**基于“物-事-证”的多维特征提取与描述算法研究报告**

**课题名称： 海关通关监管时空风险分析与预警技**

**术研究**

**起止时间： 2023年11月-2026年10月**

**承研单位： 浙江大学**

**课题负责人： 席萌**

**评审意见修改表**

|  |  |
| --- | --- |
| 课题名称 | 海关通关监管时空风险分析与预警技术研究【2023YFC3306303】 |
| 课题成果名称 | 基于“物-事-证”的多维特征提取与描述算法研究报告 |
| 评审意见修改内容 | |
| 针对专家郭平提出的 “继续修改完善” 建议，将进一步深化 “物 - 事 - 证 - 场 - 人” 多维特征算子与描述算法的普适性研究，结合更多海关实际监管案例（如空运快件、陆运中欧班列等）验证算法性能，优化算子逻辑与阈值设置，提升复杂场景下的风险识别精度。  针对专家提出的 “算法优化与完善、系统集成与推广” 建议，将深化加权组合算子、多维相似性算子等核心算子的泛化能力研究，通过引入自适应权重调整机制提升普适性；同时优化大数据处理模块，提升算法在海量数据场景下的运行效率；加快与海关现有监管系统的接口开发，推动算法在更多关区、更多监管场景（如跨境电商、保税监管）的落地复用，并建立系统迭代优化机制。  针对专家孙敏提出的 “强化算法优化与测试，引入先进技术” 建议，将收集全国各关区不同时期、不同运输方式（海运、空运、陆运）的通关数据，构建大规模、多样化测试数据集；基于数据集对算法结构与参数进行迭代优化，降低复杂环境下的误差；同时引入迁移学习解决不同场景数据分布差异问题，引入强化学习提升算法对新风险模式的自适应能力，增强鲁棒性。  针对专家王斌提出的 “场景拓展、系统融合” 建议，将重点验证算法在空运口岸危险品监管、陆运中欧班列换货风险防控等场景的适用性，优化算法实时处理模块以满足高并发监管需求；加快与海关 “智慧海关” 系统的集成，建立 “算法输出 - 监管反馈 - 参数调整” 的动态反馈机制，同时联合法规部门梳理算法应用的合规边界，确保技术应用符合现行海关法规要求。  针对专家张祥林提出的 “完善多维特征算子与描述算法体系，扩展应用场景” 建议，将进一步补充 “场 - 人” 维度的特色算子（如人员行为序列分析算子、监管场地风险叠加算子），完善算法体系；同时在现有海运危化品监管、进口大宗散货监管场景基础上，拓展至旅客行李监管、快件监管等场景，通过标准化描述算法提升指标跨场景复用效率。  修改人：杨灏哲  2025年7月10日 | |

**目录**

[一、研究背景 1](#_Toc2062539744)

[二、业务现状 2](#_Toc424447534)

[三、解决方案 3](#_Toc1896512251)

[3.1 多维风险特征提取方法 4](#_Toc1729113783)

[3.1.1 算子设计与符号定义 4](#_Toc1466639677)

[3.1.2 属性算子 5](#_Toc995751073)

[3.1.3 异常指标算子 11](#_Toc248222840)

[3.2 多维风险特征命名规则与描述算法 15](#_Toc1468029406)

[3.2.1 风险指标统一标识 16](#_Toc730606259)

[3.2.2 风险指标描述方法 17](#_Toc2135385114)

[四、应用案例 20](#_Toc670902334)

[4.1 应用场景说明 20](#_Toc1566380788)

[4.2 应用场景案例 20](#_Toc159875343)

[4.2.1 输入报关单信息 20](#_Toc522847404)

[4.2.2 复杂算子组合与风险指标构建 21](#_Toc2140719151)

[4.2.3 风险指标描述生成流程 22](#_Toc125749019)

[五、总结 24](#_Toc339853685)

## **研究背景**

海关监管在维护国家经济安全和国际贸易秩序中发挥着至关重要的作用。2023年，中国海关进出口总额达到41.76万亿元人民币，同比增长0.2%。然而，巨大的通关量和复杂的业务条线给海关监管带来了前所未有的挑战。2022年仅走私方面海关立案侦办案件已达4509起，案值1210亿元，前三季度扣留侵权嫌疑货物4万批，扣留侵权嫌疑货物5977万件。这些风险事件不仅严重扰乱了市场秩序，还对国家税收安全造成了巨大威胁。海关通关涉及范围广、内容复杂、业务条线众多，大量通关业务仅靠人工进行监管任务重、压力大。因此，智慧海关等一系列数字化智能化赋能工作迫在眉睫。

海关监管面临的核心挑战在于多时空要素风险关联的内隐性、态势判别的复杂性以及跨部门联动的局限性。从多时空要素风险关联关系内隐角度来看，海关业务涵盖的时间跨度长、空间范围广，不同要素间的关联关系不仅复杂而且内隐，例如存在一批违规货物分散从不同海关报关进入的情况，这种跨关区的风险漂移状况依靠单一维度的数据很难有效识别。在态势判别方面，由于缺乏综合性的多维特征提取与描述方法，海关面对复杂业务场景时，难以快速且精准地判别风险态势，如在海运口岸出口危化品监管中，对企业货物申报信息、物流信息等多源数据的综合分析能力不足，导致伪瞒报风险不能及时被发现。而在联动分析能力上，当前海关系统在跨部门、跨区域的数据共享和联动分析环节存在明显短板，比如在陆运口岸中欧班列监管中，防控公路运输过境集装箱换货风险需要多个海关协同配合，但现有技术手段难以实现高效的信息共享和指挥调度。

具体到风险指标模型体系的建设中，当前主要存在两方面的挑战：1）单指标阈值预警误报率高，以陆运监管为例，仅凭"运输车辆中途停车时间＞1小时"的简单规则触发预警，导致大量误报信息上传，一线监管人员不得不选择性忽视该指标，导致存在风险指标的无效建设；2）指标命名描述体系混乱，各关区自主开发风险评估模型和风险指标存在大量命名重复或描述歧义，如"运输异常"在既可指轨迹偏离，也可以指速度波动，这种碎片化建设导致既有指标的可读性、可复用性下降，大量重复建设的同时风险指标模型体系难以完善。

鉴于上述背景，本工作围绕"物-事-证"展开多维特征提取与描述算法研究。其中物维度涵盖商品属性、运输工具、包装特征等实体要素，如危化品的UN编号、集装箱铅封状态等；事维度包括物流轨迹、企业行为、交易模式等动态特征，如中欧班列中途停靠异常等；证维度涉及许可证件、企业资质等合规要素，如原产地证书真伪鉴别等。具体研究内容包括：1）多维风险指标算子设计，基于单维风险评估指标设计多维风险指标计算算子，包括属性算子、二元逻辑算子、三元逻辑算子等；2）建立标准化的多维风险指标命名规范，解决指标命名冲突、重复建设等问题；3）研发多维风险指标智能化描述方法，实现自动化的自然语言描述生成，提高指标模型研发效率，增加指标模型的可读性和可复用性。最终提升海关通关监管工作的精准度与效率，为构建全国统一的海关智能风控体系提供核心技术支撑。

## **业务现状**

目前海关系统在风险防控方面主要采用"一控二控"分级管控体系，各直属海关及业务部门自主开发风险指标模型。以上海海关为例，其开发的"跨境物流风险指数"包含12类核心指标，而深圳海关的"通关安全评分"体系则侧重7类运输异常特征。在实际运行中，关员需综合参考HS编码风险参数、企业信用等级、物流轨迹评分等多维度数据，通过人工经验判断风险等级。这种模式下，全国形成了20余套差异化预警模型，日均处理超10万条报关单的评估工作，虽然能结合本地特点开展监管，但普遍存在指标定义不统一、阈值设置差异大等问题。具体问题总结如下：

1）单指标误报率高易被规避。一线人员多依赖手工定制风险指标来开展相关工作，对于每个手工定制的风险指标，还需设置报警阈值。但由于缺乏科学、精准的设定依据和方法，多是凭借经验来操作，这使得单个指标普遍较为简单，难以全面、综合地考量复杂多变的实际业务情况。导致误报率居高不下，常常出现大量不必要的报警信息，不仅干扰了正常的监管工作流程，也增加了后续筛选、甄别有效风险信息的难度。而且，单维指标容易被不法分子察觉、规避，需要在获取众多风险指标的相关反馈后，人工进行统筹分析。由于数据来源多样、指标繁杂，人工去梳理、整合和分析这些信息难度极大，容易出现疏漏或者误判的情况，很难从整体上准确把握风险态势，无法及时、精准地发现真正的风险隐患。总署 2022 年对 15 个主要关区的调研显示，单指标平均误报率高达 35.2%，如某口岸 "运输车辆中途停车超 1 小时" 指标触发的预警中仅 8.7% 为真实风险，其余 91.3% 为司机休息、交通管制等正常情况。深圳海关 2021 年 "申报金额异常波动" 指标日均产生 2,300 余条无效报警，误报占比达 38.5%。上海海关 2022 年查处的伪瞒报案件中，72% 的涉案企业通过分拆订单、分散申报等方式规避单指标阈值监管，某企业曾利用单指标漏洞实施低报行为达 17 次。技术设备层面的数据同样佐证了这一问题，这些数据表明，单指标监管模式已难以适应复杂业务场景，亟需通过多维特征分析提升风险识别效能。

2）复杂指标可复用性差、体系维护困难。在风险指标定制环节，一线人员往往需要针对不同情况手工进行设置，这就导致大量重复设计的情况出现，缺乏统一、高效的标准规范来指导指标定制，使得整个过程效率较为低下。珠三角某海关投入开发的"水客行李关联分析模型"，虽能通过"旅客历史记录+行李X光图像特征"实现85%的识别准确率，但因指标定义未标准化，无法在其他口岸复用。调研显示，全国各关区近三年开发的380个复杂指标中，仅有15%实现了跨关区共享，导致同类场景重复开发成本超2000万元。典型案例是某跨境电商监管指标，长三角、粤港澳等5个关区分别开发了功能重叠但算法迥异的"订单拆单风险模型"，不仅造成资源浪费，还因标准不统一导致同一企业在不同关区风险评级差异达40%。

## **解决方案**

当前，海运口岸出口危化品监管的风险识别方法主要依赖单一特征和单风险指标，存在准确率和召回率低的问题，同时缺乏系统的多维特征提取方法与标准化命名规则，导致风险指标管理混乱，难以实现高效分析和协同监管。多维风险特征是基于“物-事-证-场-人”中的多个维度进行风险特征的组合计算设计，通过多维指标统筹考虑货物基础属性（如类别、来源地）、关键事件（如报关、运输）、证件信息（如合法性与使用情况）、时间特性（如周期规律、高峰期）、空间分布（如运输路径和区域风险）、以及相关人员的行为记录（如发货人、承运人历史风险）。通过设计多维特征提取方法，结合标准化命名规则与智能描述算法，实现对多维数据的统一描述与管理。风险评价指标体系旨在服务于出口危化品监管场景，通过全面量化与动态预警，实现高风险货物的精准识别，提升查获效率，同时优化海关监管流程，确保物流安全与合规。

### 3.1 多维风险特征提取方法

本节将算子按照输入数据的类型分为两类：以属性为输入数据的算子和以异常指标为输入数据的算子。其中，以异常指标为输入数据的算子分为二元和三元算子。所有输出数据均为是否有异常，每类算子根据其特点进行定义和描述。

#### 3.1.1 算子设计与符号定义

（1）属性算子

表 1 属性算子表

|  |  |  |
| --- | --- | --- |
| 算子名称 | 符号 | 定义 |
| 对比算子 |  | 比较两个特征值的大小关系。 |
| 差值算子 |  | 计算两个特征值的差值。 |
| 乘法算子 |  | 计算两个特征值的乘积。 |
| 比值算子 |  | 计算两个特征值的比值。 |
| 差分比率算子 |  | 计算差分值相对于基准值的比率。 |
| 子集判断算子 |  | 判断一个集合是否为另一个集合的子集。 |
| 均值算子 |  | 计算两个特征值的平均值。 |
| 方差算子 |  | 计算两个特征值的方差。 |
| 欧几里得距离算子 |  | 计算两点间的欧几里得距离。 |
| 加权组合算子 |  | 计算多个特征值的加权综合值。 |
| 交叉偏差算子 |  | 计算多个特征值的交叉偏差。 |
| 多变量方差算子 |  | 计算多个特征值之间的方差。 |
| 三维欧几里得算子 |  | 计算三维空间中点之间的欧几里得距离。 |
| 联合概率算子 |  | 计算多个事件或特征的联合发生概率。 |
| 多维相似性算子 |  | 计算三维空间的余弦相似度。 |

（2）异常指标算子

表 2 二元异常指标算子表

|  |  |  |
| --- | --- | --- |
| 算子名称 | 符号 | 定义 |
| 逻辑与运算算子 |  | 对两个特征值进行逻辑与运算。 |
| 逻辑或运算算子 |  | 对两个特征值进行逻辑或运算。 |
| 逻辑异或运算算子 |  | 对两个特征值进行逻辑异或运算。 |
| 逻辑蕴含运算算子 |  | 对两个特征值进行逻辑蕴含运算。 |
| 逻辑与非运算算子 |  | 对两个特征值进行逻辑与非运算。 |
| 逻辑或非运算算子 |  | 对两个特征值进行逻辑或非运算。 |
| 逻辑等价运算算子 |  | 对两个特征值进行逻辑等价运算。 |

表 3 三元异常指标算子表

|  |  |  |
| --- | --- | --- |
| 算子名称 | 符号 | 定义 |
| 三元逻辑与运算算子 |  | 对两个特征值进行逻辑与运算。 |
| 三元逻辑或运算算子 |  | 对两个特征值进行逻辑或运算。 |
| 三元逻辑异或运算算子 |  | 对两个特征值进行逻辑异或运算。 |
| 三元逻辑蕴含运算算子 |  | 对两个特征值进行逻辑蕴含运算。 |

#### 3.1.2 属性算子

（1）对比算子

定义：比较两个特征值的大小关系，常用于阈值判断。

数学表达：

应用场景：

判断货物重量是否超出港口限重，提前预警超载风险。

检查危险化学品的申报数量是否超过区域配额。

示例：

青岛港的限重为 50 吨，货物申报重量为 55 吨，判定为超标。

某批货物的危险化学品申报数量为 120 吨，区域配额为 100 吨，触发超量预警。

（2）差值算子

定义：计算两个特征值的差值，常用于时间、空间或数值偏差的分析。

数学表达：

应用场景：

分析申报时间与实际到货时间的差异，用于检查是否存在延迟或伪报。

计算当前时间与证件有效期的差值，判断证件是否过期。

示例：

申报时间为 2025-01-01，实际到货时间为 2025-01-03，差值为 2 天，标记为轻度延迟。

当前时间为 2025-01-07，证件有效期截至 2025-01-05，差值为 -2 天，判定为过期。

（3）乘法算子

定义：计算两个特征值的乘积，用于特定的加权或综合计算场景。

数学表达：

应用场景：

计算货物运输成本（单价 × 重量）用于异常成本分析。

分析货物风险等级与频率的乘积，用于量化高风险货物的重要性。

示例：

运输单价为 500 元/吨，货物重量为 10 吨，总成本为 5000 元。

某批货物的风险等级为 4，申报频率为 5 次，风险乘积为 20，触发高优先级检查。

（4）比值算子

定义：计算两个特征值的比值，用于衡量相对变化幅度或占比情况。

数学表达：

应用场景：

计算货物申报金额与实际价值的比值，判断是否存在低报。

分析高风险货物占总货物数量的比率，用于判断查验资源分配。

示例：

申报金额为 8 万元，实际价值为 10 万元，比值为 80%，触发低报预警。

港口货物总量为 1000 吨，其中高风险货物为 300 吨，高风险占比为 30%。

（5）差分比率算子

定义：计算差分值相对于基准值的比率。

数学表达：

应用场景：

计算当前货物运输延误时间相对于总运输时限的比率。

分析申报金额差值占总金额的比例，用于识别高风险货物。

示例：

当前延误时间为 5 小时，总时限为 50 小时，延误比率为 10%。

某批货物的申报金额差值为 1 万元，总金额为 10 万元，差分比率为 10%。

（6）子集判断算子

定义：判断一个集合是否为另一个集合的子集，用于特征范围的包含性分析。

数学表达：

应用场景：

判断港口货物清单是否包含在全国高风险清单中。

检查货物的类别是否属于允许清单。

示例：

青岛港货物清单与全国高风险清单对比，结果为是。

危险化学品类别属于允许清单，结果为是。

（7）均值算子

定义：计算两个特征值的平均值，用于评估平滑特征变化。

数学表达：

应用场景：

分析不同时间段的货物申报量的均值。

计算运输成本与申报金额的平均值。

示例：

港口日均申报量50吨，高峰期均值80吨。

运输成本与申报金额平均值为5万元。

（8）方差算子

定义：计算两个特征值的方差，用于评估其波动性或稳定性。

数学表达：

应用场景：

分析港口货物流量的波动性。

评估时间节点上货物申报的稳定性。

示例：

港口货物流量日波动方差为300吨。

申报时间与到货时间的波动方差为2小时。

（9）欧几里得距离算子

定义：计算两点之间的欧几里得距离，用于空间或数值特征的差异分析。

数学表达：

应用场景：

计算两个港口之间的运输路径距离。

分析货物申报金额与实际价值的差异。

示例：

港口A到港口B的实际距离为20公里。

申报金额8万元，实际价值10万元，差异距离为2万元。

（10）加权组合算子

定义：对三个特征值赋予不同权重，计算加权综合值，用于综合风险评估。

数学表达：

应用场景：

结合运输时效、货物风险等级和路径复杂性生成动态风险评分。

根据权重调整提高重点监管货物的优先级。

示例：

权重：运输时效 0.3，风险等级 0.4，路径复杂性 0.3；加权风险评分为 75 分。

（11）交叉偏差算子

定义：计算三个特征值的交叉偏差，用于多维比较分析。

数学表达：

应用场景：

比较申报金额、实际交易金额与货物市场价值的偏差。

分析运输路径、停留时间与计划时长之间的偏差。

示例：

偏差5%，判定为低报行为。

偏移2公里，标记异常。

（12）多变量方差算子

定义：计算三个特征值的方差，评估特征间的波动性。

数学表达：

应用场景：

比较不同港口或区域之间的货物流量变化，用于识别波动大的港口并优化资源配置。

评估货物属性、时间分布和事件频率的变化幅度，用于监控潜在风险货物的特征一致性。

示例：

港口1：500吨，港口2：600吨，港口3：550吨，计算得到方差为833吨。偏移2公里，标记异常。

某批货物的属性波动方差为0.2，时间分布波动方差为0.4，事件频率波动方差为0.1，结合分析得出特征稳定性较差。

（13）三维欧几里得距离算子

定义：计算三维空间中三个点之间的欧几里得距离，用于空间分析。

数学表达：

应用场景：

计算三港之间的运输距离，用于路径优化。

分析货物实际路径偏离程度。

示例：

实际距离50公里，推荐距离40公里，偏差10公里。

（14）联合概率算子

定义：计算三个事件或特征的联合发生概率，基于逻辑运算和比例计算。

数学表达：

应用场景：

分析货物类别、运输路径和报关时段的联合概率，用于多维高风险识别。

评估高危化学品在不同时间和港口的分布联合概率。

示例：

某货物类别为危险化学品，运输路径经过高风险区域，申报时间为夜间，联合概率为 85%，触发高风险警告。

（15）多维相似性算子

定义：计算三维空间的余弦相似度，用于多维特征的模式匹配。

数学表达：

应用场景：

将待分析货物特征与高风险货物的标准模板进行对比，计算相似度。用于快速识别潜在高风险货物。

对不同港口的货物流通特征进行相似性分析，优化货物分类和监管策略。

对比货物的运输路径与历史高风险路径的相似性，用于路径监控和优化。

示例：

某货物特征向量：[0.8, 0.7, 0.6]，高风险模板：[0.9, 0.8, 0.7]，余弦相似性为0.99，判断为高相似度。

当前运输路径特征向量：[0.3, 0.4, 0.5]，历史高风险路径：[0.4, 0.5, 0.6]，相似度为0.95，触发路径风险警报。

#### 3.1.3 异常指标算子

二元算子

（1）逻辑与运算算子

定义：对两个特征值进行逻辑关系计算。

数学表达：

应用场景：

检查货物是否为高风险类别且申报时间在高峰期。

判断货物是否同时符合合法证件状态和合规申报金额。

示例：

货物为危险化学品，申报时间为夜间高峰期，结果为满足。

证件有效，申报金额符合标准，逻辑结果为满足。

（2）逻辑或运算算子

定义：对两个特征值进行逻辑关系计算。

数学表达：

应用场景：

检查货物是否为高风险类别或运输路线经过高风险区域。

判断是否符合时间条件（延迟或高峰期）之一。

示例：

货物类别为普通货物，但运输路径经过高风险区域，逻辑结果为满足。

货物运输延迟但不在高峰期，逻辑结果为满足。

（3）逻辑异或运算算子

定义：对两个特征值进行逻辑关系计算。

数学表达：

应用场景：

判断货物是否为高风险类别但未通过高风险路径。

检查证件状态是否有效但申报金额不合规。

示例：

货物为高风险，但路径未经过高风险区域，逻辑结果为满足。

证件已过期，但申报金额符合标准，逻辑结果为满足。

（4）逻辑蕴含运算算子

定义：对两个特征值进行逻辑蕴含运算，当第一个特征值为真且第二个特征值为假时，结果为假；其他情况为真。

数学表达：

应用场景：

判断货物的运输路径是否安全（路径风险为假，货物风险为真，则不满足）。

检查证件状态是否合法（合法性为真，使用状态为假，则不满足）。

示例：

货物运输路径风险为高（假），货物风险等级为高（真），逻辑结果为满足。

证件为有效状态，但未实际使用，逻辑结果为不满足。

（5）逻辑与非运算算子

定义：对两个特征值进行逻辑与非运算，当两个特征值都为真时，结果为假；其他情况为真。

数学表达：

应用场景：

检查货物是否既不超重又不超过限量配额。

判断时间和路径是否同时符合高风险条件（两者同时满足为假）。

示例：

货物重量未超标，限量配额未超过，逻辑结果为满足。

时间符合高峰期，路径经过高风险区域，逻辑结果为不满足。

（6）逻辑或非运算算子

定义：对两个特征值进行逻辑或非运算，当两个特征值都为假时，结果为真；其他情况为假。

数学表达：

应用场景：

判断货物是否既不属于高风险类别，也未经过高风险路径。

检查是否既没有超时，也没有延迟申报。

示例：

货物类别为普通货物，路径未经过高风险区域，逻辑结果为满足。

时间未超时，申报未延迟，逻辑结果为满足。

（7）逻辑等价运算算子

定义：对两个特征值进行逻辑等价运算，当两个特征值相等时，结果为真；否则为假。

数学表达：

应用场景：

判断货物是否具有一致性特征（如申报重量是否与实际重量一致）。

检查时间是否一致（如运输时间是否匹配计划时间）。

示例：

申报重量为50吨，实际重量也为50吨，逻辑结果为满足。

计划到达时间为2025-01-10，实际到达时间为2025-01-10，逻辑结果为满足。

（8）三元逻辑与运算算子

定义：对三个特征值进行逻辑与运算，当且仅当三个特征值都为真时，结果为真；否则为假。

数学表达：

应用场景：

检查货物是否为高风险类别、运输路径经过高风险区域、且申报时间处于高峰期。

示例：

货物类别为危险化学品，路径为高风险区域，申报时间为夜间高峰期，逻辑结果为满足。

（9）三元逻辑或运算算子

定义：对三个特征值进行逻辑或运算，当至少一个特征值为真时，结果为真；当三个都为假时，结果为假。

数学表达：

应用场景：

判断货物是否属于高风险类别、运输路径是否经过高风险区域、或者申报时间是否处于高峰期。

示例：

货物类别为普通货物，但路径为高风险区域，逻辑结果为满足。

（10）三元逻辑异或运算算子

定义：对三个特征值进行逻辑异或运算，当三个特征值中奇数个为真时，结果为真；否则为假。

数学表达：

应用场景：

检查货物是否在某个维度存在特定的唯一性条件（如仅一个维度异常）。

示例：

货物类别为高风险，路径未经过高风险区域，申报时间处于高峰期，逻辑结果为满足。

（11）三元逻辑蕴含运算算子

定义：对三个特征值进行逻辑蕴含运算，当任意两个前提为真且结论为假时，结果为假；其他情况为真。

数学表达：

应用场景：

检查货物类别、路径和时间是否一致符合高风险逻辑关系。

示例：

货物为危险化学品，路径为高风险区域，时间不在高峰期，逻辑结果为不满足。

### 3.2 多维风险特征命名规则与描述算法

通过清晰、规范的命名方式，对多维特征进行统一描述，便于数据管理、分析与共享。具体包括以下几方面。

#### 3.2.1 风险指标统一标识

根据输入数据的类型，风险指标命名分为三类：属性算子、输入数据为异常指标的二元算子和输入数据为异常指标的三元算子。

（1）属性算子

定义：这类算子处理属性数据，通常涉及数值比较或计算，结果根据阈值判断是否异常。命名格式为 算子名称(特征1, 特征2, 阈值)。

示例：

差值算子：

格式：DIFF(特征1, 特征2, Threshold)

示例：

特征 1：申报重量（物 - 进口大宗散货 - 数量和重量不符）

特征 2：实际重量（物 - 进口大宗散货 - 数量和重量不符）

阈值：5%（重量偏差率阈值）

应用场景：对比进口大宗散货的申报重量与实际重量，若偏差率≥5%，触发 “重量申报异常” 风险。

（2）输入数据为异常指标的二元算子

定义：这类算子处理两个异常指标之间的逻辑关系，通常用于判断两个异常指标是否同时满足某种条件。命名格式为 算子名称(特征1, 特征2)。

示例：

逻辑或运算算子：

格式：OR(特征1, 特征2)

示例：

特征 1：旅客行李数量不符（物 - 行李 - 数量不符）

特征 2：旅客频繁出入境（人 - 旅客 - 频繁出入境记录）

应用场景：若旅客存在 “行李数量不符” 或 “频繁出入境记录” 任一情况，触发 “高风险旅客预警”。

（3）输入数据为异常指标的三元算子

定义：这类算子处理三个异常指标之间的逻辑关系，通常用于判断三个异常指标是否同时满足某种条件。命名格式为 算子名称(特征1, 特征2, 特征3)。

示例：

三元逻辑与运算算子：

格式：AND3(特征1, 特征2, 特征3)

示例：

特征 1：危化品运输路线异常（物 - 危化品 - 运输路线异常）

特征 2：运输工具历史不良记录（物 - 运输工具 - 运输工具的历史记录异常）

特征 3：承运人资质不全（事 - 生产企业备案 - 资质证书不齐全或无效）

应用场景：当 “危化品运输路线异常”“运输工具历史不良记录”“承运人资质不全” 同时存在时，触发 “高危运输复合风险预警”。

#### 3.2.2 风险指标描述方法

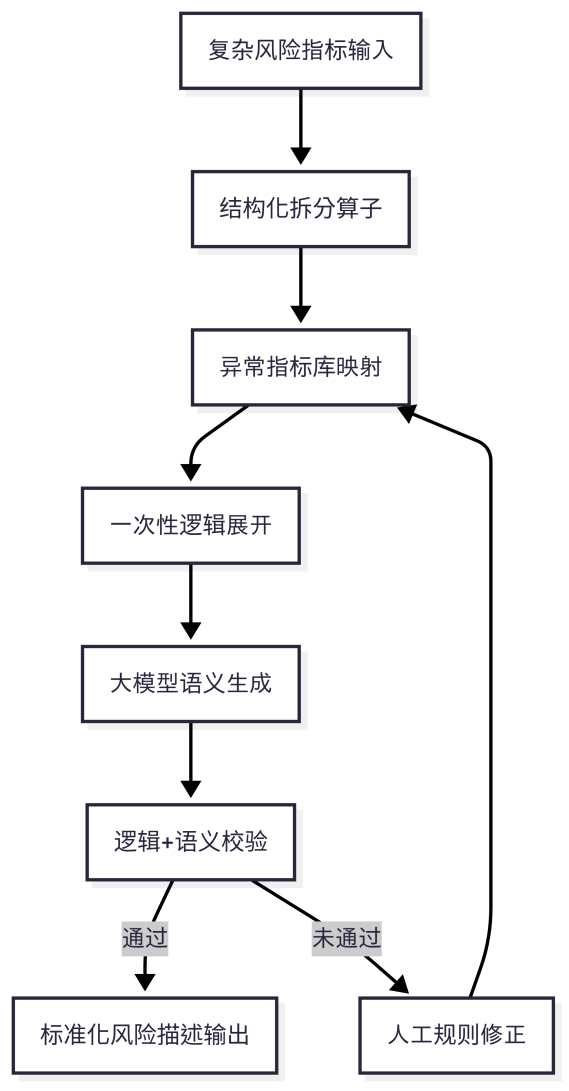


图 1 算法流程图

针对复杂多维风险指标的语义化描述需求，本研究提出 “算子展开 - 业务匹配 - 大模型总结” 的描述方法，通过将复杂指标拆解为基础算子组合，匹配海关业务场景含义，最终由大模型生成自然语言描述，确保指标可读性与监管实用性。

传统递归展开方法在处理多层级复合算子时效率较低，且易导致业务语义碎片化。本方法采用“一次性展开” 策略，直接解析复杂指标的所有子算子及其关联特征，通过业务场景映射明确算子含义，再借助大模型强大的语义整合能力，生成完整、易懂的风险指标描述，具体流程如下：

1. 复杂指标结构化拆分

输入包含多算子组合的风险指标，通过递归展开技术识别指标类型（属性算子 / 二元算子 / 三元算子），并提取所有子算子及输入特征。

1. 业务场景关联与算子含义匹配

将拆分后的子算子及其特征，与异常指标库进行映射，明确每个算子在海关监管中的实际意义。

1. 一次性逻辑展开与语义框架构建

不采用递归分层解析，而是将所有子算子的 输入特征、阈值条件、风险关联 一次性罗列，形成 “特征 - 算子 - 规则 - 后果” 的基础语义框架。

1. 大模型语义总结与自然语言生成

调用微调后的 DeepSeek 海关监管大模型，输入上述语义框架及业务场景上下文（如监管对象、风险维度、历史案例），由模型自动整合多算子逻辑，生成符合海关业务语言习惯的风险描述。

1. 逻辑与语义双重校验

逻辑校验：检查算子条件是否冲突、阈值是否符合海关业务标准。

语义校验：对比术语库，确保 “物 - 事 - 证” 维度用词一致。

校验通过：输出标准化描述；未通过：触发人工干预，补充业务规则或修正算子映射关系。

**应用示例**如下：

以《海关风险指标系统报告》中 “海运口岸出口危化品伪瞒报风险” 为例：

复杂指标：AND(危化品货物描述不一致, 运输路线异常, 许可证过期)

一次性展开：

货物描述不一致（物 - 危化品 - 货物描述不一致）：申报名称与实际危化品不符；

运输路线异常（物 - 危化品 - 运输路线异常）：绕路高风险区域或多次转运；

许可证过期（证 - 出口许可证 - 许可证过期或无效）：危化品运输许可证失效。

大模型总结描述：  
“当出口危化品出现申报名称与实际货物不符、运输路线偏离常规路径且运输许可证过期时，判定为‘伪瞒报高危风险’。此类风险可能导致违禁危化品非法运输，威胁运输安全及目的地国家环境安全，需立即触发人工查验与许可证溯源。”

## **应用案例**

### 4.1 应用场景说明

海关在多种监管场景中需要对报关单及相关信息进行风险评估，以识别潜在的风险因素并生成多维风险指标，为后续查验提供决策支持。这些场景涵盖了海运口岸出口危化品监管、空运口岸入境旅客及行李监管、海运口岸进口大宗散货监管、进口快件监管以及陆运口岸中欧班列监管。在海运口岸出口危化品监管中，海关需识别危化品的伪瞒报、超重、低报、证件过期以及运输路径异常等风险；在空运口岸入境旅客及行李监管中，重点关注高风险旅客及其行李中的违禁品、低报价值等风险；海运口岸进口大宗散货监管则聚焦于重量不符、过度包装、辐射超标等问题；进口快件监管需防范跨境电商中的人员、企业及途经地风险，特别是集货风险；而陆运口岸中欧班列监管则需识别公路运输过境集装箱换货风险，实现跨关区协同防控。这些场景涉及多种风险类型，需要综合考虑货物、事件、证件、时空等多个维度的信息，以实现精准的风险预警。

### 4.2 应用场景案例

#### 4.2.1 输入报关单信息

报关单编号：预录入编号与海关编号均为 531620221160109770，申报日期为 2022 年 11 月 07 日，申报单位为深圳市坤祥物流供应链有限公司（税号：91440300MA5FPMYXXH）。

企业信息：境内收货人 / 消费使用单位：深圳市美星建筑装饰材料有限公司（税号：914403005571519063），地址与企业名称一致。

境外发货人：RUST-OLEUM CORPORATION（地址未详细标注）。

合同与运输：合同协议号 20220831-01，提运单号 ONEYRICCJKE685900，运输工具为水路运输（航次号：UN9337482/22W），启运港与经停港均为美国洛杉矶（USA264），入境口岸为盐田（代码：470601）。

商品详情（共 6 项，HS 编码均为 3209100090，属 “以合成聚合物为基本成分的油漆”）：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 编号 | 商品名称及规格 | 数量 / 单位 | 单价 / 总价（美元） | 原产国 | 最终目的国 |
| 1 | 肌理漆 | 2250 千克 |  | 美国 | 中国 |
| 2 | 久白防霉乳胶漆 - 蛋壳光 | 6720 千克 |  | 美国 | 中国 |
| 3 | 防霉乳胶漆 - 蛋壳光 | 165 千克 |  | 美国 | 中国 |
| 4 | 室内乳胶漆 - 白色蛋壳光（样品赠送） | 11 千克 |  | 美国 | 中国 |
| 5 | 室内乳胶漆 - 深基科蛋壳光（样品赠送） | 11 千克 |  | 美国 | 中国 |
| 6 | 室内乳胶漆 - 浅基科壳蛋光（样品赠送） | 11 千克 |  | 美国 | 中国 |

重量与价格：毛重 9551.8 千克，净重 9173 千克，成交方式为 C&F（代码 2），运费率 1%（运费 0.0/3/1），保费及杂费均为 0，特殊关系确认、价格影响确认等均为 “否”，自报自缴为 “是”。

监管信息：监管方式为一般贸易（0110），征免性质为一般征税（101），许可证号显示 “无”。

随附单证：包括反倾销措施排除代码（PC011201210140010）、代理报关委托协议（电子）、发票、装箱单、提 / 运单、合同及企业声明；备注标注 “有 IPPC 标识”（国际植物保护公约标识），集装箱号为 NYKU3820960。

报关人员：报关人员证号 53154949，联系电话未填写，申报单位与境内收货人不一致（前者为物流企业，后者为货主）。

商品备注：零售包装，第 1-6 项含聚合物成分，第 2-6 项成分详见附页；流向龙岗海关查验科，集装箱标箱数为 1。

#### 4.2.2 复杂算子组合与风险指标构建

AND(RATIO(样品单价, 正常单价, 0.5), AND(许可证缺失, DIFF(毛重, 净重, 5%)))

其中包含属性算子和异常指标算子。RATIO(样品单价, 正常单价, 0.5)：计算样品单价与正常商品单价的比值（属于 “物 - 报关单 - 价格申报明显偏低或偏高”）。DIFF(毛重, 净重, 5%)：计算毛重与净重的偏差率（属于 “物 - 进口大宗散货 - 数量和重量不符”）。AND(许可证缺失, DIFF...)：连接许可证缺失与重量偏差风险。外层AND用来进一步连接价格比值异常与上述二元风险。

#### 4.2.3 风险指标描述生成流程

**步骤 1：结构化拆分算子**

**子算子提取**：

RATIO（属性算子）：特征 1 = 样品单价（11 美元），特征 2 = 正常单价（取第 3 项 165 美元），阈值 = 0.5（即单价＜正常价 50% 触发）。

DIFF（属性算子）：特征 1 = 毛重（9551.8kg），特征 2 = 净重（9173kg），阈值 = 5%（偏差率阈值）。

两层AND（二元算子）：第一层连接 “许可证缺失” 与 “重量偏差”，第二层连接 “价格比值异常” 与第一层复合风险。

**步骤 2：业务场景匹配**

|  |  |  |
| --- | --- | --- |
| 算子 / 特征 | 关联异常指标 | 业务含义 |
| RATIO | 物 - 报关单 - 价格申报明显偏低或偏高 | 样品单价显著低于正常商品，可能存在低报或伪报 “样品” 逃避关税 |
| DIFF | 物 - 进口大宗散货 - 数量和重量不符 | 毛重与净重偏差率 4%（未超 5% 阈值，单独不触发，但需结合其他风险） |
| 许可证缺失 | 证 - 进口许可证 - 许可证过期或无效 | 未提供进口许可证，可能涉及禁止 / 限制进口货物 |

**步骤 3：一次性逻辑展开**

**价格风险：第 4-6 项 “样品赠送” 单价 11 美元，仅为第 3 项正常单价（165 美元）的 6.67%，低于 50% 阈值，触发 “低报价格风险”。**

**重量风险：毛重与净重偏差率 4%（＜5% 阈值，单独不触发，但作为复合风险因子）。**

**证件风险：许可证号 “无”，且商品 HS 编码 3209100090（涉化工品）可能需许可证，存在 “许可证缺失风险”。**

**步骤 4：大模型语义总结**

**输入提示词：  
 “当前指标包含比值算子（RATIO，样品单价 / 正常单价＜50%）、差值算子（DIFF，毛重 / 净重偏差 4%）及逻辑与算子（AND，许可证缺失 + 重量偏差），对应《海关风险指标系统报告》中‘价格低报’‘重量不符’‘许可证缺失’。请结合报关单中‘样品赠送’标注，总结该指标的风险含义。”**

**生成描述：**

**“通过比值算子（RATIO）分析，报关单中‘样品赠送’商品单价（11 美元）仅为正常同类商品（165 美元）的 6.67%，低于 50% 阈值，触发‘样品价格异常风险’；同时，差值算子（DIFF）显示毛重与净重偏差 4%（接近 5% 预警线），且未提供进口许可证（许可证缺失）。该复合指标表明，企业可能利用‘样品’名义低报价格（实际单价显著偏离市场合理区间），虽重量偏差未达单独预警标准，但结合许可证缺失，存在以‘样品’伪报逃避许可证监管及关税的嫌疑。需重点核查样品真实性、许可证必要性及价格构成，防范借样品渠道走私或违规进口化工品。”**

**步骤 5：校验与输出**

**逻辑校验：价格比值计算准确，许可证缺失与商品 HS 编码关联性（3209100090 涉涂料，可能需危化品许可证）；**

**语义校验：术语 “低报价格”“许可证缺失” 与海关风险指标系统报告一致。**

## **总结**

本研究提出的基于“物-事-证-场-人”多个维度的多维特征提取框架，全面整合了货物基础属性、关键事件、证件信息、时间规律、空间分布和人员行为记录，实现了多维特征的系统化提取与描述。通过设计标准化的算子（如加权组合算子、多维相似性算子等）和规范的命名规则与描述算法，显著简化了风险指标的构建过程，提升了多维数据的处理能力和精准性，为风险评估和决策提供支持，提升海关监管的智能化水平和决策效率。

同时，本方法有效解决了传统海关通关监管中存在的挑战。首先，本研究提出的基于"物-事-证-场-人"多维风险特征提取算子，有效解决了单指标误报率高易被规避的问题。通过建立多维度特征关联分析机制，将原本孤立的单指标检测转变为综合风险评估，既避免了单一阈值判断的局限性，又通过特征间的协同验证显著提升了风险识别的可靠性，在保证查获率的同时大幅降低了误报率。此外，通过标准化风险指标命名和描述算法，实现了风险指标的快速构建与跨场景复用，大幅提升了指标体系的适应性和可维护性。具体应用案例展示了在海运口岸出口危化品监管中，如何设计、计算多维风险指标并进行应用，为海关通关监管提供了高效、精准的风险评估与预警工具。

后续将进一步完善多维特征提取算子与描述算法体系，将算法集成到海关监管系统中，推广至更多监管场景；建立持续改进和反馈机制，结合现行法律法规，为新技术应用提供依据和保障，进一步提升海关监管的智能化水平和决策效率。