

隐私计算之联邦三部曲





傅致晖
算法专家

2015年毕业于上海交通大学，曾就职于众安科技从事深度学习相关研究和落地，2019年加入同盾科技人工智能研究院，目前主要从事隐私计算的算法研究和平台开发。





主要内容

- 时代背景和数据孤岛问题
- 打破数据孤岛的案例简介
- 隐私计算的多样技术方案
- 平台孤岛问题与FIRM体系

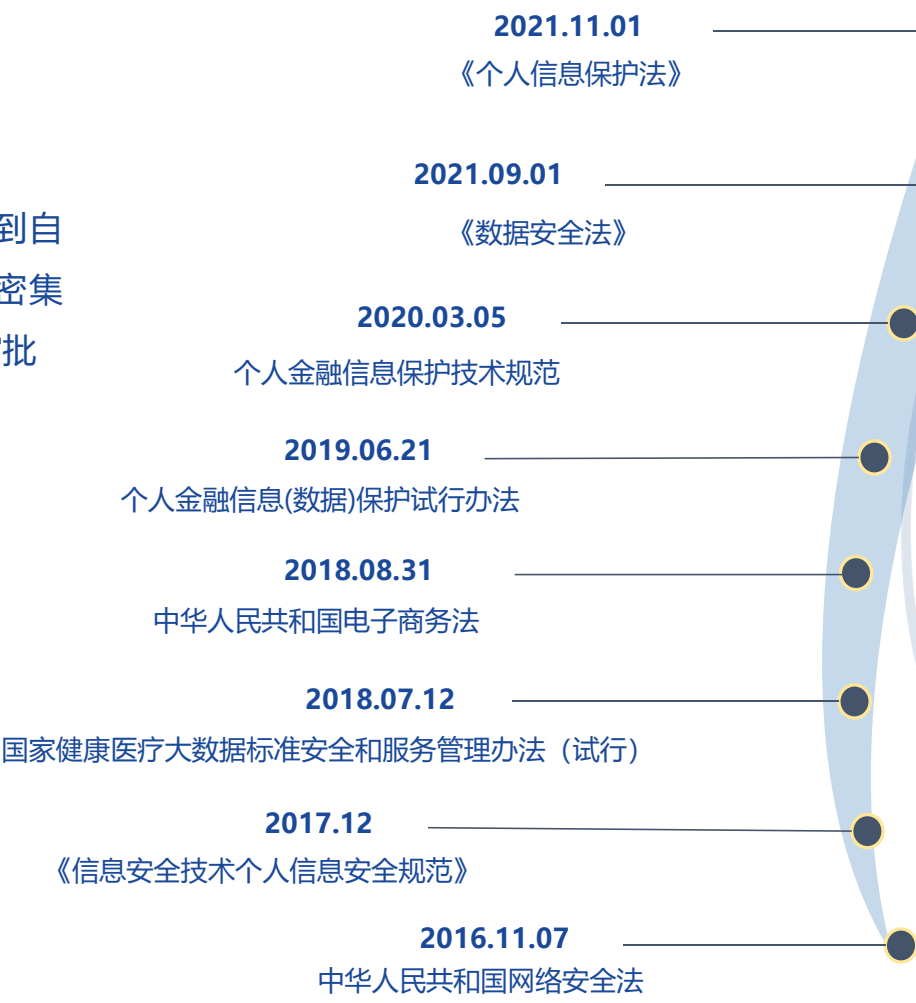


数据安全法规逐步完善

道德是上限，法律是底线。确保数据安全，立法是关键！

数据管控严格化、全面化

数据控制方责任明确，刑罚到自然人，各领域数据管理细则密集出台，用户授权+监管部门审批



数据安全法规进化史

数据成为新的生产要素

- **数据可用性**：即开放性。如何充分利用各方的数据，让数据对外开放，对决策服务有用
- **数据不可见性**：即不共享性。数据不离开各机构或个人，保证数据不对外直接共享

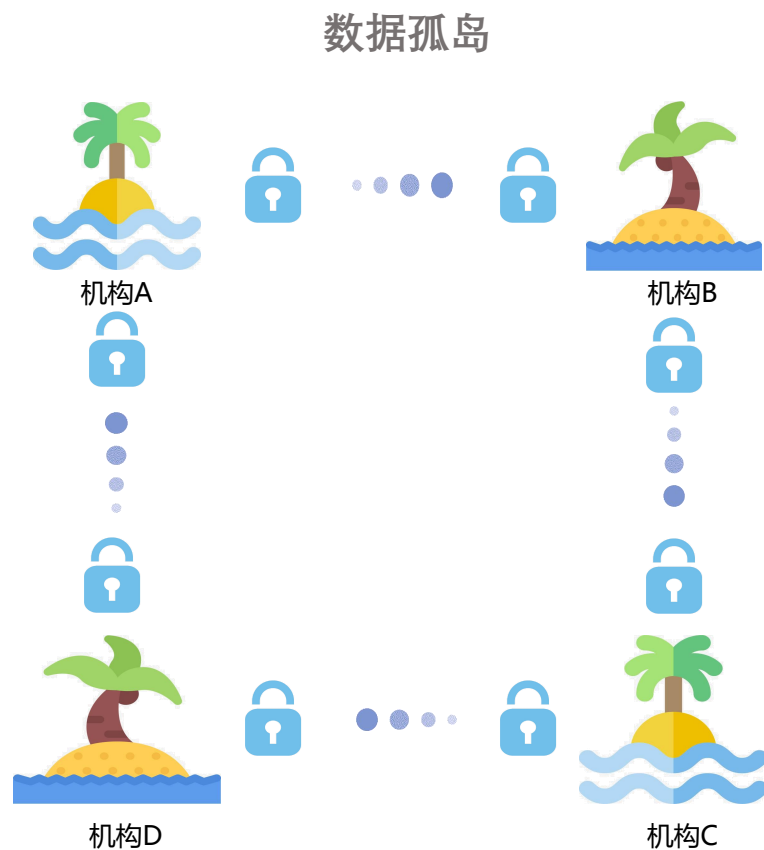
**数据资产化
资产价值化**

**数据被列为一种新型生产要素
与土地、劳动力、资本、技术等传统要素并列**

4月9日《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》

数据孤岛问题

所谓数据孤岛问题，指不同机构各自持有数据，且不能互相分享。然而，众所周知，建模精度几乎依赖于数据量及特征的丰富度和精确性。如何打通不同机构的数据，同时保护数据的隐私，是各种数据相关行业普遍遇到的问题。





打破数据孤岛的案例简介

- 保险金融联合营销
- 消费金融信贷风控
- 政务经济大数据融合



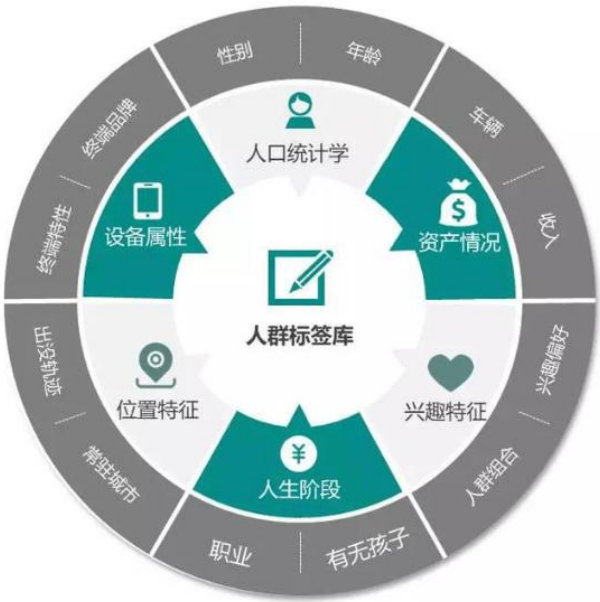
保险金融联合营销

业务目标

基于银行与保险公司有紧密业务关系，从银行的海量客群中挖掘潜在保险用户，有效提升银保营销转化率，提高保险销量产品和渠道的多样性。

解决方案

通过智邦平台联合银行保险客户，采用复合型联邦方式构建精准用户画像。



实施步骤

- 数据探查：探查两方数据，细粒度划分人群标签，定制衍生变量
- 联邦建模：通过两方部署节点实现亿级的训练和预测
- 场景策划：基于保险营销场景，精准推送高潜力客户



健康险场景

保险意识/保险消费能力/保险偏好



年金险场景

借贷情况/投资理财/信用风险



车险场景

汽车品牌/新车/二手车/车险偏好

结果验证

经验证，相比于单方模型，K-S 值提升 5.X%

对比方法	入模指标维度	K-S值
联邦学习模型	2XX	0.4X (提升5.X%)
保险单方模型	8X	0.3X

实际营销中，相比于联邦建模前，同渠道投放的营销转化率提升3倍。

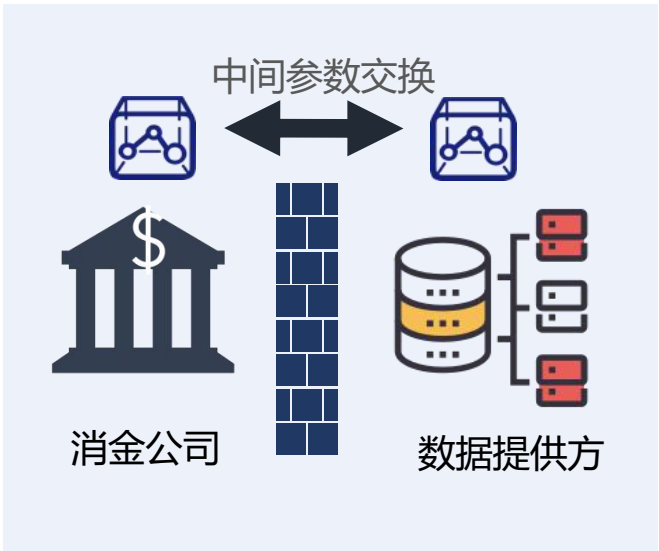
消费金融信贷风控

业务目标

需在确保安全的基础上，引入外部数据，提高消金公司个人信贷审核信用评估准确率，降低逾期风险。

解决方案

基于智邦平台，采用跨特征的联邦方式训练学习符合消金公司任务需求的信用评估模型，提升信用预测模型准确率。



实施步骤

通过数据聚合对齐用户2X万条，共补充特征维度到2XX维。



结果验证

经实验验证，联邦学习方法可使模型K-S值提升30%；部署上线后，该联邦模型月调用量约50万次，风控能力显著提升。

对比方法	入模指标维度	K-S值
联邦学习模型	12X	0.4X (30%)
消金单方模型	3X	0.2X

政务经济大数据融合

业务目标

在安全合规的基础上，解决地方大数据局因数据孤岛而对地方企业缺乏细粒度的了解，监控企业健康度与区域经济发展状况，从而实现政府部门可对地方经济状况的监控与对资源的统筹规划等目标。

解决方案

基于智邦平台的多方安全计算实现安全统计、分析，输出实时可见监测业务看板。
基于结果给政务可展示、可解读、可应用的数据解决方案，推进城市的数字化转型。



实施步骤

- 联通委办单位数据
搭设智邦平台，安全联合分析各委办单位的企业数据
- 联通互联网数据
打通本地互联网平台数据，丰富分析维度



结果验证

经实际验证，试点市在政务处理效率普遍可提高2*%以上。在疫情发生后，政府市场复产复工专项扶持项目，平台专项扶持款在一个半月内增长1**亿。





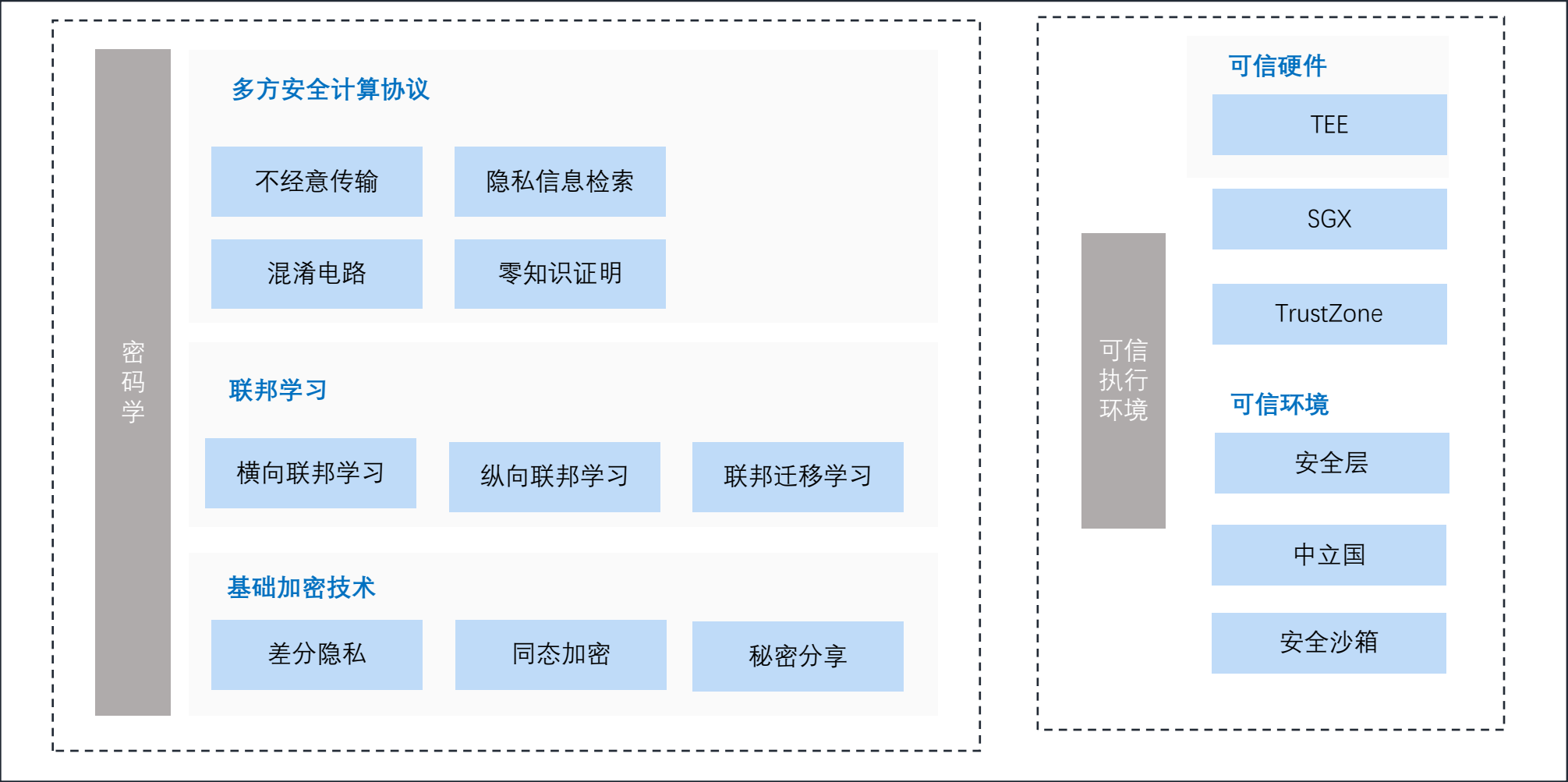
隐私计算的多样技术方案

- 隐私计算技术的组成
- 联邦学习介绍
- 纵向逻辑回归方案枚举



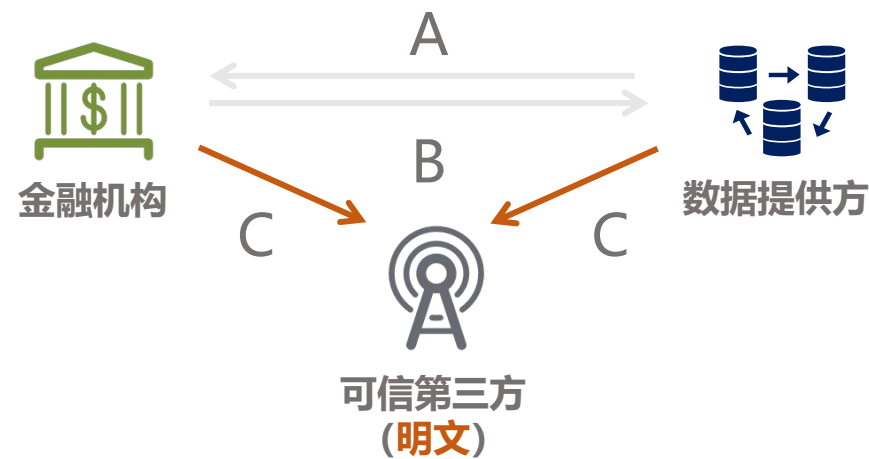
隐私计算技术组成

技术类别



联邦学习与传统建模差异

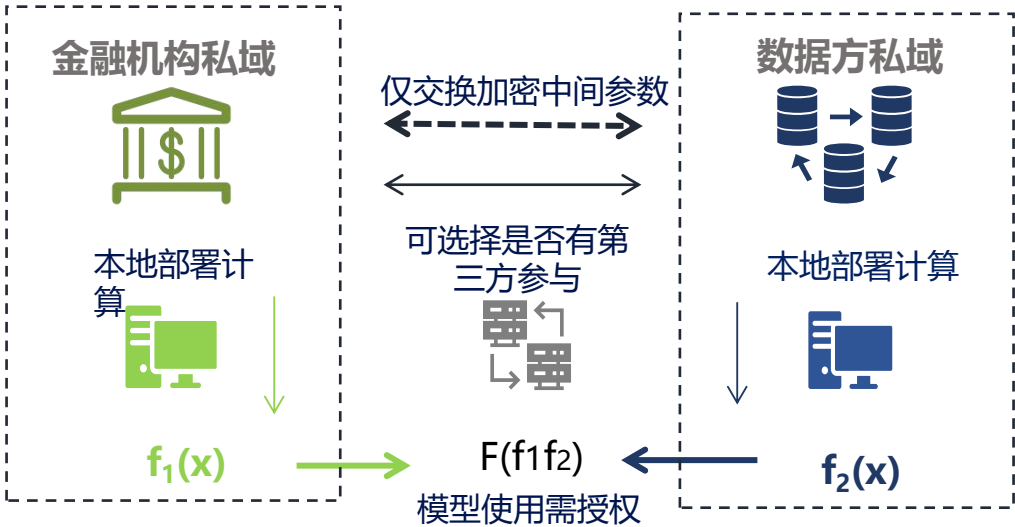
传统建模方式



传统建模方式弊病

成本高	安全性差	第三方难管控
<ul style="list-style-type: none">部署成本高人员成本高时间成本高	<ul style="list-style-type: none">金融机构易泄漏所查询用户信息风险不合规模型易暴露	<ul style="list-style-type: none">第三方可获得原始数据与模型难确保第三方安全难确保合规

联邦学习方式



联邦学习方式优势

成本低	安全性高	生态共创共享
<ul style="list-style-type: none">本地部署效率高，即插即用人员成本低	<ul style="list-style-type: none">数据不出域仅传输加密参数，确保安全各方仅有各自模型参数，防止模型泄露	<ul style="list-style-type: none">可选取有/无第三方两种模式采用多方安全计算方式模型碎片化存储

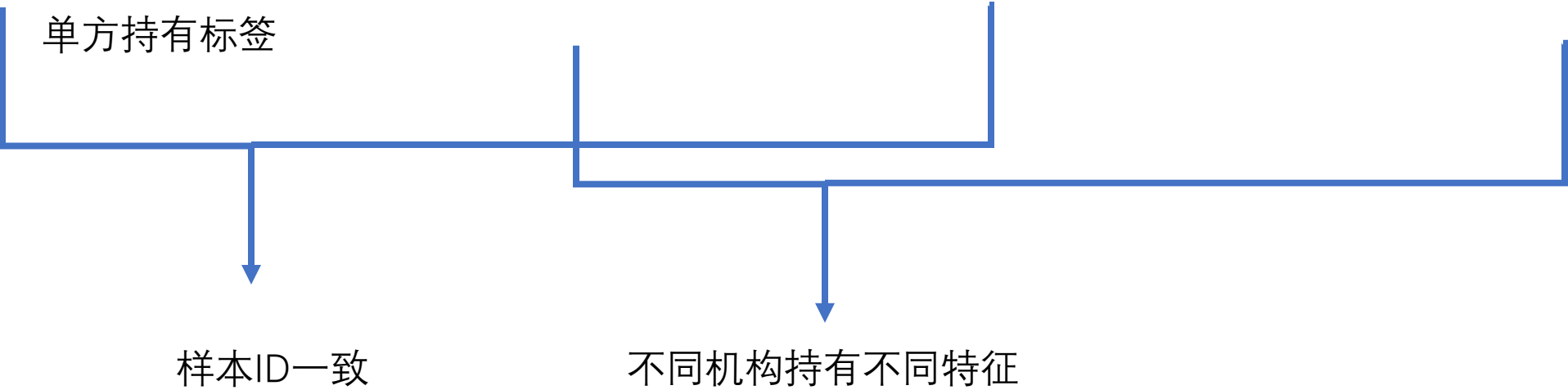
纵向数据分布

机构A

id	y	x0	x1
0	0	-0.9061604274807564	0.9472087306838816
1	0	-1.1530145734460011	-0.1871729827099886
2	1	-0.16559798958501856	-0.6409256680675369
3	1	0.4515373753280957	0.7203323880051073
4	1	-0.9061604274807564	-1.321554696103859
5	1	1.4389539591890783	0.2665797026475599
6	1	0.32811030234547217	-1.0946783534250846
7	1	2.179516397084815	-0.1871729827099886
8	0	-0.7827333544981329	2.308466786756525

机构B

id	x2	x3
0	-1.309829668289983	-1.2848585589733093
1	-1.309829668289983	-1.2848585589733093
2	0.22169257106748505	0.17345038156389506
3	0.9590921677951552	1.4991857820522625
4	-0.40226093385592776	-0.09169669853377836
5	0.5620308464802559	0.3060239216127316
6	1.0725382595994115	0.3060239216127316
7	1.6397687186206964	1.234038701954589
8	-1.2531066223878546	-1.4174320990221458



基础加密技术-加性同态加密和秘密分享

加性同态加密


$$\mathbf{D}(\mathbf{E}(a) + \mathbf{E}(b)) = a + b$$

加性秘密分享

	机构A		机构B	
5	=	-1234	+	1239
		(随机数)		

纵向逻辑回归方案枚举-基于同态加密的近似梯度计算

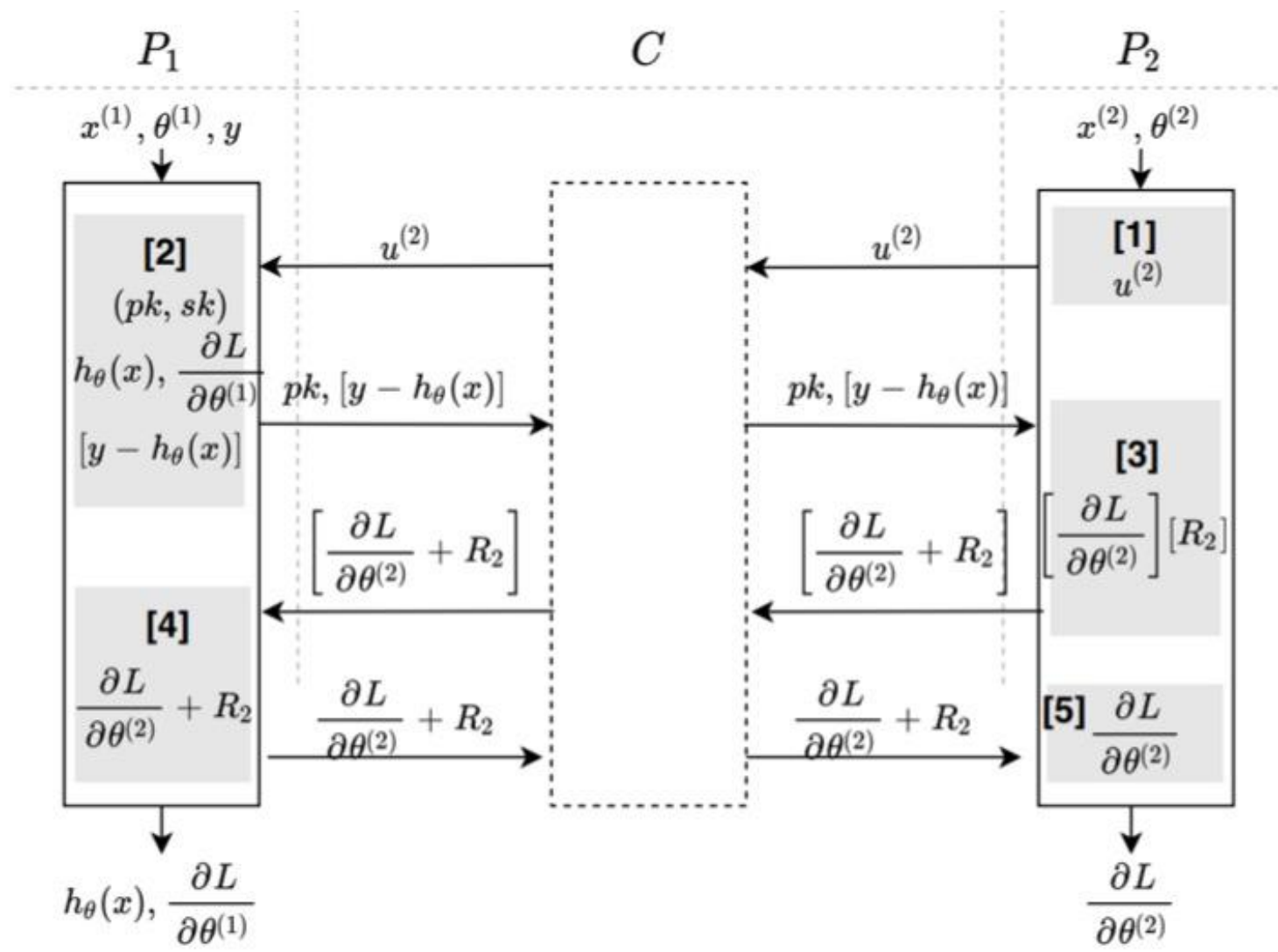
$$L(\theta) = \frac{1}{n} \sum^n \log(1 + e^{-y_i \theta^T x_i}) \approx \log 2 - \frac{1}{2} \mathbf{y} \theta^T \mathbf{x} + \frac{1}{8} (\theta^T \mathbf{x})^2$$



$$\nabla L(\theta_k) = (\sum_k \frac{1}{4} \theta_k^T \mathbf{x}_k - \frac{1}{2} \mathbf{y}) \mathbf{x}_k$$

	Participant A	Participant B	Participant C
Step1	Generate the encryption key pair $(Q, (s_A, s_B, s_C))$		
Step2	initialize model parameters θ_A	initialize model parameters θ_B	initialize model parameters θ_C
Step3	compute u_A , encrypt to $[u_A]_{ABC}$	compute u_B , encrypt to $[u_B]_{ABC}$	compute u_C , encrypt to $[u_C]_{ABC}$
Step4	Calculate $[w]_{ABC} = \sum_k [u_k]_{ABC}$ and $[\nabla L(\theta_k)]_{ABC} = [w]_{ABC} \mathbf{x}_k$		
Step5	send $[\nabla L(\theta_A)]_{ABC}$ to B, get $[\nabla L(\theta_C)]_{ABC}$ and partially decrypt it to $[\nabla L(\theta_C)]_{BC}$	send $[\nabla L(\theta_B)]_{ABC}$ to C, get $[\nabla L(\theta_A)]_{ABC}$ and partially decrypt it to $[\nabla L(\theta_A)]_{AC}$	send $[\nabla L(\theta_C)]_{ABC}$ to A, get $[\nabla L(\theta_B)]_{ABC}$ and partially decrypt it to $[\nabla L(\theta_B)]_{AB}$
Step6	send $[\nabla L(\theta_C)]_{BC}$ to B, get $[\nabla L(\theta_B)]_{AB}$ and partially decrypt it to $[\nabla L(\theta_B)]_B$	send $[\nabla L(\theta_A)]_{AC}$ to C, get $[\nabla L(\theta_C)]_{BC}$ and partially decrypt it to $[\nabla L(\theta_C)]_C$	send $[\nabla L(\theta_B)]_{AB}$ to A, get $[\nabla L(\theta_A)]_{AC}$ and partially decrypt it to $[\nabla L(\theta_A)]_A$
Step7	send $[\nabla L(\theta_B)]_B$ to B, get $[\nabla L(\theta_A)]_A$ and finally decrypt it to $\nabla L(\theta_A)$	send $[\nabla L(\theta_C)]_C$ to C, get $[\nabla L(\theta_B)]_B$ and finally decrypt it to $\nabla L(\theta_B)$	send $[\nabla L(\theta_A)]_A$ to A, get $[\nabla L(\theta_C)]_C$ and finally decrypt it to $\nabla L(\theta_C)$
Step8	update θ_A	update θ_B	update θ_C
obtain	θ_A	θ_B	θ_C

纵向逻辑回归方案枚举-假设内积安全的方案





新的孤岛问题与FIRM体系

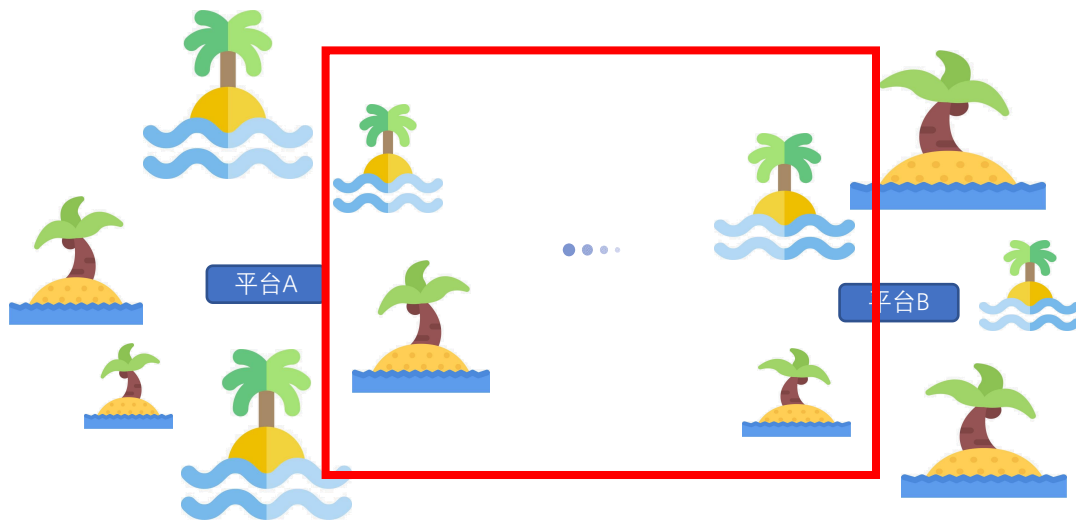
- 新的孤岛-平台孤岛
- FIRM体系的理论基础-互联互通的逐层解藕
- 数据交换层FLEX
- 应用层与算法层互通



新的孤岛-平台孤岛

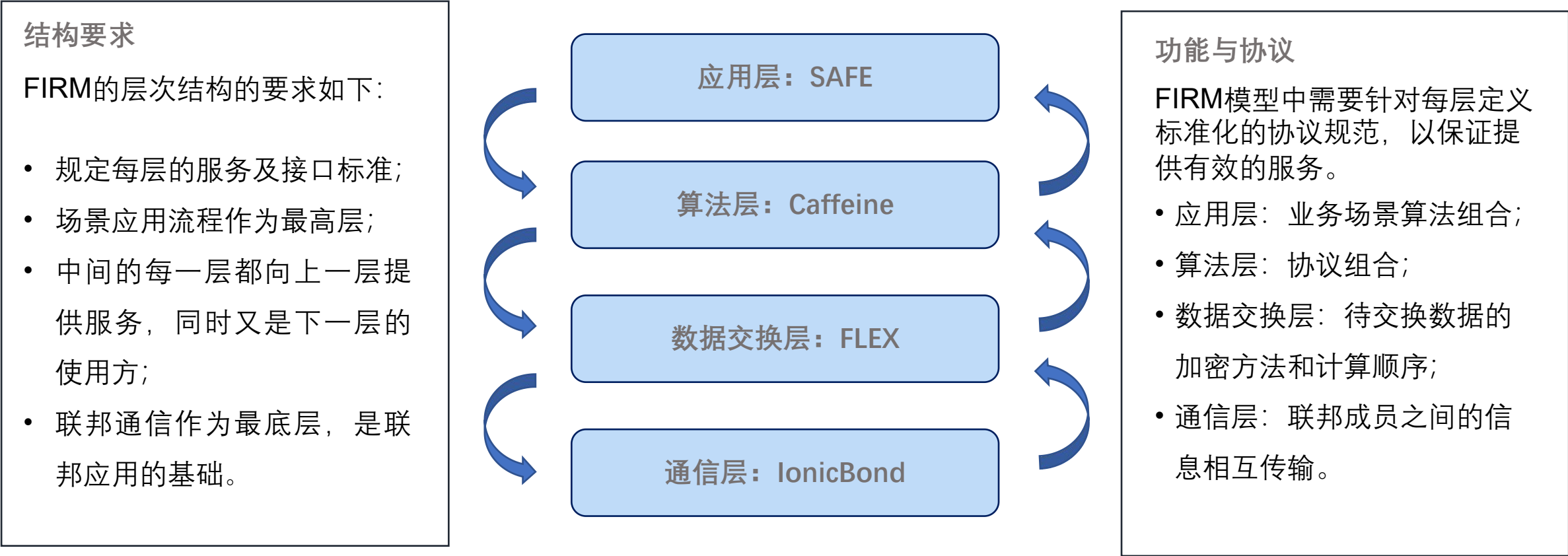
随着隐私计算技术的创新和隐私计算产业的成长，数据孤岛问题逐渐解决，然后新的孤岛问题：平台孤岛出现了；既采用不同联邦学习平台实现的机构之间难以互通，真正意义上的数据互联、互通、共享仍未实现。

平台孤岛



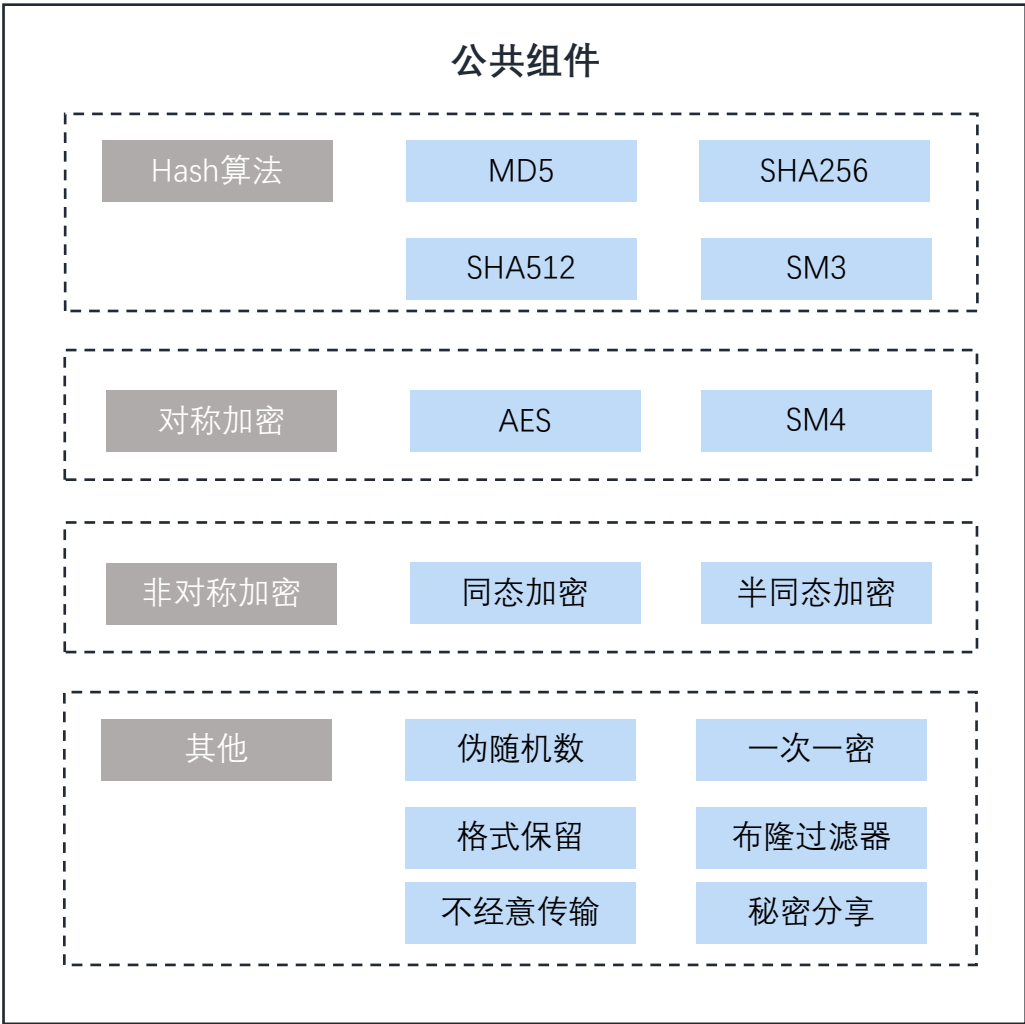
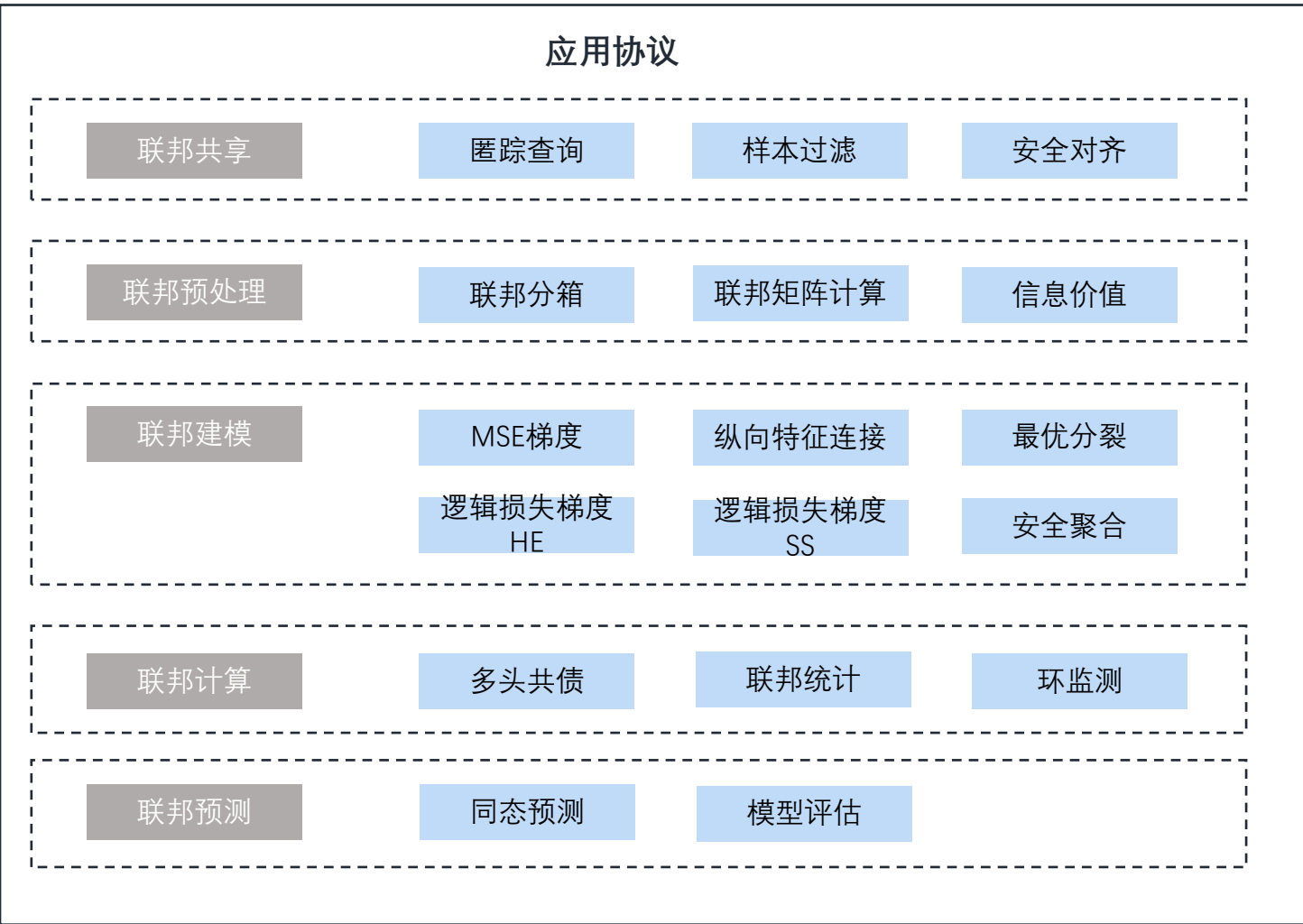
FIRM体系理论基础-互联互通的逐层解藕

同盾人工智能研究院首次提出了“开放联邦系统互联参考模型”，即FIRM (open Federated system Interconnection/Reference Model)。它将联邦系统互联互通协议分为四层：通信层、数据交换层、算法层和应用层。



数据交换层FLEX

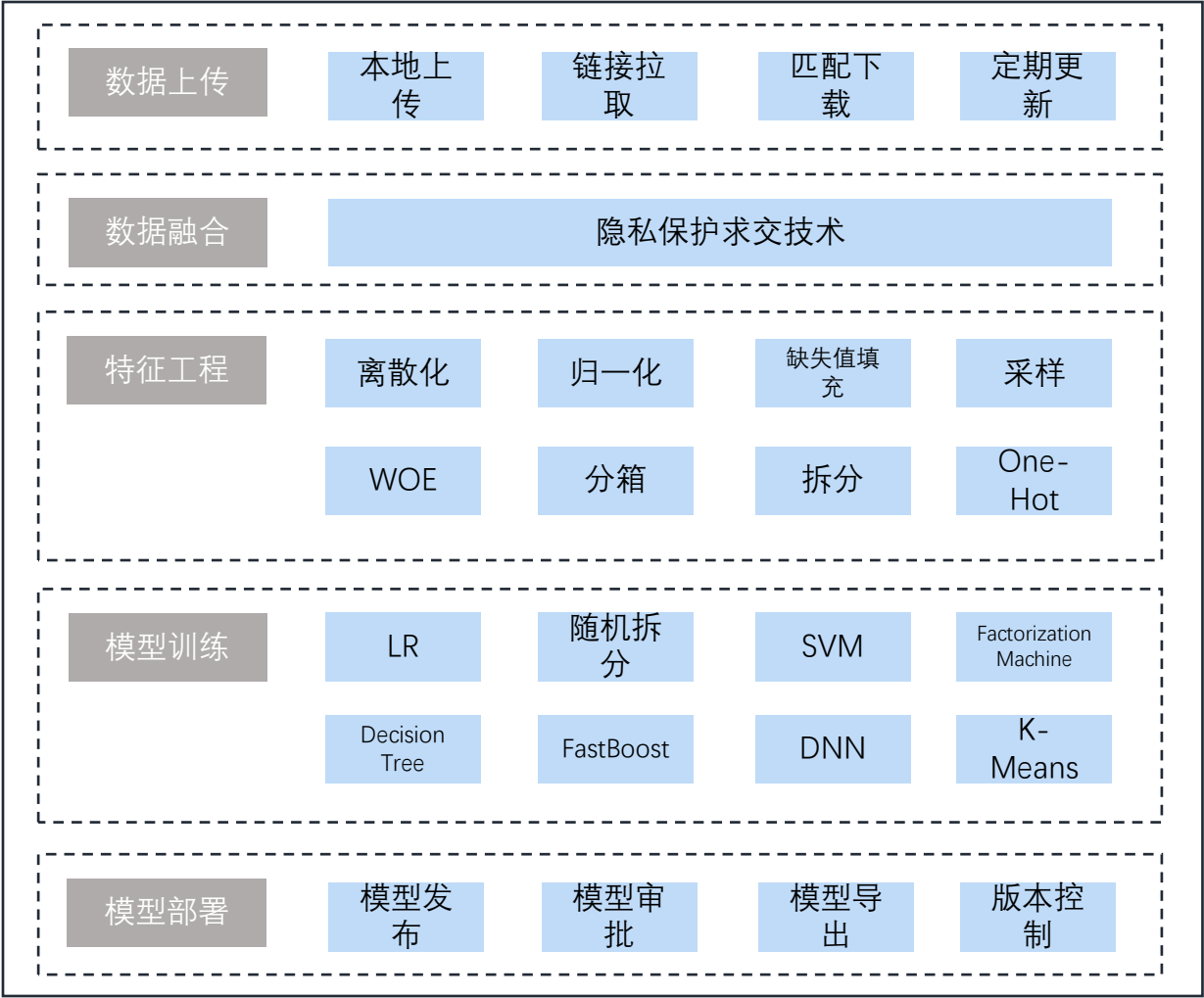
FLEX协议



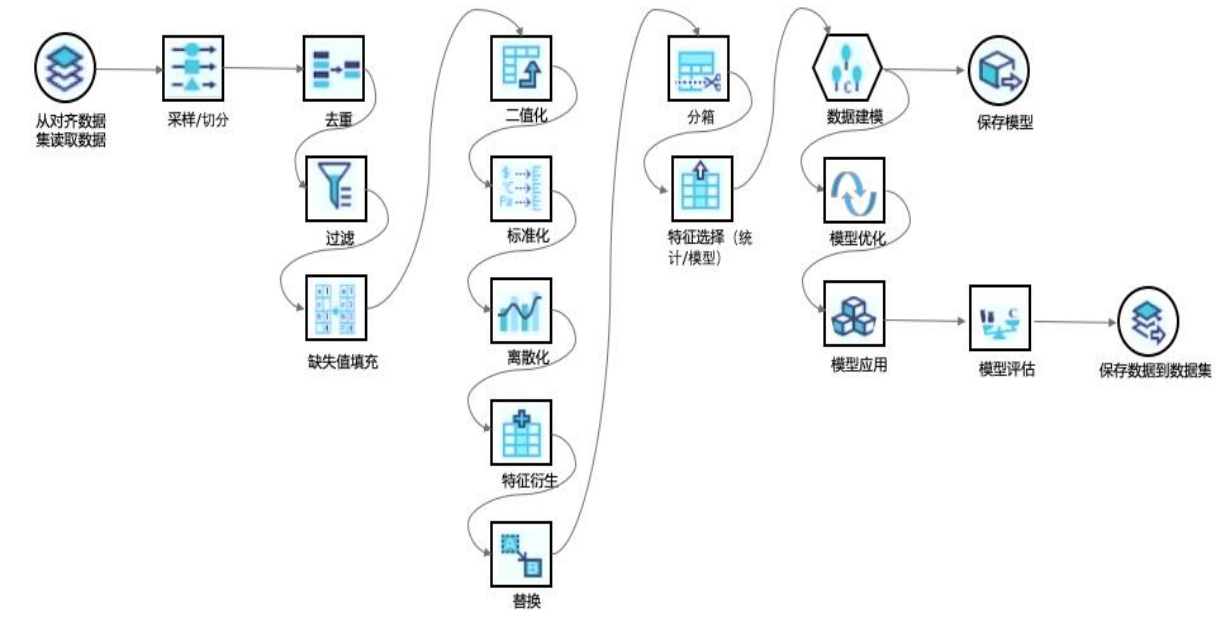
协议白皮书和参考实现开源地址：<https://github.com/tongdun/iBond-flex>

算法层与应用层互通

- 算法层：提供了数据建模各模块所需调用的算法集合，可根据业务应用场景自由选取调用。
- 应用层：基于业务经验，在营销，风控等多个领域也定制了标准流程模版，保证模型效果的同时，实现了快速交付落地。



以评分卡调用流程为例，应用层通过配置业务流，可从算法层的算子库（如：Caffeine）中快速选取建模各部分所需算子。流程全自动化，方便，简易，快捷。





Take Home Message

- 数据管控越来越严格化、全面化，数据孤岛问题愈发严重
- 我们可以使用隐私保护技术打破数据孤岛
- 隐私保护技术具有多样的实现方式
- 多样实现方式会引起平台孤岛问题
- 可以利用FIRM分层体系应对平台孤岛问题





谢谢!





关注msup公众号
获取更多AI落地实践

麦思博(msup)有限公司是一家面向技术型企业的培训咨询机构，携手2000余位中外客座导师，服务于技术团队的能力提升、软件工程效能和产品创新迭代，超过3000余家企业续约学习，是科技领域占有率第1的客座导师品牌，msup以整合全球领先经验实践为己任，为中国产业快速发展提供智库。