Modeling Users' Behavior Sequences with Hierarchical Explainable Network for Cross-domain Fraud Detection

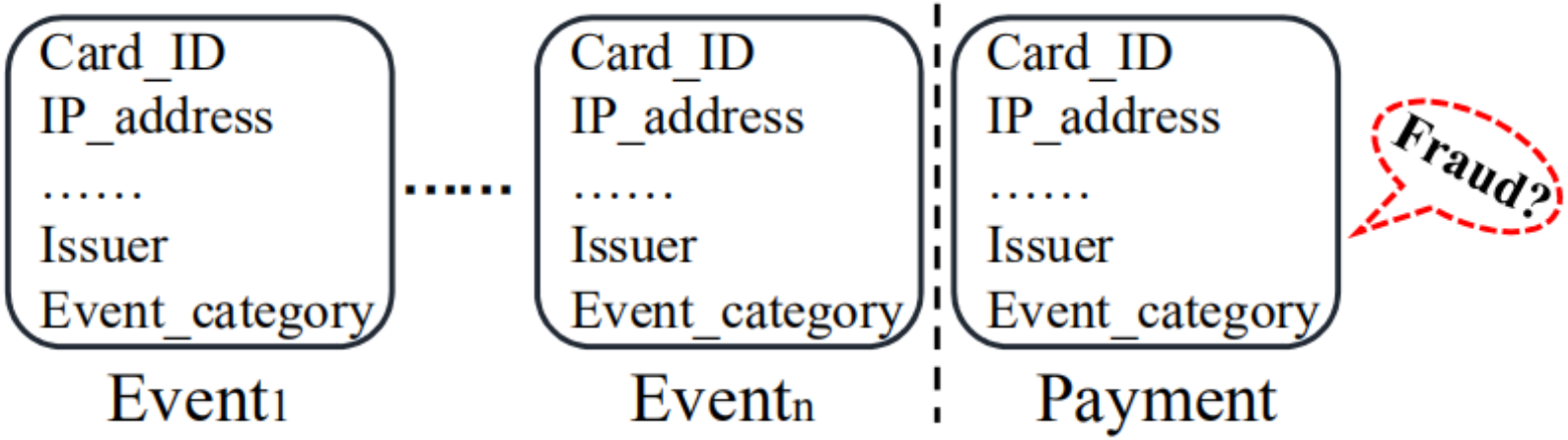
Yongchun Zhu[1, 2, 3], Dongbo Xi[1, 2, 3], Bowen Song[3], Fuzhen Zhuang[1, 2], Shuai Chen[3], Xi Gu[3], Qing He[1, 2]

1. Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS) , Institute of Computing Technology, CAS, Beijing, China

2. University of Chinese Academy of Sciences, Beijing, China

3. Alipay (Hangzhou) Information & Technology Co., Ltd.

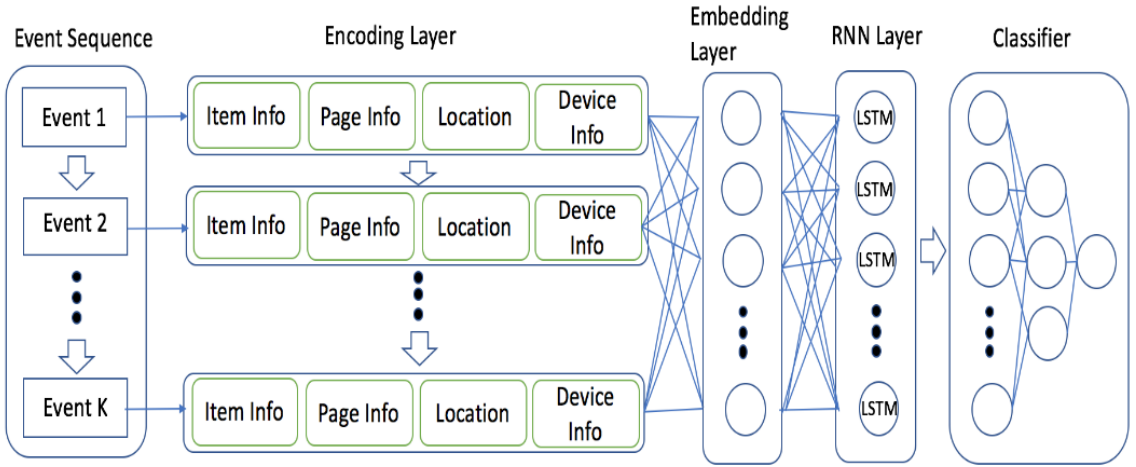
Motivation



Sequence-based Fraud Detection Task: Exploit the users' historical behavior sequences to help differentiate fraudulent payments from regular ones.

Motivation

Existing Method



Shortcommings:

- The prediction results are difficult to explain for these methods.
- They focus more on the sequential information of the behaviors, but fail to thoroughly exploit the internal information of each behavior, e.g., only the first-order information of fields' embeddings is used to represent events.

In order to tackle the above two problems, we propose a Hierarchical Explainable Network (HEN) to model users' behavior sequences, which could not only improve the performance of fraud detection but also help answer this "why".

Motivation

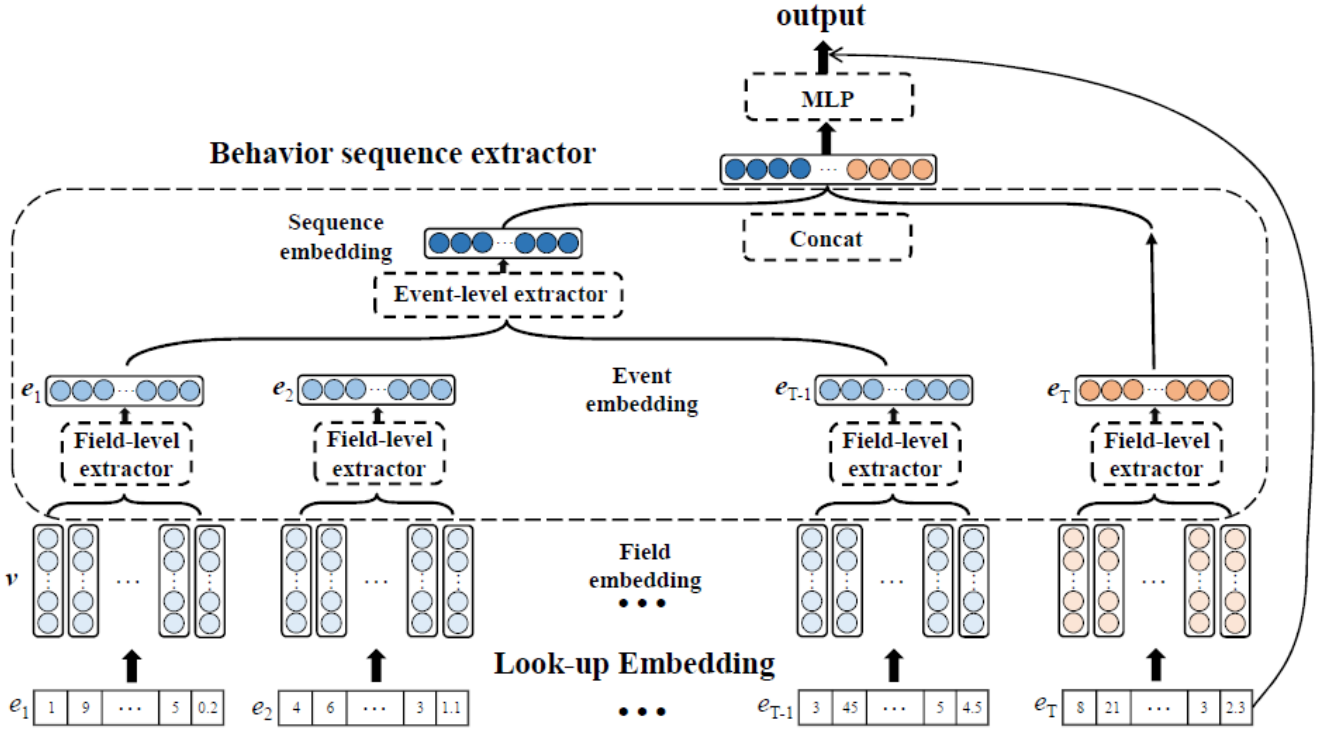
New challenge:
Our company is one of the world-leading cross-border e-commerce companies. As its e-commerce business expands to new domains, e.g., new countries or new markets, one major problem for modeling users' behavior in fraud detection systems is the limitation of data collection, e.g., very few data/labels available. (The need of Lazada)

Due to the new challenge, we introduce **cross-domain fraud detection**, which aims to transfer knowledge from existing domains (source domains) with enough and mature data to improve the performance in the new domain (target domain).

Contributions

- To solve the sequence-based fraud detection problem, we propose hierarchical explainable network (HEN) to model users' behavior sequences, which could not only improve the performance of fraud detection but also give reasonable explanations for the prediction results.
- For cross-domain fraud detection, we propose a general transfer framework that can be applied upon various existing models in the Embedding & MLP paradigm.
- We perform experiments on real-world datasets of four countries to demonstrate the effectiveness of HEN. In addition, we demonstrate our transfer framework is general for various existing models on 90 transfer tasks. Finally, we conduct case study to prove the explainability of HEN and the transfer framework.

Hierarchical Explainable Network



Problem Statement:

Event sequence:

$$\boldsymbol{E} = [\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_T]$$

Event presentation:

$$\boldsymbol{e}_t = [x_1^t, x_2^t, ..., x_n^t]$$

Structure:

- Look-up embedding
- Field-level extractor
- Event-level extractor
- Wide layer
- MLP

Hierarchical Explainable Network

1、Look-up embedding

- Look-up embedding has been widely adopted to learn dense representations from raw data for prediction.

$$\boldsymbol{v}_i^t = \begin{cases} \Phi_i[x_i^t] & \text{the } i\text{-th field is categorical field} \\ x_i^t \times \Phi_i & \text{the } i\text{-th field is numerical field} \end{cases},$$

2、Field-level extractor

- The field-level extractor aims to extract the event representations from the embedding vectors of fields.

$$e_t = \sum_{i=1}^{n} w_i^t \boldsymbol{v}_i^t + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \boldsymbol{v}_i^t \odot \boldsymbol{v}_j^t, \qquad w_i^t = \frac{\exp(a^i_{x_i^t})}{\sum_{j=1}^{n} \exp(a^j_{x_j^t})},$$

Hierarchical Explainable Network

3、Event-level extractor

- The event-level extractor aims to extract the sequence embedding from the historical event embedding vectors.

$$s = \sum_{t=1}^{T-1} u_t f_1(e_t), \quad \hat{u}_t = \frac{< f_2(e_t), f_3(e_t) >}{\sqrt{k}}, u_t = \frac{\exp(\hat{u}_t)}{\sum_{j=1}^{T-1} \exp(\hat{u}_j)},$$

4、Wide layer

- Wide layer is the linear part just like the “wide” part in Wide & Deep to capture the first-order information.

$$l(e_T) = \sum_{i=1}^{n} c_i(x_i^T) + c_0,$$

Hierarchical Explainable Network

5、Prediction and learning

$$\hat{y} = Sigmoid\left(MLP([s, e_T]) + l(e_T)\right),$$

$$L(\theta) = -\frac{1}{N} \sum_{(E,y)\in\mathcal{D}} (y \log \hat{y} + (1-y)\log(1-\hat{y})),$$

Explainability：

- Field-level extractor: Our field-level extractor is able to choose informative fields. **w** indicates the attention distribution of field embeddings that could explain which field embedding is more important to represent event embedding.
- Event-level extractor: Our event-level extractor is able to score the event importance, and from the score, we could find which event is more important to represent the sequence embedding.
- Wide layer: The wide layer of HEN could help to identify specific value of fields with high-risk/low-risk, which could be used as whitelist/blacklist.

General Transfer Framework

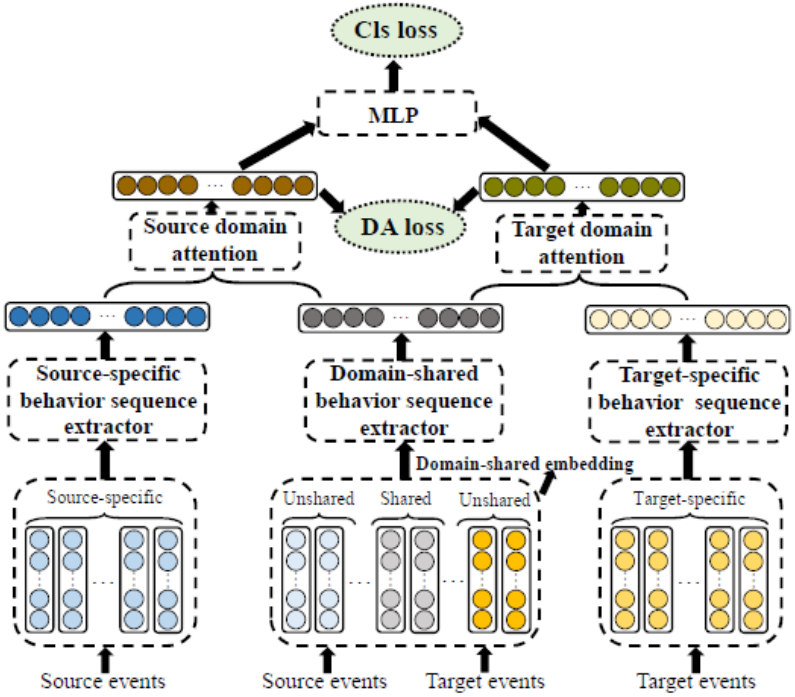


Figure 3: The structure of the general transfer framework for cross-domain fraud detection.

Problem Statement:

There are a source domain and a target domain. The number of labeled samples in target is far less than the source domain.

Structure:

- The Strategy of Embedding
- Shared and Specific behavior sequence extractor
- Domain Attention
- Aligning Distributions

Explainability:

- Domain Attention: Domain attention module is capable of learning the importance of domain-shared and domain-specific representations.

General Transfer Framework

1、The Strategy of Embedding

- We divide the embedding layer into domain-shared and domain-specific.

2、Shared and Specific behavior sequence extractor:

- We also divide the behavior sequence extractor into domain-shared and domain-specific.

3、Domain Attention

- By combining the domain-shared and domain-specific representations (z_{share} and z_{spe}), it could contain more useful knowledge. (o denotes the domain label)

$$z^o = b^o_{share} g_1(z^o_{share}) + b^o_{spe} g_1(z^o_{spe}), \qquad \hat{b}^o_p = \frac{< g_2(z^o_p), g_3(z^o_p) >}{\sqrt{k}}, b^o_p = \frac{\exp(\hat{b}^o_p)}{\exp(\hat{b}^o_{share}) + \exp(\hat{b}^o_{spe})},$$

General Transfer Framework

4、Aligning Distributions

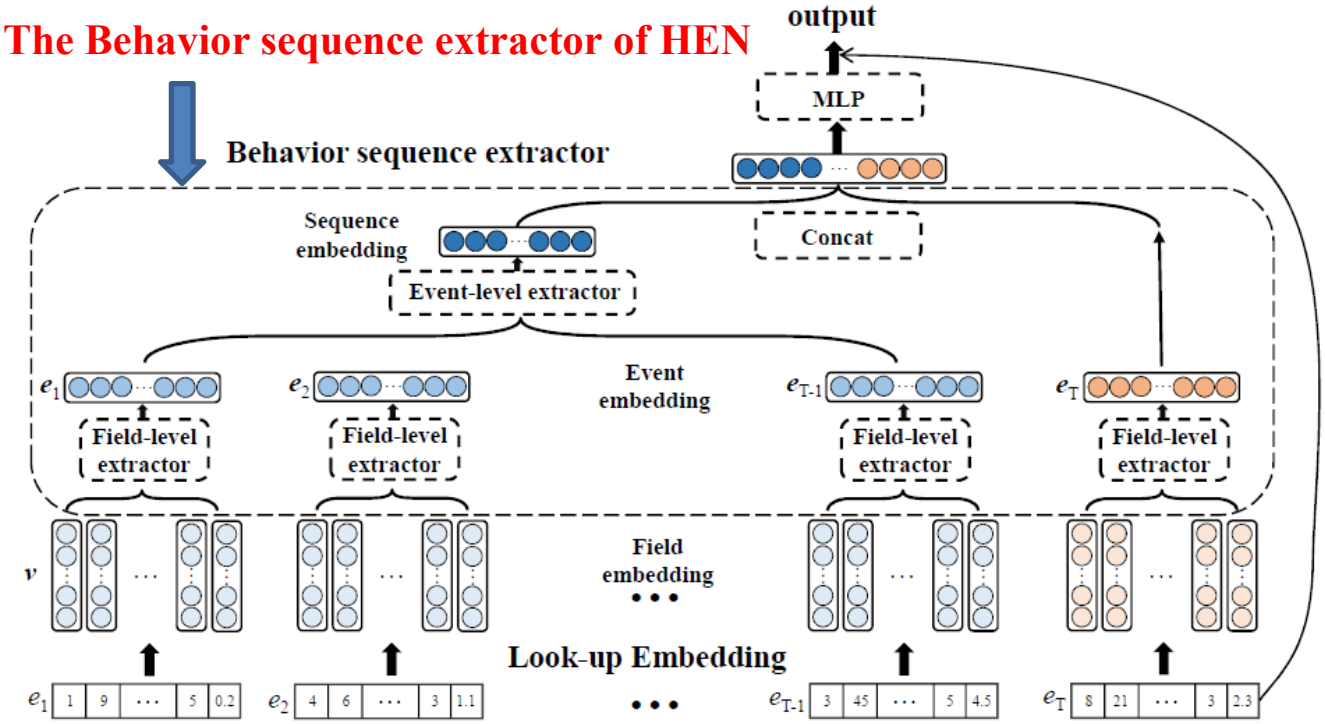
To feed z^{src} and z^{tgt} into the same MLP, we need to align the distibutions of them. Most existing transfer approaches aim to align the marginal and conditional distributions. However, in our scenario, the class distribution is extremely unbalanced that the number of non-fraud samples is about 100 times more than fraud samples. Hence, aligning the marginal and conditional distributions would lead to unsatisfying performance. For our scene, we propose Class-aware Euclidean Distance which explicitly takes the class information into account and measures the intra-class and inter-class discrepancy across domains.

Euclidean Distance: $$d(\mathcal{D}_{c_1}^{o_1}, \mathcal{D}_{c_2}^{o_2}) = \left\| \frac{1}{N_{c_1}^{o_1}} \sum_{i=1}^{N_{c_1}^{o_1}} z_i^{o_1} - \frac{1}{N_{c_2}^{o_2}} \sum_{i=1}^{N_{c_2}^{o_2}} z_i^{o_2} \right\|^2,$$

Class-aware Euclidean Distance: $$\Delta(\mathcal{D}^{src}, \mathcal{D}^{tgt}) = \frac{\sum_{c \in \{0,1\}} d(\mathcal{D}_c^{src}, \mathcal{D}_c^{tgt})}{\sum_{o_1 \in \{src,tgt\}} \sum_{o_2 \in \{src,tgt\}} d(\mathcal{D}_{c=0}^{o_1}, \mathcal{D}_{c=1}^{o_2})},$$

General Transfer Framework

The Behavior sequence extractor of HEN



The general transfer framework could be applied upon various existing models in the Embedding & MLP paradigm. The most important thing to apply the transfer framework is to define the behavior sequence extractor. For our HEN, the behavior sequence extractor contains a field-level extractor and an event-level extractor as shown in the left.

Experiments

Dataset:

Dataset	#pos	#neg	#pos ratio	#fields	#events	#avglen
C1	10K	1.93M	5.2‰	56	3.57M	4.94
C2	4.2K	1.48M	2.8‰	56	3.91M	4.73
C3	15K	1.37M	10.8‰	56	4.28M	14.97
C4	5.7K	174K	31.7‰	56	353K	4.91

The fraud detection dataset is collected from one of the world-leading cross-border e-commerce company, which utilizes the risk management system to detect the transaction frauds.

Experiments

Baselines

- W & D[1]:In real industrial applications, Wide & Deep model has been widely accepted.
- NFM[2]: It is a recent state-of-the-art simple and efficient neural factorization machine model.
- LSTM4FD[3]: Some work has applied LSTM for the sequence-based fraud detection tasks, and we called these methods as LSTM4FD.
- M3R[4]：It is a most recent hierarchical sequence-based model (M3R and M3C) which deals with both short-term and long-term dependencies with mixture models. We choose the better hierarchical model M3R as the baseline.

[1] Cheng, Heng-Tze, et al. "2016. Wide & deep learning for recommender systems." In DLRS, 2016.
[2] He, Xiangnan, et al. "Neural factorization machines for sparse predictive analytics." In SIGIR, 2017.
[3] Wang S, et al. "Session-based fraud detection in online e-commerce transactions using recurrent neural networks"//ECML&EKDD, 2017.
[4] Tang, Jiaxi, et al. "Towards neural mixture recommender for long range dependent user sequences." In WWW, 2019.

Experiments

In binary prediction tasks, AUC (Area Under ROC) is a widely used metric. However, in our real card-stolen fraud detection scenario, we should increase the recall rate, while avoid disturbing the normal users as few as possible. In other words, the task is improving the True Positive Rate (TPR) on the basis of low False Positive Rate (FPR). Therefore, we adopt the standardized partial AUC ($\mathrm{SPAUC}_{\mathrm{FPR}\le\mathrm{maxfpr}}$)

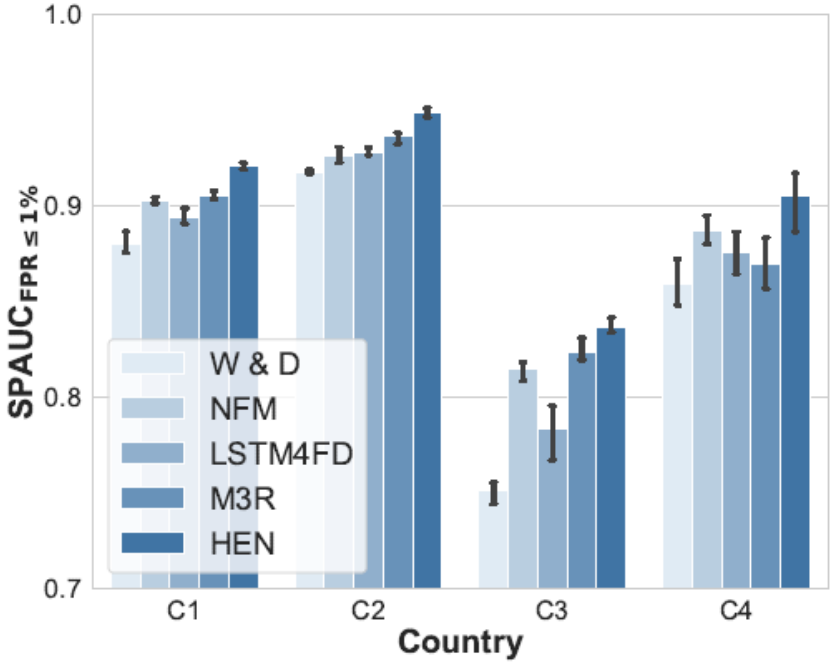
$$\mathrm{SPAUC}_{\mathrm{FPR}\le\mathrm{maxfpr}} = \frac{1}{2}\left(1 + \frac{\mathrm{AUC}_{\mathrm{FPR}\le\mathrm{maxfpr}} - \mathrm{minarea}}{\mathrm{maxarea} - \mathrm{minarea}}\right),$$

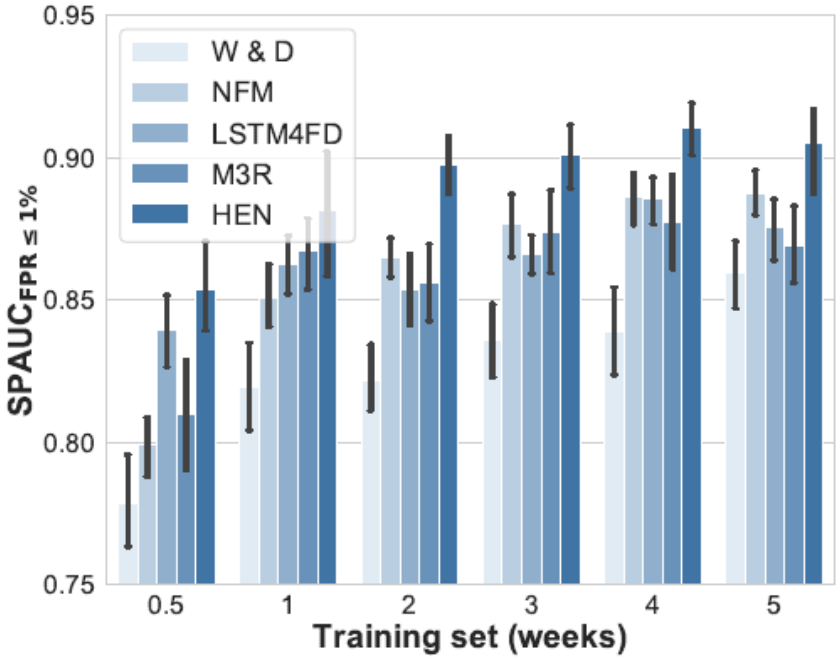
$$\text{where} \quad \mathrm{maxarea} = \mathrm{maxfpr},$$

$$\mathrm{minarea} = \frac{1}{2} \times \mathrm{maxfpr}^2.$$

Experiments

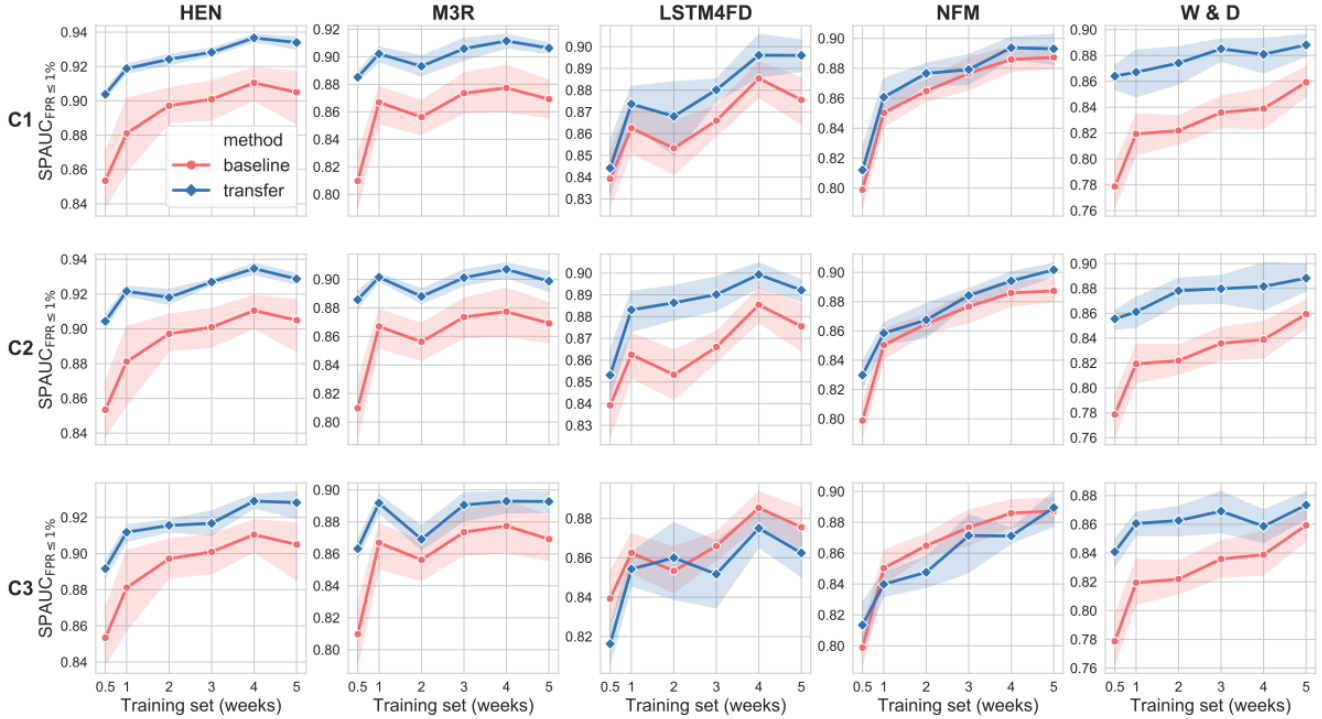
Standard supervised prediction tasks.





Experiments

Transfer tasks



We observe that the transfer framework is able to improve the performance of base models with different sizes of training sets, which proves that the transfer framework is compatible with many models in the Embedding & MLP paradigm.

Experiments

Abalation study:

Table 2: Results ($SPAUC_{FPR \leq 1\%}$) of ablation study on transfer tasks from dataset C1 to dataset C4 based with HEN.

Methods	3days	1week	2weeks	3weeks	4weeks	5weeks	avg
target	0.8534±0.0173	0.8811±0.0232	0.8971±0.0111	0.9009±0.0117	0.9105±0.0099	0.9050±0.0171	0.8913
source	0.5724±0.0230	0.5732±0.0272	0.5909±0.0576	0.5877±0.0334	0.5474±0.0393	0.5705±0.0419	0.5737
pretrain	0.8639±0.0047	0.8920±0.0066	0.8983±0.0115	0.9120±0.0051	0.9172±0.0064	0.9191±0.0054	0.9004
domain-shared	0.8811±0.0110	0.8995±0.0119	0.9025±0.0153	0.9157±0.0109	0.9204±0.0068	0.9194±0.0067	0.9065
our structure	0.8801±0.0101	0.9071±0.0042	0.9038±0.0085	0.9152±0.0094	0.9226±0.0081	0.9233±0.0113	0.9087
Coral	0.8566±0.0156	0.8938±0.0096	0.9028±0.0112	0.9163±0.0128	0.9236±0.0080	0.9221±0.0081	0.9025
Adversarial	0.8763±0.0113	0.8963±0.0135	0.8951±0.0109	0.9116±0.0099	0.9289±0.0053	0.9260±0.0069	0.9057
MMD	0.8744±0.0107	0.8964±0.0105	0.9049±0.0091	0.9185±0.0058	0.9249±0.0075	0.9213±0.0057	0.9067
CMMD	0.8797±0.0136	0.8897±0.0132	0.8963±0.0145	0.9121±0.0135	0.9219±0.0069	0.9146±0.0122	0.9024
ED	0.8767±0.0148	0.9010±0.0088	0.9180±0.0063	0.9186±0.0073	0.9258±0.0076	0.9251±0.0046	0.9109
CED	0.8866±0.0091	0.9076±0.0073	0.9140±0.0059	0.9220±0.0070	0.9305±0.0076	0.9298±0.0059	0.9151
ours	**0.9038**±0.0052	**0.9188**±0.0052	**0.9242**±0.0063	**0.9283**±0.0039	**0.9367**±0.0041	**0.9340**±0.0059	**0.9243**

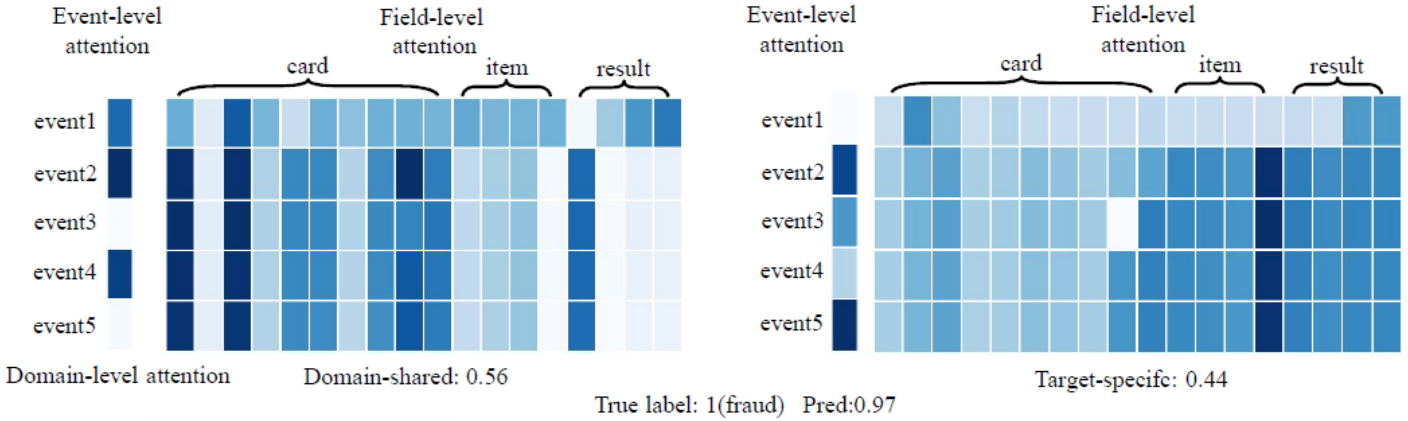
Experiments

Case Study:

	Email	Card	Issuer
High-Risk	email1(123/131) email2(27/27) email3(54/59)	card1(33/33) card2(42/54) card3(77/78)	issuer1(602/607) issuer2(76/108) issuer3(77/90)
Low-Risk	email4(0/382) email5(0/298) email6(0/471)	card4(0/2365) card5(0/245) card6(0/5972)	issuer4(27/12491) issuer5(0/789) issuer6(1/725)

Explanation of Wide Layer:

Three-level Attention:



Thanks  Q & A

Name: Yongchun Zhu | Email: zhuyongchun18s@ict.ac.cn