IEEE 版权声明

版权所有 (c) 2022 IEEE

允许将本材料用于个人用途。对于任何其他用途，无论使用当前还是未来的任何媒体，均须事先获得 IEEE 的许可；此类用途包括为广告或促销目的重印或再版本材料、创作新的合集作品、转售或重新分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护部分。

录用发表在：第九届计算科学与计算智能年会（CSCI'22：2022年12月14日至16日，美国）

引用格式：

O. Izunna、H. Shane 和 K. Jess。《用于评估公共危机的机器学习方法：荟萃分析》，2022年计算科学与计算智能国际会议（CSCI），美国内华达州拉斯维加斯，2022年。

BibTeX：

@InProceedings{okpala2022methods, author = {Okpala, Izunna and Halse, Shane and Kropczynski, Jess}, title = {用于评估公共危机的机器学习方法：荟萃分析}, booktitle = {2022 计算科学与计算智能国际会议（CSCI）}, month = {12月14日 – 16}, year = {2022}, publisher = {IEEE},}

arXiv:2302.02267v1 [cs.LG] 5 Feb 2023

用于评估公共危机的机器学习方法：荟萃分析

1st IzunnaOkpala信息技术学院 辛辛那提大学
okpalaiu@mail.uc.edu

2nd ShaneHalse信息技术学院，辛辛那提大学
halsese@ucmail.uc.edu

3rd JessKropczynski信息技术学院，辛辛那提大学
kropczjn@ucmail.uc.edu

摘要—**本研究考察了危机管理中使用的机器学习方法。分析危机中发现的模式，需要通过自动化手段收集和评估历史数据集或近实时数据集。本文采用元综述法，分析了运用机器学习技术评估危机期间人类行为的科学文献。研究人员从三个学术数据库中获取已发表成果，并通过系统的文献评估，将所选研究归纳为若干主题和新兴趋势。结果显示，在所评估的文章中，社交媒体数据最为突出，占27%；其次是灾害管理、健康（COVID）和危机信息学等众多主题。此外，监督式机器学习方法占据主导地位，应用比例达到69%。分类技术在其他机器学习任务中尤为突出，使用比例为41%。发挥主要作用的算法包括支持向量机、神经网络、朴素贝叶斯和随机森林，贡献比例分别为23%、16%、15%和12%。**

索引词——**危机信息学，灾害管理，机器学习，学习算法，元分析**

I. 引言

过去十年来，IEEE 和危机响应与管理信息系统（ISCRAM）等科学界团体开展了许多研究，利用实时信息源支持大规模事件期间的态势感知 [1], [2]。机器学习领域在自动化流程以筛选海量数据方面取得了进展 [3]。本研究探讨学术文章中采用的多种机器学习解决方案，以了解人类面对危机时的行为。研究还参考了探讨灾害管理、健康、政治及其他形式危机的研究；这些研究利用社交媒体以外的数据，并以主要学术数据库中许多聚焦于分析人类行为的学术文章作为证据。对于无法掌控主流媒体的个人而言，社交媒体平台是首个互动媒介，但也存在用于追踪危机的本地信息来源。人们往往会通过自己熟悉的社交网络，或经核实的本地报道机构和新闻媒体，报告事件或讨论正在发生的事件 [4], [5]。科学研究人员利用大量涌现的社交媒体数据 [6], 和本地报告，开展预测等机器学习程序。本文旨在考察学者在危机管理中采用的主流机器学习方法、这些方法的实施方式以及数据来源。

如前所述，历史数据和实时数据在危机管理中发挥着重要作用；在本文中，它们用于评估危机 [7]。为了规划应对未来危机的措施、减轻其影响并加以避免，建议调查历史上或现有的解决方案。人类行为的概念基于某些成因，即人在采取行动之前必有原因 [8]。人类所处的环境是推动危机发生的主要因素，而人类管理环境资源的方式对危机的发生起着重要作用。可用于根据数据源确定人类行为的一些概念包括情感、认知和/或态度 [9]。认知能够独特地识别基于意见的想法和印象 [10], [11]，并且可以与情感和态度区分开来；情感强调情绪 [12], [13], ，而态度则会引导行动 [7]。本研究旨在借助同行评审文章，展示当今世界上最普遍用于评估或预测危机的机器学习方法。研究重点是公众或人类针对危机所采取的行动，涉及态度和/或认知等关键概念。

A. 研究问题

1. 管理危机时占主导地位的机器学习方法有哪些？**2.** 探讨危机的科学研究中经常使用哪些关键词？

II. 背景

统计方法是最早用于评估危机的工具之一 [14]。1932 年，Patrick [15] 对一些组织完成多变量分析后，金融利益相关者开始开发模型，以评估其组织发生危机的可能性。此后，学者们设计了多种定量方法来检测和评估危机。一些定量分析方法，例如 t 检验，已成功用于量化比率 [16]。Altman [17] 设计了一种评分，用于将观测结果归类为良好或不良。多重判别分析 (MDA) 也在一些高级分析中发挥了作用，用于压缩数据集之间的方差 [18]。尽管这类模型在学术界和工业界都得到了广泛应用，但事实证明，它们关注的都是数字和数量，因此需要加以改进，以便

超越数字的局限 [19]。为了解决这些模型的局限性，采用模式匹配方法的各类研究已在机器学习领域得到广泛开展 [20]。其中一些研究证明了机器学习模型处理不平衡数据集 [21],、图像数据和文本数据 [22]的能力。甚至可以检测出参数方法与非参数方法在风险分析方面的差异 [23]。

为回答这些研究问题，作者探讨了与自动编码、模式匹配和文本分析相关的一些背景文献。大多数文章讨论了如何利用机器学习、数据转换/缩放和自然语言处理来处理危机管理问题。机器学习帮助 IT从业者在极短时间内完成任务 [24]。它似乎是快速识别中断事件、获取真实信息流或实时检测周期性事件的一种选择 [25]。学习此类模式也是一项重大变革，因为这种方法试图在给定数据集中映射感兴趣的模式或相似性 [26],，同时还展现出学习并生成准确结果的能力 [27]。理解人类行为的关键模式，是体现偏好的线索。这种偏好可以表现为情绪、观点、看法或特定注释，有助于解释人们为何会有那样的行为 [28]。在评估或预测行为时，需要留意的一些关键概念包括认知、态度、情感等。“认知”一词可能会被误解为与态度甚至情感含义相同。简单来说，情感关注的是人们对某一事件的感受，即对事件的正面或负面感受 [13], ；认知和态度关注的则是人们对某一事件的看法 [11], [28]。情绪有助于理解认知，但二者的区别在于，认知或态度可以基于事实形成，而不单单取决于情绪 [29]。态度具有反应性（可能引发即时行动），而认知则是暗示未来行为或态度的内在线索 [30]。聚焦于认知，其正式定义是组织和解释感官输入以理解事件的过程 [11]。

为了检测或预测危机，需要确保用于分析的数据集具有可操作性。可操作数据的特点是包含能够据以采取行动的信息，或能够充分洞察未来的数据 [31]，也就是说，这些数据能够为行动提供参考，从而支持有价值的决策。换言之，它不仅仅是存放在数据仓库中的数据，还经过了分析和数据处理，并以清晰、易懂且通常具有视觉吸引力的方式呈现 [32]。这使研究人员能够发现错误或潜在危机，把握新机遇，改进未来的行动，并为未来更快地做出更明智的决策 [33]。

A. 用于危机检测与管理的机器学习

危机事件有多种理由值得使用机器学习进行分析。其中一个原因是防止此类事件在未来再次发生，另一个原因是研究人们在此类情况下及时采取行动的模式。COVID-19疫情是2020年- 2021 [34],期间全球面临的重大危机，许多学者尝试评估实时或历史数据，以了解疫情如何加剧，以及据称导致其加剧的各项指标。其他研究人员则探究为何有些人愿意接种疫苗 [11] ，而另一些人不愿意。除了COVID-19大流行之外，危机信息学研究还涉及各种形式的危机，例如灾害管理、911或311事件、政治运动、自然灾害，以及各种形式的袭击，如强奸等。鉴于我们的研究聚焦于危机场景，机器学习在科学界展现出良好的应用前景，这一点从大量运用自动化方法检测、预测或预防危机事件的论文中可见一斑。几乎所有主要的无监督和有监督算法，如神经网络、支持向量机、Na¨ıve贝叶斯算法、K均值聚类、K近邻、决策树和梯度提升算法，都已在不同场景中得到应用。使用经过训练且准确率较高的模型，可以及时预防或应对多种危机事件，尤其是在人工介入解决问题几乎不可能或非常耗时的情况下。

借助多种数据源和机器学习方法，规避危险变得更加容易，也更具表现力。Twitter 平台就是一个可以从中获取可操作数据的渠道。数千名学者研究过该平台，因为它可用于在危机期间生成洞见 [35]。此外，在处理来自不同来源的数据时，数据结构问题会限制某些流程，但通过恰当应用机器学习调优或数据转换，可以缓解这一问题。另一些数据源则专门归档特定主题的数据，例如健康危机——世界卫生组织数据（WHO） [36], ； 金融危机——世界银行开放数据 [37], ； 政府灾害预警——data.gov[38],; 以及儿童死亡率和孕产妇死亡率——联合国儿童基金会 [39]。

只要拥有可靠且可操作的数据源，机器学习便能高效完成通常需要耗费人类大量时间的计算任务。随着时间推移，机器学习不断发展，如今任何具有内存的计算设备都可以通过训练来遵循特定模式 [40]。传统的学习方式是让人们接受专门材料的训练，并通过测试来判断他们对某一主题的掌握程度；机器学习的概念正是由此产生的。机器学习描述了机器具备某种智能并能够从经验中学习的能力。西洋跳棋就是机器通过经验进行学习的一个例子 [41]。除了西洋跳棋，机器还利用BERT模型接受训练，以区分真实新闻和假新闻，以及垃圾邮件和真实邮件 [42]。部署这类适用于机器的系统，有助于无缝预测危机事件。

III. 方法

本研究采用的方法是元综述法。该方法能够在一组主题相关的文章中识别可行变量。这种方法通常用于分析以特定语言发表的文献；本研究的综述聚焦于以英语撰写并发表于 ISCRAM、ScienceDirect 和 IEEE Xplore 数据库的学术文章。选择这些数据库，是因为其中设有利用机器学习开展危机管理研究的子类别。这并不是说不存在其他侧重危机研究的数据库；本研究主要聚焦于这三个数据库。图1所示的检索与筛选标准展示了各个组成部分及其间的数据流。检索词实质上是针对特定数据库查询所需使用的关键词。检索方法一节对检索词作了更深入的说明。查询结构中使用了 AND 和 OR 运算符，因为它们有助于筛除不符合查询范围的文章。图2进一步展开了图1所示的纳入与排除标准，以呈现纳入与排除机制的不同组成部分，以及数据在此过程中如何被截取或缩减，最终得出结果。

A. 检索方法

检索方法对于确保本次评估性综述的严谨性至关重要。研究论文是通过在多个不同的电子数据库中进行自动检索找到的。表I展示了所检索的三个科学数据库。

表I：检索范围。

索引	NAME	URL
DB1	IEEEXplore 数字图书馆	https://ieeexplore-ieee-org/
DB2	ScienceDirect 图书馆	https://www.sciencedirect.com/
DB3	ISCRAM 数字图书馆	http://idl.iscram.org/

通过采用结构化的搜索模式，本研究旨在获取与研究问题相关的文章。在搜索阶段，我们从相关来源筛选文献，并选择合适的关键词。文章中包含的关键词包括：机器学习、危机和灾害。关键词准备步骤如下：

1) 根据研究问题确定检索词。2) 同时确保识别检索词的不同拼写形式、反义词和同义词。3) 对检索词执行布尔运算（AND、OR）。4) 确定检索查询的日期。

论文中使用了以下反映我们研究问题的关键词和运算符：(灾害) OR (危机) AND (机器学习)。

搜索依据数据库对查询结构的偏好进行。我们使用

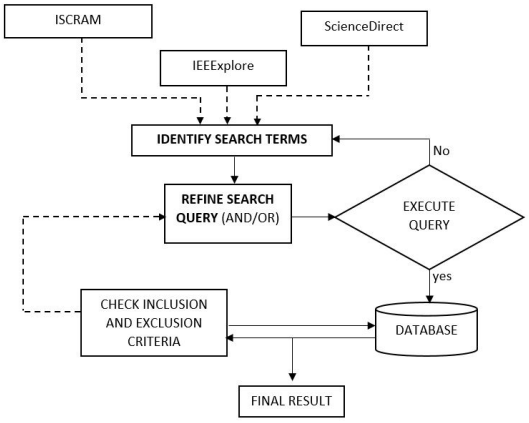


图1：数据检索与筛选算法

ISCRAM 专用的上下文查询语言（CQL）进行查询，代码见表II。借助其基于网页的门户网站输入搜索参数，IEEE Xplore 和 ScienceDirect 的操作则简单得多；参数中使用了表II所示的“AND”和“OR”运算符。

表II：数据库的检索方式

索引	查询类型	查询
1	IEEE Xplore	“危机 AND 机器学习 OR 灾害”
2	ScienceDirect	“危机 AND 机器学习 OR 灾害”
3	ISCRAM (CQL)	“所有 摘要 机器学习 危机 灾害”

此外，我们将搜索范围限定在2010年到2021年之间发表的论文（近期出版物），并限制了文献类别，仅包括同行评审期刊和会议论文。

B. 数据筛选

按照图1所示的逐步方法，我们仔细筛选了所评述的55篇文章。如前所述，选取2010年到2021年，主要是为了反映机器学习实践在危机领域的当前趋势和发展。根据表II所示的查询格式，各数据库返回的初始结果分别为：ScienceDirect有2,274篇文章，IEEE Xplore有16,236篇文章，ISCRAM有76篇文章。鉴于文章数量庞大，我们缩小了发表文献的范围，以突出我们的研究领域（危机）的重要性。具体而言，对于ScienceDirect数据库，我们只选择计算机科学领域的期刊和会议文章，因为该学科在机器学习领域显然占据主导地位，相关贡献也更多。最终得到186篇文献。IEEE检索也采用了相同策略，但高级搜索参数有所不同。通过将主题类别限制为灾害，并仅选择会议和期刊文章，IEEE Xplore返回的16,236篇初始结果减少至5,770篇。应用计算机科学学科筛选后，共得到476篇文章。

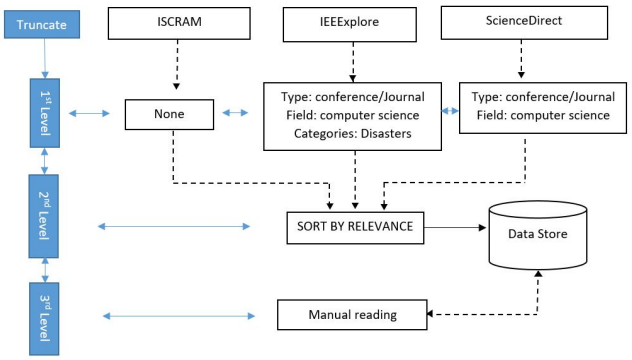


图2：纳入与排除流程图

ISCRAM的检索结果无需进一步缩减，因为我们只得到76篇文章。通读所有检索到的文章几乎不可能，而且非常耗时。因此，我们尝试利用各服务器现有的系统化方法，而非随机抽样，选出最相关的前30篇文章。相关性体现为文章对知识体系的贡献程度以及被引频次。为平衡三个数据库之间的数量，我们最终选取了90篇文章。随后，如图2所示，我们通过人工阅读这90篇文章进行进一步筛选，以确保每篇文章都采用了机器学习方法，与危机响应相符，并提出了恰当的研究问题来探讨危机。为此，我们采用了以下标准：

1）纳入：根据摘要，研究展示了定义明确的方法论。2）纳入：根据结论，研究至少指出一个检索词，并提供了能证明该研究的指标。3）排除：研究所采用的方法未体现机器学习。4）排除：研究缺乏有力的验证或验证依据。

表III：数据项收集表。

索引	字段	描述
DI-1	标题	文章标题
DI-2	Year	文章的出版年份
DI-3	数据库	出版来源
DI-4	技术	本文采用的机器学习方法 文章
DI-5	研究领域	研究涵盖的感兴趣领域 研究（计算机科学）

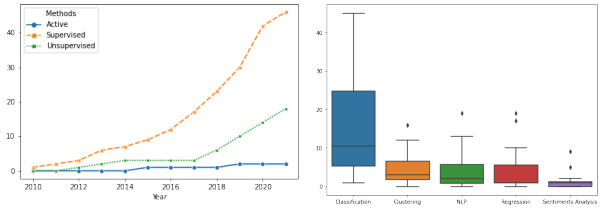
表III展示了从每篇手稿中提取的组成部分。这不仅体现了这些内容与研究问题之间的联系，也提供了一种机制，将数据提取限定在相关字段内。标题、年份、数据库、机器学习技术和所涵盖的研究问题均属于此类字段。

C. 数据提取

研究人员对数据进行人工编码，以回答我们的两个研究问题。标注分四批进行，分别针对方法、任务、关键词和算法。根据筛选标准，本研究考虑的所有出版物均聚焦于危机响应和机器学习技术的应用。

IV. 结果

分析显示，所审阅的文章中，大多数支持分类机器学习任务，而非回归或聚类任务。图 3b 显示，41% 的审阅论文使用了分类任务，而回归和聚类任务分别占 18% 和 16%。如图 3a 所示，将监督式机器学习方法与无监督学习和主动学习一同统计时，监督式机器学习占 69%，无监督学习和主动学习分别占 27% 和 3%。这并不意味着一种方法优于另一种方法，而是反映了不同科学界的偏好。对监督学习的偏好可归因于危机领域特有训练数据的可用性，这些数据来自 **CrisisNLP** [43], https://crisisLex.org - CrisisLexT26、SoSItalyT4 和 BlackLivesMatterU/T1 [44]。有时，危机数据的传播格式可能不适合作为训练模型的语料库。在这种情况下，无监督学习方法便派上了用场 [45]。所考察的一些研究采用了混合方法，根据其部分分析中训练数据的可用情况，使用监督方法和无监督方法。主动学习似乎并未广泛应用于危机研究。造成这一现象的一个因素可能是，主动学习是一种半监督学习形式，需要借助外部来源为数据集添加标签。因此，半监督学习并未得到全面应用。从方法的时间分布来看，显然，随着年份推移，三个数据库中监督式机器学习的使用都有所增加，如图 3 所示。



(a) 机器学习方法　(b) 机器学习任务

图 3：机器学习方法和任务的图表示

图 3a 中的方法在图 3b 中进一步细分为分类、聚类、回归、自然语言处理和情感分析。图中将自然语言处理和情感分析单独列出，因为它们可以是分类、聚类或回归任务的子类。需要注意的是，有时它们会独立存在

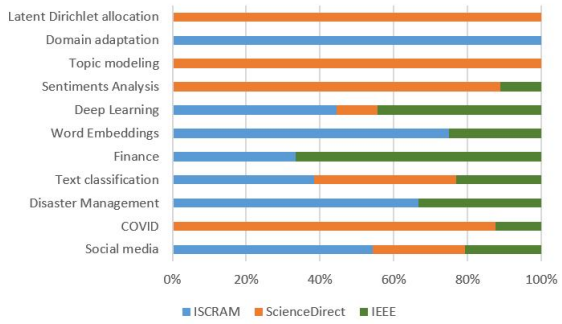


图4：所评述文章中关键词的分布

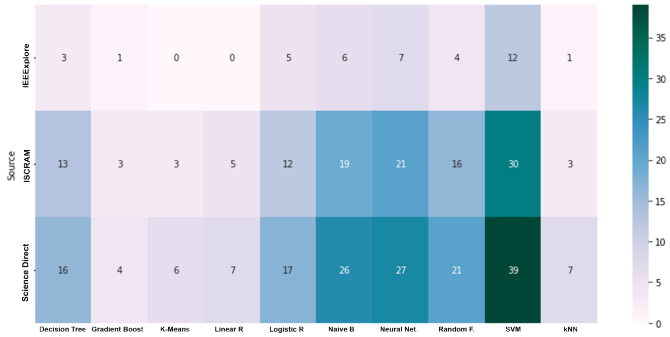


图 5：综述文章中采用的算法（热图）

（如变换器），而不遵循上述三类任务中的算法过程。

分类使用上的一致性进一步强化了先前的论述，即监督学习在研究人员中占主导地位。其次是聚类，其中K近邻（K近邻）、K均值，以及一些采用随机森林和决策树算法的任务，均如图5和图6所示。线性回归和逻辑回归等回归任务表明，它们在当今的研究中仍然具有相关性，所占比例分别为4%和10%。分析还表明，机器学习方法在危机管理或评估中的应用于2020年和2021年达到最高水平。再次，机器学习分类技术使用量的增长，可能与训练数据集的可用性、算法易于实施，以及训练数据标签易于理解有关。复杂技术的使用也可能是一个影响因素，因为此类分析可能会耗费人工分析人员大量时间。自动化数据处理、预测或分析的需求已成为常态，能够为人类提供帮助。

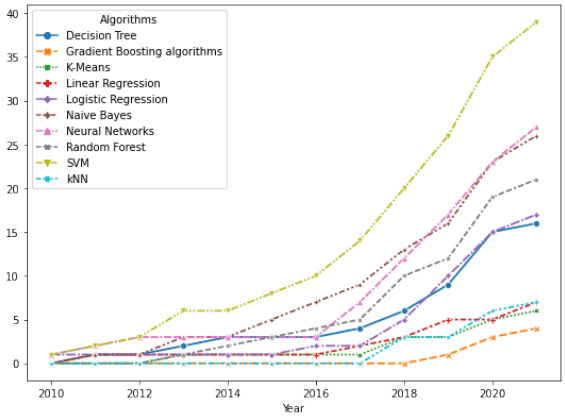


图 6：综述文章中采用的机器学习算法（时间序列）

图 4 中关键词的分布得出了一些值得关注的结果。在所审阅的出版物中，“COVID、灾害管理和社交媒体”出现得更频繁，并在 2020 年达到峰值。社交媒体被频繁提及，以及它们被用于汇总危机数据，这两者之间呈现出明显的模式。我们的分析还发现，“社交媒体”是 2019 年、2020 年和 2021 年最受关注的关键词，这表明在现代社会，社交媒体是收集危机相关信息的有用工具。我们可以将 2021 年对健康危机和 COVID 的提及与近期的 COVID-19 大流行病联系起来，也可以联系到研究人员如何利用智能系统汇总来自各种媒体的数据。灾害管理也被用作多篇文章中的关键词，但在 2021 年较少出现。此外，文本分类、领域自适应和主题建模也受到关注，这些都与分析危机的子方法有关。主题建模在机器学习领域虽鲜少出现，却十分重要，并且已在一些研究中得到应用。这与我们先前的论述相呼应：自然语言处理任务可以在无监督环境中充当转换器（例如，主题建模会用到无监督方法）。

本研究的另一项重要发现是，当我们进一步按照算法标准或表明其运作方式的术语对这些方法进行细分时，之前的判断便得到了进一步印证。支持向量机算法似乎在综述文章中占据主导地位，其次是神经网络。这也进一步印证了我们先前的说法，即分类任务和监督方法占据主导地位。支持向量机的优势体现在结构风险最小化、训练过程中的高效内存管理，以及数据集所需的高维空间处理能力上。内存管理是深度学习算法面临的问题；深度学习算法的准确率高于支持向量机，但具体取决于数据量。然而，对危机领域的研究人员而言，支持向量机的优势似乎更具吸引力。神经网络在所研究的文章中也十分突出。在危机研究中出现的神经网络变体包括卷积神经网络和长短期记忆网络（长短期记忆网络），后者获得了更多关注。图 6 所示的朴素贝叶斯、随机森林、决策树和逻辑回归也展现出良好的应用前景。

V. 讨论

我们对2010年至2021年间发表的科学文章进行了这项评述，重点关注

危机管理和机器学习。危机可能以不同形式出现，例如健康危机、自然灾害、经济危机、粮食危机和政治事件等。这些变化增加了开发自动化方法来管理危机以及追踪新兴事件的复杂性，而此类方法有助于快速决策。在描述元综述流程，以及我们如何收集和评估有关机器学习在危机管理中应用的不同出版物方面，所采用的方法足够明确。文献综述和结果部分对机器学习概念的阐述，展示了多种机器学习技术和算法如何用于管理危机。从结果中可以明显看出，所有算法都充分体现了其所带来的价值。机器学习领域不断发展，致力于帮助分析大量数据，通过自动化流程减轻数据科学家的工作，并改变数据提取和解读的方式。

所有纳入综述的文章都根据所研究问题的结构、所应对的危机类型、数据来源和数据量得出了结果。图 3a 所示的百分比分布反映了人们偏好的机器学习任务。在自动化危机管理方面，自然语言处理任务（如情感分析）被视为至关重要；这类任务可归为监督学习或无监督学习。分类、聚类和回归任务广受青睐，其中分类位居首位，这归功于训练数据集易于获取，以及通过 Python、R 和 Java 等流行编程语言中的预封装库，能够轻松实现相关算法。此外，人们普遍认识到，当前科学领域正面临可重复性危机。机器学习技术通常更易于用于分析。可重复性危机是指由不同研究人员团队重复进行同一实验时，研究结果无法复现的情况日益增多。这一问题会影响多个领域，因为这些领域都利用机器学习来开展发现。

A. 局限性

本研究未回顾所有科学研究数据库中的文献，而是从三个符合纳入与排除标准的数据库中筛选文章。因此，我们回顾的研究反映了三个来源中有关公众行动、危机和机器学习的研究。本研究结果可能受到论文所采用的检索策略、研究人员的偏见、已发表期刊或会议论文集分布不均，以及数据提取失实等因素的影响。

本研究采用了自动化和人工检索技术。正如图2所示，首次迭代共找到数百条数据。初步检索后，研究人员根据研究论文的内容制定了人工检索流程。三位研究人员选取并分析了可能相关的研究。检索结果可能遗漏了相关研究。因此，本综述的范围可能受到限制。因此，本研究的效度仅限于纳入这项评估性综述的55篇关键论文。

VI. 结论

研究结果表明，相当大比例的文章（41%）采用了分类，而非回归或聚类，这得益于训练数据/语料库和预封装机器学习库的可用性。69%的文章采用了监督式机器学习方法（RQ1），表明各科学界在应对危机时普遍偏好这种方法。因此，27%的研究采用了无监督学习技术，其余4%采用了主动学习方法。为回答RQ2，我们的分析揭示了回顾文章中使用的机器学习方法和常见关键词。这表明，支持向量机、神经网络、朴素贝叶斯和随机森林等算法在危机管理领域的研究人员中颇受欢迎（RQ2）。关于RQ2，危机信息学这一关键词在所考察的科学文献中也引起了广泛关注。一些值得关注的研究方向，如健康和灾害管理，到2020年已受到越来越多的关注；社交媒体则在2019、2020和2021年最受关注，这表明在当今时代，社交媒体有助于收集危机相关数据。

参考文献

[1] T. Nguyen、C. P. Lim、N. D. Nguyen、L. Gordon-Brown 和 S. Na- havandi，《航空环境中的态势感知评估方法综述》，*IEEESystems Journal*，第13卷第3期，第3590–3603页，2019年。[2] K. Steen-Tveit 和 J. Radianti，《多种紧急响应情境下共同作战图景与态势感知分析》。载于第*18*届 *ISCRAM* 会议，2019年。[3] M. Y. Kabir 和 S. Madria，《Emocov：利用新冠肺炎推文进行情绪检测、分析和可视化的机器学习方法》，*Online Social Networks and Media*，第23卷，文章编号100135，2021年。[4] H. Hwang 和 K.-O. Kim，《社交媒体作为社会运动工具：社交媒体使用与社会资本对参与社会运动意愿的影响》，*International Journal of Consumer Studies*，第39卷第5期，第478–488页，2015年。[5] O. Oh、M. Agrawal 和 H. R. Rao，《社区情报与社交媒体服务：对社会危机期间推文的谣言理论分析》，*MIS quarterly*，第407–426页，2013年。[6] E. Nathan、M. Pranav、K. Monica、H. Shane 和 K. Jess，《探索利用社交媒体考察大流行病期间当地居民对戴口罩态度的方法》，载于第*18*届 *ISCRAM* 会议 – 布莱克斯堡，弗吉尼亚州，美国，*2021*年*5*月。ISCRAM，2021年，第1–14页。[7] S. Praveen、R. Ittamalla 和 G. Deepak，《分析印度公民对新冠肺炎疫苗的态度——一项文本分析研究》，*Diabetes & Metabolic Syndrome: Clinical Research & Reviews*，第15卷第2期，第595–599页，2021年。[8] M. Fritts，*Actions and Causes: A Methodological Inquiry into the Explanation of Action*。威斯康星大学麦迪逊分校，2020年。[9] B. Chen、J. M. Zhang、Z. Jiang、J. Shao、T. Jiang、Z. Wang、K. Liu、S. Tang、H. Gu 和 J. Jiang，《中国“乙肝疫苗危机”期间媒体与公众对疫苗接种的反应》，*Vaccine*，第33卷第15期，第1780–1785页，2015年。[10] M. R. Haupt、A. Jinich-Diamant、J. Li、M. Nali 和 T. K. Mackey，《利用无监督机器学习和社交网络分析刻画新冠肺炎期间“解放”运动的 Twitter 用户话题与传播网络动态》，*Online Social Networks and Media*，第21卷，文章编号100114，2021年。

[11] O. Izunna、R. Guillermo、Romera、Z. Weibing、H. Shane 和 K. Jess，《感知分析：利用自然语言处理和机器学习对支持与反对疫苗的观点进行分类》，第55届夏威夷国际系统科学会议论文集，第2981–2990页，2022年。[12] P. Sv、J. Tandon、H. Hinduja 等，《印度公民对新冠肺炎疫苗副作用的看法——一项机器学习研究》，Diabetes & Metabolic Syndrome: Clinical Research & Reviews，2021年。[13] O. U. Izunna、I. P. Ngozi 和 E. A. Obiajulu，《评价情感分析：一种伪语言学方法及用于数据决策的在线接受度指数——以尼日利亚选举为例》，Internet of Things and Cloud Computing，第7卷第2期，文章编号39，2019年。[14] C. Young，《模型不确定性与科学危机》，Socius，第4卷，文章编号2378023117737206，2018年。[15] P. Patrick，《成功工业企业与失败企业比率的比较》，《注册会计师》，第2卷，第598–605页，1932年。[16] R. Sonmez，《偶尔加班对建筑业劳动生产率的影响：定量分析》，《加拿大土木工程期刊》，第34卷第7期，第803–808页，2007年。[17] E. I. Altman，《财务比率、判别分析与企业破产预测》，《金融期刊》，第23卷第4期，第589–609页，1968年。[18] J. A. Ohlson，《财务比率与破产概率预测》，《会计研究期刊》，第109–131页，1980年。[19] X. Song、A. Mitnitski、J. Cox 和 K. Rockwood，《机器学习技术与经典统计模型在预测健康结局方面的比较》，载于 MEDINFO 2004。IOS出版社，2004年，第736–740页。[20] F. Barboza、H. Kimura 和 E. Altman，《机器学习模型与破产预测》，《专家系统与应用》，第83卷，第405–417页，2017年。[21] M. A. H. Farquad 和 I. Bose，《使用支持向量机预处理不平衡数据》，《决策支持系统》，第53卷第1期，第226–233页，2012年。[22] D. Faure 和 C. N´edellec，《利用机器学习从技术文本中获取谓词论元结构知识：ASIUM系统》，载于国际知识工程与知识管理会议。施普林格，1999年，第329–334页。[23] D. G. Altman 和 J. M. Bland，《参数方法与非参数方法在数据分析中的比较》，《英国医学杂志》，第338卷，2009年。[24] S. R. Chowdhury、M. Imran、M. R. Asghar、S. Amer-Yahia 和 C. Castillo，《Tweet4act：利用事件特定档案对危机相关消息进行分类》，载于 ISCRAM。Citeseer，2013年。[25] K. Starbird、G. Muzny 和 L. Palen，《向人群学习：利用协同过滤技术识别大规模突发事件中的一线推特用户》，载于 ISCRAM。Citeseer，2012年。[26] L. Szczyrba、Y. Zhang、D. Pamukcu 和 D. I. Eroglu，《一种量化脆弱性在飓风灾害中作用的机器学习方法》，载于 ISCRAM 2020会议论文集——第17届国际危机响应与管理信息系统会议，2020年。[27] F. Angaramo 和 C. Rossi，《用于 Twitter 实时事件检测的在线聚类与分类》，载于 ISCRAM，2018年。[28] S. Wiggins 和 J. Potter，《态度与评价实践：日常食物评价中的类别与项目，以及主观与客观建构》，《英国社会心理学期刊》，第42卷第4期，第513–531页，2003年。[29] K. D. Landreville 和 C. Niles，《“这可是事实！”：政治意识形态、PSRS 与感知来源可信度在估计党派新闻事实内容中的作用》，《广播与电子媒体期刊》，第63卷第2期，第177–194页，2019年。[30] J. O. Faronbi、O. Adebowale、G. O. Faronbi、O. O. Musa 和 S. J. Ayamolowo，《护理专业学生对老年患者护理的感知、知识与态度》，《非洲护理科学国际期刊》，第7卷，第37–42页，2017年。[31] J. Kropczynski、R. Grace、J. Coche、S. Halse、E. Obeysekare、A. Mon- tarnal、F. Benaben 和 A. Tapia，《识别社交媒体上可供紧急调度使用的可操作信息》，ISCRAM亚太会议论文集，文章编号11，2018年。[32] N. Spasojevic 和 A. Rao，《识别社交媒体上的可操作消息》，载于2015年IEEE国际大数据会议（Big Data）。IEEE，2015年，第2273–2281页。[33] M. Avvenuti、S. Cresci、F. Del Vigna、T. Fagni 和 M. Tesconi，《Crismap：一种基于损害检测与地理解析的大数据危机地图系统》，《信息系统前沿》，第20卷第5期，第993–1011页，2018年。

[34] X. Wu, R. C. Nethery, B. M. Sabath, D. Braun, 和 F. Dominici，《美国空气污染暴露与 COVID-19 死亡率》，*MedRxiv*，2020。[35] J. L. Hellerstein、S. Ma 和 C.-S. Perng，《发现事件数据中的可操作模式》，*IBM Systems Journal*，第 41 卷，第 3 期，第 475–493 页，2002。[36] B. Karo、C. Haskew、A. S. Khan、J. A. Polonsky、M. K. A. Mazhar 和 N. Buddha，《世界卫生组织在孟加拉国罗兴亚危机期间的早期预警、警报和响应系统，2017–2018》，*Emerging Infectious Diseases*，第 24 卷，第 11 期，第 2074 页，2018。[37] C. E. Weaver，《促进发展的开放数据：世界银行、援助透明度与国际金融机构的善治》，载于*Good Governance and Modern International Financial Institutions*。Brill Nijhoff，2019，第 129–183 页。[38] C. Birchall，《“data. gov-in-a-box”：界定透明度》，*European Journal ofSocial Theory*，第 18 卷，第 2 期，第 185–202 页，2015。[39] 世界卫生组织等，《孕产妇死亡率趋势：1990-2015：世界卫生组织、联合国儿童基金会、联合国人口基金、世界银行集团和联合国人口司的估算》，孕产妇死亡率趋势：*1990-2015*：世界卫生组织、联合国儿童基金会、联合国人口基金、世界银行集团和联合国人口司的估算。世界卫生组织，2015。[40] C. Stauffer 和 W. E. L. Grimson，《利用实时跟踪学习活动模式》，*IEEE Transactions on Pattern Analysis and Machine Intelligence*，第 22 卷，第 8 期，第 747–757 页，2000。[41] A. L. Samuel，《使用西洋跳棋进行机器学习的若干研究》，*IBM Journal of Research and Development*，第 3 卷，第 3 期，第 210–229 页，1959。[42] J. Cao 和 C. Lai，《基于多语言 BERT 的双语多类型垃圾信息检测模型》，载于*GLOBECOM 2020-2020 IEEE* 全球通信会议。IEEE，2020，第 1–6 页。[43] R. ALRashdi 和 S. O’Keefe，《用于危机响应推文分类的深度学习与词嵌入》，*arXiv* 预印本 *arXiv:1903.11024*，2019。[44] A. Olteanu、C. Castillo、F. Diaz 和 S. Vieweg，《CrisisLex：用于收集和筛选危机期间微博通信的词典》，载于第八届 *AAAI* 网络日志与社交媒体国际会议，2014。[45] C. Arachie、M. Gaur、S. Anzaroot、W. Groves、K. Zhang 和 A. Jaimes，《大规模灾害中子事件的无监督检测》，载于*AAAI* 人工智能会议论文集，第 34 卷，第 01 期，2020，第 354–361 页。