Supplementary material of "TPFN: Apply Outer Product along Time to Multimodal Sentiment Analysis Fusion on Imperfect Data"

Binghua Li[1,2,⋆,[0000−0002−2595−4762]], Chao Li[1,⋆], Feng Duan[2,⋆⋆], Ning Zheng[1,[0000−0002−2282−1200]], and Qibin Zhao[1,[0000−0002−4442−3182]]

[1] RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan
{chao.li, ning.zheng, qibin.zhao}@riken.jp
[2] College of Artificial Intelligence, Nankai University, Tianjin, China
nkvhua@outlook.com, duanf@nankai.edu.cn

In this part, we introduce more details in our paper, and show more results on additional experiments. We 1). introduce the details of the classification task and the parameters searching range in our experiments, 2). give and analyse the detailed results on CMU-MOSEI, 3). show how the sliding stride effects the performance and 4). show how our approach performs on other metrics.

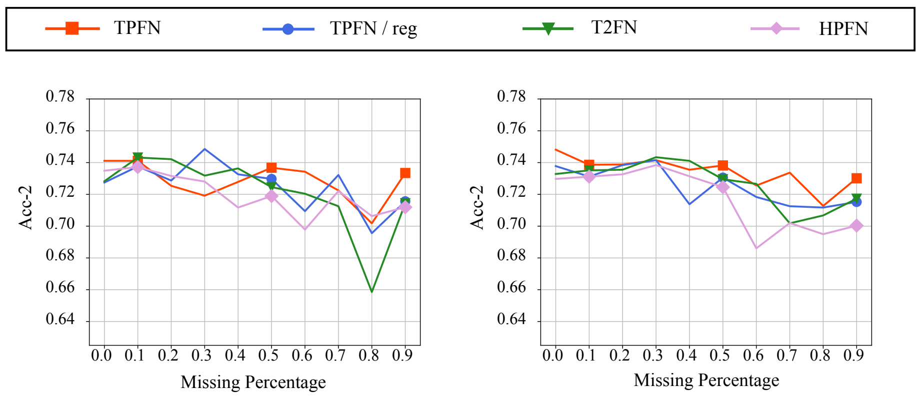
1 Experiment settings

Below we introduce the details of our experiments in the paper. Similar to the work in [2], we try to learn a mapping from multimodal data to regression value ranging from $[-3, 3]$. Loss function is calculated by Mean Absolute Error (MAE) for all models, and additional norm-induced regularization is added in T2FN and our TPFN for low-rank constrain. The binary accuracy is obtained with the decision boundary $y = 0$. We then select hyper-parameters that can achieve the maximum mean binary classification accuracy. The hyper-parameter search range is listed in Table 1. To be fair, all data under each p is the same for all methods. To be reproducible, results on TPFN and TPFN/reg are obtained with random seed 0.

Table 1. Hyper-parameter search range statistics in our experiments

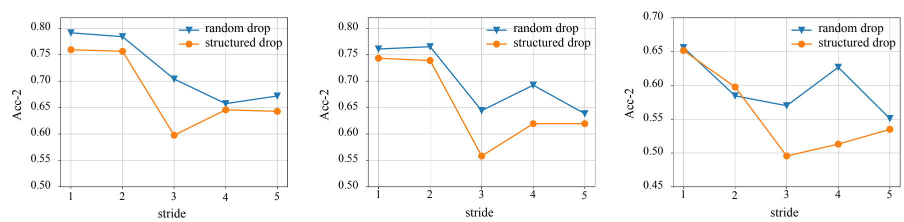
Hyper-parameter Search range			
Hidden units for acoustic	8, 16	Hidden units for visual	4, 8, 16
Hidden units for language	64, 128	Dropout probability	0.0, 0.1, 0.3, 0.5
Learning rate	0.0003, 0.001, 0.003	Rank	4, 8, 12, 16, 24, 32
Batch size	4, 16, 64, 128	Decay on learning rate	0.0, 0.01, 0.001

⋆ Equal contribution.
⋆⋆ Feng Duan is the main corresponding author.



(a) Random drop (b) Structured drop

Fig. 1. Results of CMU-MOSEI on TPFN, TPFN/reg, T2FN and HPFN.



(a) Low (b) Medium (c) High

Fig. 2. Comparison on CMU-MOSI as stride varies in the random drop task and the structured drop task

2 Details on CMU-MOSEI

We give the results on CMU-MOSEI under all p in Fig. 1. Unlike the remarkable results on CMU-MOSI, our TPFN achieve an averagely better results on CMU-MOSEI. It is because the CMU-MOSEI is such a big dataset that numerous entries are still retained even if $p = 0.9$, and therefore the results are relatively stable for all models we compare with.

3 Effect of stride

To show the effect on performance of the sliding stride, we keep all parameters unchanged except for the stride. We change the it from $\{1, 2, 3, 4, 5\}$, and results are shown in Fig. 2.

As shown in Fig. 2, the performance declines as stride grows. It is expected because a larger stride means we sample from M_t and apply less data from the series to calculate M.

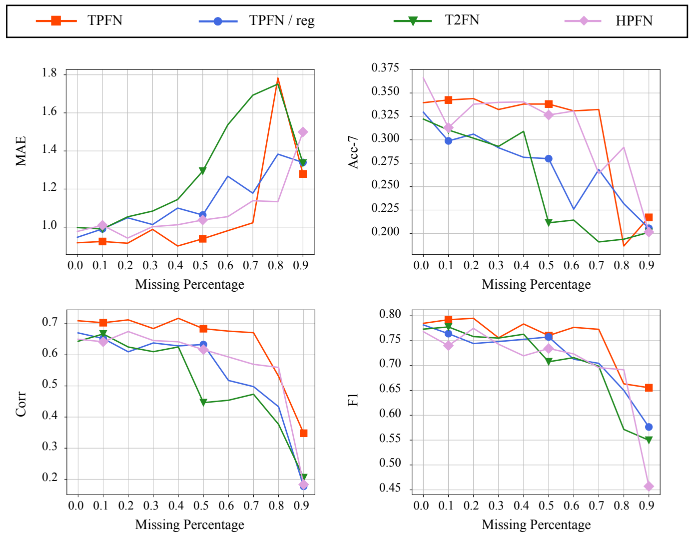


Fig. 3. Metrics of CMU-MOSI on structured drop task. Results of (a) Mean absolute error, (b) accuracy of multiclass classification, (c) Pearson correlation and (d) F1 score are reported. The marker points corresponds to the results when $p = 0.1, 0.5, 0.9$

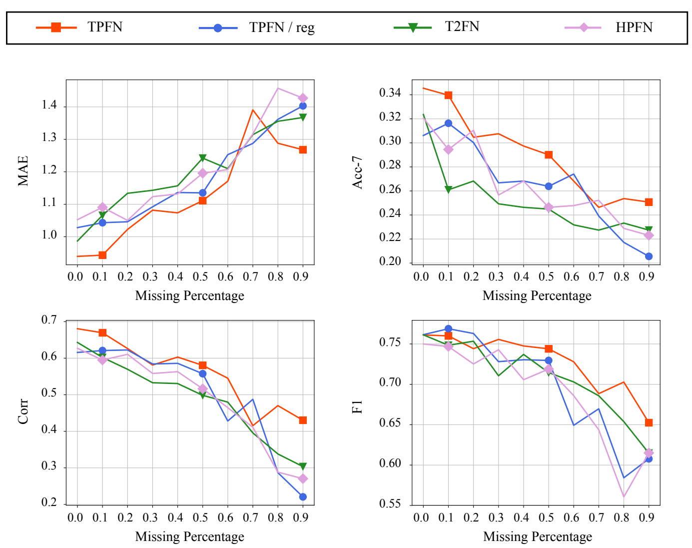


Fig. 4. Metrics of CMU-MOSI on structured drop task. Results of (a) Mean absolute error, (b) accuracy of multiclass classification, (c) Pearson correlation and (d) F1 score are reported. The marker points corresponds to the results when $p = 0.1, 0.5, 0.9$

4 Other metrics on CMU-MOSI.

The same as previous works [1, 2], we will give the results on other metrics on CMU-MOSI in this part. We report F1 score, MAE, Pearson correlation (Corr) and the accuracy of multiclass classification (ACC-7, 7 is the number of classes) in Fig. 3 and 4. Higher values denote better performance for all metrics except for MAE. Overall, our TPFN also fares best in all these metrics. More supplemental materials and codes are available in the webpage https://qibinzhao.github.io.

References

1. Hou, M., Tang, J., Zhang, J., Kong, W., Zhao, Q.: Deep multimodal multilinear fusion with high-order polynomial pooling. In: Proc. NeurIPS (2019)
2. Liu, Z., Shen, Y., Lakshminarasimhan, V.B., Liang, P.P., Zadeh, A., Morency, L.P.: Efficient low-rank multimodal fusion with modality-specific factors. arXiv preprint arXiv:1806.00064 (2018)