



开放科学
(资源服务)
标识码
(OSID)

航空安全信息风险主题语义图谱构建

刘俊杰 叶英豪 董立映

中国民航大学 安全科学与工程学院 天津 300300

摘要: [目的/意义] 为挖掘航空安全信息中蕴含的风险信息和事件演化规律并进行知识可视化展示,提升航空安全信息中安全风险相关内容的获取效率,实现数据驱动的安全风险管理。[方法/过程] 通过依存句法分析和语义角色标注对信息内容进行语义三元组抽取,基于 LDA 主题模型和主题强度对语义信息进行风险主题建模和核心风险主题确定,构建航空安全信息知识模型,形成风险主题语义图谱。[结果/结论] 经实例验证,本研究识别了确定的核心风险信息以及构建了基于主题划分的航空安全信息语义图谱,为航空安全信息的深度分析与应用提供了有效的解决方案。

关键词: 航空安全; 信息分析; 知识可视化; 风险管理

中图分类号: TP391

Construction of Risk Topic Semantic Map of Aviation Safety Information

LIU Junjie YE Yinghao DONG Liying

School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China

Abstract: [Objective/Significance] In order to mine the risk information and event evolution law contained in aviation safety information and carry out visual display of knowledge, improve the acquisition efficiency of safety risk related content in aviation safety information and realize data-driven safety risk management. [Methods/Processes] Semantic triples are extracted from the information content through dependency parsing analysis and semantic role labeling. Based on LDA topic model and topic strength, risk topic modeling and core risk topic determination are carried out for the semantic information, and aviation safety information knowledge model is constructed to form a risk topic semantic map. [Results/Conclusions] Through the example verification, this study identifies the determined core risk information and constructs the semantic map of aviation safety information based on subject division, which provides an effective solution for the in-depth analysis and application of aviation safety information.

Keywords: Aviation safety; information analysis; knowledge visualization; risk management

作者简介 刘俊杰(1971-), 硕士, 副研究员。研究方向为航空安全数据挖掘、民航安全、飞机性能, E-mail: billows.liu@126.com; 叶英豪(1996-), 硕士研究生, 研究方向为航空安全数据挖掘与分析。

引用格式 刘俊杰, 叶英豪, 董立映. 航空安全信息风险主题语义图谱构建 [J]. 情报工程, 2022, 8(4): 31-40.

引言

航空安全信息的分析与利用对识别安全隐患,探寻事故发生规律,提升安全水平有着极其重要的作用,已成为行业共识。根据人为因素分析与分类系统(Human Factors Analysis Classification System, HFACS),人为因素导致的主观类不安全事件是不安全行为、不安全行为的前提、不安全的监督、组织影响4个层次的失效造成的^[1],不安全事件发生后经过调查并在规定的时限内将基本事实信息和调查报告进行上报,因此在此类航空安全信息中隐藏着大量安全风险相关信息。目前,对这类航空安全信息的分析与利用侧重于基本数据统计与人工分析。其中,统计分析难以深入挖掘不安全事件发生的具体原因和隐藏的安全隐患,人工分析则过于依靠专业人员的经验和专业水平,且耗费大量时间精力。因此,如何快速、高效、客观地获取大量人为因素航空安全信息中蕴含的风险信息和潜在规律,形成具有指导意义的知识,为不安全事件的预防提供基于历史数据的支持便成为了亟待解决的问题。

航空安全信息的分析研究主要涉及以下三个方面:一是基于HFACS等人为因素差错分析模型,分析导致不安全事件发生的原因,或以上述模型为基础结合贝叶斯网络等构建新的人为因素分析模型^[2-5];二是基于不安全事件的文本描述人工分析文本信息中的不安全事件原因、后果和飞行阶段等,以建立不安全事件发展演化机制,制定相应的预防控制措施^[6-8];三是借助信息处理技术对大量的文本类安全报告进行聚类、分类、关联分析等,以识别安全报告中

的风险因素、词条间的语义关系以及不安全事件发生的原因等^[9-13]。

综上,传统的人为因素分析模型和事件文本的分析方法在对大量航空信息中隐含的风险规律的整合和利用方面还存在不足,且因过度依赖专家判断,主观性较强;基于文本信息处理技术的研究数据大多是英文航空安全报告,对符合我国实际情况的中文航空安全信息的研究还较少,且侧重于利用文本聚类或分类方法识别和预测粗粒度的风险因素,对知识可视化和细粒化风险因素挖掘的研究还存在提升空间。

随着深度学习技术的发展,知识组织与知识可视化技术在挖掘与展示专业领域文本信息精炼的核心语义信息方面已有广泛应用^[14-19]。本文拟借助知识组织与知识可视化技术高效、客观地挖掘大规模中文航空安全信息知识内涵和潜在的风险规律,促进知识显化,构建风险主题语义图谱,为中文航空安全信息的知识挖掘、知识发现与知识可视化研究提供参考,为风险管理和安全监管提供基于数据的支持。

1 航空安全信息风险主题建模

1.1 研究框架

首先,将样本数据文本切割成句,形成语料库;其次,利用语言技术平台^[20](Language Technology Platform, LTP)结合自定义词典对语料库进行分词、词性标注、命名实体识别、依存句法分析和语义角色标注;然后,基于知识三元组抽取规则从语料库中抽取<主语,谓语,宾语>结构化形式的航空安全信息知识三元组

并利用 LDA 主题模型进行风险主题建模；基于主题强度确定核心风险主题；最后，可视化展

示和分析核心风险主题及其所包含的知识三元组。

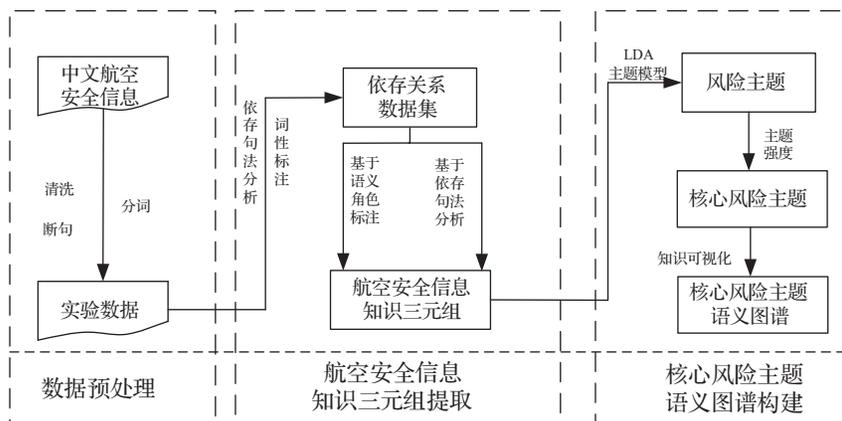


图 1 风险主题语义图谱构建流程

1.2 数据预处理

考虑到中文分词的准确性问题，结合民航领域词库添加民航专业词汇和习惯搭配，构建了一份含有 4891 个专业词汇的自定义词典。

考虑到一义多词干扰问题，引入文本概念模型中的概念词典^[21]。将样本中表达含义相同的领域词汇统一到同一个领域概念之下，实现概念消歧。部分内容见下表 1 所示。

表 1 概念词典

领域概念	原文表达
CRM管理	CRM管理、机组资源管理、驾驶舱资源管理
SOP	SOP、标准程序、标准操作程序、SOP程序
高度改变三步法	高度调整（改变、调定）三步法、高度（层）改变程序（方式）
...	...

1.3 航空安全信息知识三元组获取

依存句法分析（Dependency Parsing, DP）是指用依存语法体系自动对句子的组块序列进

行依存关系分析，以表示句子语法成分及其从属关系的一种句法分析方法。如下图 2 例句，有向弧代表不同句子成分之间的依存关系。根据 LTP 依存句法关系标注集的定义，HED 代表“欠缺”是核心谓词也是整个句子的根节点，SBV 代表“机组”和“欠缺”之间是主谓关系，VOB 代表“欠缺”和“意识”之间是动宾关系，“ATT”代表定中关系，“WP”代表标点。

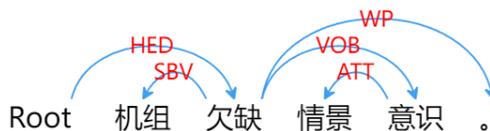


图 2 依存句法分析示意图

语义角色标注（Semantic Role Labeling, SRL）被定义为识别出句子中给定的谓语动词和其他成分，并为其分配语义角色标签，包括施事者、受事者等核心语义角色，还包括时间、地点等附属语义角色。在上图 2 例句中，“欠缺”是核心谓词，“机组”和“情景意识”分别为“欠缺”的两个语义角色，即施事者和受事者。

三元组作为常用的知识表示方法，可以描述真实世界中存在的各种实体或概念及其复杂联系^[22]。因此，本研究结合中文启发式语法规则，利用知识三元组表示非结构化的航空安全信息中所包含的知识条目，进而展开风险主题挖掘及可视化研究，具体规则如下：

规则 1：若存在语义角色标注返回值，基于施事者、受事者、影响范围三种核心语义角色，直接获取知识三元组。

规则 2：若不存在语义角色标注返回值，则依据依存句法分析结果，抽取以谓词为中心的知识三元组：

(1) 主谓宾关系抽取：若句中含有主谓及动宾关系，则抽取知识三元组。

(2) 含有动宾关系的主谓动补结构抽取：若句中仅含有主谓关系，则寻找动补结构或介宾关系，抽取知识三元组。

(3) 定语后置抽取：若句中仅含有动宾关系，则寻找定中关系，抽取知识三元组。

(4) 谓词并列结构处理：以句中谓语为核心，判断主谓宾及并列关系，若并列要素为名词，则共享同一个谓词，抽取知识三元组；若并列要素为谓词，则共享同一个主语，抽取知识三元组。

(5) 宾语并列结构处理：确定句中同核心谓词之间的关系分别为主谓和动宾关系的两个词语 A 和 B。若有其他词语与 A 之间为并列关系，则知识三元组客体相同；若有其它词语与词语 B 之间为并列关系，则知识三元组主体相同。

规则 3：主语、宾语内容补全。根据依存句法分析结果，发现句子结构中修饰主语或宾

语的修饰词，即定中关系，对知识三元组结果中的主语和宾语进行扩展。

规则 4：谓词含义补全，若核心谓词前有其发生状中关系的词语，则将其按顺序组合添加到核心谓词前，补全核心谓词含义。

1.4 航空安全信息风险主题挖掘

航空安全信息包含了大量运行过程中产生的风险信息，是发现并纠正当前潜在系统缺陷、改善航空安全的重要依据。如果将航空安全信息中包含的潜在风险模式抽象为不同的文本主题，即以一组具有内在联系的关键特征词表示一个文本主题和一种风险模式，那么同一类型信息中对应的概率较高的几个文本主题则反映了该类不安全事件中的几种主要风险模式，因此可将基于文本挖掘技术得到的文本主题抽象为风险主题。

1.4.1 LDA主题模型

LDA (Latent Dirichlet Allocation, LDA) 主题模型^[23]，即潜在狄利克雷分配模型，其原理见图 3。

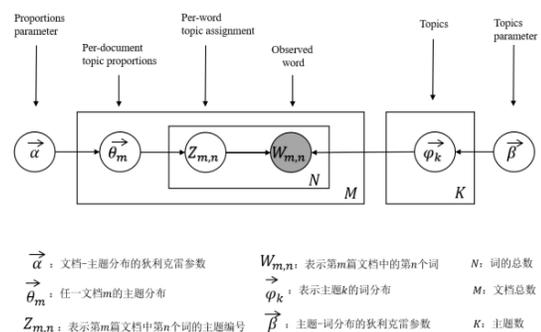


图 3 LDA 原理图

各个参数的具体含义见图 3。LDA 主题模型的生成过程主要包括两个过程：

(1) $\vec{\alpha} \rightarrow \vec{\theta}_m \rightarrow Z_{m,n}$ 过程表示依照参数为 $\vec{\alpha}$ 的“文档-主题”多项式分布所对应的先验狄利克雷分布, 获得多项式分布参数为 $\vec{\theta}_m$ 的文档 m 及其所对应的主题, 然后依照这个参数为 $\vec{\theta}_m$ 的多项式分布生成第 n 个词语的主题 $Z_{m,n}$ 。

(2) $\vec{\beta} \rightarrow \vec{\varphi}_k \rightarrow W_{m,n} | Z_{m,n}$ 过程表示依照参数为 $\vec{\beta}$ 的“主题-词”多项式分布所对应的先验狄利克雷分布, 采样多项式分布参数为 $\vec{\varphi}_k$ 的主题 k 及其所对应的词, 然后再依照这个参数为 $\vec{\varphi}_k$ 的多项式分布获得文档 m 中对应于主题 $Z_{m,n}$ 的第 n 个词。

LDA 主题模型能够良好地表达文本内各词语特征项之间的内在联系。而航空安全信息知识三元组也可看作一个个词语特征项, 比如“[机组, 欠缺, 情景意识]”, 因此利用 LDA 主题模型将知识三元组组织起来, 刻画其内在联系, 挖掘信息中潜在的风险主题是合理的, 有助于形成具有指导意义的潜在规律。

1.4.2 主题强度

主题强度指在每一篇文档中特定主题出现的概率的累计值。对于第一篇文档的 K 个主题构造过程来说, 有 $M_1=[p_1, p_2, p_3, \dots, p_k]$, 其中 p_k 代表第一篇文档属于第 K 个主题的概率, 由此 M 篇文档便可构造一个 $M \times K$ 的主题分布矩

阵 $A_{M \times K} = \begin{bmatrix} p_{11} & \dots & p_{1k} \\ \vdots & & \vdots \\ p_{m1} & \dots & p_{mk} \end{bmatrix}$, 将矩阵中不同主题的概率值累加起来后再除以文档总数, 得到各个主题

出现的总体的概率, 即主题强度, 公式如下:

$$T_k = \frac{\sum_{m=1}^M p_{mk}}{M} \quad (1)$$

式中, T_k 代表主题强度, p_{mk} 代表不同主题在主题分布矩阵中的概率值, M 代表文档总数。

从概率的角度讲, 某一个或某几个主题强度较大, 则代表该主题在样本语料中出现频繁, 具有较为重要的地位, 可作为确定核心主题的依据。因此在对航空安全信息进行风险主题挖掘时, 主题强度可用于从大量的风险主题中辅助判断核心风险主题, 即判断航空安全信息中的核心风险模式。

2 实例分析

选取 2017-2020 年“偏离姿态/高度”类型事件信息共 1066 条。此类航空安全信息属于强制报告信息, 内容真实可靠; 且格式有《事件信息填报和处理规范》(AC-396-03R1) 的要求, 较为规范, 能准确和客观地反映事件发生经过。

2.1 知识三元组抽取结果

利用 LTP 对经过预处理的 1066 条数据进行依存句法分析和语义角色标注操作, 根据抽取规则获取相应知识三元组, 共耗时 4870 秒。清洗无法表达完整事件的无效数据, 共获得 4420 组知识三元组, 截取的部分实验结果如图 4 所示。

```
[[['误认为', '需要', '继续上升'], ['机组', '欠缺', '规章意识'], ['机组', '改变', '高度']]
[['机组', '未严格监控', '高度'], ['机组', '未发现', '高度非指令性上升'], ['机组', '建立', '着陆构型'], ['着陆构型', '早于', '手册'], ['机组', '未执行', '管制指令']]
[['机组', '提前预位', 'APP'], ['机组', '未按', '要求'], ['机组', '未执行', '程序'], ['未有效监控飞机状态', '及时干预', '飞机非预期性下降'], ['飞机', '穿越', '目标高度']]
[['机组', '误听', '管制指令'], ['机组', '实施', '标准喊话和交叉检查'], ['机组', '证实', '指令'], ['机组', '错过', '避免错误最佳时机'], ['了', '避免', '错误']]
[['飞行员', '自行改变', '高度']]
[['机组', '误碰', '高度窗旋钮'], ['机组', '未证实', '对飞行控制面板动作'], ['机组', '担心', '穿越区域']]
[['机组', '错误设置', '目标高度']]
[['机组', '误听', '管制指令']]
[['机组', '未及时转换', '气压基准']]
[['机长', '未发现', '副驾驶操作偏差']]
[['机组', '未严格执行', 'SOP']]
```

图 4 知识三元组抽取结果

2.2 LDA风险主题建模结果

利用 Python 编程调用 LDA 主题模型，结

合困惑度与专家咨询法设置 $K=6$ 。主题挖掘结果分别如下表 2 和图 5 所示。

表 2 主题 - 词概率分布表

风险主题	关键特征词
动作幅度差错 (主题0)	油门、带杆、修正量、改平迟缓、操纵、抬轮、模式
注意力分配 (主题1)	误听、未合理分配、情况、注意力、欠缺、未严格监控、CRM管理
动作时间差错 (主题2)	未及时设置、高度、气压基准、未及时修正、提前预位、偏晚、早于
遗漏 (主题3)	忘记、交叉检查、未进行、设置标准气压、气压基准、忽略、未落实
动作目标差错 (主题4)	高度、错误设置、气压基准、误碰、TOGA电门、错误选择、航向旋钮
违规 (主题5)	交叉检查、未严格执行、SOP、标准喊话、高度、违反、未遵守

将样本信息所包含的潜在风险模式抽象为不同的文本主题，则同类型信息对应的概率较高的几个主题则反映了该类不安全事件中的几种主要风险模式，因此可将挖掘得到的文本主题抽象为风险主题。基于对主题词概率分布表中的每一组主题关键特征词以及原始数据的归纳分析，将主题 0-5 分别定义为动作幅度差错、注意力分配、动作时间差错、遗漏、动作目标差错、违规六个风险主题。

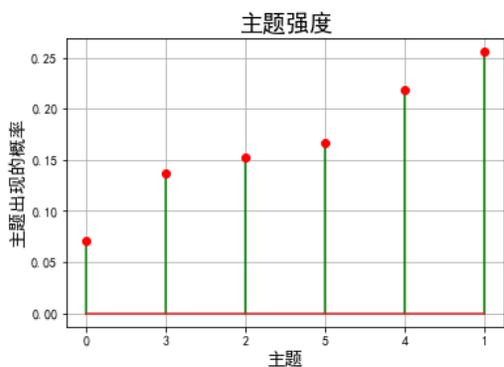


图 5 风险主题强度

对样本中各个风险主题出现的概率进行排序，挖掘样本中相对重要的风险主题。从图 5 可看出各个风险主题的重要度从高到低依排序依次为主题 1、4、5、2、3、0。其中风险主题 1、4、

5 在样本语料中出现的概率相对来说高于 2、3、0，表明三个主题在样本语料中是主要风险主题，是“偏离姿态 / 高度”类型不安全事件风险精准管理重点。

2.3 风险主题语义图谱构建及分析

在航空安全信息知识可视化方面，采用 Gephi 可视化分析软件实现风险主题语义图谱的呈现。Gephi 是一款数据可视化工具，可将文本数据中的关系进行精确的量化分析^[24]。清洗有缺失的无效数据、合并重复数据，将知识三元组 < 主语，谓语，宾语 > 中的主语和宾语视作头节点和尾节点，将谓语视作定义节点之间关联关系的边，输入到 Gephi 中形成风险主题语义图谱。

2.4.1 违规风险主题

将涉及“违规”风险主题的知识三元组数据筛选出来，构建该风险主题语义图谱，设置节点标签和边标签，如图 6 所示。

Gephi 将 PageRank 算法引入到节点的重要性评估当中。将网络中节点之间的链接关系抽象为网页之间的链接关系，通过计算节点的

PageRank 值即可对网络中各个节点的重要性进行排序,可有效挖掘关键节点,PageRank 值越高,代表该节点越重要^[24]。

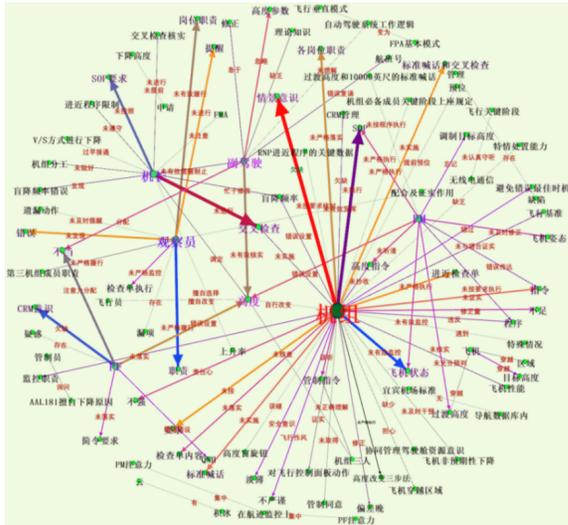


图 6 违规风险主题语义图谱

表 3 违规风险主题节点的 PageRank 值

节点标签	PageRank
高度	0.0243
高度指令	0.0232
交叉检查	0.0159
CRM管理	0.0155
SOP要求	0.0154
飞机状态	0.0109
职责	0.0105
情景意识	0.0094
简令要求	0.0094
检查单执行	0.0093

图 6 和表 3 反应了“违规”风险主题中核心的风险信息,即机组地不安全操作行为特征。违规是指违反程序规定,包括违反 SOP 和管制员许可,主要产生于决策和执行阶段。飞行高度改变是一个程序化的过程,有严格的 SOP 要求,机组在未取得管制许可指令前不得擅自改变高度,且改变高度时需要严格遵守“高度层改变三步法”。但是有部分机组在遇到特殊情

况后,如天气、任务需求等,没有取得管制许可,就“擅自决策,自行改变高度”,做出错误决策并执行;还有部分机组虽按照管制指令改变高度,但是在执行指令时却出现习惯性的“未落实高度改变三步法”、“未进行交叉检查”“违反 SOP、简令要求”、“忽略飞机状态”等行为,需通过专项教育、负向强化、修订 SOP 等方式加以关注和预防。

2.4.2 动作目标差错风险主题

筛选出涉及“动作目标差错”风险主题的知识三元组,构建该风险主题语义图谱,如图 7 所示,图谱中部分节点的 PageRank 值如表 4 所示。

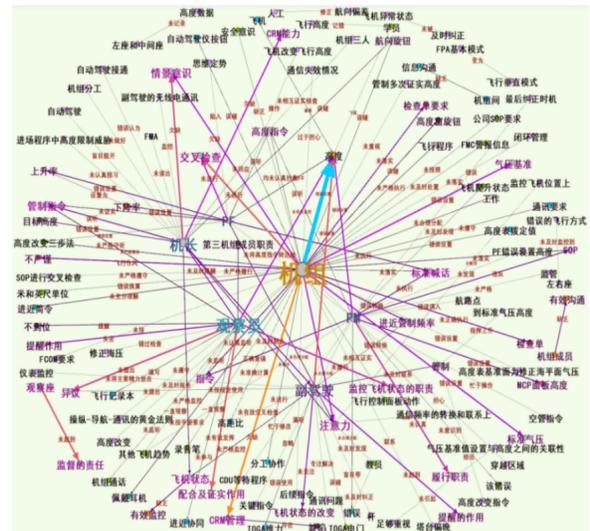


图 7 动作目标差错风险主题语义图谱

表 4 动作目标差错风险主题节点的 PageRank 值

节点标签	PageRank
高度	0.0147
气压基准	0.0135
高度指令	0.0133
高度改变三步法	0.0125
注意力	0.0110
标准气压	0.0109
交叉检查	0.0081
情景意识	0.0081
CRM管理	0.0078
SOP	0.0074

图 7 和表 4 反应了“动作目标差错”风险主题中核心的风险信息。从机组、观察员等几个语义社区出发可以捕捉到核心风险信息中属于直接原因的有：机组、副驾驶等错误设置高度、气压基准、上升率、下降率，误碰高度窗旋钮、航向旋钮、TOGA 电门等，这些知识三元组表明机组动作执行目标出现了错误，即动作转到邻近、类似或者不相关的目标上，是促使不安全事件发生的直接原因。此外，部分知识三元组则表明了直接原因产生的促成因素：机组欠缺 CRM 管理、情景意识，未严格监控飞机状态，未落实检查单要求，未执行标准喊话、高度改变三步法、交叉检查，观察员未有效发挥配合及证实作用等可以归结为不安全的监督和不安全行为的前提，是促成因素。

2.4.3 注意力分配风险主题

筛选出涉及“注意力分配”风险主题的三元组数据，构建该风险主题语义图谱，如图 8 所示，节点的 PageRank 值如表 6 所示。

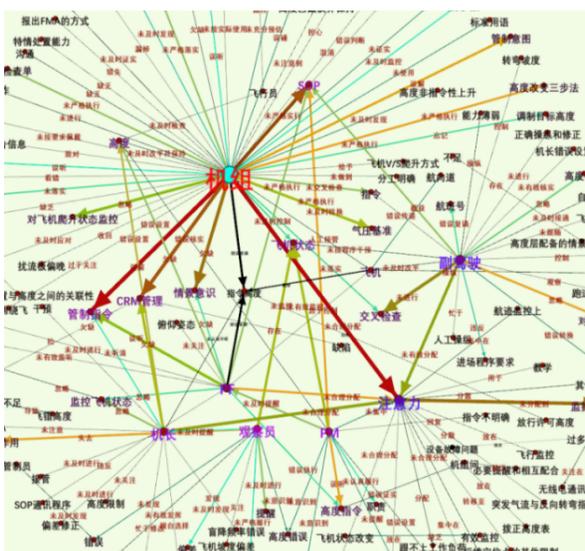


图 8 注意力分配风险主题语义图谱

表 5 注意力分配风险主题节点的 PageRank 值

节点标签	PageRank
注意力	0.0299
飞机状态	0.0184
管制指令	0.0131
监控飞机位置	0.0124
CRM管理	0.0123
指令高度	0.0115
SOP	0.0114
交叉检查	0.0103
高度	0.0102
情景意识	0.0078

图 8 和表 5 反应了“注意力分配”风险主题中的核心风险信息。注意力分配在执行动作过程中由于注意力分配不当产生的差错。高度改变是多任务过程，需要机组间相互配合，需要一定的“CRM 管理”，尤其是遇到诸如雷雨天气等的“特殊情况”，更需要机组“合理分配注意力”，保持“情景意识”。此外，在巡航阶段，飞机接通自动驾驶仪，Norman Mackworth 的“性能衰减”理论表明人很难保持持续的注意力，因此机组设置高度之后会出现“未严格监控”飞机状态、“未进行交叉检查”的情况，直到经管制提醒飞机偏离指令高度，机组才注意到存在“误听指令”、“错误设置高度”等问题。

综上，构建的航空安全信息风险主题语义图谱可将一定时间范围内同类型不安全事件数据核心内容聚焦于清晰明了的可视化图形中。从信息分析人员的角度看，风险主题语义图谱能做到对航空安全信息的知识化凝练，无需对大量的信息进行人工筛选处理，通过计算机即可快速、客观地获取信息中地核心风险，提升

了对航空安全信息资源的利用效率。从监管者的角度来看,风险主题语义图谱能够为不安全事件的宏观把握与整体趋势分析提供有效的解决方案,为局方或企业安全管理人员有针对性的管控安全风险提供准确的语义信息,具有明确的语义指导价值。

3 总结

通过构建航空安全信息风险主题语义图谱,为基于航空安全事件信息的智慧化民航安全风险管理与监管提供数据支撑,为航空安全信息资源的分析与利用提供新思路。基于依存句法和语义角色标注的知识三元组抽取规则可从无序的、非结构化的航空安全信息中获取精炼的、结构化的核心语义特征;LDA主题模型与主题强度可确定航空安全信息中的核心风险主题;PageRank值可辅助确定核心风险节点,便于安全管理人员对各类不安全行为进行针对性管控。通过以上过程构建的风险主题语义图谱能够详细、全面地展示不安全事件发生过程中机组的相关操作行为特征。同时,航空安全信息知识三元组抽取和风险主题挖掘过程由计算机程序执行,可提升航空安全信息资源分析与利用效率。

由于本研究所提出的信息风险主题语义图谱构建方法是基于Python编程的自动化过程,且中文的基本语法结构是固定的,所以知识三元组抽取和风险主题挖掘过程无需做太大改变,未来可便捷地拓展到其他类型的航空安全信息,包括强制报告信息和自愿报告信息等。

需要指出的是,语料库表述规范性和自然

语言处理工具性能的限制会对知识三元组抽取的准确性造成干扰。未来可考虑采用机器学习与深度学习的方式扩展语料库和训练知识三元组抽取模型,进一步提高准确性。

参考文献

- [1] Salmon P, Cornlissen M, Trotter M. Systems-based accident analysis methods: a comparison of Accimap, HFACS, and STAMP[J]. Safety Science, 2012, 50(4): 1158-1170.
- [2] 向维,李明,吴超,等.航空不安全事件人为因素分析 R-S-TER 模型的构建与应用研究[J].中国安全科学学报,2009,19(2):152-159.
- [3] 甘旭升,崔浩林,高文明,等.基于 HFACS 的空中相撞事故分析及建议[J].中国安全生产科学技术,2015,11(10):96-102.
- [4] Omole H, Walker G. Offshore transport accident analysis using HFACS[J]. Procedia Manufacturing, 2015(3):1264-1272.
- [5] Zhou T, Zhang J Y, Dashzeveg B. A hybrid HFACS-BN model for analysis of mongolian aviation professionals' awareness of human factors related to aviation safety[J]. Sustainability, 2018, 10(12):4522.
- [6] 刘俊杰,张丽娟.飞行疲劳事件语义网络分析[J].中国安全科学学报,2016,26(1):34-39.
- [7] 崔振新,孙文龙.基于 WFA 的中美航空事故调查报告安全建议对比[J].中国民航大学学报,2020,38(2):51-54.
- [8] 刘俊杰,赵永正,张丽娟.基于知识地图的民用航空安全信息事件分析[J].安全与环境学报,2018,18(2):640-645.
- [9] 刘梦娜.基于文本挖掘的航空安全事故报告致因因素分析和风险预测[D].合肥:安徽建筑大学,2019.
- [10] 周秀婷.基于文本挖掘技术的民航批量不安全事件风险管控研究[D].北京:北京邮电大学,2020.
- [11] Kuhn K D. Using structural topic modeling to identify latent topics and trends in aviation incident reports[J]. Transportation Research Part C: Emerging Technologies, 2018(87): 105-122.
- [12] Nektarios K, Jeffrey N. The controllability

- classification of safety events and its application to aviation investigation reports[J]. Safety Science, 2018(108): 89-103.
- [13] Liu J, Ye Y, Du Y. Text feature extraction and clustering analysis of events caused by the cockpit crew[C]. 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT). IEEE, 2020.
- [14] 李浩飞. 民航突发事件实体识别和关系抽取方法的研究[D]. 天津: 中国民航大学, 2020.
- [15] 马晓君, 刘亚雪, 魏晓雪, 等. 航空公司微博评论的意见信息抽取研究——以国航、南航和东航为例[J]. 系统科学与数学, 2017, 37(4):1072-1091.
- [16] 王红, 祝寒, 林海舟. 航空安全事故因果关系抽取方法的研究[J]. 计算机工程与应用, 2020, 56(11):265-270.
- [17] 陈芳, 陈茜, 徐碧晨. 基于文本挖掘的管制运行风险主题分析[J]. 中国安全生产科学技术, 2020, 16(11):47-52.
- [18] 景国勋, 王远声, 郭昕曜, 等. 2019年国内航空安全领域研究现状分析[J]. 安全与环境学报, 2020, 20(5):2024-2028.
- [19] 程程, 徐吉辉. 社会网络分析法在国内航空安全管理体系研究现状的应用与分析[J]. 航空工程进展, 2018, 9(3):405-410.
- [20] Che W X, Li Z H, Liu T. LTP: A Chinese Language Technology Platform[C]. In Proceedings of the Coling 2010: Demonstrations. Beijing: ACM, 2010: 13-16.
- [21] 陈龙, 范瑞霞, 高琪. 基于概念的文本表示模型[J]. 计算机工程与应用, 2008, 44(20): 162-164.
- [22] 安宁, 安璐. 突发公共卫生事件舆情环境下的群体智慧涌现研究[J]. 情报学报, 2022, 41(1):96-106.
- [23] Blei D M, Ng A Y, Jrodan M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003(3): 993-1022.
- [24] 刘勇, 杜一. 网络数据可视化与分析利器: Gephi中文教程[M]. 北京: 北京电子工业出版社, 2017.