基于关联规则挖掘的重大事故属性相关性评估模型

【摘 要】识别严重事故的属性之间的相关性对预防它们可能至关重要。如果能动态地了解这种关系，就有可能对事故采取预防措施。本文的目的是开发一种适用于每种类型的数据分析模型，以创建适用于任何计算系统的预防措施。本模型以连贯的方式集体地显示了属性之间的关系，以避免严重事故。在这方面，关联规则挖掘被用作识别属性之间相关性的技术。本研究采用实证主义的方法，通过个案研究和定量测量，客观地坚持九种不同事故类型的事实知识。以九种不同类型的施工事故为例，验证了该模型的适应性。结果表明，每一种事故类型具有不同的属性组合特征，并通过数据集实现了适应变化的分析模型。最终，专业人员可以有效地识别因果关系，制定预防措施，打破事故原因因素之间的联系。

引言

由于技术的重大发展，建筑行业已经演变成难以置信的复杂结构。这种发展带来了一些严重的问题，换句话说，复杂性增加了职业健康和安全方面的问题。纵观历史，人们曾试图采取预防措施，以预防死亡或严重伤害。在这方面，职业健康安全正成为一个需要详细调查的重大问题。

2015年，欧盟国家共发生320万起非致命事故，每一起事故造成至少4个工作日的损失，3786起致命事故(欧盟统计局，2015年)。此外，2017年美国发生280万起非致命事故，导致882730个工作日损失，5147人死亡(美国劳工统计局，2017年)。根据欧盟统计局(2015)和美国劳工统计局(2017)的数据，建筑业的致命职业事故发生率分别为21%和20.7%。这些统计数字表明事故发生率如此之高，不容忽视。

建设项目具有不同的属性，如范围、地理条件和风险因素。这样的变化使得每个项目都是独特的，而且过于复杂，无法创建系统的方法来解决工业问题，如上所述，建筑中最大的挑战之一是减少甚至避免对人或环境的伤害。虽然没有简单的方法来防止安全故障，但意外是可以通过充分执行下列程序来避免的。这些是确定导致风险的因素，并准备相应的可行风险评估，采取必要的预防措施来控制风险评估的结果，创建和应用严格的安全程序，并利用从系统中吸取的教训创建一个可持续的环境。几十年来，研究人员也为此付出了巨大努力。

预防事故的第一步是准确地确定导致职业事故的属性。各种方法已经被应用于确定哪个属性或属性组导致事故。例如，Mohammadi等人在审查了侧重于探索安全因素的论文后，应用定性内容分析对建筑事故的潜在原因进行了分类，他们开发了一个层次结构框架来衡量这些属性的重要性。为了通过关联规则挖掘来调查建筑工地职业事故的成因因素之间的相关性，之前已经进行了一些研究。其他一些研究建立了预测模型，并引入了无法适应数据变化的简单预防措施。目前研究的不足之处在于它们需要通过个体检查来捕捉潜在的风险。这意味着模型的适应过程不是很好。此外，现有的模型还存在着对主观获得的数据有使用问题。职业健康安全专业人士将他们的意见放在记录过程中，所以现实往往与之背道而驰，此外，这些研究没有就事故的个别类型或它们的整体相似性提出质疑。

与文献相反，所提出的研究通过明确检查最常见的建筑工地事故来填补文献中的空白，从而对文献做出了贡献。本研究旨在引入一种可适应的分析模型来提取属性之间的相关性，从而诱发建筑事故。它推出了一种基于属性的预防系统，可以自动应对任何类型的问题。

通过案例研究，结合九种不同事故类型的事实知识，建立了一个基于臂的分析模型，以显示属性之间的联系。使用关联规则挖掘的目的是，手臂是一种流行的数据挖掘算法，它产生了重要的关系随数据结构。工作逻辑依赖于频繁项集挖掘，它可以成功地与事件数据合作，因为数据集封装了大量具有可变观察率的属性。此外，无论事故类型如何，都进行了额外的分析，以获得对高严重事故的全面了解，并验证所提出的模型对不同数据集的适应性。最终，该研究揭示了导致这些不同类型事故的属性之间有意义的关系。在这项研究的帮助下，职业健康安全专业人员可以更具体地关注每一种事故类型，并通过采取必要的预防措施，将识别出的显著相关性最小化危险后果。此外，从不同类型的事故中得到有意义的结论，也证明了模型对任何计算系统的遵从性。此外，研究结果及建议的方法将与以往研究的事故预防系统进行比较。

1.在文献和实践中的职业健康和安全挑战

几十年来，事故后果的严重性一直吸引着研究人员的注意。他们投入了大量的精力，通过识别事故的属性来发现事故的特征。理解事故触发属性之间的潜在相关性将为预防建筑工地常见的与工作相关的安全故障提供巨大机会。

研究人员根据安全风险、安全表现和管理、具体事故类型的研究和分析模型的发展等热门题目，研究建造业的安全问题。虽然他们的重点是防止事故，但研究的方法和解决问题的方法往往在每一项研究不同。这些研究已经建立了许多关于安全问题结果的分析或专家模型，但所提出的或构建的模型的成功取决于对属性之间的相关性的推导。下面的部分显示了理解事故案例的内在结构的重要性，无论研究的主题或焦点是什么。

**1.1安全风险**

研究人员最关注的话题之一是基于建设项目的安全风险。Camino López等人研究了发生在西班牙的事故。他们调查了影响属性之间的关系，并探讨了这些属性如何影响事故的严重程度。在另一项研究中，工作类型与事故类型之间的关系是相关的，并对它们之间的相关性进行了详细的调查。Aminbakhsh等人在职业健康安全专家的帮助下，采用层次分析法对安全风险要素进行排序。提出了另一种安全风险评估模型，分析不同安全风险水平下的不同施工现场布局。研究旨在调查建筑项目利益相关者与职业健康安全专业人员之间的安全与风险认知的相似性。正如上述研究所讨论的，了解属性之间的相关性对于主动预防事故至关重要，因为职业健康安全专业人员应该学习如何克服安全风险，以及如何提前管理风险评估。

**1.2 安全绩效与安全管理**

施工现场的安全绩效和安全管理是施工安全的另一个重要课题。与安全风险过程类似，解释属性之间的相关性对于满足更好的安全性能和安全评估变得至关重要。Choi等人提出了一种方法来确定可穿戴传感器的有效性，该传感器测量工人的生理反应。研究表明，在低风险活动和高风险活动中，工人的反应有显著差异。在另一项研究中，讨论了无人机用于安全检查的适用性。该研究基于无人机拍摄的照片的分析，以捕捉危险的工作条件，然后用于实施预防措施。

实施安全管理系统是满足施工现场安全环境的关键因素。足够的安全管理系统要求对造成事故的属性进行全面的调查。Yiu等人研究了香港建筑业实施安全管理系统的效率，Kheni等人对发展中国家进行了类似的研究。根本原因分析通常用于探索问题背后的核心原因。贝叶斯网络和社会网络分析用于确定致命事故的根本原因。另一方面，Shao等人发现致命事故多发于夏季、周一以及10点-11点和15点-16点的时间段内，其中有一半是高空坠落。此外，研究表明，由于她们的生理和职业的不同，妇女在工作场所处理职业危害的方式很独特。

从事故调查的结果中吸取的教训促进了安全表现的显著进步。在这方面，安全培训开始在事故预防中发挥重要作用。几项研究对安全培训的有效性提出了质疑。提供安全培训是向员工传授安全理论知识，树立职业安全意识的最有效途径。Evanoff等人为缺乏经验的建筑工人设计了一项培训计划，以增强他们的跌倒预防知识。此外，还提出了一些提高少数民族工人安全绩效的策略。Olson等人为主管开发了易于使用的工具箱材料。除了它们的有效性外，还审查了在全球范围内实施安全措施的困难。安全领导是影响SMS正确实施的重要因素。一项关于在铁路项目中如何通过有效的安全领导提升安全绩效的调查，以及Grill和Nielsen提供了详细的指导，使安全从业者能够积极影响安全绩效。

1.3 具体事故类型的研究

研究人员已经开始关注建筑工地最频繁和最严重的事故，以降低死亡率和严重伤亡率。从高处坠落是建筑行业中最常见的事故类型。因此，人们进行了许多研究来调查和减少坠落事故的数量。Winge和Albrechtsen评估了事故的类型，以确定最常见的事故，并分析了障碍故障。该研究主要集中在近端属性上，但存在潜在的相互作用和相关性，需要进行更深层次的研究。在另一项研究中，试图找出美国致命坠落事故背后的原因，并提出可行的解决方案。研究中使用的事故数据源没有提供有关事故的一些关键因素的信息，如工作条件和环境以及安全气候因素。因此，结果可能不能反映事故的所有方面。与坠落相关的设备，包括桅杆攀爬工作平台也对其稳定性进行了评估。还研究了干墙安装人员的坠落事故，以识别其特征。

失去平衡与工人的姿势稳定性有关;因此，我们确定了平衡损失的相关因素，并使用惯性测量单元传感器对建筑工人进行姿势稳定性分析。除了这些分析，还提出了一项长期研究，旨在提高工头培训的有效性，以减少坠落事故。

有一些研究关注的是坠落事故之外的其他特定类型的事故。还调查了电力事故和土方设备相关事件。Suárez-Cebador等对13个显著性水平低于0.05的不同属性进行了电力事故调查。然而，这项研究并没有涉及导致建筑工地事故发生的人为因素或工作场所因素。Tokdemir和Ayhan研究了异物损伤，并开发了ANN-AHP混合模型来预测事故的严重程度。在预测过程中，研究了这类事故中最常观察到的属性，以帮助专业人员采取必要的预防措施来预防事故。虽然许多研究人员分别论述了建筑事故的问题，但还没有研究将一种以上的事故类型的调查概括为其共同点和不同点。原因是，事故可能具有某些特征，可以是普遍的或独特的，可能需要进行重大的分析来揭示。所提出的方法旨在揭示各种事故类型的属性之间的相关性，因此在本文的剩余部分将给出最常见事故的不同点和相似点。

1.4 建筑安全分析模型

人工智能(AI)是基于创造超级机器模仿人类大脑并处理复杂信息的想法。Patel和Jha提出了基于人工神经网络(ANN)的模型，用于评估安全气候效率和评估员工行为的安全水平。结合模糊神经网络技术和多元统计分析方法，确定俄罗斯的环境安全水平。选择卷积神经网络和长短期记忆网络检测施工现场的安全和不安全行为。另一项研究侧重于通过深度学习检查工人的认证，可以识别每个工人的脸，并检查他是否拥有有效的认证。采用模糊德尔菲法和决策试验评价实验室相结合的方法来发现职业事故的影响因素及其相互关系。Mohandes和Zhang开发了一种基于混合模糊的综合风险评估模型，以消除现有安全管理体系方法的缺陷。Ayhan和Tokdemir也提出了一种模型，该模型有人工神经网络和模糊逻辑组成，用于预测事故的后果。该模型可以处理从真实事故中收集的数据，使职业健康安全专业人员在施工前和施工阶段采取必要的预防措施，防止工伤。Ayhan和Tokdemir改进了他们之前的研究，采用聚类技术来获得更好的预测精度。作为最后一步，他们引入了一个启发系统，作为在致命事件中观察到的属性频率的预防措施。虽然研究结果对预防事故有希望，但应根据不同的分析模型进行更详细的调查，以获得更可靠的结果。

基于分析模型的基于属性的安全评价策略也是目前研究的热点之一。文献对建筑事故属性之间的相关性进行了研究。建设项目的备案机制大多是主观的，专家的个人意见可能会诱发现实。Esmaeili等人提出了一种基于属性的风险评估系统，该系统使用了1812起撞击事故伤害报告的可靠国家数据库。它们的目的是克服现有研究中要求个别评价和调查的问题。所介绍的方法只针对撞击事故的施工事故。虽然这项研究仅限于评估特定类型的事故，但它促进了加强主动安全管理策略。

Esmaeili等人推进了他们的研究，该研究概括了基于广义线性回归分析的预测模型的开发。专家可以主动预测项目的风险级别。尽管有这些贡献，该研究在某些方面可能仍存在不足。拟议的系统在动工前无法实施。此外，它只处理意外事故。

Mistikoglu等人进行了决策树分析，这是一种监督数据挖掘方法。该研究调查了美国职业安全与健康管理局的一起高空坠落事故。结果描述了属性之间的关联。坠落距离、伤亡原因、安全训练、施工操作等因素是导致坠落事故增加的主要因素。

现有研究的局限性可以描述为，如前所述，模型依赖于主观记录的数据。为了解决这一限制，Tixier等人提出了一个框架，包括自然语言处理，以标准化事件的基本属性。目的是从多维数据挖掘中获得新的安全知识。从大量的数据中获得了少量的相关观察结果。

之前提出的显示了可能的结果，实现了标准的方式，提取有价值的见解，从伤害报告应用无监督的方法。研究人员只关注了三种不同类型的事故，包括被物体击中、在同一水平面上坠落、夹在物体之间。它可能缺乏对剩余事故类型的洞察。

关联规则挖掘是另一种用于指示属性之间关系的有用技术，它是一种应用于建筑行业的方法。Lin和Fan考虑990个公共建设项目的检查等级，确定缺陷类型与检查类型之间的关系，并结合遗传算法和关联规则挖掘提取缺陷模式。还调查了2000年至2007年台湾和韩国发生的职业事故和死亡报告。Verma等人使用关联规则对印度一家钢铁公司2010年3月至2013年7月期间导致事故发生的经营因素进行了调查和识别。Guo等将预先开发的结构化学习算法与加权关联规则相结合，创建了一个融合模型，生成一个危险关联网络。Liao等将人的动力学分析与关联规则挖掘相结合，寻找中国武汉某地铁施工现场工人不安全行为时间分布的时间统计规律。关联规则用于调查导致交通事故的因素。

此外，在路易斯安那州发生的工作区域撞车事故、车辆与行人相撞事故以及在雨天发生的撞车事故都是明确的重点。其他研究人员试图发现中国和伊朗铁路事故的有意义的关系、模式和趋势。同样，Zhang和Liu建立了海上交通事故初级数据库，并利用关联规则来确定事故背后因素的依赖性。

本研究明确地将重点放在建筑工地常见的事故上，以弥补上述文献中所述的缺失。其目的是证明如何通过揭示这些属性之间的相关性来预防严重事故。

2 方法研究

研究的重点是一致性地检验属性之间的相关性，以避免严重事故的发生。调查了九种不同事故类型所涉及的因素，并提出了一个分析模型，通过控制各种属性组合来揭示每种事故类型的特征。该研究有三个阶段，如图1所示。首先进行数据准备过程。事故记录是从同意分享事故信息的建筑公司获取的。然后，利用关联规则挖掘分析对数据集进行分析建模。在最后阶段，对具有强联系的相关性进行解释，以使对每种严重事故提出必要的预防建议。

2.1 数据预处理

研究首先界定了建筑工地最常见的意外类型，图2显示了每一宗意外。作者联系了主要的建筑公司，他们在欧洲和亚洲地区有几个建筑工地。关于这些国家的事件记录汇总如下。事故记录中最重要的部分来自俄罗斯和土耳其，分别占整个数据集的36%和31%。波兰占9%，土库曼斯坦占11%。数据集中的其余国家是伊拉克、沙特阿拉伯和哈萨克斯坦，它们的比例分别为7%、3%和3%。两家公司同意在保密的基础上分享它们高度严重的事故文件，包括过去和现在的记录;因此，本研究未涉及任何个人或组织信息。收集事故数据后，对事故记录进行过滤，确保所有事故都属于急救、医疗干预、工作日损失或死亡类别的高严重级别。再次过滤后，保留3870条事故记录作为输入关联规则挖掘的数据。

数据准备的下一个步骤是组织描述事故的属性列表。统计分析需要一个具有序数或名义变量的有意义的表达式，以便计算模型能够进行必要的计算以从数据集提取结果。文献中有重要的研究可以合理地解释事故案例。因此，首先根据Ayhan和Tokdemir确定属性类别。然后，在OHS专家的帮助下，对属性进行重新排列，并结合Ayhan和Tokdemir的研究，形成一个更加连贯的列表，并将另外三个类别(涉及事故人员的职业、年龄和经验持续时间)纳入本研究。

表1描述了从上述研究中获得的属性。将它们编码为二进制格式的原因是，基于关联规则挖掘的分析模型可以很容易地适应于二进制表达式。此外，在一个情况下可以同时观察到多个属性，因此该方法允许对它们进行分类表示。因此，如果事故案例中存在表1中指定的属性之一，则赋值将变为1。还有一些连续和类别属性需要转换为二进制表达式。表2显示了关联规则分析前的属性组及其特征。这些也应该通过编码为1和0转换成二进制格式。这是通过将属性划分为子组来实现的，如表2所示，以适应定义的选项。为了说明这一点，当谈到年龄时，我们创建了四个子属性来反映四个年龄组。

执行这些操作后，属性类别的总数增加到9个，这些类别总共由74个属性组成。属性类别和属性的详细信息如表1所示。因此，9个不同的数据集以及所有数据集都可以在关联规则挖掘中进行分析，如图1所示。

2.2基于关联规则挖掘的分析模型开发

关联规则挖掘是Agrawal等人为了调查客户的购买倾向而引入的一种数据挖掘技术。它通常用于确定隐藏模式和数据集中变量之间的关系以及这些变量之间的依赖关系。在本研究中，作者使用了Apriori算法，这是关联规则中应用最广泛的算法之一。算法由两个阶段组成。在第一个阶段，分析从数据集中捕获最经常观察到的项目或属性，然后通过生成属性之间的逻辑关联规则来推进分析。图3为本研究采用的关联规则算法的步长。基于Rapidminer Studio 9.2.0软件对算法步骤进行了演示。本研究使用Rapidmainer Studio和RStudio(2019)。

规则以“X→Y”的形式定义和呈现，其中X是前件，Y是后件。换句话说，这些规则保留了if-then语句的特点，其中if和then子句分别代表先行词和后句。  
在生成关联规则时，支持度、置信度和提升度是用于识别最强规则的阈值，如图3所示。支持度表示该项在数据集中有多常见，置信度显示生成的if-then关联为真有多频繁。最后，提升度表示关系有多强，提升度值越高，关系越好。

支持度、置信度和提升度用方程(1)、(2)和(3)计算。

其中N为数据集中项目的总数，F(X,Y)为项目的频率，同时包含X,Y,F(X)和F(Y)分别为X,Y的频率。

如前所述，使用RapidMiner Studio 9.2.0和RStudio进行关联规则分析，对于所有事故类型，支持度、置信度、提升都分别为0.15，0.6，1.1.采用Gephi0.9.2软件获取图像。

2.3 解释的结果

研究的一个重要部分是解释关联规则分析，以防止在建筑工地发生职业事故。如上所述，关联规则分析的结果由前因和后因两个要素组成。这些属性之间的规则表明它们之间有很强的相关性，并据此对施工事故进行观察。如果通过采取必要的预防措施，打破属性之间构成的规则，就可以避免事故的危险后果。因此，所有9种事故类型的结果都是根据其提升度进行排序的，这表明前件和后件之间的相关性有多强。同时对所有事故类型实施关联规则，已检测每种事故类型规则的变化。

3 讨论结果

3.1 包括所有事故类型的数据分析结果

通过寻找事故属性之间的相关性，用关联规则评估事故属性之间的关系。事故数据由九种最常见的事故类型组成，如图2所示。将关联规则应用于这9种事故类型，用于完整的事故数据观察属性之间关系的偏差。首先，实现对完整数据集的关联规则来捕获不同事故类型的公共点。然后，对九个事故组进行了分析，以获得对个别高重特大建筑事故的更详细的解释。在开始分析之前，作者确定了置信度、支持度和提升度的阈值标准分别为0.6、0.15和1.1。

分析始于对事故的理解。为此，计算了属性的频率。图4显示了3870个事故案例中超过15%的属性条目频率。结果表明，RB-1、HA-1和RB-3是观测到的最深刻的属性。虽然这一数字不能表明统计结论，但它提供了一个线索，说明哪些属性可能需要有相当程度的意图来衡量预防性行动。此外，预计关联规则产生的规则将与经常观察到的属性高度相关。

当支持度为0.15，置信度为0.6时，利用关联规则分析生成202条规则。结果表明，从不同事故类型中获得的202条规则中，有160条规则最多只对一种事故类型具有唯一性。换句话说，160条规则不会出现在一种以上的事故类型中。分析结果表明，事故的多重因果性也体现在复杂的压载操作中。因此，每一项都应该单独进行调查。并将整个数据集的分析结果与个别事故分析结果进行对比，考察事故类型是否存在偏差。结果显示，12条规则中没有一条在9种事故类型中都出现，而至少有两条规则在2种事故类型中出现。因此，可以肯定，单独评价事故类型更有益，因为根据所有数据分析的结果采取预防措施可能不足以有效避免事故。

图5解释了规则是如何形成的。没有标签的节点表示规则。根据事故类型，将表示属性之间连接的边缘用不同的颜色进行分类。由于事故有共同之处，它们也有明显的特征，如图5所示。带标签的节点表示属性，并根据大小和颜色进行标记。尺寸越大、颜色越深的事故类型支持值越高，观测率越高。如上所述，规则与频率大于0.5的属性高度相关。

下一步是解释提升度、支持度和置信度。图6显示可置信度和支持度的散点图。图上的点用不同的形状来表示事故类型。颜色刻度表示置信值。从图中可以看出，大多数规则发生在支持度在0.15-0.2之间，而提升度在1.1-1.3之间变化。图6表明在这些范围内获得了最高的置信度。由于所选提升度足以获得高置信度规则，因此应用提升度应等于或大于1.1的规则对关联规则结果进行过滤。这意味着在解释前因后果之间的相互关系时考虑了提升度大于1.1的规则。

排除提升度较低的规则后，所有事故类型只剩下12条规则。图7显示了过滤后的规则，并考虑了已定义的标准。规则用气球图表示，气球的大小代表频率。前因和结果被标记为LHS和RHS，因此解释规则之间的关系变得更容易。图8和表3进一步说明。图8中草拟了过滤规则的网络可视化，以便更有效地表达属性之间的关系。此外，为了更好地理解，表3给出了逻辑规则的文本表达式。

根据这些结果，可以得出以下关于所有事故类型的一般结论。“无法感知外部风险”的RB-1是事故数据和所有类型事故中最常见的属性，且事故发生频率最高的时间段是9:00-12:00和21:00-00:00。不完善的事故分析系统可能导致无法感知和评估外部风险，以及培训方面的问题，从而导致违反安全工作政策。究其原因，是对问题的根源没有深入调查，事故的教训没有很好地传递给工人。当工作负荷不平衡时，这种违规行为会导致人们倾向于走捷径，结果导致了严重后果的事故发生。

3.2 每种事故类型的个别分析结果

为所有事故类型实现了关联规则，以捕获关于属性之间相互关系的更多信息。所有事故的属性之间的潜在关系已经被揭示出来，这揭示了哪些连接应该被打破，以防止事故。之后，对每一次事故都进行同样的处理，得到的结果如下:表4总结了关联规则生成的规则，并根据事故类型进行分类。分析全面探讨了每一次事故的特点。它们所有分析和个别分析之间的重要区别是置信值，它代表了所描述条件的发生的可能性。与表3中的规则相比，单独的关联规则分析产生了更高的置信水平。几乎所有的事故类型都包含置信水平高于0.80的规则。这意味着分别应用关联规则可以更精确地推导出每个事故最有可能的触发点。

对于夹在两个物体之间的事故，无法感知外部风险一般被观察到是关联规则的结果。建筑设备操作员和有生理残疾的工人无法察觉工地上的危险。此外，这类事故发生时，有12至24个月工作经验的工人正在建筑工地进行日常工作，如果工作环境不合适，他们无法评估外部风险。使用捷径通常涉及日常活动，当这种情况与过度的工作负荷相结合时，事故就会发生。

接触化学品的事故多发生在21:00 - 00:00之间。在没有充分评估外部风险的时间间隔内，违反了安全工作政策。这类事故的受害者大多是年龄在35岁至45岁之间的雇员，他们在工作环境不适宜时接触了这些化学物质。此外，在钢筋/模板安装过程中出现了不正确的物理移动。当这些类型的运动与不平衡的工作相结合时，也导致了不合适的工作环境。此外，如果在工作场所缺乏沟通，就违反了安全工作政策。

如果分别涉及以下几个属性:教育程度不高、技能和感知能力不强、从业人员经验6 - 12个月，则无法正确判断和评估从高处坠落的外部风险。此外，当手工工具或施工设备使用不符合要求时，违反安全作业规定。如果使用不合适的个人防护装备时更倾向于走捷径，也会违反这些政策。此外，在控制或跟踪不足、违反安全工作政策的情况下，以及在存在不平衡的工作量和与教育水平相关的问题的情况下，会使用捷径。此外，缺乏监督和无法感知外部风险导致了不平衡的工作量。

由于工作环境不适宜，事故分析体系不健全，无法评估同级事故类型坠落的外部风险。此外，在9:00-12:00的工作期间，无法感知这些风险最有可能表明工作区域不合适。由此可以推断，在不合适的工作环境下，外部风险无法很好地理解，反之亦然。

在火灾或爆炸事故类别中，在21:00至00:00之间观察到违反安全工作政策和不平衡的工作量。这类意外的受害者一般在45至65岁年龄组，而且工作负荷也不均衡。此外，建筑应用程序中与程序有关的问题导致无法评估外部风险和不平衡的工作量。

异物损坏通常是由于只有3到6个月经验的工人无法感知和检查外部风险。此外，在被飞行或下落物体击中时，由于不正确的物理运动和施工过程中的问题，导致无法识别外部风险。在钢筋或模板安装过程中，由于无法感知外部风险，加之工作环境不当，往往导致事故发生。

被移动的物体击中也会造成危险的后果。不正确的安全政策或缺乏安全工作政策导致无法评估外部风险。对于需要使用手动工具或其他设备的工作，不完善的事故分析系统也获得了这一结果。此外，如果使用手动工具或受害者的工作量不平衡，就违反了安全工作政策。

被物体撞击是最常见的事故类型，结果表明，它通常发生在高温作业中，违反或没有安全作业政策是极其危险的。技能不足、无法感知风险和安全措施不足通常会导致违反工作安全政策。与其他类型的事故一样，缺乏沟通也导致无法感知外部风险。

结果表明，根据事故类型的不同，前因和后果的组合发生了显著的变化。尽管乍一看，结果似乎很熟悉，考虑到如此多类型的事故，捕捉动态变化的组合以防止严重的事故是至关重要的。

3.3 对知识体系的贡献

提出的研究为安全防范过程提供了一种新的思路。以往研究中定义的预防措施对未来的工作具有启发意义。图9表示最近研究中预防措施程序的可视化发展。对促进积极预防程序的研究进行了调查，并在发展方面与提议的模式出现了详细的分歧。预防部分在每次研究中都完成了预测过程，并在每次工作中进行了改进。最初，仅根据施工事故的严重程度制定了预防措施。将施工案例分为三组，并根据严重程度形成预防措施。第二项研究补充了一个风险矩阵，使其能够对工作团队和公司的风险因素进行评级。在此之后，在制定预防措施时也考虑了施工安全方面的个人经验。之后，通过引入致命事故分析，使预防措施先进行。在没有执行任何分析模型的情况下检查致命工作事件，以了解哪些属性对致命事件有影响。

然而，事故分析需要更全面的调查，以揭示根本原因。像基于关联规则的推理系统这样的分析模型对分类和控制属性之间的相关性有重要贡献，从而发现隐藏的关系—所提出的模型关注事故的每个方面，包括严重程度、个人信息和属性。Apriori算法客服了大量数据中存在的问题，从中挖掘出有意义的结论。因此，关联规则支持更深层次的理解，并允许设置更有效的预防措施，因为属性之间隐藏的关系被揭露。此外，该模型能够适应数据的变化，获得事故类型的不同特征。这意味着该模型可以动态地与任何系统一起工作，并且可以成为之前研究中已经开始研究的通用系统的一部分。

此外，研究者还通过关联规则研究了事故属性及其之间的相关性，但没有详细检查事故类型。换句话说，这些研究并没有关注与各种事故类型相关的因素。因此，本研究涵盖了图2中给出的最常见的建筑事故。对这些事故进行了个别分析，以揭示其属性之间的关系。

最后，我们采用了一种数据挖掘技术关联规则挖掘来研究建筑工地职业事故的共同影响因素之间的相关性。通过不同建筑公司的职业健康安全专业人员收集了3870起建筑事故案例，包括急救、医疗干预、工作日损失案例和死亡案例。关联规则分析确定了不同类型施工事故的不同属性。由于事故类型不同，其特点也不同，因此必须采取不同的措施来防止伤亡。分析整个数据集(包括每一种类型的事故)也支持这一观察结果，并且与个别分析的结果有明显的差异。

结论

职业事故仍然是建筑业严重关注的问题，因为许多在建筑工地工作的人都遭受了事故的危险后果。预防事故变得越来越重要，第一个里程碑是了解和评估导致职业事故的因素。必须确定主要因素，并提前采取预防措施，避免施工事故的发生。文献中对建筑安全进行了广泛的研究，对建筑安全做出了重要贡献。例如，Heinrich提出了多米诺骨牌理论，认为可控的不安全行为会导致事故的危险程度。Reason引入了瑞士案例模型(SCM)。SCM描述了当一系列的潜在条件结合在一起时，就会在危险和事故之间创建一条路径。虽然SCM无疑是最常见的事故因果模型，Heinrich多米诺骨牌理论在工业上仍被广泛应用，但这些模型在描述部件之间的复杂交互作用，即事件的属性时仍然需要改进。

事故分析需要深入的调查，以找到根本原因。分类或控制属性的频率不足以提取原因并采取预防措施。例如，确定事故的主要原因是无法评估现有风险的状态是可预测的。如果调查停留在这个水平，就无法查明真正的原因。因此，关联规则挖掘可以发挥重要的作用。关联规则挖掘结果也支持了深入的理解，因为单独评估施工事故增加了从分析中获得的置信度值，如表3和表4所示的信息所示。

因此，提出的模型引入了一种适应性分析过程，以更全面的方式调查职业事故的因果关系。职业健康安全专业人员可以在不受数据变化影响的情况下，立即识别出因果关系，并制定预防措施，打破关联规则挖掘对每种常见事故类型所获得的事故要素之间的联系。最终，本研究可以帮助职业健康安全专业人员通过使用信心等统计数据向他们展示属性之间的关系，从而减轻安全风险，这些数据可以作为具体事故预防的指导或工具。

本研究存在以下局限性。首先，文献中没有关于关联规则分析阈值选择的指南。所选值越低;生成的规则数量越高。根据职业健康安全专业人员的反馈，可以重新安排阈值以获得更多或更少的规则。在未来的研究中，可以利用关联规则分析的结果设计一个模糊决策工具。此外，该模型仅使用施工数据进行测试，对预防研究有很好的理解，因此还需要针对不同行业进行测试，这将揭示不同的属性组合，以帮助专业人员预防事故。