

· 专题一：双清论坛“大规模商务场景的统计管理理论” ·

基于可泛化模型的复杂商务场景数据分析*

史颖欢^{1†} 郭金涛^{1†} 李泽昆¹ 祁磊² 高阳^{1**}

1. 南京大学 计算机软件新技术全国重点实验室, 南京 210023

2. 东南大学 计算机科学与工程学院, 南京 210096

[摘要] 随着全球经济合作与竞争形势的快速发展, 复杂商务场景日益涌现。其中, 数据的规模和复杂性也呈现显著增长, 在带来可利用数据资源的同时, 也为有效挖掘和分析潜在商务模式带来挑战。本文首先针对复杂商务场景下的数据特点进行分析, 其次调研了商务分析方法的研究进展, 包括: 大规模商务数据集、多模态预训练与特征融合方法以及多场景下可泛化商务模型技术。此外, 对复杂商务场景数据分析在未来的应用发展进行思考与总结, 讨论了目前亟需解决的关键问题, 包括: 对更多模态的适应性、商务分析模型可解释性、动态环境下的分布变化鲁棒性等。进一步分析了复杂商务场景可泛化模型的研究现状和挑战, 旨在为复杂商务分析领域的研究提供初步参考, 实现复杂商务场景的全面智能化。

[关键词] 商务场景; 可泛化数据分析; 深度学习; 多模态融合

1 复杂商务场景数据分析研究意义

全球经济格局变化迅速, 经济模式日新月异, 逐渐形成了以数字经济为代表的新型商务场景。近年来, 我国也在数字经济领域积极布局, 奋勇争先, 抢占制高点。需要指出的是, 数字经济发展速度之快、辐射范围之广、影响程度之深前所未有, 正在成为重组全球要素资源、重塑全球经济结构、改变全球竞争格局的关键力量^[1]。随着技术的进步和创新, 人工智能已经成为当今世界经济和社会发展的支撑技术。2022年1月公布的《“十四五”数字经济发展规划》^[2], 明确了“十四五”时期推动数字经济健康发展的指导思想、基本原则、发展目标、重点任务和保障措施。数字贸易是数字时代的象征, 是科技赋能的标志, 是未来贸易发展的方向。据统计, 全球服务贸易中一半以上已实现数字化^[3]。数字经济的发展离不开对现有商务场景的深入剖析。随着我国大数据产业的崛起, 商务场景逐渐从简单演变到复杂, 从单一到多源, 从分布孤立到竞争协作, 此外交易数量和



高阳 南京大学教授, 博士研究生导师, 南京大学智能科学与技术学院书记兼执行院长, 中国人工智能学会会士, 江苏省人工智能学会常务副理事长。主持国家自然科学基金重点项目、重大项目等课题。曾获吴文俊人工智能自然科学奖二等奖、江苏省科学技术奖二等奖等荣誉。



史颖欢 南京大学计算机学院副教授、博士研究生导师, 南京大学健康医疗大数据国家研究院医疗AI平台负责人。主持国家自然科学基金优秀青年科学基金项目、国家重点研发计划、科技创新2030—新一代人工智能重大项目等课题。曾获江苏省科学技术奖二等奖、吴文俊人工智能优秀青年奖等。



郭金涛 南京大学计算机学院在读博士研究生, 主要研究方向为模型泛化性、模式识别与人工智能。

收稿日期: 2023-12-30; 修回日期: 2024-03-04

† 共同第一作者

* 本文根据国家自然科学基金委员会第344期“双清论坛”讨论的内容整理。

** 通信作者, Email: gaoy@nju.edu.cn

本文受到国家自然科学基金项目(62222604, 62206052, 62192783)的资助。

规模快速膨胀,跨足广泛的产业和领域,涉及众多企业、客户和供应链。不同于仅涉及零售交易、小型餐饮等较少交易的简单商务场景,复杂商务场景包括跨国贸易、大型电商平台等,涉及大量商业活动、参与者和复杂交易,其数据规模和复杂性更高,具有实时可得、体量大、类型多样等关键特点。这些海量数据覆盖了产品销售、用户反馈、供应链运作、市场竞争等多个领域信息,具有辅助企业管理决策的重要价值,已成为推动企业运营和经济增长的战略资源。然而,大规模商务场景数据的多样性、异构性与实时性,也给数据统计分析方法提出严峻挑战。传统分析方法面向简单商务场景设计,难以处理具有高维异构特性的海量商务数据。因此,如何分析复杂商务场景中的数据资源,利用先进的数据科学和人工智能技术,将其转变为有效信息以辅助管理决策过程,挖掘潜在商业模式,服务于国家与社会,已成为当今国际形势日趋复杂动荡情形下,急需面对的现实挑战。

复杂商务场景涉及复杂多样化的商业应用任务,覆盖市场营销、客户服务、供应链管理、商品推荐等众多领域。这些任务的多样性不仅极大增加了商务数据的量级,而且涉及的信息类型和业务场景多种多样。例如,以电商平台淘宝为例,每天有超过6000万固定访客,伴随产生数十亿条用户行为数据,营销渠道涵盖数百个在线广告平台和设计媒体;平均每分钟售出4.8万件商品,需要进行供应链管理智能管理,实时检测商品库存和追踪数千万个货物物流。商务场景任务的复杂性和互相关联性使得商务数据的分析变得更加具有挑战性。面向这些多样化任务,数据主要具有三个关键特点:覆盖范围广、涉及链条多和数量规模大。首先,商务数据覆盖企业内外的各个环节,涉及的维度和层次非常多,需要综合考虑不同层面的数据信息;其次,商务数据涉及链条多,数据之间存在复杂的关联和依赖关系;最后,商务数据数量规模庞大,不同领域数据分布也会具有较大差异性。上述所提及特点给商务分析模型的鲁棒性与泛化性带来挑战。

面向海量异构的复杂商务数据,单一模态数据分析的局限性逐渐显现,多模态数据分析成为关键技术。在电子商务、在线教育等领域,用户产生的数据包括文字、图像、视频等多种模态,这使得单一模态的分析方法难以胜任。因此,有效融合多模态数据、提高模型泛化性成为当前迫切需要解决的问题。多模态分析方法的研究进展将在解决商务场景数据

的多样性和复杂性方面发挥关键作用。因此,本文将从复杂商务数据技术出发,深入剖析多模态分析方法的研究进展,并介绍多模态数据分析在不同应用场景中的应用,总结当前研究面临的挑战。通过对泛化方法、多模态融合、预训练模型等方面的深入讨论,提出可行的研究建议,展望未来在复杂商务场景数据分析领域的发展方向。通过本文的研究,旨在为调研复杂商务场景下的数据分析问题,以此提供新的视角和方法。共同推动智能商务领域的进一步发展,为促进我国经济发展,占领国际经济高地,提供有力支撑。

2 复杂商务场景的可泛化模型研究进展

围绕复杂商务场景的数据分析问题,本文首先从商务应用场景出发,围绕商务数据多模态和动态性特点,详细介绍了商务背景下的多模态特征提取与融合技术,以及多模态预训练范式,并且对商务场景下的多模态分析方法进行了深入探讨。然后介绍了多场景商务应用中可泛化模型的研究进展。图1总结了基于可泛化模型的复杂商务领域数据分析方法的通用范式。通过对可泛化商务应用模型进行了系统的总结和归纳,揭示了相关研究的重要价值和意义,以剖析当前国际学术前沿以及未来发展趋势。

2.1 复杂商务数据分析应用场景

复杂商务领域涉及金融科技、智慧零售、互联网消费等多个商业场景,在复杂商务活动和交易的过程中,会产生海量的实时数据。如何挖掘和分析复杂商务场景中的数据资源,提取关键有效信息,是管理决策智能化的关键研究问题。由于商务环境的复杂性,需要面临市场竞争、供应链管理、定价策略等多方面挑战,复杂商务场景下的任务呈现多样化特点。典型的商务数据分析任务包括市场趋势分析、客户行为分析、供应链优化、风险管理、营销优化、人力资源管理等。例如,在电子商务场景,现有任务可基于以商品为中心和以用户为中心两个主要维度进行划分。其中,以商品为中心的数据分析任务包括商品实例检索^[4]、商品分类和标签^[5]、广告生成^[6]等;以用户为中心的数据分析任务则主要涉及个性化推荐^[7]、用户行为分析^[8]、情感分析^[9]等。

在以商品为中心的任务中,商品信息涵盖了多模态的异构数据,例如文本、图像、视频、音频等不同结构信息。多模态已经成为商务数据的最显著特征,如何从多模态商务数据中提取通用信息,并建立各模态之间的关联已经成为商务数据分析的关键挑

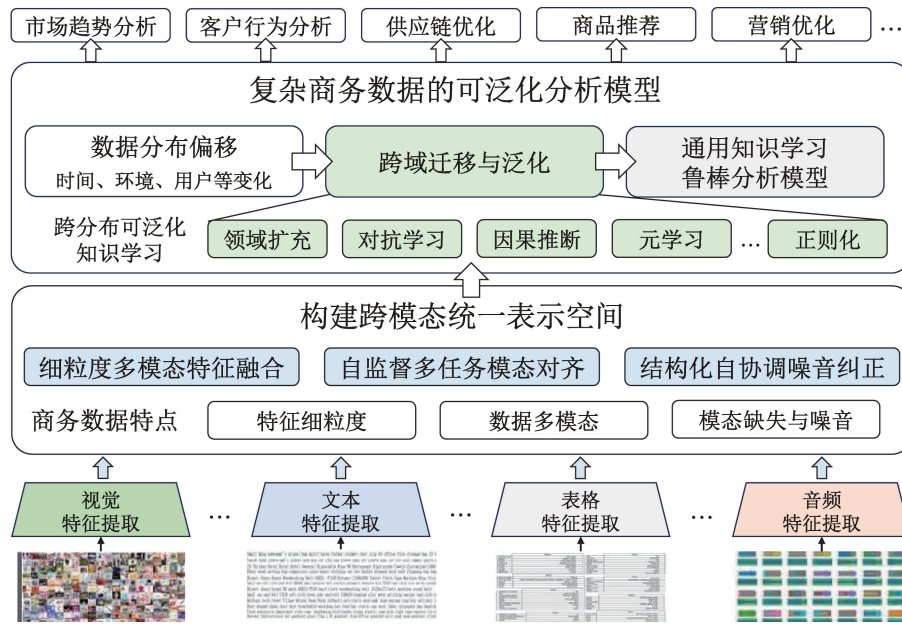


图1 基于可泛化模型的复杂商务领域数据分析方法示意图

战之一。相关研究表明^[10-12],多模态预训练与融合技术已成为解决这一问题的关键方法。在以用户为中心的任务中,用户信息除了涉及历史评论、浏览行为、购买记录等多种结构数据外,还具有个性化需求、时间动态性、新用户冷启动、隐私保护性强等特点。如何从有限的用户历史数据中,构建能够适应环境时间变化的预测模型,并且实现模型对新用户的泛化成为商务数据分析的关键问题。构建可泛化商务应用模型已成为解决该挑战的关键技术^[13, 14]。因此,从复杂商务代表性场景出发,聚焦于商务数据多模态和动态性特点,本文围绕商务数据分析中的预训练范式和多模态融合技术,以及可泛化数据分析方法展开梳理和讨论。

2.2 多模态商务数据处理

随着信息技术的不断发展,商务场景数据的形式日趋复杂,体现为覆盖范围广泛、涉及链条复杂、数据规模巨大等。现阶段,在电子商务领域的数据通常以多个模态呈现,例如,一个商品常包含详情图、评论图、商品属性等多种模态数据。这种多模态数据的呈现方式为商务分析和决策提供了更全面的信息基础。然而,不同模态数据之间也存在潜在的关联性和异构性,这种模态差异也给模型学习通用的特征表示带来挑战。如何从海量多模态商务数据中学习得到可泛化的通用特征,已经成为目前研究的关键问题。

2.2.1 多模态商务基准数据集

现有的电子商务数据集主要从社交网络(如

Twitter、Facebook等)上收集而来,并且涉及图像、文本、视频、音频、表格等多种模态数据。表1总结了代表性多模态商务基准数据集。其中,Dress Retrieval^[15]、RPC checkout^[16]和 Product1M^[17]是典型的电子商务多模式数据集。Dress Retrieval^[15]数据集包含约2万个样本,涉及50个服装类别。RPCcheck-out^[16]提供3万个简单背景的小型零售商品样本。Product1M^[17]提供458个化妆品类别的118万个样本。然而,这些数据集仅限于为特定任务收集的图像/文本两种模态,对于复杂商务场景中丰富且多样化的下游任务,有限的类别、数量、模态并不足以稳健地验证下游任务的性能。为了解决模态多样性不足和规模有限的问题,SCALE^[4]构建了一个超大规模的电子商务多模态产品集M5Product,包含来自6232个类别的600多万多模态样本,涉及图像、文本、视频、音频和表格5种模态的数据。M5Product包含不完整的模态对和噪声,同时还具有长尾分布,这与现实世界中的大部分商务问题相似。ECLIP 100M^[12]从电子商务网站中收集了约1亿图像文本对,包含1500万个不同的商品,涵盖了服装、日用品、仪器等约9000个不同的类别。对于每个商品样本,都有相应的文字描述和来自商品详情页、客户评论和附加广告视频的图片。这些多模态数据集可用于全面地评估多模态预训练范式的商务应用和泛化能力,促进商务场景下多模态研究的新方向。

表 1 多模态商务基准数据集

数据集	样本数	类别数	模态
Dress Retrieval ^[15]	30 000	200	图像/文本
RPC checkout ^[16]	20 200	50	图像/文本
ProductM ^[17]	1 182 083	458	图像/文本
MEP-3M ^[18]	3 012 959	599	图像/文本
M5Product ^[4]	6 313 067	6 232	图像/文本/音频/ 视频/表格
ECLIP 100M ^[12]	~100 000 000	~9 000	图像/文本/视频

2.2.2 通用多模态预训练范式

近年来,随着对视觉—文本多模态学习的不断深入研究,多模态预训练模型在各个领域取得了显著的进展。这些模型主要分为两类:一是单流模型,其转换层在对视觉样本进行特征提取之后,直接与文本输入进行串联拼接,然后输入到网络中进行多模态学习,代表性的工作有 Image-BERT^[19]、MMT^[20]和 UNITER^[21]。这些工作使用大量训练数据和多个预训练任务,旨在得到通用的图文多模态模型。UNITER^[21]在四个图像和文本数据集上进行大规模预训练,并且设计了四种预训练任务,包括:遮蔽语言建模、遮蔽区域建模、图像—文本匹配和词—区域对齐任务,在多个图文多模态任务中达到先进水平。二是双流模型,其特点是图像和文本输入会经过独立的编码器模块,然后通过交叉编码器来实现不同模态信息的融合,代表性的工作有多模态预训练模型(Contrastive Language-image Pre-training, CLIP)^[22]和图像生成系统 DALL-E^[23]。这些模型通过基于 Transformer^[24]的大规模多模态数据进行预训练,随后通过微调适应各种下游任务,包括但不限于图像分类、文本生成、目标检测等^[25, 26]。其中,CLIP^[22]具有强大的通用性和零样本泛化能力,其核心思想是通过对比学习进行视觉和自然语言表征的对齐。DALL-E^[23]则是在自回归生成模型的基础上,加入了文本输入,从而实现了零样本的按文生图。尽管在架构、特征和主干网络上存在差异,视觉—语言预训练的共同目标是通过大规模的图像和文本语料库获得隐式预先构建的模型,使其具备理解多模态数据的能力,并在多个下游任务中展现出卓越的性能^[27, 28]。多模态对齐与预训练范式的研究已验证了更多的数据和更大的模型这一思路对模型性能的有效性。在商务领域,多模态预训练大模型也逐渐应用于多种复杂任务中,包括跨模态商品检索、用户推荐、欺诈检测等。然而,相比于通用场景,商务场景数据往往具有细粒度、多模态的特

点,并且具有严重的模态缺失和模态噪音等问题。将现有的多模态预训练方法直接应用于商务场景,会严重降低多模态信息学习的性能^[29],这也给模型向多样复杂的下游任务泛化提出挑战。因此,面向商务场景复杂多样的任务,如何结合商务数据集特点设计商务场景多模态预训练方法,已经成为亟需解决的研究问题。

2.2.3 商务场景多模态预训练方法

受到 BERT(Bidirectional Encoder Representations from Transformers)^[30]在自然语言处理方面成功的启发,关于视觉和语言表示学习的研究受到极大发展。其中,考虑到商务下游应用的广泛性,现有的方法主要聚焦于以产品为中心的多模态数据预训练上^[10-12, 31, 32],致力于挖掘图像模态信息与文本模态信息之间的关联。商务场景数据分析方法早期通常关注于跨模态检索任务,依赖于在特定任务上预训练的模型来提取细粒度的多模态特征。例如在电商服装领域,衣物的面料、花纹等属性细节对检索任务非常重要,而已有通用的预训练模型无法很好地捕捉到这些细粒度特征。此外,电商图像也存在感兴趣区域(Region of Interests, ROI)单一,包含大量对商品匹配无用的噪音等问题。因此, FashionBERT^[31]聚焦于服装领域的跨模态图文匹配任务,提出在视觉模态将图片切分为固定块以抽取特征,从而在保留图像所有细节信息的同时,保证不会出现重复或过多无用的 ROI 的问题。FashionBERT 以预训练 BERT 模型为骨干网络,并且利用掩码语言建模、掩码图块建模、文本图像对齐等多任务来预训练模型,从而增强模型对图像和文本中细粒度语义信息的感知能力。进一步地, Kaleido-BERT^[32]在视觉模态上往更细粒度的方向进行了探索。Kaleido-BERT 提出提取商品图像最显著的前景区域,以降低空白背景的影响,并且进一步提出多尺度的图片块切分策略,以提取多粒度的图片块特征。在预训练策略上, Kaleido-BERT 提出一种基于图片文本预对齐的掩码任务,以解决随机掩码策略中模态间监督信息缺失问题。Kaleido-BERT 也设计了多样的自监督预训练任务,为图片端提供更充分的自监督学习任务。CAPTURE^[17]提出一种混合流架构,分别编码不同模式的数据,并以统一的方式进行特征融合。此外, CAPTURE 引入了跨模态的对比学习损失来实现图像与文本的对齐,从而生成具有判别性的实例特征,在实例级产品检索任务中取得了优异的性能。这些技术针对商务

场景数据的细粒度、多模态特点,主要解决了通用预训练模型在商务数据细粒度特征捕捉能力不足的问题。

尽管这些针对图像和文本模态提出的多模态预训练方法表现出良好性能,但在面对多模态商品数据有大量缺失和噪声的实际场景时,这些方法的效果仍会严重下降。针对模态缺失和模态噪声的两大难题,K3M^[10]提出将结构化的商品数据表示为知识三元组,并构建知识图谱,将知识图谱作为一个新的模态和图像文本等数据融合,用于纠正噪声并补充图像和文本的模态缺失问题。SCALE^[4]则考虑了结构数据以及视频、音频模态所提供的额外补充信息,探索了一种可扩展到大量模态的跨模态表征学习方法。SCALE提出了一种自协调的对比学习框架,通过自适应特征融合机制将不同的模态整合到统一的模型中。此外,为区别于之前给所有模态分配相同贡献的方法,SCALE提出从模态嵌入中自适应学习各模态重要性,从而有效地提升了模型对模态噪声的鲁棒性。为了进一步提升多模态预训练模型在广泛商务任务上的泛化性,CommerceMM^[11]提出了一种全检索预训练的新方法,以建立任务模态之间的关系和学习跨模态通用表示。不同于之前在图像—文本对数据上预训练的工作,CommerceMM提出了5种图像文本的预训练任务,以及9种跨模态和跨对检索任务,旨在学习多模态到多模态数据上的对齐,以应对更常见和多样化的商务数据。ECLIP^[12]则考虑在商务数据中,同一样本的外观会因为视角或场景的变化而有所不同,但它们都包含相同的产品实体,因此提出了一种以实例为中心的多模态预训练范式,从而有效地提高了模型对背景偏移的泛化性。这些方法都采用了多模态预训练的方式,并结合结构化的商品数据表示和知识图谱构建等技术,以应对多模态商品数据中的缺失和噪声问题。上述商务场景下的预训练方法为学习可泛化通用特征提供了坚实的基础。然而,在面对商务环境中的复杂多样化的下游任务时,仍须结合特定模型的能力以及任务的具体特征,来深入研究如何提升模型的泛化性能,从而确保模型在复杂商务场景应用中的高效性与可靠性。

2.3 可泛化商务模型构建

在复杂商务场景中,受用户行为存在差异、用户群体不断扩展等诸多因素影响,商务数据分布也表现出持续动态变化的特性,使得模型在部署阶段所面对的测试数据相较于训练数据存在分布偏移,造

成性能显著下降。如何增强模型在未见数据分布上的可泛化性成为推进深度学习落地应用的关键问题。这一问题在近年来已经得到广泛关注,各种深度模型可泛化性增强技术不断涌现。同时,相关技术进一步被用于复杂商务场景下的多项具体任务,以构建可泛化商务应用模型。

2.3.1 通用深度模型的可泛化性研究

当前大多数深度模型的优异性能高度依赖于训练数据与测试数据服从独立同分布这一假设。当测试数据采样自不同未知分布时,深度模型的性能往往会受到显著负面影响^[33]。为了应对训练数据与测试数据之间的分布偏移,研究者们希望仅在已知数据(源域)上训练的深度模型能够具备泛化到分布外新数据(目标域)上的能力,称之为领域泛化问题^[34]。一类典型的领域泛化方法是通过对抗学习^[33]、因果学习^[35]、元学习^[36]等技术手段,使得模型学习到与领域无关的特征表示。另一类重要方法则是领域扩充,即通过在输入空间^[37]或特征空间^[38]扩展源域数据多样性来增强模型的可泛化性。正则化方法在领域泛化中的应用也受到关注。受经典正则化方法 Dropout^[39]的启发,PLACE^[40]和 DomainDrop^[41]等方法通过抑制部分中间层特征通道的方式来减少领域特定特征的产生,进而缓解模型对源域的过拟合。此外,集成学习和自监督学习的技术在增强模型可泛化性方面也展现出非凡的潜力。例如,DAEL^[42]针对不同源域数据训练多个领域特定模型,并在集成后用于目标域任务。Bucci等^[43]进一步研究了其它多种自监督代理任务增强模型可泛化性的作用。

除了领域泛化之外,还有一些其它的研究问题关注模型在新数据分布上的可泛化性。迁移学习关注如何将已有模型的知识迁移到另一个相关但不同的下游任务上^[44]。不同于领域泛化问题假设目标域数据在训练过程中不可见,迁移学习需要在目标数据上进行微调以适应下游任务。领域自适应^[45]假设模型在训练时能够访问一部分稀疏标注或无标注的目标数据,以帮助近似估计目标域的数据分布,从而实现从源域到目标域的泛化。在复杂商务场景下,可以根据具体的训练条件,即是否有目标域的数据及标注,来选取对应的技术手段帮助构建可泛化的商务应用模型。

2.3.2 代表性商务场景中的可泛化模型

在各种商务场景中,数据分布偏移现象及其带来的挑战也普遍存在。例如,消费者的购物行为会

随着季节的更替和地域的不同而存在显著差异,时空变换带来的分布偏移可能影响相关应用系统的实际部署;智能化的商务数据分析往往需要用户在多个不同场景下的行为记录以实现准确的用户画像,但不同平台之间的领域差异为此增加了难度;动态变化的用户群体和商品集合使得目标数据分布在训练时无法完全被很好地覆盖。因此,面向具体的商务应用系统,对其中可泛化模型的研究也得到了广泛关注。

推荐系统是电子商务场景中的一个典型应用,目的是通过用户的历史交互行为挖掘用户的兴趣所在,进而为其推荐可能感兴趣的产品。然而,推荐系统面临着两大挑战:其一,用户与产品之间的交互数量往往是不足的,难以从中挖掘出可靠的用户兴趣;其二,对于不断加入的新用户,系统中并不具备任何历史交互记录。这两大挑战,通常概括为数据稀疏性问题与冷启动问题,长久以来限制着推荐系统的性能。因此,跨域推荐^[8]问题应运而生,旨在借助用户在其它平台上的交互记录,来提升当前平台上的推荐性能。跨域推荐可以被建模为迁移学习或领域自适应问题^[9, 46],其中,其它平台的丰富信息被视为源域数据,而当前平台的推荐任务被视为目标域。借助于不同领域之间用户和产品的重合,跨域推荐方法学习领域之间的映射关系以实现知识迁移。RC-DFM^[47]将评论文本和产品信息作为辅助信息,并使用堆叠去噪自编码器融合辅助信息与评分矩阵。同时,RC-DFM在隐空间中训练一个多层感知器,用于实现推荐结果的跨域非线性映射。DDTCDR^[48]将源域中用户与产品的特征嵌入迁移到目标域,以帮助目标域中的用户建模。具体地,DDTCDR采用双向迁移学习机制,在源域与目标域之间进行交替训练直至学习过程稳定。SSCDR^[49]尝试在跨域推荐中引入半监督学习技术,将不同领域之间重合的用户视为有标注数据,而将所有的产品视为无标注数据,以此学习跨域隐空间中的距离编码函数,从而实现跨域推荐。CDIE-C^[50]提出使用协同聚类的方法挖掘不同领域中用户行为相关性,进而将聚类簇层级的相关性信息嵌入到用户—产品层级的特征空间中。尽管跨域迁移的研究已经取得了非凡的进展,但这一设定的局限性体现在所有的源域和目标域数据在训练阶段都是已知且可见的,于是相关方法均依赖先验的用户重叠情况与领域相关知识,而这些先验在实际商务场景下往往难以获得。为应对潜在的未知目标域,Zhang等^[7]进行了

初步探索,提出领域泛化推荐任务设定,并引入了HIRL+方法。在领域泛化推荐设定下,目标域的用户—产品交互信息以及源域—目标域的重叠情况在训练过程中不可见,推荐系统需要仅根据用户和产品资料完成推荐。HIRL+通过分层不变表示学习从用户和产品的属性中提取共同模式,通过可学习的环境分配模块动态调整细粒度的环境层次划分,通过对抗训练范式增强模型对分布偏移的鲁棒性。

欺诈检测系统是另一个重要的电子商务应用,对于预防欺诈风险,维护网络金融安全意义重大^[51]。对于用户而言,欺诈检测系统通过分析用户的行为数据和交易模式,识别可能存在的欺诈行为或非法交易;对于产品而言,欺诈检测系统通过分析电子商务平台上的公共数据,识别出具有欺诈行为的商品,包括虚假购买、伪造评分等^[52]。在新的领域扩展新的电子商务业务时,由于数据和标签有限,欺诈检测系统与推荐系统会面对相似的挑战。因此,跨域(跨平台)欺诈检测^[52, 53]问题同样受到学界和业界的关注。CATS^[52]通过广泛地收集公共数据并进行训练,尝试将跨域的样本特征嵌入到一个统一的隐空间中,实现跨平台的产品欺诈行为检测。HEN^[53]关注跨域用户欺诈行为检测,采用层次化结构分别捕捉领域共享和领域特定的知识,并基于领域注意力机制自动学习两种表示的权重,同时通过最小化类别感知欧几里得距离减少跨领域样本之间的差异。ADAPT^[54]关注与欺诈检测相关的跨域金融风险预测问题,对于不同领域之间用户行为分布不均衡的现象,采取自适应特征加权的方法加以解决。

上述工作代表性地展示了深度模型可泛化性增强的通用技术在复杂商务场景具体任务中的应用。与多模态商务数据处理方法高度依赖数据自身特点相类似,在构造可泛化商务模型时,研究者们仍然需要针对特定的任务场景及相关领域知识,借助不同的深度学习技术手段,设计特定的模型架构和训练方式。

3 复杂商务场景分析研究挑战与关键科学问题

正如上文所概括,当前基于泛化模型的复杂商务场景分析研究主要从数据和应用两个角度展开。对商务多模态数据恰当的收集、处理与分析是构造可泛化商务应用模型的基础;所得到的可泛化商务应用模型进而构成复杂商务场景智能系统的骨架。

前一章分别介绍了复杂商务场景分析各个环节的具体工作。这些工作针对某一类具体场景,从不同的动机出发,解决了数据处理和模型构建中的特定问题。本章着眼于复杂商务场景分析中尚未妥善解决的共性挑战,将其概括为亟待解决的三大关键科学问题。如图2所示,第一,现有商务数据分析方法以图像—文本两种模态为主,对更多模态的适应性仍然不足;第二,基于深度学习的商务模型难以解释和理解,限制了其在敏感重点领域的应用;第三,复杂商务环境动态多变,模型对分布变化的鲁棒性仍需加强。要实现复杂商务场景分析的智能化,需针对这些研究挑战进行重点突破。

3.1 商务通用多模态预训练方法

预训练模型作为现有商务数据分析的核心方法,如何适应复杂商务场景中多样化的下游任务是关键问题。在自然语言处理领域,目前已有较为成熟的预训练方法,将多种任务统一为包含掩码的生成任务,经过掩码语言建模预训练的模型,通过提示学习等方法适配于不同的下游任务。而在复杂商务领域,往往涉及海量的多模态数据,尚缺少类似的通用预训练任务。目前主流的多模态预训练任务为图文对比学习,代表性工作为 CLIP、BLIP,与下游任务有较大差别,难以采用统一范式包含所有自监督预训练任务。由于复杂商务场景的下游任务种类繁多,包括产品检索、广告生成、个性化推荐等,其输入和输出等模态均不相同,难以使用一套统一的框架。此外,商务数据包含图像、文本、音频、视频等多模态数据,模态内部包含大量冗余连续信息,模态之间也存在复杂的关联性,因此,从商务数据中学习通用规律知识需要海量的训练数据,学习难度也更大。

针对复杂商务领域下游任务的多样性难点,Dong 等^[4]提出了包含 5 种模态的大规模商务数据,并且设计了一种自协调的对比学习框架 SCALE,针对不同模态特点提出不同的掩码建模任务,并使用自适应跨模态对比学习方法实现不同模态之间的语义对齐。Yu 等^[15]引入图像/文本/多模态两两之间的对比约束,构建了一种全检索预训练方法,并通过

模态随机化方法来动态地调整模型训练过程。尽管这些方法研究了更多模态数据的处理方法,但对于多模态输入仍不够灵活,难以适应下游商务任务愈加复杂的多模态数据。针对海量商务数据,如何对不同模态输入进行统一且有效的自监督预训练学习,提高预训练模型的跨模态泛化能力,仍是目前亟待解决的问题之一。

3.2 可解释商务分析方法

基于深度学习的人工智能模型被认为是“黑箱”模型,其决策过程难以解释和理解^[55]。由此产生的高风险是阻碍深度学习产品获得商务应用的主要问题之一。因此,提升商务分析模型的可解释性至关重要。当前对模型可解释性的研究主要关注两个方面。其一,是模型输出结果的可解释性。以典型的推荐系统应用为例,解释某些产品为何被推荐给特定的用户,有助于增强推荐结果的可信度,并帮助纠正不恰当的推荐结果^[56]。其二,是模型本身的可解释性。研究模型如何产生输出结果,哪些结构起到关键作用,有助于进一步优化模型设计,减少冗余结构,进而降低成本^[57]。

然而,对于复杂场景下的商务分析模型而言,海量多模态、跨领域商务数据的引入给模型的可解释性带来了新的挑战:首先,复杂商务场景中存在的数据稀疏、模态失配、领域偏移等问题可能显著影响现有可解释性方法的性能。其次,可解释性方法需要同时考虑解释不同领域、不同模态的信息分别如何影响模型的输出结果。在目前已有工作中,Zhu 等^[53]在跨域欺诈检测框架 HEN 中,借助注意力权重评估各领域用户行为序列的不同层次和不同部分与最终欺诈检测结果的相关性。Yu 等^[58]提出可解释的跨域推荐系统 AdaReX,使用不同的编码器在不同领域上执行评分预测和评论重构任务,并通过判别器模型进行连接,借助提取得到的领域不变共享特征,目标编码器只需要学习与个性化解释生成相关的最小特征,从而有效地利用其他领域的知识,增强目标领域的可解释性。上述工作主要关注应对复杂商务场景下的领域偏移挑战,但同时考虑数据稀疏、模态失配等工作尚且缺乏,需要在未来的研究中加以关注。

3.3 鲁棒商务分析模型

复杂商务领域数据具有持续变化的突出特点,其中时空分布的不断变化和用户行为的动态演化给分析模型带来显著挑战。如何应对复杂多变的数据分布变化,以保证模型的有效性和泛化性是关键问

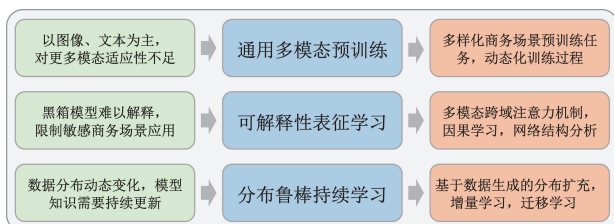


图2 复杂商务场景数据分析的难点及未来发展方向

题。目前,用于研究的商务数据集并不能完全覆盖实际应用遇到的数据类型,不同时间、不同环境或不同用户群体都会造成商务数据的分布差异。因此,训练数据集不能涵盖所有潜在的数据,会导致商务分析方法的泛化性下降。此外,不同于其他领域,复杂商务场景下的分析任务与时空和用户具有强相关性,比如在推荐系统中,即使是面向同一用户,推荐的商品也会随着季节的变化而变化^[29]。由于商务场景会持续产生海量数据,这也对模型提出快速适应新任务或数据,且在更新过程中不会遗忘历史知识的要求。因此,从商务数据中构建可持续学习方法,实现对持续变化的商务场景中时空和用户变化的高效泛化,已经成为目前研究的关键科学问题。

针对复杂商务数据持续变化的难点,目前的研究尝试通过增加数据集的多样性来提高模型的泛化性^[37, 38],但这往往受到数据收集和标记的成本和难度的限制。此外,一些研究考虑使用增量学习算法^[18, 19],例如知识蒸馏、迁移学习等技术,但由于新数据可能与旧数据存在概念漂移,仍可能导致模型性能下降。因此,现有方向尚未充分解决在复杂商务场景中时空分布变化和用户变化的泛化性问题。未来的研究方向应聚焦于研究新的持续学习框架和算法,在时空动态变化和用户行为演化的情境下实现持续学习的高效泛化。此外,如何构建学习算法,使得模型能够跟随用户需求不断演进更新,得到更个性化和更相关的结果,实现以用户为中心的专用商务模型,也是未来研究的关键方向。

4 总结与展望

本文讨论了基于泛化模型的复杂商务场景数据分析方面的相关研究。首先,从复杂商务场景数据分析的重要意义出发,介绍了目标数据的内在特点,以及相应分析方法所面临的共性挑战。其次,从数据和应用两个角度,分别详细介绍了当前在多模态商务数据通用特征的学习及多场景可泛化商务应用模型的构建两方面研究进展。最后,讨论了复杂商务场景数据分析方向仍待解决的关键技术问题。目前,对面向真实世界的多模态、可泛化、可解释、强鲁棒的数据分析方法研究仍处于探索阶段,距离实现复杂商务场景的全面智能化任重而道远。希望本文调研能为未来的相关研究提供初步参考,共同支撑我国在复杂商务场景数据中的探索,为提高我国国际经济地位、提升人民生活水平做出贡献。

参 考 文 献

- [1] 梅宏. 大数据与数字经济. 求是, 2022(2): 7.
- [2] 国务院. “十四五”数字经济发展规划. (2021-12-12)/[2023-12-28]. file:///C:/Users/Qiang/Downloads/ad75a0bdc6784c15beec545db98c61ab-3.pdf.
- [3] 怀进鹏. 打造数字贸易新引擎. 人民日报, 2020-09-06(05).
- [4] Dong X, Zhan XL, Wu YX, et al. M5Product: self-harmonized contrastive learning for E-commercial multi-modal pretraining. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 21220—21230.
- [5] Pratap S, Daultani Y, Dwivedi A, et al. Supplier selection and evaluation in e-commerce enterprises: a data envelopment analysis approach. Benchmarking, 2022, 29(1): 325—341.
- [6] Liu K, Xue F, Guo D, et al. Multimodal graph contrastive learning for multimedia-based recommendation. IEEE Transactions on Multimedia, 2023, 25: 9343—9355.
- [7] Zhang ZY, Gao HY, Yang H, et al. Hierarchical invariant learning for domain generalization recommendation. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach, CA, USA. ACM, 2023: 3470—3479.
- [8] Sturman O, von Ziegler L, Schläppi C, et al. Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 2020, 45(11): 1942—1952.
- [9] Mykhalchuk T, Zatonatska T, Dluhopolskyi O, et al. Development of recommendation system in e-commerce using emotional analysis and machine learning methods// 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. Virtual; IEEE, 2021: 527—535.
- [10] Zhu YS, Zhao HX, Zhang W, et al. Knowledge perceived multi-modal pretraining in e-commerce// Proceedings of the 29th ACM International Conference on Multimedia. Chengdu: ACM, 2021: 2744—2752.
- [11] Yu LC, Chen J, Sinha A, et al. CommerceMM: Large-scale commerce multimodal representation learning with omni retrieval// Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington: SIGKDD, 2022: 4433—4442.

- [12] Jin Y, Li YZ, Yuan ZH, et al. Learning instance-level representation for large-scale multi-modal pretraining in E-commerce// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 11060—11069.
- [13] Jang J, Ye S, Yang S, et al. Towards continual knowledge learning of language models// International Conference on Learning Representations. Virtual; Ithaca, 2022.
- [14] Sun Y, Wang SH, Li YK, et al. ERNIE2.0: a continual pre-training framework for language understanding// Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 8968—8975.
- [15] Corbière C, Ben-Younes H, Ramé A, et al. Leveraging weakly annotated data for fashion image retrieval and label prediction// Proceedings of the IEEE international conference on computer vision workshops. Venice: IEEE, 2017: 2268—2274.
- [16] Wei XS, Cui Q, Yang L, et al. RPC: A large-scale retail product checkout dataset. (2019-01-22)/[2023-12-28]. <https://arxiv.org/pdf/1901.07249.pdf>.
- [17] Zhan XL, Wu YX, Dong X, et al. Product1M: towards weakly supervised instance-level product retrieval via cross-modal pretraining. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 11762—11771.
- [18] Liu F, Chen DL, Du XY, et al. MEP-3M: a large-scale multi-modal E-commerce product dataset. Pattern Recognition, 2023, 140: 109519.
- [19] Qi D, Su L, Song J, et al. Imagebert: Cross-modal pre-training with large-scale weak-supervised image-text data. (2020-01-22)/[2023-12-28]. <https://arxiv.org/pdf/2001.07966.pdf>.
- [20] Gabeur V, Sun C, Alahari K, et al. Multi-modal transformer for video retrieval// Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020: 214—229.
- [21] Chen YC, Li LJ, Yu LC, et al. Uniter: Universal image-text representation learning// Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020: 104—120.
- [22] Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision// International conference on machine learning. Vienna: PMLR, 2021: 8748—8763.
- [23] Ramesh A, Pavlov M, Goh G, et al. Zero-shot text-to-image generation// International Conference on Machine Learning. Vienna: PMLR, 2021: 8821—8831.
- [24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need// Proceedings of the 31st International Conference on Neural Information Processing Systems. California: ACM, 2017: 6000—6010.
- [25] Li JN, Selvaraju RR, Gotmare AD, et al. Align before fuse: Vision and language representation learning with momentum distillation. Advances in neural information processing systems, 2021: 9694—9705.
- [26] Zhang PC, Li XJ, Hu XW, et al. Vinvl: revisiting visual representations in vision-language models// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE, 2021: 5579—5588.
- [27] Zhou LW, Palangi H, Zhang L, et al. Unified vision-language pre-training for image captioning and VQA// Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 13041—13049.
- [28] Jia C, Yang YF, Xia Y, et al. Scaling up visual and vision-language representation learning with noisy text supervision// International conference on machine learning. Vienna: PMLR, 2021: 4904—4916.
- [29] Yang C, Wu Q, Wen Q, et al. Towards out-of-distribution sequential event prediction: a causal treatment. Advances in neural information processing systems, 2022: 22656—22670.
- [30] Devlin J, Chang MW, Lee K, et al. Bert: pre-training of deep bidirectional transformers for language understanding. (2019-05-24)/[2023-12-28]. <https://arxiv.org/pdf/1810.04805.pdf>.
- [31] Gao D, Jin L, Chen B, et al. Fashionbert: text and image matching with adaptive loss for cross-modal retrieval// Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Virtual; ACM, 2020: 2251—2260.
- [32] Zhuge MC, Gao DH, Fan DP, et al. Kaleido-bert: vision-language pre-training on fashion domain// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual; IEEE, 2021: 12647—12657.
- [33] Zhu W, Lu L, Xiao J, et al. Localized adversarial domain generalization// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 7098—7108.
- [34] Zhou KY, Liu ZW, Qiao Y, et al. Domain generalization: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4396—4415.

- [35] Lv FR, Liang J, Li S, et al. Causality inspired representation learning for domain generalization// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022; 8036—8046.
- [36] Zhang J, Qi L, Shi YH, et al. Mvdg: a unified multi-view framework for domain generalization// Proceedings of the 17th European Conference on Computer Vision. Cham: Springer, 2022; 161—177.
- [37] Zhou KY, Yang YX, Hospedales T, et al. Deep domain-adversarial image generation for domain generalisation// Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020; 13025—13032.
- [38] Li P, Li D, Li W, et al. A simple feature augmentation for domain generalization// Proceedings of the IEEE/CVF International Conference on Computer Vision. Virtual: IEEE, 2021; 8866—8875.
- [39] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 2014, 15 (1): 1929—1958.
- [40] Guo JT, Qi L, Shi YH, et al. PLACE dropout: a progressive layer-wise and channel-wise dropout for domain generalization. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(3): 65.
- [41] Guo JT, Qi L, Shi YH. Domaindrop: Suppressing domain-sensitive channels for domain generalization// Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023; 19057—19067.
- [42] Zhou KY, Yang YX, Qiao Y, et al. Domain adaptive ensemble learning. *IEEE Transactions on Image Processing*, 2012, 30: 8008—8018.
- [43] Bucci S, D'Innocente A, Liao YJ, et al. Self-supervised learning across domains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44 (9): 5516—5528.
- [44] Zhuang FZ, Qi ZY, Duan KY, et al. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 2021, 109(1): 43—76.
- [45] Zhang L, Gao XB. Transfer adaptation learning: a decade survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(1): 23—44.
- [46] Zhang YX, Liu Y, Xiong H, et al. Cross-domain disentangled learning for E-commerce live streaming recommendation// 2023 IEEE 39th International Conference on Data Engineering (ICDE). Anaheim: IEEE, 2023; 2955—2968.
- [47] Fu WJ, Peng ZH, Wang SZ, et al. Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems// Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu: AAAI, 2019; 94—101.
- [48] Li P, Tuzhilin A. Ddtcdr: deep dual transfer cross domain recommendation// Proceedings of the 13th International Conference on Web Search and Data Mining. Houston: ACM, 2020; 331—339.
- [49] Kang S, Hwang J, Lee DH, et al. Semi-supervised learning for cross-domain recommendation to cold-start users// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019; 1563—1572.
- [50] Wang YQ, Feng CY, Guo CL, et al. Solving the sparsity problem in recommendations via cross-domain item embedding based on co-clustering// Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. Melbourne: ACM, 2019; 717—725.
- [51] 蒋昌俊. 算力网络与交易风控. *重庆邮电大学学报(自然科学版)*, 2023, 35(1): 1—7.
- [52] Weng HQ, Ji SL, Duan FZ et al. Cats: cross-platform e-commerce fraud detection// 2019 IEEE 35th International Conference on Data Engineering (ICDE). Macao: IEEE, 2019; 1874—1885.
- [53] Zhu YC, Xi DB, Song BW, et al. Modeling users' behavior sequences with hierarchical explainable network for cross-domain fraud detection// Proceedings of The Web Conference 2020. Taipei: ACM, 2020; 928—938.
- [54] Zeng GX, Chi JF, Ma R, et al. ADAPT: adversarial domain adaptation with purifier training for cross-domain credit risk forecasting. *International Conference on Database Systems for Advanced Applications*. Cham: Springer, 2022; 353—369.
- [55] 张钲, 朱军, 苏航. 迈向第三代人工智能. *中国科学: 信息科学*, 2020, 50(9): 1281—1302.
- [56] 李学龙. 多模态认知计算. *中国科学: 信息科学*, 2023, 53(1): 1—32.
- [57] Bai X, Wang X, Liu XL, et al. Explainable deep learning for efficient and robust pattern recognition: a survey of recent developments. *Pattern Recognition*, 2021, 120: 108102.
- [58] Yu Y, Sugiyama K, Jatowt A. AdaReX: cross-domain, adaptive, and explainable recommender system// Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region. Beijing: ACM, 2023; 272—281.

Generalizable Analysis Methods for Complex Business Data

Yinghuan Shi[†] Jintao Guo^{1†} Zekun Li¹ Lei Qi² Yang Gao^{1*}

1. *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023*

2. *School of Computer Science and Engineering, Southeast University, Nanjing 210096*

Abstract With the rapid development of big data technology, business scenarios are rapidly evolving, witnessing a substantial increase in both the scale and intricacy of data. While it brings accessible data resources, it concurrently presents challenges in effectively excavating and analyzing business data. This paper first introduces the characteristics of data in complex business scenarios, then introduces the research advancements in business analysis methodologies, encompassing large-scale business datasets, multimodal pre-training, feature fusion and generalizable business models across diverse scenarios. The paper provides an outlook on the future research directions of data analysis in complex business scenarios, analyzing the key issues that need to be solved urgently, including adaptability to more modalities, interpretability of business analysis model, and generalizability in dynamic environments. The paper summarizes the research progress and challenges encountered by generalizable models tailored for complex business scenarios, aiming to promote scientific research progress in complex business analysis and attaining comprehensive intelligence within complex business scenarios.

Keywords business scenarios; generalizable data analysis; deep learning; multi-modal fusion

(责任编辑 张强)

[†] Contributed equally as co-first authors.

* Corresponding Author, Email: gaoy@nju.edu.cn