
Supplemental material: Sharing Knowledge for Meta-learning with Feature Descriptions

Anonymous Author(s)

Affiliation

Address

email

1 Experimental results

2 Figure 1 shows the test mean squared error with different rates of observed feature descriptions by
3 the proposed method. The error decreased as the observed rate increased especially in e-Stat data.

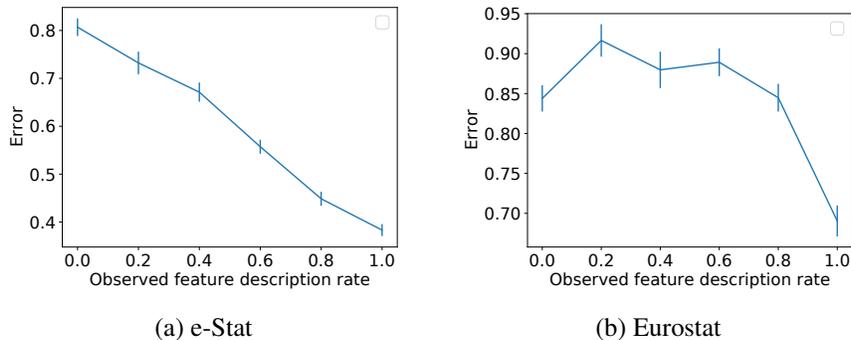


Figure 1: Average test mean squared errors with different rates of observed feature descriptions. The bar shows the standard error.

4 Table 1 shows the test mean squared error by the proposed method using word2vec [1] for sentence
5 embeddings instead of BERT. We used <https://github.com/Kyubyong/wordvectors> for e-Stat,
6 and <https://code.google.com/archive/p/word2vec> for Eurostat, and the sentence embed-
7 ding was obtained by the average of the word embeddings in the sentence. The error with word2vec
8 was higher than that with BERT. This result indicates that the meta-learning performance depends on
9 the performance of the sentence encoder.

Table 1: Average test mean squared error and its standard error by the proposed method with word2vec instead of BERT.

(a) e-Stat			
#Support	5	10	20
Ours+word2vec	0.500 ± 0.016	0.442 ± 0.012	0.390 ± 0.013
(b) Eurostat			
#Support	5	10	20
Ours+word2vec	0.865 ± 0.021	0.800 ± 0.016	0.680 ± 0.017

10 **2 Limitations**

11 The proposed method requires feature descriptions for meta-training datasets. When there are not
12 enough meta-training datasets, the performance can be comparable with methods that are trained
13 using support instances for each task. When meta-training and meta-test tasks are significantly
14 different, there is a risk of degrading the performance. The evaluation of the proposed method
15 with classification tasks and numerical feature values are not performed since the data used in our
16 experiments were regression tasks with categorical feature values.

17 **3 Potential negative societal impacts**

18 Since the proposed method uses various datasets for meta-training, there is a potential risk that users
19 might include biased datasets without careful thought, which might result in biased predictions. We
20 encourage research to automatically detect biased datasets. Although the proposed method improves
21 performance, the prediction might be not always correct. We encourage research to estimate the
22 uncertainty of the prediction by meta-learning from datasets with feature descriptions.

23 **References**

24 [1] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in
25 vector space. *arXiv preprint arXiv:1301.3781*, 2013.