

Supplementary materials

Minsu Kim, Sangseon Lee, Sangsoo Lim, Doh Young Lee, and Sun Kim*

January 16, 2021

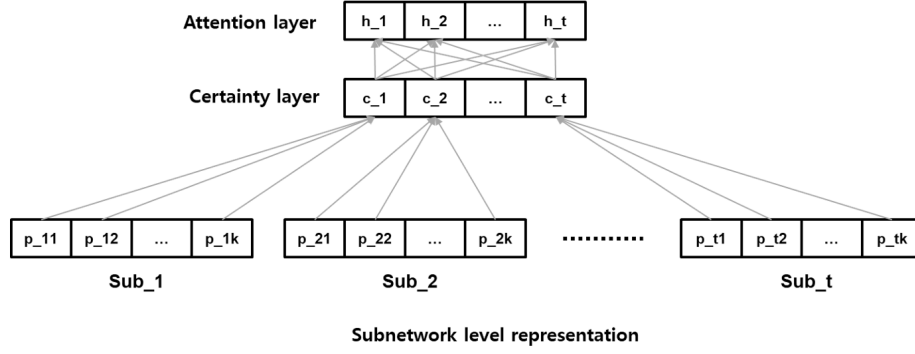


Fig. 1: Attention layer structure

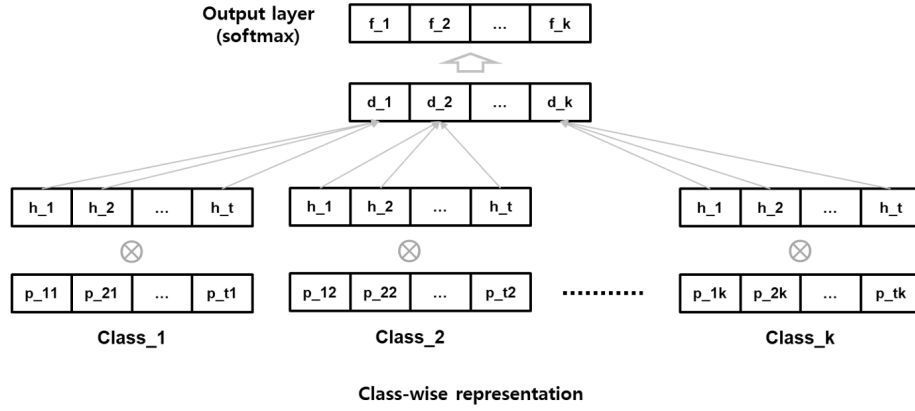


Fig. 2: Decision making process

Table 1: Description of hyperparameters used for the proposed method evaluation

Parameter	Options
Optimization ¹	SGD, RMSprop, AdaGrad, AdaDelta, Adam, AdaMax, and Nadam
Loss ²	Mean squared error, Mean absolute error, Mean absolute percentage error, Hinge, Mean squared logarithmic error, Logcosh, Squared hinge, Categorical hinge, Categorical crossentropy, Poisson, Sparse categorical crossentropy, Binary crossentropy, Cosine proximity, and Kullback-Leibler divergence
Dropout rate ³	Not-reduced, 1, 2, ..., 100
Split ratio ⁴	100, 1000, and 10000
Feature selection ⁵	True and False
Total	4,900 combinations

1 Type of algorithm for optimizing attention layer

2 Type of loss function for optimizing attention layer

3 Dropout rate for each back propagation for optimizing attention layer

4 Validation split ratio of training data for early stopping

5 Whether using feature selection or not

Table 2: Description of parameters used for the baseline performance evaluation

Parameter	Options
Algorithm ¹	LR, KNN, RF, and SVM
Kernel ²	linear, polynomial, RBF, and sigmoid
Size of dimension ³	Not-reduced, 1, 2, ..., 100
Number of estimator ⁴	100, 1000, and 10000
TPM logarization ⁵	True and False
Feature selection ⁶	True and False
Total	12,024 combinations

1 Type of algorithm used for constructing prediction model

2 Type of kernel used for Kernel Principal Component Analysis (KPCA)

3 Size of dimension used for KPCA

4 Number of estimator in case of using RF algorithm

5 Whether using logarization or not

6 Whether using feature selection or not