Persistence Homology Distillation for Semi-supervised Continual learning

Yan Fan, Yu Wang*, Pengfei Zhu, Dongyue Chen, Qinghua Hu
College of Intelligence and Computing, Tianjin University, China
Haihe Laboratory of Information Technology Application Innovation, China
fyan_0411@tju.edu.cn, wang.yu@tju.edu.cn, zhupengfei@tju.edu.cn
dyuechen@tju.edu.cn,huqinghua@tju.edu.cn

Data Benefit Society 机器学习与数据挖掘实验室 Lab of Machine Learning and Data Mining

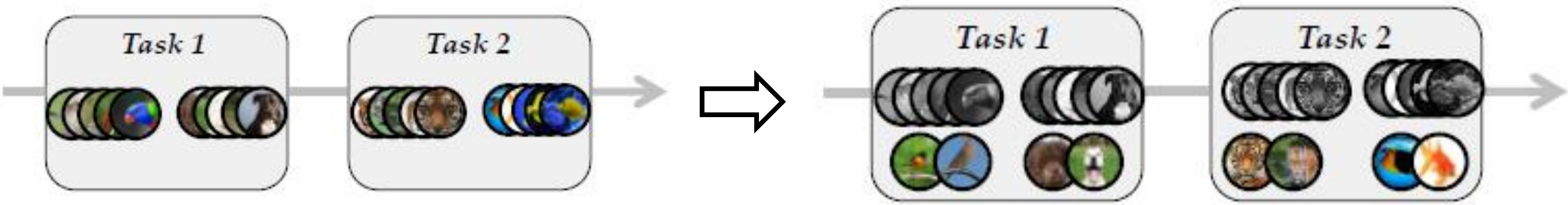
NEURAL INFORMATION PROCESSING SYSTEMS

Persistence Homology Distillation for Semi-supervised Continual learning

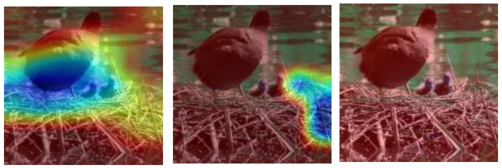
Background

Continual learning

Semi-supervised Continual Learning



Model Evolution



Attention Deficiency ☹

Model Evolution

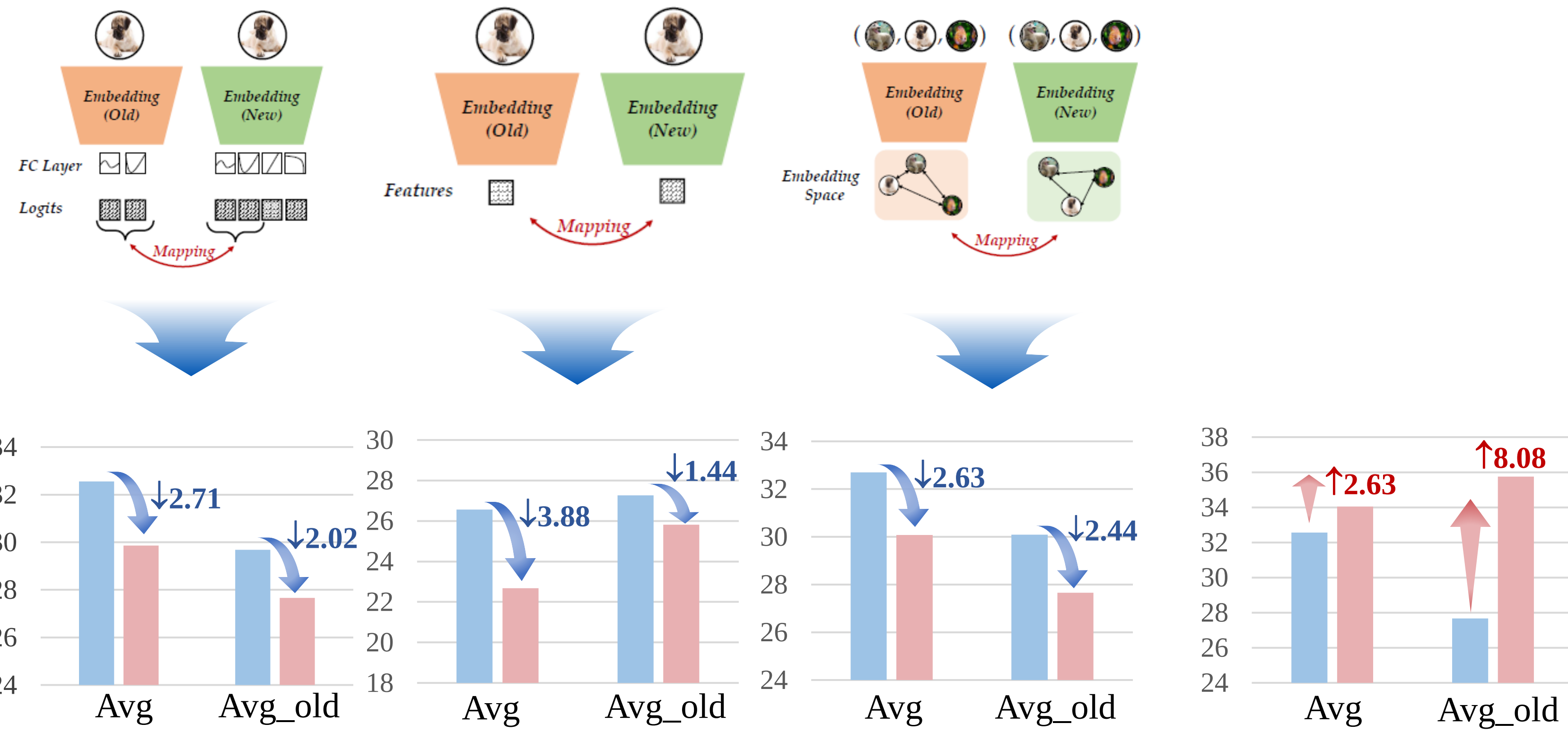
Attention Bias ☹

Fan Y, Wang Y, Zhu P, Chen D, Hu Q. **Persistence Homology Distillation for Semi-supervised Continual Learning**. NIPS, 2024.

Enable stable model evolution by using knowledge of unlabeled data

Persistence Homology Distillation for Semi-supervised Continual learning

Knowledge Distillation

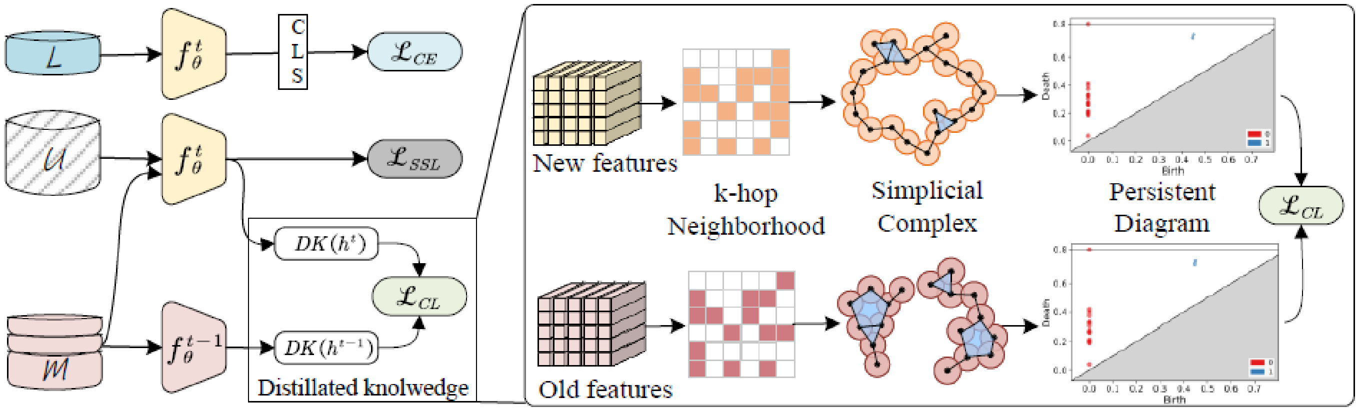


(a) Logits distillation (b) Feature distillation (b) Relation distillation (b) PH distillation (Ours)

Fan Y, Wang Y, Zhu P, Chen D, Hu Q. **Persistence Homology Distillation for Semi-supervised Continual Learning**. NIPS, 2024.

A more powerful noise insensitive distillation is needed for SSCL

Main contribution: persistence homology distillation



(a) (b)

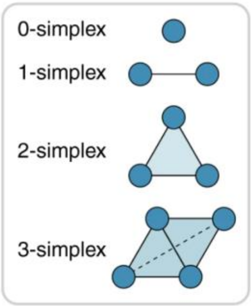
- We utilize simplicial complexes to approximate vision data and explore stable topological feature representation of unlabeled data in semi-supervised continual learning.
- • We propose a novel persistence homology distillation strategy for SSCL that is insensitive to noise information interference, and devise an accelerating algorithm to reduce computation costs.
- • We demonstrate that our method outperforms existing methods on several benchmarks, and highlights the potential of utilizing unlabeled data to overcome catastrophic forgetting.

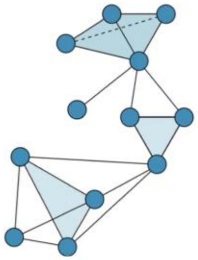
Fan Y, Wang Y, Zhu P, Chen D, Hu Q. **Persistence Homology Distillation for Semi-supervised Continual Learning**. NIPS, 2024.

Persistence Homology Distillation for Semi-supervised Continual learning

Preliminary of Persistence Homology

- **Simplicial complex** : built from a set of simplices that satisfies Closure under Faces and Intersection Property.
  - **k-simplex**： k-dimensional polytope



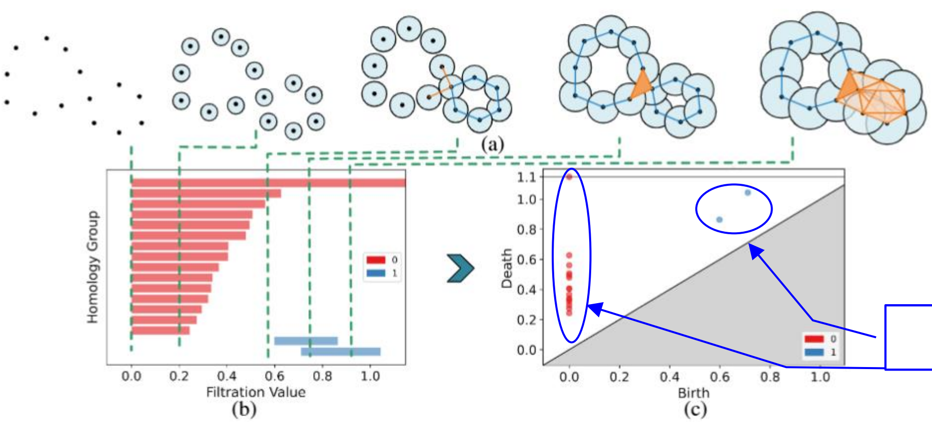


V-R simplicial complex

- **Persistence homology**： quantifying topological features over a range of length scales by defining a filtration on it.

Filtration: $K_r = \delta_X^{-1}((-\infty, r]) = \{y \in \mathbb{R}^d | \delta_X(y) \leq r\} = \cup_{x \in X} B(x, r)$

$$\emptyset = K_0 \subset K_1 \subset \cdots \subset K_m = K,$$



(x_σ, y_σ) : one simplex σ born in x_δ and dies in y_δ

$$Dgm = \{(x_\sigma, y_\sigma) \in \mathbb{R}^2 : x_\sigma < y_\sigma\} \cup \Delta$$

Value: Breaking the limitations of binary topological relationships

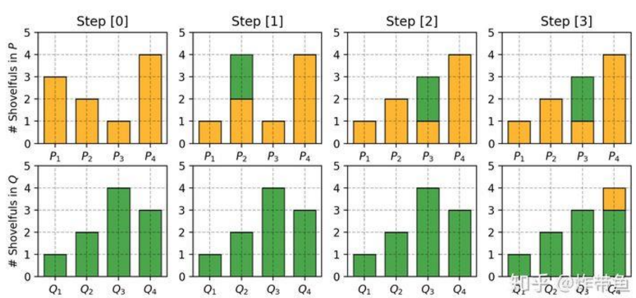
Persistence Homology Distillation

Topological representation：

- Adjcency Matrix： $G(V,E)$ $\quad W_{ij}=\begin{cases}1, if\ s_{ij}\geq\beta\\0, if\, s_{ij}<\beta\end{cases}$ $\quad s_{ij}=<h_i,h_j>=\frac{h_ih_j}{\|h_i\|\|h_j\|}$
- k-hop neighborhood： $\mathcal{N}(x_i,k)$ $x_j\in\mathcal{N}(x_i,k)$ if the shortest path between x_i and x_j is k.
- h-dim local persistence diagram： $Dgm_h(\mathcal{N}(x_i,k))$

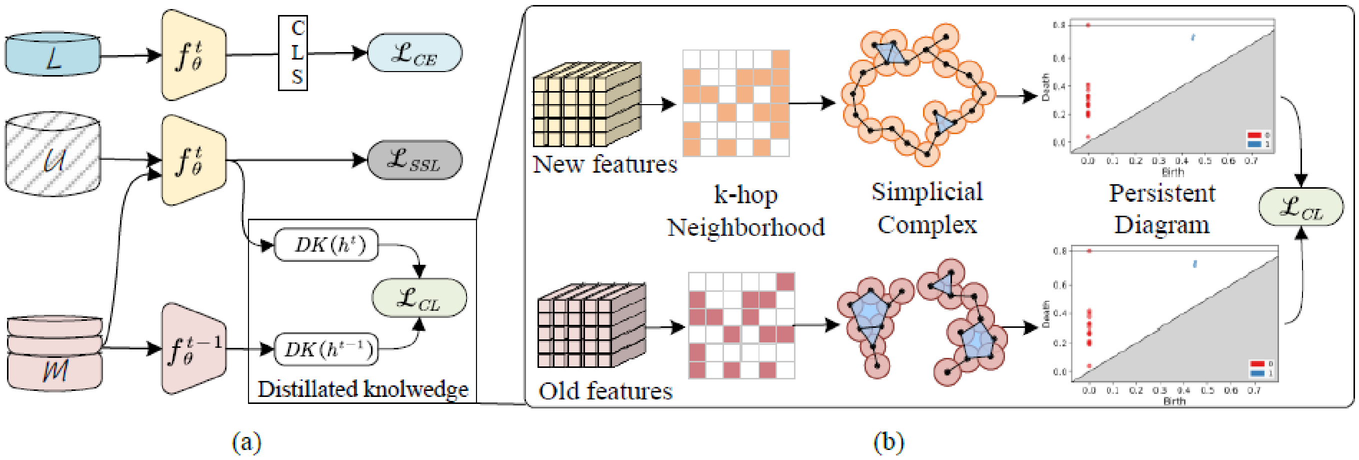
Topological distance : p-Wasserstein distance（Optimal transport distance)

$$d_{\mathcal{N}_x^k}(f_{old},f_{new})=inf_\gamma(\sum_{u\in Dgm_h(\mathcal{N}_x^k,f_{old})\cup\triangle}\|u-\gamma(u)\|^p)^{\frac{1}{p}}$$



Capture high-order relationships in vision data distribution

■ Persistence Homology Distillation



Overall Loss：

$$\mathcal{L}_{sscl} = \mathcal{L}_{CE} + \mathcal{L}_{SSL} + \lambda \mathcal{L}_{CL}$$

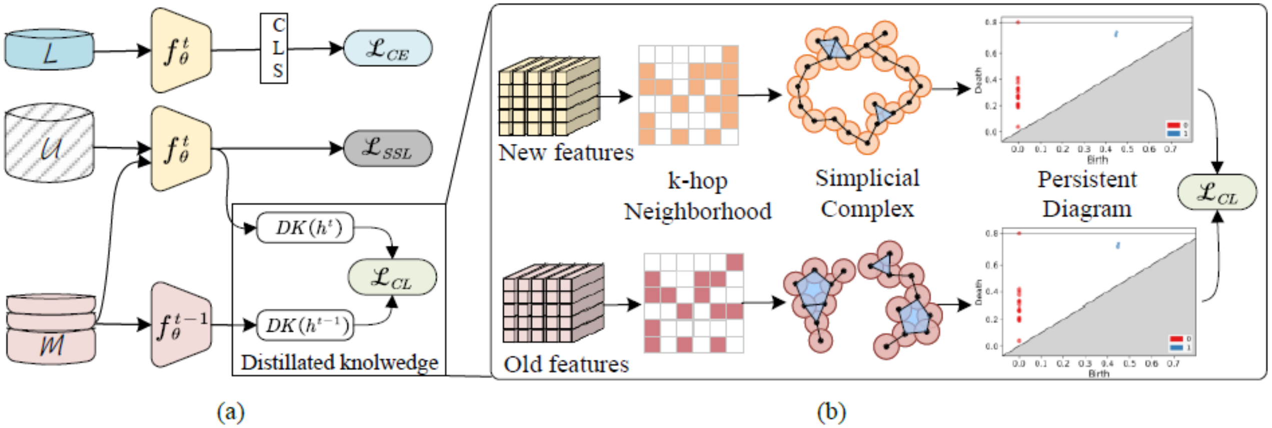
Persistence Homology Distillation loss：

$$\mathcal{L}_{hd} = \frac{1}{|S|} \sum_{x \in S} d_{\mathcal{N}_x^k}(f_{old}, f_{new})$$

Fan Y, Wang Y, Zhu P, Chen D, Hu Q. **Persistence Homology Distillation for Semi-supervised Continual Learning**. NIPS, 2024.

Preserce global knowledge through high-order homology features

Stability Analysis



Theorem 1. *Given the two feature space of point data as $\mathcal{P}_0$ and $\mathcal{P}_1$, for all $p \geq 1$ and $h \in \mathbb{Z}^+$, we have*

$$W_p(Dgm_h(\mathcal{R}(\mathcal{P}_0)), Dgm_h(\mathcal{R}(\mathcal{P}_1))) \leq \binom{M-1}{k}^{\frac{1}{p}} W_p^{pair}(\mathcal{P}_0, \mathcal{P}_1), \tag{7}$$

where $Dgm_h(\mathcal{R}(\mathcal{P}_0))$ and $Dgm_h(\mathcal{R}(\mathcal{P}_1))$ are the h-dimensional persistence diagram for the Vietoris-Rips filtration on the point set $\mathcal{P}_0$ and $\mathcal{P}_1$ respectively. W_p^{pair} represents the pair-wise distance between the point set $W_p^{pair}(\mathcal{P}_0, \mathcal{P}_1) = inf_\phi(\sum_{u,v\in\mathcal{P}_0} |\, \|v-u\| - \|\phi(v)-\phi(u)\|\, |^p)^{\frac{1}{p}})$, where ϕ is a bijection between the two set $\mathcal{P}_0$ and $\mathcal{P}_1$.

Fan Y, Wang Y, Zhu P, Chen D, Hu Q. **Persistence Homology Distillation for Semi-supervised Continual Learning**. NIPS, 2024.

Robustness to noise interference compared to similarity distillation

Experiment Results

- Better adaptable to unlabeled data compared to other knowledge distillation methods

Table 3: Comparison of different knowledge distillation methods applied on SSCL.

Method	Type	CIFAR10 5%	CIFAR10 25%	CIFAR100 5%	CIFAR100 25%
iCaRL[34]	logits	79.2	78.8	31.3	41.4
Foster[40]	logits	75.0	70.4	24.6	38.8
LUCIR[20]	feature	75.2	74.7	32.0	32.6
Podnet[12]	feature	57.9	69.0	21.4	21.1
R-DFIL[16]	relation	78.5	78.8	34.7	34.2
DSGD[14]	relation	79.1	79.0	35.9	43.1
TopKD[22]	topology	78.7	79.8	35.4	41.9
PsHD	topology	**81.5**	**81.4**	**38.8**	**43.2**

- Stability to noise interference

Table 4: Comparison of distillation methods with Gaussian noise interference. σ is the standard deviation.

σ	Podnet[12]		LUCIR[20]		R-DFCIL[16]		DSGD[14]		TopKD[22]		PsHD	
	BWT↓	ACC↑	BWT↓	ACC↑	BWT↓	ACC↑	BWT↓	ACC↑	BWT↓	ACC↑	BWT↓	ACC↑
0.2	31.8	55.6	21.8	71.5	19.2	77.5	18.1	76.7	19.7	76.5	**14.0**	**78.7**
1.0	44.4	27.9	34.3	57.3	28.2	62.1	27.1	64.2	23.0	67.2	**17.7**	**70.9**
1.2	49.8	34.6	36.4	56.4	26.4	61.8	23.7	64.1	22.8	65.8	**18.3**	**67.8**

Fan Y, Wang Y, Zhu P, Chen D, Hu Q. **Persistence Homology Distillation for Semi-supervised Continual Learning**. NIPS, 2024.

Robustness to noise interference compared to other distillation

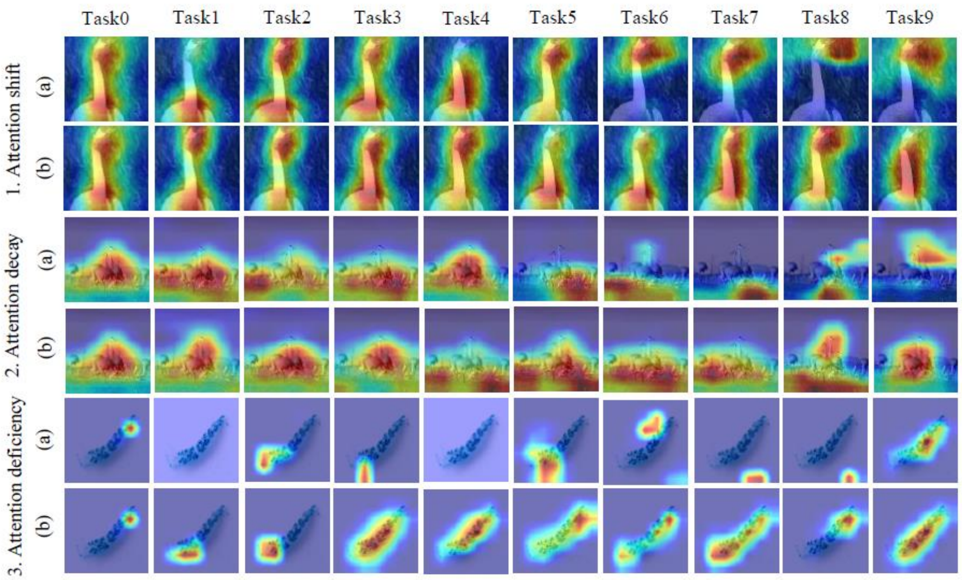
Persistence Homology Distillation for Semi-supervised Continual learning

Experiment Results

Ablation Study

Table 5: Ablation study of proposed persistence homology distillation

Method	$\mathcal{L}_{SSL}$	$\mathcal{L}_{hd}$	CIFAR10_5%		CIFAR10_25%		CIFAR100_5%		CIFAR100_25%	
			avg	last	avg	last	avg	last	avg	last
iCaRL	√		83.7	79.2	82.9	78.8	54.2	36.1	58.4	41.1
	√	√	$85.3_{+1.5}$	$80.8_{+1.6}$	$85.4_{+2.6}$	$80.5_{+1.7}$	$56.4_{+2.2}$	$38.8_{+2.7}$	$59.2_{+0.8}$	$42.2_{+1.1}$
DER	√		86.1	81.2	84.9	81.4	57.8	46.3	68.1	57.2
	√	√	$86.2_{+0.1}$	$81.8_{+0.6}$	$87.9_{+3.0}$	$83.4_{+2.0}$	$58.2_{+0.4}$	$47.2_{+0.9}$	$68.3_{+0.2}$	$57.7_{+0.5}$



Fan Y, Wang Y, Zhu P, Chen D, Hu Q. **Persistence Homology Distillation for Semi-supervised Continual Learning**. NIPS, 2024.

Stable evolution under insufficient labeling is achieved

Thanks for your listening

Yan Fan, Yu Wang*, Pengfei Zhu, Dongyue Chen, Qinghua Hu
College of Intelligence and Computing, Tianjin University, China
Haihe Laboratory of Information Technology Application Innovation, China
fyan_0411@tju.edu.cn, wang.yu@tju.edu.cn, zhupengfei@tju.edu.cn
dyuechen@tju.edu.cn,huqinghua@tju.edu.cn

Data Benefit Society 机器学习与数据挖掘实验室 Lab of Machine Learning and Data Mining

NEURAL INFORMATION PROCESSING SYSTEMS

Homepage

Paper

Code