

港航业大模型语料库建设亟需提质提效

□ 徐凯 李燕 郭胜童



天津港智慧码头。 本文图片由本报资料室提供

在当今人工智能技术迅猛发展的时代,产业大模型的应用已经渗透到各个领域,其中语料库在产业大模型应用中起到了关键作用。

虽然最近一年港航领域的大模型应用成果丰富,但语料库的建设还不尽如人意,需要提质提效,以促进港航业大模型高质量持续发展。

港航领域的大模型应用突飞猛进

最近一年,港航领域大模型落地案例频出,拥抱更多应用场景。

2023 年 6 月,西井科技发布了集装箱物流供应链大模型 TerminalGPT,它能够通过自主学习优化港口运营,提高作业效率和安全性;禾斯科技推出了 56GPT 人工智能产品,这是专为货代公司设计的产品,可通过自然语言处理和深度学习模型提供一站式服务,提高了工作效率和员工满意度。

2023 年 10 月,天津港集团联合多家企业启动了 PortGPT 的研发,覆盖了港口生产、物流服务等多个应用场景。天津港还创造了数字人助手“天天”,它在数据分析挖掘能力上已超越普通员工,能够提供实时的问题解决方案。

2023 年 11 月,招商轮船开发了基于“商道”行业大模型的 ShippingGPT,用于船舶管理、市场分析等多个领域,并通过“世界航商”APP 向全球用户开放。

今年 3 月,山东港口日照港与百度合作开发大模型,专注于件杂货码头的自动化和智能化,通过智能化系统,日照港的运转效率、设备利用率、堆场周转率和利用率均有显著提升。就在同一个月,广州港南沙三期开展了基于大模型的港口 GPT 一站式服务平台的研究与应用,实现了服务全流程的智能化,并通过大模型技术提升了智能服务水平。

此外,港联航科技自主研发了港口大模型系统——PortGLM,建立了港口码头领域的垂直语料库,包括港口码头相关法律法规、基础知识与生产操作知识、事故案例分析、风险评估模型等。该系统能与港联航的“神谋”超级自动化 iPaaS 平台结合,配置港口数字员工,提高港口运营智能化水平。

作为全球领先的数字员工机器人公司,壹沓科技聚焦大供应链领域,推出了“运小沓”供应链数字员工超自动化平台,实现供应链业务的全局超自动化,包括营销、履约执行和财务结算等方面。同时,壹沓科技还推出了“运小沓·CubeAgent”,提供虚拟数字员工专家团队,助力企业提高人才密度和实现生产力变革。

值得一提的是,新加坡投资 7000 万元开发了东南亚首个大型语言模型生态系统,以满足多元文化和多语需求,并助力新加坡成为人工智能解决方案的全球领导者。

这些应用展示了大模型在港口航运领域的多样化用途,包括提高作业效率、优化资源配置、增强决策支持和提升客户服务水平等。随着技术的进步,大模型在港口航运领域的应用将更加广泛和深入。

港航垂直语料库需要体系化建设

港口航运垂直语料库的体系化建设对于航运领域的发展至关重要,应重视其建设和维护,以期对港口航运业的发展提供有力支持。

港口航运垂直语料库的建设需要全面考虑各类数据。这不仅包括传统的结构化数据,如统计数据和业务数据,还应涵盖非结构化数据和多模态数据。这些数据类型在航运领域的应

用日益广泛,对于理解和分析港口运营状况、预测市场趋势以及优化物流流程至关重要。

上海国际航运研究中心正在受上港集团、山东港口集团、南京港口集团、广州港口集团等单位的委托起草《港口大数据分类与编码》,已经将语料数据考虑在大数据体系之中。

港口航运垂直语料库的构建需要一个

科学的目录体系。上海国际航运研究中心已对港口领域的语料进行了细致的梳理,形成了 12 个一级分类和 72 个二级分类。一级分类包含港口概述、港口管理与运营、港口设施与设备等,每个一级分类下又细分为货物装卸设备、装卸工艺类型与特点、港口物流链优化等不同的二级分类。这种分类体系有助于清晰地展示港口航运领域的

知识结构和信息内容。

港口航运垂直语料库需要专业词向量和知识图谱的构建。通过对专业词汇的词向量进行梳理,才可以更好地理解词汇之间的关系和含义,进而构建出知识图谱。同时,还需梳理词向量中与其他数据对接的关联概念,如统计数据中的货类细分等,以实现数据之间的有效衔接和利用。

专业语料要重“量”更要重“质”

专业领域的语料库建设,往往更看重语料的专业性和质量,而不是一味的追求语料的数量。

首先,语料库的规模是衡量其价值的基础。一个庞大的语料库能够为大模型提供丰富的训练数据,从而提高模型的性能和泛化能力。然而,规模并非唯一重要的因素。如果仅仅追求数量而忽视质量,那么语料库中可能充斥着大量低质量的、重复的或者无关紧要的数据,这不仅无法提升模型的效能,反而可能引入噪声,降低模型的准确性和可靠性。

其次,语料质量是决定模型效能的关键。高质量的语料能够提供准确、一致的信息,有助于模型学习到更加真实和有效的知识。而低质量的语料

则可能导致模型学习到错误的信息或者产生误导性的输出。因此,在语料库建设中,必须对语料进行严格的质量控制,确保其准确性和可靠性。

最后,内容分布的均衡性也是评价语料库的重要指标。一个均衡的内容分布能够保证模型在各种情境下都能够做出合理的预测和决策,而内容分布的不均衡则可能导致模型在某些方面的过度拟合或者在另一些方面表现欠佳。因此,在语料库建设中,需要注重内容的多样性和均衡性,以确保模型在各种场景下的适用性。

如果不进行科学评价,很可能会有人利用 GPT 等工具批量生成大量内容重复而只有形式差异的语料,无法为大模型带来真正的价值。而“综合语料

价值指数”(Comprehensive Corpus Value Index,简称 CCVI)综合考虑了语料的质量、所属分类的权值以及数量,可以更客观地反映垂直领域大模型语料的资源价值(CCVI=语料质量×语料分类权值×语料数量)。

CCVI 评价的引入,不仅能够帮助用户更加科学地评价和选择语料库,还能够引导在语料库建设中注重质量、多样性和均衡性,从而为大模型提供更具优质、有效的训练数据。

在未来的发展中,业内人士需要更加注重语料库的质量和多样性,同时也应关注其持续更新和扩充,以适应知识的不断发展变化。只有这样,才能构建出更加精准、高效的人工智能模型,为港航业的发展提供有力支持。



广州港南沙港区。

货代业的五大新趋势

□ 王禹



俯瞰盐田港。本报资料室供图

当前,全球经济进入高通胀、低增长的新阶段。受此影响,航运市场面临需求收缩,叠加新船交付导致的供给上升,供需关系趋紧。在这一背景下,货代企业将不得不直面贸易前景不明、地缘局势不稳定和技术颠覆带来的各种难题,但与此同时,挑战中也蕴藏着机遇,能够把握行业趋势的企业,将在新一轮发展中受益。

产能增加需求波动

2024 年,运输需求下降和运力规模增加的双重压力,将继续影响港航物流业。根据波罗的海国际航运公会(BIMCO)的预测,2024 年,全球集装箱海运需求增长 3%—4%,而全球海运船队运力则将增长 7.8%,达到创纪录的 270 万 TEU。马士基也预计,2024 年全球集装箱海运货量增长 2.5%—4.5%,市场将面临严重的运力过剩。对货代业而言,市场供需失衡将加剧企业间的竞争,同时可能会动摇客户的忠诚度。这意味着,货代企业要想在新一轮竞争中脱颖而出,需要更加努力地修炼内功,例如提高客户服务的响应速度、满足客户的个性化需求、提供更多增值服务等。同时,要想取得理想的业绩,还要进一步提高运营效率,降低各种成本,做好过苦日子的准备。

突发事件破坏性增强

近几年,货代领域的突发事件频发。2024 年,这些事件仍将影响货代业。但与以往不同的是,这些事件将从意外的、随

机的事件,转变为气候变化以及地缘紧张局势等方面,对行业的破坏性可能更强。

有研究表明,干旱、飓风和野火等自然灾害正变得越来越频繁。例如,2023 年,低水位影响了包括巴拿马运河、中国长江和欧洲莱茵河等在内的全球多条水路运输通道。特别是低水位使得巴拿马运河限制船舶通行,造成大量船舶排队、拥堵,更多船舶不得不更改航线,加剧了全球供应链紧张的局面。

此外,全球多地暴发地区冲突,地缘紧张局势加剧,并且有进一步扩大的趋

势,这些都将影响供应链运营和投资。正如经济学家智库提到的那样:当前对供应链的投资必须要进行调整,以将政治风险降至最低。

在这种市场环境下,促使货代企业优先考虑运输的灵活性和可见性,以便能快速有效地应对供应链紧张。

兼并浪潮来袭

2000 年以来,货代业的并购就日益活跃,2024 年依然没有放缓的迹象,甚至有

业内机构预测,2024 年将是货代业整合的一年。

在一个充满挑战的市场中,小型货代企业的经营状况堪忧,特别是容易面临资金短缺的问题。而大型货代企业将抓住机会,收购竞争对手及其客户。

进入 2024 年,德迅、嘉里物流等知名货代企业,都官宣了新收购,分别收购了亚洲跨境公路物流服务商 City Zone Express 和法国国际货运公司 Business By Air SAS,以增强自身在新兴市场的竞争力。

而近期最引人关注的货代市场收购,当属德国国有铁路运营商德铁集团旗下物流子公司德铁信可(DBSchenker)出售案。货代巨头 DSV 及航运巨头马士基,都对德铁信可表现出浓厚兴趣。特别是 DSV 曾明确表示,有兴趣收购德铁信可。以业务量计算,一旦 DSV 和德铁信可合并,将成为全球最大的货代企业,改写全球货代业格局。

可持续发展

2024 年是港航物流业脱碳之旅的重要一年。

欧盟碳排放交易体系(EU-ETS)已于 2024 年 1 月起正式对航运业实施,运营涉及欧盟国家航线的船公司,将面临越来越严格的碳排放要求,运营成本也随之增加。同时,国际海事组织(IMO)也将更密切地关注行业减排,并进一步明确已出台的部分法规的实施细则,以进一步规范船公司运营。

当前,越来越多的企业致力于提高自

身供应链的可持续性。根据联合国全球契约组织针对全球企业 CEO 的大规模调查显示,近一半企业的 CEO 已将供应链责任视为企业可持续发展战略的一部分。

不过,在艰难的市场环境下,这种势头能否保持下去,还有待观察。同时,面对减排带来的额外成本,有多少客户能够接受,也有较大变数。

数字化建设

数字化发展成为港航物流业的大趋势已毫无争议,对货代企业的重要性也不言而喻。无论是自身发展的需要,还是客户需求的推动,越来越多的传统货代企业都在数字化领域进行大量投资。根据知名上市咨询公司埃森哲的研究,76%的货代物流企业认为,不注重数字化能力建设,将“严重危及自身业务”。

目前,许多货代企业已经在货运过程中实现了追踪功能,并且不断提高数字平台的服务。这一趋势将会持续下去,因为货代企业越来越意识到,只有提供更好的数字化服务,才能满足现有客户的需求,并获得更多客户的青睐。同时,数字工具还可以最大限度地提高货代企业预测市场和应对危机的能力,便于及时做出决策。

不过,在进行大量数字化投资后,企业的盈利能力仍需关注。以美国知名科技型货代公司飞协博(Flexport)为例,其迫于经营压力进行了架构调整,2023 年至今已进行了多轮裁员。数字化之路该怎么走,数字化的步子要迈多大,仍需要货代企业进一步探索。