

基于 TEM 模型的空管安全信息相关性分析

王洁宁¹, 刘锦睿¹, 左燕², 郭庆²

(1. 中国民航大学空中交通管理学院, 天津 300300; 2. 中国民用航空华东地区管理局, 上海 200335)

摘要: 针对空管系统危险源控制清单中关键信息识别与潜在隐患管理的问题, 本文提出了一种基于威胁与差错管理(TEM, threat and error management)模型的空管安全信息相关性分析方法。首先, 依据 TEM 模型的内在分类构建 TEM 模型本体; 其次, 利用信息抽取技术从危险源控制清单文本中抽取结构化事件, 利用深度自编码器网络模型和改进的余弦相似度公式进行文本语义相似度匹配; 最后, 基于贝叶斯网络对 TEM 模型本体类别的概率关系进行定量分析, 得到威胁、差错和非期望的航空器状态之间的相关性。通过对民航空管系统危险源控制清单中数据的实例分析, 识别并抽取出了文本中的关键信息、与空管运行安全强相关的实体词和事件类型, 基于抽取结果对空管各运行环节安全性的提升提出合理建议。

关键词: 空管安全管理; 威胁与差错管理模型; 模型本体; 信息抽取; 相关性分析

中图分类号: V355.1; X949 **文献标志码:** A **文章编号:** 1674-5590(2026)02-0032-09

Correlation analysis of air traffic control safety information based on TEM model

WANG Jiening¹, LIU Jinrui¹, ZUO Yan², GUO Qing²

(1. College of Air Traffic Management, CAUC, Tianjin 300300, China; 2. CAAC East China Regional Administration, Shanghai 200335, China)

Abstract: To address the problems of identifying key information and managing potential risks in the hazard control list of the air traffic control (ATC) system, this paper proposes a correlation analysis method for ATC safety information based on the Threat and Error Management (TEM) model. Firstly, the TEM model ontology is constructed according to the internal classification of the TEM model. Then, information extraction techniques are employed to extract structured events from the text of the hazard control list. The deep autoencoder network model and an improved cosine similarity formula are used to match the text semantic similarity. Finally, a Bayesian network is applied to quantitatively analyze the probabilistic relationships among the TEM ontology categories, obtaining the correlations between threats, errors, and undesired aircraft states. Through a case study using data from a civil aviation ATC hazard control list, key information, entity words and event types strongly related to ATC operational safety are identified and extracted. Based on the extraction results, reasonable suggestions are proposed for enhancing the safety of various ATC operational phases.

Key words: air traffic control (ATC) safety management; threat and error management (TEM) model; ontology; information extraction; correlation analysis

为准确、可靠地识别和管理各民航管制运行单位面临的危险源和隐患, 中国民航空管系统出台了一系列的咨询通告, 以规范危险源的管理^[1]。同时, 国际上涌现出如确定危险范围的 5M (man, machine, material, method, environment) 模型、危险与可操作性分析 (HAZOP, hazard and operability study) 模型、系统理论过程分析 (STPA, systems theoretic process analysis) 模型、

威胁与差错管理 (TEM, threat and error management) 模型等^[2-5]定性和定量地辅助危险源识别。事实上, 危险识别和风险分析的结果全部采用文字形式描述, 因此, 上述内容虽然通过 Excel 或数据库的方式实现了表单式的存储, 但其内容完全是通过中文自然语言的方式存在的。由于危险源识别的各运行单位主体存在理解和语言风格的差异, 危险源控制清单内容中的

文字描述天然存在着较大差异,这使得各单位、各专业共享危险源时,很难自然发现其内部的潜在联系,或挖掘其中的内在价值,这是摆在面前的现实性问题。

信息抽取是自然语言处理(NLP, natural language processing)技术的重要组成部分,主要功能是从文本中抽取特定的事实信息,这些文本可以是结构化、半结构化或非结构化的数据^[6-8]。Sizov 等^[9]基于案例推理的方法,利用案例推理(CBR, case-based reasoning)系统对加拿大运输安全委员会事故报告中的事件分析进行了评估,证明了系统利用 NLP 技术可以从文本事件报告中自动抽取对过去事件的解释信息,避免了人工构造。国际民航组织(ICAO, International Civil Aviation Organization)^[10]提出了《空中交通管制的威胁与差错管理》咨询通告,称为 TEM。其在分析机组人员资源管理(CRM, crew resource management)过程中,以人为因素的研究为基础,建立了 TEM 框架,兼具可分析、可使用的特点,兼顾人与环境,已广泛应用于 CRM 和空管班组资源管理(TRM, team resource management)的审计、监督和安全培训。李芳国等^[11]从危险事件角度设计构建了 HAZOP 模型本体,并在双向长短期记忆网络-条件随机场(BiLSTM-CRF, bidirectional long short-term memory-conditional random field)命名实体识别模型的基础上,实现了关键 HAZOP 模型信息的自动提取和知识图谱构建。Zhao 等^[12]基于 TEM 模型对 2011 年空中交通管制不规则事件建立了贝叶斯网络模型,对空中交通管制运行中的威胁、差错和非期望的航空器状态之间的相关性进行了精确的定量分析。

以上研究虽已使用 NLP 技术进行信息抽取,但多使用自然语言分词,缺少空管领域专业分词的词库;并且在危险源识别模型的应用方面不够具体。因此,本文构建本体模型来针对性地解决上述问题。

综上,首先,基于 TEM 模型,引入本体概念,结合 NLP 中的信息抽取方法及知识图谱,构建空管领域自定义词库;然后,基于贝叶斯网络计算不同实体之间的概率关系,对 TEM 模型本体类别的概率关系进行定量分析,得到威胁、差错和非期望的航空器状态之间的相关性;最后,使用空管系统危险源信息数据作为实例进行分析,从相关性结果角度给出有助于提升空管运行安全的合理建议。

1 TEM 模型本体构建流程

结合 NLP 中的信息抽取方法及知识图谱,对 TEM 模型进行自上而下和自下而上相结合的本体构建,并

使用空管系统危险源信息数据作为实例进行分析,研究 TEM 模型各分类之间的相关性,验证 TEM 模型本体构建方法的可行性。本体构建流程如图 1 所示,具体步骤如下。

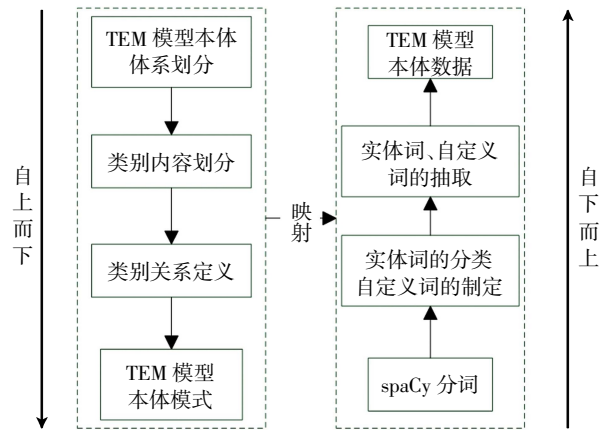


图 1 TEM 模型本体构建流程

Fig.1 Ontology construction process of the TEM model

1) 自上而下构建 TEM 模型本体模式层

首先给出 TEM 模型本体体系划分的标准,构建 TEM 模型知识图谱的大致框架;然后将 TEM 模型中不同类别的内容与结构详细划分,对类别之间的关系进行定义;最终建立清晰、准确的 TEM 模型框架^[11]。

2) 自下而上构建 TEM 模型本体数据层

首先通过 NLP 技术,对非结构化的文本数据进行分词操作;然后根据已有的危险源信息文本数据,结合空管安全专业知识,对危险源信息文本实体进行分类,并设计命名实体标注规则;随后制定事件抽取规则,以自定义词典为依据抽取出其他词、人因词、后果词、措施词等的词性和句法依存关系;最后以深度自编码器(DAE, deep autoencoder)网络模型为基础,采用改进后的余弦相似度计算方法对所得实体、短文本描述与 TEM 模型框架中的事件进行文本相似度映射,结合 TEM 模型框架类别间的关系,构建 TEM 模型本体数据层。

3) 实例分析

根据现有的危险源信息数据,按照已经搭建好的 TEM 模型框架类别间的关系,基于贝叶斯网络计算不同实体之间的概率关系,对 TEM 模型本体类别的概率关系进行定量分析,得到威胁、差错和非期望的航空器状态之间的相关性。

2 模式层构建

2.1 TEM 模型关键要素与属性定义

TEM 模型有助于理解在动态和具有挑战性的运

行环境中安全和人员行为之间的相互关系。TEM 模型的主要目标是识别飞机运行时可能导致安全裕度降低的动态工作环境和人为因素,并适当提出建议。因此,“可能导致飞行安全裕度降低的事件”是构成 TEM 模型信息的基本结构。从空中交通管制员的角度来看,一般认为,TEM 模型框架由 3 个基本部分组成,分别是威胁、差错和非期望的航空器状态。ICAO 的 Doc9868 文件从飞行机组的角度,增加了确保飞机运行的“对策”。所以根据 TEM 模型的特性,结合 Doc9868 文件中对 TEM 模型描述,对可能导致飞行安全裕度降低的事件、威胁、差错、非期望的航空器状态和确保飞机运行的对策进行定义。

可能导致飞行安全裕度降低的事件:TEM 模型框架中信息的基本结构,由威胁、差错、非期望的航空器状态和确保飞机运行的对策构成。

威胁:发生在空中交通管制员影响范围之外的事件或差错,会增加飞机运行的复杂性,并且为保持安全裕度必须对其进行管理。

差错:空中交通管制员无意中导致偏离组织或飞机运行预期的作为或不作为。

非期望的航空器状态:航空器运行过程中出现的航空器位置、速度、姿态或配置等问题,是偏离运行期间通常经历的、与安全裕度下降有关的参数。

确保飞机运行的对策:保证威胁、差错、非期望的航空器状态不会影响航空器运行所需安全裕度的必要手段,是管制员日常职责的一部分,存在于威胁、差错与非期望的航空器状态的管理之中。

2.2 TEM 模型本体规则设计

设计 TEM 模型本体的规则是构建 TEM 模型本体的核心工作。构建 TEM 模型本体的步骤主要有定义类及类的层次结构、类的属性等。类是本体构建中表示共性概念的集合,是直接影响本体构建质量的重要因素,因此,完整的 TEM 模型本体中应该包含以下类:可能导致飞行安全裕度降低的事件及其包含的威胁、差错、非期望的航空器状态、确保飞机运行的对策。TEM 模型框架如图 2 所示。

2.2.1 事件概念划分与定义

TEM 模型事件概念的划分是对可能导致飞行安全裕度降低的事件进行类型划分。根据 TEM 模型内部基本事件的元素构成及其关系,将可能导致飞行安全裕度降低的事件划分为以下不同类型。

(1)因果型事件。这类事件的构成有 3 种可能,分

别是:威胁类事件导致非期望的航空器状态;差错类事件导致非期望的航空器状态;威胁类事件导致差错类事件,差错类事件接着导致非期望的航空器状态,威胁类事件与差错类事件均只有一件发生。

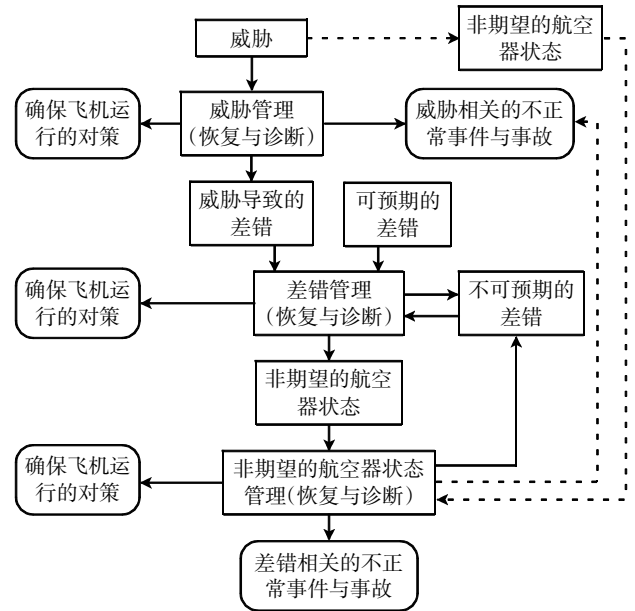


图 2 TEM 模型框架

Fig.2 TEM model framework

(2)并列+因果型事件。这类事件的构成为:威胁类事件与差错类事件各有一件发生但不存在因果关系,两类事件共同导致了非期望的航空器状态。

(3)时序+因果型事件。这类事件的构成为:威胁类或差错类事件有多个发生,发生时间存在先后顺序,一系列事件共同导致了非期望的航空器状态。

(4)时序+并列+因果型事件。这类事件的构成为:威胁类事件与差错类事件有多个发生且存在时序关系,两类事件共同导致了非期望的航空器状态。

2.2.2 类的属性定义

根据 2022 年 TEM 模型对某空管单位某一季度系统危险源控制清单中事件的分析,对可能导致飞行安全裕度降低的事件(M)中的威胁(T)、差错(E)、非期望的航空器状态(U)和确保飞机运行的对策(S)4 类事件的数量进行统计,结果如表 1 所示。对 4 类事件进行更细化地分类和编号:威胁(T)类型包括内部威胁(T_I)、外部威胁(T_X)、空中威胁(T_A)和环境威胁(T_E);差错(E)类型包括设备操作差错(E_E)、程序差错(E_P)和通信差错(E_C);非期望的航空器状态(U)类型包括地面非期望的航空器状态(U_G)和空中非期望的航空器状态(U_A)。根据上述分类,构建 TEM 模型本体模式层概念体系,如图 3 所示。

表 1 可能导致飞行安全裕度降低的事件(M)分类与统计

Tab.1 Classification and statistics of events (M) that may reduce flight safety margins

| 事件类型及编号 | 事件细分类型及编号 | 类型示例 |
|--------------|--|---|
| 威胁(T) | 内部威胁(T _i) 外部威胁(T _x) 空中威胁(T _a) 环境威胁(T _e) | 运行压力:延误、晚点、设备变化 天气雷暴、湍流、结冰、风切变、侧风/顺风、超低温/超高温 — — |
| 差错(E) | 设备操作差错(E _e) 程序差错(E _p) 通信差错(E _c) | 向自动化系统输入错误的信息 检查清单:遗漏的项目,检查清单未使用或在错误的时间使用 — |
| 非期望的航空器状态(U) | 地面非期望的航空器状态(U _c) 空中非期望的航空器状态(U _a) | 飞机在停止时继续滑行 未经批准的空域穿越 |
| 确保飞机运行的对策(S) | — | 利用所有可用资源来管理威胁 不允许工作量超负荷 |

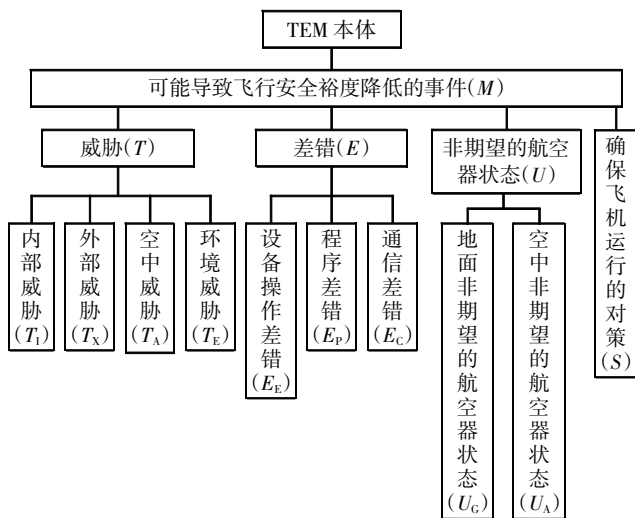


图 3 TEM 模型本体模式层概念体系

Fig.3 Concept system at the ontology pattern layer of TEM model

3 数据层构建

3.1 空管系统危险源生语料库构建

空管系统危险源原始数据中存在大量的非结构化和半结构化文本,记载了航空器运行过程中的事故征候、不安全事件等信息。本文收集了民航空管系统 2022 年危险源控制清单中 4 个季度的危险源数据,包含:8 个危险源要素,即危险源编号、危险源所在部门、专业类别、重大危险源标记及空管四大风险标记、危险源的名称、触发因素、后果、现有防控机制;4 个空管专业,即管制、技保、气象、站级;3 个核心业务部门,即塔台、进近和管制部门。其中,第 1 季度有 136 条,第 2 季度有 175 条,第 3 季度有 184 条,第 4 季度有 205 条,共 700 条数据作为生语料库构建的基础。

3.2 文本信息标注与抽取

以危险源要素为指导,对危险源语料要素进行划分,如表 2 所示。其中危险源编号、危险源所在部门、

专业类别、重大危险源标记及空管四大风险标记、危险源的名称这 5 个要素已经能够从结构化的文本语料中清晰地看到对应的信息;而触发因素、后果、现有防控机制这 3 个要素中的文本依然不能结构化,但这 3 项包含的安全管理信息与 TEM 模型中涉及的 4 类信息相吻合,所以本文将使用 TEM 模型对这 3 项数据进行分析与讨论。

表 2 危险源语料要素划分示例

Tab.2 Examples of element division of hazard source corpus

| 危险源要素 | 危险源语料要素 |
|------------------|--|
| 危险源编号 | WXY-ZSAM-GZ-QY-202109-002A |
| 危险源所在部门 | 区域 |
| 专业类别 | 管制 |
| 重大危险源标记及空管四大风险标记 | 人为因素 |
| 危险源的名称 | 席位之间协同不到位可能造成信息、动态错漏忘 |
| 触发因素 | 1.口头通报的信息在传递中出现不到位,如信息传递方在未确认收到情况下即认为已经完成信息传递 2.复杂环境下(大流量,复杂活动)扇区工作负荷大,席位协同难度大 3.席位之间人员搭配不合理 |
| 后果 | 管制指挥、协调被动,并可能导致不安全事件 |
| 现有防控机制 | 1.管制信息通报规程 2.防止航空器危险接近、防止航空器空中相撞操作细则 |

要将生语料库中的信息结构化以便进一步转化为计算机能够理解的命令,需要将原始语料数据进行一些人工处理。将这些数据中有关航路、航线、管制、塔台等敏感信息剔除,将生语料中无实际意义的介词、副词、助词等进行停用词过滤,再将不好识别的标点符号、语句标号等进行改动,初步实现对生语料的结构化。本文采用 spaCy 环境进行分词,由于 spaCy 环境下要求输入文本为 json 格式,所以将已实现半结构化的语料转化为 json 格式的分析数据文件。

此外,由于 spaCy 环境自带的中文分词库不能对空管系统危险源语料中的专有名词实现准确划分,所以要在精确分词步骤中加入自定义词库,即对民航系统中的专业词汇,如“管制室”“塔台电子进程单”等实体词做出标注和分类:将仅出现在“触发因素”一栏中涉及外部环境(如设备、程序、天气等)的

实体词均定义为“其他词”,涉及人员、单位、席位和因人工处理发生的差错等实体词均定义为“人因词”;将仅出现在“后果”一栏中的有代表性的动词定义为“后果词”;将仅出现在“现有防控机制”一栏中的有代表性的动词定义为“措施词”。基于此,形成自定义词典,如表 3 所示。

表 3 自定义词典汇总

Tab.3 Custom dictionary summary

| 自定义词典名称 | 来源语料要素 | 涉及的语料要素 | 涉及的 TEM 本体类别 | 举例 |
|---------|--------|----------------------|-----------------|-------------------|
| 其他词 | 触发因素 | 触发因素 后果 现有防控机制 | 威胁 非期望的航空器状态 | 空管自动化系统、塔台电子进程单系统 |
| 人因词 | 触发因素 | 触发因素 后果 现有防控机制 | 差错 非期望的航空器状态 | 管制员、航班计划 |
| 后果词 | 后果 | 后果 | 非期望的航空器状态 | 导致、造成 |
| 措施词 | 现有防控机制 | 现有防控机制 | 确保飞机运行的对策 | 确保、核实 |

将自定义词典与 spaCy 环境中的中文分词库相结合,实现空管系统领域的专业分词,确定事件中的基

本元素(即分词结果)、词性标注和依存句法,结果如图 4、图 5 所示。

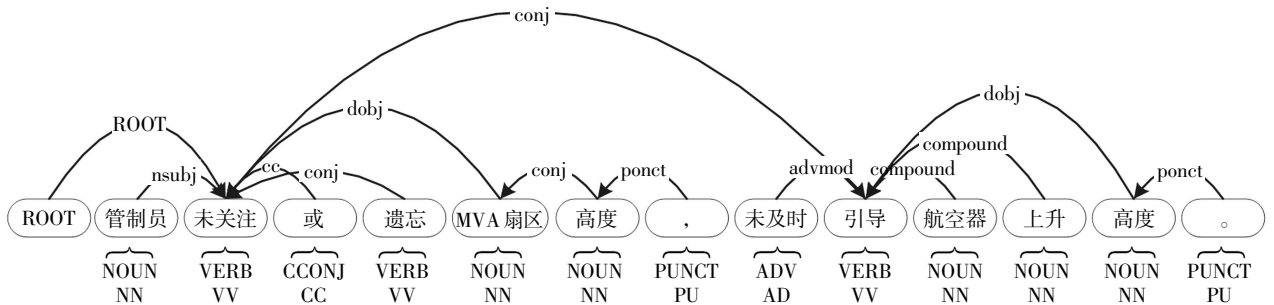


图 4 句法依存树示例

Fig.4 Example of a syntactic dependency tree

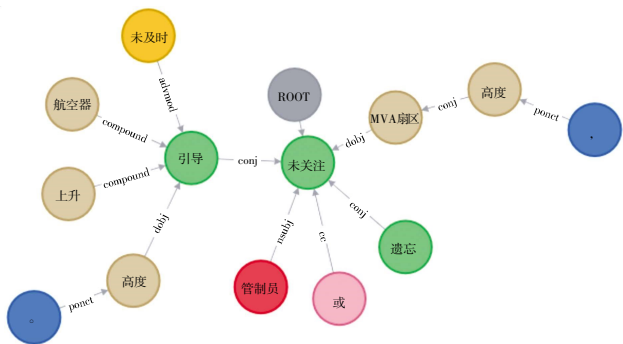


图 5 句法依存分析示例

Fig.5 Example of syntactic dependency analysis

图 4 中“ROOT”代表中心词和根节点,“NOUN”“NN”代表常用名词,“VERB”“VV”代表常用动词,“CCONJ”“CC”代表连词,“PUNCT”“PU”代表停顿和标点符号,“ADV”“AD”代表副词,“nsubj”代表名词性主语,“cc”“conj”“dobj”“ponct”“advmod”“compound”分别代表不同词性之间的依存关系。标注标准是 spaCy 环境中 POS(part-of-speech)、TAG(part-of-speech tag)、

DEP(dependency relation)以及依存关系库^[13]。图 5 中的节点,红色是实体名词,绿色是动词,粉色是连词,棕色是其他名词,黄色是介词,灰色是根节点,蓝色是标点。

根据分词词性和句法依存关系,对语料中的关键信息进行抽取。抽取规则为:

- (1)遍历语料中的所有分词结果,以自定义词典中的触发词为识别标准,抽取出名词性主语(nsubj);
- (2)抽取作为根节点的谓语动词和动词的直接宾语(dobj);
- (3)抽取与根节点相邻的否定副词(ADV);
- (4)抽取自定义词典“其他词”和“人因词”中的其他名词;
- (5)按照语序将抽取结果进行排列。

抽取“触发因素”一栏的要素时采用(1)~(5)条规则;抽取“后果”“现有防控机制”两栏的要素时采用(2)~(5)条规则。抽取结果示例如表 4 所示。

表 4 依据规则抽取自定义实体词结果示例

Tab.4 Examples of custom entities extracted based on rules

| 输入文本 | 文本所在的要素栏 | 抽取结果 | 抽取自定义实体词个数 |
|---|----------|---------------------------------|------------|
| 管制员未关注或遗忘 MVA 扇区高度,未及时引导航空器上升高度 | 触发因素 | 管制员、未关注、MVA 扇区、航空器 | 3 |
| 人工处理计划动态发生错误,存在自动化系统无法自动相关、飞行计划关键信息错误等情形,导致飞行不安全事件,存在较高运行安全风险 | 后果 | 自动化系统、飞行计划、导致、飞行不安全事件、存在、运行安全风险 | 4 |
| 制定 XX 区域管制室与 XX 塔台实施 AIDC 协调移交管制协议 | 现有防控机制 | 制定、XX 区域管制室、XX 塔台、AIDC 协调移交管制协议 | 3 |

3.3 文本相似性匹配与事件类型判断

根据民航空管系统 2022 年危险源控制清单中 4 个季度的危险源数据,从威胁(T)、差错(E)、非期望的航空器状态(U)、确保飞机运行的对策(S)4 个事件类型中,剖析出所有可能导致飞行安全裕度降低的事件,结合 Doc 9868 文件中的“威胁与差错管理”,识别出相应的事件。威胁(T)、差错(E)、非期望的航空器状态(U)、确保飞机运行的对策(S)共有 10 个事件细分类型,如表 5 所示,其中共包含了 29 个可能导致飞行安全裕度降低的事件。

表 5 TEM 事件分类与编号

Tab.5 TEM event classification and numbering

| 事件类型及编号 | 事件细分类型及编号 | 事件识别关键词及编号 |
|---------------------|----------------------|-----------------------|
| 威胁(T) | 内部威胁(T_i) | 设备(T_{i1}) |
| | | 工作环境(T_{i2}) |
| | | 程序(T_{i3}) |
| | | 其他管制员(T_{i4}) |
| | 外部威胁(T_x) | 机场布局(T_{x1}) |
| | | 助航设施(T_{x2}) |
| | | 空域基础设施/设计(T_{x3}) |
| | | 相邻管制单位(T_{x4}) |
| | 空中威胁(T_A) | 飞行员(T_{A1}) |
| | | 航空器性能(T_{A2}) |
| 无线电通话传输(T_{A3}) | | |
| 空中交通情况(T_{A4}) | | |
| 环境威胁(T_e) | 天气情况(T_{e1}) | |
| | 地理环境(T_{e2}) | |
| | 其他干扰项(T_{e3}) | |
| 差错(E) | 设备操作差错(E_e) | 雷达使用(E_{e1}) |
| | | 自动化系统(E_{e2}) |
| | | 无线电/对讲机通话(E_{e3}) |
| | | 飞行进程单(E_{e4}) |
| | 程序差错(E_p) | 工作岗位交接(E_{p1}) |
| | | 信息传递(E_{p2}) |
| | | 文件(E_{p3}) |
| | | 检查单(E_{p4}) |
| | | 最小间隔(E_{p5}) |
| | 通信差错(E_c) | 管制员对外部(E_{c1}) |
| | 管制员之间(E_{c2}) | |
| 非期望的航空器状态(U) | 地面非期望的航空器状态(U_c) | — |
| | 空中非期望的航空器状态(U_A) | — |
| 确保飞机运行的对策(S) | — | — |

模型基于条件随机场序列标注的命名实体识别技术,以自定义词典中的词语为事件触发词,识别文本中的事件及其分类。由于 dependency parser 判别函数只允许当前词依存一个词,所以本文选择 semantic dependency parser 判别函数,依据 SEMEVAL2016 数据集,判别句子中不同词语间的语义依存关系,如图 6 所示。将结果可视化表示为知识图谱,如图 7 所示。

为探究 TEM 模型中威胁、差错和非期望的航空器状态之间的相关性,需要采用改进的余弦相似度计算方法,将空管系统危险源语料通过文本语义相似度匹配,分别映射到上述 TEM 模型中 29 个可能导致飞行安全裕度降低的事件上。DAE 能很好地编码语料中所包含的语义信息,可以在一定程度上解决文本语义相似度计算中的语义理解问题和词向量相似度计算中的词序问题。基于 DAE 网络的词向量生成模型如图 8 所示。

本文的余弦相似度计算公式^[14]为

$$\cos(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{k=1}^K \mathbf{X}_k' \mathbf{Y}_k'}{\sqrt{\sum_{k=1}^K \mathbf{X}_k'^2} \sqrt{\sum_{k=1}^K \mathbf{Y}_k'^2}} \quad (1)$$

式中: \mathbf{X}, \mathbf{Y} 为不同的词语抽取结果; $\mathbf{X}_k', \mathbf{Y}_k'$ 为对应的输出向量; k 代表输入的文本编号; K 为参与运算的文本总数。

将文本归一化到[0, 1]区间,即

$$\text{sim}(\mathbf{X}, \mathbf{Y}) = 0.5 + 0.5 \cos(\mathbf{X}, \mathbf{Y}) \quad (2)$$

词向量间的语义相似度是基于词语语义依存关系,以训练得到的词向量值为数据,使用余弦相似度公式计算得到的。

经改进后的余弦相似度为

$$\text{sim}(D_l, D_p) = \left(\frac{\sum_{w_m \in D_l} \sum_{w_q \in D_p} H_{m,q} \times R(w_m) \times R(w_q)}{\sqrt{\sum_{w_m \in D_l} \sum_{w_o \in D_l} H_{m,o} \times R(w_m) \times R(w_o)} \times \sqrt{\sum_{w_e \in D_p} \sum_{w_q \in D_p} H_{e,q} \times R(w_e) \times R(w_q)}} \right) \quad (3)$$

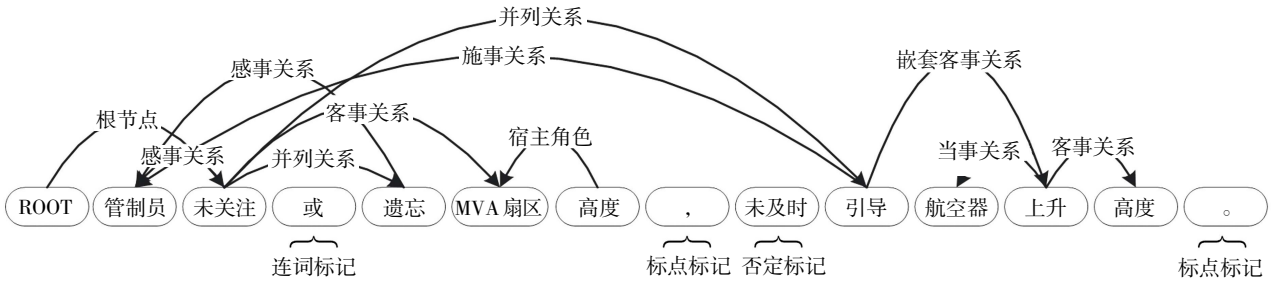


图6 语义依存树示例

Fig.6 Example of a semantic dependency tree

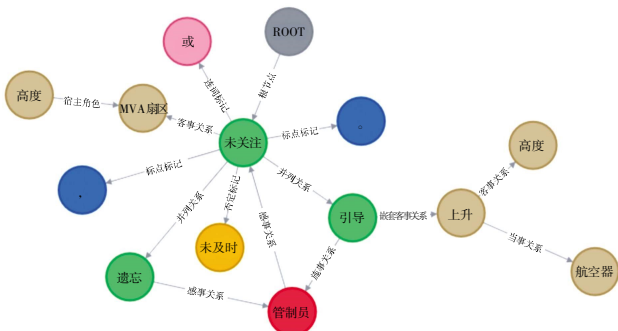


图7 语义依存分析示例

Fig.7 Example of semantic dependency analysis

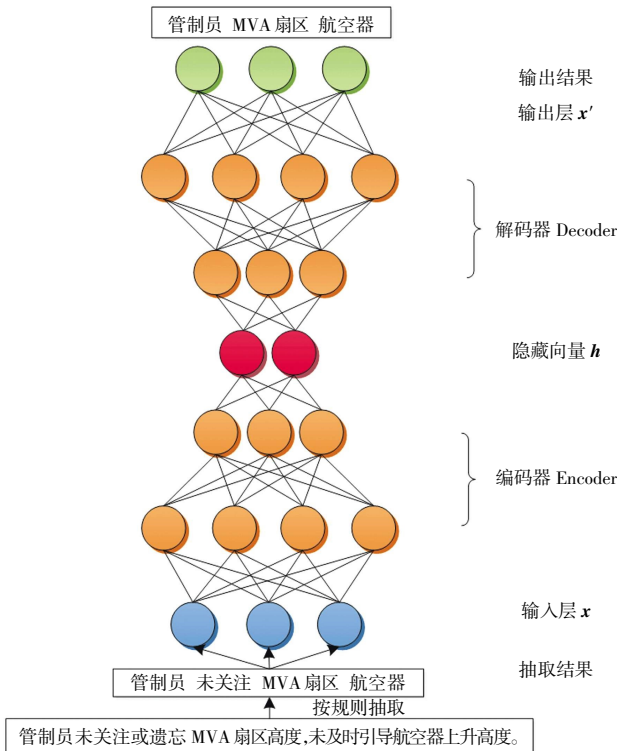


图8 DAE网络模型

Fig.8 DAE network model

式中： D_l, D_p 是不同的两句语料； w_m, w_o 是 D_l 中的词； w_e, w_q 是 D_p 中的词； H 为词向量语义相似度； $R(w_m)$ 是 w_m 的训练词向量值。计算结果在 $[0, 1]$ 区间，数值越大说明两句语料越相似。依据文本语料，判定文本间余弦

相似度不小于 0.667 的两句语料之间具有共指关系，即两句语料属于 TEM 模型本体模式层(图 3)的同一分类。

根据 2.2.1 节事件概念划分，不同事件间的关系类型有因果、并列、时序 3 种。事件关系类型的判别依据如下。

因果：同一条数据的触发因素与后果之间为因果关系；在触发因素中，同一句话里先出现“其他词”，后出现“人因词”，基于两类触发词识别的事件之间为因果关系。

并列：一条数据的某项要素中出现同一词典中的多个触发词，且不同触发词在不同的分点叙述中，基于多个触发词识别的多个事件之间为并列关系；多条数据的同一项要素的事件之间为并列关系。

时序：一条数据的某项要素中出现同一词典中的多个触发词，且不同触发词在同一句话中，基于多个触发词识别的多个事件之间为时序关系。

上述关系示例如图 9 所示，其中，不同颜色的节点代表不同词典的实体词，粉色是“人因词”，绿色是“后果”一栏的“其他词”，橘色是“触发因素”一栏的“其他词”。图 9 中选择了 3 条数据，节点为实体词所代表的事件。

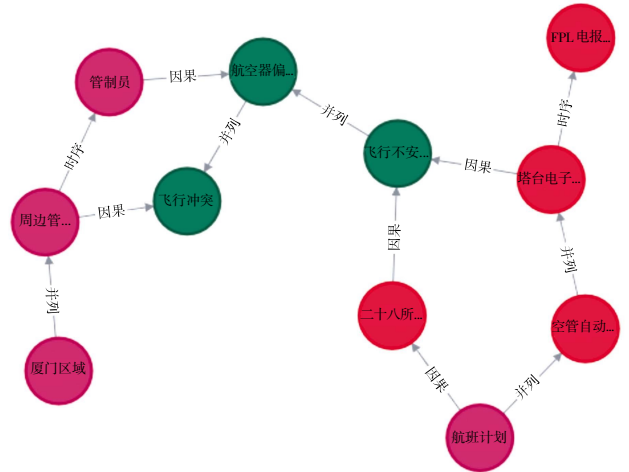


图9 实体词所代表的事件关系示例

Fig.9 Examples of event relationships represented by entities

4 实例分析

4.1 贝叶斯网络在事件关系中的应用

贝叶斯网络是一种具有网络结构的定向图形模型,是人工智能、概率论、图论和决策理论的结合,其使用方向图来表示因素之间的关系和影响程度,并使用节点变量来表示不同因素。本文中每个节点代表不同类型的事件,使用方向弧来表示不同事件的关系,条件概率表用于表示不同事件之间的关联影响程度。

设事件集为 $M = \{M_1, M_2, \dots, M_j\}$, 其中 M_i 表示贝叶斯网络的第 i 个事件节点。所有事件的联合概率分布为

$$P(M_1, M_2, \dots, M_i) = \prod_{j=1}^i P(M_j | M_{j-1}, M_{j-2}, \dots, M_1) \quad (4)$$

式中, j 代表第 j 个事件编号。同时,可以计算每个随机变量的边际概率,即

$$P(M_j) = \sum_{\mathcal{M}_j \in \mathcal{M}_j} P(\mathcal{M}_j, M_j) \quad (5)$$

式中: \mathcal{M}_j 表示事件 M_1, M_2, \dots, M_i 发生与否的一种事件组合; $\mathcal{M}_j = \{\mathcal{M}_j\}$ 表示事件 M_1, M_2, \dots, M_i 发生与否各种事件组合的全集。

通过联合概率和边际概率,可以计算条件概率,即

$$P(M_j | M_i) = \frac{P(M_j, M_i)}{P(M_i)} \quad (6)$$

通常,对于多值随机变量,贝叶斯表达式为

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)} \quad (7)$$

式中: $P(A)$ 、 $P(B)$ 分别表示事件 A 、事件 B 发生的概率; $P(A|B)$ 是先验概率; $P(B|A)$ 是后验概率。为了计算节点之间的相关性,使用先验概率来计算后验概率,而后者表示在特定情况下发生事件的概率。

不同事件类型的实体词识别结果如图 10 所示。实体词个数略大于事件个数,因为可能出现某几个实体词描述同一事件的情况。根据识别结果,结合语料信息进行同一事件的合并,各类型事件个数占总事件数的比例为初始概率,依据上述方法计算得到后验概率,即事件发生概率,如表 6 所示。

4.2 结果分析

所有可能导致飞行安全裕度降低的事件中,威胁不仅会导致差错,还会导致非期望的航空器状态。为了研究威胁、差错和非期望的航空器状态之间的相关性,在民航空管系统 2022 年危险源控制清单数据分析的基础上,按照第 3 节中文本语义相似度匹配的方法将语料填入相应的本体模型分类中。然后挑选存

在因果关系的事件,并选择内部威胁(T_I)、外部威胁(T_X)、空中威胁(T_A)、环境威胁(T_E)、程序差错(E_P)、通信差错(E_C)、地面非期望的航空器状态(U_C)和空中非期望的航空器状态(U_A)作为贝叶斯网络中的节点,形成具有 8 个节点的方向图,如图 11、图 12 所示。图中圆圈表示节点,箭头出发的节点为父节点,箭头指向的节点为子节点,箭头的方向表示因果关系的方向,箭头的粗细和数字代表事件个数,即关系权重。图 11 为第 1 季度未形成完整链路的节点关系图,即此图所示的因果关系只能两两形成,不能形成连续的演化链路。图 12 为第 1 季度已形成完整链路的节点关系图,即在此图中可以任意组成 $T \rightarrow E \rightarrow U$ 的完整链路。

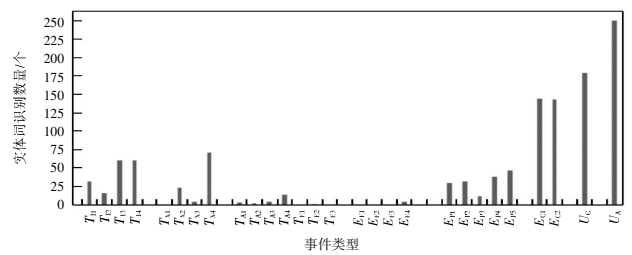


图 10 不同类型事件的实体词识别个数

Fig.10 The number of recognized entities across different event types

表 6 威胁、差错和非期望的航空器状态的发生概率

Tab.6 Probability of occurrence for threats, errors, and undesired aircraft states

| 事件类型及编号 | 事件细分类型及编号 | 事件数/个 | 发生概率 |
|------------------|----------------------|-------|------|
| 威胁(T) | 内部威胁(T_I) | 161 | 0.56 |
| | 外部威胁(T_X) | 102 | 0.35 |
| | 空中威胁(T_A) | 11 | 0.04 |
| | 环境威胁(T_E) | 16 | 0.05 |
| 差错(E) | 设备操作差错(E_E) | 4 | 0.01 |
| | 程序差错(E_P) | 145 | 0.33 |
| | 通信差错(E_C) | 288 | 0.66 |
| 非期望的航空器状态(U) | 地面非期望的航空器状态(U_C) | 180 | 0.41 |
| | 空中非期望的航空器状态(U_A) | 259 | 0.59 |

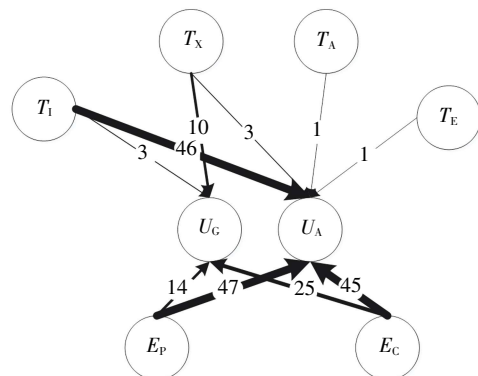


图 11 各节点间的关系权重(非完整链路)

Fig.11 The relationship weights between nodes(incomplete links)

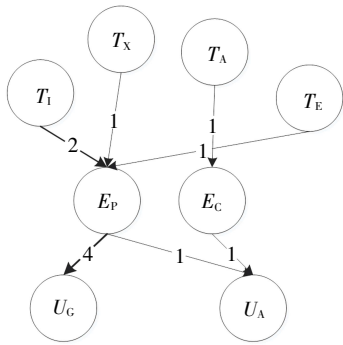


图 12 各节点间的关系权重(完整链路)

Fig.12 The relationship weights between nodes (complete links)

贝叶斯网络模型是在对民航空管系统 2022 年危险源控制清单中所有可能导致飞行安全裕度降低的事件进行研究的基础上建立的。观察后验概率和先验概率之间的差异,并从后验概率中获得威胁、差错和非期望的航空器状态之间的相关性,如表 7—表 9 所示。贝叶斯网络模型能够更实际地获得事件不同因素的权重,从而找到关键因素并采取相应的预防措施。在实际空管运行实践中,可根据本文提出的方法对管制员、管制单位等提出合理化建议。

表 7 威胁关于非期望的航空器状态的后验概率(第 1 季度)

Tab.7 Posterior probability of threats regarding undesired aircraft states (first quarter)

| $P(T U)$ | T_1 | T_x |
|----------|-------|-------|
| U_C | 0.062 | 0.196 |
| U_A | 0.322 | 0.021 |

表 8 差错关于非期望的航空器状态的后验概率(第 1 季度)

Tab.8 Posterior probability of errors regarding undesired aircraft states (first quarter)

| $P(E U)$ | E_P | E_C |
|----------|-------|-------|
| U_C | 0.273 | 0.484 |
| U_A | 0.335 | 0.314 |

表 9 4 个季度的威胁、差错关于非期望的航空器状态的后验概率

Tab.9 Posterior probabilities of threats, errors regarding undesired aircraft states for four quarters

| 后验概率 | 第 1 季度 | 第 2 季度 | 第 3 季度 | 第 4 季度 | 均值 |
|--------------|--------|--------|--------|--------|-------|
| $P(T_1 U_A)$ | 0.896 | 0.896 | 0.900 | 0.897 | 0.897 |
| $P(E_P U_A)$ | 0.522 | 0.512 | 0.418 | 0.456 | 0.477 |
| $P(E_C U_A)$ | 0.478 | 0.488 | 0.582 | 0.544 | 0.523 |

(1)由表 9 可知,在空中非期望的航空器状态(U_A)发生的条件下内部威胁(T_1)发生的概率均值为 0.897,可见其他的威胁类因素对空中非期望的航空器状态(U_A)的影响很小。说明空管系统内部的设备(T_{11})条件、工作环境(T_{12})因素、工作交接程序(T_{13})的合理性、其他

管制员(T_{14})与正在工作的管制员的配合协调性需要进一步提升。

(2)由表 9 可知,在空中非期望的航空器状态(U_A)发生的条件下程序差错(E_P)发生的概率均值为 0.477,通信差错(E_C)发生的概率均值为 0.523。空管系统内部,这两项差错均以人因作为触发词,并且计算得到的概率大小相似,说明管制单位在这两项上预防非期望的航空器状态的措施要并重,措施如多注意管制员工作状态、信息传递准确性等。在事件抽取部分,本文抽取了这两项相关数据的处理措施涉及的实体词,在抽取的结果中涉及与程序差错、通信差错相似度高的实体词数量较少,说明空管单位在这两项差错上的处理措施不够有力,可以继续加强。

(3)由图 12 可知, $T \rightarrow E \rightarrow U$ 的完整链路很少。说明空管系统的各项有效措施可及时阻断事件演化。在 700 条数据中抽取出的 2 334 个事件中,像这样的完整链路仅有 80 条,且其中 63 条为 $T_A \rightarrow E_C \rightarrow U_A$,说明空中威胁发生时,易引发通信差错,进而导致空中非期望的航空器状态的发生。经更详细地抽取和分析发现,无线电通信为空中威胁和通信差错中重要组成部分,这可能是连接演化链路的一项重要因素。所以管制单位需要从无线电通信各环节可能存在的隐患着手,切断演化链路,防止不安全事件的发生。

5 结语

(1)本文提出了基于 TEM 模型的危险源信息抽取方法。针对空管安全系统危险源控制清单文本中不同语料要素构建了自定义词典,在 spaCy 环境下得到空管安全领域词库的词汇、分类、词频等信息。调用了 Python 中的 POS、TAG、DEP 以及依存关系库,得到基于依存关系的特征词抽取结果。结合 TEM 模型和空管领域专业知识构建 TEM 模型本体,对本体中的关键要素、属性等做出规定;选择 semantic dependency parser 判别函数,依据 SEMEVAL2016 数据集,进行语义相似度判别,并与 TEM 模型相似度映射,构建 TEM 模型支持的关键词抽取知识图谱。

(2)本文完成了威胁、差错和非期望的航空器状态之间的相关性分析。利用 DAE 网络模型和改进的余弦相似度公式进行文本语义相似度匹配;基于贝叶斯网络对 TEM 模型本体类别的概率关系进行定量分析,得到威胁、差错和非期望的航空器状态之间的相关性,识别并抽取出文本中的关键信息、与空管运行

(下转第 53 页)