

智慧海关税收征管局干部能力需求分析

——基于文本分析技术

闻萍^{1*} 侯彩虹²

1. 海关总署税收征管局（广州） 广东省广州市 510623

2. 上海海关学院海关与公共管理学院 上海市 201204

摘要：智慧海关建设背景下，海关税收征管人员面临数字化转型挑战，需具备复合能力以应对复杂环境。本文以海关税收征管局干部为研究对象，使用TF-IDF方法对智慧海关、智慧税收征管相关的政策文件、制度和规范性文件中进行研究，通过词频（TF）和逆文档频率（IDF）分析了智慧海关税收征管局的胜任力要素。首先搭建了包括态度能力、专业能力、技术能力和管理能力四个维度的胜任力评价指标，其次抽取了每种能力所需的前5位高频词。该研究为评估海关税收征管局干部胜任力提供了理论的框架，为推进智慧海关建设奠定了良好的基础。

关键词：海关税收征管；胜任力评价；文本分析法；智慧海关

1. 引言

在全球贸易数字化与海关智能化背景下，智慧海关成为海关治理能力建设的主要内容。智慧海关不仅仅是使用先进技术，更是对海关管理模式和业务流程的深度变革。智慧税收征管是智慧海关创新发展的重要需求，也是税收征管工作满足智慧化建设的重要体现。《智慧海关建设总体方案》提出构建以数据驱动、风险防控为核心的征管体系，要求海关人员转型为“技术+专业”复合型人才。然而，随着全国通关一体化改革，海关税收征管局（简称“税管局”）的人力资源和能力结构尚未满足政策需求。现有海关胜任力评价研究多侧重于通用素质评价。统计数据显示，海关税管局干部具备专业良好的关税与数据分析能力，但对新技术使用的适应度不足。现有研究多集中于技术如何赋能海关管理，如信息化流程创新（杜晓英等，2023）、区块链监管（Segers et al., 2019）等，但未就人才应有的技术能力需求开展研究。世界海关组织（WCO）提出“技术-管理-道德”三维能力建设框架，立足各个国家海关，侧重宏观层面要求，难以指导具体岗位的能力建设（WCO, 2019）。基于此，本文以税管局推进海关总署智慧海关建设要求的相关的制度规范为基础，构建“态度-专业-技术-管理”四维胜任力模型，并通过实证分析识别能力短板，据此提出优化策略，为海关人力资源管理和智慧政府建设提供参考。

2. 智慧海关税收征管局能力评价研究现状

2019年，中国海关提出“智慧海关、智能边境、智享联通”理念，推动国际海关合作，构建智能化治理模式。海关税收征管不仅关乎国家财政收入的保障，也服务于对外开放、推动产业结构升级、促进经贸合作等方面，具有重要意义。智慧海关要求税收征管人员具备复合能力：技术层面需掌握数据分析和智能系统；专业层面需具备关税技术与新业态知识；管理层面需适应扁平化组织并协调多方合作。然而，现行海关人力资源体系配置仍停留在传统要求，侧重“职能分工”为导向，缺乏对上述复合能力的系统性整合，难以满足智慧海关建设对复合型人才的需求。海关干部的技术素养和创新能力依然是制约智慧海关落地的主要制约因素（庄妙慧，2024；徐少晖，2023）。

自McClelland（1973）提出胜任力理论以来，该理论逐步从企业扩展到公共部门。国内外公共部门人力资源研究中广泛应用着胜任力模型。西方政府胜任力模型研究在公务员能力标准的建立上取得了显著进展，如美国的行政核心资格和英国的PSG胜任力框架。而国内公共部门胜任力模型研究经历了从企业引入到政府，再由推广再到价值回归的发展过程，涵盖了各类公共组织人员。国内研究方法主要包括定性和定量相结合，应用广泛于干部选拔、培训开发等领域。国内公共部门胜任力指标的识别包括通用能力指标和特色

指标两种,但针对特定岗位的研究仍有待加强。

经典的“冰山模型”将胜任力分为表层技能与深层特质,而“洋葱模型”则强调动机、自我认知等内隐因素的动态交互 (Spencer & Spencer, 1993)。这些理论为胜任力的多维分析提供了基础,揭示了胜任力不仅包括外显的技能和知识,还涉及内隐的动机和态度。现有研究尽管已构建了适用于海关岗位的资格体系,但仍存在一定局限,尚未充分涵盖智能海关背景下所需的新兴能力,如数据分析等 (宁波海关课题组, 2020; Xie, 2024)。

TF-IDF 作为经典文本特征量化方法,是信息检索与自然语言处理领域的基石性算法,该算法核心假设有两个,其一包括词的重要性与其在文档内频率 (TF) 正相关,其二词的重要性与语料库中分布广度 (IDF) 负相关。该算法为文本表征提供了可解释的统计框架。TF-IDF 凭借低成本与可解释性,在高维稀疏场景 (如法律文书分类) 和社交网络分析中使用广泛;此外,使用时序加权 TF-IDF 算法可以成功捕捉舆情演化关键节点 (Zhang et al., 2020);在金融风险预警中,熵加权 TF-IDF 成功识别年报文本中的风险语义标记 (Loughran & McDonald, 2016)。本研究选用 TF-IDF 算法,因其既能衡量词语在单篇文本中的重要性,又能综合考虑其在整体语料库中的分布特征,能够突出核心术语,剔除常见词汇,适用于提取与海关干部胜任力需求密切相关的关键信息。

3. TF-IDF 原理及其在税收征管局胜任力评价中的应用

TF-IDF (Term Frequency-Inverse Document Frequency) 是一种常用于文本分析的统计算法,主要用于衡量一个词在文档中的重要性。它结合了词频 (TF) 和逆文档频率 (IDF) 两个概念,通过综合考虑某个词在单个文档中的频率以及该词在整个文档集中的普遍性,来评估该词对特定文档的重要程度。

词频 (TF, Term Frequency): 表示某个词在单篇文档中出现的频率。词频越高,表明该词在该文档中越重要。但如果某个词在整个文档集中都很常见,它的重要性就会降低。词频计算公式为:

$$TF(t, d) = \frac{\text{词 } t \text{ 在文本 } d \text{ 中出现的次数}}{\text{文本 } d \text{ 中的总词数}} \quad (3-1)$$

逆文档频率 (IDF, Inverse Document Frequency): 用来衡量某个词在整个文档集中的普遍性。如果某个词在很多

文档中都出现,那么它的区分度较低, IDF 值就会较小。相反,出现较少的词则具有较大的 IDF 值,表明其在特定文档中具有更大的辨别力。IDF 的计算公式为:

$$IDF(t, D) = \log \left(\frac{\text{文本集 } D \text{ 中的文本总数}}{1 + D \text{ 包含词 } t \text{ 的文本数量}} \right) \quad (3-2)$$

TF-IDF 值的计算: 通过将词频与逆文档频率相乘,得到 TF-IDF 值,计算公式为:

$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3-3)$$

LAD (Latent Dirichlet Allocation) 模型用于从文档中自动发现主题。它假设每篇文档是由多个主题混合而成,而每个主题又由一组单词组成。LDA 通过推断每个文档中各主题的比例和每个主题中单词的分布,来实现主题聚类。该方法广泛应用于文本挖掘、推荐系统和信息检索等领域,能有效地帮助研究者深入分析文本数据的潜在结构 (张涛、马海群, 2024)。

本研究构建海关税管局四维胜任力框架,通过“通用-特定”双层次结构应对多维岗位需求。通用胜任力聚焦基础素养: 基于“冰山模型”的政治忠诚、廉洁自律等态度要素,以及融合“洋葱模型”的跨部门协作、行政管理等管理能力;特定胜任力突出业务适配: 对接 WCO 标准的税收政策执行、风险防控等专业能力,以及支撑智慧海关转型的数据分析、参数化模型应用等技术能力。该框架既整合经典理论要素,又细化智能技术场景下的新型能力标准,为后续模型开发提供结构化基础。

4. 实证分析

本研究通过文本分析法,从政策文件、海关制度及内部规章中提取胜任力要素。文本资料来源于三个方面,具体包括: 国家政策文件,包括《公务员法》《公务员平时考核办法 (试行)》等,共 19 份;海关总署制度,如《“十四五”海关发展规划》《海关专业人才发展规划》等,共 16 份;税管局内部规章,涵盖《内设处室职责分工》《绩效评估指标体系》等,共 7 份。总计 42 份文档,形成 7.4 万字的文本库。

研究使用 Python 的 Jieba 库进行分词,剔除停用词 (如“的”“了”) 及低频词 (出现次数 < 5 次),保留 10695 个有效词语。

首先,计算词语权重,保留 TF-IDF 值 > 0.3 的关键词 (表 1),得到专业胜任力中的“风险防控”,技术胜任力中的“数据分析”、态度胜任力中的“廉洁自律”与管理胜任力中的“改

革攻坚”，对应的TF-IDF值分别为0.81、0.85、0.65和0.77。

表1 TF-IDF高频词示例(部分)

关键词	TF-IDF值	所属主题类别
风险防控	0.81	专业胜任力
数据分析	0.85	技术胜任力
廉洁自律	0.65	态度胜任力
改革攻坚	0.77	管理胜任力

LAD (Latent Dirichlet Allocation) 模型用于从文档中自动发现主题。它假设每篇文档是由多个主题混合而成，而每个主题又由一组单词组成。LDA通过推断每个文档中各主题的比例和每个主题中单词的分布，来实现主题聚类。该方法广泛应用于文本挖掘、推荐系统和信息检索等领域，能有效地帮助研究者深入分析文本数据的潜在结构(张涛、马海群, 2024)。通过困惑度(Perplexity)确定最优主题数为4，生成“态度、专业、技术、管理”四类主题词(表2)。将提取的主题词与《海关专业发展标准》(WCO, 2019)对比，一致性达82%(Kappa系数=0.75)，表明结果有效。

表2 LDA主题词分布(部分)

主题类别	高频词(前5位)
态度胜任力	政治能力、责任担当、廉洁自律、服务意识、心理调适
专业胜任力	关税技术、风险研判、行业分析、国际合作、减免税管理
技术胜任力	数据挖掘、参指模应用、系统操作、技术适应、安全保密
管理胜任力	改革攻坚、科学决策、团队协作、调查研究、选人用人

基于文本分析结果，采用德尔菲法对胜任力要素进行多轮筛选，确保指标的科学与适用性。组成8人的专家小组，人员涵盖税管局处级及以上领导(3人)、关税专家(2人)、海关人事管理者(2人)、公共管理学者(1人)，均具有副高级以上职称，平均从业年限18年。专家权威系数(Cr=0.87)，表明咨询结果可信。通过三轮德尔菲法，最终形成4个一级维度、16项二级指标(表3)，专家协调系数(W=0.36, p < 0.01)通过检验。

表3 海关税管局胜任力指标体系

一级维度	二级指标
态度胜任力	政治能力、责任担当、廉洁自律、心理调适
专业胜任力	关税技术、风险防控、行业管理、国际合作
技术胜任力	数据分析、参指模应用、系统操作、技术适应
管理胜任力	改革攻坚、科学决策、调查研究、团队协作

5. 结论

本文通过TF-IDF文本分析法，对海关税收征管局干部的胜任力要素进行了研究，构建了涵盖态度能力、专业能力、技术能力和管理能力的四维胜任力评价框架，并识别出各维

度的关键高频词。研究为评估海关税收征管局干部的胜任力提供了理论框架，明确了智慧海关建设背景下干部所需的核心能力要素，为海关人力资源管理和智慧海关建设提供了科学依据和参考。

参考文献:

- [1] 杜晓英, 王晓林. 我国海关促进贸易便利化的措施研究[J]. 对外经贸实务, 2023,(06): 64-69.
- [2] Segers L, Ubacht J, Tan Y H, et al. The use of a blockchain-based smart import declaration to reduce the need for manual cross-validation by customs authorities[C]. Proceedings of the 20th Annual International Conference on Digital Government Research. 2019: 196-203.
- [3] World Customs Organization. ProfessionalStandards[EB/OL].(2019-04-01)[2025-02-04].
<https://www.wcoomd.org/-/media/wco/public/global/pdf/topics/capacity-building/activities-and-programmes/picard/professional-standards/omd-normes-prof-uk-basse-def.pdf?la=fr>
- [4] 庄妙慧. 新时代海美国门生物安全人才队伍建设面临的困境与对策研究[J]. 海关与经贸研究, 2024,45(03):57-68.
- [5] 徐少晖. 互联网众筹税收征管问题研究[D]. 上海海关学院, 2023.
- [6] McClelland D C. Testing for competence rather than for "intelligence." [J]. American psychologist, 1973, 28(1): 1.
- [7] Spencer L M, Spencer P S M. Competence at Work models for superior performance[M]. John Wiley & Sons, 2008.
- [8] 宁波海关课题组, 王继蓬. 基于岗位胜任力构建海关资质考核测评体系研究[J]. 海关与经贸研究, 2020,41(06):35-47.
- [9] 谢竹歆. 基于胜任力模型的DX海关培训需求识别研究[D]. 广西大学, 2024.DOI:10.27034/d.cnki.ggxixu.2024.001104.
- [10] Zhang Y, Zhou Y, Yao J T. Feature extraction with TF-IDF and game-theoretic shadowed sets[C]//Information Processing and Management of Uncertainty in Knowledge-Based Systems: 18th International Conference, IPMU 2020, Lisbon, Portugal, June 15 - 19, 2020, Proceedings, Part I 18. Springer International Publishing, 2020: 722-733.

[11] Loughran T, McDonald B. Textual analysis in accounting and finance: A survey[J]. *Journal of Accounting Research*, 2016, 54(4): 1187-1230.

[12] 张忠, 马浩. (2024) 政策文本计算: 理论、方法与实证研究. 中国社会科学出版社, (63).

作者简介: 闻萍 (199206 —), 女, 汉族, 江苏靖江人, 大学本科学历, 工学学士学位, 现工作于海关总署税收征管局 (广州)。主要研究方向为干部胜任力评价与税收征管信息化。