T/EJCCCSE

ICS 35.240.01

CCS L 70

团 体 标 准

2024-04-27实施

2024-03-27发布

发 布

中国商业股份制企业经济联合会

人工智能大规模画像及推荐系统技术规范

Technical specifications for large-scale artificial intelligence portrait and recommendation systems

（征求意见稿）

T/EJCCCSE 061-2024

目 次

前 言 II

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

4 缩略语 3

5 系统设计要求 4

6 推荐系统框架 7

7 技术要求 8

8 测试方法 12

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由深圳市新风向科技股份有限公司提出。

本文件由中国商业股份制企业经济联合会归口。

本文件起草单位：深圳市新风向科技股份有限公司。

本文件主要起草人：×××

人工智能大规模画像及推荐系统技术规范

1. 范围

本文件规定了人工智能大规模画像及推荐系统的术语和定义、缩略语、系统设计要求、推荐系统框架、技术要求和测试方法。

本文件适用于人工智能大规模画像及推荐系统。

1. 规范性引用文件

本文件没有规范性引用文件。

1. 术语和定义

下列术语和定义适用于本文件。

 大规模画像

指通过收集、整合、分析用户多维度数据，构建出全面反映用户特征、兴趣、需求等信息的综合模型。

 推荐系统

指基于用户画像和物品信息，通过一系列算法为用户精准推荐可能感兴趣的内容或服务的智能化系统。

 准确率 Accuracy

正确推荐的物品数量占总推荐物品数量的比例。

召回率 Recall

正确推荐的物品数量占用户实际感兴趣物品数量的比例。

F1值

准确率和召回率的调和平均数，用于综合评估推荐性能。

1. 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

API ：应用程序接口（Application Programming Interface）

NLP：自然语言理解（Nature Language Process）

1. 系统设计要求
	1. 设计原则
		1. 系统稳定性原则
			1. 系统应确保24/7稳定运行，即在高负载、大数据量的情况下也能保持性能稳定。
			2. 系统应具备故障恢复能力，能够在出现故障时快速恢复服务。
		2. 数据安全性原则

用户数据是系统的核心资产，必须采取严格的数据加密、访问控制和安全审计措施，确保数据不被非法获取、篡改或泄露。同时，应遵守相关法律法规，对用户数据进行合规处理。

* + 1. 用户友好性原则

系统的用户界面应简洁、直观、易于操作，以满足不同用户群体的需求。此外，系统应提供清晰的反馈和错误信息，帮助用户快速解决问题。

* + 1. 高效处理性原则

系统应能够高效处理大规模的用户画像和推荐任务，具备强大的计算能力和处理能力。通过优化算法和硬件资源分配，实现快速响应和低延迟。

* + 1. 可扩展性原则

随着用户量和数据量的增长，系统应能够无缝扩展，以应对未来可能的增长需求。这包括硬件资源的扩展、算法的优化以及系统架构的调整等。

* + 1. 可维护性原则

系统应具有良好的模块化和结构化设计，方便后期维护和升级。同时，应提供详细的日志和监控功能，帮助运维人员快速定位问题并进行修复。

* + 1. 技术创新性原则

系统设计应充分考虑当前和未来的技术发展趋势，采用前沿的算法和技术手段，确保系统具有竞争力和生命力。通过不断的技术创新，提升系统的性能和用户体验。

* 1. 系统基础架构设计
		1. 微服务架构

推荐系统宜采用微服务架构，将系统拆分为多个独立的微服务，每个微服务负责特定的功能或业务逻辑。云原生由微服务架构、DevOps和以容器为代表的敏捷基础架构组成，实现微服务的动态管理和自动扩展。架构图见图1。

 

1. 微服务架构图
	* 1. 容器化技术
			1. 宜采用Docker等容器化技术，将微服务打包为可移植的容器镜像。
			2. 宜使用Kubernetes等容器编排工具，实现容器的自动化部署、管理和扩展。
		2. API网关
			1. 构建统一的API网关，作为外部访问系统的入口点。
			2. 在API网关中实现认证授权、流量控制、请求路由等功能，提高系统的灵活性和安全性。
		3. 基础架构功能
			1. 应用管理
				1. 概述

用于将项目生成的镜像部署到测试环境或生产环境上，镜像升级可自动触发升级，对于运行的镜像可停止、扩缩容、升级回滚等操作。

* + - * 1. 应用管理

应用是一组服务的集合，集群采用Kubernetes作为分布式容器管理工具，用户可将应用的镜像部署到生产或测试环境中。用户可管理正在运行的业务，进行升级回滚、扩容缩容、日志查看、异常恢复、启动停止等操作。具体功能包括：

1. 新建应用：新建应用信息，并添加成员的权限；
2. 删除应用：删除已定义的应用；
3. 应用详情：点击详情进入应用管理界面，可新增应用的服务以及管理应用的部署配置和运维操作。
	* + - 1. 服务管理

列出所有应用的服务，并对服务的部署配置进行管理。具体功能包括：

1. 新增服务：新增服务信息，填写部署到集群环境的K8S部署配置以及负载均衡信息；
2. 删除服务：修改已创建的服务信息以及部署配置信息；
3. 重启服务：重启正在运行的服务；
4. 停用流服务：停用正在运行的服务；
5. 启用服务：启用停用的服务；
6. 升级/回滚：镜像升级可触发自动升级，升级后可回滚到指定旧版本的配置。修改部署配置后保存就对部署配置版本升级一个新版本，会提示用户是否升级；
7. 扩容/缩容：正在运行的服务可点击扩容缩容按钮对实例进行手动伸缩服务的实例数量。填写伸缩实例的数量即可完成伸缩功能；
8. 自动伸缩：填写是否开启自动伸缩的开关以及CPU或内存的阈值，当业务访问需求变化触发阈值则进行实例的自动伸缩以满足业务需求；
9. 部署详情：展示部署的信息，列表显示部署触发的事件以及错误信息。
	* + - 1. 容器管理

管理所有服务部署产生的所有容器实例。应能查看容器日志以及采用web界面ssh登录到容器的控制台。具体功能包括：

1. 日志查看：查看容器内服务产生的日志；
2. 控制台：点击控制台则以root身份ssh登录到容器的控制台查看容器内服务状态。
	* + 1. 运维管理
				1. 集群管理

管理多个K8S集群，对于拥有developer权限以上的用户，可为该集群添加主机并设置主机的工作场景、主机标签、是否用于构建等信息。具体功能包括：

1. 新增集群：新增集群信息，填写K8S集群的APISERVER、etcd、dns、日志开启等信息；
2. 集群设置：对添加的集群信息进行查看和修改；
3. 删除集群：删除已定义的集群信息；
4. 命名空间：添加集群命名空间，不同团队可以拥有不同的命名空间，不同命名空间的服务名、部署名、POD名称可以相同；
5. 集群成员：添加成员并设置集群成员的权限。
	* + - 1. 主机管理

维护集群中的节点，可添加K8S节点以及管理节点标签。具体功能包括：

1. 添加主机：通过选择主机类型是centos或ubuntu类型，将生成的命令拷贝到主机上执行即添加一个K8S的节点；
2. 标签设置：管理该主机的标签，可添加、删除用于该主机的标签，用于服务部署时编排到该主机；
3. 主机设置：设置该主机的工作环境是测试环境还是正式生产环境以及是否用于CI构建镜像；
4. 实例列表：列出运行于该主机的容器实例，可以查看该实例的日志和ssh进日控制台查看服务状态。
	* + - 1. 告警管理

告警管理用于添加报警策略，用户可以自定义主机组以及用户组，当主机组满足告警策略定义的阈值时，发送邮件或短信给用户组中定义的用户。具体功能包括：

1. 新建模板：新建告警策略模板，用户可选择主机或容器实例进行监控，设置告警逻辑，当告警产生时发送的邮件模板和用户组；
2. 添加主机组：添加运用于告警模板的主机组合。对不同的主机组可设置不同的告警策略模板；
3. 添加用户组：添加接收告警邮件或短信的用户组；
4. 未恢复告警：列表显示未恢复的告警，只有master权限才可以删除未恢复的告警。
5. 推荐系统框架
	1. 概述

人工智能大规模画像及推荐系统的技术架构应采用分层设计，接口采用HTTP标准接口协议；系统开发语言采用主流开发语言及成熟的数据库及操作系统等中间件，从上往下可分为应用层、技术服务层、算法层、数据层、基础层。技术架构图如图 2 所示。



1. 技术架构图
	1. 应用层

应用层主要负责与用户进行交互，开发多种交互方式的用户接口，如移动应用、网页端等，实现推荐结果的展示和用户反馈的收集。具体包括：

1. 开发移动应用、网页端等多端交互界面，支持用户登录、浏览、搜索、购买等操作；
2. 集成推荐结果展示组件，根据用户画像和推荐算法生成个性化推荐列表；
3. 实现用户反馈收集机制，包括评分、评论、点赞等功能，用于优化推荐效果。
	1. 技术服务层

技术服务层作为构建服务框架，包括推荐服务、搜索服务、用户交互服务等，提供统一的API接口，实现服务的调用和管理。具体包括：

1. 构建统一的API网关，提供服务调用、认证授权、流量控制等功能；
2. 实现服务治理机制，包括服务注册与发现、负载均衡、熔断降级等，确保服务的稳定性和可用性；
3. 集成日志收集与分析系统，用于监控和诊断服务运行状况。
	1. 算法层

算法层利用机器学习、深度学习等技术对用户数据和物品信息进行处理和分析，生成个性化的推荐结果。算法层需要包含多种算法模型，如基于用户的推荐、基于物品的推荐等，并能够根据数据特点和业务需求进行选择和调整。具体包括：

1. 集成多种推荐算法库，如TensorFlow、PyTorch等，为推荐系统提供算法支持；
2. 实现算法调度与配置管理功能，根据业务场景和数据特点选择合适的算法模型；
3. 搭建算法训练与评估平台，支持模型的离线训练、在线评估和性能调优。
	1. 数据层

数据层负责数据的收集、存储、处理和分析。应建立分布式数据仓库，实现用户数据、物品数据的存储和管理。采用ETL工具进行数据清洗和整合，为算法层提供高质量的数据支持。具体包括：

1. 部署分布式数据仓库，如Hadoop、HBase等，用于存储用户数据、物品数据等；
2. 实现数据ETL流程，包括数据采集、清洗、整合和加载，确保数据的准确性和一致性；
3. 构建数据服务接口，为上层提供数据查询和分析功能。
	1. 基础层

基础层包括支持通用计算单元或异构计算单元的硬件、网络设备和存储设备等计算资源构成的计算资源池。具体包括：

1. 利用云计算平台（如SaaS、PasS、IaaS等）提供的计算和存储资源，部署虚拟机或容器化服务；
2. 配置网络和安全设施，确保系统的网络连通性和安全性。
3. 技术要求

应用流程人工智能大规模画像及推荐系统的流程见图2。



1. 应用流程图
	1. 数据采集
		1. 数据收集
			1. 数据源确定与采集方式
				1. 应根据业务需求，确定所需数据类型，包括但不限于：
2. 用户行为数据；
3. 内容数据；
4. 交易数据；
5. 社交数据；
6. 反馈数据等。
	* + - 1. 应根据数据类型和来源，选择合适的采集方法，包括但不限于：
7. API接口调用；
8. 爬虫抓取；
9. 人工补充；
10. 传感器采集等。
11. 对于API接口调用，需与数据提供方协商接口规范和数据传输协议。
	* 1. 数据处理
			1. 数据验证与清洗
				1. 对采集到的数据进行初步验证，确保数据的完整性和准确性。
				2. 对采集到的数据清洗，去除重复、无效和异常数据，提高数据质量。
				3. 将来自不同数据源的数据进行整合，形成统一的数据视图，方便后续分析。
			2. 数据存储与管理
				1. 应选用数据存储系统，如分布式存储系统、关系型数据库等，根据数据特性和业务需求进行数据存储和管理。
				2. 在数据处理过程中，采取必要的安全措施，如数据加密、访问控制等，防止数据泄露和非法访问。
	1. 情报分析（画像模型）
		1. 概述

情报分析（画像模型）基于多源数据的融合和NLP（自然语言处理）技术，对目标对象形成多维度、全面的画像，包括形成画像、线索分析、情报报告三个步骤。

* + 1. 形成画像

形成画像的过程是情报分析的基础。具体要求为：

1. 需采集与目标对象相关的多源数据，包括但不限于社交媒体信息、通讯记录、公开资料等；
2. 利用NLP技术对采集到的文本数据进行处理，包括分词、词性标注、实体识别等，以提取出关键信息；
3. 将NLP处理后的数据与其他来源的数据进行融合，形成一个完整的数据集。这个过程需考虑不同数据源之间的关联性和互补性，确保画像的准确性和全面性；
4. 基于融合后的数据集，构建多维度任务画像。这些维度可能包括目标对象的身份信息、社交关系、行为模式、兴趣爱好等。
	* 1. 线索分析

线索分析主要是对目标对象的最新整体态势状况进行深入挖掘和分析。具体要求为：

1. 分析目标对象的行为变化、社交动态、情绪变化等方面，以发现可能存在的异常情况或潜在风险；
2. 应结合历史数据和背景信息，对目标对象的线索进行关联分析和趋势预测。
	* 1. 情报报告

在完成线索分析后，形成人员整体状况初步报告。情报报告应全面、客观地反映目标对象的最新态势和关键信息。具体要求为：

1. 报告内容应包括目标对象的身份信息、行为特征、社交关系、潜在风险等方面的描述和分析；
2. 报告还应提供对目标对象未来可能的动态和趋势的预测，以及针对这些预测的建议和措施。
3. 情报报告应具有一定的可读性和可理解性，以便决策者能够迅速了解目标对象的整体状况并作出相应决策。
	1. 目标优选（优选模型）
		1. 明确目标

在进行目标优选时，应先明确优选的目标和条件。优选条件包括但不限于：

1. 用户的兴趣；
2. 行为模式；
3. 购买力；
4. 潜在价值等。
	* 1. 优选方式
			1. 人员优选
5. 首先设定明确的优选条件，基于业务需求、岗位要求、个人能力等多个维度自动筛选出符合条件的人员，并进行初步推荐；
6. 除预设优选条件，还应允许用户根据实际需求进行精确条件的指定；
7. 筛选出符合条件的人员后，应能根据评估指标（如经验、技能、绩效等）对人员进行优先级排序；
8. 查看推荐人员特性，人员信息包括但不限于教育背景、工作经历、技能专长、性格特点等。
9. 用户可以通过对比不同人员的特性进行判断并完成推荐。
	* + 1. 指标评分

按照人员画像模型和推荐场景条件，计算人员评分并进行排序推荐。

* + - 1. 算法模型
				1. 大模型评估法

原理

利用大规模的预训练模型进行目标优选。包含大量参数及网络结构，使得它们能够从海量的数据中学习到丰富的特征和信息。

应用

通过输入特定的条件或特征，大模型能够输出与目标优选相关的预测或推荐结果。

* + - * 1. 业务专家模型法

原理

结合业务专家的知识和经验，构建特定的规则和逻辑来进行目标优选。

应用

业务专家可以根据经验和专业知识，设定筛选条件或权重，推荐模型基于相关条件和权重进行目标人员的筛选和推荐。

* + - * 1. 数据标注训练模型法

原理

需要对原始数据进行标注，即给数据打上标签，以指示其所属类别或特性。然后，利用这些标注数据训练机器学习模型，使其能够学习到数据的内在规律和模式。

应用

训练完成后，模型可以根据新的输入数据进行预测或分类，从而实现目标优选。这种方法需要足够的标注数据和计算资源，但能够提高优选的准确性和效率。

* 1. 干预策略（策略模型）

根据人物的人格类型和认知风格结合大模型自动推荐策略，影响策略生成过程如下：

1. 重点目标确定：根据情报分析结果，确定需要优先干预的重点目标；
2. 个性化认知分析：基于大模型，分析目标的价值观念、思维方式、立场态度等认知特征；
3. 心理活动分析：分析目标的知觉情感、精神意志、行为倾向等心理活动，了解其心理状态；
4. 生理功能分析：运用生理测量技术，分析目标的感觉机能、脑电活动、神经机制等生理功能；
5. 认知干预策略生成：综合个性化认知分析、心理活动分析和生理功能分析结果，生成认知干预策略集。
	1. 干预信息生成：利用自然语言生成技术，根据干预策略集，生成个性化的认知干预信息；
	2. 快速生成与推送：使用数字内容生成技术，快速生成干预信息，并通过合适的渠道推送给目标。
	3. 效用评估
		1. 基本要求
			1. 应建立建立评估标准，明确评估的目标和指标，确保评估结果具有可比性和可操作性。
			2. 选择合适的评估方法，根据系统特点和业务需求选择合适的评估方法，如离线评估、在线评估或A/B测试等。
			3. 持续监控和优化，定期对系统进行效用评估，根据评估结果对系统进行优化和改进，以满足用户需求和提升业务价值。
		2. 评估方法
			1. 离线评估
				1. 利用历史数据构建测试集，对推荐算法进行离线测试。
				2. 常用的准确性评估指标包括准确率、召回率、F1值等。
				3. 可通过计算平均绝对误差（MAE）或均方根误差（RMSE）来评估算法对用户评分的预测准确性。
			2. 在线评估
				1. 在实际运行环境中对推荐系统进行实时测试，通过收集用户反馈和行为数据来评估系统性能。
				2. 用户满意度调查是一种常用的在线评估方法，通过问卷或用户反馈收集用户对推荐系统的满意程度。
				3. A/B测试，将不同版本的推荐系统随机展示给用户，对比它们的性能指标和用户反馈。
			3. 实时性评估
				1. 评估推荐系统处理新数据并生成推荐结果的速度。
				2. 可以通过测量系统从接收到新数据到生成推荐结果的时间间隔来评估实时性。
			4. 业务价值评估
				1. 分析推荐系统对业务目标的影响，如提高销售额、增加用户活跃度等。
				2. 可以通过对比使用推荐系统前后的业务指标变化来评估业务价值。
	4. 算法性能指标
		1. 准确率

准确率应 ≥ 0.9 。

* + 1. 召回率

召回率应≥ 0.85 。

* + 1. F1值

F1值应≥ 0.88。

1. 测试方法
	1. 测试条件
		1. 测试环境

除特殊规定外， 基本测试环境如下：

a) 环境温度： 15 ℃ ～ 35 ℃ ；

b) 相对湿度： 25％ ～ 75％ ；

c) 大气压力： 86 Kpa ～ 106 Kpa ；

d) 照度： 300 lx ～ 1 500 lx ；

e) 环境噪声： ＜ 60 dB 。

* + 1. 测试工具

测试工具包括软硬件平台， 可由检测机构提供或由测试用户准备，其中：

1. 检测机构提供的测试工具要求如下：
	1. 硬件平台：CPU 主频大于2.0 GHz、内存大于 4 GB 的等同性能的硬件平台，测试硬件平台的算力性能上限宜根据检测业务需求确定；
	2. 操作系统：主流操作系统，如 Linux、windows、centos7 等；
	3. 调用算法库进行测试时，应提供算法库接口函数。
2. 测试用户准备测试工具要求如下：
	1. 可结合推荐系统实际应用时的配置准备测试工具；
	2. 直接提供被测系统或产品，使用检测机构提供的测试数据，被测系统或产品反馈测试结果记录。
	3. 性能测试
		1. 准确率

正确推荐的物品数量占总推荐物品数量的比例。计算公式为：

 (1)

式中：

TP代表真正例（True Positive），即预测为正例且实际也为正例的样本数。

FP代表假正例（False Positive），即预测为正例但实际为负例的样本数。

* + 1. 召回率

正确推荐的物品数量占用户实际感兴趣物品数量的比例。计算公式为：

 (2)

式中：

TP代表真正例（True Positive），即预测为正例且实际也为正例的样本数。

FN代表假负例（False Negative），即预测为负例但实际为正例的样本数。

* + 1. F1值

 准确率和召回率的调和平均数，用于综合评估推荐性能。计算公式为：

 (3)

**━━━━━━━━━━━**