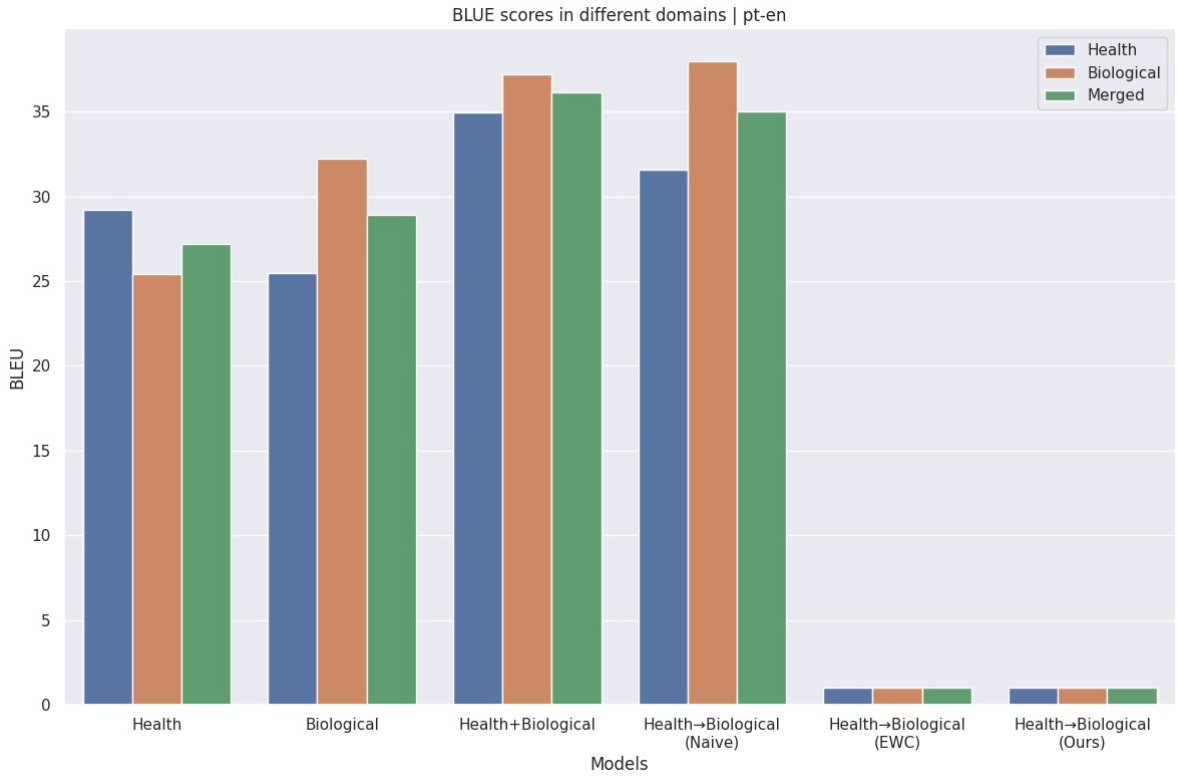
```
S-1782    Los experimentos fueron llevados a cabo usando el músc
ulo aislado de ambas especies .
T-1782    Both species experiments were carried out using the is
olated muscles .
H-1782    -0.43383994698524475    The experiments were carried o
ut using the isolated muscle of both species .
D-1782    -0.43383994698524475    The experiments were carried o
ut using the isolated muscle of both species .
P-1782    -0.5694 -0.3670 -0.1327 -0.4449 -0.0903 -0.2822 -0.154
5 -2.4607 -0.0882 -0.4919 -0.5733 -0.1386 -0.1103 -0.1698
S-1071    Los invertebrados que habitan el suelo son actores imp
ortantes en los procesos edáficos .
T-1071    Soil @-@ inhabiting invertebrates play an important ro
le in soil processes .
H-1071    -0.7320635914802551    The invertebrates that inhabit
 the soil are important actors in the edaphic processes .
D-1071    -0.7320635914802551    The invertebrates that inhabit
 the soil are important actors in the edaphic processes .
P-1071    -2.1574 -0.2994 -2.4002 -0.5484 -0.6102 -0.0960 -0.056
1 -0.2094 -0.1566 -0.5941 -0.8326 -2.4657 -0.3104 -0.0811 -0.163
4
S-768    Lo último podría limitar el restab@@ lecimiento de pobl
aciones de especies herb@@ ívoras .
T-768    The later prevent the re @-@ establishment of herbivore
species populations .
H-768    -0.8204675316810608    Finally , the latter could limit
the recovery of herbivorous species .
D-768    -0.8204675316810608    Finally , the latter could limit
the recovery of herbivorous species .
P-768    -1.6022 -0.3679 -2.0152 -1.0818 -1.2024 -0.3188 -0.1867
-1.2975 -0.1059 -0.4300 -1.1682 -0.6948 -0.1946
S-296    M@@ is resultados los compar@@ o con la información exi
stente para el planeta .
T-296    I compared my results against information for the entir
e Plan@@ et .
H-296    -0.7367602586746216    The results are compared with th
e existing information for the planet .
D-296    -0.7367602586746216    The results are compared with th
e existing information for the planet .
P-296    -4.1590 -0.0544 -1.3278 -0.6892 -0.4899 -0.4855 -0.9092
-0.0514 -0.6784 -0.1990 -0.2870 -0.0653 -0.1818
```

Constrained

Spanish-English:



Portuguese-English:



Comments:

- The results from es-en and pt-en are pretty similar across all tested domains. Therefore, the comments below can apply to both of them.

- **Health model:** This model is trained on the health domain exclusively. This time, with a balanced dataset the health model performs as expected. Very well in its domain, reasonably well on the merged domain, and poorly on the biological domain.
- **Biological model:** This model is trained on the biological domain exclusively. It performs well in its domain, a little worse in the merged domain and finally, the health domain is the one with the worst results as expected.
- **Health+Biology:** This model is trained on the health and biological domain at once. First, we observed a consistent increase in performance for all domains. This is expected, since as usual: "the more data, the better". However, this additional (balanced) data makes the model more robusts to unknown domains and the gap in terms of performance across domains is slightly decreased.
- **Health->Biology:** This model is trained sequentially. First on the health domain, and then fine-tuned on the biological domain. As a result of this sequential training, the performance in the fine-tuned domain increased but at expense of worsen its performance on a previously learned domain (health). This phenomenon is known as "catastrophic forgetting" and it is the one we will try to mitigate.

Table:

In [1]:
```
import pandas as pd
df = pd.read_csv("../data/constrained/test_data.csv")
print(df)
```

	Unnamed: 0	Model	Test domain	lang	BLEU
0	0	Health	Health	es-en	29.17
1	1	Health	Biological	es-en	23.28
2	2	Health	Merged	es-en	26.25
3	3	Biological	Health	es-en	26.65
4	4	Biological	Biological	es-en	32.98
5	5	Biological	Merged	es-en	29.67
6	6	Health+Biological	Health	es-en	35.86
7	7	Health+Biological	Biological	es-en	38.49
8	8	Health+Biological	Merged	es-en	37.26
9	9	Health→Biological\n(Naive)	Health	es-en	31.03
10	10	Health→Biological\n(Naive)	Biological	es-en	39.30
11	11	Health→Biological\n(Naive)	Merged	es-en	35.49
12	12	Health→Biological\n(EWC)	Health	es-en	1.00
13	13	Health→Biological\n(EWC)	Biological	es-en	1.00
14	14	Health→Biological\n(EWC)	Merged	es-en	1.00
15	15	Health→Biological\n(Ours)	Health	es-en	1.00
16	16	Health→Biological\n(Ours)	Biological	es-en	1.00
17	17	Health→Biological\n(Ours)	Merged	es-en	1.00
18	18	Health	Health	pt-en	29.19
19	19	Health	Biological	pt-en	25.42
20	20	Health	Merged	pt-en	27.18
21	21	Biological	Health	pt-en	25.48
22	22	Biological	Biological	pt-en	32.24
23	23	Biological	Merged	pt-en	28.92
24	24	Health+Biological	Health	pt-en	34.98
25	25	Health+Biological	Biological	pt-en	37.21
26	26	Health+Biological	Merged	pt-en	36.15
27	27	Health→Biological\n(Naive)	Health	pt-en	31.57
28	28	Health→Biological\n(Naive)	Biological	pt-en	37.97
29	29	Health→Biological\n(Naive)	Merged	pt-en	35.04
30	30	Health→Biological\n(EWC)	Health	pt-en	1.00
31	31	Health→Biological\n(EWC)	Biological	pt-en	1.00
32	32	Health→Biological\n(EWC)	Merged	pt-en	1.00
33	33	Health→Biological\n(Ours)	Health	pt-en	1.00
34	34	Health→Biological\n(Ours)	Biological	pt-en	1.00
35	35	Health→Biological\n(Ours)	Merged	pt-en	1.00

Translated examples

> **Leyend:** Source (S), Target (T), Hypothesis (H), Detokenized (D), Probabilities (P)
>
> **Translations not cherry picked**

Trained on: Health, Tested on: Health (es-en)

```
S-4277    Todas , menos una , evolucionaron favorablemente indep
endientemente del tratamiento utilizado .
T-4277    All of them , but one , evolved successfully .
H-4277    -0.9211965203285217    All , at least one evolved favo
r@@ ably independently of the treatment .
D-4277    -0.9211965203285217    All , at least one evolved favo
r@@ ably independently of the treatment .
P-4277    -0.1193 -3.0150 -1.8813 -0.1571 -0.3179 -1.0666 -1.348
7 -0.0061 -3.1623 -0.1907 -0.5136 -0.3611 -0.6346 -0.1223
S-342    Dis@@ función miccional post @-@ bra@@ quit@@ erapia en
el cáncer prostático
T-342    Objectives :
H-342    -0.8363382816314697    Post @-@ brachy@@ therapy functi
on in prostate cancer
D-342    -0.8363382816314697    Post @-@ brachy@@ therapy functi
on in prostate cancer
P-342    -4.2303 -0.1559 -2.0284 -0.5584 -0.1517 -0.1784 -0.1890
-0.0128 -0.0221
S-244    La tomografía computador@@ izada mostró imágenes con as
pecto de embolia pulmonar .
T-244    CT images suggested pulmonary embolism .
H-244    -0.3998290002346039    The computed tomography showed i
mages with pulmonary embolism .
D-244    -0.3998290002346039    The computed tomography showed i
mages with pulmonary embolism .
P-244    -0.9179 -1.2668 -0.1386 -0.1885 -0.3406 -0.0906 -0.8390
-0.0500 -0.0671 -0.0993
S-1075    Consumo de alcohol en la consulta de una sala de emerg
encia .
T-1075    Alcohol consumption in the emergency room .
H-1075    -0.4401593506336212    Alcohol consumption in the emer
gency room visits .
D-1075    -0.4401593506336212    Alcohol consumption in the emer
gency room visits .
P-1075    -0.0470 -0.1664 -0.8801 -0.3831 -1.0671 -0.0174 -1.227
7 -0.0822 -0.0905
```

Trained on: Health, Tested on: Biological (es-en)

```
-2781    Las dietas de har@@ inas mezcl@@ adas mostraron los mej
ores resultados en longitud .
T-2781    The mixed meal @-@ based diets showed better productio
n results .
H-2781    -0.6066553592681885    The diets of mixed diets showed
the best results in length .
D-2781    -0.6066553592681885    The diets of mixed diets showed
the best results in length .
P-2781    -1.9708 -0.1460 -0.5205 -1.3207 -1.2588 -0.7932 -0.682
1 -0.2032 -0.1533 -0.4996 -0.0840 -0.1498 -0.1047
S-1315    El intervalo de tiempo promedio entre aportes de presa
fue de 7@@ 98 s.
T-1315    The mean time interval between pre@@ y deliveries was
 7@@ 98 s .
H-1315    -0.8942710757255554    The average time interval betwe
en contributions was 7@@ 98 .
D-1315    -0.8942710757255554    The average time interval betwe
en contributions was 7@@ 98 .
P-1315    -0.3406 -0.8814 -0.3741 -0.8798 -0.3730 -3.1739 -1.802
5 -0.1505 -0.0582 -1.7010 -0.1020
S-1161    El factor m@@ ór@@ fico para la especie fue de 0.7@@ 6
0@@ 41 .
T-1161    The form @-@ factor for the species was of 0.7@@ 60@@
 41 .
H-1161    -0.6691439747810364    The morphic factor for the spec
ies of the species was 0.7@@ 41 .
D-1161    -0.6691439747810364    The morphic factor for the spec
ies of the species was 0.7@@ 41 .
P-1161    -0.5770 -1.9056 -0.9096 -0.2214 -0.1486 -0.8225 -0.840
1 -0.7356 -1.0623 -0.0298 -1.6293 -0.2767 -0.0661 -0.1434
S-1051    El análisis de los resultados se hizo mediante una pru
eba t pare@@ ada .
T-1051    The analysis of results was done by a paired t @-@ tes
t .
H-1051    -0.5913573503494263    The analysis of the results was
performed using a paired t @-@ test .
D-1051    -0.5913573503494263    The analysis of the results was
performed using a paired t @-@ test .
P-1051    -0.6477 -1.3748 -0.1423 -0.5766 -0.0618 -0.5341 -2.149
1 -0.6055 -0.5726 -0.0360 -0.4455 -1.4269 -0.0169 -0.1689 -0.111
7
```

Trained on: Biological, Tested on: Health (es-en)

```
S-3484    Un paciente HIV positivo fal@@ leci@@ ó debido a epide
r@@ mo@@ lisis .
T-3484    One HIV positive patient died of epider@@ mol@@ y@@ si
s .
H-3484    -0.6549513339996338    HIV positive patient died due t
o epider@@ mo@@ lysis .
D-3484    -0.6549513339996338    HIV positive patient died due t
o epider@@ mo@@ lysis .
P-3484    -0.2747 -0.6397 -1.1379 -0.5288 -0.4474 -0.1497 -3.030
8 -0.1033 -0.6284 -0.1305 -0.1334
S-3556    Este estudio tiene por objetivo presentar una revisión
sobre el asun@@ to .
T-3556    This study aimed at revie@@ wing this subject .
H-3556    -0.5915473699569702    This study aims to present a re
view on the matter .
D-3556    -0.5915473699569702    This study aims to present a re
view on the matter .
P-3556    -0.7122 -0.3697 -0.6416 -0.6107 -0.5905 -0.1485 -0.198
4 -1.1848 -0.4627 -1.4014 -0.6620 -0.1160
S-4888    Los valores plasmáticos del sodio y del potasio no se
 alter@@ aron .
T-4888    Plas@@ ma sodium and potassium were not changed .
H-4888    -0.38854897022247314    Plas@@ matic values of sodium
 and potassium were not altered .
D-4888    -0.38854897022247314    Plas@@ matic values of sodium
 and potassium were not altered .
P-4888    -0.9361 -1.3466 -0.8944 -0.2965 -0.0703 -0.1775 -0.047
4 -0.1599 -0.3272 -0.1757 -0.1317 -0.0993
S-11    In@@ dic@@ aciones médicas y quirúrgicas de los potencia
les ev@@ oc@@ ados .
T-11    Medical and surgical indications of ev@@ o@@ ked potenti
als .
H-11    -1.183539867401123    Dental and surgical indicators of
 the potential ev@@ o@@ king .
D-11    -1.183539867401123    Dental and surgical indicators of
 the potential ev@@ o@@ king .
P-11    -3.3431 -0.3797 -0.0375 -2.9420 -0.2289 -2.6478 -0.4204
 -0.1158 -1.3230 -1.9889 -0.6540 -0.1213
```

Trained on: Biological, Tested on: Biological (es-en)

```
S-2258    Isla Fuer@@ te es un complejo arrecifal ho@@ loc@@ éni
co parcialmente emerg@@ ido .
T-2258    Isla Fuer@@ te is a partially emerged Holoc@@ enic ree
f @-@ complex .
H-2258    -0.5470016598701477    Isla Fuer@@ te is a partial ree
f complex that emerged .
D-2258    -0.5470016598701477    Isla Fuer@@ te is a partial ree
f complex that emerged .
P-2258    -0.2208 -0.1517 -0.0103 -0.1791 -0.5054 -0.1410 -0.537
2 -0.0172 -3.8636 -0.0996 -0.7024 -0.1356
S-513    Se registraron seis especies en la comunidad de Cerro T
e@@ <unk> in@@ tle .
T-513    Six species of carnivorous were recorded .
H-513    -0.965829074382782    We recorded six species in the Ce
rro El Density community .
D-513    -0.965829074382782    We recorded six species in the Ce
rro El Density community .
P-513    -1.2235 -0.4278 -0.5862 -0.2635 -0.7154 -1.2146 -0.1618
-1.6539 -4.5431 -0.4244 -0.2450 -0.1308
S-1573    Me@@ todo para mont@@ aje permanente de huevos de helm
intos entero@@ para@@ sitos .
T-1573    Def@@ ini@@ tive preservation of helmin@@ thic eggs .
H-1573    -1.3171203136444092    Me@@ an@@ while for permanent c
olo@@ urs of entero@@ parasites were observed .
D-1573    -1.3171203136444092    Me@@ an@@ while for permanent c
olo@@ urs of entero@@ parasites were observed .
P-1573    -1.7694 -0.6177 -0.4050 -1.4824 -1.9134 -3.8539 -0.674
1 -0.7958 -1.4900 -0.9137 -2.1089 -1.6854 -0.5953 -0.1348
S-1520    Sin embargo , fal@@ leci@@ ó debido a neum@@ onia intr
a@@ hospital@@ aria .
T-1520    However , he died due to intra@@ hospital pneumonia .
H-1520    -1.145051121711731    However , died due to intra@@ ho
spital stay .
D-1520    -1.145051121711731    However , died due to intra@@ ho
spital stay .
P-1520    -0.5008 -0.1761 -3.0102 -0.5419 -0.1582 -3.2985 -0.636
3 -2.8702 -0.1446 -0.1138
```

In [ ]: