同济大学
TONGJI UNIVERSITY

毕业设计（论文）

课题名称：基于文本挖掘的高铁列车晚点影响因素分析

副 标 题：

学　　院：交通运输工程学院

专　　业：交通工程

学　　号：1852127

姓　　名：赵冠华

指导教师：邢莹莹

日　　期：2022 年 06 月 01 日

同济大学本科毕业设计（论文）原创性声明

本人郑重声明：所呈交的毕业设计（论文），是本人在导师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本毕业设计（论文）的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本毕业设计（论文）所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本毕业设计（论文）原创性声明的法律责任由本人承担，并承诺已遵从《同济大学交通运输工程学院本科毕业设计（论文）检测及处理办法（试行）》提交论文检测报告，文字重合比符合答辩要求，检测报告真实有效。

毕业设计（论文）作者签名：赵冠华

2022 年 5 月 26 日

基于文本挖掘的高铁列车晚点影响因素分析

装 订 线

摘 要

正点率作为高铁服务水平的重要评价指标，受到铁路运营管理部门的高度重视。然而，受到设备故障、突发事故、自然环境、人为失误等不确定性因素的干扰，高铁列车晚点故障时有发生，严重影响旅客的行程安排，降低高铁列车的服务质量。因此，开展高铁列车晚点影响因素的研究十分必要。

高铁故障记录是获得高铁列车晚点影响因素的数据来源，主要以自然语言的形式记录。这种记录形式具有不规范性和随意性的特点，导致高铁列车晚点故障征兆中存在大量的语义不明确的现象，使得从高铁故障记录中提取高铁列车晚点的影响因素成为亟待解决的问题。本文基于 2016 年至 2019 年 3 月高铁故障记录数据，以卡方值表示相关性知识，提取晚点词项与晚点致因模式的主题-词分布；然后，建立基于先验相关性知识的 LDA 主题模型，完成高铁晚点故障数据的语义级特征提取；最后，构建高铁列车晚点模糊事故树模型，通过定量与定性分析得到高铁晚点影响因素对顶上事件的影响程度，探明高铁晚点的关键影响因素，并提出降低高铁晚点率的有效措施与治理优先级。本文的研究成果可为高铁列车的正点高效运行和列车调度指挥提供理论支撑，有利于提高乘客满意度与高铁服务水平，促进高速铁路的健康可持续发展。

关键词：列车晚点，文本挖掘，LDA 主题模型，模糊事故树

Analysis of influencing factors of high speed railway train delay based on Text Mining

ABSTRACT

As an important evaluation indicator of high-speed railway service level, punctuality rate is attached with great importance by railway operation management department. However, due to uncertain factors such as equipment failure, accidents, natural environment and human error, delays of high-speed railway break out from time to time, which seriously affects the schedule of passengers and lowers the quality of high-speed railway's services.

The existing high-speed railway fault record, which is recorded in the form of natural language, is the data source to obtain the key factors of train delay. The form of recording exists great liberalization and substandard, leading to a large number of semantic ambiguities in the symptoms of high-speed train delay and thus making it an urgent problem to extract the key factors of high-speed train delay from high-speed train fault records. Based on the record data of high-speed railway from 2016 to March 2019, this paper uses chi-square measure to represent the relevancy and extracts the topic-word distribution of delayed word items and delayed causes; after that, the LDA topic model based on prior knowledge is established to extract the semantic level features; finally, the fuzzy fault tree model of high-speed railway train delay is constructed. Through quantitative and qualitative analysis, the impact that high-speed railway delay factors have on top events is obtained, the key factors affecting the delay are found, and the effective measures to reduce the high-speed railway delay rate and governance priorities are proposed. The result of this work will supply the theoretic support for the on-time and efficient operation of high-speed railway trains and train dispatching command, improve passenger satisfaction and high-speed railway service level, and promote the healthy and sustainable development of high-speed railway.

Key words： Train Delay, Text Mining, LDA Topic Model, Fuzzy Fault Tree

目 录

1 引 言 1
1.1 研究背景及意义 1
1.1.1 研究背景 1
1.1.2 研究意义 2
1.2 国内外研究现状 2
1.2.1 列车晚点研究现状 2
1.2.2 文本挖掘算法研究现状 3
1.2.3 研究现状总结 4
1.3 研究内容与技术路线 4
1.3.1 研究内容 4
1.3.2 技术路线 4
1.4 创新点 6
1.5 本章小结 6
2 相关理论与方法 7
2.1 隐含狄利克雷分布 7
2.1.1 产生式模型 7
2.1.2 吉布斯采样 7
2.1.3 最佳主题数目的确定方法 8
2.2 模糊事故树 8
2.2.1 事故树基本理论 8
2.2.2 定性与定量分析 10
2.2.3 三角模糊数 11
2.3 本章小结 12
3 基于先验 LDA 的高铁列车晚点影响因素特征提取 13
3.1 文本数据简介与预处理 13
3.1.1 数据简介 13
3.1.2 分词 14
3.2 基于相关性的先验知识提取 16
3.2.1 卡方值矩阵 16
3.2.2 先验知识提取 17
3.3 先验知识与 LDA 模型的整合 19
3.3.1 主题更新公式 19
3.3.2 算法流程 20
3.3.3 建模结果可视化 22
3.4 本章小结 25
4 基于模糊事故树的高铁晚点影响因素分析与应对措施 26
4.1 模糊事故树构建 26
4.1.1 晚点致因分析 26
4.1.2 建立事故树 27
4.2 基本事件三角模糊处理 29
4.2.1 可统计的基本事件 29
4.2.2 不可统计的基本事件 30
4.2.3 基本事件汇总 31
4.3 定性分析 32
4.3.1 最小割集 32

4.3.2 最小径集 33
4.3.3 结构重要度 33
4.4 定量分析 34
4.4.1 顶上事件模糊概率 34
4.4.2 模糊概率重要度 34
4.4.3 临界重要度 37
4.5 降低晚点率的有效措施 39
4.6 本章小结 40
5 结论和展望 41
5.1 论文工作 41
5.2 不足与展望 42
参考文献 43
谢 辞 45

装 订 线

1 引言

1.1 研究背景及意义

1.1.1 研究背景

截止 2021 年 12 月 30 日，我国高速铁路运营里程超过 4 万公里[1]，“四纵四横”高铁网全面建成，“八横八纵”高铁网正在加密形成，高铁已经覆盖了全国 92%的 50 万人口以上的城市，中国成为全球唯一高铁成网运行的国家。除此之外，在“一带一路”的背景之下，在“走出去”的战略指导下，中国高铁作为高铁市场中的“后起之秀”，已经成为祖国自主创新的名片。

高速铁路之所以如此令人喜爱并受国家的重视，离不开其运载量大、价格较低、安全、快速、便捷以及区间带动的特点。尽管铁路与航空都是长途运输的主要交通方式，但是实际上航空旅客的运输量远低于铁路运输量，在部分与高铁线路重叠的航班上，已经出现高铁票价高于航班票价的现象，如表 1.1 所示[2]。高时速的高铁缩小了二者之间的时速差距，同时由于高铁受天气影响小、准点率高，因此很多旅客更愿意选择高铁出行。

表 1.1 广州-长沙路段高铁、航空日运营情况对比表（20190812）[2]

交通工具	班次（每日最大值）	定员数	日运输量（人次）	日运输量占比	票价（元）	速度
铁路	163	556	90628	99.18%	98 元（火车）；高铁二等、一等、商务舱票价分别为 314 元、504 元、995 元	8~10 小时（火车）；2 小时 40 分钟（高铁）
航空	4	150 或 295	745	0.82%	票价不定：远期：240 元，270 元，290 元；近期：880 元，950 元，1280 元	1 小时 20 分钟

然而，高铁列车在运行时不可避免的会受到其他因素的干扰，导致高铁晚点时有发生。一旦高铁列车发生晚点，不仅会导致车站的客流滞留，还可能会导致乘客被困车厢内，降低旅客对高铁的信任度。同时鉴于列车与线路之间复杂的耦合关系，当出现极端天气或特殊状况时，往往会导致列车晚点的链式传播，甚至影响到整个铁路网，造成一定的经济损失，带来负面的社会影响。因此，降低高铁列车晚点率是提高高铁服务水平的关键。

在这样的背景下应当高度重视高铁列车晚点问题。目前关于高铁列车晚点问题的研究多侧重于晚点延误的传播和晚点后的恢复，并没有太多关于从根本上减少高铁列车晚点现象发生的研究。所以，有必要对高铁列车晚点影响因素进行分析，掌握高铁列车晚点影响因素之间的关联，提出降低高铁列车晚点率的有效措施，以促进高速铁路健康可持续发展。

1.1.2 研究意义

当前，高速铁路故障写实表与处置过程均以自然语言的形式记录，而该形式具有不规范性和随意性的特点，这使得文本中存在大量的语义不明确的现象。因此，如何从大量高铁故障记录数据中提取高铁列车晚点的影响因素是亟需解决的问题之一。本课题基于 2016 年-2019 年高铁列车故障记录数据，采用文本挖掘技术来建立故障特征和列车晚点之间的联系，探明高铁列车晚点的影响因素，提出降低高铁列车晚点率的有效措施。本文理论意义如下：

（1）本文提出基于先验 LDA（Latent Dirichlet Allocation）模型进行高铁列车晚点的语义级特征提取。在分析导致晚点产生的列车故障文本数据的基础上，得到晚点因素词库、最优主题个数以及各类词项之间的相关性，将 LDA 模型与先验知识整合，以此来实现特征提取。先验 LDA 模型作为一种半监督模型，在一定程度上克服了 LDA 模型在文本挖掘过程中的盲目性，为故障特征和列车晚点之间联系的探究提供了一种新的方案。

（2）本文对高铁列车故障数据中找出的高铁晚点影响因素进行模糊事故树分析，完成定性与定量分析，以此为依据提出针对性的改善措施，充实了高铁晚点问题的研究内容，丰富了其在高速铁路研究中的具体实践。

同时，本文的现实意义如下：

本文通过对高铁列车晚点影响因素的分析，为决策者更加全面的了解高铁晚点问题的实际情况提供现实、客观和公正的信息；通过分析掌握了各类影响因素的影响程度，有助于铁路局在高速列车调度指挥过程中对各种异常情况的影响进行宏观把控；通过分析提出降低高铁晚点率的有效措施，有助于调整运行图，改善运营组织环境，提高乘客的满意度，提高高铁服务水平。

1.2 国内外研究现状

本文将从列车晚点影响因素与文本挖掘算法两个方面进行国内外研究的综述，以为本文基于文本挖掘的高铁列车晚点影响因素分析提供理论支撑。

1.2.1 列车晚点研究现状

国际铁路联盟[3]曾在标准 450-2《铁路间相互合作交换国际客、货列车行车的统计和分析数据》中列出了 50 个列车晚点的可能原因。

Preton[4]参考了 2006-2007 英国的国家铁路网，数据表明 40%的列车延误可能与列车运营商（如列车故障和乘务员短缺）、基础设施管理部门（如轨道和信号设备故障）以及外部因素（如极端天气、自杀）有关，剩余 60%的延误则由延误传播导致。

Ghaemi[5]认为列车故障、信号故障等中断是导致乘客延误的重要原因，并提出了基于贝叶斯网络的中断长度模型、基于 MILP 的短转弯模型及乘客分配模型三种模型组成的独特的迭代方法以获得可靠的中断长度，降低中断的负面影响。

Wang[6]等人调查了冬季气候（包括冰/雪/降雨）对 Botnia-Atlantica 地区高速客运列车性能的影响，并通过 Cox 模型研究其主要延迟与到达延迟，结果表明温度与湿度对主要延迟的发生和到达延迟与非延迟之间的过渡强度都有显著的影响。

翟恭娟[7]在对广珠城际轨道交通主线进行数据分析后，认为导致列车晚点的主要因素有六类：

自然灾害（如地震、水害、泥石流）、天气异常（如大雾、雷雨、暴风雪）、突发事件（如站车发生火灾、爆炸）、设备故障（包括固定设备与动车组车辆故障）、作业延误（如固定设备施工延长而占用了列车运行的正常时间）以及晚点传播，并对这六种影响因素的可预测性、可避免性、影响程度进行了分类汇总。

汪静等[8]通过对专家进行问卷发放的方式对高铁列车晚点影响因素之间的相互关系进行调查，并结合相关文献与数据建立解释结构模型，总结出列车晚点的 14 个直接因素（如高铁列车故障、各类设备故障）、间接因素（如碰撞异物）与根源因素（如规章制度不完善），由此进行了高铁列车晚点的风险应对探讨。

纪媛媛[9]使用 TF-IDF 方法对武广高铁线下行线路的晚点记录进行关键词提取，在分析整理后得到包括无线超时制动、地面设备故障、安全软件故障、大风、大雨、大雪与异物侵限等在内的 18 个影响因素，对晚点特征进行数值化处理，以方便列车晚点预测。

石睿[10]结合武广高铁历年的非正常时间数据将列车晚点致因分为非正常事件因素与列车间相互作用因素，他认为前者是导致晚点的直接原因（包括环境因素、系统性因素与人为因素），后者是晚点传播的重要原因。

装

1.2.2 文本挖掘算法研究现状

文本挖掘指抽取有效、新颖、有用、可理解的、散布在文本文件中的有价值知识，并且利用这些知识更好地组织信息的过程。由于写作文本的词汇表是非常巨大的，但是一个给定的文档可能只有几百个词，因此常常使用聚类的方法进行文本挖掘。

订

文本聚类算法可以分为层次聚类算法、K 均值聚类算法以及概率聚类算法[11]。由于大量的高铁文本数据均以自然语言的形式记录，具有一定的不规范性和随意性，而通常情况下的特征选择方法仅仅是对词进行提取，没有考虑特征词之间的关联信息（例如同义词现象），因此常选用概率聚类算法中的主题模型完成文本挖掘任务。

线

1999 年，Hofmann[12]提出概率潜在语义分析（pLSA），在隐性语义索引（LSA）的基础上使用概率图模型，引入了一个隐变量主题，同时解决了语义分析问题中的同义词与一词多义现象，被认为是第一个主题模型。

然而，pLSA 只是对已有文档的建模，其参数量随着文档数量线性增长，容易出现过拟合的现象。2003 年，Blei 等人[13]提出隐含狄利克雷分布 （LDA），将 pLSA 固定在贝叶斯框架之下，使参数满足狄利克雷分布，主题模型得到了进一步优化。

目前，LDA 已被广泛应用于各种领域，如文档建模[14]与上下文感知推荐[15]等；也创建出许多 LDA 的变种。然而，基础的 LDA 模型是无监督无层次结构。在这一基础上， Blei 等人于 2004 年[16]提出无监督有层次的 hLDA 模型，又在 2007 年[17]提出了有监督无层次结构的 sLDA 模型，同时，Petinot 等人[18]还于 2011 年提出了有监督有层次结构的 hLLD 模型。

除此之外，毛先领[19]于 2012 年提出 SSHLDA 模型，将有标签的主题融合于文本生成过程之中，以获得主题的层次结构。王峰[20]于 2016 年提出的整合先验知识的 LDA 模型用于车载设备的故障特征提取。这两者都属于半监督模型，一定程度上克服了属于无监督学习的 LDA 模型的盲目性，同时以较小的标签量完成了大量未标记数据的分类任务。

1.2.3 研究现状总结

总的来说，截止目前国内外学者对列车晚点影响因素取得了一定的成果，但这些研究也存在着一些不足，主要体现在：国内外对列车晚点的研究大多注重于导致晚点产生的单一原因，以及如何降低这些原因对列车晚点的影响；除此之外还注重于各类晚点致因的可预测性、可避免性以及影响程度，并以此进行列车晚点预测，而对高铁列车晚点影响因素之间关系的分析相对较少。

鉴于目前文本挖掘算法的发展，高铁文本数据所具有的不规范性与随意性，以及本课题所依靠的 2016 年至 2019 年高铁列车故障记录的数据规模，为提升聚类的准确度，本文决定采用半监督的 LDA 模型进行高速铁路列车晚点影响因素的研究。

1.3 研究内容与技术路线

1.3.1 研究内容

当前，高速铁路列车快速发展，在运输系统中占据越来越重要的作用。而高铁列车晚点率作为高铁服务水平的重要评价指标，降低晚点率十分关键。高速铁路故障写实表与处置过程记录了运输过程中出现的各类意外事故以及具体处置方案，为知识挖掘提供了一定的数据基础，对探明晚点致因之间的相互关系、降低高铁列车的晚点率具有重要参考意义。

然而，高速铁路故障写实表与处置过程均以自然语言的形式记录，文本中存在大量的语义不明确的现象，如何有效地从中提取晚点致因成为亟需解决的一个问题。本课题面向这一需求，以 2016 年至 2019 年高铁列车故障记录数据为基础，拟探究故障特征与高铁列车晚点之间的联系，拟研究各类晚点致因间的内在关联，提出降低高铁列车晚点率的有效措施。论文的主要内容如下：

（1）对高速铁路故障写实表与处置过程文本描述进行预处理，筛选出造成高速铁路列车晚点的记录。综合考虑国内外现有研究结果以及国内高速铁路实际运营情况提取先验知识。最后，基于先验知识和写实数据建立半监督的 LDA 模型，实现晚点致因的语义级特征提取。

（2）对晚点致因进行模糊事故树建模，运用逻辑关系辨识与评价高速铁路列车晚点影响因素中各基本事件的危险性，确定各因素的相互影响关系，并通过定性与定量分析总结出高铁列车晚点的直接原因与潜在原因。

（3）针对以上研究成果提出降低高铁列车晚点率的有效措施与治理风险因素的优先级。

1.3.2 技术路线

本文的主要内容包括基于先验 LDA 模型的高铁列车晚点影响因素特征提取和基于模糊事故树的高铁晚点影响因素分析两个部分。技术路线如图 1.1 所示。

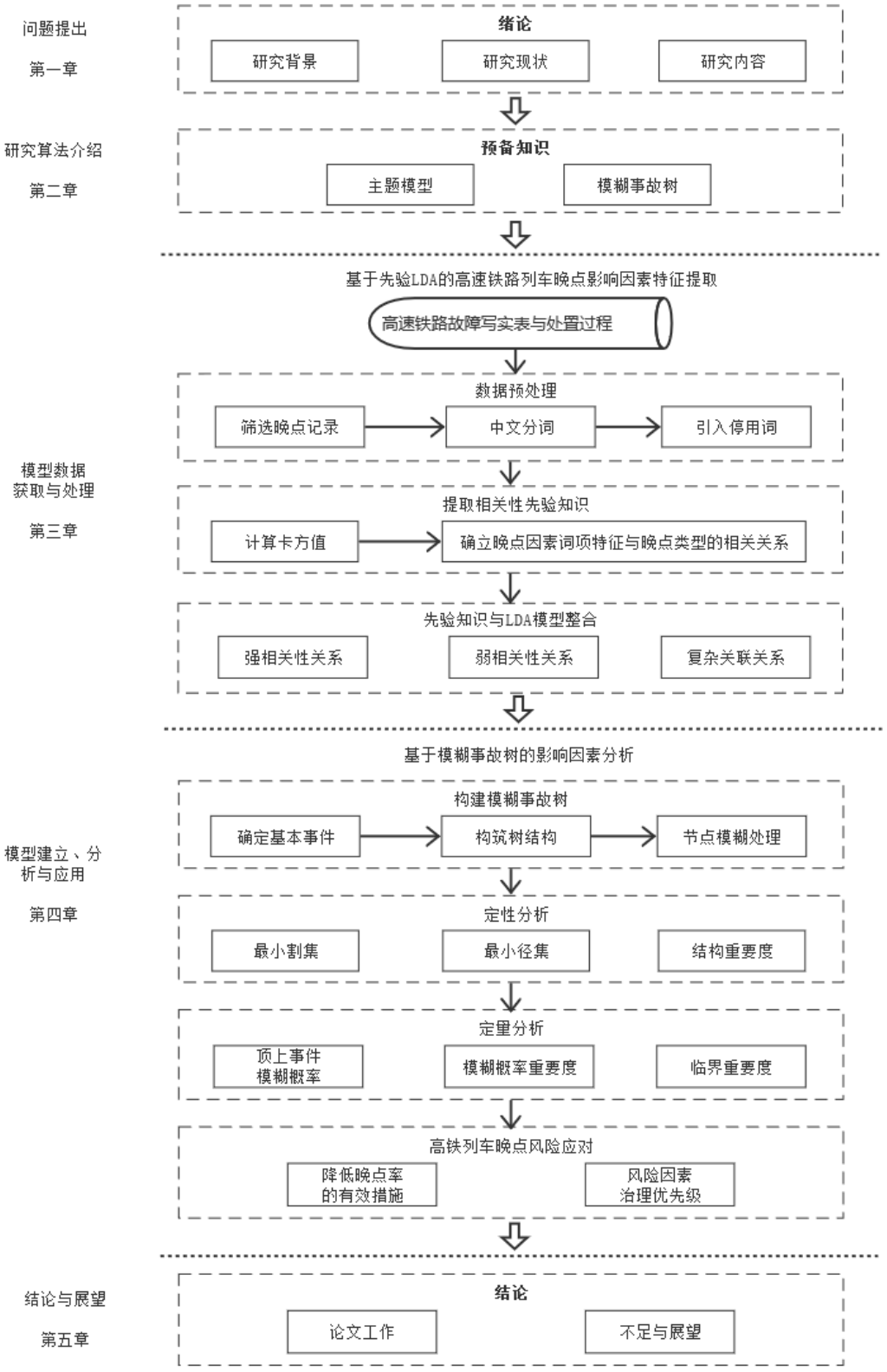
问题提出
第一章
绪论
研究背景
研究现状
研究内容
研究算法介绍
第二章
预备知识
主题模型
模糊事故树
基于先验LDA的高速铁路列车晚点影响因素特征提取
高速铁路故障写实表与处置过程
数据预处理
筛选晚点记录
中文分词
引入停用词
模型数据
获取与处理
第三章
提取相关性先验知识
计算卡方值
确立晚点因素词项特征与晚点类型的相关关系
先验知识与LDA模型整合
强相关性关系
弱相关性关系
复杂关联关系
基于模糊事故树的影响因素分析
构建模糊事故树
确定基本事件
构筑树结构
节点模糊处理
模型建立、分
析与应用
第四章
定性分析
最小割集
最小径集
结构重要度
定量分析
顶上事件
模糊概率
模糊概率重要度
临界重要度
高铁列车晚点风险应对
降低晚点率
的有效措施
风险因素
治理优先级
结论与展望
第五章
结论
论文工作
不足与展望

图 1.1 技术路线

1.4 创新点

本篇论文的创新点如下：

（1）提出基于先验 LDA 模型进行高铁列车晚点的语义级特征提取。在分析导致晚点产生的列车故障文本数据的基础上，得到晚点因素词库、最优主题个数以及各类词项之间的相关性，将 LDA 模型与先验知识整合，以此来实现特征提取。

（2）对高铁列车故障数据中找出的高铁晚点影响因素进行模糊事故树分析，得到高铁晚点影响因素危险性，为风险治理措施提供依据。

1.5 本章小结

本章介绍了论文的研究背景与论题，对论题所涉及的领域进行了国内外研究综述，指明本论文的研究内容、技术路线以及研究意义与创新点。

2 相关理论与方法

2.1 隐含狄利克雷分布

LDA 模型是自然语言处理中十分常用的一个主题模型，常常应用于文本主体识别、文本分类与文本相似度计算等方面。本节将介绍产生式模型、吉布斯采样、算法流程以及最佳主题个数的判断方法。

2.1.1 产生式模型

LDA 属于产生式模型，它认为主题可以由词汇分布来表示，文章可以用主题分布表示。在原始的论文中，文档的生成是这样描述的：

（1）从一个全局的泊松参数为β的分布中生成一个文档的长度N。

（2）从一个全局的狄利克雷参数为α的分布中生成一个当前文档i的主题θ_m。

（3）对当前文档长度N的每一个字执行：①从主题的多项式分布θ_m中取样生成文档i第j个词的主题$z_{m,n}$；②根据已分配的主题$z_{m,n}$，得到对应的词语分布$\varphi_{z_{m,n}}$；以$\varphi_{z_{m,n}}$和 $z_{m,n}$共同为参数的多项分布中产生一个字$w_{m,n}$。

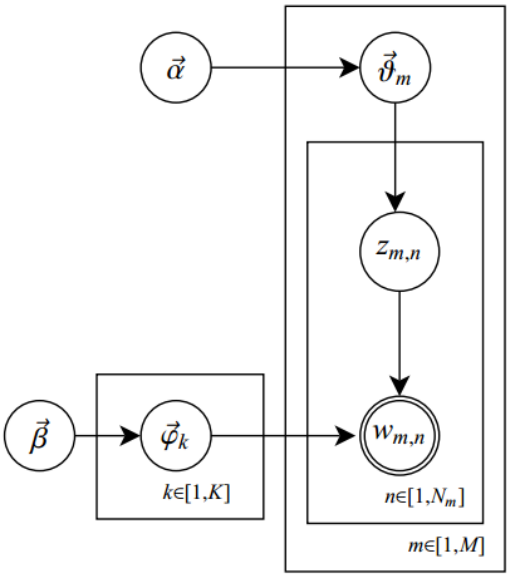

图 2.1 LDA 模型的图模型结构

因此，该过程可以理解为：先以一定的概率选取某个主题，然后再以一定的概率选取该主题下的某个词，不断重复这两步，直到完成整个文档。

2.1.2 吉布斯采样

LDA 的生成模型需要从狄利克雷先验中随机抽取主题分布，那么假定文档已经产生，为了获得文档的主题参数，常采用基于贝叶斯估计的吉布斯采样法获取未知的分布参数θ、φ。

吉布斯采样的步骤是：初始随机给每个文档的每个词赋予一个主题，然后根据式（2.1）计算

对每个词w_i的主题概率$P(z_i = j \mid z_{-i}, w, \alpha, \beta)$，根据主题概率采样一个新主题赋予该词，然后同样方法更新下个词直至收敛。式中，α控制一个文档中的主题数量，α越大，文档中包含的主题越多；β控制每个主题中词的数量，β越大，主题中包含的词越多。

$$P(z_i = j \mid z_{-i}, w, \alpha, \beta) \propto \left(\frac{n_{-i,j}^{w_i} + \beta}{\sum_w^W n_{-i,j}^{w_i} + W\beta}\right)\left(\frac{n_{-i,j}^{d_i} + \alpha}{\sum_j^T n_{-i,j}^{d_i} + T\alpha}\right) \tag{2.1}$$

2.1.3 最佳主题数目的确定方法

LDA 模型作为无监督贝叶斯模型，只需要输入文档和主题数量 K 即可完成训练，模型的效果与主题数量 K 相关。因此，超参数 K 的选择成为主题模型训练前的重中之重。

传统的最优主题个数判断方法有五种：

① 基于经验：需要人类主观判断、不断调试，操作性强且最为常用；但是引入人工成本极高，同时评价过程具有模糊性和主观性；

② 基于困惑度：在 LDA 模型原始论文中，Blei[13]通过计算模型的困惑度大小判断不同主题个数 K 下的模型效果，认为困惑度越小，模型效果越好；然而不少研究表明基于困惑度选择的主题语义在很多场景下与人工的判别有一定差距，主题辨识度不高；

③ 基于贝叶斯统计标准方法：Griffiths[21]使用 Log-边际似然函数确定最优主题个数；然而该方法计算复杂度高，无法体现模型的泛化能力；

④ 基于非参数方法：Teh[22]提出了基于狄利克雷过程的非参数贝叶斯模型 HDP，该方法可以自动从数据中选择最优主题个数 K；然而该方法需要同时建立两个模型，较为复杂；

⑤ 基于主题相似度：曹娟[23]与 Arun[24]分别通过计算主题之间的余弦距离、KL 距离计算主题结构的平均相似度，当平均相似度最小时，对应的模型最优。

除去以上这五种方法，最近几年，通过计算主题连贯性选择最优主题个数的研究越来越多。如果说困惑度是评价模型预测准确性的指标，那么连贯性就是评价主题质量的指标。2010 年，David[25]提出 C_uci 方法自动评估主题连贯性，基于滑动窗口，对给定主题词中的所有单词对的点态互信息进行计算；2011 年，Mimno[26]则提出 C_umass 方法优化语义一致性，基于文档并发计数，利用 one-preceding 分割和对数条件概率计算连贯度；2015 年，Roder[27]在对比以上内容的基础上探索主题连贯性度量的空间，提出 C_v 方法对主题词进行 one-set 分割，并使用归一化点态互信息和余弦相似度间接获得连贯度。

2.2 模糊事故树

事故树模型能对系统的危险性进行评价，在分析得到事故直接原因的基础上，还可以深入探查事故的潜在原因，同时完成定性与定量分析，由此依据事故发生的各个途径提出针对性的改善措施。

2.2.1 事故树基本理论

A. 术语与符号

事故树中每一个节点都代表一个事件，包含顶上事件、中间事件与基本事件。顶上事件指事

故树分析中所关心的结果事件，位于事故树顶端，符号如图 2.2（a）所示；基本事件指导致顶上事件发生的最基本的或不能再向下分析的原因，符号如图 2.2（b）所示；中间事件指位于顶上事件与基本事件之间的结果事件，符号如图 2.2（c）所示。

事故树各事件由逻辑门连接，主要是与门与或门，可以连接多个输入事件和一个输出事件。与门连表示当所有输入事件均发生时，输出事件才发生，符号如图 2.2（d）所示；或门表示至少一个输入事件发生时，输出事件就发生。

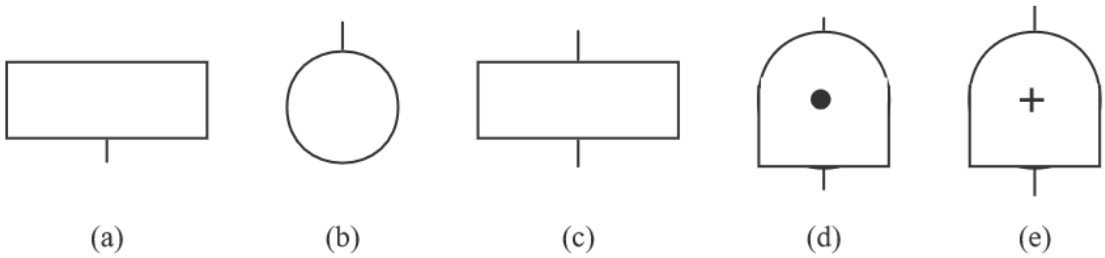

图 2.2 事故树常用符号

B. 运算法则

由于事故树各基本事件是由逻辑门连接的，因此其运算满足布尔逻辑运算法则，即满足表 2.1。

表 2.1 布尔逻辑运算的基本法则

逻辑关系	运算公式
逻辑和运算	$A+B=B+A$
	$A+(B+C)=(A+B)+C$
	$A+A+\cdots+A=A$
	$A+1=1$
	$A+0=0$
逻辑积运算	$AB=BA$
	$A(BC)=(AB)C$
	$A\cdot A\cdot \ldots\cdot A=A$
	$A\cdot 1=A$
	$A\cdot 0=0$
对偶律	$(A+B+\cdots+K)'=A'+B'+\cdots+K'$
	$(A\cdot B\cdot \ldots\cdot K)'=A'\cdot B'\cdot \ldots\cdot K'$

续表 2.1

逻辑关系	运算公式
吸收律	$A+A'B=A+B$
	$A+AB=A$
	$A(A+B)=A$
分配律	$AB+AC+BC=AB=AC$
	$A(B+C)=AB+AC$
	$A+BC=(A+B)(B+C)$
对合律	$AB+A'B=A$
	$(A+B)(A+B')=A$

2.2.2 定性与定量分析

A. 定性分析

通过对事故树的定性分析，可以弄清系统（或设备）出现顶事件的可能性，该步骤包含对最小割集、最小径集与结构重要度的求解。

a. 最小割集

割集指可以导致顶事件发生的基本事件的集合，最小割集指若去掉该集合中任意一个基本事件则该集合不再是割集的集合。最小割集常用布尔代数法求解，步骤为：先建立事故的布尔代数表达式；再根据布尔逻辑运算的基本法则化简。

最小割集可以表示系统的危险性，最小割集越多，系统危险性越大；它的布尔代数表达式表明了顶上事件发生的原因组合，方便人类掌握事故的发生规律。

b. 最小径集

径集指使顶事件不发生的基本事件的集合，最小径集指若去掉该集合中任意一个基本事件则该集合不再是径集的集合。最小径集常用对偶树法求解，步骤为：先根据对偶原理将事故树转化为成功树；成功树的最小割集就是事故树的最小径集。

最小径集可以表示系统的安全性，最小径集越少，系统越安全；同时，可以根据最小径集中包含基本事件的个数、技术的难易程度与耗费的精力、物力、财力确定最经济有效的控制事故的方案。

c. 结构重要度

结构重要度是在不考虑其发生概率大小的情况下，单单通过观察事故树结构获得的基本事件对顶上事件的影响程度。结构重要度可以通过计算结构重要度系数准确排序，但当事故树较为复杂时，该方法计算量较大；因此大多情况下，可以通过最小径集根据式 2.2 求得近似值。

$$I(i)=\frac{1}{k}\sum_{j=1}^{n}\frac{1}{n_j}(j\in k_j) \tag{2.2}$$

k为最小径集总数，k_j为第 j 个最小径集，n_j为第 j 个最小径集包含的基本事件总数。

B. 定量分析

通过对事故树的定量分析，可以求得顶上事件的发生概率、概率重要度与临界重要度。

a. 顶上事件发生概率

顶上事件的发生概率有多种方法可以计算。在事故树规模不大、又没有重复的基本事件时，可以从底部逻辑门连接的事件算起，逐渐向上，直至得到顶上事件的发生概率。除此之外，还可以根据式（2.3）通过最小径集逼近法近似计算，该方法适用于径集数目较少的情况。

$$P(T)=1-\sum_{r=1}^{k}\prod_{x_i\in P_r}(1-q_i)+\sum_{1\le r<s\le k}\prod_{x_i\in P_r\cup P_s}(1-q_i)-\cdots+(-1)^{k-1}\prod_{\substack{r=1\\ x_i\in P_1\cup P_2\ldots\cup P_k}}^{k}(1-q_i) \quad (2.3)$$

式（2.3）中，P_r为最小径集$(r=1,2,\ldots\ldots k)$； r、s为最小径集的序数，且$r<s$；k为最小径集数；$x_i\in P_r$指属于第r个最小径集的第i个基本事件；$x_i\in P_r\cup P_s$指属于第r个或第s个最小径集的第i个基本事件。

通过求解顶上事件发生概率可以估算系统的可靠性，同时与实际生产生活统计数据对比，验证事故树模型在该顶上事件分析的适用性。

b. 概率重要度

概率重要度与结构重要度一样反映了基本事件对顶上事件发生的影响程度，但与结构重要度不同，它定义为基本事件概率变化引起顶事件概率变化的程度。根据定义，结构重要度可以根据式（2.4）求解。

$$Q(i)=\frac{\partial P(T)}{\partial q_i} \quad (2.4)$$

通过求解概率重要度可以了解到减少哪个基本事件的发生概率就可有效地降低顶上事件的而发生概率，即得到减少事故发生的有效措施。

c. 临界重要度

当各基本事件的发生概率不相等时，一般情况下，改变概率大的基本事件比改变概率小的基本事件容易，而基本事件的概率重要度无法反映这一事实。因此，引入了临界重要度再次描述基本事件对顶上事件发生概率的影响，根据定义，临界重要度可以通过式（2.5）求解。

$$C_i=\frac{\partial lnP(T)}{\partial lnP(q_i)}=\frac{P(q_i)}{P(T)}Q(i) \quad (2.5)$$

通过求解基本事件的临界重要度可以得到治理的优先级顺序。

2.2.3 三角模糊数

模糊事故树是指通过将所有节点模糊化，以模糊数表示基本事件的发生概率，从而降低数据缺失与低可靠性造成的数据误差的改进事故树模型。它基于模糊集理论，一定程度上克服了很难得到大量、准确、可靠的基本事件的发生概率的问题。

三角模糊数与梯形模糊数常应用于事故树分析，本文主要采用前者对各节点模糊化处理。

首先，应当了解三角模糊数的隶属函数。对于三角模糊数$\tilde{A}\in\xi(X)$，其隶属函数$\mu_A(x)\in[0,1]$定义为：

$$\mu_A(x)=\begin{cases}0, x<\alpha\\ \dfrac{x-\alpha}{m-\alpha}, \alpha\le x<m\\ \dfrac{\beta-x}{\beta-m}, m\le x<\beta\\ 0, x>\beta\end{cases} \tag{2.6}$$

式（2.6）中，α、β分别对应方案评价中三角模糊数的下界与上界。

2.3 本章小结

本章主要对主题模型以及模糊事故树的相关理论和指标的计算方法进行简单介绍，同时给出主题模型的算法流程和模糊事故树的分析流程，为第三章、第四章的数据提取和模型建立与分析提供基础知识。

3 基于先验 LDA 的高铁列车晚点影响因素特征提取

为了建立适合度更高的事故树，本章将通过基于先验知识的主题模型完成数据集的语义特征提取工作。文本特征指最能概括文本主旨的词的集合[28]。通过对其进行提取，可以提高信息处理的效率，快速获得文档主题。同时，通过主题模型得到数据集的“主题-词”分布，方便理解事故树各基本事件之间的关系，有利于事故树的建立与后续分析。

3.1 文本数据简介与预处理

由于需要对广铁集团的高铁列车故障记录进行高铁列车晚点影响因素特征提取，而特征提取要求输入词项文档矩阵，因此必须对原始数据进行预处理与分词，方便下一步的进行，预处理流程如图 3.1 所示。

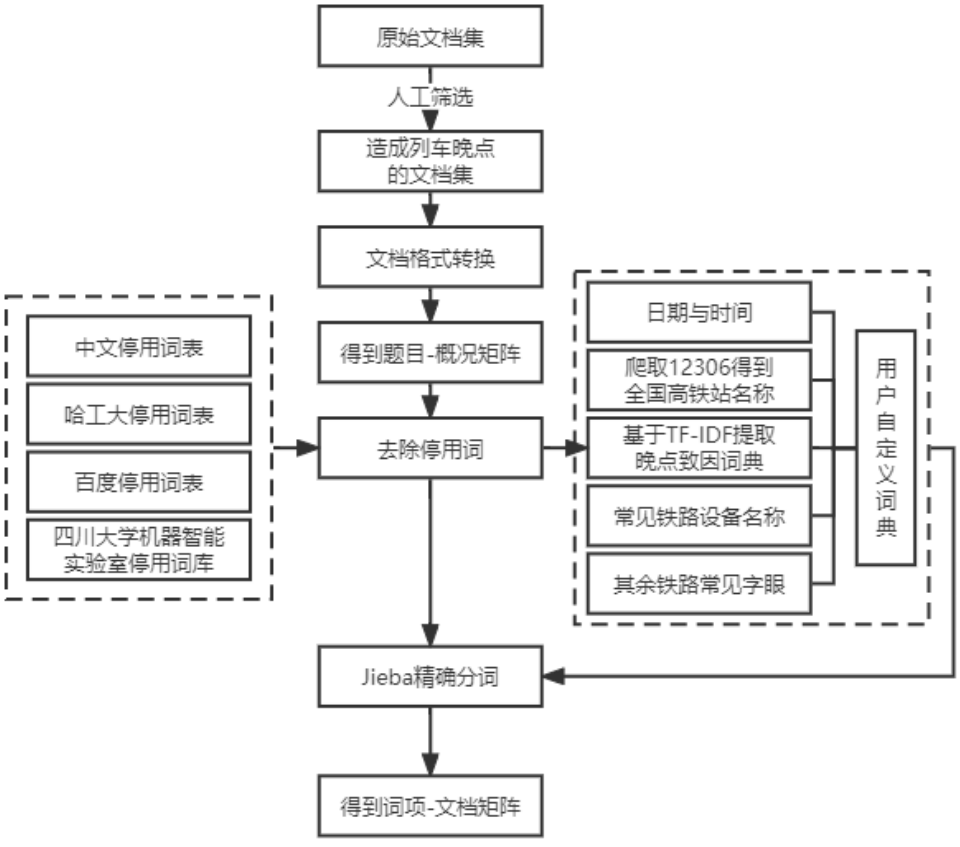

图 3.1 数据集预处理流程

3.1.1 数据简介

本论文的数据来源于 2016 年 1 月至 2019 年 3 月共 39 个月内广铁集团突发事件的应急处置

写实表与故障处置过程，由广铁集团工作人员记录。其中，应急处置写实表由相关工作人员信息、基本概况、处置方案、时间节点与产生影响五部分组成，而故障处置过程由故障概况、处置过程、处置依据三个部分组成，具体如图 3.2 所示。可以发现，“概况”涵盖了本设计想要获取的所有信息，因此需要对其进行提取。

2017 年 2 月 1 日京广高铁武广三台 G1303 次异响处置过程

一、故障概况：

2 月 1 日 22 时 16 分，G1303 次（上海虹桥-广州南，上海动车段 CRH380B-3727+3588 号，广州机务值乘）行至英德西站至清远站间下行线 K2171+500M 处因 01 车底部异响停于 K2174+477M 处（5.4‰下坡道），经随车机械师下车检查确认为撞小鸟且无异常后于 22 时 36 分开。影响动车组 10 列。

二、处置过程：

22:16，G1303 次行至英德西站至清远站间下行线 K2171+500M 处因 01 车底部异响停于 K2174+477M 处,不确定影响邻线。

22:17，通知后续已进入区间的后续列车停车，未进入区间列车扣停站内。通知值班副主任、动车调度及工务、供电等部门。

22:20，G1303 次司机汇报：随车机械师申请邻线限速 160KM/H 下车检查。

22:21，确认符合下车条件，同意机械师下车检查。

22:24，设置上行线 K2175+477M 至 K2173+477M 限速 160KM/H 列控限速，并下达限速命令。

22:25，通知 0G9682 次运行至 K2173+500M 至 K2169+500M 限速 160KM/H。

22:29, G1303 次司机汇报：确认是撞小鸟。

广铁集团突发事件应急处置写实表

2016 年 1 月 5 日

	姓名	分管/处室	职务	通知时间	到岗时间
领导					
业务处					

基本概况：1 月 5 日 20 时 30 分，厦深线 D686 次（深圳北-泉州，广州动车段 CRH1A-1152 号，广州机务段值乘）潮汕站 6 道营业时车组报 08 车 BCU 轴 3 速度故障（代码：8263），经随车机械师下车切除 08 车制动并通过滚动试验后于 20 时 57 分开。21 时 16 分列车到达饶平站 4 道（营业），随车机械师下车检查确认无异常后于 21 时 24 分开。影响本列。

处置方案：依据《铁路技术管理规程》（高速铁路部分）第 407 条“动车组列车运行途中发生车辆故障应急处理”及《关于防止动车组（机车）轮对擦伤的补充规定》（广铁运电【2015】79 号）中“动车组因故障需进行制动切除时的规定”进行处置。

时间	处置经过
20:30	D686 次在潮汕站 6 道营业时司机反映 08BCU 轴 3 速度故障，机械师申请下车检查，邻线要求封锁。

图 3.2 原始数据实例

由于并非所有的故障与突发事件都最终导致了高速铁路列车晚点事件的发生，所以要先对数据集进行人工筛选，得到所有与晚点相关的突发事件应急处置表和故障处置过程，总计 2569 条。

其中，突发事件应急处置表和故障处置过程均以“.doc”的格式储存。为方便通过编程对其进行处理，使用 Python 脚本将其格式均转换为“.docx”。

接着，观察突发事件应急处置写实表和故障处置过程的基本格式，寻找标志性语言，并利用“python-docx”“os”“re”等包分离文件标题与内容，同时得到概况。这样便成功将多个文件的有价值内容集合在一个表格中。

文本预处理的最后一步是去除停用词。停用词实际上是语言中最常见却无明确意义的词，例如语气助词、副词、连词等，它们在文本中的出现频率很高，但实际意义并不大。去除停用词的目的是方便关注更重要的信息，减少训练时间。本次预处理的停用词表来源于“中文停用词表”“哈工大停用词表”“百度停用词表”与“四川大学机器智能实验室停用词库”，对其合并去重，删掉其中的英文并整理了其中的标点符号，最终得到 2803 个停用词。

3.1.2 分词

经历过文本预处理，本文得到了导致高铁列车晚点的去掉停用词的相关事件描述，但仍未完成获得词项文档矩阵的目标，因此还需要对上一步的结果进行分词操作。这里主要运用 Jieba 分

词器的精确分词模式，将句子最精确地切分开以进行文本分析。

Jieba 分词属于概率语言模型分词，它基于前缀词典的 Trie 树模型实现词图扫描，生成句中所有可能词情况的有向无环图，动态规划查找最大概率路径，对于词典中不涵盖的词则基于隐马尔可夫模型来预测分词，从而得到分词结果。

由于本次文本分析所在领域为交通运输领域，更具体些的描述是高铁相关，而 Jieba 自带的含有两万多条词的词典在该领域涉猎较少，因此，为了提高分词的准确性，本文定义了新的用户词典。

该自定义词典主要由五个部分组成：

① 日期与时间：“某年某月某日”“某时某分”；

② 全国高铁站名称：爬取 12306 网站得到，共 2866 条，格式如“太原南”“太原南站”；

③ 常见铁路设备名称：如“道岔”“断路器”等；

④ 其余铁路常见字眼：如“京广高铁”“贵广台”等；

⑤ 晚点致因词典：综合词频-逆文件频率的数据处理结果与经验得到。

这里的词频-逆文件频率方法（TF-IDF）认为一个词语在一篇文章中出现次数越多，同时在所有文档中出现次数越少，其重要性越大，越能够代表该文章，基于 TF-IDF 方法对上一步的数据进行关键词提取，得到词云图，如图 3.3 所示。

图 3.3 TF-IDF 得到的词云图

将其与经验中常见的晚点致因综合，形成更为全面的晚点致因词典，如图 3.4，以方便分词后对先验知识的提取。

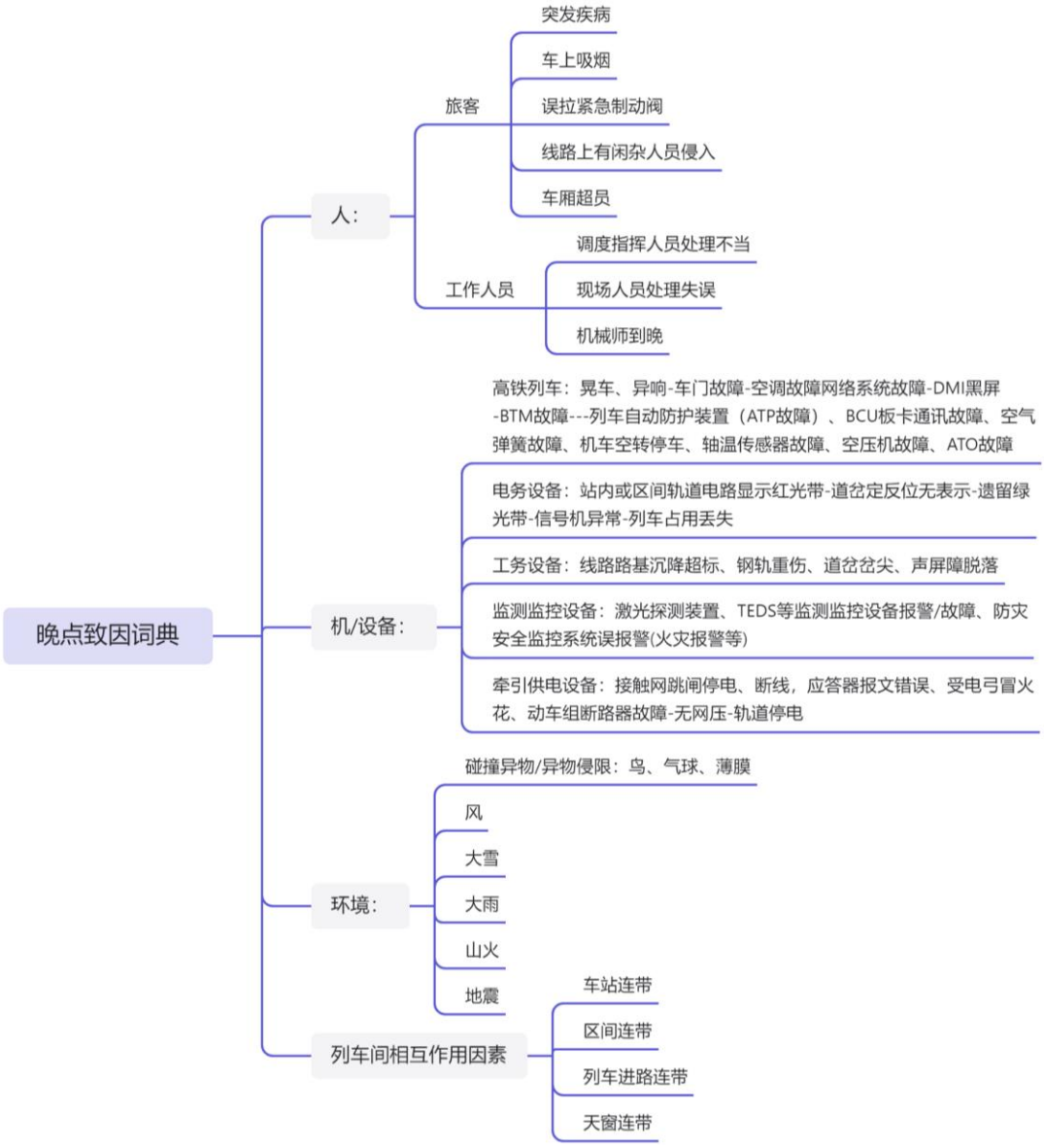

图 3.4 TF-IDF 后形成的晚点致因词典

载入新的用户词典，对数据进行精确分词，并去除其中的单字与重复词，获得分词结果——晚点描述词项。

3.2 基于相关性的先验知识提取

LDA 主题模型是无监督的贝叶斯模型，在训练过程中具有一定的盲目性。若把基于相关性的先验知识整合入 LDA 模型之中，将有利于主题挖掘，提高挖掘精度。因此，将在本节讲述相关性先验知识的提取流程。

3.2.1 卡方值矩阵

卡方值是量化相关性的重要指标。在本算法中，可以假设词项特征w与晚点致因模式c之间的

非独立关系是类似一维自由度的卡方分布，通过计算获得二者的相关性矩阵T。

首先，需要对 2567 条数据进行晚点致因模式标记，序号与对应模式类别如表 3.1 所示。

表 3.1 晚点致因模式分类

模式序号	模式名称
0	旅客相关
1	工作人员相关
2	高铁列车相关
3	电务设备相关
4	工务设备相关
5	监测监控设备相关
6	牵引供电设备相关
7	环境
8	列车间相互作用
9	其他原因

对于词项特征w与晚点致因模式c，可以根据式（3.1）计算二者的卡方值。

$$\chi^2(w,c_i)=N*\frac{[P(w,c_i)P(w',c_i')-P(w',c_i)P(w,c_i')]^2}{P(w)P(w')P(c_i)P(c_i')} \tag{3.1}$$

式中，$P(w,c_i)$代表词项t与致因模式c同时在一篇文档中出现的概率；$P(w',c_i')$代表词项t与致因模式c同时不在一篇文档中出现的概率；$P(w',c_i)$代表词项t出现，而致因模式c不出现在这篇文档中的概率；$P(w,c_i')$则代表词项t不出现，而致因模式c出现在这篇文档中的概率。卡方统计值越高，词项对模式的相关性越大；若卡方统计值为 0，则词项与模式不相关。

通过以上计算，获得词项特征w与晚点致因模式c的相关性矩阵T。

设词项特征共有m个，晚点致因模式共有n个，采用式（3.2）对计算出的卡方值矩阵T进行初始化，以方便对比。

$$\bar{T}(w_i,c_j)=\frac{T(i,j)^2}{\sum_{i=1}^{n}T(i,j)*\sum_{j=1}^{m}T(i,j)} \tag{3.2}$$

通过以上计算，获得初始化后词项特征w与晚点致因模式c的相关性矩阵为$\bar{T}$。

3.2.2 先验知识提取

在得到相关性矩阵$\bar{T}$后，采用 K 均值法对矩阵中的各值进行聚类，以判断其相关性强度。由于 C-H 分数由聚类后的分离度与紧密度的比值进行计算，分数越高代表类自身越紧密，类间越分散[29]，因此根据 C-H 指标确定聚类的目标簇数为 12 簇。降序排列这 12 个聚类类簇。

定义词项特征w与晚点致因模式c之间存在三种相关关系：强关联、弱关联与复杂相关关系。认为前三个类簇对应的词项特征与致因模式属于强相关关系；后三个类簇对应的词项特征与致因模式属于弱相关关系；中间六个类簇则为复杂相关关系，并记录这六个类簇的中心点为θ_i，$i=$

1,2,3,4,5,6。

根据$\bar{T}$设置同样大小的矩阵S描述判断出来的相关关系，命其初值为 0。若词项特征w与晚点致因模式c为强相关关系，则令其值为 2；若二者是弱相关关系，则令其值为10^{-12}；若二者是复杂相关关系，则令其值通过式（3.3）计算。

$$S = \frac{\theta_i - t}{t'} \tag{3.3}$$

式中，t与t'均是给定的阈值，t由卡方检验的P值确定，t'由中间类簇的中心点确定。

已知卡方检验返回卡方值与P值两个统计量，同时常使用 0.05 作为P值的显著性水平。因此，在自由度为 1 的情况下，得到对应P的阈值为 3.841。由于在之前的步骤中对卡方值进行了初始化，那么对该阈值进行同样的操作后便获得了阈值 t。

为了使复杂关系得到的S值落在（0, 1）之间，t'取类簇 3（最大中间类簇）与类簇 6（最小中间类簇）对应中心点的均值。

总结以上对应关系，得到表 3.2。

表 3.2 词项特征与致因模式相关关系的阐述

相关关系	类簇	词项特征分配至对应致因模式主题下的概率	S
强	0,1,2	加大	2
弱	9,10,11	减小	10^{-12}
复杂	3,4,5,6,7,8	综合考虑	$(\theta_i - t)/t'$

为每个晚点致因模式c预分配 10 个潜在主题特征z，设晚点描述词项特征w与晚点致因主题特征z之间的相关性矩阵为Γ。将词项特征w与晚点致因模式c 对应的S的值，赋给词项特征w与晚点致因模式c所包含的所有潜在主题特征z对应的Γ。

经过以上操作便完成从词项特征w与晚点致因模式c之间的相关性到词项特征w与主题特征z之间相关性的计算，图 3.5 总结了这一过程。

Algorithm 1 ：相关性知识提取算法步骤

输入数据： 处置过程与处置写实表记录D; 高铁常见晚点致因词典Ω; 晚点致因模式集C; 潜在晚点致因主题特征集Z 。

输出结果： 晚点描述词项特征$w_i \in W$与晚点致因主题特征$z_k \in Z$相关性集合$\Gamma(w_i, z_k)$。

1: $W \leftarrow$ 词库：由加入常见晚点致因词典Ω的中文分词工具对D进行中文分词得到;
2: $M \leftarrow$ 词项文档矩阵：基于W和D,由向量空间模型（VSM）表达得到；
3: for $w_i \in W$ 且$c_j \in C$ do；
4: $T(i,j) \leftarrow$词项特征w_i和晚点致因模式c_j的相关性，由公式(3-1)求得。
5: end for
6: $\overline{T} \leftarrow$通过公式(3-2)对T进行归一化。
7: $\Xi \leftarrow$k(k=12)个聚类类簇：由K—means对$\overline{R}$聚类得到，并降序排列。
8: $\Theta \leftarrow$复杂相关程度集合：Ξ中除最高三个和最低三个类簇外，Θ_i(i=1,2,3,4,5,6)表示剩下k-6=6个类簇中第i个的中心点。
9: for $w_i \in W$ 且$c_j \in C$ do；
10: if $\overline{T}(w_i,c_j)$属于Ξ最高的三个或最低的三个类簇then
11: $\overline{T}(w_i,c_j)$被指定为词项特征w_i和晚点致因模式c_j为强关联关系或弱关联关系,并且将对应的S矩阵的值分别设置为正数(>1)或极小数(≈0且>0)。
12: else
13: $\overline{T}(w_i,c_j)$被指定为词项特征w_i和晚点致因模式c_j为复杂关联关系，S的值由公式(3-3)求得。
14: end if
15: end for
16: 为每个晚点致因模式预分配m(m = 10)个相应的潜在主题特征$z_{10*i}, z_{10*(i+1)}, \ldots, z_{10*(i+m)}(1 \le i \le |C|)$。
17: $\Gamma(w_i, z_k) \leftarrow$初始化词项特征$w_i$与主题特征$z_k \in Z$相关性为0。
18: for $w_i \in W$ 且$z_k \in Z$ do；
19: if $z_k \in c_j$, then
20: 将$S(w_i,c_j)$的值分配给$\Gamma(w_i, z_k)$。
21: end if
22: end for

图 3.5 相关性知识提取算法流程图

3.3 先验知识与 LDA 模型的整合

在本节中，选择对主题更新公式进行略微的修改，以方便将上一步提取出的先验知识整合在 LDA 模型中。本文将结合 2.1 隐含狄利克雷模型中的基本算法与最优主题个数选择两部分的内容对整合过程进行介绍。

3.3.1 主题更新公式

隐含狄利克雷模型的原始算法选择了基于吉布斯采样的式（2.1）更新词w_i的主题概率$P(z_i = j \mid z_{-i}, w, \alpha, \beta)$，根据主题概率采样一个新主题赋予该词，然后同样方法更新下个词直至收敛，也就是说它由此更新了词项对应的主题编号与语料库中该词项的编号。除此之外，基于式（2.1）

由式（3.4）与（3.5）可以获得主题的多项式分布θ与词语分布φ。

$$\theta = \frac{n_{-i,j}^{w_i} + \beta}{\sum_w^W n_{-i,j}^{w_i} + W\beta} \tag{3.4}$$

$$\varphi = \frac{n_{-i,j}^{d_i} + \alpha}{\sum_j^T n_{-i,j}^{d_i} + T\alpha} \tag{3.5}$$

由于已经获得词项特征w与主题特征z之间相关性矩阵Γ，为了将其融合到主题更新公式中，根据$\Gamma(\mathrm{w_i}, \mathrm{z_k})$的值选择不同的主题概率更新公式：

$$P(z_i = j \mid z_{-i}, w, \alpha, \beta) = \left(\frac{n_{-i,j}^{w_i} + \beta}{\sum_w^W n_{-i,j}^{w_i} + W\beta}\right)\left(\frac{n_{-i,j}^{d_i} + \alpha}{\sum_j^T n_{-i,j}^{d_i} + T\alpha}\right) * \Gamma(\mathrm{w_i}, \mathrm{z_j}) \tag{3.6}$$

$$\varphi' = \frac{(1 + \Gamma(\mathrm{w_i}, \mathrm{z_j}))n_{-i,j}^{d_i} + \alpha}{\sum_j^T (1 + \Gamma(\mathrm{w_i}, \mathrm{z_j}))\, n_{-i,j}^{d_i} + T\alpha} \tag{3.7}$$

$$P(z_i = j \mid z_{-i}, w, \alpha, \beta) = \varphi'\left(\frac{n_{-i,j}^{d_i} + \alpha}{\sum_j^T n_{-i,j}^{d_i} + T\alpha}\right) \tag{3.8}$$

① 若$\Gamma(\mathrm{w_i}, \mathrm{z_k})$值为$2$：证明晚点描述词项特征$w$与对应的晚点致因潜在主题特征$z$为强相关关系，需要加大词项分配到主题的概率，则其对应主题概率更新公式如式（3.6）；

② 若$\Gamma(\mathrm{w_i}, \mathrm{z_k})$值为$10^{-12}$：证明晚点描述词项特征$w$与对应的晚点致因潜在主题特征$z$为弱相关关系，需要降低词项分配到主题的概率，则其对应主题概率更新公式同样为式（3.6）；

③ 若$\Gamma(\mathrm{w_i}, \mathrm{z_k})$值其它值：证明晚点描述词项特征$w$与对应的晚点致因潜在主题特征$z$为复杂相关关系，由此需要对主题概率更新公式中的词项分布φ进行一定的修改，修改后的φ'如式（3.7）所示，新的主题概率更新公式如式（3.8）所示。

3.3.2 算法流程

除去主题概率公式，在整个模型建立之前，还需要确定最佳主题个数。

根据 2.1.3 最佳主题数目的确定方法， U_mass Coherence 优化了语义一致性；C_v Coherence 探索了主题连贯性度量的空间，相对比而言，前者表示方便，后者表示了精度。因此，在本模型中选择这两个指标确定最优主题个数。

将修改后的主题概率更新公式应用至 LDA 模型算法之中，遍历主题个数，范围为 3~110，分别获得 C_v Coherence 与 U_mass Coherence 评分，绘制主题连贯性-主题个数折线图，如图 3.6 与 3.7 所示。

对于 C_v Coherence -主题个数折线图，最高点对应的主题个数为最优主题个数；对于 U_mass Coherence-主题个数折线图，最低点对应的主题个数为最优主题个数。观察图 3.5 与 3.6，确定本论文的最优主题个数为 74。

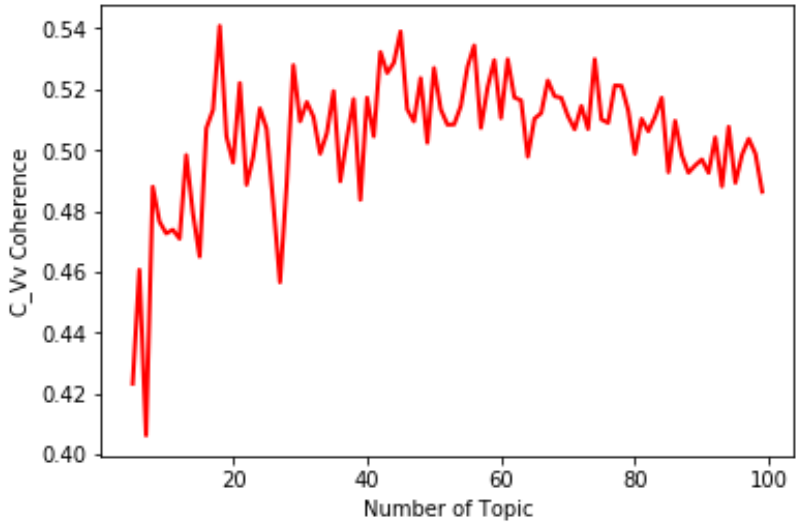

图 3.6 C_v Coherence -主题个数折线图

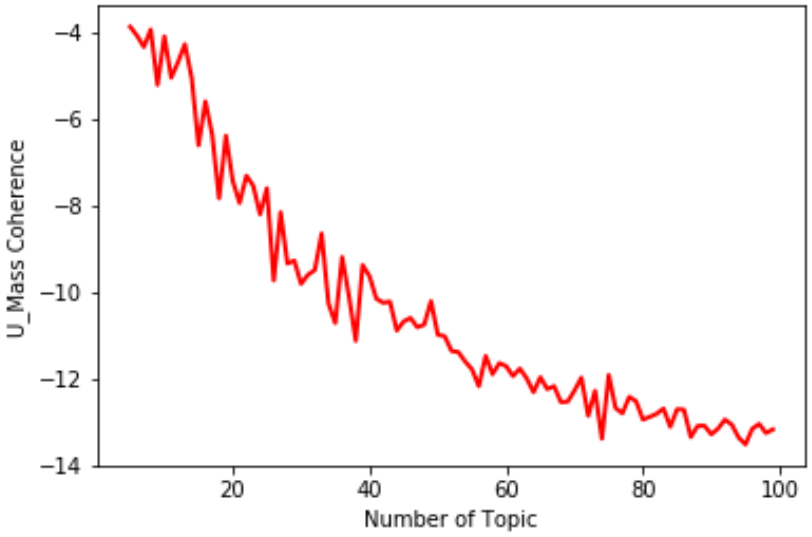

图 3.7 U_mass Coherence -主题个数折线图

输入词项特征w与主题特征z之间相关性矩阵Γ，最优主题个数$N = 74$，以及词项-文档矩阵M，训练融合先验知识的 LDA 主题模型，获得高速铁路列车晚点的文档-主题分布θ与主题-词分布φ。

整个流程如图 3.8 所示。

Algorithm 1 ：整合先验知识的LDA主题模型建立步骤

输入数据： 晚点描述词项特征$w_i \in W$与晚点致因主题特征$z_k \in Z$相关性集合,$\Gamma(w_i, z_k)$; 处置过程与处置写实表记录(分词后)D; 超参数α, β; 最优主题个数K; 迭代次数$iter\,times$;

输出结果： 文档主题分布θ_d；LDA主题与词的分布ϕ_k。

1: $K \leftarrow$模型最优主题个数：根据主题连贯性指标求得$K = 74$。

2: 选择合适的超参数向量α, β。

3: 变量申请：

概率向量p，；词在类上的分布nw；每个类上的词的总数$nwsum$；每篇文章中，各个类的词个数分布nd；每篇文章中的词的总个数$ndsum$；每个词分派一个类Z; 文章→类的概率分布θ；类→词的概率分布φ

4: 初始化阶段：对应语料库中每一篇文档的每一个词，随机的赋予一个主题编号z。

```
for x in 文章数do：
    统计ndsum[文章id][词的个数]
    for y in 每篇文章的词个数do:
        给所有词随机分派一个类
    end for
end for
```

5: 重新扫描语料库，对于不同相关关系的词项，选用修正后的Gibbs采样公式更新其主题编号，并更新语料中该词的编号。

```
for i in 迭代次数:
    for m in 文章数do:
      for v in 文章中词:
        取topic = Z[m][v]
        判断词项与主题相关关系，根据公式(3.6)(3.7)(3.8)计算概率p
        for k in (1,类的个数-1) do:
            p[k] += p[k-1]
        end for
        再随机分派一次，记录被分派的新的topic
      end for
    end for
end for
```

6: 重复上一步基于坐标轴轮换的Gibbs采样，直到Gibbs采样收敛。

7: 统计语料库中各个文档各个词的主题，得到文档主题分布θ_d，统计语料库中各个主题词的分布，得到LDA主题与词的分布ϕ_k。

图 3.8 整合先验知识的 LDA 主题模型建立算法流程

3.3.3 建模结果可视化

利用 pyLDAvis 包将建模结果可视化，其中一页结果（主题 4 相关展示）如图 3.9 所示。

气泡大小指出对应主题在库中的重要性，与图例进行比较，发现主题 4 的出现频率约在 4%~5%，该主题出现频率降序排列第四。

对于每个气泡，右侧的直方图列出了前 30 个最相关的词汇。对于主题 4 而言，该主题最相关的词语是 ATP 故障，其余相关词语还包括“重启”“停车”“冒进”“紧急制动”“显示”等。已知，ATP 发生故障往往引起列车的紧急制动，司机会尝试重启 ATP，观察是否恢复正常，从而采取其他措施；除此之外，当 ATP 显示冒进信号时，列车也会立刻停车，这是为了防止造成列车冲突、脱轨等事故。这两种导致列车停车从而晚点的原因都与 ATP 装置有关，LDA 模型将其归为一类。

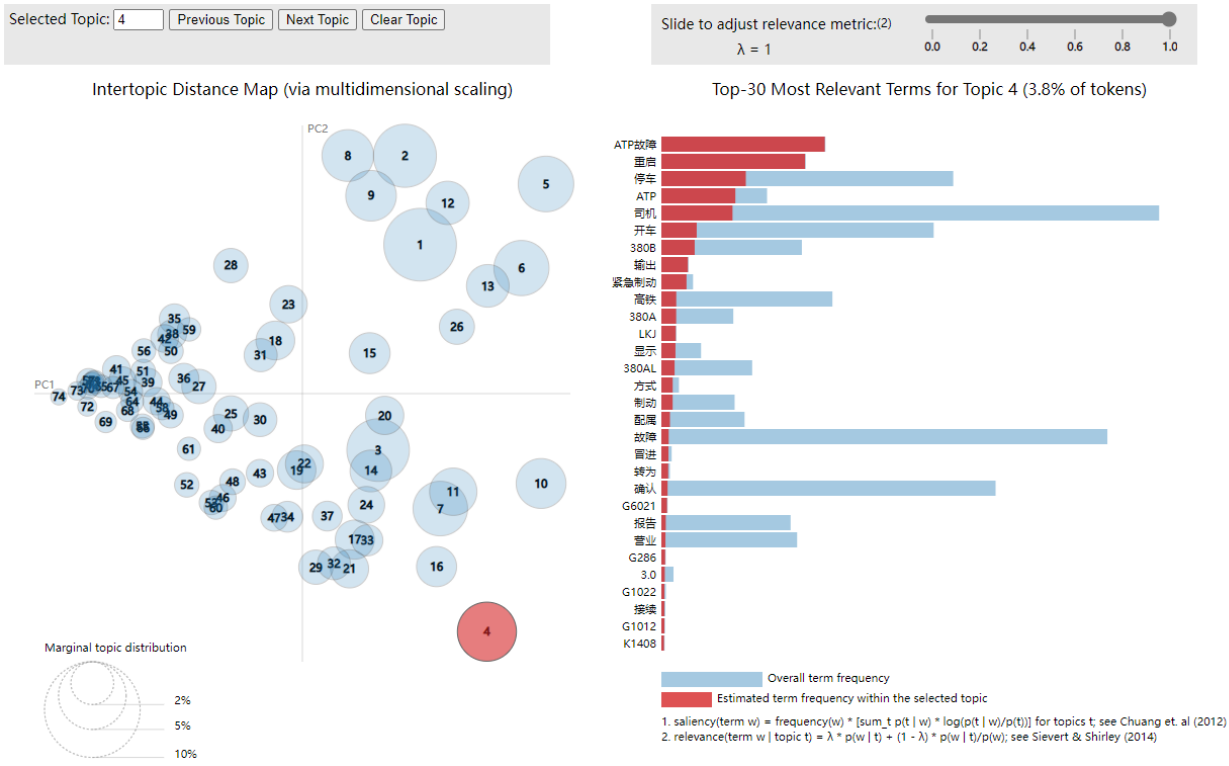

图 3.9 建模结果可视化展示

相关词语右侧的浅蓝色条表示这个词在整个文档中出现的权重，深红色条表示这个词在这个主题中所占的权重。通过调整参数λ可以变化词语权重的计算公式，如式（3.9），以得到不同的主题-词语排序。如果λ接近 1，那么在该主题下更频繁出现的词，跟主题更相关；如果λ接近 0，那么该主题下更特殊、更独有的词跟主题更相关。以主题 6 为例，分别取$\lambda = 0.6$与$\lambda = 1$时主题 6 的相关词，如图 3.10 所示。观察发现，两种情况下主题 6 最相关的词汇均为“受电弓”，“受电弓”是该主题的关键词语；随着参数的增大，词汇“承力索”“接触网”的权重上升，词汇 “换弓”等的权重下降，这有助于对该主题进行总结。

$$relevance(term\ w|topic\ t) = \lambda * p(w|t) + (1-\lambda) * \frac{p(w|t)}{p(w)} \tag{3.9}$$

气泡中心距离代表主题相似性，圆圈有重叠说明对应主题存在词汇交叉。例如，观察图 3.9 发现，主题 2、主题 8 和主题 9 彼此接近。查看各主题相关词汇，发现共用词汇“工务”，且主题-词汇的相关性不低，均在相关性降序前十名。除此之外，选择词汇“工务”，可以获得其在各主题中的分布，如图 3.11 所示，圆圈大小代表了相对频率的大小。这一结果也可以证明主题 2、主题 8 和主题 9 的相近性。

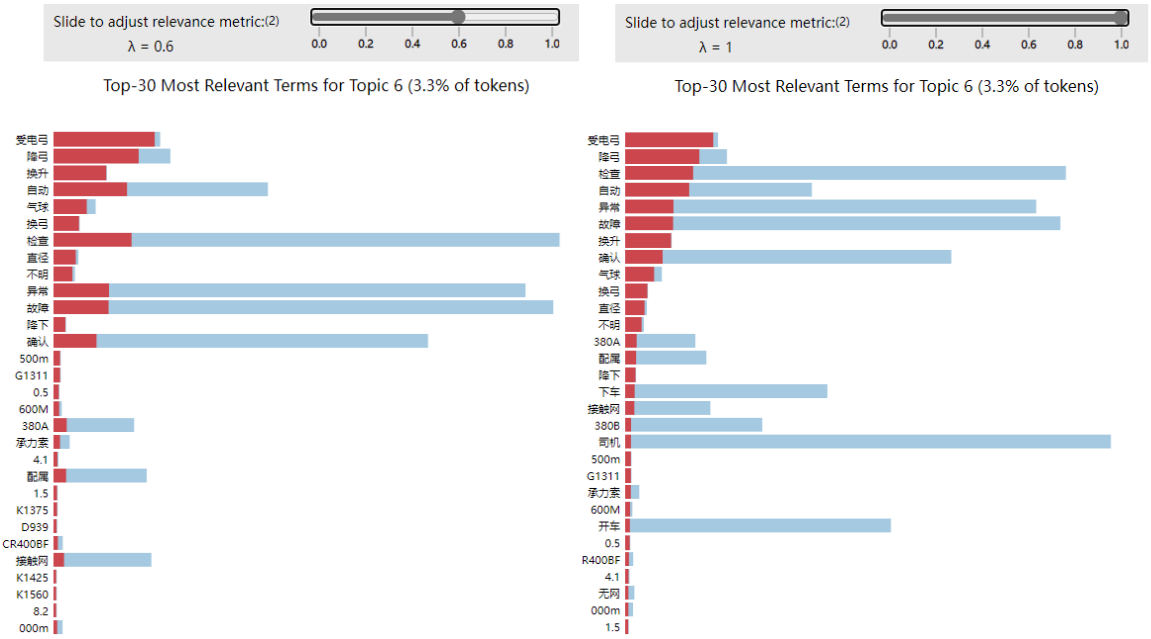

图 3.10 $\lambda = 0.6$与$\lambda = 1.0$时主题 6 的相关词

图 3.11 与词汇“工务”相关性较高的主题分布

3.4 本章小结

本章节完成了对原始数据集的筛选与预处理工作；成功从中提取了高铁列车晚点词项与晚点致因模式的相关性先验知识；并将相关性先验知识融合在 LDA 主题模型之中。

经过以上操作从 2016 年至 2019 年 3 月广铁集团突发事件的应急处置写实表与故障处置过程数据集中获得了高速铁路列车晚点致因主题-词分布，有利于理解晚点致因之间的相互关系，有利于事故树的建立与进一步分析。

4 基于模糊事故树的高铁晚点影响因素分析与应对措施

在获得高速铁路列车晚点致因主题-词分布之后，将在本章对晚点影响因素进行三角模糊事故树建模，分析导致高铁晚点的直接原因与潜在原因；同时以定性与定量分析结果为依据提出针对性改善措施。模糊事故树的分析流程如图 4.1 所示。

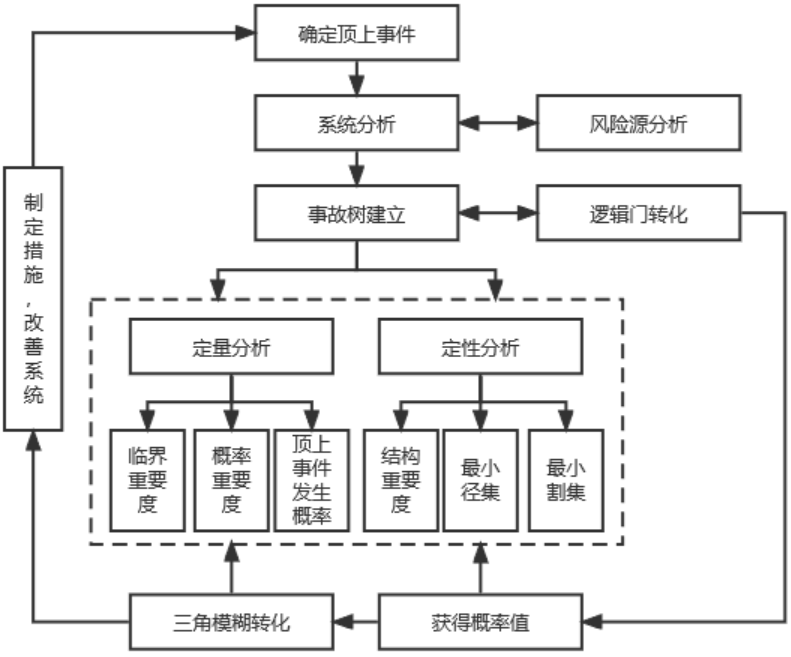

图 4.1 模糊事故树分析流程

4.1 模糊事故树构建

4.1.1 晚点致因分析

晚点指运行或到达晚于规定的时间。在铁路系统中，高速铁路列车晚点指列车在运行过程中由于某种原因比列车运行图上规定的时刻晚到发的现象，同时称该原因为晚点致因。

通过整理国内外文献、专家知识和通过以 2016 年至 2019 年广铁集团的高铁列车故障记录数据为基础的主题模型的提取内容，将高速铁路列车在运营过程中发生的延误事件致因进行不同类型的划分，将晚点致因进一步归纳以下类别：

A. 人类因素

现代城市交通以人为主导，高速铁路列车的运营需要人的参与。在导致列车晚点的原因方面，人类因素可以进一步划分为乘客因素与管理因素。

高速铁路列车的乘客包括工作人员与旅客，二者均有造成列车晚点的可能。常见的乘客因素包括旅客的不良行为、乘客突发疾病、车厢超员等。例如，部分旅客缺乏安全意识，在高速列车内部吸烟，导致烟雾报警器被触发，使得列车立即降速运行或紧急停车，从而发生晚点。除此之

外，还存在部分旅客因坐过站或其它原因误拉紧急制动阀的现象，然而紧急制动阀是为了避免突发事件引起的行车事故而设置的，一旦拉起阀门，列车将立即停车。其余原因，包括闲杂人员侵入线路、乘客突发疾病需要救治、车厢超员、工作人员工作失误(如负责接发列车的客运值班员误打信号)等也都会增大高速铁路列车晚点的可能性。

导致列车晚点的管理因素则主要产生自铁路运输系统内部的工作人员，包括驾驶员、随车机械师、调度员、现场工作人员、车站票务与站务等。当列车或线路内已经产生突发事件，现场人员处理不及时、工作人员误报警、调度员处理不当或者列车运行图不合理[30]都有可能无法挽回流逝的时间，无法恢复晚点，在少数情况下，错误的调度甚至会导致晚点事件的传播。

B. 系统因素

高速铁路列车运行涉及多个部门与多种设施设备，因此高铁列车晚点的系统致因包含了列车自身、电务设备、工务设备、监测设备、牵引供电设备的故障与报警。

列车故障包含晃车、车体故障（如车门故障、空气弹簧故障、轴温传感器故障、空压机故障）与车载系统故障等；电务设备故障则主要表示为站内或区间轨道电路显示红光带、道岔定反位无表示、遗留绿光带、信号机异常等；除此之外，线路路基沉降超标、钢轨重伤、道岔岔尖发生故障以及声屏障脱落等也是导致高速铁路列车晚点的重要因素；同时，TEDS 系统报警、激光探测装置与防灾安全监控系统(误)报警等监测监控设备的故障也是晚点致因的一大组成部分；至于牵引供电设备发生的故障则包括接触网跳闸停电、受电弓冒火花、轨道停电、无网压、动车组短路器故障。

C. 环境因素

高速铁路列车晚点的环境因素主要包括自然灾害与异物碰撞、侵限事件。自然灾害由台风、雷雨、暴雪、山火、地震等组成，一旦发生将导致大量列车晚点，更严重的还会取消行程。同时，鸟、气球、薄膜等异物侵限接触网，或小狗、小鸟甚至是人与列车发生碰撞，这些因素都会干扰轨道交通的正常运营，促使列车停车以致晚点产生。

D. 列车间相互作用因素

我国高速铁路列车的实际追踪间隔小，当调度员无法及时对前行列车产生的干扰采取有效措施时，为了避免发生行车事故，后续列车必须限速追踪运行，从而导致后续列车的晚点。列车间的相互作用包括车站连带、区间连带、列车进路连带、天窗连带[31]。在实际生产生活中，运行图中设置的缓冲时间足够大部分晚点列车恢复晚点，避免了晚点传播现象；但是存在小部分受干扰较大的列车无法恢复晚点，从而导致连带晚点的出现。

4.1.2 建立事故树

尽管本数据集涵盖 2563 条有效数据，但是实际生产生活中各晚点影响因素的发生频率与统计数据仍存在一定的区别，为了对高速铁路列车晚点成因进行更可靠的分析，选择三角模糊事故树对晚点影响因素进行建模。

结合 2.2 章节中事故树的基本知识与晚点致因分析，以列车晚点事件为顶上事件，上述四大方向为事故树的次顶事件，探索挖掘出所有的基本事件与逻辑关系，完成高速铁路列车晚点模糊事故树的建立。事件介绍如表 4.1 所示，树结构如图 4.2 所示。

表 4.1 晚点事故树事件阐述

编号	事件名称	编号	事件名称
M_1	非管理因素	X_{10}	晃车、异响
M_2	系统因素	X_{11}	车体故障
M_3	环境因素	X_{12}	车载系统故障
M_4	乘客因素	X_{13}	信号机异常
M_5	管理因素	X_{14}	轨道电路故障
M_6	列车故障	X_{15}	道岔定反位无表示
M_7	电务设备故障	X_{16}	列车占用丢失
M_8	工务设备故障	X_{17}	路基沉降超标
M_9	监测监控系统/设备故障	X_{18}	钢轨重伤
M_{10}	牵引供电设备故障	X_{19}	声屏障脱落
M_{11}	非调度因素	X_{20}	激光探测装置报警/故障
X_1	乘客吸烟	X_{21}	TEDS 系统报警/故障
X_2	闲杂人员侵入线路	X_{22}	防灾安全监控系统报警/故障
X_3	车厢超员	X_{23}	接触网跳闸停电、受电弓冒火花
X_4	乘客误操作	X_{24}	断路器故障
X_5	乘客突发疾病	X_{25}	轨道停电
X_6	现场人员处理失误	X_{26}	自然灾害
X_7	运行图不合理	X_{27}	异物碰撞
X_8	调度指挥人员处理不当	X_{28}	异物侵限
X_9	列车间相互作用因素		

装 订 线

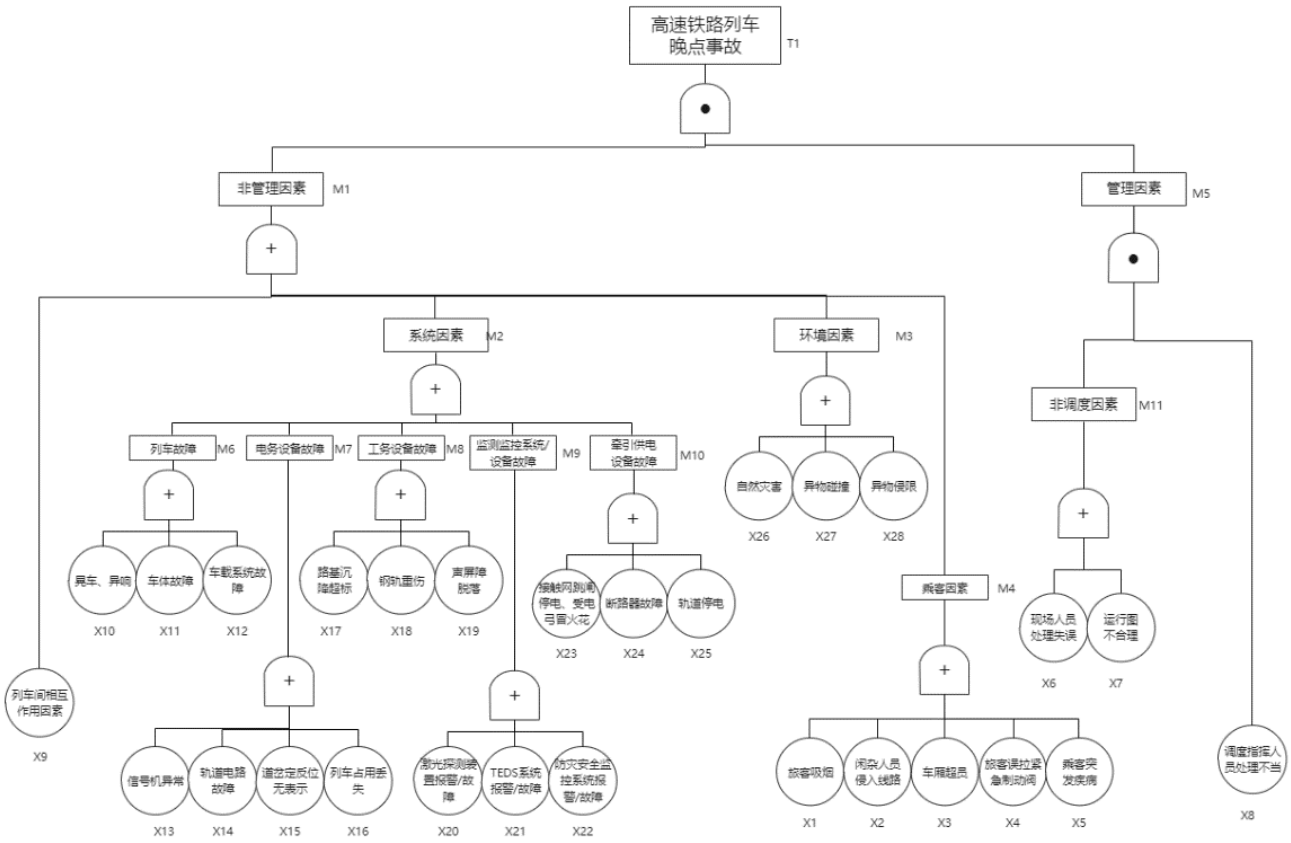

图 4.2 高铁晚点模糊事故树

4.2 基本事件三角模糊处理

根据图 4.1，在建立高铁晚点事故树之后，需要对 28 个基本事件的模糊概率进行计算，并将其转化为三角模糊数。基本事件分为可统计概率的基本事件与不可统计概率的基本事件。

4.2.1 可统计的基本事件

对于可统计的基本事件，根据数据集数据统计获得其发生概率。由于事故发生的随机性以及人类操作可能出现失误，实际应用中可统计事件的发生概率也在范围内波动，因此同样需要对其进行模糊化处理。本论文基于三角模糊数法则进行模糊化处理，将统计确定的精确概率值赋给 m 值，上、下边界 α 与 β 则取误差范围为±5%，即 α=0.95m,β=1.05m。计算结果如表 4.2 所示。

表 4.2 可统计的基本事件三角模糊概率值

编号	事件名称	α	m	β
X_1	乘客吸烟	0.0090	0.0095	0.0100
X_2	闲杂人员侵入线路	0.0319	0.0336	0.0353
X_3	车厢超员	0.0023	0.0024	0.0025
X_4	乘客误操作	0.0131	0.0138	0.0145

续表 4.2

编号	事件名称	α	m	β
X_5	乘客突发疾病	0.0165	0.0173	0.0183
X_9	列车间相互作用因素	0.6447	0.6786	0.7125
X_{10}	晃车、异响	0.0473	0.0498	0.0523
X_{11}	车体故障	0.1449	0.1525	0.1601
X_{12}	车载系统故障	0.1610	0.1695	0.1780
X_{13}	信号机异常	0.0199	0.0209	0.0220
X_{14}	轨道电路故障	0.0537	0.0565	0.0593
X_{15}	道岔定反位无表示	0.0642	0.0676	0.0709
X_{16}	列车占用丢失	0.0206	0.0217	0.0228
X_{17}	路基沉降超标	0.0030	0.0032	0.0033
X_{18}	钢轨重伤	0.0056	0.0059	0.0062
X_{19}	声屏障脱落	0.0015	0.0016	0.0017
X_{21}	TEDS 系统报警/故障	0.0165	0.0174	0.0183
X_{23}	接触网跳闸停电、受电弓冒火花	0.1306	0.1375	0.1444
X_{24}	断路器故障	0.0135	0.0142	0.0149
X_{25}	轨道停电	0.0079	0.0083	0.0087

4.2.2 不可统计的基本事件

轨道交通作为一个复杂系统，含有大量基本事件不可统计，需要进行概率模糊化，本论文采用 3σ 表征法完成以上工作。

3σ 表征法指由三人及以上专家组成的团队根据经验对基本事件可能发生的概率进行打分，将打分结果的平均值赋给 m 值；同时根据正态分布原则，将三倍标准差 3σ 赋给上、下边界 α 与 β，即该类基本事件的最终模糊概率为(m-3σ,m,m+3σ)。本事故树中不可统计的基本事件专家打分结果如表 4.3 所示。

表 4.3 不可统计的基本事件专家打分结果

编号	事件名称	专家 1	专家 2	专家 3	m	σ
X_6	现场人员处理失误/不及时	0.2764	0.2455	0.2639	0.2619	1.269E-02
X_7	运行图不合理	0.0134	0.0119	0.0142	0.0132	9.534E-04
X_8	调度指挥人员处理不当	0.1433	0.1342	0.1313	0.1363	5.112E-03
X_{20}	激光探测装置报警/故障	0.0132	0.0127	0.0119	0.0126	5.354E-04
X_{22}	防灾安全监控系统报警/故障	0.0079	0.0088	0.0086	0.0084	3.859E-04

将其转化为三角模糊概率，结果如表 4.4 所示。

表 4.4 不可统计的基本事件三角模糊概率值

编号	事件名称	α	m	β
X_6	现场人员处理失误/不及时	0.2239	0.2619	0.3000
X_7	运行图不合理	0.0103	0.0132	0.0160
X_8	调度指挥人员处理不当	0.1209	0.1363	0.1516
X_{20}	激光探测装置报警/故障	0.0110	0.0126	0.0142
X_{22}	防灾安全监控系统报警/故障	0.0073	0.0084	0.0096

4.2.3 基本事件汇总

汇总所有基本事件的三角模糊数，得到表 4.5。

表 4.5 所有基本事件三角模糊概率值

编号	事件名称	α	m	β
X_1	乘客吸烟	0.0090	0.0095	0.0100
X_2	闲杂人员侵入线路	0.0319	0.0336	0.0353
X_3	车厢超员	0.0023	0.0024	0.0025
X_4	乘客误操作	0.0131	0.0138	0.0145
X_5	乘客突发疾病	0.0116	0.0122	0.0129
X_6	现场人员处理失误/不及时	0.2239	0.2619	0.3000
X_7	运行图不合理	0.0103	0.0132	0.0160
X_8	调度指挥人员处理不当	0.1209	0.1363	0.1516
X_9	列车间相互作用因素	0.6447	0.6786	0.7125
X_{10}	晃车、异响	0.0473	0.0498	0.0523
X_{11}	车体故障	0.1449	0.1525	0.1601
X_{12}	车载系统故障	0.1610	0.1695	0.1780
X_{13}	信号机异常	0.0199	0.0209	0.0220
X_{14}	轨道电路故障	0.0537	0.0565	0.0593
X_{15}	道岔定反位无表示	0.0642	0.0676	0.0709
X_{16}	列车占用丢失	0.0206	0.0217	0.0228
X_{17}	路基沉降超标	0.0030	0.0032	0.0033
X_{18}	钢轨重伤	0.0056	0.0059	0.0062
X_{19}	声屏障脱落	0.0016	0.0021	0.0026
X_{20}	激光探测装置报警/故障	0.0110	0.0126	0.0142
X_{21}	TEDS 系统报警/故障	0.0165	0.0174	0.0183
X_{22}	防灾安全监控系统报警/故障	0.0073	0.0084	0.0096

续表 4.5

编号	事件名称	α	m	β
X_{23}	接触网跳闸停电、受电弓冒火花	0.1306	0.1375	0.1444
X_{24}	断路器故障	0.0135	0.0142	0.0149
X_{25}	轨道停电	0.0079	0.0083	0.0087
X_{26}	自然灾害	0.0405	0.0427	0.0448
X_{27}	异物碰撞	0.0417	0.0439	0.0460
X_{28}	异物侵限	0.0950	0.1000	0.1050

4.3 定性分析

对高速铁路列车晚点事故树的定性分析包括最小割集、最小径集的求解与基本事件结构重要度的计算三个部分。其中，最小割集通过布尔代数法求解，最小径集基于对偶理论求解，结构重要度则通过最小径集近似法计算，计算原理见 2.2.2。通过对这三项指标的分析，可以了解高铁晚点的形成途径与控制途径，认识各基本事件在结构上对顶上事件的影响程度。

4.3.1 最小割集

基于布尔逻辑运算法则基本法，得到高速铁路列车晚点事故树结构表达如下：

$$
\begin{aligned}
T = M_1 * M_5 &= (M_2 + M_3 + M_4 + X_9) * (M_{11} * X_8) \\
&= (M_6 + M_7 + M_8 + M_9 + M_{10} + M_3 + M_4 + X_9) * [(X_6 + X_7) * X_8] \\
&= [(X_{10} + X_{11} + X_{12}) + (X_{13} + X_{14} + X_{15} + X_{16}) + (X_{17} + X_{18} + X_{19}) \\
&+ (X_{20} + X_{21} + X_{22}) + (X_{23} + X_{24} + X_{25}) + (X_{26} + X_{27} + X_{28}) + (X_1 + X_2 \\
&+ X_3 + X_4 + X_5) + X_9] * [(X_6 + X_7) * X_8] \\
&= [(X_6 + X_7) * X_8] * (X_1 + X_2 + X_3 + X_4 + X_5 + X_9 + X_{10} + X_{11} + X_{12} \\
&+ X_{13} + X_{14} + X_{15} + X_{16} + X_{17} + X_{18} + X_{19} + X_{20} + X_{21} + X_{22} + X_{23} \\
&+ X_{24} + X_{25} + X_{26} + X_{27} + X_{28}) = \\
&= X_6X_8X_1 + X_6X_8X_2 + X_6X_8X_3 + X_6X_8X_4 + X_6X_8X_5 + X_6X_8X_9 + X_6X_8X_{10} \\
&+ X_6X_8X_{11} + X_6X_8X_{12} + X_6X_8X_{13} + X_6X_8X_{14} + X_6X_8X_{15} + X_6X_8X_{16} \\
&+ X_6X_8X_{17} + X_6X_8X_{18} + X_6X_8X_{19} + X_6X_8X_{20} + X_6X_8X_{21} + X_6X_8X_{22} \\
&+ X_6X_8X_{23} + X_6X_8X_{24} + X_6X_8X_{25} + X_6X_8X_{26} + X_6X_8X_{27} + X_6X_8X_{28} \\
&+ X_7X_8X_1 + X_7X_8X_2 + X_7X_8X_3 + X_7X_8X_4 + X_7X_8X_5 + X_7X_8X_9 + X_7X_8X_{10} \\
&+ X_7X_8X_{11} + X_7X_8X_{12} + X_7X_8X_{13} + X_7X_8X_{14} + X_7X_8X_{15} + X_7X_8X_{16} \\
&+ X_7X_8X_{17} + X_7X_8X_{18} + X_7X_8X_{19} + X_7X_8X_{20} + X_7X_8X_{21} + X_7X_8X_{22} \\
&+ X_7X_8X_{23} + X_7X_8X_{24} + X_7X_8X_{25} + X_7X_8X_{26} + X_7X_8X_{27} + X_7X_8X_{28}
\end{aligned}
$$

对化简后结果表示高速铁路列车晚点事件的最小割集共 50 个，如果割集中的每个基本事件都发生则会引起高铁列车晚点，即存在 50 种方式导致高铁列车晚点。这说明晚点事故极易发生且发生路径多，应当引起注意。除此之外，每个最小割集中包含的基本事件数量少，这说明引

发高铁列车晚点事故的路径较为简单，风险性相对较大。

4.3.2 最小径集

根据对偶理论将晚点事故树中的“与门”和“或门”互换，获得高铁列车晚点成功树，如图 4.3 所示。

基于成功树结构得到其成功树的最小割集，可表示为：

$$T' = M_1' + M_5'$$
$$= (X_9' * M_2' * M_3' * M_4') + (M_{11}' + X_8')$$
$$= (X_9' * M_6' * M_7' * M_8' * M_9' * M_{10}' * M_3' * M_4') + (X_6' * X_7' + X_8')$$
$$= X_1' * X_2' * X_3' * X_4' * X_5' * X_9' * X_{10}' * X_{11}' * X_{12}' * X_{13}' * X_{14}' * X_{15}' * X_{16}' * X_{17}' * X_{18}' * X_{19}' * X_{20}'$$
$$* X_{21}' * X_{22}' * X_{23}' * X_{24}' * X_{25}' * X_{26}' * X_{27}' * X_{28}' + X_6'X_7' + X_8'$$

化简后结果表示高速铁路列车晚点事故的最小径集有 3 个，即存在 3 种方案使高速铁路列车晚点事故不发生。

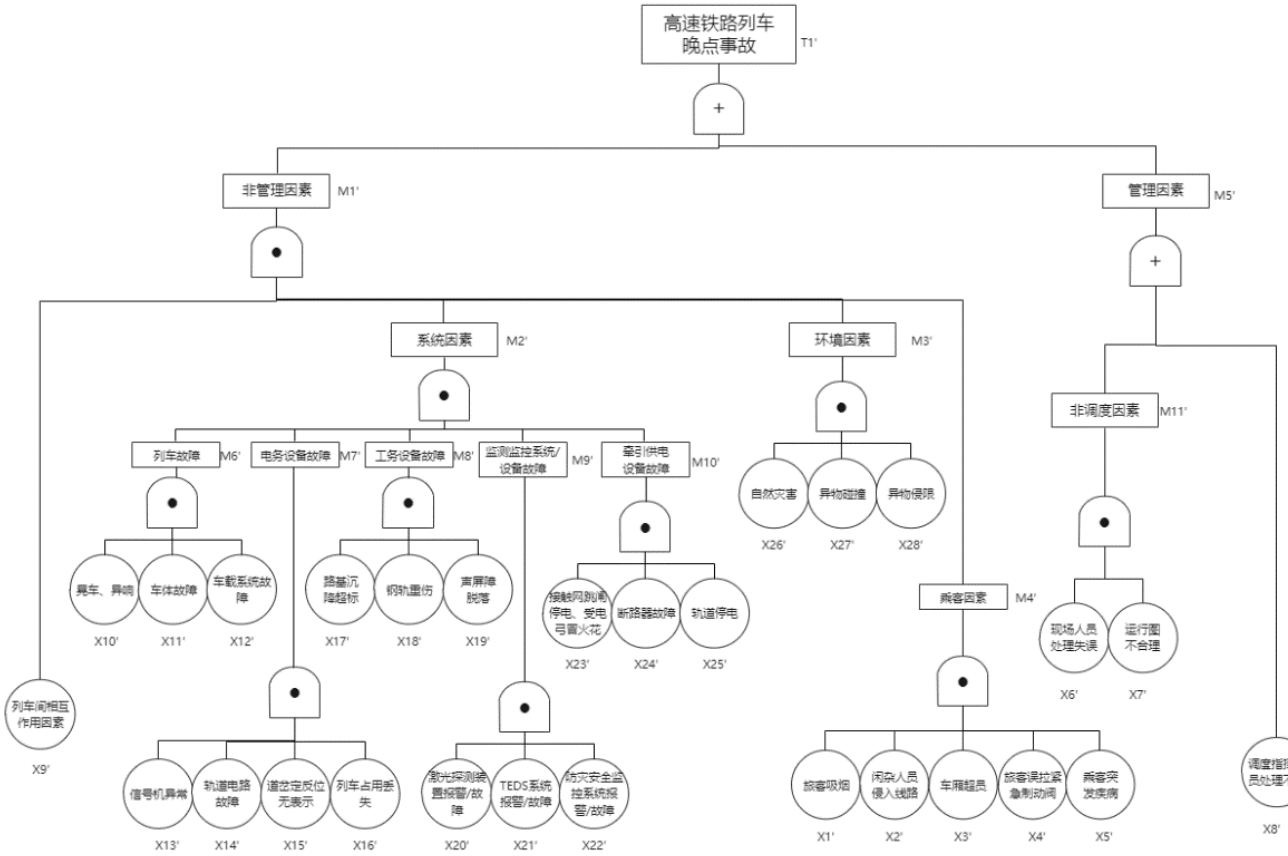

图 4.3 高铁晚点成功树

4.3.3 结构重要度

由于本事故树中基本事件共有 28 个，而最小径集数量较少，因此采用最小径集法。根据式

（2.2）近似求解结构重要度并得到重要度排序如下：

$$I[X_8] > I[X_6] = I[X_7] > I[X_1] = I[X_2] = I[X_3] = I[X_4] = I[X_5] = I[X_9] = I[X_{10}] = I[X_{11}]$$
$$= I[X_{12}] = I[X_{13}] = I[X_{14}] = I[X_{15}] = I[X_{16}] = I[X_{17}] = I[X_{18}] = I[X_{19}] = I[X_{20}]$$
$$= I[X_{21}] = I[X_{22}] = I[X_{23}] = I[X_{24}] = I[X_{25}] = I[X_{26}] = I[X_{27}] = I[X_{28}]$$

观察结构重要度排序，X_8最重要；X_6、X_7次之，处于同等地位；除X_8、X_6、X_7以外的剩余25个基本事件处于同等地位，最不重要。

4.4 定量分析

对高速铁路列车晚点事故树的定量分析包括顶上事件发生概率、事件的概率重要度和临界重要度的求解三个部分。其中，顶上事件发生概率通过最小径集逼近法求解，概率重要度和临界重要度则基于定义求解，计算原理见2.2.2。通过对这三项指标的分析，可以估算系统的可靠性，验证事故树模型在该顶上事件分析上的适用性，得到减少列车晚点发生的有效措施与治理的优先级顺序。

4.4.1 顶上事件模糊概率

由于本事故树的最小径集数量少，因此选用最小径集逼近法近似计算顶上事件“高速铁路列车晚点”事件的发生概率。参照表4.5中各基本事件的三角模糊数，计算得到顶上事件发生的模糊概率为(0.02438, 0.03289, 0.04275)。也就是说，基于本事故树结构，每百辆高铁列车的运行过程中存在2~4辆列车会发生晚点；高铁列车准点的概率为96.7%，波动范围为95.7%~97.6%。

根据中国铁路总公司负责人的介绍，2015年全国铁路动车组始发正点率达98.8%，终到正点率达95.4%[32]；根据世界银行发布的《中国的高速铁路发展报告》，目前（指2019年7月8日止）中国的高铁发车准点率已超过98%。到达准点率超过95%；复兴号列车具备更好的准时性，其出发和到达的准点率分别为99%和98%。

而本论文的数据集来源于2016年至2019年3月年高速铁路列车故障记录，事故树顶上事件发生概率的计算结果与实际生产生活统计结果相近，同时给出了概率值的波动区间，验证了模糊事故树模型在高速铁路列车晚点方面的适用性。

4.4.2 模糊概率重要度

对于三角模糊数（α，m，β）而言，定义中位数z，使经过x=z的分界线使得被分割的模糊函数图的左右两边面积A_1、A_2相等。结合公式（2.6），得到中位数z的计算公式，如公式（4.1）所示。

$$z = \begin{cases} m - \sqrt{(m-\alpha)(2m-\alpha-\beta)}, A_1 > A_2 \\ m, A_1 = A_2 \\ m + \sqrt{(\beta-m)(\alpha+\beta-2m)}, A_1 < A_2 \end{cases} \tag{4.1}$$

由于顶上事件发生的模糊概率为(0.02438, 0.03289, 0.04275)，对应中位数$Z_T = 0.0365265 \approx 0.0365$，记基本事件$x_i$模糊数的中位数为$Z_{T_i}$。参考章节2.2中概率重要度的定义，对于模糊事故树而言，其基本事件的模糊概率重要度Q_i常采用中值法进行计算，计算公式如式（4.2）所

示。

$$Q_i = Z_T - Z_{T_i} \tag{4.2}$$

得到各基本事件的中位数Z_{T_i}、模糊概率重要度Q_i与模糊重要度的降序排序如表 4.6 所示，注意：Q_i越大则认为该基本事件更重要。

表 4.6 所有基本事件的模糊概率重要度

编号	事件名称	Z_{T_i}	Q_i	Q_i排序
X_1	乘客吸烟	3.649E-02	3.909E-05	22
X_2	闲杂人员侵入线路	3.638E-02	1.417E-04	14
X_3	车厢超员	3.652E-02	9.752E-06	27
X_4	乘客误操作	3.647E-02	5.698E-05	19
X_5	乘客突发疾病	3.648E-02	5.096E-05	21
X_6	现场人员处理失误/不及时	1.836E-03	3.469E-02	2
X_7	运行图不合理	3.524E-02	1.283E-03	4
X_8	调度指挥人员处理不当	7.473E-17	3.653E-02	1
X_9	列车间相互作用因素	2.710E-02	9.431E-03	3
X_{10}	晃车、异响	3.631E-02	2.136E-04	11
X_{11}	车体故障	3.579E-02	7.365E-04	6
X_{12}	车载系统故障	3.569E-02	8.362E-04	5
X_{13}	信号机异常	3.644E-02	8.754E-05	16
X_{14}	轨道电路故障	3.628E-02	2.441E-04	10
X_{15}	道岔定反位无表示	3.623E-02	2.950E-04	9
X_{16}	列车占用丢失	3.644E-02	9.035E-05	15
X_{17}	路基沉降超标	3.651E-02	1.245E-05	26
X_{18}	钢轨重伤	3.650E-02	2.416E-05	25
X_{19}	声屏障脱落	3.652E-02	6.521E-06	28
X_{20}	激光探测装置报警/故障	3.647E-02	5.385E-05	20
X_{21}	TEDS 系统报警/故障	3.645E-02	7.215E-05	17
X_{22}	防灾安全监控系统报警/故障	3.649E-02	3.648E-05	23
X_{23}	接触网跳闸停电、受电弓冒火花	3.587E-02	6.522E-04	7
X_{24}	断路器故障	3.647E-02	5.862E-05	18
X_{25}	轨道停电	3.649E-02	3.403E-05	24
X_{26}	自然灾害	3.635E-02	1.811E-04	13
X_{27}	异物碰撞	3.634E-02	1.864E-04	12
X_{28}	异物侵限	3.607E-02	4.538E-04	8

同时，可以得到各中间事件的模糊概率重要度如表 4.7 所示。

表 4.7 所有中间事件的模糊概率重要度

编号	事件名称	Q_i	Q_i排序
M_1	非管理因素	0.014004	3
M_2	系统因素	0.003454	4
M_3	环境因素	0.000821	6
M_4	乘客因素	0.000298	9
M_5	管理因素	0.0725	1
M_6	列车故障	0.001786	5
M_7	电务设备故障	0.000717	8
M_8	工务设备故障	4.31E-05	11
M_9	监控监测设备故障	0.000162	10
M_{10}	牵引供电设备故障	0.000745	7
M_{11}	非调度因素	0.035973	2

为了方便分析，在事故树结构图中根据重要度排序结果为各基本事件与中间事件赋予绿、红色阶背景色，结果如图 4.4 所示。

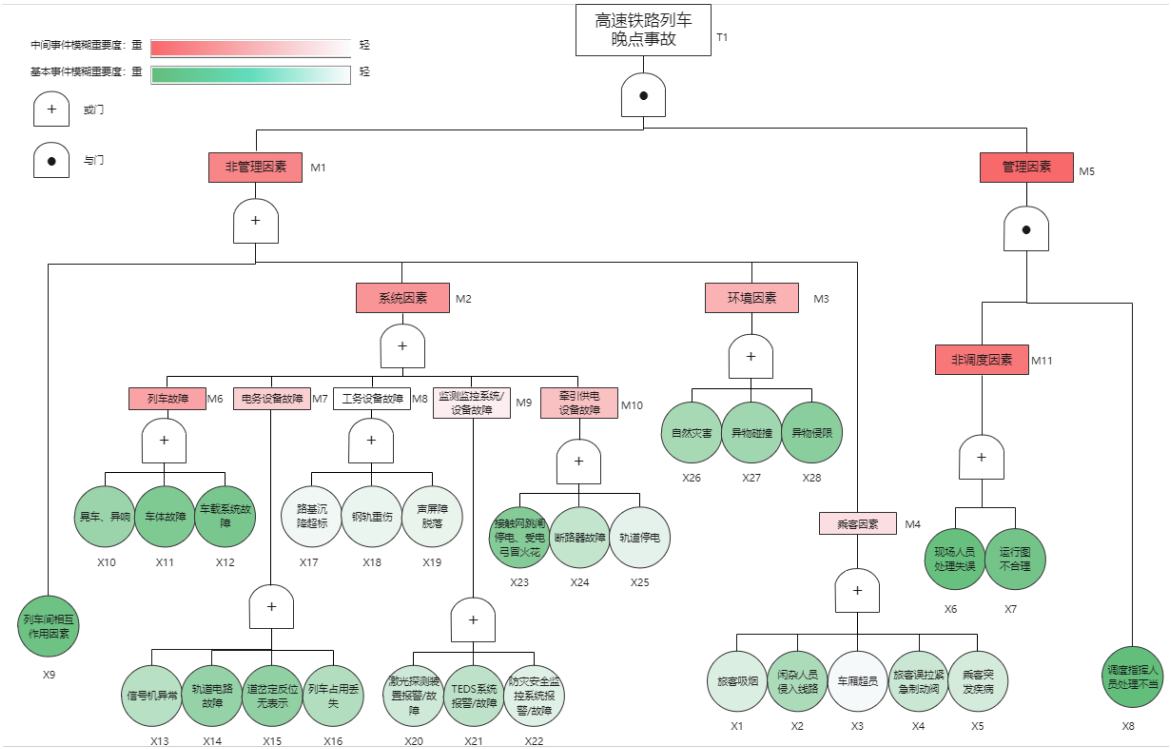

图 4.4 模糊概率重要度分析-高铁晚点事故树

观察图 4.4，发现：

（1）管理因素对顶上事件的影响程度是中间事件中最大的，无论是现场人员处理失误/不及时、运行图不合理，还是调度指挥人员处理不当都极大的影响着晚点的发生。从因果角度粗略地将管理因素分为两个部分，一是运输系统内工作人员的失误导致晚点发生，二是晚点发生后工作人员处理不力，未能成功恢复。基于其对顶上事件的影响程度，后者更为关键，毕竟无法保证所有的晚点致因均不发生，但是合适的处理，例如运行图的缓冲时间与调度员的指挥，则有可能消灭尚未成熟的晚点现象。

（2）对于非管理因素，可以发现基本事件X_9“列车间相互作用因素”的模糊重要度甚至高于其余三个中间事件的模糊重要度。这是因为我国高铁开行密度大，计划性强，例如，京津城际铁路上高峰时段的班次间隔低于 10 分钟，最短发车间隔仅为 3 分钟。一旦列车的初始晚点未被及时恢复，列车间的相互作用就可能会使后续列车被迫降速、变更接发车股道影响到车站的正常组织，甚至影响整条线路上的列车行程。

（3）乘客因素在所有二级中间事件中的模糊重要度最低，这说明在乘客行为较小影响着列车晚点。其中，车内吸烟、误拉紧急制动阀（按下紧急按钮）、无故侵入线路等乘客违规行为的影响程度较大。

（4）环境因素对顶上事件有一定的影响，并且环境因素一旦发生，其晚点时长往往超过半小时，甚至可能造成一定的经济损失，可谓后果严重。这是因为，异物侵限与异物碰撞发生后，司机应立即停车并报告，同时要求后续列车停车、邻线列车限速行驶直至机械师检查处理完毕、确认安全或请求到支援为止；而狂风、暴雨、大雪等天气因素与地震火灾等自然灾害会严重威胁行车安全，列车需要降速通过甚至会封锁线路，禁止列车开行。

（5）系统因素是包含基本事件数量最多的中间事件。其中，列车故障对列车晚点发生的影响程度最高，牵引供电设备故障与电务设备故障次之，且二者影响程度相似，监测监控系统/设备故障与工务设备故障对晚点发生的影响程度则较低。

在所有系统因素中影响程度最大的是列车车载系统故障，主要包括列车自动监控系统(ATC)、列车自动防护子系统(ATP)与列车自动运行系统(ATO)。安全性是其需要保证的首要性能，因此，根据故障-安全原则，一旦运行过程中系统出现故障，列车往往会紧急制动并重启设备，导致列车受阻并晚点。

除此之外，设备故障也极大影响着晚点的发生，这是由于设备状态不明或数据异常均会触发 ATP 从而降速运行，甚至停车，同时还需要正确判断故障点并完成设备的维修与更换。其中，接触网跳闸停电影响最大，因为极限温度、异物侵限等多种原因都会导致其跳闸；同时，停电区域内的列车应立即停车，区域外的列车则被扣停；并且，供电部门需要前往现场检查、处理；即使跳闸重合或送电成功后，如果原因不明，后续首列列车也需要降速至 80km/h 运行。

4.4.3 临界重要度

根据公式（2.5），代入顶上事件的发生概率$p(T) = 0.03289$，得到各基本事件的临界重要度，计算结果与临界重要度排序如表 4.8 所示。

表 4.8 所有基本事件的临界重要度

编号	事件名称	$P(q_i)$	Q_i	C_i	C_i排序
X_1	乘客吸烟	0.0095	3.909E-05	1.129E-05	22
X_2	闲杂人员侵入线路	0.0336	1.417E-04	3.235E-04	14
X_3	车厢超员	0.0024	9.752E-06	3.415E-03	27
X_4	乘客误操作	0.0138	5.698E-05	4.310E-03	19
X_5	乘客突发疾病	0.0122	5.096E-05	5.563E-05	21
X_6	现场人员处理失误/不及时	0.2619	3.469E-02	4.193E-04	1
X_7	运行图不合理	0.0132	1.283E-03	6.062E-04	9
X_8	调度指挥人员处理不当	0.1363	3.653E-02	5.961E-05	3
X_9	列车间相互作用因素	0.2619	3.469E-02	4.193E-04	1
X_{10}	晃车、异响	0.0132	1.283E-03	6.062E-04	9
X_{11}	车体故障	0.1363	3.653E-02	5.961E-05	3
X_{12}	车载系统故障	0.6786	9.431E-03	1.211E-06	2
X_{13}	信号机异常	0.0498	2.136E-04	4.335E-06	11
X_{14}	轨道电路故障	0.1525	7.365E-04	3.172E-07	5
X_{15}	道岔定反位无表示	0.1695	8.362E-04	1.447E-04	4
X_{16}	列车占用丢失	0.0209	8.754E-05	2.063E-05	16
X_{17}	路基沉降超标	0.0565	2.441E-04	3.817E-05	10
X_{18}	钢轨重伤	0.0676	2.950E-04	9.318E-06	8
X_{19}	声屏障脱落	0.0217	9.035E-05	2.727E-03	15
X_{21}	TEDS 系统报警/故障	0.0032	1.245E-05	2.531E-05	26
X_{23}	接触网跳闸停电、受电弓冒火花	0.0059	2.416E-05	8.589E-06	25
X_{24}	断路器故障	0.0016	6.521E-06	2.352E-04	28
X_{25}	轨道停电	0.0126	5.385E-05	2.488E-04	20

根据基本事件临界重要度的数量级将其制作成六阶金字塔图，临界重要度越大，基本事件越重要，其所在阶梯越在上方，结果如图 4.5 所示。

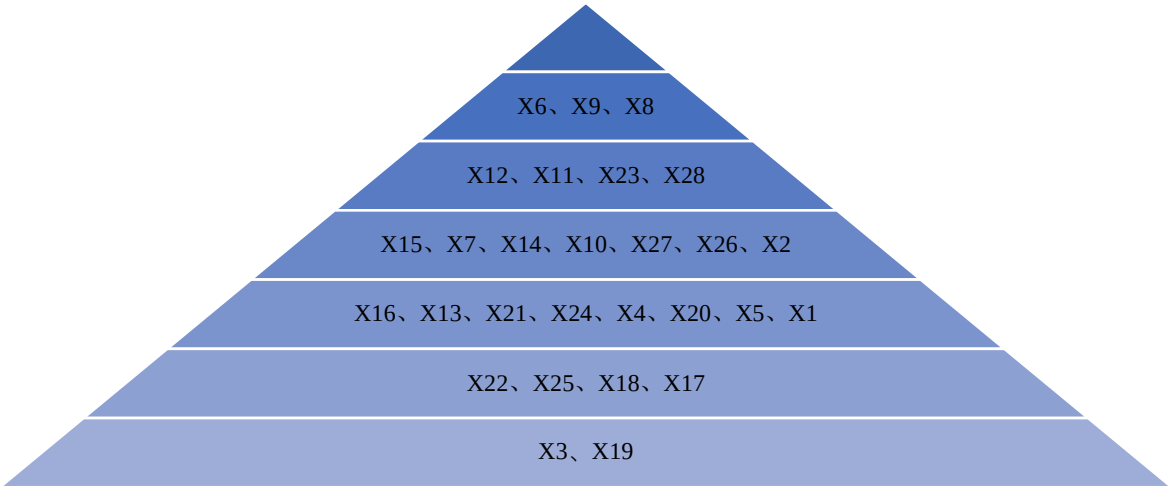

图 4.5 临界重要度金字塔图

观察图 4.5，基本事件X_6、X_9、X_8的临界重要度处于第一阶梯，X_{12}、X_{11}、X_{23}、X_{28}处于第二阶梯，X_{15}、X_7、X_{14}、X_{10}、X_{27}、X_{26}、X_2处于第三阶梯，X_{16}、X_{13}、X_{21}、X_{24}、X_4、X_{20}、X_5、X_1处于第四阶梯， X_{22}、X_{25}、X_{18}、X_{17}处于第五阶梯，X_3、X_{19}处于第六阶梯。因此需要更加关注对第一、第二阶梯内基本事件的治理。除此之外，与概率重要度相比，基本事件X_7的临界重要度排名明显下降，这说明该基本事件的危险度降低。

4.5 降低晚点率的有效措施

针对以上分析，可以将降低高铁晚点率的有效措施划分为晚点致因发生前与发生后两个节点来实施。当高铁列车正常运行在线路之中时，应当减小晚点致因的发生概率。当晚点致因已经发生，则需要及时发觉、采取恰当方式避免晚点传播并努力去恢复列车的晚点。本文将从乘客、设备、环境、管理与列车间相互作用五个方面提出降低高铁列车晚点率的有效措施。

尽管乘客因素的模糊重要度较低，但作为较好管控的因素，可以从改正乘客不良行为入手降低晚点概率，其主要措施包括：加强对旅客的高铁安全教育与宣传、通过广播或手机应用程序提醒乘客到站信息、加重对违规行为的处罚等。除此之外，为减少闲杂人员侵入线路，还可以在基础设施上设置安全警示标志。

无论是轨旁设备、车载设备还是基础设施，若要降低高速铁路列车晚点的发生概率，则一定要降低其发生故障的概率。因此，在设计阶段要选用恰当的型号与合适的技术，建设阶段保证安装建造的质量，运营阶段要时常检查与维护，后者可具体表现在动态的特制检查车、每日天窗时间的静态检查以及地面监控、列车数据检查与车载远程监控三位一体的综合监测方式上。

尽管人类不能控制天气与自然灾害的发生、无法预料小动物的行动轨迹，但是可以通过天气与灾害预测报警、异物侵限报警系统尽可能降低损失，这要求传感器稳定有效、系统准确可靠。同时，有关机构应设置各类成熟预案，做好预防措施。

从管理因素来讲，晚点事件的应对措施可以概括为列车运行图的优化和工作人员的培训。

当高速铁路列车由于各种晚点致因处于晚点状态时，列车运行图中所设置的缓冲时间在一定程度上可以阻断晚点传播。因此，可以通过改善列车运行图的方式降低列车间相互作用与横向传播的可能性，降低晚点的发生概率。在“一日一图”[33]组织模式的背景下，这就要求根据不同时期的客流与动车组资源尽可能平均分配列车密度、优化缓冲时间设置，以增强其抗干扰能力。

除此之外还应该增强对司机、随车机械师、乘务等工作人员的培训力度，以降低工作人员操作失误的概率并提高其对非正常情况的应急处理能力。例如，调度员需要了解列车从出库起至运行完毕再入库的每一个环节。这样，当晚点故障发生时，他们才可能综合考虑线路状况与技术设备、运行时间等数据，通过区间范围内列车赶点，站点范围内压缩停站时间、变更接车股道，变更到发顺序组织性，调度列车在备用线路上行驶等措施合理调度列车，降低晚点时长与后果，甚至避免晚点发生。

为降低列车间相互作用的影响，除去改善列车运行图之外，还可以根据以往的晚点数据挖掘列车运行调整与晚点分布的规律[34]，预测晚点恢复策略及其结果，为调度员提供参考。

装

4.6 本章小结

订

在本章中，通过对晚点影响因素的模糊事故树建模，确定出其基本事件与中间事件，并进行了定性与定量分析，验证了模糊事故树模型在高铁晚点事件上的适用性；同时得到高铁列车晚点的形成途径、控制途径；从结构、概率与临界三个方面认识到各类事件对晚点发生的影响程度；并在最后根据分析结果提出了降低高铁晚点率的针对性有效措施。

线

5 结论和展望

5.1 论文工作

本论文数据集为 2016 年至 2019 年 3 月广铁集团高铁列车突发事件处置过程与故障写实表，通过先验 LDA 模型建模与事故树分析，得到降低高铁晚点率的有效措施。论文的主要工作如下所示：

（1）基于先验 LDA 的高铁列车晚点影响因素特征提取。首先，人工筛选出数据集中造成晚点现象的记录；对记录进行文字提取、中文分词、去停用词等数据预处理工作，得到词项-文档矩阵。接着，通过相关性知识提取算法步骤，以卡方值表示晚点词项与晚点致因模式之间的相关关系，得到表示为相关性矩阵的先验知识。然后，将这一先验知识融入在 LDA 模型的主题更新公式之中，以指导 LDA 主题模型完成晚点影响因素的语义级特征提取工作。最终，通过整合先验知识的 LDA 主题模型算法步骤，获得高铁列车晚点影响因素主题-词分布，方便进一步理解晚点致因之间的相互关系，方便接下来事故树的建立与分析。

（2）基于模糊事故树的高铁列车晚点影响因素分析与应对措施。首先，确定顶上事件为高铁列车晚点，结合从先验 LDA 模型中获得的晚点影响因素主题-词分布、文献综述与专家知识，建立事故树结构。接着，基于统计数据与专家评分获得各基本事件发生概率的三角模糊数表示，完成节点模糊处理工作。然后，对事故树进行最小割集、最小径集与顶上事件模糊概率的求解，计算各基本事件的结构重要度、模糊概率重要度与临界重要度，对这六个指标进行分析，得到高铁晚点的形成途径、控制途径与各晚点致因对顶上事件的影响程度。最终，基于以上分析提出降低高铁晚点率的有效措施与治理优先级。

本文的研究结论如下：

（1）从中间事件层面来讲，导致高铁晚点事件发生的最大因素是管理因素，包括调度处理、运行图设置与现场工作人员的工作均对晚点发生概率产生较大的影响；包含基本事件数量最多的因素是系统因素，该方式引起列车晚点的途径最为多样；而环境因素一旦发生，其造成结果较为严重；应当高度关注这几个方面。从基本事件层面来讲，除去上述已经提到的事件外，还应当关注列车间相互作用因素，其模糊概率重要度甚至大于某几个中间事件的模糊概率重要度。

（2）提出降低高铁列车晚点率的有效方式。这些方式可以粗略分成两个时间节点去实施，晚点致因发生前与晚点致因发生后。当高铁列车正常运行在线路之中时，应当减小晚点致因的发生概率，其措施包含加强对旅客的安全教育力度、对设备与基础设施的维护与检修、对工作人员的培训力度和设置应对各种突发事件的预案等。当晚点致因已经发生，则需要及时发觉、采取恰当方式避免晚点传播并努力去恢复列车的晚点，主要内容包括传感与报警系统的完善、列车运行图的改善和调度员的适当调度等等。这些措施还可以概括为乘客、设备、环境、管理与列车间相互作用五个角度。

（3）关注属于临界重要度金字塔图中的第一、二阶梯的基本事件，包括现场人员处理失误/不及时、列车间相互作用因素、调度指挥人员处理不当、车载系统故障、车体故障、接触网跳闸停电、受电弓冒火花、异物侵限，应当优先提出这些因素的针对性治理手段。

5.2 不足与展望

针对高速铁路列车晚点论题，本文通过先验 LDA 模型完成了特征提取，基于模糊事故树建模对晚点影响因素进行了分析，并提出了针对性措施。尽管获得一些成果，研究过程却仍存在不足：

（1）作为较为重要的晚点影响因素，列车间相互作用因素包含多种类型，然而本论文的数据集无法满足细分类型的需要。因此，在以后的工作中，希望可以通过更新数据集完善论文，获得更加精确的结论。

（2）在通过三角模糊事故树对高铁列车晚点影响因素进行分析时，采用 3 σ 表征法确定不可统计的基本事件的发生概率，评分来源自三位专家，具有一定的主观性。在未来的工作中，可以通过增加专家位数来降低主观影响，或直接采用更适合的评分方式。

参考文献

[1] 陆娅楠. 我国高铁运营里程超 4 万公里[N]. 人民日报, 2021-12-31(001).

[2] 吴漫云. 航空—高铁竞合关系分析[J]. 民航管理, 2019(10):11-15.

[3] UIC450-2. Assessment of the performance of the network related to rail traffic operation for the purpose of quality analyses - delay coding and delay cause attribution process[S]. Paris, France : International Union of Railways, 2009.

[4] John Preston et al. Impact of Delays on Passenger Train Services : Evidence from Great Britain[J]. Transportation Research Record, 2009, 2117(1) : 14-23.

[5] Nadjla Ghaemi et al. Impact of railway disruption predictions and rescheduling on passenger delays[J]. Journal of Rail Transport Planning & Management, 2018, 8(2) : 103-122.

[6] J Wang, Granlf M , J Yu. Effects of winter climate on high speed passenger trains in Botnia-Atlantica region[J]. Journal of Rail Transport Planning & Management, 2020.

[7] 翟恭娟. 高速铁路列车运行调整优化研究[D]. 西南交通大学, 2013.

[8] 汪静，彭一川，陆键. 基于 ISM 的高铁列车晚点影响因素分析[J]. 中国铁路, 2020, (01):48-52.

[9] 纪媛媛. 基于特征选择与机器学习的列车晚点预测方法研究[D]. 北京交通大学, 2020.

[10] 石睿. 基于数据驱动的高速铁路列车晚点分析及预测方法研究[D]. 北京交通大学, 2021.

[11] Allahyari M , Pouriyeh S , Assefi M , et al. A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques[J]. 2017.

[12] Hofmann T. Probabilistic latent semantic indexing: Proceedings of the 22nd Annual International SIGIR Conference. New York: ACM Press, 1999:50-57.

[13] Blei D ,Ng A, Jordan M. Latent Dirichlet Allocation .[J]. Journal of Machine Learning Research, 2003, 3 : 993-1022.

[14] Chemudugunta C , Holloway A , Smyth P , et al. Modeling Documents by Combining Semantic Concepts with Unsupervised Statistical Learning[J]. Springer-Verlag, 2008.

[15] Allahyari M , Kochut K . Semantic Context-Aware Recommendation via Topic Models Leveraging Linked Open Data[J]. 2016.

[16] Blei D M , Griffiths T L , Jordan M I , et al. Hierarchical Topic Models and the Nested Chinese Restaurant Process[J]. Advances in neural information processing systems, 2004, 16.

[17] Blei D M , Mcauliffe J D . Supervised topic models. In NIPS, 2007, Vol 7.121-128.

[18] Petinot Y , Mckeown K R , Thadani K . A hierarchical model of web summaries[C]// The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA - Short Papers. DBLP, 2011.

[19] X Mao, Z Ming, T Chua, et al.; SSHLDA: A Semi-Supervised Hierarchical Topic Model[C].EMNLP, 2012.

[20] 王峰. 基于文本挖掘的高铁车载设备故障诊断方法研究[D]. 北京交通大学, 2016.

装 订 线

[21] Griffiths T L, Steyvers M. Finding Scientific Topics[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(S1): 5228-5235.

[22] Teh Y, Jordan M, Beal M, et al. Hierarchical Dirichlet Processes [J]. Journal of the American Statistical Association, 2007, 101(476): 1566-1581.

[23] 曹娟，张勇东，李锦涛，等. 一种基于密度的自适应最优 LDA 模型选择方法[J]. 计算机学报，2008，31(10)：1780-1787.

[24] Arun R， Suresh V， Madhavan C E V , et al. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations[J]. Springer-Verlag, 2010.

[25] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin.Automatic evaluation of topic coherence. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 100–108. 2010

[26] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum. Optimizing semantic coherence in topic models. In Proc. of the Conf. on Empirical Methods in Natural Language Processing, pages 262–272. 2011.

[27] Michael Roder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In the Eighth ACM International Conference on Web Search and Data Mining, pages 39– 408.

[28] 税昌锡. 语义特征分析的作用和语义特征的提取[J]. 北方论丛, 2005(03):66-70.

[29] Calinski, T, Harabasz, et al. A Dendrite Method for Cluster Analysis[J]. Comm in Stats Simulation & Comp, 1974.

[30] 胡思继. 列车运行图编制理论与方法[M]. 2013.

[31] 袁强. 高速铁路列车晚点分布及传播模型研究[D]. 北京交通大学, 2020.

[32] 人民网. 正点率达 98.9% 累计 50 亿人次坐高铁出行[EB/OL]. (2016-7-22)[2022-5-7]. http://m.cnr.cn/news/20160722/t20160722_522759873.html.

[33] 罗强. 面向“一日一图”的列车运行图与动车组交路协同优化研究[D]. 北京交通大学, 2021.

[34] 李津，文超，杜雨琪，徐传玲. 基于梯度提升回归树的武广高铁区间晚点恢复策略研究[J]. 中国铁路, 2021, (10):76-84.

装 订 线

谢 辞

时光荏苒，岁月如梭，伴随着新冠肺炎，属于我的四年大学生活一晃就过去了。尽管由于疫情有两个学期未能实现线下上课，我仍然非常感激母校对我的培养，感激母校给我提供的一切资源，感谢这个校园。哈佛校长曾这样谈论大学教育：“好的教育之所以好，是因为它让你坐立不安，它强迫你不断重新认识自己和周遭的世界，从而不断做出改变。”在远离家乡求学的这四年中，我在同济校园中结识到志趣相投的朋友，遇到生活与学业上的恩师，他们的出现提升了我的眼界，也塑造了我为人处世的方式。

感谢导师邢莹莹在完成毕业设计过程中给予我的支持。从论文开题、文献综述、资料收集、算法建模、结果分析到最终毕业论文的定稿，邢老师都给了我很多建议，给予我很大的帮助。由于疫情，每周的开会都是在线上完成的，以致到目前为止我已经半年多未与邢老师见面了。不久之前，我才了解到邢老师刚刚完成生产，而在这之前，她从未像我透露过分毫。因此，我十分敬佩邢老师的学术精神与园丁精神。

感谢班主任黄世泽老师、辅导员许青老师、整个 2018 届交通信息班的同学与陪伴我校园生活四年的舍友这四年来对我的照顾与帮助，是你们见证了我这四年的喜乐哀怨，见证了我的成长。

最后，感谢我的父母。从高考结束，我的父亲就对我讲，自己的人生自己做主。因此，无论是学校还是专业，他们都从未干涉我的选择，同时成为了我坚强的后盾。在此，我十分感激父母对我的培养、信任与支持！

装 订 线