纵向联邦学习
Vertical Federated Learning

Yang Liu[1]*, Yan Kang[2], Tianyuan Zou[1], Yanhong Pu[1], Yuanqin He[2], Xiaozhou Ye[3], Ye Ouyang[3], Ya-Qin Zhang[1] and Qiang Yang[2,4]

1、 Institute for AI Industry Research, Tsinghua University, Beijing, China.
2、 Webank, Shenzhen, China.
3、 AsiaInfo Technologies, Beijing, China.
4、 Hong Kong University of Science and Technology, Hong Kong, China.

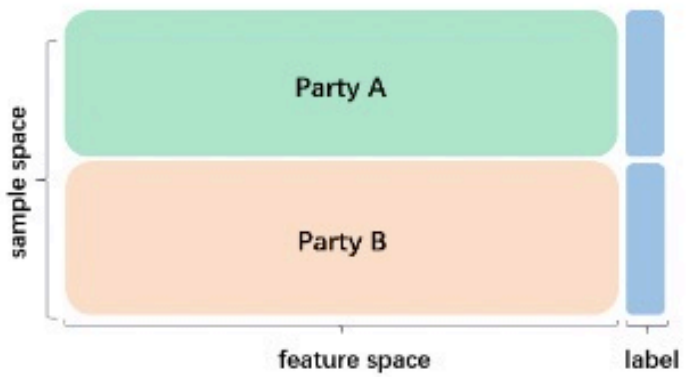
Available online at

目录

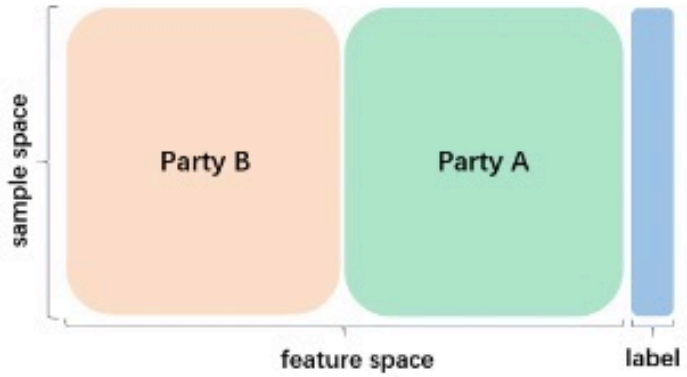
CONTENTS

Available online at

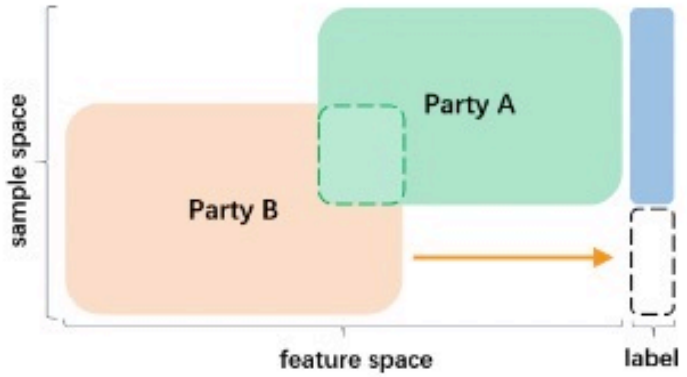
联邦学习的分类



(a) Horizontal Federated Learning



(b) Vertical Federated Learning

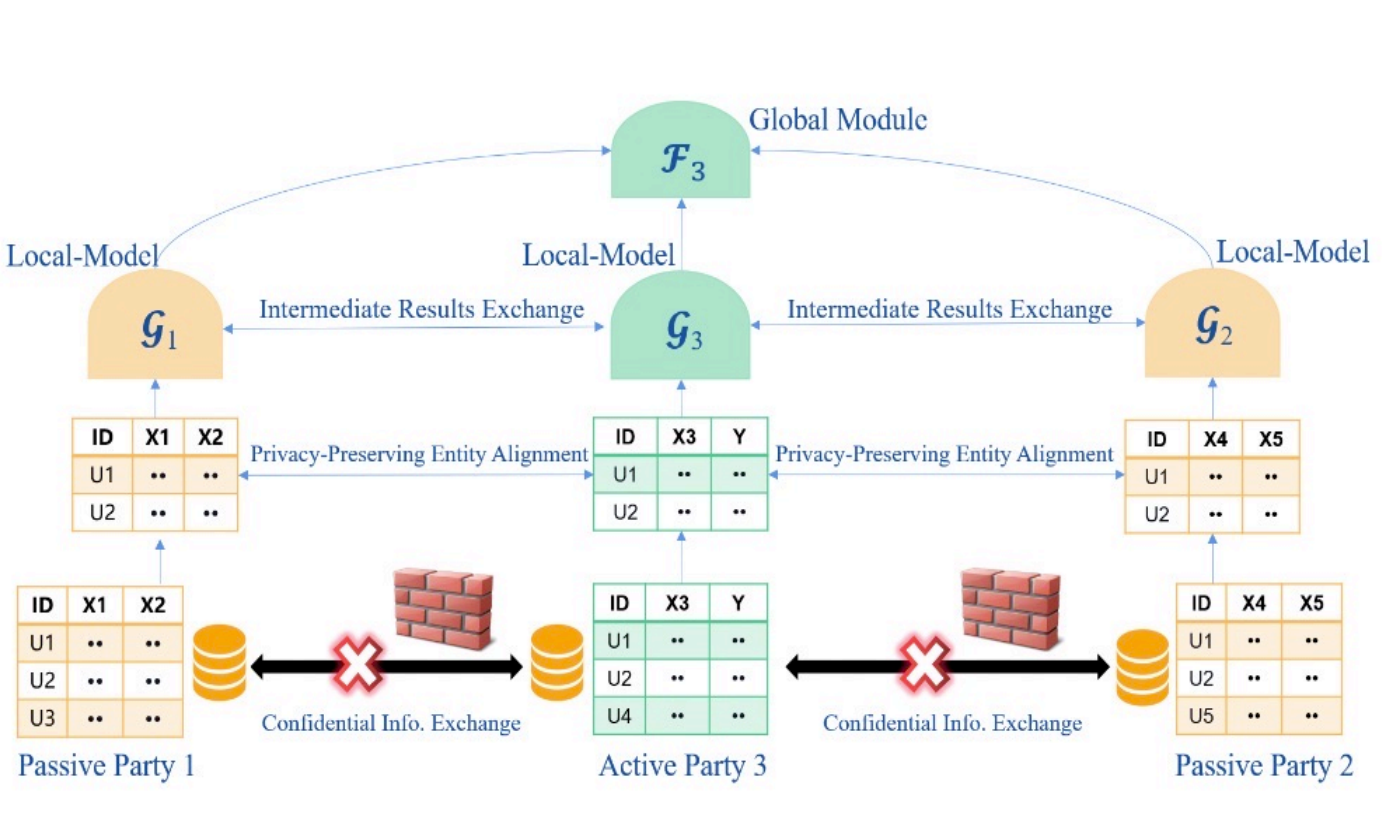


(c) Federated Transfer Learning

	HFL	VFL	FTL
Data is different in	Sample space	Feature space	Both
Scenarios	Cross-device/ Cross-silo	Cross-silo	Mostly Cross-silo
What is exchanged among parties?	Model parameters or gradients	Intermediate results	Intermediate results
What is kept private?	Local data	Local data and model	Local data and model
Each party obtains	A shared global model	A local model	A local model
Require collaboration during inference?	No	Yes	No

Available online at

纵向联邦学习训练目标与基本过程



$$f(\mathbf{\Theta}; \mathbf{x}_i, y_i) = \mathcal{L}\left(\mathcal{F}_K\left(\psi_K; \mathcal{G}_1(\mathbf{x}_{i,1}, \theta_1), ..., \mathcal{G}_K(\mathbf{x}_{i,K}, \theta_K)\right), y_{i,K}\right)$$

Algorithm 1 A General VFL Training Procedure.

Input: learning rates η_1 and η_2

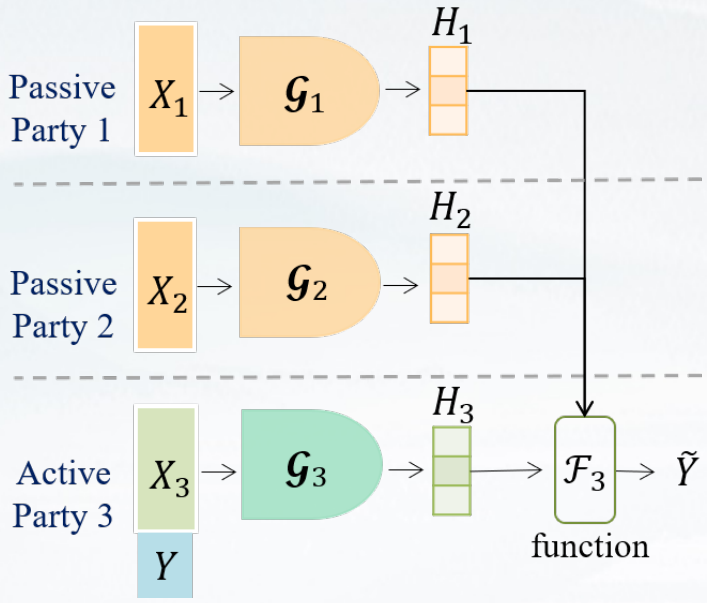
Output: Model parameters θ_1, θ_2 ... θ_K, ψ_K

Party 1,2,...,K, initialize θ_1, θ_2, ... θ_K, ψ_K.
for each iteration $j = 1, 2, ...$ **do**
  Randomly sample a mini-batch of samples $\mathbf{x} \subset \mathcal{D}$
  for each party k=1,2,...,K in parallel **do**
    Party k computes $H_k = \mathcal{G}_k(\mathbf{x}_k, \theta_k)$;
    Party k sends $\{H_k\}$ to party K;
  end for
  Active party K updates $\psi_K^{j+1} = \psi_K^j - \eta_1 \frac{\partial \ell}{\partial \psi_K}$;
  Active party K computes and sends $\frac{\partial \ell}{\partial H_k}$ to all other parties;
  for each party k=1,2,...,K in parallel **do**
    Party k computes $\nabla_{\theta_k} \ell$ with Equation (6);
    Party k updates $\theta_k^{j+1} = \theta_k^j - \eta_2 \nabla_{\theta_k} \ell$;
  end for
end for

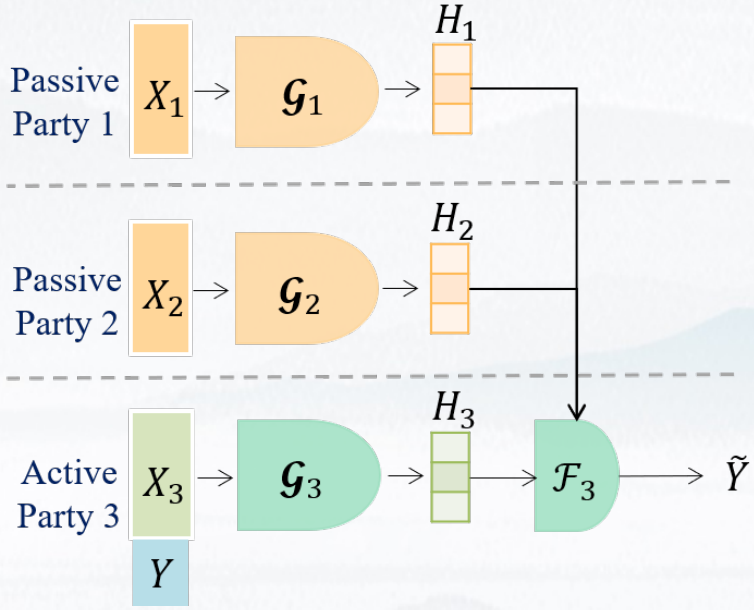
Available online at

纵向联邦学习训练目标与基本过程

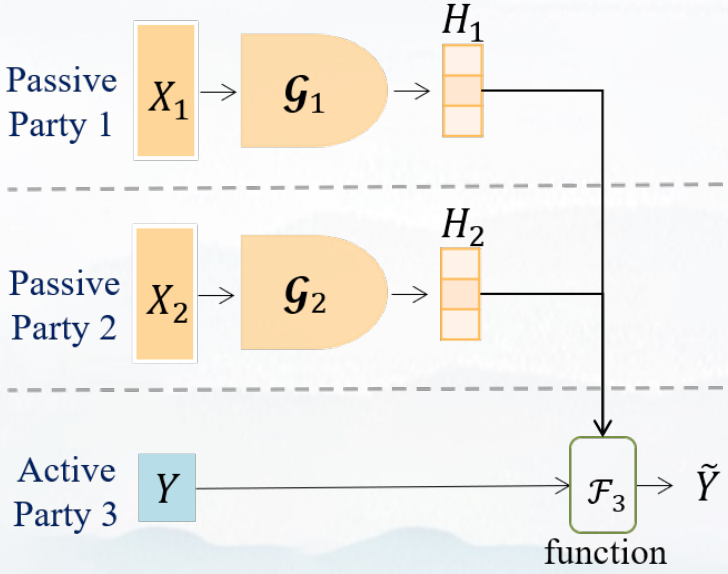
aggVFL



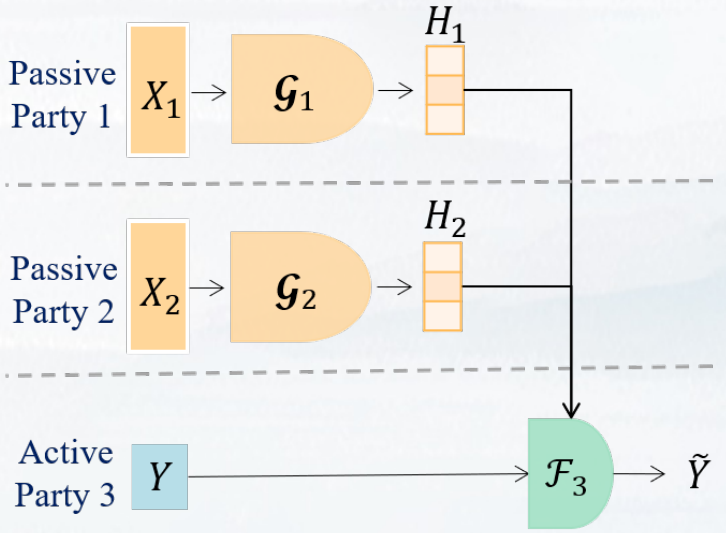
splitVFL



aggVFL$_c$



splitVFL$_c$



Available online at

纵向联邦学习传输优化方法

TABLE 4: Summary of existing works that aim to improve the efficiency of VFL. In the Model column, the LR denotes logistic regression, NN denotes Neural Network and GBDT denotes gradient boosting decision tree. In the Convergence Rate column, T represents the total number of local iterations, Δ represents stochastic variance, and ϵ represents the accuracy.

Category	Existing Work	VFL Setting	Model	Convergence Rate	Core Method
Multiple Client Updates	FedBCD [9]	splitVFL & aggVFL	LR & NN	$O(1/\sqrt{T})$	Block coordinate descent w/ multiple local updates
	Flex-VFL [29]	splitVFL$_c$	NN	$O(1/\sqrt{T})$	Customized # of local updates constrained by time
	AdaVFL [57]	aggVFL$_c$	NN	$O(1/\sqrt{T})$	Customized # of local updates through optimization
	VIMADMM [58]	splitVFL & aggVFL	NN	-	Alternative direction method of multipliers
	CELU-VFL [59]	splitVFL	NN	$O(\Delta/\sqrt{T})$	Cache-based mechanism for local updates
Asynchronous Coordination	GP-AVFL [60]	aggVFL	LR & NN	-	Asynchronous training with gradient prediction
	AVFL [61]	aggVFL	LR	-	Backup-based straggler-resilient scheme
	T-VFL [62]	splitVFL$_c$	NN	-	Channel-aware user scheduling poicy
	VAFL [63]	splitVFL$_c$	LR & NN	$O(1/\sqrt{T})$	Asynchronous query-response strategy
	FDML [18]	aggVFL$_c$	LR & NN	$O(1/\sqrt{T})$	Asynchronous local updates w/ the same data order
	AFAP [64]	aggVFL	LR	$O(e^{-T})$	Tree-structured asynchronous communication (TSAC)
	AsySQN [65]	aggVFL	LR	$O(e^{-T})$	TSAC & quasi-Newton method
	VFB^2 [66]	aggVFL	LR	$O(e^{-T})$	TSAC & bi-level parallel update
	FDSKL [67]	aggVFL	LR	$O(1/T)$	TSAC & random features & doubly stochastic gradient
	FedGBF [68]	aggVFL	GBDT	-	Use RT as the base learner for learning GBDT
	VF^{2}Boost [69]	aggVFL	GBDT	-	Concurrent protocol & customized Paillier HE
One-Shot Coordination	FedOnce [70]	splitVFL	NN	-	Unsupervised learning by predicting noise
	CE-VFL [71]	splitVFL	NN	-	Unsupervised learning using PCA & autoencoder
Compression	AVFL [61]	aggVFL	LR	-	Principle component analysis
	CE-VFL [71]	splitVFL	NN	-	Autoencoder and principle component analysis
	SecureBoost+ [72]	aggVFL	Tree	-	Encode encrypted first-order and second-order gradients into a single message
	eHE-SecureBoost [73]	aggVFL	Tree	-	
	C-VFL [28]	splitVFL	NN	$O(1/\sqrt{T})$	Arbitrary compression scheme
	GP-AVFL+DESC [60]	aggVFL	LR & NN	-	Double-end sparse compression

Available online at

纵向联邦学习建模优化方法

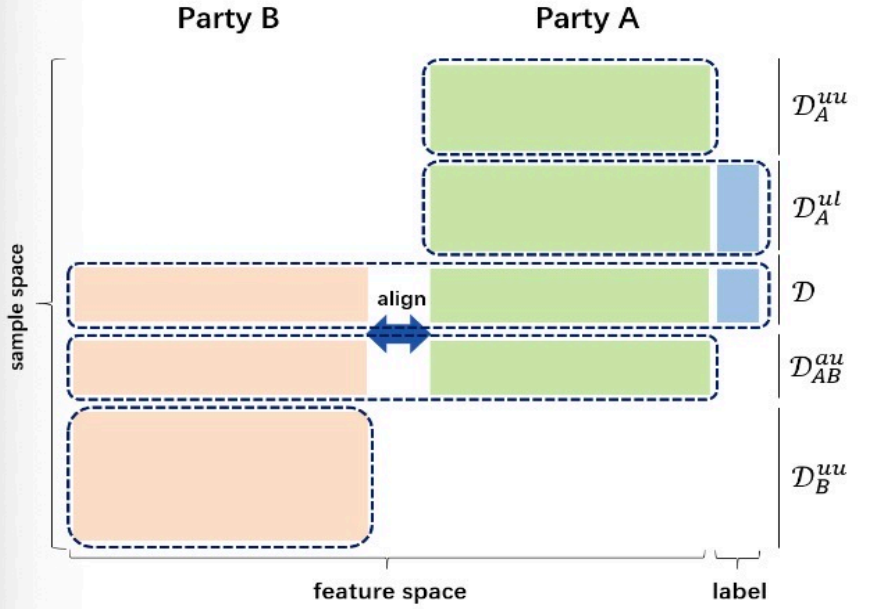
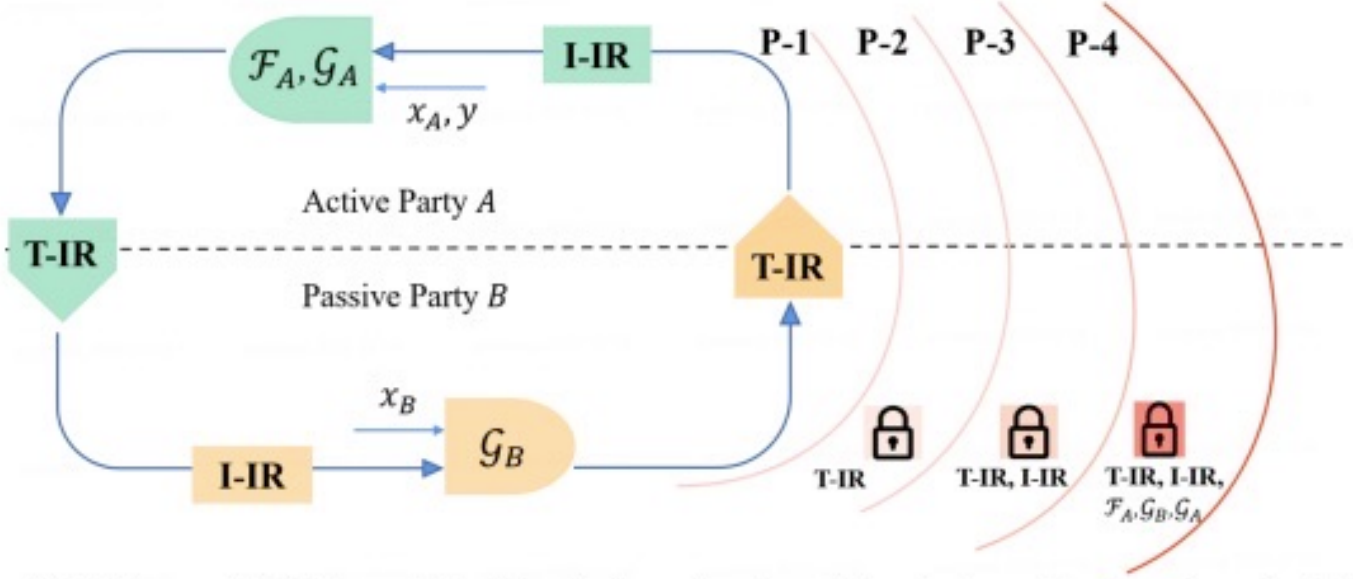


Fig. 3: A general virtual dataset of VFL.

Objective	Existing Work	Data Used					Approach	Party	Data Type
		Aligned		Unaligned					
		$\mathcal{D}^{au}$	$\mathcal{D}$	$\mathcal{D}_B^{uu}$	$\mathcal{D}_A^{uu}$	$\mathcal{D}_A^{ul}$			
Build a joint model	FedCVT [55]	-	√	√	-	√	Semi-SL	2	tabular, image
	FedMC [76]	-	√	√	-	√	Semi-SL	2	tabular
	VFLFS [77]	-	√	√	√	-	Self-SL	≥ 2	tabular
	VFed-SSD [78]	√	√	-	-	-	Self-SL	2	tabular
	FedHSSL [56]	√	√	√	√	-	Self-SL & DA	≥ 2	tabular, image
	SS-VFNAS [54]	-	√	√	√	-	Self-SL & NAS	≥ 2	image
Build a local model for active party A	VFL-KT [79]	√	-	-	-	√	KD & SVD	≥ 2	tabular
	SVFL-Infer [80]	-	√	-	-	-	KD	2	tabular
	VFed-SSD [78]	√	√	-	-	-	KD & Self-SL	2	tabular
	VFed-JPL [81]	-	√	-	-	√	KD	2	tabular
Build a local model for passive party B	MMVFL [82]	-	-	√	-	-	TL	≥ 2	tabular
	SFHTL [83]	-	√	√	-	√	TL & Semi-SL	≥ 2	tabular
	SFTL [84], [85]	√	√	-	-	√	TL	2	tabular
	PrADA [86]*	√	√	-	-	-	TL	3	tabular

Available online at

纵向联邦学习安全分级



T-IR: Transmitted Intermediate Results (e.g., local model outputs and backward gradients)

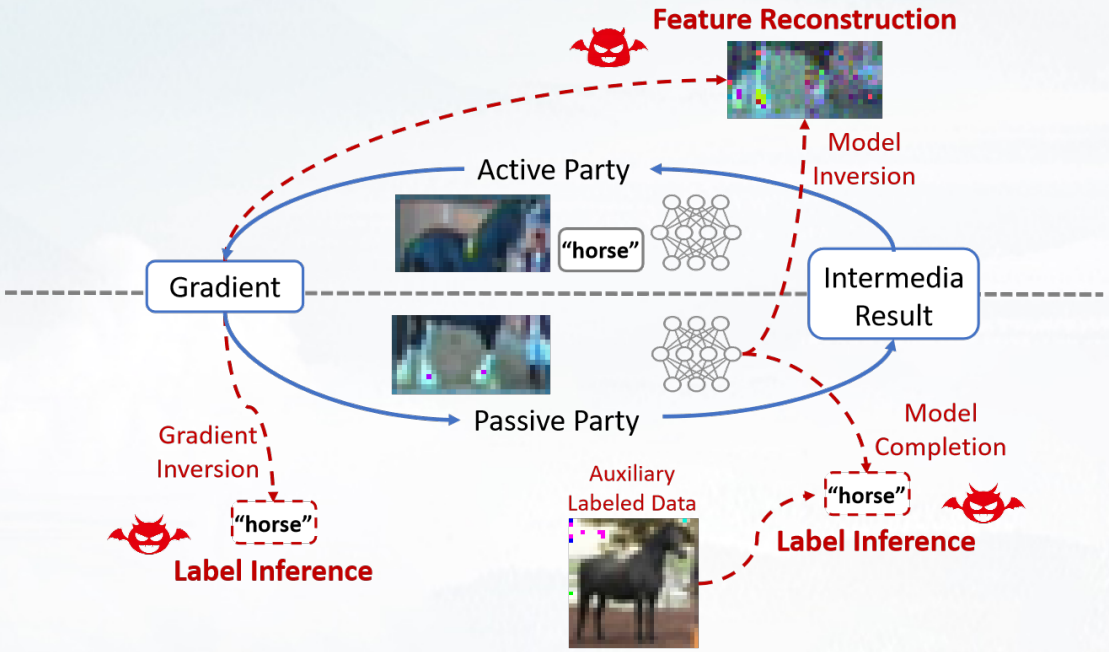
I-IR: Internal Intermediate Results (e.g., local trainable model parameters/gradients)

- **Basic Protocol (P-1): Keeping Private data and models local.**
- **Standard Protocol (P-2): Protecting Exchanged Intermediate Results**
- **Enhanced Protocol (P-3): Protecting Entire Training Protocol**
- **Strict Protocol (P-4): Protecting Training Protocol and Results**
- **Relaxed Protocol (P-0): Nonprivate label or model.**

Available online at

纵向联邦学习框架的攻击防御

数据反演攻击



后门攻击

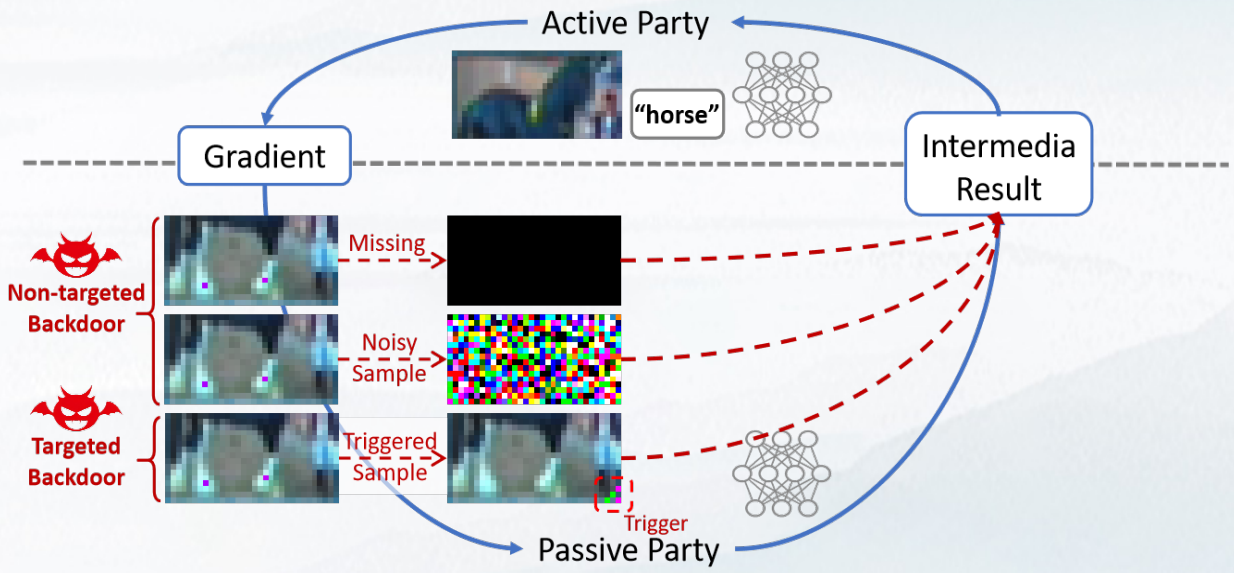


Figure credit : Tianyuan Zou & Yang Liu

Available online at

纵向联邦学习框架的攻击方法

	Attacking Method	VFL Setting	Model	Against Protocol	Attacking Phase	Auxiliary Requirement
Label Inference Attack	Direct Label Inference (DLI) [19], [108]	aggVFL	NN	P-1	Training	–
	Norm Scoring (NS) [109]	splitVFL$_c$	NN	P-1	Training	–
	Direction Scoring (DS) [109]	splitVFL$_c$	NN	P-1	Training	–
	Residual Reconstruction (RR) [110]	aggVFL	LR	P-2	Training	–
	Gradient Inversion (GI) [108]	aggVFL	NN	P-2	Training	–
	Gradient Inversion (GI) [111]	splitVFL$_c$	NN	P-2	Training	Label Prior Distribution
	Passive Model Completion (PMC) [19]	splitVFL	NN	P-3	Inference	Labeled Data
	Active Model Completion (AMC) [19]	splitVFL	NN	P-3	Inference	Labeled Data
Feature Inference Attack	Binary Feature Inference Attack (BFIA) [112]	splitVFL	NN	P-1	Training	Binary Features
	Reverse Multiplication Attack (RMA) [113]	aggVFL	LR	P-2	Training	Corrupted Coordinator
	Protocol-aware Active Attack(PAA) [114]	aggVFL	LR	P-2	Training	–
	Reverse Sum Attack (RSA) [113]	aggVFL	GBDT	P-2	Training	–
	Equality Solving Attack (ESA) [100]	aggVFL	LR	P-0(g)	Inference	–
	Path Restriction Attack (PRA) [100]	aggVFL	Tree	P-0(g)	Inference	–
	Generative Regression Network (GRN) [100]	aggVFL	NN	P-0(g)	Inference	–
	White-Box Model Inversion (MI) [101], [102]	aggVFL & splitVFL$_c$	LR & NN	P-0(g)	Inference	–
	Black-box Model Inversion (MI) [101], [102]	aggVFL & splitVFL$_c$	LR & NN	P-1	Inference	Labeled Data
	Catastrophic Data Leakage in VFL (CAFE) [23]	aggVFL$_c$	NN	P-0(g)	Training	–

	Attacking Method	VFL Setting	Against Protocol	# of Classes	Attacking Phase	Auxiliary Requirement
Targeted Backdor Attack	Label Replacement Backdoor by replacing gradients (LRB) [138]	aggVFL	P-2	≥ 2	Training	At least one label of clean samples
	Adversarial Dominant Input attack (ADI) [139]	VLR/splitVFL$_c$	P-0(g)/P-1	≥ 2	Inference	a few samples from the other party
Non-targeted Backdoor Attack	Adversarial attack [24], [140]	splitVFL/aggVFL	P-1	≥ 2	Training	–
	Missing attack [24]	splitVFL/aggVFL	P-3	≥ 2	Training	–

Available online at

纵向联邦学习框架的防御方法

Cryptographic Defense

Defense Work	VFL Setting	Model	Defense Scheme	Protocol	Party	Require Coordinator	Adversarial Assumption
GasconLR [17]	aggVFL	LR	GC+SS	P-3	≥ 2	✓	SH
HardyLR [94]	aggVFL	LR	HE	P-2	≥ 2	✓	SH
BaiduLR [107]	aggVFL	LR	HE	P-2	≥ 2	✗	SH
SecureLR [108]	aggVFL	LR	HE+SS	P-2	≥ 2	✗	SH
CAESAR [19]	aggVFL	LR	HE+SS	P-3	$= 2$	✗	SH
HeteroLR [97]	aggVFL	LR	HE+SS	a :P-3, p :P-4	$= 2$	✗	SH
FedV [21]	aggVFL	LR/SVM	FE	P-2	≥ 2	✓	SH
SecureBoost [15]	aggVFL	XGB	HE	P-2	≥ 2	✗	SH
SecureBoost+ [33]	aggVFL	XGB	HE	P-2	≥ 2	✗	SH
SecureXGB [35]	aggVFL	XGB	HE+SS	P-3	$= 2$	✗	SH
MP-FedXGB [38]	aggVFL	XGB	SS	P-3	≥ 2	✓	SH
SecureGBM [34]	aggVFL	LGBM	HE	P-2	$= 2$	✗	SH
Pivot [39]	aggVFL	RF / GBDT	HE+SS	P-3	≥ 2	✗	SH, $\leq K$-1 colluded parties
Enhanced Pivot [39]	aggVFL	DT	HE+SS	P-4	≥ 2	✗	SH, $\leq K$-1 colluded parties
FedSGC [109]	aggVFL$_c$	GNN	HE	P-2	$= 2$	✗	SH
ACML [110]	splitVFL$_c$	NN	HE	P-1	$= 2$	✗	SH
PrADA [79]	splitVFL	NN	HE	P-1	≥ 2	✗	SH
BlindFL [96]	splitVFL	NN	HE+SS	a :P-2, p :P-4	$= 2$	✗	SH
SFTL [77]	aggVFL	NN	HE	P-2	$= 2$	✗	SH
SFTL [77]	aggVFL	NN	SS	P-3	$= 2$	✗	SH
SEFTL [78]	aggVFL	NN	HE+SPDZ	P-3	$= 2$	✗	MA,dishonest majority
N-TEE [111]	aggVFL	XGB	TEE	P-3	≥ 2	✗	SH

Emerging Defense

	Defense Work	VFL Setting	Model	Defense Scheme	Against Attack	Defending Party
Defenses against Label Inference Attack	MARVELL[98]	splitVFL$_c$	NN	Add Noise	NS, DS	Active party
	Max-Norm[98]	splitVFL$_c$	NN	Add Noise	NS, DS	Active party
	CAE [30]	aggVFL	NN	HE+Disguise Label	DLI, MC	Active party
	DCAE [30]	aggVFL	NN	HE+Disguise Label+DG	DLI, MC	Active party
	PELoss [113]	splitVFL$_c$	NN	Potential Energy Loss	MC	Active party
	dCorr [101]	splitVFL$_c$	NN	Minimize Correlation	SA	Active party
	RM [114]	aggVFL	LR	HE+Random Mask	RR	Active party
Defenses against Feature Inference Attack	FG [28]	splitVFL	NN	Random Fake Gradients	CAFE	Passive party
	DRAVL [115]	splitVFL$_c$	NN	Adversarial Training	MI	Passive party
	MD [102]	splitVFL	NN	Masquerade	BFIA	Passive party
	DP-Paillier-MGD [104]	aggVFL	LR	HE+DP	PAA	Passive party

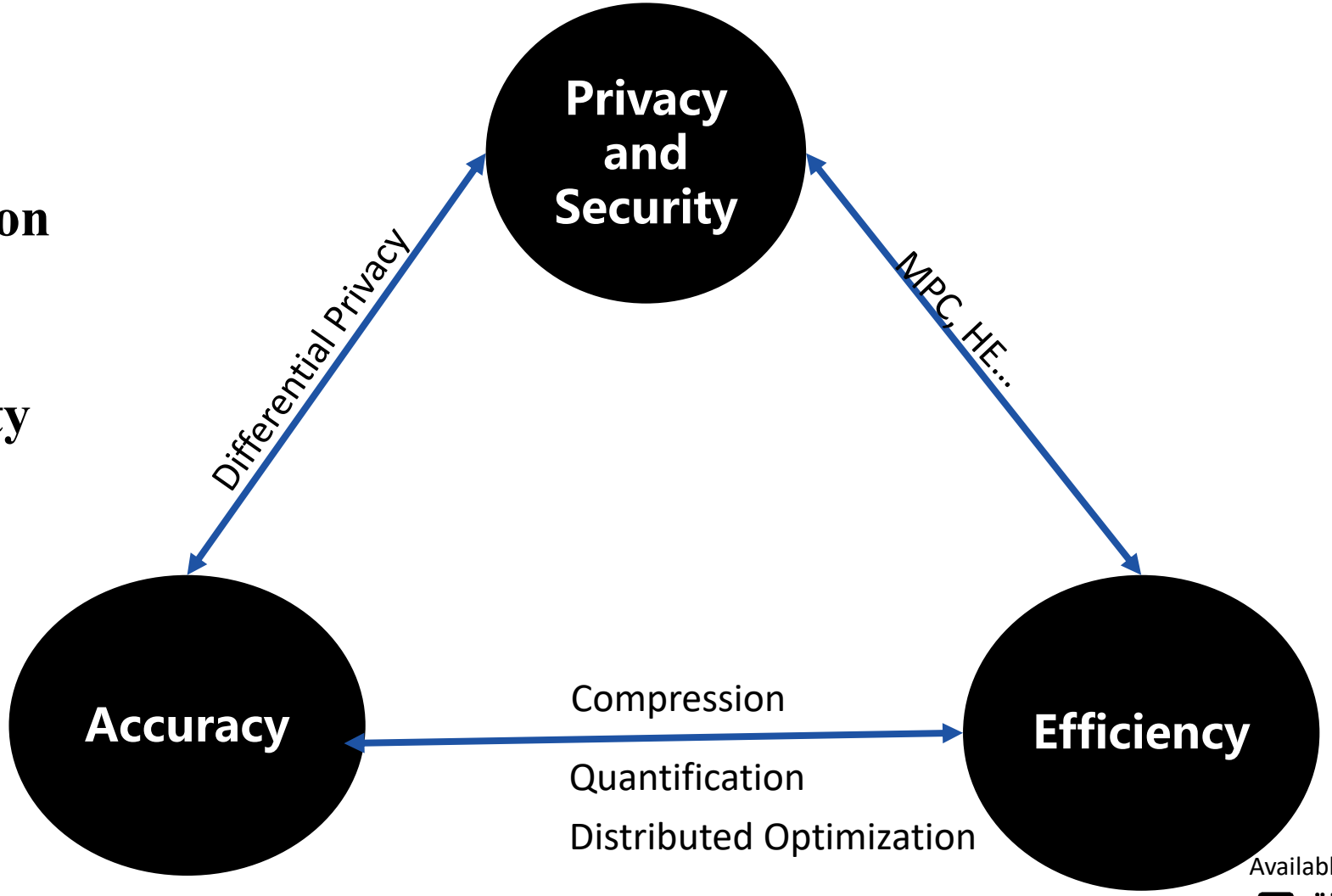
Table 9: Summary of defense strategies for defending against backdoor attacks.

Defense Work	VFL Setting	Defense Scheme	Against Attack
DP[30]	aggVFL	Add Noise	Targeted
GS [30]	aggVFL	Sparsify Gradient	Targeted
CAE [30]	aggVFL	HE+Disguise Label	Targeted
DCAE [30]	aggVFL	HE+Disguise Label+DG	Targeted
RVFR [29]	splitVFL	Robust Feature Sub-space Recovery	Targeted/Non-targeted

Available online at

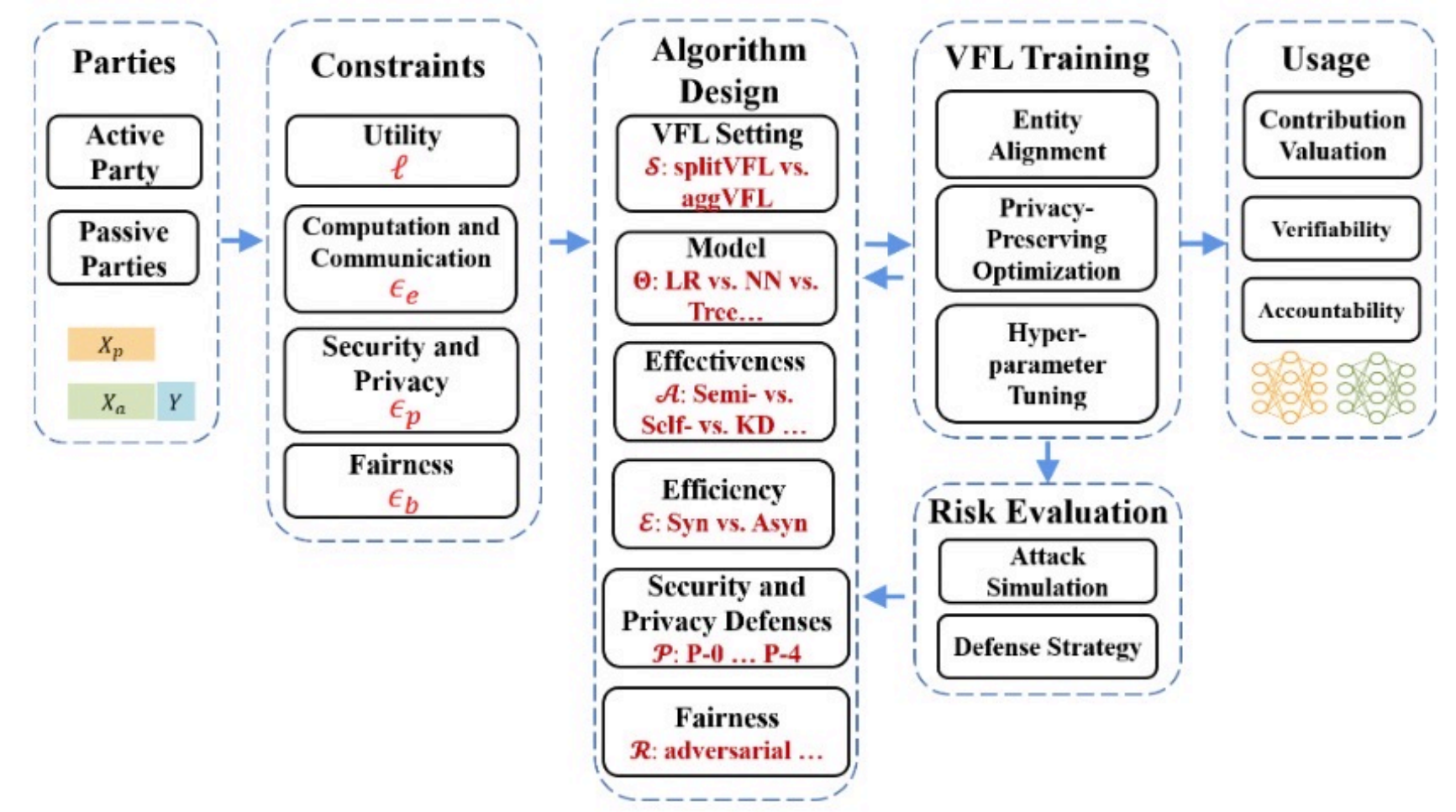
其他关键问题

- 数据价值 Data Valuation
- 可解释性 Explainability
- 数据公平性 Fairness



Available online at

VFLow：A VFL Design Framework for Trustworthy VFL



$$\min_{\Theta} \ell(\Theta; \mathcal{S}, \mathcal{A}, \mathcal{E}, \mathcal{P}, \mathcal{R}, \mathcal{D})$$

$$s.t. \quad M_p(\Theta; \mathcal{K}, \mathcal{P}) \leq \epsilon_p, M_e(\Theta; \mathcal{E}, \mathcal{P})) \leq \epsilon_e, M_b(\mathcal{R}, \mathcal{D}) \leq \epsilon_b$$

Available online at

纵向联邦学习数据集

Dataset	Data Type	Size	Description
Income [140]	Tabular	48842	Demographics and income features
Bank [141]	Tabular	41188	Demographics and economic features
Credit Card [142]	Tabular	30000	Demographics and payments
Give Me Some Credit [143]	Tabular	250000	Debt features
MIMIC III [144]	Tabular	42276	Medical records
Breast Cancer [145]	Tabular	569	Breast tumor features
Diabetes [146]	Tabular	400	Patient records
Avazu [147]	Tabular	4M	Click-through data
Criteo [148]	Tabular	4.5M	Click-through data
Vehicle [149]	Tabular	98528	Acoustic and seismic signals
Drive [150]	Tabular	58509	Electric current drive signals
Cover type [151]	Tabular	581012	Digital spatial data
NUSWIDE [152]	Tabular	269648	Image and the associated tags from Flickr
Handwritten [153]	Tabular	2000	Handwritten digit features
Epsilon [154]	Tabular	500000	Synthetic data
BHI [155]	Image	277524	Medical images
CheXpert [156]	Image	65240	Medical images
Modelnet [157]	Image	20000	Multi-views of 3D objects
Cora [158]	Graph	2708/5429	Citation network
Citeseer [158]	Graph	3327/4732	Citation network
PubMed [158]	Graph	19717/44338	Citation network
Yahoo Answers [159]	Text	1.46M	Corpus of questions and answers
News20 [160]	Text	19928	Newsgroup documents

Available online at

应用领域

Major Applications

- 推荐系统-营销领域 **Recommendation systems and Advertising**
- 金融领域 **Finance**
- 医疗领域 **Healthcare**
- 通讯及其他领域 **Wireless Communication**

Open-Source Projects

- FATE
- PyVertical
- FedLearner
- FedML
- Fedtree
- PaddleFL

……

Available online at

挑战与展望

- 互联互通 **Interoperability**
- 可信联邦学习 **Trustworthy**
- 自动化 **Automated**
- 区块链结合 **Blockchain**