

基于贝叶斯网络的空管信息抽取相关性偏差分析

华 炜, 郭 庆, 左 燕

民航华东地区空管局, 上海

收稿日期: 2024年8月8日; 录用日期: 2024年9月18日; 发布日期: 2024年9月26日

摘 要

本文首先结合事件类型识别出危险源文本中相应的事件, 接着识别对相关性的存在影响的中性词语, 建立基于后门调整的主题分析模型, 剔除主题混杂和语义相似性偏差带来的相关性分析不准确的问题。最后, 基于贝叶斯网络计算不同节点的发生概率并得到非完整链路和完整链路的演化分析结果。

关键词

空中交通管理, 安全管理, 贝叶斯网络模型, 信息抽取, 语义相似性

Bayesian Network-Based Correlation Bias Analysis for Air Traffic Control Information Extraction

Wei Hua, Qing Guo, Yan Zuo

Civil Aviation East China Air Traffic Control Bureau, Shanghai

Received: Aug. 8th, 2024; accepted: Sep. 18th, 2024; published: Sep. 26th, 2024

Abstract

In this paper, we first identify the corresponding events in the text of the hazard source by combining the event types, then identify the neutral words that have an influence on the correlation, and establish the theme analysis model based on backdoor adjustment to eliminate the inaccuracy of the correlation analysis caused by the theme mixing and semantic similarity deviation. Finally, based on Bayesian network, we calculate the occurrence probability of different nodes and obtain the evolution analysis results of non-complete and complete links.

Keywords

Air Traffic Management, Safety Management, Bayesian Network Model, Information Extraction, Semantic Similarity

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在我国,根据《民航空中交通管理安全管理体系(SMS)建设指导手册》进行民航安全管理[1]。毛佳静[2]使用HAZOP方法识别平行跑道使用RNP AR进近程序的危险源,重点研究了切入五边时的碰撞风险,以填补我国在EoR技术运行风险评估方面的研究空白。李海等[3][4]则利用AHP方法研究了通航运行中危险源的危险程度评估问题,解决了评估结果无量化的问题,对危险源的评估具有重要意义。Raskin等[5]利用本体论和基于意图的语义本体论,分析了如何避免特内里费岛灾难,并将这种方法用于理解自然语言和计算机处理信息的抽取。王洁宁等[6]、刘继新等[7]建立了空中交通管制人为因素的本体模型,研究了危险空中交通管制中的人为失误,并发现了危险事故之间的潜在联系。Zhao等[8]基于TEM模型,建立了贝叶斯网络模型,对空中交通管制运行中的威胁、错误和不期望状态之间的相关性进行了精确的定量分析。徐一旻等[9]将故障树模型与贝叶斯模型结合,分析了影响跑道入侵的主要人因要素,对跑道入侵安全领域提出了合理化建议。

以上研究中,虽已使用NLP方法进行信息抽取,但多是使用自然语言分词,以开放域语料作为基础语料库,缺少空管领域专业分词的词库;以往的文本信息相关性分析多是基于词向量,本文将语义相似度引入相似度计算表达式中,使空管领域相关文本的相似性分析更准确。

2. 抽取信息的语义相似性偏差分析

2.1. 危险源抽取信息事件分类

在进行相似性匹配和事件类型判断时,语义相似性是重要的参考线索。抽取的关键词中有类似“风切变”这样明确指向“环境威胁”的实体词;有类似“管制员”这样明确指向“通信差错”的实体词;也有类似“飞行冲突”这样明确指向“空中非期望状态”的实体词。但是抽取的关键词中有许多词语具有“中性词”特征,比如“军航”、“进程单”等,在威胁、差错、非期望的航空器状态的不同主题里,这些词语并不具有明显的主题偏向。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,剖析出所有可能导致飞行安全裕度降低的事件,结合Doc 9868中的“威胁和差错管理”,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种事件类型,如表1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(M)。

这些中性词有的对最终的相似性分析影响较大,有的影响很小,所以在对语义相似判别的实验中,受到影响较大的中性词词向量相似的干扰,中性词的词频会影响最终的相似性匹配结果。模型在进行实验的过程中,尽管给出正确的相似性匹配结果,但是模型并未真正区分当前是因为“中性词”词频相似程度高所以文本相似;还是因为不同主题不同分词的词向量相似所以文本相似。所以要针对前文的实验结果,使用词向量、先验概率等线索建立语义相似性去偏模型,剔除由语义偏差和主题混杂带来的虚假

相关影响，提高相关性分析的准确率。

Table 1. Event classification and numbering
表 1. 事件分类与编号

类型	种类及编号	事件识别关键词及编号
T	内部威胁(TI)	设备(TI1)、工作环境(TI2)、程序(TI3)、其他管制员(TI4)
	外部威胁(TX)	机场布局(TX1)、助航设施(TX2)、空域基础设施/设计(TX3)、相邻管制单位(TX4)
	空中威胁(TA)	飞行员(TA1)、航空器性能(TA2)、无线电传输(TA3)、空中交通情况(TA4)
	环境威胁(TE)	天气情况(TE1)、地理环境(TE2)、其他干扰项(TE3)
E	设备操作差错(EE)	雷达使用(EE1)、自动化系统(EE2)、无线电/对讲机通话(EE3)、飞行进程单(EE4)
	程序差错(EP)	工作岗位交接(EP1)、信息传递(EP2)、文件(EP3)、检查单(EP4)、最小间隔(EP5)
	通信差错(EC)	管制员对外部(EC1)、管制员之间(EC2)
U	地面非期望状态(UG)	1) 飞机在停止时继续滑行；2) 滑向错误的滑行道、停机坪、登机门或等待点等
	空中非期望状态(UA)	1) 未经批准的空域穿越；2) 超出航空器限制规定的运行等
S	——	1) 飞行计划处理工作流程；2) 扇区监控席、指挥席和协调席的增加配合，相互提醒等

2.2. 主题混杂

Landeiro 等(2019) [10]曾提出，在 IMDB 电影评论数据集中，由于选择偏差的存在，在涉及不同类型的电影评论时会倾向于表达不同的感情色彩。在本文的相关性研究中，由于中性词语义选择偏差的存在，在涉及不同主题的相关性时会出现偏差。使用威胁、差错、非期望的航空器状态的不同主题为混杂因子，采用关键词的词向量的训练值作为输入特征，使用简单的逻辑回归模型学习相关性预测，影响较高的词汇列表如表 2 所示。

Table 2. Neutral words and numbers with high impact on correlation
表 2. 对相关性影响较大的中性词及编号

威胁和差错涉及的中性词	非期望的航空器状态涉及的中性词
告警 a01、军航 a02、航班 a03、机场 a04、鸟击事件 a05、航空器 a06、航线 a07、高度限制 a08、流量 a09、席位 a10、组员 a11、人员 a12、军民航协调 a13	不安全事件 b01、飞行安全 b02、安全事件 b03、安全运行风险 b04、飞行计划 b05

从表 2 中可知，“鸟击事件”对“威胁”和“差错”主题的贡献系数较高，而这一实体词本身并无主题偏向，因此，在涉及不同主题的文本分析时，模型也许会做出错误的判断。本章将模型学习到“鸟击事件”的虚假相关性的原因解释为：主题影响评论文本分布的同时，对语义相似也有偏向性。根据关键词抽取结果，实体词“鸟击事件”更容易出现在“威胁”主题的上下文中，与此同时，源文本在“威胁”主题下涉及内容倾向于与“鸟击事件”语义高度相似的。因此，实体词“鸟击事件”出现时，更容易是“差错”主题。

使用 BERTopic 模型获取训练集的主题概率分布以及主题表示向量。BERTopic 主题模型主要分为 3 步：

- 1) 获取上下文语义的句子表示：所利用的句子编码可以表示任意句子向量表示模型，比如 Glove 向量表示、BERT 或 RoBERTa 等预训练语言模型的句子表示都可以作为其向量表示基本模型。
- 2) 对句子表示进行降维：文档的表示向量维度较高，在词嵌入空间非常稀疏，不易实现聚类，因此，

该工作使用 UMAP 算法进行降维。

3) 对降维后的句子表示进行聚类：使用基于层次和密度的经典聚类算法 HDBSCAN 进行聚类。

下面是四个季度的信息抽取中涉及“威胁”和“差错”主题的关键词，用表 3 来直观地说明在考虑主题混杂与不考虑主题混杂的两种情况下，所得到的抽取关键词特征与“威胁”、“差错”的分主题之间的相关性是截然不同的。

Table 3. Correlation distribution in the hypothetical example

表 3. 假设例子中的相关分布

512 条关键词抽取数据	威胁(T)主题(358 条)	未提到“鸟击事件”(58 条)	内部威胁(T) (29 条)
			空中威胁(TA) (29 条)
	差错(E)主题(154 条)	提到“鸟击事件”(300 条)	外部威胁(TX) (149 条)
			环境威胁(TE) (151 条)
		未提到“鸟击事件”(0 条)	设备操作差(EE) (0 条)
		提到“鸟击事件”(154 条)	程序差错(EP) (77 条)
			通信差错(EC) (77 条)

关键词抽取数据一共 512 条，其中包括 358 条属于威胁(T)主题以及 154 条属于差错(E)主题，这样的抽取结果是符合常理的，提到“鸟击事件”的主题更容易是威胁(T)主题，因为鸟击事件更多与外部威胁、环境威胁有关；但“鸟击事件”这个实体词实际上并没有主题偏向，在更多因为人因导致的差错主题中，也有很大一部分包含关键词“鸟击事件”。

不考虑主题的情况下，在未提到“鸟击事件”的 58 条数据中，威胁(T)主题占比 100%；而提到“鸟击事件”的 454 条数据中，威胁(T)主题占比 66%，差错(E)主题占比 34%。因此，是否提到“鸟击事件”与属于威胁(T)主题或差错(E)主题有强烈的相关性。

当我们按照不同主题分层，则会得到不同的结论：

1) 威胁(T)主题下，在未提到“鸟击事件”的 58 条数据中，内部威胁(TI)与空中威胁(TA)各占比 50%；提到“鸟击事件”的 300 条数据中，外部威胁(TX)与环境威胁(TE)占比各为 50%。因此，在威胁(T)主题中，是否提到“鸟击事件”与威胁分主题的文本相似匹配不存在相关性。

2) 差错(E)主题下，在提到“鸟击事件”的 154 条数据中，程序差错(EP)与通信差错(EC)占比也各为 50%。因此，在差错(E)主题中，是否提到“鸟击事件”与差错分主题的文本相似匹配也不存在相关性。

因此，按照主题分层讨论时，是否提到“鸟击事件”与分主题的文本相似匹配是不存在相关性的，这更符合认知，即“鸟击事件”一词是中性的，本不应该对分主题的文本相似匹配有影响作用。

2.3. 基于后门调整的主题分析模型

在社会科学领域，用来控制混杂的方法包括：匹配、分层，以及回归等。协变量调整(又称后门调整)方法可以去除混杂变量的影响。Landeiro 首先利用后门调整方法提升文本分类任务的鲁棒性。本节与其不同点在于，为了利用预训练语言模型优异的语义表示能力，使用主题的连续性表示作为混杂变量。后门调整是通过 do 算子来实现的。do 算子在因果推断中可以更好地理解量化变量之间的因果关系，尤其是在存在潜在混淆因素的情况下。通过合理地选择和调整后门变量，以及正确地应用 do 算子，可以更准确地估计和处理效应，从而得出更可靠的因果结论。

后门调整思路的核心是，在数据生成过程满足事件间的关系类型为因果关系的前提下，通过后门调整公式将原数据分布转换为新分布，如图 1 所示。

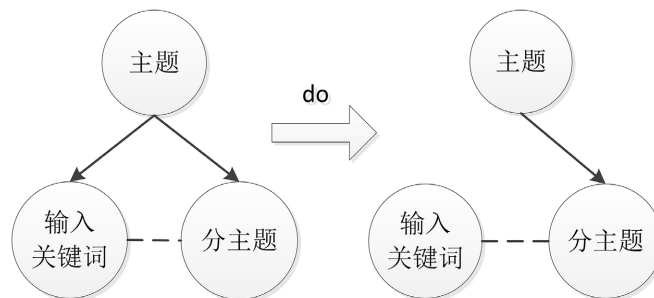


Figure 1. Causal diagram representation after backdoor adjustment
图 1. 后门调整后的因果图表示

在新分布下，混杂(主题)与处置变量(输入文本)之间的相关性不再存在，此时学习到的相关性是抽取信息特征与不同分主题之间的因果相关性。

这里从数据加权的角度解释算子(干预操作)能够进行分布变换的原因。如公式(1)所示，将后门调整公式的分子和分母同时乘以条件概率 $P(X|Z)$ ，可发现后门调整的操作，等价于将总体数据中每条数据 $(Y = y, X = x, Z = z)$ 的分布概率用因子 $\frac{1}{P(X = x|Z = z)}$ 来放大，也就是说，在本节的语义偏向性分析下，如果对于某条数据，主题 z 生成其输入文本 x 的概率越高，那么该样本的权重越低。在该因子的作用下，去除了每个主题与输入文本的相关性。因此，使用网络结构对干预概率(而非条件概率)进行建模，再对干预概率与分类目标之间的损失进行优化训练，模型学习到的相关性才是比较接近因果的相关性，进而可以提升模型的稳健性能。

$$\begin{aligned}
 P(Y | do(X)) &= \sum_z P(Y | X, Z) \otimes P(Z) \\
 &= \sum_z \frac{P(Y | X, Z) P(X | Z) P(Z)}{P(X | Z)} \\
 &= \sum_z \frac{P(Y, X, Z)}{P(X | Z)}
 \end{aligned} \tag{1}$$

如表 4 所示，后门调整的作用是将总体数据中每条数据 $(Y = y, X = x, Z = z)$ 的分布概率按照因子 $\frac{1}{P(X = x|Z = z)}$ 来进行复制，形成新的数据分布，在这个分布下，716 条威胁(T)主题关键词抽取数据中，提到“鸟击事件”和未提到“鸟击事件”的数据各占 358 条；在 308 条差错(E)主题关键词抽取数据中，提到“情节”的数据程序差错(EP)和通信差错(EC)各占 77 条。即主题与是否包含“鸟击事件”这个实体词特征无关，阻断了从主题混杂到输入关键词之间的箭头传播。

Table 4. Data distribution after backdoor adjustment
表 4. 后门调整后的数据分布

512 条关键词 抽取数据 → 1024 条关键词 抽取数据	威胁(T)主题 (358 条→716 条)	未提到“鸟击事件” (58 条→58*358/58 条)	内部威胁(T) (29 条→29*358/58 条) 空中威胁(TA) (29 条→29*358/58 条)
		提到“鸟击事件” (300 条→300*358/300 条)	外部威胁(TX) (149 条→149*358/300 条) 环境威胁(TE) (151 条→159*358/300 条)

续表

差错(E)主题 (154 条→308 条)	未提到“鸟击事件” (0 条)	设备操作差(EE) (0 条)
	提到“鸟击事件” (154 条)	程序差错(EP) (77 条) 通信差错(EC) (77 条)

基于因果干预的主题分析模型的整体框架如图 2 所示。

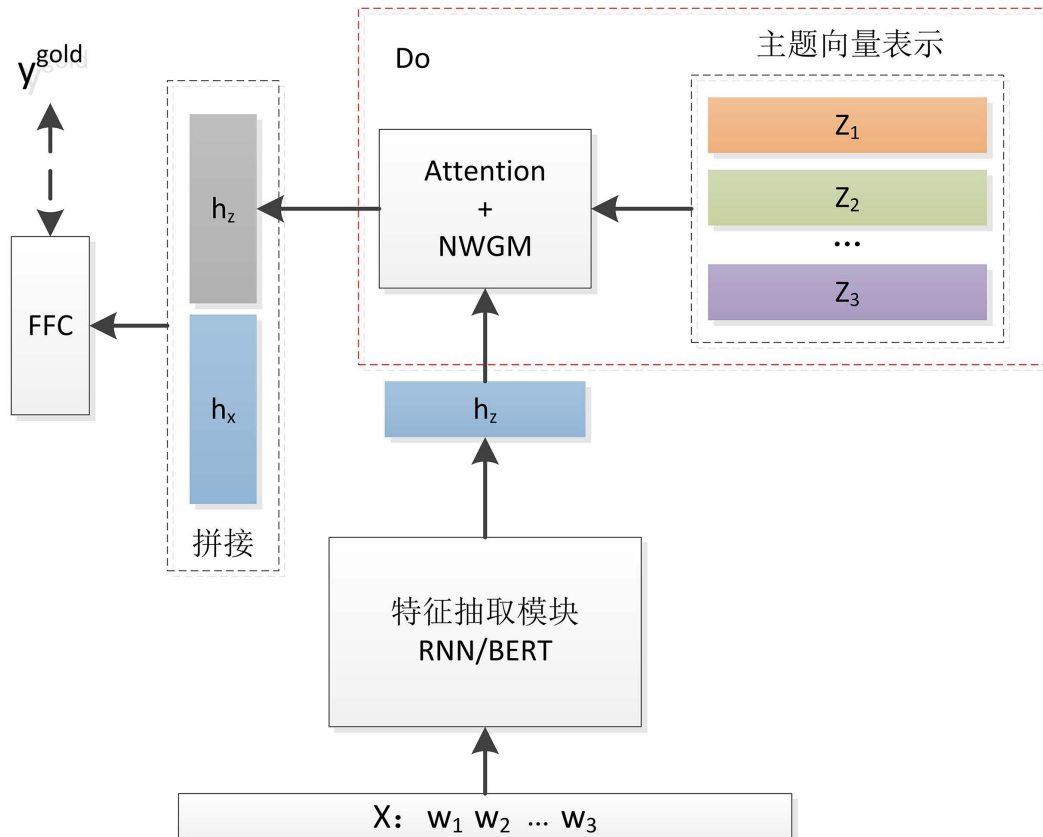


Figure 2. Theme analysis model based on causal intervention

图 2. 基于因果干预的主题分析模型

已知文本分类任务的数据为 $D: \{(X_i, y_i)\}_{i=1,2,\dots,N}$ ，其中 X_i 表示输入文本， y_i 表示分类标签， $y_i \in \{0, 1, \dots, c\}$ 。首先，使用特征抽取模块(RNN-att 或预训练语言模型)对句子进行编码表示： $H_\theta(X)$ ，其中， θ 是文本编码模块的参数。

1) 主题表示：使用文本数据训练主题模型 BERTopic 获取各主题的先验概率以及嵌入表示，这里为保证主题的嵌入表示与文本的编码表示在同一语义空间，BERTopic 句子编码模型与文本分类模型的编码模型保持一致，具体来说，当文本分类模型的特征抽取模块是 BERT 预训练模型(或 RNN、RoBERTa)时，BERTopic 使用的句子编码模型也是 BERT (或 Glove、RoBERTa)，式(2)中 $P(Z)$ 代表各个主题的先验概率， $H(Z)$ 代表各个主题的向量表示，维度取决于 BERTopic 模型所使用的编码模型。

$$P(Z), H(Z) = \text{BERTopic}(\{X_1, X_2 \dots X_N\}) \quad (2)$$

2) 干预概率：通过后门调整计算的干预概率为：

$$P(Y | do(X)) = \sum_z P(Y | X, Z) \otimes P(Z) \tag{3}$$

下面来对干预概率进行建模，具体来说，最后一层预测标签的网络是 Softmax 层，因此有：

$$P(Y | X, Z) = \text{Softmax}(f_Y(X, Z)) \tag{4}$$

汇总来看，后门调整后的干预概率计算方式为公式(5)，这里应用归一化加权几何平均来近似上述期望的计算，即公式(6)：

$$P(Y | do(X)) := E_z [\text{Softmax}(f_Y(X, Z))] \tag{5}$$

$$E_z [\text{Soft max}(f_y(X, Z))] \approx \text{Soft max}(E_z [f_y(X, Z)]) \tag{6}$$

在本章，假定函数 $f_Y(X, Z)$ 的形式为公式(7)， $[\cdot]$ 代表向量的拼接操作， \otimes 代表乘运算符， $\partial_y(Z)$ 代表各个主题对于输入特征来说的重要程度，计算方式为公式(8)，其中， σ 为隐藏层向量表示的维度。

$$f_y(X, Z) = W_c [H_\theta(X); \partial_y(Z) \otimes H(Z)] + b \tag{7}$$

$$\partial_y(Z) = \frac{(W_q H_\theta(X))^T (W_z H(Z))}{\sqrt{\sigma}} \tag{8}$$

最后，可以得到建模后的干预概率：

$$P(Y | do(X)) \approx \text{Softmax} \left[W_c \left[H_\theta(X); \sum_z P(Z) \otimes \partial_y(Z) \otimes H(Z) \right] + b \right] \tag{9}$$

对于干预概率与分类目标之间的损失进行优化训练，这里使用交叉熵函数来计算损失，整体的损失表示为：

$$\frac{1}{N} \sum L_{CE}(P(Y_i | do(X_i)), Y_i^{gold}) \tag{10}$$

3. 抽取信息的相关性分析

将贝叶斯网络模型应用于空管系统运行数据分析(表 5~7)，表明了该模型在分析空管系统危险源事件中的价值。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。

Table 5. Correlation rate of threat and undesirable state (first quarter)

表 5. 威胁和非期望状态的相关率(第一季度)

P (T U)	TI	TX
UG	0.062	0.196
UA	0.322	0.021

Table 6. Correlation rate of error and undesirable state (first quarter)

表 6. 差错和非期望状态的相关率(第一季度)

P (E U)	EP	EC
UG	0.273	0.484
UA	0.335	0.314

Table 7. Correlation rates for four quarters of threat, error and undesirable state
表 7. 四个季度的威胁、差错、非期望状态相关率

P (TI UA)/P (EP UA)/P (EC UA)	一季度	二季度	三季度	四季度	均值	准确率
P (TI UA)	0.896	0.896	0.900	0.897	0.897	0.892
P (EP UA)	0.522	0.512	0.418	0.456	0.477	0.881
P (EC UA)	0.478	0.488	0.582	0.544	0.532	0.902

3.1. 非完整链路分析

如图 3 所示为四个季度中所有抽取事件未形成完整链路的节点关系图，即在此图中所示的因果关系只能两两形成，不能形成连续的演化链路。

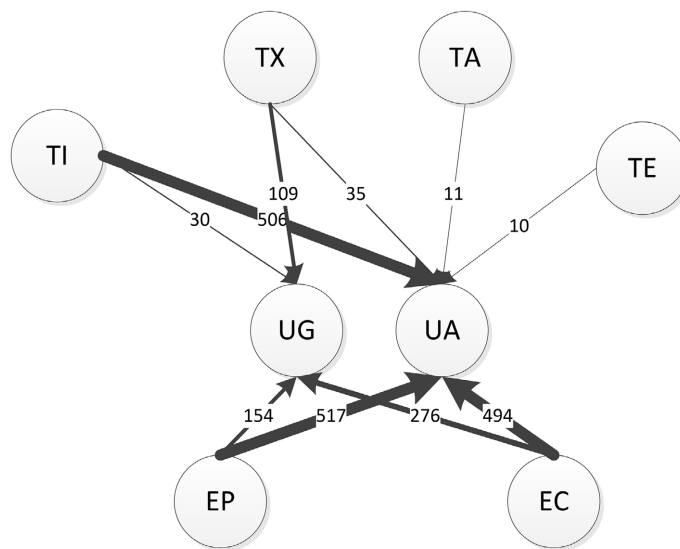


Figure 3. Relationship weight between each node in the fourth quarter (incomplete links)
图 3. 四季度各节点间关系权重(非完整链路)

在空中非期望状态(UA)发生的条件下，内部威胁(TI)发生的概率为 0.885，其他的威胁类因素对空中非期望状态(UA)的影响很小，说明空管系统内部的设备(TI1)条件、工作环境(TI2)因素、工作交接程序(TI3)的合理性、其他管制员(TI4)与正在工作的管制员的配合协调性需要进一步提升。

在空中非期望状态(UA)发生的条件下，程序差错(EP)发生的概率为 0.473，通信差错(EC)发生的概率为 0.523，说明空管系统内部的工作岗位交接(EP1)、信息传递(EP2)、文件(EP3)、检查单(EP4)、最小间隔(EP5)与管制员对外部(EC1)、管制员之间(EC2)的重要程度相当。这两项差错均为以人为触发词，并且计算得到的概率大小相似，所以管制单位在这两项上预防非期望状态的措施要并重。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，本章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。

3.2. 完整链路分析

如图 4 所示为四个季度中所有抽取事件能形成完整链路的节点关系图，即在此图中可以任意组成 T→E→U 的完整链路。

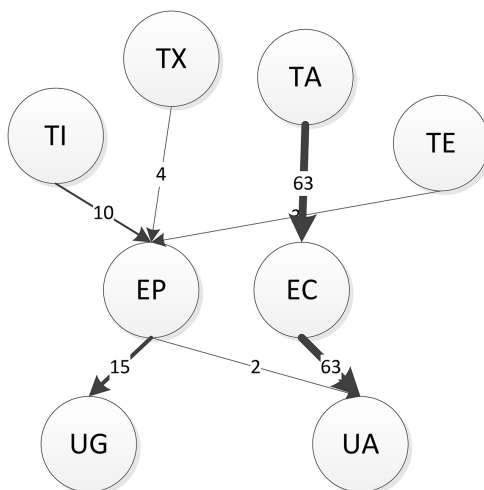
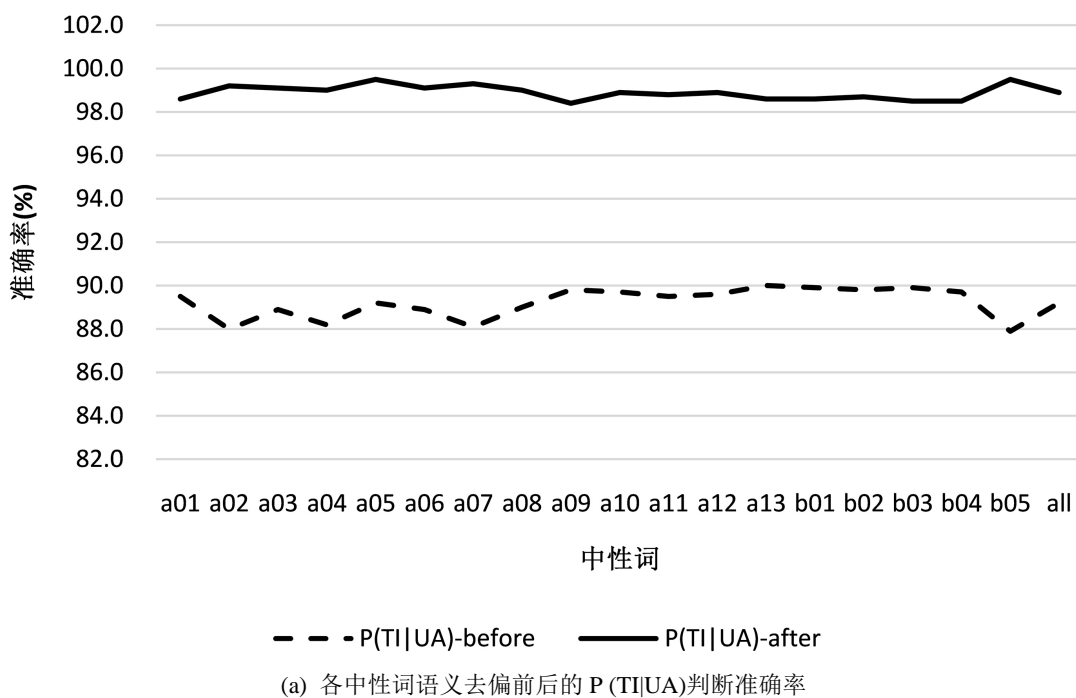


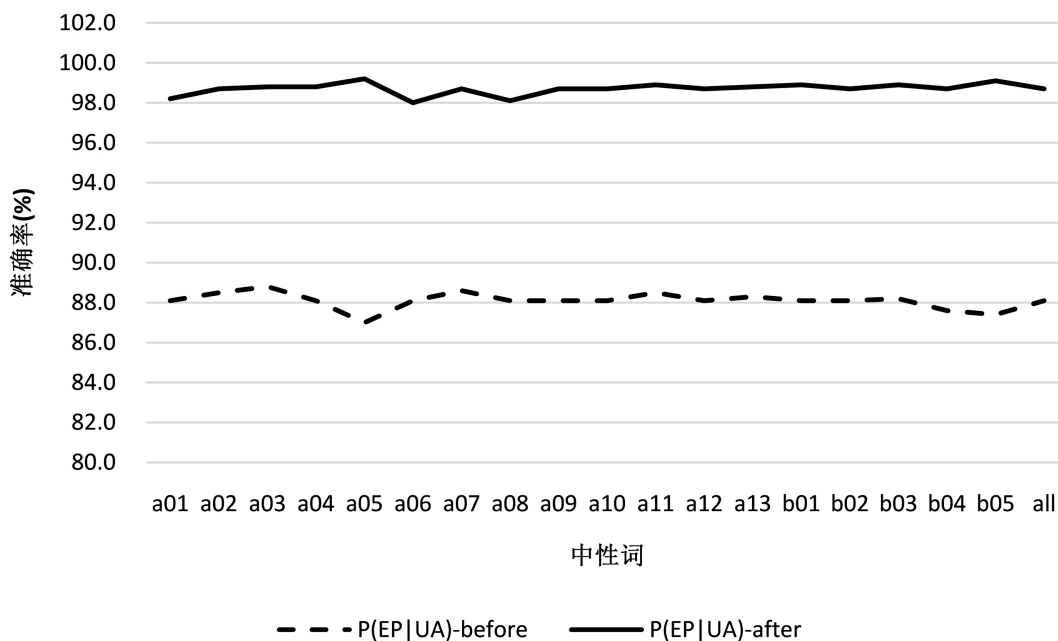
Figure 4. Relationship weight between each node in the fourth quarter (complete links)
图 4. 四个季度各节点间关系权重(完整链路)

T→E→U 的完整链路很少，说明空管系统的各项措施有效，可及时阻断事件演化，在 700 条数据抽取出的 2334 个事件中，像这样完整的链路仅有 80 条，且其中 63 条为 TA→EC→UA，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状态的发生。经更详细的抽取和分析，空中威胁与通信差错中无线电通信均为重要组成部分，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。

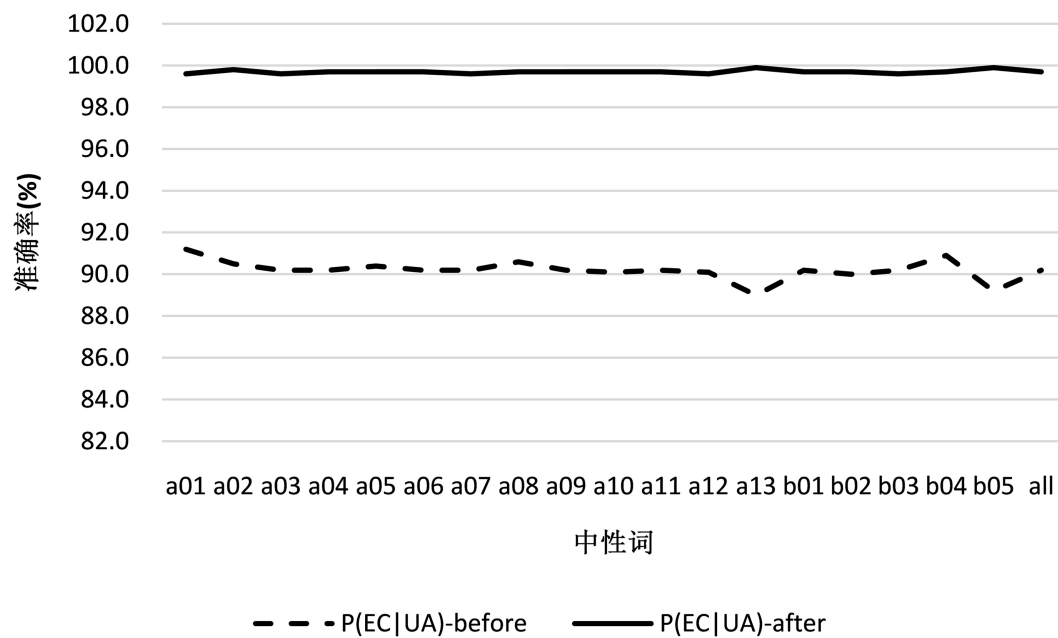
3.3. 相关性准确率分析

经过上节的语义去偏，可以得到更为准确的分主题相关性预测结果。





(b) 各中性词语义去偏前后的 P(EP|UA)判断准确率



(c) 各中性词语义去偏前后的 P(EC|UA)判断准确率

Figure 5. Accuracy rate of correlation judgment before and after semantic debiasing

图 5. 语义去偏前后的相关性判断准确率

如图 5 所示为不同中性词语义去偏前后的相关性。图 5(a)~(c)分别为 P(TI|UA)、P(EP|UA)、P(EC|UA) 语义去偏前后相关性判断准确性。从图中可看出，基于后门调整的主题分析模型在去除不同的中性词时，对相关性的准确率分别有不同程度的提升。

根据观察后验概率和先验概率之间的差异,将所有中性词进行语义去偏后,从先验概率、因果关系、语义相似偏向中获得威胁、错误和不期望状态之间更准确的相关性,四个季度的相关度分析准确率提升结果如表 8 所示。

Table 8. Different correlations for the four quarters

表 8. 四个季度的不同相关性

后验概率	第一季度	第二季度	第三季度	第四季度	均值	准确率	准确率提升
P (TI UA)	0.791	0.792	0.789	0.793	0.791	0.989	10.9%
P (EP UA)	0.418	0.412	0.410	0.420	0.415	0.987	12.0%
P (EC UA)	0.481	0.488	0.470	0.473	0.478	0.997	10.5%

4. 结论

贝叶斯网络的研究结果表明,外部威胁、程序差错、通信差错是造成地面非期望状态的主要因素,内部威胁、程序差错、通信差错是造成空中非期望状态的主要因素,内部威胁、外部威胁、环境威胁是导致程序差错的关键因素,必须立即加以解决。还应采取进一步措施,避免外部威胁和内部威胁。首先对相关分析理论进行简要概述,结合 TEM 模型事件类型和 Doc 9868 中的“威胁和差错管理”,识别出危险源文本中相应的事件,接着采用基于 Deep AutoEncoder 深度自编码器网络的方法生成词向量模型,并且通过改进的余弦相似度计算方法,将空管系统危险源语料通过对文本的相似性映射,根据事件类型划分对映射结果进行图谱表示。然后,识别对相关分析存在影响的中性词语,建立基于后门调整的主题分析模型,剔除主题混杂和语义相似性偏差带来的相关性分析不准确的问题,相关度的准确率各提升了 10.9%、12.0%、10.5%。最后,基于贝叶斯网络计算不同节点威胁、差错和非期望状态的发生概率(即相关度),对管制单位提出合理化建议并得到计算结果中的非完整链路和完整链路的演化分析。

区别于以往的文本信息相关性分析,本文不仅仅基于词向量,而是将语义相似度引入到余弦相似度计算表达式中,使文本的相似性映射更加准确;并且利用预训练语言模型优异的语义表示能力,构建了基于后门调整的主题分析模型,解决了主题混杂和语义相似性偏差带来的相关性分析不准确的问题,不同主题间相关度的准确率有所提升。

基金项目

民航华东空管局科技项目(KJ2101)。

参考文献

- [1] 余雅莉. 民航空管危险源识别及其应用研究[D]: [硕士学位论文]. 南京: 南京航空航天大学, 2018.
- [2] 毛佳静. EoR 技术应用的风险评估[D]: [硕士学位论文]. 德阳: 中国民用航空飞行学院, 2022.
- [3] 李海, 陈勇刚, 刘雨佳. 基于 AHP 的通航运行中危险源危险程度评估方法研究[J]. 西安航空学院学报, 2018, 36(5): 41-44.
- [4] 蒋美芝, 吕靖, 王爽. 基于动态贝叶斯网络的海上通道风险预警[J]. 运筹与管理, 2023, 32(10): 63-68.
- [5] Raskin, V., Taylor, J.M. and Hempelmann, C.F. (2013) Meaning- and Ontology-Based Technologies for High-Precision Language and Information-Processing Computational Systems. *Advanced Engineering Informatics*, 27, 4-12. <https://doi.org/10.1016/j.aei.2012.12.002>
- [6] 王洁宁, 杨海滨. 面向空管安全系统中人为因素的本体建模分析[J]. 中国民航大学学报, 2009, 27(4): 33-36.
- [7] 刘继新, 沈丽楠. 空管安全系统中人为因素的本体建模研究[J]. 武汉理工大学学报(信息与管理工程版), 2011, 33(4): 656-659.

- [8] Zhao, Y., Hua, S. and Ren, X. (2017) Relevance Research of Threat/Error and Undesired States in Air Traffic Management Based on Bayesian Network Model. *Journal of Air Transport Management*, **60**, 45-48. <https://doi.org/10.1016/j.jairtraman.2017.01.001>
- [9] 徐一旻, 田梦莹, 李治, 等. FTA-BN 在机场跑道入侵事故影响因素分析中的应用[J]. *安全与环境学报*, 2023, 23(5): 1361-1367.
- [10] Landeiro, V., Tran, T. and Culotta, A. (2019) Discovering and Controlling for Latent Confounds in Text Classification Using Adversarial Domain Adaptation. In: *Proceedings of the 2019 SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics, 298-305. <https://doi.org/10.1137/1.9781611975673.34>