

Transwarp Sophon白皮书

Transwarp Sophon是一款通用的人工智能平台,以帮助企业级用户快速实现人工智能的落地为目的。对于大多数企业用户,要通过自己的力量从无到有打造AI人工智能平台的代价是巨大的,需要有技术实力强大的大数据团队和AI团队作为基础。而Sophon打通了大数据平台和人工智能平台,业务分析师和数据分析师可以通过自动建模以及内置的行业模板轻松构建对应AI模型,从而提升业务价值。

主要功能

Sophon主要具有以下的特点:一站式图形化人工智能开发环境、自动建模的能力、高性能的分布式机器学习算法、高效率的交互式体验、完善的深度学习支持以及大量行业应用模板。

- 一站式图形化人工智能开发环境。整个流程都可以通过拖拽式操作包括:数据导入、数据探索与预览、数据预处理、特征工程、算法选择、模型训练、模型发布、模型管理。

1. **数据导入**:支持多种数据源。用户可以通过多种方式导入数据,除了跟Transwarp TDH做深度对接以外,还支持SQL数据库、其他Hadoop集群、多种开源格式,以及目前流行的云存储格式。

2. **数据预览与探索**:丰富的可视化展示和统计分析。用户可以对数据进行多种可视化展示,并进行数据质量和数据特征分析,从而为后续的数据预处理和特征工程做准备。

3. **数据预处理**:完善和多样的预处理功能。Sophon提供了多达50余种算子,包括但不限于数据清洗、数据规约、数据转换等。

4. **特征工程**:灵活的特征提取手段。Sophon提供了多种特征提取和特征转换方法。

5. **算法选择**:高性能的分布式算法。用户可以选择单机和分布式算法,包括常见的分类、回归、聚类、推荐、时序、统计等机器学习算法,也包括多种经典的神经网络,以及NLP和图像相关领域的算法。同时, Sophon通过与SQL流式引擎Transwarp Slipstream结合提供流式机器学习。

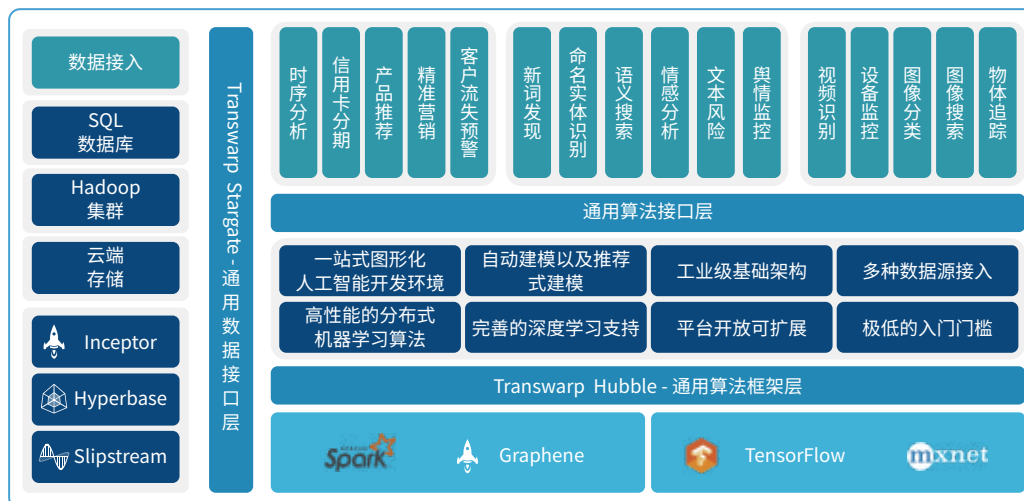
6. **模型训练**:高效的训练方式和多种模型评估手段。Sophon提供了多种底层运行平台,通过Kubernetes+Docker的模式进行资源管控空和调度。对于训练好的模型,用户可以通过多种指标和模型展示方式进行评估,从而得到最优模型。

7. **模型发布和模型管理**:强大的模型发布和管理功能。用户可以选择满意的模型进行发布,既可以将模型通过PMML格式导出,也可以将其API服务进行在线部署,或者通过工作流调度的方式进行定时调度。训练好的模型可以版本化管理。

- 自动建模的能力。Sophon提供了一键式自动建模以及推荐式自动建模两种方式。一键式自动建模就是Sophon平台通过内置的智能化算法进行数据类型检测、数据预处理、特征选择、模型选择以及参数调优。用户可以不做任何处理就训练出一个精度较高的模型。对于精度要求更高的场景，用户可以使用推荐式自动建模。Sophon通过多行业类似场景的模型积累，以及对当前数据和场景的深入挖掘，自动推荐了合适的算子作为用户下一步可选择的算子，从而通过人工和自动相结合的方式进一步提升模型精度。
- 高性能的分布式机器学习算法。支持100多种分布式算法，对于常见的机器学习算法，Sophon中对应算法的性能相对开源实现有3到10倍的提升。Sophon还支持几种常见算法的流式预测，与业界最强大的SQL流式引擎Transwarp Slipstream结合可以支持流式机器学习。此外，用户可以通过Python、Scala等语言进行自定义算子的封装，轻松扩展现有算法。

算法分类	解释
分类	除了常见的分类算法，还额外整合lightgbm、xgboost等框架
聚类	对常用的KMeans算法，相对于开源实现有接近5倍的性能提升
回归	与分类算法类似，额外整合lightgbm、xgboost等框架
推荐	目前包括因子分解机和最小二乘，算法持续添加中
统计	包括常见的一元二元指标统计
关联规则	高效的分布式实现
深度学习	常见的网络层级，以及多种经典网络的整合
NLP	覆盖了NLP相关的几乎所有操作
图像	支持基本的图像处理，算法持续添加中
时序分析	支持多种常见的时序分析方法
预处理	支持PCA、SVD、归一化、离散化等算法

- 高效率的交互式体验。用户通过Sophon能快速进行数据探索、迭代建模过程，大大缩短建模周期，模型能更及时地在生产环境进行A/B测试。得益于Sophon的易用性，公司多部门之间可以通过统一的人工智能平台进行协助，提供企业级支持包括：多租户、权限管理、安全控制、团队协作、模型共享、分布式调度等。
- 完善的深度学习支持。用户可以通过拖拽的方式直接构造复杂的神经网络。构造好的神经网络层次分明，结构清晰，非常易于理解。Sophon通过Hubble抽象出统一的算法框架，同时对接Tensorflow和MXNet，保证与复杂的预处理算子无缝连接。此外，Sophon在分布式GPU算法上进行了针对性优化，有一倍的性能提升。Sophon还内置了多种经典的神经网络结构，其中包括将深度学习和广度学习结合起来的深广模型。
- 大量的行业模板。在普通的机器学习领域，Sophon中集成了如证券时序分析、信用卡分期、金融产品推荐、精准营销、客户流失预警等模板。对于NLP和图像两个深度学习领域，Sophon也集成常见的解决方案。在NLP领域，Sophon支持新闻新词发现、命名实体识别、海量数据的语义搜索、舆情监控、企业文本风险管理、问答系统等。而在图像领域，Sophon支持人脸识别、车牌识别、图像分类、图像搜索、物体检测等。



技术特点

Sophon主要有八大技术特点。

- 易用性。通过一站式的界面操作，普通的数据分析师和业务人员能够迅速上手机器学习。即使对于资深的数据科学家，高效率的交互式体验也使得模型精度提升的周期缩短。在设计和实现任何功能的时候，研发团队始终以平台的易用性作为最高的指导目标。
- 智能化。Sophon基于平台的数据和模型积累，自身具备了人工智能的能力，通过全自动建模和推荐式建模等智能化建模方式，大大降低用户的建模难度。
- 多数据源接入。用户的底层平台可以选择使用Transwarp TDH，同时也可以自由地选择其他的数据源。数据源接入之后会统一的视图和规范，从而保证用户的精力能集中在建模上，而不必浪费到数据平台之间的数据整合上。
- 平台开放性。用户可以通过Python或者Scala轻松扩展算法，同时对于整合其他技术或者框架，Sophon提供了通用的算法框架层和算法接口层，用户可以轻松接入。
- 图形化和可视化。Sophon将可视化做到极致，无论是建模过程的拖拽，还是数据预览、模型评估和管理。
- 高性能。Sophon针对开源的算法实现有多达3到10倍的性能优化，对分布式GPU有额外并行化优化。
- 企业级特性。相对于个人用户，企业用户在考虑产品会更多地考虑产品是否支持多租户，是否有权限管理和安全控制，是否支持团队间协作，能否进行模型共享，如何做好资源管理等问题。而Sophon能出色地解决这些问题。
- 预制模板。对于大多数用户来说，平台提供的预制模板至关重要，这能帮助他们快速开展业务，而不需要长时间的培训和学习。

适用场景

基于星环在多个行业的数据挖掘经验的积累，Sophon适用的场景包括但不限于：

金融行业：信用风险、客户生命价值、商圈分析、用户画像、智能投顾、风险控制、量化投资、实时推荐、流失预警、问答系统、语义搜索、知识图谱、图像识别

电信行业：垃圾短信检测、套餐推荐、智能网络优化、用户行为分析、精准营销、用户流失分析

电商行业：用户行为分析、用户画像、产品推荐

工业行业：生产设备故障检测、智能可靠性维护、设备安全监控

交通能源行业：车牌识别、图像识别、物体追踪、设备安全监控

政府机构：文本挖掘、舆情分析、新词发现



◀关于我们:

星环科技是全球领先的大数据与人工智能基础平台供应商,专注于提供企业级容器云计算、大数据和人工智能核心平台的研发和服务,打造大数据和人工智能技术的“中国心”。公司以上海为总部,以北京、广州为区域总部,在南京、郑州、成都设有支持中心,同时在深圳、天津、武汉等地设有办事机构,并在加拿大和美国设有子公司。经过多年自主研发,星环科技建立了五大产品系列:基于容器的智能大数据云平台Transwarp Data Cloud (TDC)、一站式大数据平台Transwarp Data Hub (TDH)、量子人工智能平台Transwarp Sophon、数据库(分析型数据库ArgoDB,图数据库StellarDB,交易型数据库KunDB)和超融合大数据一体机TxData Appliance,并拥有多项专利技术。2016年被Gartner评为全球最具有前瞻性的数据仓库及数据管理解决方案厂商,2017年被IDC评为中国大数据市场领导者。公司产品已经在十多个行业应用落地,是国内落地案例最多的大数据和人工智能平台供应商。目前星环科技已完成C轮融资,由腾讯领投。

◀核心技术:

基于Hadoop构建企业级数据仓库和数据集市
高性能、可扩展的分布式数据库
融合低延时的事件驱动机制和复杂批处理编程模型的流处理引擎
具备统计、机器学习和深度学习完备算法和实践模型的人工智能平台
基于Docker和Kubernetes的多租户PaaS平台

◀应用行业:

金融、电信、交通、物流、政府、公共安全、媒体、电力、能源、零售、制造业、医疗、教育等。

◀部分用户:



☎ 电话: 4008-079-976 北京分公司: 010-62682815
🌐 网址: www.transwarp.io
@ 咨询: sales@transwarp.io 技术支持: support@transwarp.io
📍 上海: 徐汇区虹漕路88号B座11F&12F
📍 北京: 海淀区西直门北大街甲43号金运大厦B座1101室
📍 广州: 天河区体育东路140-148号南方证券大厦1015-1016室

