Multi-Memory enhanced Separation Network for Indoor Temperature Prediction

Zhewen Duan[1,2,3], Xiuwen Yi[2,3] ⋆, Peng Li[2,3], Dekang Qi[2,3,4], Yexin Li[2,3], Haoran Xu[2,3], Yanyong Huang[5], Junbo Zhang[2,3], and Yu Zheng[1,2,3]

[1] Xidian University, Xi'an
zwdanzer@gmail.com
[2] JD Intelligent Cities Research
[3] JD iCity, JD Technology, Beijing, China
xiuwenyi@foxmail.com, {qidekang,lipeng367,liyexin}@jd.com, ryanxhr@gmail.com, {msjunbozhang, msyuzheng}@outlook.com
[4] Southwest Jiaotong University, Chengdu
[5] Southwestern University of Finance and Economics, Chengdu
huangyy@swufe.edu.cn

Abstract. Indoor temperature prediction is vital to predictive control on district heating systems. Due to the data collection in practice, there always exist residential areas with limited historical data. Transferring the knowledge from residential areas with sufficient data is of great help to address the data scarcity problem. However, it is still challenging as the data distribution shifts among residential areas and shifts over time. In this paper, we proposed a **M**ulti-**M**emory **e**nhanced **S**eparation **N**etwork (**MMeSN**) to predict indoor temperature for residential areas with limited data. MMeSN is a parameter-based multi-source transfer learning method, mainly consisting of two components: *Source Knowledge Memorization* and *Memory-enhenced Aggregation*. Specifically, the former component jointly decouples the domain-independent & domain-specific information which separately memorize the specific historical patterns for each source. The latter component memorizes the historical relationships between the target and multiple sources and further aggregates the domain-specific & domain-independent information. We conduct extensive experiments on a real-world dataset, and the results demonstrate the advantages of our approach.

Keywords: Time series prediction · Transfer learning · Deep Learning · Indoor Temperature · Urban Computing.

1 Introduction

District heating system is widely used during winter, supplying heat to residents for keeping the house warm. For monitoring the performance of heating services, heat companies have deployed some temperature sensors in the house,

⋆ Xiuwen Yi is the corresponding author.

generating real-time indoor temperature records. Accurate predicting future indoor temperature is important for the predictive control of heating systems. By controlling the heat supply intelligently, the indoor temperature can be sustained at a comfortable level while reducing energy consumption.

However, considering the real-world data collection mechanism, over-fitting problem often happened with traditional time-series prediction approaches when there is not sufficient training data. For example, we can only get limited amount of data for the residential area with newly-deployed temperature sensors due to the cold-start problem.

One feasible idea is to transfer knowledge from a data-rich domain to the data-poor domain[10]. Besides, multi-source transfer learning has become a hotspot since more source information will contribute to a robust model. In this paper, we propose a multi-memory enhanced separation network, named MMeSN, to predict residential indoor temperature with limited data. MMeSN is a multi-source parameter-based transfer learning method, which mainly contains source knowledge memorization and memory-enhanced aggregation. The former memorizes useful knowledge of multiple sources, and the latter adapts the source knowledge to the target domain. Our main contributions are summarized as follows:

- *Multiple memory modules.* We design multiple source memories to learn the specific patterns of each source and design a target-source memory to capture the correlations between target and source domains, which is robust for historical knowledge memorization with better transfer generalization.
- *Joint decomposition architecture.* We design the joint decomposition architecture for decoupling independent & specific information among all domains, which helps distill individual historical knowledge and alleviates the complexity of knowledge transfer.
- *Real evaluation.* We conduct extensive experiments on a real-world dataset with four residential areas and the results demonstrate the advantages of our approach over several state-of-the-art baselines.

2 Overview

2.1 Problem Definition

Formulation of prediction task. For a residential area, we specify $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^m) \in R^m$ as m different sensor readings at time interval i, including *indoor temperature, heating temperature, outdoor temperature, outdoor humidity, wind speed, and wind direction.* Besides, we specify y_i as the indoor temperature. The prediction task can be defined as predicting the next indoor temperature y_{L+1}, given the historical observations $\mathcal{X} = \{\mathbf{x}_i | i = 1, ..., L\}$.

Formulation of transferring task. Given n source residential areas $\{\mathcal{X}^s | s = 1, 2, ..., n\}$ with sufficient historical data and one target residential area $\mathcal{X}^t$ with limited data. The transferring task can be defined as predicting the next indoor temperature of the target residential area $y^t_{L^t+1}$, where L^t is the length of historical records of the target area.

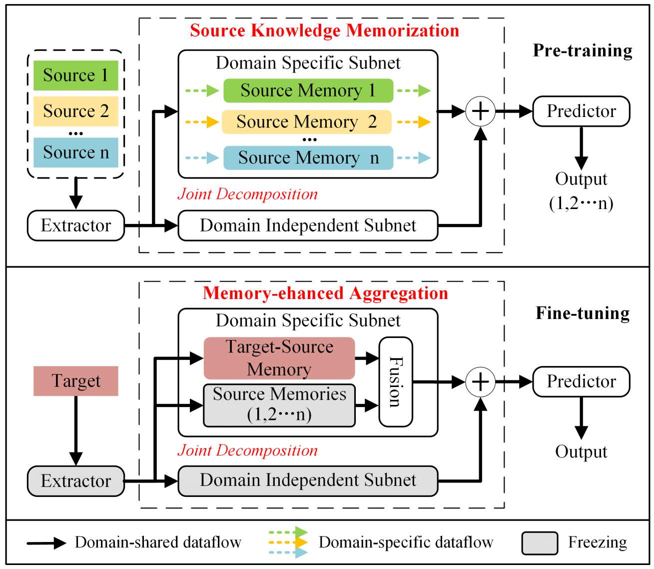


Fig. 1. Framework of our proposed MMeSN.

3 Methodology

Figure 1 illustrates the framework of MMeSN method, trained by two stages: 1) *source data pre-training*, which jointly trains the network by multiple source data to memorize the source knowledge; 2) *target data fine-tuning*, which re-trains the network with only target data to transfer the knowledge from multiple sources.

In the pre-training stage, we feed all source data into the network separately to jointly train the parameters. Firstly, the extractor learns the feature representation from the original data. Then, in source knowledge memorization component, we propose a joint decomposition architecture which contains a domain independent subnet and a domain specific subnet with multiple memories to learn the shared and specific historical information. Finally, the concatenation of these two outputs is fed into the predictor to forecast the next value.

In the fine-tuning stage, the network parameters are re-trained only using the target data. To utilize the source knowledge, we freeze the parameters of the extractor, domain independent subnet, and multiple source memories. Then, in memory-enhanced aggregation component, we reconstruct the domain specific subnet by target-source memory and fusion module for target specific knowledge learning. After that, the outputs are fed into a fusion module for knowledge aggregation. Likewise, the shared and specific knowledge are aggregated together to make the final prediction by the predictor.

MMeSN mainly consists of four components: extractor, source knowledge memorization, memory-enhanced aggregation, and predictor. Here, the extractor extracts the feature representations from the input data, capturing the interactions between multiple factors. We implement it with a feature embedding layer followed by a flatten layer and two fully connected layers with the activation function ReLU, which is denoted as $f_E(\cdot)$ for short. While the predictor receives the concatenation of the output from domain specific and independent subnet, then generates the prediction utilizing a linear transformation with Sigmoid function. In the following, we detail the key components: source knowledge memorization and knowledge-enhanced aggregation.

3.1 Source Knowledge Memorization

To learn the historical patterns of different residential areas, we design the source knowledge memorization component, consisting of a joint decomposition architecture and multiple source memories. The former architecture jointly decouples the domain-independent & domain-specific information for all sources, and the latter separately memorize the specific historical patterns for each source by the memory network.

Though the temperature changes differently among residential areas, it still obeys the same underlying heat exchange rules in the real-world scenario. For example, the indoor temperature rises with the increment of heating temperature and outdoor temperature. Motivated by such fact, we design the joint decomposition architecture, which consists of the domain independent subnet (DI) and domain specific subnet (DS) with n branches for independent and specific knowledge learning, thus alleviating the complexity of knowledge transfer.

In our task, given the source data $\mathbf{X}^s \in R^{K \times m}$ with a fixed window size K, we specify the source input as the feature representations $f_E(\mathbf{X}^s)$ learned by the extractor. Then the extracted domain independent knowledge by DI and domain specific knowledge by DS-s are further concatenated as the output of source knowledge memorization. We implement the domain independent subnet with two fully connected layers followed with ReLU and batch normalization. The output of this module is denoted as $\mathbf{z}^s_{DI} = f_{DI}(f_E(\mathbf{X}^s))$.

As for the domain specific subnet, we design n branches for n different sources. Considering the distribution of each source data changes over time, we implement each branch with a memory network trained by the corresponding source data to memorize the historical information. Memory network is effective to model the sequential data, storing the long-term dependencies. The structure of memory network is illustrated in Figure 2 (a). We construct a memory matrix $\mathbf{M}^s \in R^{V \times d}$ to store the historical information for source s, which contains V memory representations with dimension d. Each row of the memory matrix can be regarded as one distribution pattern of historical data. For each input source data $\mathbf{X}^s$, we get the key vector $\mathbf{k}^s = f_E(\mathbf{X}^s) \in R^d$, which is then utilized to calculate the similarity score p^s_j with each slice of memory representation $\mathbf{M}^s_j$. The output of domain specific subnet $\mathbf{z}^s_{DS}$ is calculated by the weighted sum of memory slices shown in equation 1.

$$p^s_j = \frac{exp(<\mathbf{k}^s, \mathbf{M}^s_j>)}{\sum_{i=1}^{V} exp(<\mathbf{k}^s, \mathbf{M}^s_i>)}, \ \mathbf{z}^s_{DS} = \sum_{j=1}^{V} p^s_j * \mathbf{M}^s_j \tag{1}$$

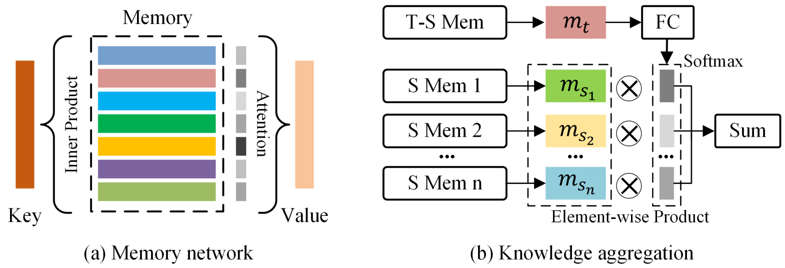


Fig. 2. The process of knowledge aggregation

When retrieving the historical patterns by the memory network, it generates the most relevant memory representation by merging all memory slices with the similarity scores. Compared with the fully connected network learning the mapping function weights, the memory network is more general and robust in transferring scenarios as it learns the historical patterns stored in memory slices.

3.2 Memory-enhanced Aggregation

After the source data pre-training stage, the shared and specific knowledge from multiple sources is memorized within the DI and DS modules. For better adapting the source knowledge to the target domain, in memory-enhanced aggregation, we freeze the parameters of domain independent subnet and multiple source memories. Then, we reconstruct the domain specific subnet, appending the target-source memory and fusion module to help transfer the source specific knowledge to the target domain.

Similar to the procedure of source knowledge memorization, in this component, the shared knowledge behind all residential areas is learned by the domain independent subnet with the output $\mathbf{z}_{DI}^t = f_{DI}(f_E(\mathbf{X}^t))$. While the specific knowledge of the target domain can be viewed as the aggregation of all source knowledge, whose detail is shown in figure 2 (b). Given the target input data $\mathbf{X}^t \in R^{K\times m}$, we get the key vector from the extractor $\mathbf{k}^t = f_E(\mathbf{X}^t) \in R^d$ with window size K, which is then fed into the domain specific subnet to generate n source memory representations $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$. Besides, we utilize the target-source memory to help incorporate the source knowledge which shares the same structure of the source memory network with different parameters. Here, the key vector $\mathbf{k}^t$ is also fed into the target-source memory to generate the correlation representation $\mathbf{v}^t$. Then, a linear transformation is adopted to output the normalized similarity scores c with a Softmax function. After that, the specific knowledge of the target domain $\mathbf{z}_{DS}^t$ can be calculated by the weighted sum of aggregating source memory representations shown in equation 2.

$$\mathbf{c} = \text{Softmax}(\mathbf{W}_p \cdot \mathbf{v}^t + \mathbf{b}_p),\ \mathbf{z}_{DS}^t = \sum_{i=1}^{n} c_i * \mathbf{v}^i \tag{2}$$

4 Experiments

4.1 Settings

Dataset We conduct experiments on a real-world indoor temperature dataset with four residential areas, collected from 2018/12/15 to 2019/03/15 with hourly time intervals, including *indoor temperature, heating temperature, outdoor temperature, outdoor humidity, wind speed*, and *wind direction*. For evaluation, we use 4-fold cross-validation, where one residential area is regarded as the target area and the others as source areas.

Baselines We compare MMeSN with 7 baselines. **No Transfer**, **MFSAN**[9], **CoDATS**[6], **DANN**[1], **TL-MLP**[2], **SHL-DNN**[4], **TL-SMI**[5]. In our setting, we adjust the implementations of MFSAN, DANN and CoDATS by changing the CNN layers to fully connected layers and the classification loss to regression loss. We also compare several model variants by removing some components (**only DI, only DS, one branch**) and replace the memory module with fully connected layers (**w/o Mem, w/o S Mem. w/o T-S Mem**).

Model Details We use min-max normalization to normalize the continuous value to [0,1]. The extractor is implemented by a feature embedding layer with unit size 8 followed by a flatten layer and 2 fully connected layers with unit size $\{64, 16\}$. As for the source and target memory size, we set the source memory as 7×16 and target memory as 6×16. For domain independent subnet, we use two fully connected layers with sizes $\{16, 8\}$. The slide window size is set as 12.

4.2 Model Comparison

Comparison among different baselines As shown in Table 1, we compare MMeSN with multiple baselines on four areas. When directly trained with the observed target data, the DNN model (No Transfer) performs worst, showing the importance of transfer learning for the data scarcity problem. The next three approaches (MFSAN, CoDATS, and DANN) achieve a relatively higher MAE, as these approaches are designed to learn the domain-invariant features by aligning the distribution of source and target observed data. However, the future distribution is unforeseeable. The performance decreases when the distribution shift happens for the target domain. TL-MLP, SHL-DNN, and TL-SMI are parameter-based methods for time-series prediction, which achieve a smaller error margin. Note that TL-SMI achieves comparatively better performance since this approach is designed for thermal load prediction and selects the optimum source to pre-train the model and then fine-tunes the whole network with target data. Our proposed model achieves the best performance comparing to all baselines with the average 8.6% and 8.3% relative improvement beyond TL-SMI on MAE and MAPE, respectively. This is because MMeSN decouples the independent & specific knowledge for each domain and memorizes multiple intra- and inter-correlations between target and source domains, which improves the generation for the target prediction.

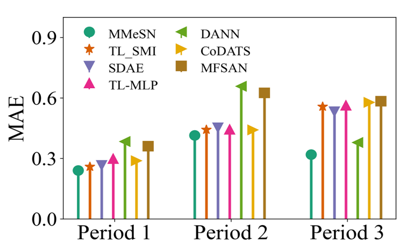
Table 1. Comparison with different baselines.

Methods	Area 1		Area 2		Area 3		Area 4		Average	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
No Transfer	0.625	3.319	0.720	3.923	0.636	2.723	1.171	6.321	0.788	4.071
MFSAN	0.341	1.798	0.474	1.589	0.167	0.715	0.577	2.876	0.390	1.744
CoDATS	0.250	1.316	0.241	1.258	0.190	0.815	0.567	2.685	0.312	1.518
DANN	0.260	1.358	0.264	1.421	0.229	0.985	0.576	2.886	0.333	1.662
TL-MLP	0.316	1.669	0.291	1.562	0.203	0.870	0.466	2.283	0.319	1.596
SHL-DNN	0.293	1.560	0.210	1.222	0.138	0.590	0.428	2.096	0.267	1.367
TL-SMI	0.259	1.371	0.230	1.232	0.122	0.514	0.413	2.021	0.256	1.284
MMeSN	**0.243**	**1.273**	**0.207**	**1.112**	**0.120**	**0.513**	**0.366**	**1.810**	**0.234**	**1.177**

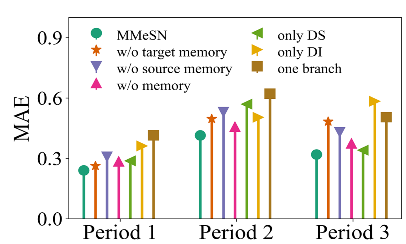
Comparison among different distributions To verify the generalization for distribution shift problem, we compare MMeSN with different baselines in Figure 3 (a) and different model variants in Figure 3 (b). From the comparison results, our proposed MMeSN could achieve the best performance both among all baselines and variants, which demonstrates the effectiveness of our model to alleviate the impact of distribution shift problem.

5 Related Work

Existing transfer learning approaches mainly can be divided into two folds. One is the domain adaptation based methods by aligning the distribution between source and target [9, 6]. The other is the parameter-based methods pre-trained by source data and fine-tuned by limited target data. For urban transfer learning, it has been widely applied in three-class applications. The first is the prediction problem, e.g. crowd flow prediction for a new city [7]. The second is the deployment problem, e.g. commercial store site recommendations [8]. The third is the detection problem which detects the objects of interest [3]. Our proposed MMeSN adopts the parameter-based methods with multiple memories to predict the indoor temperature with limited data, which is robust for memorizing historical knowledge with better transfer generalization.



(a) Baselines comparison



(b) Model variants comparison

Fig. 3. Comparison on different time periods of area 4

6 Conclusion

In this paper, we propose a multi-memory enhanced separation network, MMeSN, to predict the residential indoor temperature with limited data. For transferring multi-source knowledge, we adopt a joint decomposition architecture to decouple the domain independent & specific information and utilize multiple source memories and target-source memory to learn the historical patterns. Experimental results demonstrate the advantage of our approach over several baselines.

Acknowledgments

This work was supported by Beijing Nova program (Z211100002121119) and the Youth Fund Project of Humanities and Social Science Research of Ministry of Education (No. 21YJCZH045).

References

1. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**, 59:1–59:35 (2016)
2. Gao, N., Shao, W., Rahaman, M.S., Zhai, J., David, K., Salim, F.D.: Transfer learning for thermal comfort prediction in multiple cities. Building and Environment **195**, 107725 (2021)
3. He, T., Bao, J., Li, R., Ruan, S., Li, Y., Song, L., He, H., Zheng, Y.: What is the human mobility in a new city: Transfer mobility knowledge across cities. In: Proceedings of The Web Conference 2020. pp. 1355–1365 (2020)
4. Hu, Q., Zhang, R., Zhou, Y.: Transfer learning for short-term wind speed prediction with deep neural networks. Renewable Energy **85**, 83–95 (2016)
5. Lu, Y., Tian, Z., Zhou, R., Liu, W.: A general transfer learning-based framework for thermal load prediction in regional energy system. Energy **217**, 119322 (2021)
6. Wilson, G., Doppa, J.R., Cook, D.: Multi-source deep domain adaptation with weak supervision for time-series sensor data. In: KDD. pp. 1768–1778 (2020)
7. Yao, H., Liu, Y., Wei, Y., Tang, X., Li, Z.: Learning from multiple cities: A meta-learning approach for spatial-temporal prediction. In: WWW. pp. 2181–2191 (2019)
8. Yao, Y., Liu, P., Hong, Y., Liang, Z., Wang, R., Guan, Q., Chen, J.: Fine-scale intra-and inter-city commercial store site recommendations using knowledge transfer. Transactions in GIS **23**(5), 1029–1047 (2019)
9. Zhu, Y., Zhuang, F., Wang, D.: Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In: AAAI. vol. 33, pp. 5989–5996 (2019)
10. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q.: A comprehensive survey on transfer learning. Proceedings of the IEEE **109**, 43–76 (2021)