

# 基于数据中台的智能标签关键技术研究与应用

盛妍<sup>1</sup>, 朱青<sup>1</sup>, 张明杰<sup>1</sup>, 宋灿<sup>1</sup>, 陈泽<sup>2</sup>

(1. 国家电网公司客户服务中心, 天津 300000; 2. 江苏瑞中数据股份有限公司, 江苏 南京 210000)

**摘要:** 随着数字化转型浪潮的来临, 标签在企业生产中的作用越发重要, 但在日益增长的业务需求中, 传统标签构建方法暴露出了诸多问题。介绍了基于数据中台建设智能标签的研究和设计方法, 包括对现存问题的描述、智能标签的设计及智能标签的展望。基于数据中台所带来的优势, 设计构建智能标签的建设方案, 能为企业建设智能标签及数字化转型提供相应思路。

**关键词:** 智能标签; 数据中台; 数字化转型

中图分类号: TM711

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.211375

中文引用格式: 盛妍, 朱青, 张明杰, 等. 基于数据中台的智能标签关键技术研究与应用[J]. 电子技术应用, 2022, 48(3): 73-77.

英文引用格式: Sheng Yan, Zhu Qing, Zhang Mingjie, et al. Research and application on key technology of intelligent tag based on data middle platform[J]. Application of Electronic Technique, 2022, 48(3): 73-77.

## Research and application on key technology of intelligent tag based on data middle platform

Sheng Yan<sup>1</sup>, Zhu Qing<sup>1</sup>, Zhang Mingjie<sup>1</sup>, Song Can<sup>1</sup>, Chen Ze<sup>2</sup>

(1. State Grid Customer Service Centre, Tianjin 300000, China;

2. Jiangsu Real-time Database Limited Corporation, Nanjing 210000, China)

**Abstract:** With the coming of the trend "digital transformation", the role of tag in production is becoming more and more important. However, in the growing business requirements, the traditional methods of creating tag have exposed many problems. This paper introduces the research and design method of intelligent tag based on data middle platform, including the description of existing problems, the design of intelligent tag and the prospect. Based on the advantages of data middle platform, the method of intelligent tag is designed and created, which can provide ideas for enterprises to build intelligent tags and digital transformation.

**Key words:** intelligent tag; data middle platform; digital transformation

### 0 引言

数据表达一种事实, 标签表达一种结论和评判, 是对数据高度提炼和归纳得到的特征标识, 将信息系统数据及其分析结果自动翻译并转化为定性的、日常生产语境中大的语言文本, 实现了普通业务人员无障碍理解和使用大数据分析结果<sup>[1]</sup>。正是由于这种特性, 标签可支撑业务快速做出决策, 在大数据时代发挥越来越重要的作用。伴随着业务数字化的发展, 标签也从手工时代发展到数仓时代。但日益庞大的业务所带来的数据量使得传统标签建设变得愈发庞杂冗余, 标签也急需向智能化、统一化的方向发展。近年来, 数据中台<sup>[2-4]</sup>理念的诞生, 也为标签的进一步发展提供了方向。

### 1 标签的发展

标签发展目前已经经历了两个阶段, 分别是手工时代和数仓时代。

手工时代的标签基本通过业务人员在业务过程中手动生成, 服务于单一应用, 并通过传统数据库进行存储, 整体生成过程为数据建模—手动打标签—存档。

数仓时代的标签脱胎于对手工标签的业务提取, 通过对业务指标的统计, 从而生成相应的标签。标签的存储计算依赖于数据仓库, 整体生成过程为数据汇聚—数仓建模—规则统计—存档。

标签的下步发展可称之为智能标签时代。相较于传统标签构建方式, 智能标签具备以下特点:

(1) 低门槛: 把数据转化为体系化的业务标签, 降低服务人员使用大数据的门槛。

(2) 易共享: 把标签变成业务可理解的数据服务, 使数据更易共享和使用。

(3) 工厂式生产: 实现标签生产的工厂化, 降低标签生产难度, 提升标签生产效率。

(4)闭环式生态:基于上层应用标签应用成效对标签规则进行评估并开展迭代优化工作,形成标签运营生态闭环。

(5)统一平台管理:无论是自动标签还是手动标签,实时标签还是离线标签,临时标签还是稳态标签,都统一套标签系统统一进行管理并对外统一提供服务,以保障标签对外业务口径的统一。

## 2 现状描述

### 2.1 标签现阶段面临的问题

现阶段标签构建方式是针对某一业务数据,根据业务规则进行提炼,并最终对某标签主体给出结论化的定义,形成标签主体与标签名一一对应的结果集,从而支撑相应业务场景的应用。但传统的标签构建方式在实际生产应用中会出现以下的不足:

#### (1)标签体系紊乱

构建基于数据挖掘理论的客户标签体系的首要步骤是设计提炼客户标签,包括客户标签维度选择、指标细化等<sup>[5]</sup>。标签多以业务需求直接驱动,具有明确的的目的性,因标签应用方向的不同会产生多套标签,但因其相互独立且各自存在体系,各业务方自行定义标签,导致同一名称的标签会有不同含义,或同一含义的标签有不同名称,运用混乱。

#### (2)存储计算开销较大

多套标签存在各自独立的计算逻辑,且其更新频率各不相同,为保障各自业务的不中断,需独立进行更新计算,但多套标签中可能存在大量的重复计算工作,且都各自保留标签计算结果,从而造成了大量的存储、计算资源的开销。

#### (3)统一管理难、应用难

因多套标签的应用方式、标签源数据、标签体系等都相对独立,因此对其进行统一的管理和应用,从而造成多套标签体系共存,无法统一管理应用的问题。

### 2.2 中台概念的引入

#### (1)什么是数据中台

数据中台是聚焦“业务数据化”,践行“数据业务化”,集数据的规划、获取、治理、分析、应用和价值变现为一体的数据服务体系。数据中台核心目标是实现数据资产价值的最大化。数据中台与数据仓库、数据中心在核心定位、服务形式、数据汇聚、数据管理方面均存在差异。数据中台距离业务更近,更重场景应用、更快速响应需求。数据中台可以构建在数据中心之上,数据中心侧重于数据之“聚”,数据中台侧重于之后的“用”。

数据中台定位于为各专业、各单位提供数据共享和分析应用服务,根据数据共享和分析应用的需求,沉淀共性数据服务能力,通过数据服务满足横向跨专业间、纵向不同层级间的数据共享、分析挖掘和融通需求。

#### (2)数据中台为标签带来了什么

①海量的数据资源:在单一数据源的标签构建中,无法全方位地刻画客户特征,仅能从某一个层面或方向描述某一角度下的本体特征,因而构建过程中想要全方位地刻画本体特征,则需满足本体所需全部建模元语,进而能够全方位描述及刻画本体特征。相比于单一数据源系统及数据仓库系统,数据中台包含企业内部各业务系统的业务数据,不同业务系统按业务域划分存储,同时也可接入外部数据,丰富数据体系,为标签的多样化建设提供数据支撑。

②统一的数据管理:数据中台具备数据运营管理能力,能对各业务系统数据进行统一管理,实现对各类数据的企业级建模,解决各类数据间的业务口径不一的问题,为企业级标签体系的构建提供基础。

③强大的计算调度能力:标签构建中依赖平台的支撑、特殊组件调用、海量数据的处理、算法的训练、模型的运行等,单一数据计算环境无法提供上述资源,需依赖更为全面、高性能的基础平台。数据中台计算架构基于大数据技术<sup>[6]</sup>构建,引入内存计算、批计算<sup>[7]</sup>、流式计算等技术,为标签的实时及离线计算提供技术保障。

### 3 智能标签构建设计

相较于传统标签的构建方法,智能标签主要从标签体系、数据支撑、标签计算、标签平台和标签应用五方面开展设计研究。

#### 3.1 企业级标签体系设计

为保障企业多业务系统标签不存在语义歧义及业务口径不一的问题,需开展企业级标签体系的设计工作。

以国网用电业务为例,客户标签从基础属性、客户行为和客户评估三方面对电力全业务特征进行归类与提炼,形成一级主题;同时对一级主题分别进行细分,并根据电力业务特征将客户行为从业务场景视角出发,划分为传统业务(用电服务、客户服务、缴费服务、市场开拓等)和新兴业务(电动汽车、市场化售电、能源电商、能源金融、综合能源服务、分布式电源、数据商业化、互联网业务等),以业务视角形成二级主题,便于各业务线开展标签管理与应用,有助于将客户标签与业务场景进行有效对应,开展客户画像与原有业务流程嵌入,开展特定业务客群划分,增加标签实用性,最终形成如图1所示的电力客户智能标签体系。

#### 3.2 基于中台的数据规划设计

中台作为未来企业数据的主要载体,需对数据进行分层规划,以实现存储和计算资源的合理利用,为企业业务提供统一数据出口。参考阿里中台数据分层<sup>[8-10]</sup>,设计如图2所示的数据体系。

贴源层:参考ODS层(Operational Data Store)及STG(Stage)层设计,贴源层定位为中台核心数据基础层,负责从各业务系统接入操作型数据(结构与源系统保持一致的全量数据),并对数据进行全量或增量的合并及数

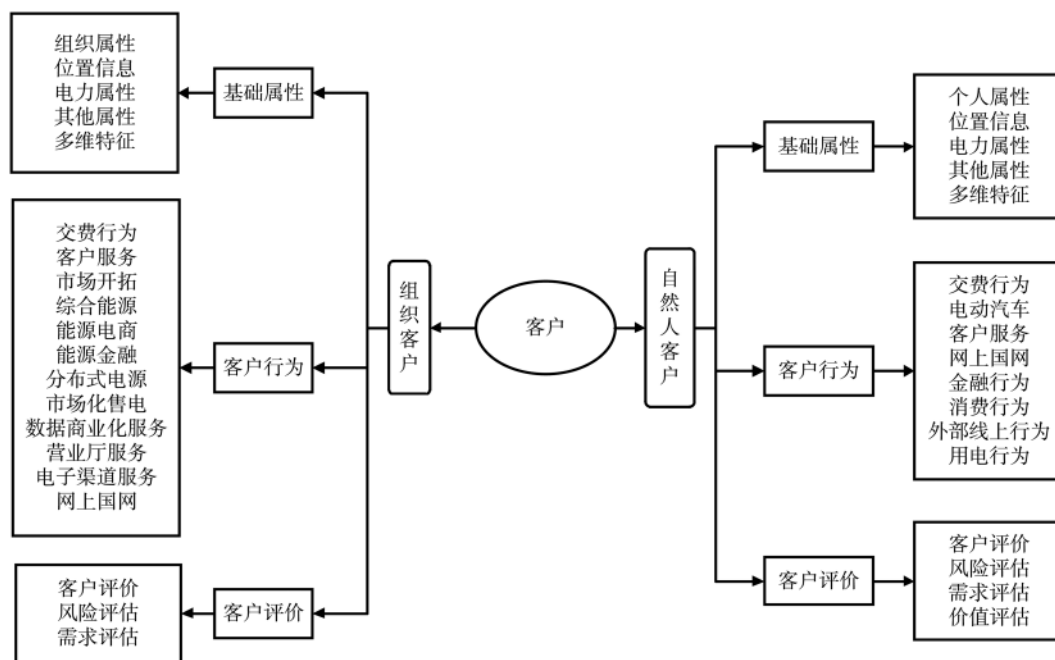


图1 电力客户智能标签体系

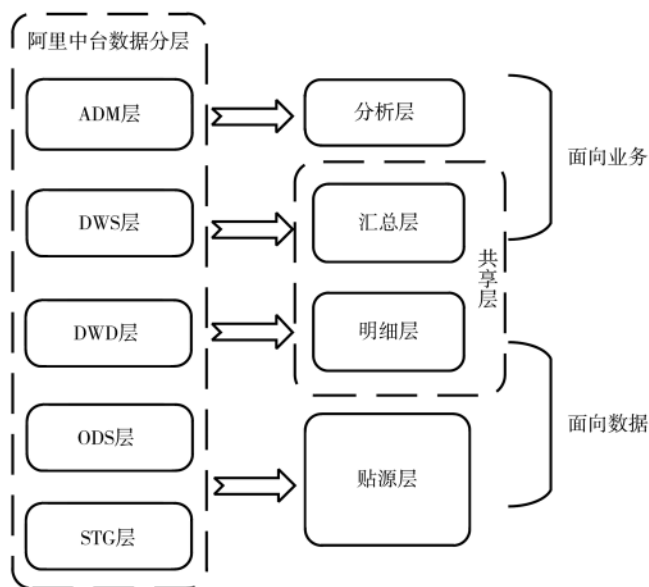


图2 中台数据体系

据清洗和标准化工作。该层基于数据主题开展标签源数据的接入,为智能标签的构建提供数据基础。

**共享层-明细区:**参考DWD层(Data Warehouse Detail)设计,共享层明细区面向数据进行数据域分类、整合、清洗,并根据企业级数据建模标准(SG-SIM4.5),构建一套标准化数据模型。

**共享层-汇总区:**参考DWS层(Data Warehouse Subject)设计,共享层汇总区按分析对象的实体进行数据整合、轻度汇总。该层在对明细区的数据进行简单的业务汇总后,支撑事实标签、规则标签的建设。

**分析层:**参考ADM层(Application Data Market)设计,完全贴合前端应用。分析层由共享层数据参照业务需

求,经过模型计算、算法分析生成对业务的重度汇总数据,支撑模型标签的构建。

### 3.3 智能标签算子设计

在标签构建过程中,通过对基础数据和业务指标的梳理分析,可定义生成标签业务算子。标签业务算子即为基于业务规则构建的单个标签的计算过程或任务,是标签生成的主要执行者,也是构建智能标签与传统标签的主要差异点。针对于复杂业务的标签组合需求,可通过对标签算子的编排组装,生成完整的业务模型,交由中台计算,实现标签快速运算。

针对已完成构建的标签业务算子和业务模型可发布成API或周期性任务运行,实现标签运算的自动化。智能标签算子整体流程如图3所示。

### 3.4 智能标签平台设计

基于数据中台的数据资源及计算资源,需设计智能标签系统,构建智能标签并进行统一管理。智能标签平台的设计主要包含以下功能:

(1)基础源数据的读取。通过读取数据中台共享层的明细数据及汇总数据,为标签业务算子的构建提供数据源支撑。

(2)标签属性生成。根据标签体系规范,结合业务逻辑,基于源数据构建标签算子。针对基础属性,可直接生成自然类标签;针对复杂属性,可通过关联计算、统计计算、挖掘计算等方式,结合业务规则,生成规则类标签、统计类标签、挖掘类标签;针对一些特殊的临时属性,可定义临时标签,支持手动输入,以支撑一些时效性业务场景。

(3)标签业务算子的发布调度。构建完成的标签算子及模型可通过定时任务或条件触发等方式实现智能化



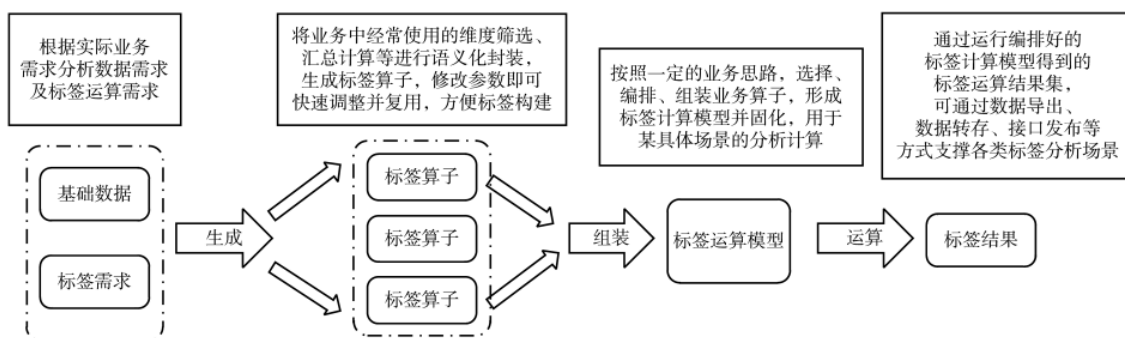


图3 智能标签算子设计流程

调度,以保证标签数据的实时性和可用性。

(4)标签数据管理查询。智能标签平台存储着构建完成的应用标签数据,可通过管理平台进行管理和快速的定位查询,也可通过API的方式发布符合标签要求的明细数据供其他业务系统使用。

(5)标签组合及智能关联分析。基于完成建设的标签,构建标签组合功能,使得业务人员能够根据具体需求自由组合标签,以满足实际业务场景的建设。同时,提供标签关联分析功能,对标签的数据价值进行二次挖掘,发掘标签群体间的关联关系,从而实现标签的应用深化。

(6)标签应用评估。基于已发布的标签及其应用成效进行评估,设置评估指标,并对评估指标不理想的标签进行下架或规则优化,实现标签的智能运营迭代。

### 3.5 智能标签应用场景设计

在完成智能标签的构建后,需开展基于业务的应用场景的设计,以发挥标签的价值。以基于客户构建的智能标签为例,通过对客户信息和相关行为数据的标签化处理 and 探索,按照业务需求圈定群体范围,实现对复杂群体数据的智能化统计、筛选、加工、沉淀,建立单实体360°全方位个性化档案,通过关系图谱展示、可视化线索分析、相似人员智能化推荐等来实现业务人员对于目标群体的精准定位,可开展如下业务应用场景:

(1)精准营销:传统的营销方式不能精准针对客户的需求进行营销,盲目的营销不仅会引起客户的反感,而且极容易引发客户投诉。基于“用户画像”的精准推荐方法可以帮助企业主动向用户推荐感兴趣或需要的商品,刺激需求,减少搜集信息成本,缩小比较范围,提高决策效率,提高成交率<sup>[11]</sup>。通过对客户行为标签和客户类型标签的分析,构建客户画像,精准圈定潜在价值客户群体,从而便捷高效地开展产品精准营销工作。

(2)差异化服务:客户在接收服务的过程中对服务的感知度高低是决定企业的差异化服务是否有效的关键。企业针对不同的客户提供不同程度的优质服务,客户的感知度越高,客户满意度越高<sup>[12]</sup>。通过对客户自然标签和行为标签的分析,结合客户近期诉求情况,设定客户服务临时标签,找出潜在投诉、举报等较高风险的

服务人群,针对性地提供精准服务,以降低客户投诉举报概率,提升客服服务满意度。

(3)新业务拓展:通过对客户理财、电商消费、业务偏好等行为标签的综合分析,评估客户消费能力,划分特征客户群体,拓展潜在客户。针对特征客户群体,推广金融、保险等相关新型业务。

(4)数据变现:可根据外部企业需求对客户标签进行组合关联,形成对外部单位有价值的标签数据,拓展面向互联网电商、金融企业等的toB和面向政府、征信单位等的toG的数据变现业务。

### 4 智能标签在数字化转型中的展望

#### (1)智能标签是业务数据化的典型实例

业务数据化指业务相关信息以数字方式存储并实现可用,即为业务生成数据。单纯的业务数据存储并无法满足数据的可用需求,需通过数据的清洗、转化、统一建模及汇总分析才可满足业务需求,支撑业务运营工作。智能标签即为业务数据化的典型实例,通过对基础业务数据的统一化建模,并根据业务指标进行汇总计算分析,生成智能标签对主体实现定性分群,支撑相关业务场景的建设,使得数据充分发挥业务价值,实现基础业务信息到可存储、可使用、可分析的数据化转变。

#### (2)智能标签<sup>[13]</sup>是数据业务化的重要支柱

数据业务化为业务数据化的延申,指的是通过对基础数据的洞察分析实现对业务的进阶探索,即为数据驱动业务。智能标签源于基础业务信息的数据化,基于业务规则对标签主体进行定性,实现简单的群体分类功能,但随着基于该主体的标签体系的不断丰富,使得标签不再只单纯承载基本的群体分类功能。通过对标签的组合及关联分析等操作,可实现对业务的深度探索,并根据这些标签衍生出更多的业务方向,实现数据对业务的反哺。

#### (3)智能标签是数字孪生化<sup>[14]</sup>的理想载体

数字孪生简单来说就是在一个设备或系统的基础上,创造一个数字版的“克隆体”。以智能标签为例,可通过标签将客户群体进行数字化<sup>[15]</sup>解析,生成相应的数字孪生体,通过对该客户群体进行模拟运营及成效评估,得出高效且合理的运营方案,从而改进实际业务策

略。同时根据实际业务运营的成效,对客户群体的模拟运营进行干预和调优。通过这种模拟运营指导实际执行、实际成效优化模拟运营方案的方式,可极大降低风险,提升整体运营服务质量。

## 5 结论

本文主要论述了在数据中台理念下的智能标签建设和应用方向,提出了一套较为完善的智能标签构建方法,为高效构建智能标签及应用提供了方向。在数字化转型的大趋势下,下一步应深入探究标签与具体业务融合,在具体业务场景下提高标签落地应用效果。

## 参考文献

- [1] 林森, 欧阳柳. 基于大数据理论的电力客户标签体系构建[J]. 电气技术, 2016, 17(12): 98-101.
- [2] 韩向东, 季献忠. 企业数字化转型八大发展趋势[J]. 管理会计研究, 2021, 4(Z1): 21-26, 101.
- [3] 崔恒志, 王翀, 吴健. 基于数据中台的数据资产管理体系[J]. 计算机系统应用, 2021, 30(3): 33-42.
- [4] 姚洪. 基于数据中台的数据治理系统的设计与实现[J]. 科学技术创新, 2020(35): 74-75.
- [5] 吕辉, 许道强, 仲春林, 等. 基于电力大数据的标签画像技术与应用研究[J]. 电力信息与通信技术, 2017, 15(2): 43-48.
- [6] 张娜, 马燕, 崔恒睿, 等. 大数据技术进展与发展趋势[J]. 通讯世界, 2019, 26(2): 78-79.
- [7] 郑钤. 基于 MapReduce 模式的大数据分布式计算态势分析[J]. 通讯世界, 2018(6): 102-104.
- [8] 阿里巴巴数据技术及产品部. 大数据之路: 阿里巴巴大数据实践[M]. 北京: 电子工业出版社, 2017.
- [9] 钟华. 企业 IT 架构转型之道: 阿里巴巴中台战略思想与架构实战[M]. 北京: 机械工业出版社, 2017.
- [10] 邓中华. 大数据大创新-阿里巴巴云上数据中台之道[M]. 北京: 电子工业出版社, 2018.
- [11] 施巍巍. 江苏联通对大数据精准营销模型的探讨[J]. 信息系统工程, 2015(1): 112-114.
- [12] 印斯佳. 电力客户差异化服务探析[J]. 城市建设理论研究, 2011(34): 2095-2104.
- [13] 宋亮. 用智能标签发现未来[J]. 标签技术, 2020(3): 16-18.
- [14] 李海峰, 王炜. 数字孪生智慧学习空间: 内涵、模型及策略[J]. 现代远程教育研究, 2021, 33(3): 73-80, 90.
- [15] 侯欣洁, 戚德祥. 数字技术驱动下的出版“走出去”数字平台建设[J]. 科技与出版, 2021(5): 60-65.

(收稿日期: 2021-02-05)

## 作者简介:

盛妍(1967-), 女, 硕士, 高级工程师, 主要研究方向: 数据挖掘分析。

朱青(1990-), 通信作者, 女, 硕士, 中级工程师, 主要研究方向: 数据分析及应用, E-mail: zhuqing2020@foxmail.com。



扫码下载电子文档

(上接第 36 页)

2017, 28(S1): 234-241.

- [4] 舒雪姣, 陈景发, 敬代波. 语音 DOA 技术在视频会议系统中的应用[J]. 电子技术与软件工程, 2014(22): 138-139.
- [5] 李伟. 智能环境下基于音频视频特征融合的多说话人跟踪研究[D]. 兰州: 兰州理工大学, 2012.
- [6] 李鹏, 高小平, 陈梅清, 等. 主动降噪技术应用于船舶舱室的可行性分析[J]. 舰船科学技术, 2020, 42(3): 69-71.
- [7] XIAO X, ZHAO S, ZHONG X, et al. A learning-based approach to direction of arrival estimation in noisy and reverberant environments[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). South Brisbane: IEEE, 2015: 2814-2818.
- [8] DIAZ-GUERRA D, BELTRAN J R. Direction of arrival estimation with microphone arrays using SRP-PHAT and neural networks[C]//2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), Sheffield, UK, 2018: 617-621.
- [9] NGUYEN T N T, GAN W S, RANJAN R, et al. Robust source counting and DOA estimation using spatial pseudo-spectrum and convolutional neural network[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing,

2020, 28: 2626-2637.

- [10] VARZANDEH R, ADILOGLU K, DOCLO S, et al. Exploiting periodicity features for joint detection and DOA estimation of speech sources using convolutional neural networks[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020: 566-570.
- [11] WANG Y, WANG D L. Towards scaling up classification-based speech separation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(7): 1381-1390.
- [12] SRINIVASAN S, ROMAN N, WANG D. Binary and ratio time-frequency masks for robust speech recognition[J]. Speech Communication, 2006, 48(11): 1486-1501.

(收稿日期: 2021-08-30)

## 作者简介:

刘雨佳(1997-), 男, 硕士研究生, 主要研究方向: 麦克风阵列、语音增强。

童峰(1973-), 通信作者, 男, 博士, 教授, 主要研究方向: 水声通信与网络、声探测与感知、智能语音处理, E-mail: ftong@xmu.edu.cn。

陈东升(1975-), 男, 硕士, 助理教授, 主要研究方向: 水声通信、声信号处理。



扫码下载电子文档

## 版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所