

面向机器学习应用的可解释性需求分析框架

裴忠一 刘 璘 王 晨 王建民

(大数据系统软件国家工程研究中心(清华大学) 北京 100084)

(清华大学软件学院 北京 100084)

(peizhyi@tsinghua.edu.cn)

An Explainability-Centric Requirements Analysis Framework for Machine Learning Applications

Pei Zhongyi, Liu Lin, Wang Chen, and Wang Jianmin

(National Engineering Research Center for Big Data Software (Tsinghua University), Beijing 100084)

(School of Software, Tsinghua University, Beijing 100084)

Abstract Data-driven intelligent software based on machine learning technology is an important means to realize industrial digital transformation. The research and development processes of data-driven intelligent software require the combined use of software requirements engineering, data and domain knowledge engineering, machine learning and so on. This process involves many subjects and roles, making it extremely challenging to clearly explain why and how the domain knowledge, business logic and data semantics relate to each other. Hence, a systematic requirements engineering approach is needed to explicitly address the explainability requirements issues of data-driven intelligence applications. It is still a fast-evolving research field which requires the proper embedding of various domain models and end-to-end machine learning technology fused into a given business processes. A key research question is how to deal with explainability as a core requirement for safety-critical scenarios in industrial, medical and other applications. We provide a research overview on requirements engineering for machine learning applications, in relation to explainability. First, the research status quo, research foci and representative research progress are reviewed. Then, an explainability-centric requirements analysis framework for machine learning applications is proposed, and some open important issues are put forward. Finally, based on the proposed framework, a case study of industrial intelligence application is discussed to illustrate the proposed requirements analysis methodological framework.

Key words requirements engineering; machine learning; domain model; industrial engineering; explainability

摘 要 基于大数据机器学习的智能软件研发过程需要综合运用软件工程、数据与领域知识工程、机器学习等多方面的知识和工具,涉及的研究主题和人员角色众多,技术实现手段复杂、研发难度大。面向智能软件的需求工程需要面对领域知识、业务知识、数据科学交织带来的挑战。然而,如何将领域知识和端到端的机器学习技术恰当地融合到给定的业务流程之中,以及如何应对工业、医疗等高可信要求场景中的可解释性需求,仍是亟待探索的重要研究问题。调研了近年来面向机器学习应用的需求工程研究文献,对该领域的发展现状、核心问题和代表性方法进行综述。提出了面向机器学习应用的可解释性需求分析框架。基于该框架,通过一个工业智能应用案例分析了未来待研究的重要问题,展望了可行的研究路径。

收稿日期: 2022-09-06; 修回日期: 2023-08-15

基金项目: 国家重点研发计划项目(2021YFB1715200); 国家自然科学基金委创新群体项目(62021002)

This work was supported by the National Key Research and Development Program (2021YFB1715200) and the National Natural Science Foundation of Innovation Group Project (62021002).

通信作者: 刘璘(linliu@tsinghua.edu.cn)

关键词 需求工程;机器学习;领域模型;工业工程;可解释性

中图法分类号 TP311

近年来,数据驱动的机器学习技术在多个行业领域得到了成功应用,包括图像识别、语音识别、交通路况预测、自动驾驶、智能助手、用户偏好预测和产品推荐等^[1].在工业领域,如工业互联网、工业大数据、工业机器学习等应用也在工业数字化转型的过程中获得长足发展.应用需求的持续驱动使智能软件技术栈不断发展、演变,形成了数据驱动的智能软件方法与技术新分支,面向智能软件需求分析方法的研究与实践也得到了广泛关注.一方面,软件工程背景的系统开发人员习惯于沿用熟悉的开发方法,希望既有开发方式能适用于数据驱动的智能软件开发;另一方面,机器学习算法开发人员倾向于基于个人经验选择开发所需的技术栈,导致智能软件的开发质量和开发效率难以得到保证.上述原因使得我们在工业以及医疗等领域中运用内嵌机器学习模型的智能应用时,可解释性和稳定性需求无法被满足.

国内外相关研究工作聚焦于数据驱动智能系统的需求分析方法和工程应用案例,分别从研究与工程实践角度,探讨现有方法、流程、策略需要的改进,以支持智能软件的研发效能提升^[2].研究主题包括机器学习应用的功能性需求定义、非功能性需求定义、质量评价标准等.其中较为重要的非功能性指标包括机器学习模型的可解释性^[3]、预测分析结果的公平性^[4]、数据密集型系统的合规性等,成为智能软件工程领域的焦点.例如,欧盟委员会于2021年提出了可信赖人工智能伦理导则^[5].

可解释性是智能软件开发中重要的主题^[6-8],一方面关注理解机器学习模型的工作原理,另一方面着重建立机器学习模型的行为与现有知识体系的关系.目前,对可解释性的研究多以机器学习技术研究为主,分析机器学习模型与现有知识体系和数据特性的关联,从需求分析中针对性地提取可解释性的依据是一项需要系统化研究的任务.本文围绕“可解释性需求分析”所面临的要点问题进行文献调研,构建方法学框架.在本文中,“可解释性需求”指对机器学习模型可解释性的要求和理解机器学习模型的需求要素.“可解释性需求分析”是对获取机器学习模型可解释性需求依据及设计决策依据的过程和方法.

智能软件需求工程涉及人工智能、领域知识工程、软件工程、数据科学及科学伦理等多方面的知

识、方法和准则.为了弥合不同知识领域之间的语义差距,需要建立跨领域的协作机制,将跨学科方法、经验和工具整合到可持续迭代的研发过程中,形成规范化、系统化的需求分析,系统设计、验证与迭代优化方法论.其中,可解释性贯穿始终,始于需求分析,落实于技术实现,最终反映在应用效果的验证过程中.

本文采用综述、归纳与案例分析相结合的研究方法.首先,从机器学习需求工程的文献调研出发,对文献的背景、问题与挑战进行归纳,以及对其中的可解释性需求的分析与处理加以总结,得到面向机器学习应用的需求工程研究路径.在此基础上,提出一种以可解释性为核心的面向机器学习应用开发的需求工程过程框架,并通过该框架指导数据驱动智能软件开发中的可解释性需求分析.

1 可解释性需求分析的相关工作

用户对智能应用的可解释性需要驱动我们在问题上下文、领域知识和机器学习模型间建立显式的联系.本文以工业领域应用场景为例,介绍智能应用需求分析的主要步骤和领域知识体系.例如如何通过人类可理解的建模方式来定义业务对象的属性和相关关系,如状态机、业务流程图、自动机、实体关系图等,以及如何用工业自动化和控制领域的模型来分析实时反应式系统等.在许多科学和工程相结合的领域中,有特定的数学、物理或化学反应过程模型,如机械工程中的力学模型、化工过程中的化学反应和工艺流程、建筑工程中的结构力学模型等.这些模型表示为数学方程、符号系统、因果网络、结构或动态行为的3维仿真建模等^[9].领域知识的显式表达是我们进行工业应用场景需求分析的重要基础,也是对系统设计进行形式化需求验证的工具^[10].

在软件需求工程中,分析师采用多种需求建模方法指导需求工程师完成需求的获取,并用来描述与领域知识体系相关的系统结构和行为约束.例如,面向目标的需求建模方法以用户和设计者的意图和目标为起点,然后通过目标分解和逐级细化所需系统的成功条件和验收标准^[11].在充分理解目标后,需求工程师采用统一建模语言(UML)和系统建模语言的多种图形化模型(SysML)^[12]等工具,建立形式化、

半形式化的业务模型,以此对领域知识体系所涉及的系统结构和行为模型等开展设计工作.跨行业数据挖掘标准流程 CRISP-DM^[13]将业务理解、数据理解作为数据驱动开发过程中2个独立的步骤.当我们在系统解决方案中引入机器学习技术,尤其是大数据驱动的深度学习技术时,数据理解对于模型可解释性有重要影响.在需求分析阶段,数据模型的定义与可视化分析是获取可解释性依据的重要手段.例如,对时序数据的周期性分析、对多个数据特征的相关性分析等.

2 机器学习需求工程研究综述

我们调研了面向机器学习的需求工程相关文献163篇,详见表1.文献研究主题包括智能应用软件开发范式的演化、机器学习应用研发所涉及的干系人角色和关注点、干系人之间的依赖关系和可解释性需求来源.

Table 1 Theme Statistics of Research Literature
表 1 调研文献的主题统计

研究主题	相关文献	文献篇数占比/%
宏观视角	[8,14–27]	9.20
需求过程模型	[28–34]	4.29
角色与协作	[35–42]	4.91
需求获取和规约	[43–51]	5.52
非功能性指标	[52–89]	23.31
数据与知识融合	[9, 90–108]	12.27
机器学习系统开发	[109–123]	9.20
可解释性	[5, 124–140]	11.04
数据流构建	[141–148]	4.91
模型溯源和验证	[149–155]	4.29
智能化应用	[156–173]	11.04

2.1 智能应用带来的软件开发范式演化

目前,软件开发与管理的典型范式以敏捷开发过程和开发运维过程 DevOps 为代表.作为数据驱动开发范式的早期里程碑,CRISP-DM将数据驱动的研发活动组织为6个阶段:业务理解、数据理解、数据准备、建模、评估和部署.CRISP-DM提出了一个规范的工作步骤,建立了数据准备、模型设计和评估的需求分析周期,形成了闭环迭代周期.CRISP-ML(Q)扩展了CRISP-DM以支持机器学习应用程序的开发,其特别关注机器学习模型的质量评估,包括鲁棒性、可扩展性、可解释性、模型复杂性和资源需求^[28].

Vogelsang 等人^[14]尝试从数据科学家的视角定义机器学习系统的需求工程特征和挑战.他们指出了开发范式中的主要变化,包括新的机器学习的性能指标和质量要求等,如可解释性、公平性和法律法规要求.Amershi 等人^[15]研究了微软公司的几个典型机器学习项目,总结了人工智能(AI)应用的重要挑战和成功要素,其包括:可持续的端到端通路;数据收集、清洗整理和可访问性;模型评估、进化和部署等.然后,提出了一个9阶段的过程模型,以解决数据相关的问题,如数据收集、清洗和标记等,以及模型相关的问题,如模型需求、特征工程、模型训练、模型评估、模型部署和模型监控.其中,从模型评估和模型监控可以向前反馈到最初的步骤,从模型训练也可以反馈到特征工程,如在表示学习中构建反馈回路.

Nalchigar 等人^[29]提出了一种目标驱动的需求建模方法,将典型机器学习目标与任务的匹配过程表示为业务分析的解决方案模式.该方法将业务决策目标映射为几个不同抽象层次的问题,然后通过将机器学习算法应用于给定的数据集获得洞察来回答这些问题.Washizaki 等人^[110]回顾了机器学习系统的常用架构模式 and 设计模式,内容涵盖多个机器学习子任务,例如用于数据存储的“数据湖模式”、用于数据分析的“原始数据提取模式”“业务逻辑与机器学习工作流的解耦模式”“事件驱动的微服务架构模式”“机器学习模型的版本管理模式”等,是针对机器学习应用程序设计过程中特定问题的可复用的局部解决方案.例如,机器学习应用需要依从特定法律法规、特定领域的物理定律和化学反应等.因此,模型可解释性、科学伦理、公平性、领域定律的依从性设定与评估已成为机器学习应用需求分析的重点.

有研究者提出将科学知识 with 端到端机器学习相结合用于工程和环境相关研究,以及将机器学习与仿真相结合的混合建模方法^[92].知识与机器学习的整合可以双向进行,可以使用机器学习增强因果关系不明显的领域模型,也可以使用物理定律、专家规则和领域知识模型改善特定领域的学习模型^[124].该方向也被称为物理感知机器学习(physical-aware machine learning)^[90]或者具有先验知识的机器学习(informed machine learning)^[91].

在软件开发范式演化涉及到的系列相关研究工作中,我们将调研文献梳理分为11个大类,如表1所示.其中,数量占比超过10%的有4类,分别是非功能性指标、数据与知识融合、可解释性、智能化应用.

非功能性指标、数据与知识融合这2个主题中包含大量可解释性相关的具体研究议题. 所以综合来看, 可解释性是机器学习需求工程中非常重要的议题之一.

2.2 机器学习应用研发的干系人及关注点

Sothilingam 等人^[174]对3个机器学习软件项目组织进行了基于案例的实证研究, 并基于面向主体的*i**框架建模分析了机器学习项目团队组织关系的差异性. 一般软件应用需求分析涉及的主要角色有业务专家、需求分析师/产品经理、软件研发工程师. 需求流程从定义业务问题开始, 通过确定软件使用的范围来识别利益相关者. 需求分析师通过与利益相关者, 特别是与业务专家的沟通, 在需求获取和识别之后, 进一步建立需求规约. 当涉及到机器学习功能的需求时, 数据科学家将参与需求分析. 同时, 领域专家在工业智能应用研发中也发挥着不可替代的作用, 因为领域知识对于理解应用场景和业务目标是必不可少的.

我们在表2中总结了面向机器学习应用的需求

工程中不同角色关注的主要问题. 表2中列举了检索文献中提及的多个挑战. 其中, 最受关注的需求问题包括可解释性、可信度、风险评估和稳定性等. 这4个问题中有相当一部分源于机器学习模型, 尤其是深度学习模型的黑盒形态, 再加上机器学习模型的使用中存在一定不确定性, 使得人们在安全攸关的场景中无法完全信赖机器学习模型的推理结果. 除上述主要围绕可解释性的需求问题以外, 业务专家的需求中还会考虑公平性问题, 因为机器学习模型可以通过选择有利于特定群体的训练数据集而引入偏差. 以上这些问题成为人们在科研方向上探索的2个主要动机: 一方面, 研究如何突破机器学习技术瓶颈, 让技术本身具备可解释性、稳定性等特性; 另一方面, 研究如何建立相对完备的数据驱动智能软件的需求分析方法论, 从需求分析的方法和规范、多角色合作机制、更全面的质量保障过程等出发, 将机器学习技术和软件工程方法与技术相融合. 其中第1个方面的研究以机器学习专家为主, 第2个方面则以软件工程专家为主.

Table 2 Focus and Challenges of Different Roles

表2 不同角色的关注焦点及挑战

角色	关注点	相关文献
业务专家	可信度、稳定性、公平性、可解释性	[83, 175–176]
需求工程师	非功能需求、业务建模、稳定性、风险评估	[2, 76, 177–179]
软件工程师	系统架构、快速原型、模型运用、数据正确处理	[22, 25, 173, 180–181]
领域专家	机理正确性、知识表达方式、数据使用方式、可解释性	[51, 182–184]
数据科学家	目标形式化、可用资源、评价指标、数据流	[40, 42, 117, 185]

表2所述的不同角色的关注点存在一些共性, 但也存在明显的差异, 这说明每个角色有其独特的视角, 也代表着在工作中要应对不同的挑战. 例如, 对于业务专家来说, 建立对相关技术的合理预期有非常重要的意义. 要为业务目标的设定提供更多依据, 避免盲目夸大机器学习技术的作用, 引入不必要的技术风险. 对于需求工程师, 除传统的业务建模外, 更关注隐私^[186]、安全^[187]、场景^[51]和目标演化^[179]等问题, 并提出了面向机器学习应用的需求建模方法. 软件工程师要应对的挑战是适应机器学习带来的软件组织和运行方式的变化, 尤其是模型的使用和维护、数据合规使用的问题. 领域专家则关注数据和知识的运用是否恰当. 而数据科学家更关注机器学习问题的定义和求解、数据质量等问题. 虽然关注重点不同, 但问题的解决普遍依赖这些角色间的协同合作.

2.3 干系人依赖关系及可解释需求

需求工程中有众多关于如何获取需求、对需求建模、规范需求、验证需求的方法、工具和最佳实践, 它们为可解释性需求的挖掘提供了重要基础. 例如: 使用KAOS^[187]进行面向目标的功能需求建模和分析, 使用NFR^[188]对非功能需求进行分析, 使用*i**^[189]对组织结构进行分析, 用交互场景来描述用例和系统行为等. 这些方法在工业智能应用的需求分析过程中依旧发挥作用. 例如, 在早期需求分析阶段, 可以采用Volere需求分析过程^[190]进行业务上下文理解、形成系统设计思想并进行验证.

虽然每个角色都有自己独特的关注点和视角, 但在可解释性需求的分析过程中, 干系人之间也存在依赖关系. 业务专家和领域专家分别承担着“业务信息挖掘”和“领域信息挖掘”的工作, 挖掘结果可作为可解释性评估的依据和来源. 数据科学家产出的

算法或者机器学习模型通常是黑盒的,往往会引入可解释性风险.可解释性的加强和减弱就体现在上述角色间的协作中,而传统需求分析方法并没有充分考虑相关问题.下面列举在需求工程的多角色协作中关注的4个焦点问题:

1)干系人共同设定合理的机器学习目标与期望.这个问题涵盖了非常广泛的领域和众多热点研究主题,其中最受关注的是可解释人工智能(explainable AI, XAI)^[81,191].在这个问题上,需要业务专家、需求工程师、领域专家、数据科学家等角色之间进行有效沟通^[3,51].首先是在业务理解阶段,需要多个角色共同来设定被业务专家认可的、达成可解释性的最终目标.同时,需求工程师则会将可解释性与可信性作为分析和评估候选模型的参照依据,保证可解释性需求不会引入不可接受的负面影响,例如过高的成本、过低的性能需求等.然后,数据科学家则要负责对候选方案进行数据层面的可解释性分析,尤其对于黑盒形态的机器学习模型,需要提供符合可解释性标准的预测结果.最终,解释的正确性需要领域专家参与研判.

2)运用领域知识提升机器学习模型的可解释性.领域知识可以是物理规律、逻辑规则或知识图谱,这些知识都为机器学习模型的可解释性提供了依据.在需求分析阶段,这些依据应该作为更细粒度的需求被提出,进而指导技术路线的决策.将这些约束、规则、关系等用于引导机器学习的特征工程、网络架构设计、损失函数设计中,可以显著增强机器学习与领域知识的联系.但为了解决这个问题,需要业务专家、领域专家和数据科学家之间的密切合作^[12,22,104].

3)结合领域特征设计适应机器学习应用的验证过程.将可解释性落实到技术路线中以后,就要考虑如何验证可解释性被正确地实现.由于软件系统和机器学习系统之间的架构差异,现有的软件验证方法不足以保障机器学习应用的可解释性需求的验证.利用领域知识,可以将机器学习模型的验证过程分为几个阶段,通过模型可视化、自动化测试、领域专家确认、用户反馈等方式逐步展开验证,在有效验证可解释性的同时降低验证成本^[20,159].

4)设计以数据流为核心的软件架构,确保机器学习的数据需求被满足.在数据驱动的机器学习模型开发中,架构设计必须考虑包括数据质量、数据处理、数据隐私等问题.其中,就需要建立对数据的正确认识.数据的可解释性依据在这一过程中发挥着重要作用.数据从采集到标注,再到训练和验证,要

经历一系列加工流程,为处理大数据量需考虑自动化解决方案,但同时也增加了数据个性化特征处理、数据质量保障的难度.数据的不断变化,也给正确理解数据、使用数据带来挑战.这需要软件工程师和数据科学家协作,在不降低模型效果和可解释性特点的前提下实现高质量、高效率的数据密集的应用软件架构^[181,192-193].

3 面向机器学习可解释性的需求分析

基于文献调研和综述分析结果,本文归纳总结了2种有助于机器学习应用需求分析的技术方向,并围绕其涉及的可解释性展开讨论.

3.1 面向业务目标的机器学习方案映射

数据驱动的智能软件需求工程需要跨越5个知识领域:将给定用例映射到相应的机器学习任务,构建模型所需要的数据流,完成机器学习模型的训练,评估模型以确认满足业务需求,并将模型部署为软件服务.此外,解决方案的设计原则被表示为一个上下文模型,模型中定义数据的状态、动机和技术约束.解决方案模式提供了数据驱动的智能需求或设计决策的多视角综合的视图,但在使用中还需要设计一个引导需求工程师按步骤推进的过程指南.该方法提供了机器学习技术的选择依据:首先,拆解后的具体业务问题可以较为直观地映射为机器学习问题;其次,在问题分解中考虑机理知识和业务上下文的影响因素,可以将一些可解释性信息融入到方案选择中;再进一步,通过分析数据现状,可以让数据中蕴含的规律、分布特点等作为可解释性信息辅助机器学习技术的选择.

该方案适合业务逻辑较复杂、层次清晰、可拆分为多个具体业务问题的用例,并能够针对各个业务问题提出明确的输入和输出.相反,如果业务逻辑较为简单,或者难以拆分为输入和输出明确的子问题,则不适合采用该方案.但是机器学习应用程序的开发通常障碍重重,失败的原因不胜枚举.例如,较差的数据条件、不正确的超参数设置或算法选择不合理等.因此,需要定义好解决方案的评估标准,包括性能指标、置信度、稳定性、训练成本等.

图1概述了从用例到机器学习算法映射过程中需要考虑的业务规则、机理、数据、算法等问题.其中机理业务上下文背后蕴含了可解释性的逻辑部分,表现为因果关系、规则约束和生活常识等.而数据现状中蕴含了可解释性的数据部分,包括数据的频域

特征、值域范围以及特征之间的相关性等.上述可解释性因素都将在选择解决方案时发挥重要作用.从用例逐步分解到数据驱动的机器学习需求必须考虑

逻辑和数据2方面的多种可解释性要素的影响,从而提供更多参考依据,使得机器学习模型可以按照预期发挥作用.

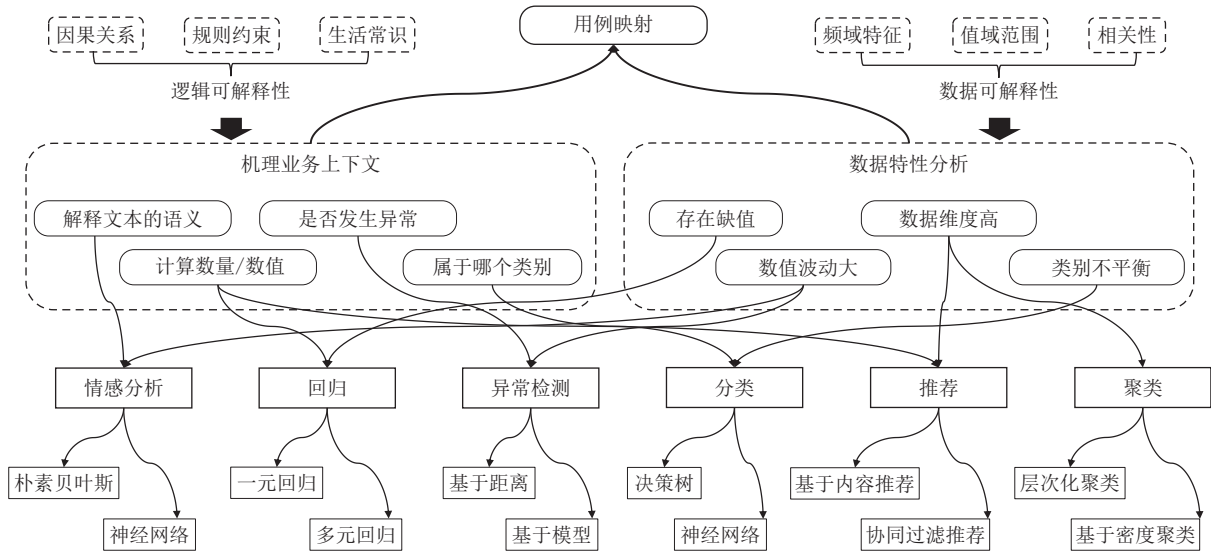


Fig. 1 Examples of mapping from use cases to machine learning tasks

图1 从用例到机器学习任务的映射示例

3.2 知识、规则与机器学习模型的融合

在大多数工业应用的开发中,领域知识起着关键作用.当谈到面向领域的需求工程案例时,领域知识对于需求工程师理解业务上下文和业务目标至关重要.例如,在安全工程领域,有多种用于评估潜在危害事件的方法和工具,如危险和可操作性分析(hazard and operability analysis, HAZOP)^[194]、故障模式和影响分析(failure modes and effects analysis, FMEA)^[195]、影响和关键性分析(effects and criticality analysis, FMECA)^[196]、保护层分析(layer of protection analysis, LOPA)^[197]、故障树分析(fault-tree analysis, FTA)^[198]和事件树分析(event tree analysis, ETA)^[199]等.这些方法被广泛应用于化工、制药和核电安全工程领域的研究与实践,领域从业者依此构建信息系统来评估和管理潜在的事故风险.近年来,相关的领域分析工具开发商也在对现有功能进行智能化扩展.为构建可用的机器学习应用,机器学习与领域知识的融合必不可少,其中有些依靠领域中的数据特性,而更显著的结合要依靠领域中的先验知识.单纯的机器学习模型开发过程,仅将数据输入机器学习数据流,经过调参验证等生成模型,未在方法论上给特定领域知识留出空间.因此,我们需要新的机器学习模型设计方法,将知识纳入机器学习的开发过程中. Von Rueden 等人^[191]完成了知识与机器学习融合的综述,

描述了不同的知识表示,如代数方程、逻辑规则和模拟结果如何用于机器学习.该工作提出了4个集成方向,包括训练数据生成、假设集定义、学习算法改进和最终假设检查.该工作描述了30多种不同知识和机器学习的整合策略.例如,科学知识可以作为深度学习模型损失函数的设计约束.基于图拉普拉斯矩阵的正则化表示可以约束相关变量在模型预测中的表现.图2中给出知识与机器学习融合的可能方式.

图2的方案适用于领域知识有多种表示形式,甚至其形式化程度也不同的场景.例如,数学方程表示足够具体,因此可以直接用于需求分析和开发.有的机理知识虽然是明确的,但在分析应用需求时,需要业务专家与领域专家综合考虑业务上下文和机理知识的关联性.另外,业务规则也是极为常见的知识表达形式,广泛存在于工业应用中.还有一些特殊形式的知识则以不同的方式发挥着各自的特殊作用.例如,知识图用于表示因果关系,领域专家的经验可以用于特征选择和模型检验.同时,领域知识通常是不完备、有缺陷的.对于可靠性及实效性较差的领域知识或复杂多变的业务规则,则不建议与机器学习技术融合运用,以免引入更多不确定性,导致技术方案难以评价和改进.同样地,可以通过机器学习模拟或者修正领域知识,持续提升对领域的认知,其可以看做通过机器学习技术实现的从数据中挖掘领域知识

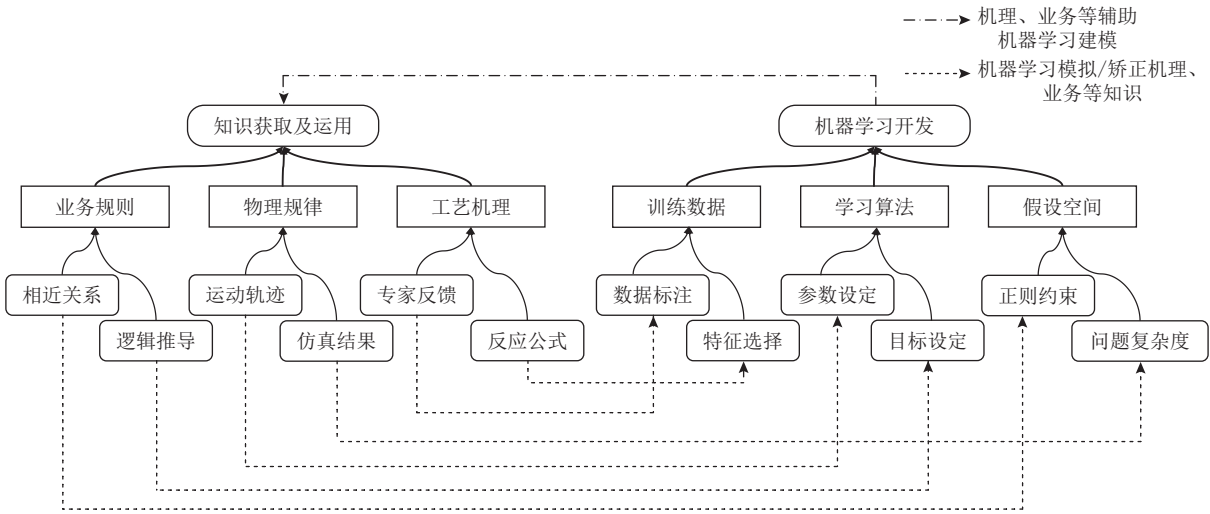


Fig. 2 Examples of integration method of knowledge and machine learning
图2 知识与机器学习融合方式举例

的方法.对于这种从数据加工得到的领域知识,其可解释性较弱,在使用时并不能完全等同于原有的领域知识,但可以降低人工数据分析的难度.典型场景如机器学习辅助数据标注,以及机器学习辅助天气预报等.

但同时知识与机器学习的融合过程需要建立知识、规则、数据间的平衡,通过该平衡降低机器学习技术的风险,加强技术方案的可解释性.但如何达到高效、适度的平衡需要较多专家经验,融合方案往往需要具体问题具体分析,尚无普适的经验,过于偏重知识和规则会限制通用机器学习技术发挥价值.为实现更好的知识融合运用,机器学习技术的发展需要深入应用领域,在应用领域中定义新问题,从而形

成面向特定领域的机器学习方法论,建立领域内的知识融合技术.

4 可解释性需求分析过程框架

本节提出一种可解释性需求分析过程框架,旨在明确机器学习应用开发中可解释性需求分析的作用,并对其中涉及的重点问题展开讨论.

通过对相关文献的调研与分析,我们将可解释性需求工程执行过程归纳为6个阶段及其对应的产出物,分别为业务上下文、技术方案、数据需求、技术指标、风险指标、业务目标等.每个阶段通过前后步骤相互衔接,形成基于多主体协作的面向机器学

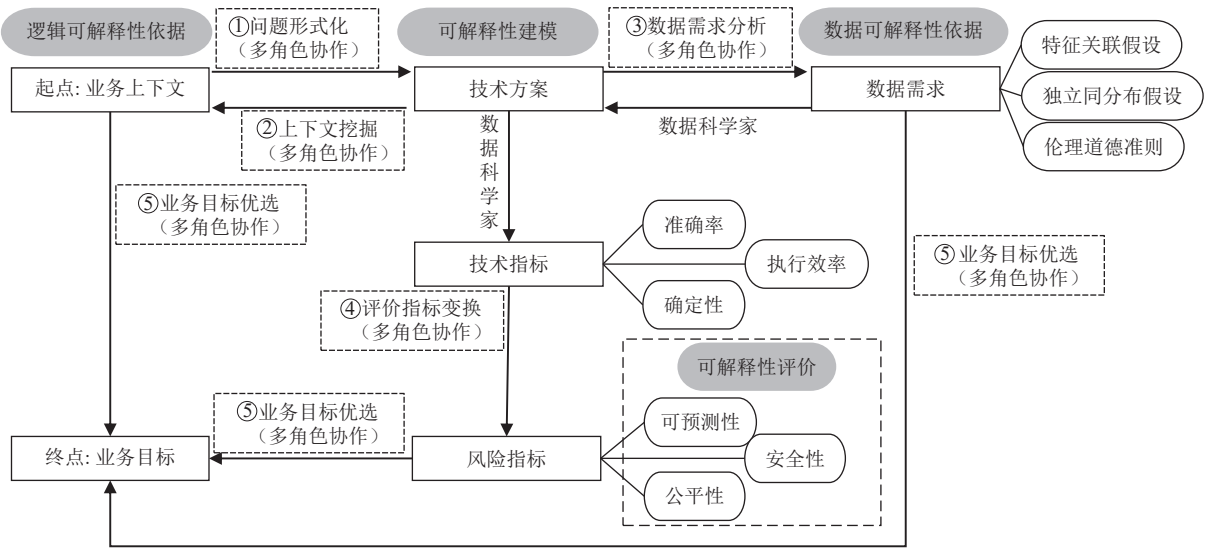


Fig. 3 Process framework of collaborative requirements engineering for machine learning
图3 面向机器学习的协作需求工程过程框架

习应用的需求工程分析过程,如图3所示.在本节框架中,可解释性需求被拆分为逻辑、数据2个方面,在每个方面上要通过需求分析得到相应的可解释性依据,并用于业务建模、技术路线决策等后续工作.其中,业务上下文中包含了可解释性需求的逻辑依据,通常体现为在问题描述中加入因果关系、逻辑约束等.在数据需求中包含了可解释性需求的数据依据,其可能体现为根据机理知识对数据特性提出假设,也可体现为通过实际数据的观测结果形成的一些数据特性的总结归纳.上述可解释性依据会用于建立业务建模、增加需求约束条件、引导技术方案的设计等.

我们对框架中需要多角色协作的步骤做了标注,同时框架中也包含一些由单一角色独立完成的工作.本文重点关注该框架中需要多角色协作的任务.

4.1 问题形式化

对问题的形式化描述是连接业务专家和数据科学家之间知识鸿沟的桥梁.不同领域的业务问题,通常有惯用的需求表示形式,例如文字描述、流程图、用例集合、示意图等.原始需求通常难以准确地转化为标准的机器学习问题.

例如,在本文的案例中,需求可以简单地表达为“通过计算机视觉技术识别视频中的挖掘机完成了多少次装载操作”.虽然语义上很清楚,但我们并不能找到一个直接解决这个问题的机器学习模型.究其原因包括2方面:一方面在于实际应用中的业务问题千变万化,很可能无法在输入和输出上适配已知的机器学习标准问题定义;另一方面,很多机器学习标准问题的求解技术尚未成熟,需要尽可能将业务问题转化为相对成熟的技术可解决的问题.业务问题的形式化,就是一个从“泛泛而谈”到“精准定位”的过程,不断压缩业务问题的不确定性空间,不断提升机器学习技术应用的可行性.

业务专家首先提出一个完整的业务问题,尽可能全面地定义业务问题的上下文.数据科学家将试图从业务问题中抽离出1个或多个标准的技术问题.典型的技术问题表达形式包括数学方程、逻辑规则和机器学习范式等.一旦技术问题确定下来,数据科学家就可以利用机器学习、数据挖掘等技术设计问题的解决方案.然而,在复杂场景中,问题形式化面临着诸多挑战.因为在项目早期,对需求的认知和对领域知识的掌握通常是不完整的,这使得我们很难清楚地识别需求的边界、技术的约束与局限、定性定量评价指标的选择等.

4.2 上下文挖掘

一次问题形式化当然不能准确捕捉全部的需求细节,一旦该工作不到位,遗漏影响问题定义的关键因素,将很可能导致后续研发工作失败.这就需要业务专家、领域专家和数据科学家充分沟通协作,在业务上下文中不断挖掘,为问题形式化提供更多参考依据,包括问题的作用域、领域知识、业务规则、常识、功能目标和细节等.

通过引入机理知识、业务规则、常识等信息,作为需求的逻辑可解释性依据,可以在更大程度上保证问题定义的完整性,降低问题定义出错的风险.最终建立从用例到机器学习任务的映射,如图1所示,其中逻辑可解释性依据往往与业务上下文紧密结合,辅助建立业务模型,加深对业务的理解.

该工作并不是单纯依靠业务专家和领域专家不断提供新的信息,更多地是需要数据科学家根据技术问题的不确定性定向地引导上下文的挖掘,才能让需求分析过程更有针对性.该工作涉及到数据科学与多个其他学科知识的交叉运用,详见图2所示的过程.

4.3 数据需求分析

上下文挖掘帮助我们更好地完成问题形式化,但要进一步选择合适的机器学习技术解决方案,就必须对数据进行充分的需求分析.实际场景中所具备的数据条件是技术方案可行性的重要评判依据,典型的数据条件包括数据量大小、数据集完备程度、标注质量高低、数据分布假设是否满足等.业界实践表明,在机器学习应用研发过程中,约80%的时间用于数据准备过程.上述比例并不包括需求分析所需时间,而这部分时间通常占研发所需时间的10%~20%,但足以说明数据理解 and 需求分析的严谨性和充分性,是避免造成数据处理过程反复,甚至影响模型的效果,带来的额外研发成本的重要保证.数据可解释性需求的分析能够在很大程度上排除对数据的错误认识,帮助深入理解数据和正确处理数据.典型数据可解释性依据是对数据机理特性、统计特性的刻画.

与上下文挖掘工作类似,数据需求分析也是一个多角色协同的过程.一方面,业务专家负责采集数据、标注数据,领域专家提供与数据相关的领域知识.另一方面,数据科学家要结合数据条件选择适用的技术方案,并提出待验证的数据假设.同时,领域专家和业务专家还需要配合数据科学家完成数据假设的验证.更复杂的情形是,业务专家需要在数据假

设中引入业务目标需要关心的伦理假设、公平性假设等. 这些数据条件将共同制约技术方案的选择. 其中通过验证的部分即可作为数据可解释性依据, 为最终成果的可解释性评价提供支持.

在模型研发中, 需要针对上述工作获得的逻辑可解释性依据、数据可解释性依据进行设计, 包括引入正则化约束、调整损失函数, 甚至可能进一步拆解问题定义, 形成由多个机器学习模型级联组成的复杂解决方案.

4.4 评估指标定性定量转化

机器学习模型与系统需要满足的技术指标由数据科学家根据技术方案提出, 用以量化评估方案的性能表现、执行效率等. 但技术指标常常不易被用户理解, 换言之, 技术指标并不直接表达需求方关心的业务目标. 同时, 在缺少可解释性依据的情况下, 评估指标的可解释性也无从谈起, 大大影响模型适应能力和用户认可度. 所以, 从技术指标向用户可以理解的风险指标进行转化, 从定量或者定性的角度呼应需求分析中得到的逻辑可解释性依据和数据可解释性依据, 在生产实践中具有重要意义. 基于可解释性依据定义评估指标, 将为软件带来更大的可靠性, 帮助最终用户加深对需求内涵和应用效果的理解.

指标的转换需要数据科学家和需求工程师双向沟通合作. 一方面, 数据科学家从技术视角提供相关的标准技术指标, 并根据需求工程师的建议加以补充和调整. 另一方面, 需求工程师从业务视角提出用户期待的风险指标, 并在数据科学家的帮助下建立风险指标和技术指标的关联, 进一步提出需要补充或调整的评估内容. 尤其针对需求分析中的可解释性依据的部分, 应尽可能设立对应的评估指标, 以全面地考量技术方案的可靠性.

4.5 业务目标优选

在需求工程师的主导下, 根据 4.1~4.4 节工作的结果得到最终业务目标. 该工作的复杂程度取决于协同 4.3 节工作中的需求分析结果, 形成协调一致的、恰当的、满足业务要求的完整需求定义. 在这个过程中, 既要根据业务需要调整风险指标及其验证标准, 又要根据技术的成熟度和作用域调整业务目标的预期范围, 在保持技术方案与业务需求一致的同时, 合理评估技术现状, 以保证方案具有足够高的可行性. 其中的典型问题包括: “哪些业务目标不在机器学习范围内” “风险指标支撑了哪些业务目标” “业务目标的数据条件有哪些” 等. 该工作由需求工程师主导, 但仍需业务专家、领域专家、数据科学家、软件工程师密切配合, 使得相关角色明晰各自职责, 建立互相支撑的协作关系, 统一各方的工作目标.

4.6 可解释性关注点

可解释性需求分析过程框架可指导各个角色明确机器学习应用的需求分析中与可解释性相关的工作. 如表 3 所示, 框架涉及业务专家、领域专家、数据科学家之间的协作, 但考虑到应用开发者和最终用户在可解释性上的重要作用, 将这 2 个角色一起列入表 3 中. 我们以工业、气象领域的机器学习应用为参考, 总结归纳出各个角色关注的逻辑可解释性依据、数据可解释依据, 以及在可解释性建模及评价上通常会关注的问题. 其中, 逻辑可解释性依据、数据可解释依据给建模阶段提供了重要的参考, 也可以作为可解释性的评价要素. 由此可见, 可解释性并不是单纯的机器学习技术问题, 而是涉及了业务规则、领域知识、数据特性、机器学习技术、应用开发等多个方面的需求分析对象. 通过在需求分析中引入可解释性需求要素, 可以增强各个角色对智能软件需求内容的理解, 保障需求分析结果的完整性、可靠性.

Table 3 Interpretability Related Steps of Each Role's Participation

表 3 各个角色参与的可解释性相关步骤

角色	逻辑可解释性依据	数据可解释依据	可解释性建模	可解释性评价
业务专家、领域专家	提供机理知识、提供业务规则、提供政策约束	提供数据描述	提供专家经验	符合机理知识、业务规则、政策约束、专家经验
数据科学家	收集逻辑可解释性依据	分析数据、提出数据分布假设、提出特征关联假设	融合建模、实现可解释性需求	验证可预测性、验证公平性
应用开发者	搭建反馈回路	提供相关数据、保护隐私安全	可操作性、控制成本	准备测试集
最终用户	提供使用反馈	提出隐私安全需求	提出公平性需求	验证失败案例

4.7 框架各步骤产出的内容

我们进一步详细列举图 3 框架主要步骤的产出内容, 并就产出内容的合规性提出若干要求, 如表 4

所示. 从表 4 中可以看出, 本文所提出的需求分析过程框架遵循机器学习软件开发的基本过程, 同时将逻辑可解释性要素、数据可解释性要素融入到需求

Table 4 Output Content of Each Step of the Demand Analysis Framework

表 4 需求分析框架各步骤的产出内容

产出内容	监督产出的角色	核心内容	合规要求
问题形式化	数据科学家	标准问题、约束条件、作用域、预期效果	准确无歧义,符合技术规范
上下文挖掘	业务专家	业务规则、领域机理、相关政策	逻辑完整,充分体现业务全貌
数据需求分析	数据科学家	数据语义、分布假设、标注规范、质量和特性	覆盖技术考量要素,满足伦理和公平性
评价指标转换	需求工程师	功能指标、非功能指标、可解释性依据、潜在风险点	用户可理解,覆盖可解释性依据
业务目标定义	业务专家	功能目标、作用范围、业务价值、非功能目标	满足业务需要,高技术可行性

分析的步骤中.其中,明确将可解释性作为需求分析中的核心要素,以及作为需求分析的产出,并在最终的评价指标中要求以可解释性依据作为必要的评价手段,为机器学习需求分析提供了可操作的参考依据.

5 面向机器学习应用需求的相关问题

面向机器学习应用的需求分析在不同的项目上下文环境中回答不同的重点问题.根据分析角色和视角的不同,需要获取和分析的需求是不同的.数据科学家的主要目标是为给定任务准备合适的算法和模型,设计一个能够适应场景的解决方案,并围绕这一目标挖掘可解释性依据.而需求工程师更关注整体需求与业务目标的一致性,以及可行性、可靠性等问题.领域专家则更倾向于找到更准确的领域知识表达形式,积累可复用的领域知识和经验,以及挖掘更严格的可解释性验证手段.上述不同角色所面对的问题中,仍然有很多场景没有被充分讨论,还潜藏一些重要问题需要研究者们展开更深入的研究.当然,很多项目中,上述角色可由一人担当或多人同时担任.下面列举 3 个典型的多角色协同解决的需求分析难题:

1)需求模型如何适应场景动态变化,建立可持续的机器学习范式.因为实际生产活动处于动态变化的场景中,数据现状会发生变化,甚至机理知识和业务规则也会发生演化.这将影响可解释性依据的可靠性.在这种情况下,各个角色的协作需要考虑更多因素,尤其是环境变化带来的不确定性因素.

2)如何有效地对项目进行可靠的成本估算.在传统的软件开发中,成本估算是必不可少的.然而,机器学习技术给这项工作带来了许多挑战.意外引入的成本可能来自过程中的任何步骤,包括数据收集、训练、模型部署和服务,以及模型的频繁更新等.在成本估算中,可解释性依据由于其对技术选型的

支撑作用,能够有效帮助降低技术成本的评估难度.

3)如何实现基于仿真原型的设计以提高方案设计效率.仿真模拟是分析复杂场景的重要手段,有助于完成可解释性依据的验证,进一步提高机器学习方案的设计效率.在游戏、仿真实验等领域的推动下,物理引擎技术已经较为成熟.但面对复杂的工业场景,在涉及化学、机械等一系列学科交叉的应用需求中,仿真技术面临着巨大挑战.

6 案例研究

结合前面综述研究和归纳的结果,尤其是表 2 和表 3 的框架,我们对一个基于视觉信息的挖掘机监控应用进行回顾性分析.在这个案例中,我们需要开发一个基于智能设备的自动化工作量计数解决方案,以取代挖掘机现场的人工监督.主要任务是统计挖掘机操作员的工作量,将一斗物料从地面拾起并装入卡车,即为 1 次挖掘装载动作,如图 4 所示.对于挖掘机租赁公司来说,跟踪每台租赁挖掘机的工作量非常重要,在过去几年中,这项任务带来了高昂的



Fig. 4 Intelligent monitoring case of construction equipment based on vision

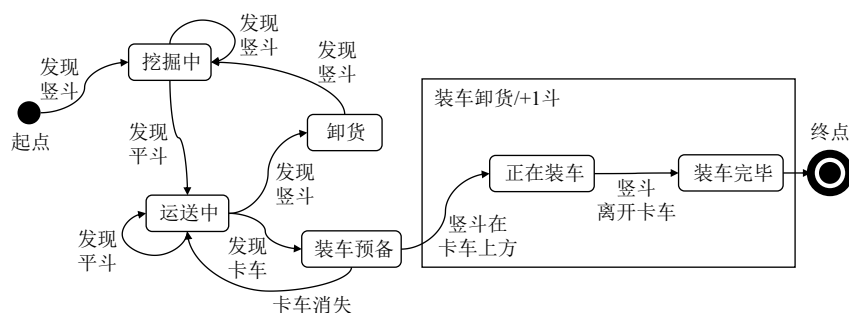
图 4 基于视觉的施工设备智能监控案例

人工成本,亟需自动化解决方案.我们将在该案例中说明如何利用可解释性需求分析手段挖掘逻辑和数据2方面的可解释性依据,以构建更具鲁棒性的技术路线.

问题需求描述和上下文挖掘考虑到机器学习在计算机视觉任务中的出色性能表现,我们尝试通过分析安装在挖掘机挡风前摄像头所采集的图像来分析挖斗动作.挖斗的操作按照视频帧的顺序记录下来,从中可以自动识别出挖斗和卡车,并完成计数.但挖掘机的工作过程不稳定,涉及到多种动作组合的可能性和异常情况,导致机器学习模型的运用存在较高的鲁棒性风险.这就需要通过可解释性需求分析手段,挖掘更多的可解释性依据,以加强技术路线的可解释性,降低鲁棒性风险.

在此案例中,我们从业务上下文中挖掘出挖斗的变化规律作为可解释性依据.在施工过程中,挖斗会在有限的几种状态之间按一定的顺序进行切换.为了建立这一可解释性依据与问题定义之间的联系,以状态图作为可解释性依据的形式化方法挖掘过程的业务逻辑,定义业务状态;以挖斗和卡车的识别作为触发状态变化的事件辅助机器学习模型的使用.

如图5所示,我们将挖掘机的工作过程分为7个状态,分别为“挖掘中”“运送中”“装车预备”“卸货”“正在装车”“卸货完毕”“装车完毕”等.状态转移事件则是依靠机器学习模型检测的具体事件,如“发现竖斗”“发现平斗”“发现卡车”“卡车离开”等.挖掘机在工作过程中,会在机器学习模型的帮助下,按照既定的状态转移关系在上述状态之间切换.



(a) 挖掘机的状态图



(b) 状态转彩记录

Fig. 5 The state diagram is used as an interpretable basis

图5 以状态图作为可解释性依据

上述工作中,我们定义了机器学习的标准问题,明确了机器学习模型的作用域,并基于机器学习模型构建了上层的业务逻辑实现方案.可解释性需求被提炼为2个具体内容:图像中的可识别状态的语义表示和描述业务推理过程的状态图.

我们通过状态机模型,将业务理解和技术路线更紧密地联系在一起.一方面通过可解释性依据建立业务模型,可以更准确地表达业务的可解释性需求,排除掉原有需求描述中不符合可解释性的矛盾点,加强需求的逻辑性和可解释性的逻辑依据;另一方面可以在更细的粒度上比较需求与技术是否匹配,通过持续跟踪当前的业务状态对机器学习模型的运行加以调试,及时发现机器学习模型的误判带来的业务状态偏差.

同时,这种基于可解释性的需求建模可以显著增强技术方案的鲁棒性.由于在每个业务状态下,机器

学习模型仅考虑该业务状态可能发生的几种状态转移事件,而不是所有事件,从而降低了误判的可能性,提升模型预测结果的鲁棒性.需求分析的具体步骤如下.

1) 数据需求分析.数据驱动的机器学习应用研发中,数据的理解、数据可获得性及质量的分析是重要的步骤.本文案例中的图像数据是通过固定在机械臂上的数字摄像机收集的.由于采用有监督学习方法,数据需求分析的挑战之一是定义训练数据的标注规则.不同人对图像数据中挖斗的标注有不同的理解.例如,有人将挖斗整体和突出物料全部放入标注框内,也有人更倾向标注挖斗的主体部分.人工标注的差异可能会忽略重要的细节,也可能引入干扰信息,严重影响机器学习模型的训练效果.因此,需要约束数据打标质量,建立符合业务目标的标注要求的可解释性依据,为机器学习模型训练提供可信的数据基础.例如,案例业务目标的实现关注挖斗

的姿态而非挖斗中的物料,所以打标工作中应该忽视物料,但突出挖斗尖端锯齿的位置.通过在数据打标中引入这种可解释性需求的约束,可以使得技术层面所需考虑的问题更加准确,减少“模糊地带”引入的需求与技术认知不一致.

另一个常见的数据需求是样本在数据类间分布的均衡性.样本数较少的目标识别准确性不高.本案例中,挖斗数据样本量相对大,但卡车数据样本量相对少,导致卡车的识别准确率偏低.类别不平衡也是一项常见的数据可解释性依据,通过在需求中补充这种可解释性依据,可以帮助技术人员更好地理解需求细节,减少错误假设带来的技术与需求的偏差.数据的多样性和图像的清晰度是图像数据的重要需求维度.这项工作由业务专家、需求工程师、数据科学家承担,有些场景中也需要领域专家的指导.

对于成熟的机器学习技术,如应用场景是较为固定的,则数据需求通常比较明确.但现实世界的开放性和复杂性导致我们通常只能应对部分典型场景,许多特殊长尾场景无法在早期的训练数据中获得.当面对新出现的场景时,模型的表现不够好,影响用户对机器学习技术的信任度,但背后的根本原因是需求发生了变化,需求中的问题范围超出了技术路线做出的假设.可见,可解释性需求,尤其是数据的理解与可解释性,是一个持续提升的认知过程,详见表3和表4中关于数据需求的分析和可解释性.在应用验证中,需要将需求分析阶段较为确定的可解释性依据作为验证手段和前提,对于超出原有目标场景的情况,要让用户知悉其不在评价范围内.此外,数据处理和数据质量改进所需的投入也是至关重要的.数据可用性、真实世界情况的复杂性和机器学习模型的通用性都可能会限制基于机器学习解决方案的表现.

2)评价指标变换.本案例中,在比较检测挖斗和卡车的多种候选机器学习算法时,一个关键的量化评估指标是物体检测算法的联合交集(intersection over union, IoU)指标.然而,该指标并非反映应用效果的“工作量计数准确率”.用户需要真阳性率和假阳性率等指标.模型的技术规范和业务最终目标的差异给技术方案的综合评价带来了挑战.本案例中,每种状态的误判率是用户关注的评价指标,也是用户可理解的反映技术方案优劣的指标,同时技术上也具有可实现性.在可解释性需求的基础上对技术指标进行评价,可以增强技术评价与原始需求的关联性,增加用户对技术方案的认可度.

3)业务目标定义.考虑到现实应用场景的变化性需求,在设定业务目标时,我们应该以训练集来约束机器学习模型的使用场景,尽可能降低未知场景带来的潜在风险.在训练集已经出现的场景中,同时采用真阳性率和假阳性率作为评判指标,并支持通过跟踪挖斗的状态变化调试模型预测错误.此外,应保留物体检测和计数记录,以备随机的人工检查.经过训练和验证的模型在部署后仍然可能失败,因为在现实世界的应用程序中经常会发生数据转移,例如光线影响、天气影响、角度影响等.为了保持满意的计数效果,需要持续补充训练数据,增加数据的丰富程度.

7 结论与未来工作

在机器学习应用的需求工程中,数据描述、性能指标、数据质量和候选解决方案必须围绕业务场景和技术可行性统一考虑,其中可解释性在多个环节起重要作用.为应对复杂的决策与预测分析任务,任何一个步骤的缺失都可能导致整个项目的失败.

本文从机器学习应用的需求工程中涉及的不同角色和视角进行了综述和分析.重点分析了业务专家、软件需求工程师、软件研发工程师、领域专家和数据科学家之间的依赖关系和协作任务,包括领域知识和机器学习模型的集成,以及如何按预期使用机器学习模型等,着重阐述了可解释性相关的问题.进一步总结了可用于支持需求活动的工具,例如从业务用例到机器学习任务的映射,以及融合先验知识和机器学习工作流的参考模式,还给出了数据驱动智能应用的案例.

未来研究工作主要是在工业领域,面向数据驱动的智能应用项目开展工程实践,基于成功和失败案例,对目前的方法框架开展广泛的实证研究;总结需求阶段要解决的关键问题,包括可解释性依据的构造方法、验证方法,并开展更深入的案例分析和共性问题研究;进一步完善和评估可解释型需求分析的过程和知识框架体系,并形成具有实际指导意义的方法指南,用于指导项目的需求定义和备选方案进行有效的协同决策.

作者贡献声明:裴忠一和刘璘为本文共同第一作者.裴忠一完成了主要调研工作和论文撰写;刘璘提出调研思路并完成了部分调研工作和论文撰写;

王晨参与文献整理和案例分析；王建民提出指导意见并修改论文。

参 考 文 献

- [1] Ashmore R, Calinescu R, Paterson C. Assuring the machine learning lifecycle: Desiderata, methods, and challenges[J]. *ACM Computing Surveys*, 2021, 54(5): 1–39
- [2] Habibullah K M, Horkoff J. Non-functional requirements for machine learning: Understanding current use and challenges in industry[C]//Proc of the 29th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2021: 13–23
- [3] Mohseni S, Zarei N, Ragan E D. A multidisciplinary survey and framework for design and evaluation of explainable AI systems[J]. *ACM Transactions on Interactive Intelligent Systems*, 2021, 11(3/4): 1–45
- [4] Mehrabi N, Morstatter F, Saxena N, et al. A survey on bias and fairness in machine learning[J]. *ACM Computing Surveys*, 2021, 54(6): 1–35
- [5] Floridi L. Establishing the rules for building trustworthy AI[J]. *Nature Machine Intelligence*, 2019, 1(6): 261–262
- [6] Wang Ye, Chen Junwu, Xia Xin, et al. Intelligent requirements elicitation and modeling: A literature review[J]. *Journal of Computer Research and Development*, 2021, 58(4): 683–705(in Chinese)
(汪烨, 陈骏武, 夏鑫, 等. 智能需求获取与建模研究综述 [J]. *计算机研究与发展*, 2021, 58(4): 683–705)
- [7] Iqbal T, Elahidoost P, Lucio L. A bird's eye view on requirements engineering and machine learning[C]//Proc of the 25th Asia-Pacific Software Engineering Conf (APSEC). Piscataway, NJ: IEEE, 2018: 11–20
- [8] Villamizar H, Escovedo T, Kalinowski M. Requirements engineering for machine learning: A systematic mapping study[C]//Proc of the 47th Euromicro Conf on Software Engineering and Advanced Applications (SEAA). Piscataway, NJ: IEEE, 2021: 29–36
- [9] Willard J, Jia Xiaowei, Xu Shaoming, et al. Integrating physics-based modeling with machine learning: A survey[J]. *arXiv preprint, arXiv: 2003.04919*, 2020
- [10] Harel D, Marron A, Sifakis J. Autonomics: In search of a foundation for next-generation autonomous systems[J]. *Proceedings of the National Academy of Sciences*, 2020, 117(30): 17491–17498
- [11] Horkoff J, Aydemir F B, Cardoso E, et al. Goal-oriented requirements engineering: An extended systematic mapping study[J]. *Requirements Engineering*, 2019, 24(2): 133–160
- [12] Wolny S, Mazak A, Carpella C, et al. Thirteen years of SysML: A systematic mapping study[J]. *Software and Systems Modeling*, 2020, 19(1): 111–169
- [13] Azevedo A, Santos M F. KDD, SEMMA and CRISP-DM: A parallel overview[C/OL]//Proc of the IADIS European Conf on Data Mining, 2008[2022-05-30]. <https://www.iadisportal.org/digital-library/kdd-semma-and-crisp-dm-a-parallel-overview>
- [14] Vogelsang A, Borg M. Requirements engineering for machine learning: Perspectives from data scientists[C]//Proc of the 27th IEEE Int Requirements Engineering Conf Workshops (REW). Piscataway, NJ: IEEE, 2019: 245–251
- [15] Amershi S, Begel A, Bird C, et al. Software engineering for machine learning: A case study[C]//Proc of the 41st IEEE/ACM Int Conf on Software Engineering: Software Engineering in Practice (ICSE-SEIP). Piscataway, NJ: IEEE, 2019: 291–300
- [16] Silva G R S, Canedo E D. Requirements engineering challenges and techniques in building chatbots[C]//Proc of the 14th Int Conf on Agents and Artificial Intelligence. Setúbal, Portugal: SCITEPRESS, 2022: 180–187
- [17] Bencomo N, Guo J L C, Harrison R, et al. The secret to better AI and better software (is requirements engineering)[J]. *IEEE Software*, 2021, 39(1): 105–110
- [18] Yoshioka N, Husen J H, Tun H T, et al. Landscape of requirements engineering for machine learning-based AI systems[C]//Proc of the 28th Asia-Pacific Software Engineering Conf Workshops (APSEC Workshops). Piscataway, NJ: IEEE, 2021: 5–8
- [19] Ahmad K, Bano M, Abdelrazek M, et al. What's up with requirements engineering for artificial intelligence systems?[C]//Proc of the 29th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2021: 1–12
- [20] Tukur M, Umar S, Hassine J. Requirement engineering challenges: A systematic mapping study on the academic and the industrial perspective[J]. *Arabian Journal for Science and Engineering*, 2021, 46(4): 3723–3748
- [21] Heyn H M, Knauss E, Muhammad A P, et al. Requirement engineering challenges for AI-intense systems development[C]//Proc of the 1st IEEE/ACM Workshop on AI Engineering-Software Engineering for AI (WAIN). Piscataway, NJ: IEEE, 2021: 89–96
- [22] Ferreira F, Silva L L, Valente M T. Software engineering meets deep learning: A mapping study[C]//Proc of the 36th Annual ACM Symp on Applied Computing. New York: ACM, 2021: 1542–1549
- [23] Dalpiaz F, Niu Nan. Requirements engineering in the days of artificial intelligence[J]. *IEEE Software*, 2020, 37(4): 7–10
- [24] Kostova B, Gurses S, Wegmann A. On the interplay between requirements, engineering, and artificial intelligence[J]. *Information and Software Technology*, 2023, 158: 107176
- [25] Belani H, Vukovic M, Car Ž. Requirements engineering challenges in building AI-based complex systems[C]//Proc of the 27th IEEE Int Requirements Engineering Conf Workshops (REW). Piscataway, NJ: IEEE, 2019: 252–255
- [26] Lwakatare L E, Raj A, Bosch J, et al. A taxonomy of software engineering challenges for machine learning systems: An empirical investigation[C]//Proc of the Int Conf on Agile Software Development. Berlin: Springer, 2019: 227–243
- [27] Bosch J, Olsson H H, Crnkovic I. It takes three to tango: Requirement, outcome/data, and AI driven development[C]//Proc of the 1st Int Workshop on Software-intensive Business: Start-ups, Ecosystems and Platforms (SiBW). Berlin: Springer, 2018: 177–192
- [28] Studer S, Bui T B, Drescher C, et al. Towards CRISP-ML (Q): A machine learning process model with quality assurance metho-

- dology[J]. *Machine Learning and Knowledge Extraction*, 2021, 3(2): 392–413
- [29] Nalchigar S, Yu E, Obeidi Y, et al. Solution patterns for machine learning[C]//Proc of the 31st Int Conf on Advanced Information Systems Engineering. Berlin: Springer, 2019: 627–642
- [30] Nalchigar S, Yu E, Keshavjee K. Modeling machine learning requirements from three perspectives: A case report from the healthcare domain[J]. *Requirements Engineering*, 2021, 26(2): 237–254
- [31] Chuprina T, Mendez D, Wnuk K. Towards artefact-based requirements engineering for data-centric systems[J]. arXiv preprint, arXiv: 2103.05233, 2021
- [32] Barrera J M, Reina Reina A, Maté A, et al. Applying *i** in conceptual modelling in machine learning[C]//Proc of the 14th Int iStar Workshop. Berlin: Springer, 2020: 56–62
- [33] Camilli M, Felderer M, Giusti A, et al. Towards risk modeling for collaborative AI[C]//Proc of the 1st IEEE/ACM Workshop on AI Engineering-Software Engineering for AI (WAIN). Piscataway, NJ: IEEE, 2021: 51–54
- [34] Alrajeh D, Cailliau A, van Lamsweerde A. Adapting requirements models to varying environments[C]//Proc of the 42nd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2020: 50–61
- [35] Deshpande A, Sharp H. Responsible AI systems: Who are the stakeholders?[C]//Proc of the 2022 AAAI/ACM Conf on AI, Ethics, and Society. New York: ACM, 2022: 227–236
- [36] Khuat T T, Kedziora D J, Gabrys B. The roles and modes of human interactions with automated machine learning systems[J]. arXiv preprint, arXiv: 2205.04139, 2022
- [37] Nahar N, Zhou Shurui, Lewis G, et al. Collaboration challenges in building ML-enabled systems: Communication, documentation, engineering, and process[C]//Proc of the 44th Int Conf on Software Engineering. New York: ACM, 2022: 413–425
- [38] Odong L A, Perini A, Susi A. Requirements engineering for collaborative artificial intelligence systems: A literature survey [C]//Proc of the 16th Int Conf on Research Challenges in Information Science. Berlin: Springer, 2022: 409–425
- [39] Barclay I, Abramson W. Identifying roles, requirements and responsibilities in trustworthy AI systems[C]//Proc of the 2021 ACM Int Joint Conf on Pervasive and Ubiquitous Computing and Proc of the 2021 ACM Int Symp on Wearable Computers. New York: ACM, 2021: 264–271
- [40] Piorkowski D, Park S, Wang A Y, et al. How AI developers overcome communication challenges in a multidisciplinary team: A case study[J]. *Proceedings of the ACM on Human-Computer Interaction*, 2021, 5(CSCW1): 131
- [41] Camilli M, Felderer M, Giusti A, et al. Risk-driven compliance assurance for collaborative AI systems: A vision paper[C]//Proc of the 27th Int Working Conf on Requirements Engineering: Foundation for Software Quality. Berlin: Springer, 2021: 123–130
- [42] Zhang A X, Muller M, Wang Dakuo. How do data science workers collaborate? Roles, workflows, and tools[J]. *Proceedings of the ACM on Human-Computer Interaction*, 2020, 4(CSCW1): 22
- [43] Berry D M. Requirements engineering for artificial intelligence: What is a requirements specification for an artificial intelligence? [C]//Proc of the 28th Int Working Conf on Requirements Engineering: Foundation for Software Quality. Berlin: Springer, 2022: 19–25
- [44] Al M, Ma Yihong, Alarcon P, et al. RESAM: Requirements elicitation and specification for deep-learning anomaly models with applications to UAV flight controllers[C]//Proc of the 30th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2022: 153–165
- [45] Villamizar H, Kalinowski M, Lopes H. Towards perspective-based specification of machine learning-enabled systems[J]. arXiv preprint, arXiv: 2206.09760, 2022
- [46] Gillani M, Niaz H A, Ullah A. Integration of software architecture in requirements elicitation for rapid software development[J]. *IEEE Access*, 2022, 10: 56158–56178
- [47] Gabriel S, Bentler D, Grote E M, et al. Requirements analysis for an intelligent workforce planning system: A socio-technical approach to design AI-based systems[J]. *Procedia CIRP*, 2022, 109: 431–436
- [48] Wang Xi, Miao Weikai. A framework for requirements specification of machine-learning systems[C]//Proc of the 32nd Conf on Software Engineering and Knowledge Engineering (SEKE). Pittsburgh, Pennsylvania: KSI Research Inc, 2022: 7–12
- [49] Ahmad K. Human-centric requirements engineering for artificial intelligence software systems[C]//Proc of the 29th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2021: 468–473
- [50] Schuh G, Scholz P, Leich T, et al. Identifying and analyzing data model requirements and technology potentials of machine learning systems in the manufacturing industry of the future[C]//Proc of the 61st Int Scientific Conf on Information Technology and Management Science of Riga Technical University (ITMS). Piscataway, NJ: IEEE, 2020: 20178653
- [51] Cirqueira D, Nedbal D, Helfert M, et al. Scenario-based requirements elicitation for user-centric explainable AI[C]//Proc of the 4th Int Cross-Domain Conf for Machine Learning and Knowledge Extraction. Berlin: Springer, 2020: 321–341
- [52] d'Aloisio G, Di Marco A, Stilo G. Modeling quality and machine learning pipelines through extended feature models[J]. arXiv preprint, arXiv: 2207.07528, 2022
- [53] Levy N, Katz G. RoMA: A method for neural network robustness measurement and assessment[J]. arXiv preprint, arXiv: 2110.11088, 2021
- [54] Siebert J, Joeckel L, Heidrich J, et al. Construction of a quality model for machine learning systems[J]. *Software Quality Journal*, 2022, 30(2): 307–335
- [55] Haindl P, Hoch T, Dominguez J, et al. Quality characteristics of a software platform for human-AI teaming in smart manufacturing[J]. arXiv preprint, arXiv: 2205.15767, 2022
- [56] Madaio M, Egede L, Subramonyam H, et al. Assessing the fairness of AI systems: AI practitioners' processes, challenges, and needs for support[J]. *Proceedings of the ACM on Human-Computer Interaction*, 2022, 6(CSCW1): 52

- [57] Perera A, Aleti A, Tantithamthavorn C, et al. Search-based fairness testing for regression-based machine learning systems[J]. *Empirical Software Engineering*, 2022, 27(3): 1–36
- [58] Chen Huaming, Babar M A. Security for machine learning-based software systems: A survey of threats, practices and challenges[J]. *arXiv preprint, arXiv: 2201.04736*, 2022
- [59] Georgieva I, Lazo C, Timan T, et al. From AI ethics principles to data science practice: A reflection and a gap analysis based on recent frameworks and practical experience[J]. *AI and Ethics*, 2022, 2: 697–711
- [60] Steimers A, Schneider M. Sources of risk of AI systems[J]. *International Journal of Environmental Research and Public Health*, 2022, 19(6): 3641
- [61] Bartels R, Dudink J, Haitjema S, et al. A perspective on a quality management system for AI/ML-based clinical decision support in hospital care[J]. *Frontiers in Digital Health*, 2022, 4: 942588
- [62] d'Aloisio G. Quality-driven machine learning-based data science pipeline realization: A software engineering approach[C]//Proc of the 44th ACM/IEEE Int Conf on Software Engineering: Companion Proceedings. Piscataway, NJ: IEEE, 2022: 291–293
- [63] Khan A, Siddiqui I F, Shaikh M, et al. Handling non-functional requirements in IoT-based machine learning systems[C]//Proc of the 7th Joint Int Conf on Digital Arts, Media and Technology with ECTI Northern Section Conf on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON). Piscataway, NJ: IEEE, 2022: 477–479
- [64] Truong H L, Nguyen T M. QoA4ML—A framework for supporting contracts in machine learning services[C]//Proc of the IEEE Int Conf on Web Services (ICWS). Piscataway, NJ: IEEE, 2021: 465–475
- [65] Muñante D, Perini A, Kifetew F M, et al. Combining risk and variability modelling for requirements analysis in SAS engineering[C]//Proc of the 29th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2021: 396–401
- [66] Eliyahu T, Kazak Y, Katz G, et al. Verifying learning-augmented systems[C]//Proc of the 2021 ACM SIGCOMM. New York: ACM, 2021: 305–318
- [67] Mauri L, Damiani E. STRIDE-AI: An approach to identifying vulnerabilities of machine learning assets[C]//Proc of the IEEE Int Conf on Cyber Security and Resilience (CSR). Piscataway, NJ: IEEE, 2021: 147–154
- [68] Dey S, Lee S W. Multilayered review of safety approaches for machine learning-based systems in the days of AI[J]. *Journal of Systems and Software*, 2021, 176: 110941
- [69] Wang Jingyi, Chen Jialuo, Sun Youcheng, et al. Robot: Robustness-oriented testing for deep learning systems[C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2021: 300–311
- [70] Baluta T, Chua Z L, Meel K S, et al. Scalable quantitative verification for deep neural networks[C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2021: 312–323
- [71] Zhang J M, Harman M. "Ignorance and Prejudice" in software fairness[C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2021: 1436–1447
- [72] Hobbs R. Integrating ethically align design into agile and CRISP-DM[C]//Proc of the Southeast Conf 2021. Piscataway, NJ: IEEE, 2021: 20631991
- [73] Tariq S, Cheema S M. Approaches for non-functional requirement modeling: A literature survey[C]//Proc of the 4th Int Conf on Computing & Information Sciences (ICCIS). Piscataway, NJ: IEEE, 2021: 21573902
- [74] Cerqueira J, Acco H, Dias E. Ethical guidelines and principles in the context of artificial intelligence[C]//Proc of the 17th Brazilian Symp on Information Systems. New York: ACM, 2021: 36
- [75] Byun T, Rayadurgam S. Manifold for machine learning assurance [C]//Proc of the 42nd IEEE/ACM Int Conf on Software Engineering: New Ideas and Emerging Results (ICSE-NIER). Piscataway, NJ: IEEE, 2020: 97–100
- [76] Ishikawa F, Matsuno Y. Evidence-driven requirements engineering for uncertainty of machine learning-based systems[C]//Proc of the 28th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2020: 346–351
- [77] Damyanova B. Quality attributes in AI-ML-based systems: Differences and challenges[D]. Stuttgart: University of Stuttgart, 2020
- [78] Balasubramaniam N, Kauppinen M, Kujala S, et al. Ethical guidelines for solving ethical issues and developing AI systems[C]//Proc of the Int Conf on Product-Focused Software Process Improvement. Berlin: Springer, 2020: 331–346
- [79] Riccio V, Tonella P. Model-based exploration of the frontier of behaviours for deep learning system testing[C]//Proc of the 28th ACM Joint Meeting on European Software Engineering Conf and Symp on the Foundations of Software Engineering. New York: ACM, 2020: 876–888
- [80] Wilhelm C, Younis A A. A threat analysis methodology for security requirements elicitation in machine learning based systems[C]//Proc of the 20th IEEE Int Conf on Software Quality, Reliability and Security Companion (QRS-C). Piscataway, NJ: IEEE, 2020: 426–433
- [81] Nakamichi K, Ohashi K, Namba I, et al. Requirements-driven method to determine quality characteristics and measurements for machine learning software and its evaluation[C]//Proc of the 28th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2020: 260–270
- [82] Humbatova N, Jahangirova G, Bavota G, et al. Taxonomy of real faults in deep learning systems[C]//Proc of the 42nd ACM/IEEE Int Conf on Software Engineering. Piscataway, NJ: IEEE, 2020: 1110–1121
- [83] Chechik M. Uncertain requirements, assurance and machine learning[C]//Proc of the 27th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2019: 2–3
- [84] Horkoff J. Non-functional requirements for machine learning: Challenges and new directions[C]//Proc of the 27th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2019: 386–391
- [85] Kuwajima H, Ishikawa F. Adapting square for quality assessment of

- artificial intelligence systems[C]//Proc of the IEEE Int Symp on Software Reliability Engineering Workshops (ISSREW). Piscataway, NJ: IEEE, 2019: 13–18
- [86] Bhatia J, Breux T D. Semantic incompleteness in privacy policy goals[C]//Proc of the 26th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2018: 159–169
- [87] Zhang Boyu, Magaña J C, Davoodi A. Analysis of security of split manufacturing using machine learning[C]//Proc of the 55th Annual Design Automation Conf. New York: ACM, 2018: 2767–2780
- [88] Hayrapetian A, Raje R. Empirically analyzing and evaluating security features in software requirements[C]//Proc of the 11th Innovations in Software Engineering Conf. New York: ACM, 2018: 9
- [89] Hoel T, Griffiths D, Chen Weiqin. The influence of data protection and privacy frameworks on the design of learning analytics systems[C]//Proc of the 7th Int Learning Analytics & Knowledge Conf. New York: ACM, 2017: 243–252
- [90] Burwinkel H, Matz H, Saur S, et al. Physics-aware learning and domain-specific loss design in ophthalmology[J]. *Medical Image Analysis*, 2022, 76: 102314
- [91] Von Rueden L, Mayer S, Beckh K, et al. Informed machine learning—A taxonomy and survey of integrating knowledge into learning systems[J]. *arXiv preprint*, arXiv: 1903.12394, 2019
- [92] Vo K, Nguyen T, Pham D, et al. Combination of domain knowledge and deep learning for sentiment analysis of short and informal messages on social media[J]. *International Journal of Computational Vision and Robotics*, 2019, 9(5): 458–485
- [93] Abbasi A, Nataraj C. Physics-informed machine learning for uncertainty reduction in time response reconstruction of a dynamic system[J]. *IEEE Internet Computing*, 2022, 26(4): 35–44
- [94] Huang Bin, Wang Jianhui. Applications of physics-informed neural networks in power systems—a review[J]. *IEEE Transactions on Power Systems*, 2022, 38(1): 572–588
- [95] Zhao W L, Gentine P, Reichstein M, et al. Physics-constrained machine learning of evapotranspiration[J]. *Geophysical Research Letters*, 2019, 46(24): 14496–14507
- [96] Lin Jing, Zhang Yu, Khoo E. Hybrid physics-based and data-driven modeling with calibrated uncertainty for lithium-ion battery degradation diagnosis and prognosis[J]. *arXiv preprint*, arXiv: 2110.13661, 2021
- [97] Cornelio J, Mohd Razak S, Cho Y, et al. Residual learning to integrate neural network and physics-based models for improved production prediction in unconventional reservoirs[J]. *SPE Journal*, 2022, 27(6): 1–22
- [98] Soleimani M, Intezari A, Pauleen D J. Mitigating cognitive biases in developing AI-assisted recruitment systems: A knowledge-sharing approach[J]. *International Journal of Knowledge Management*, 2022, 18(1): 1–18
- [99] Willard J, Jia Xiaowei, Xu Shaoming, et al. Integrating scientific knowledge with machine learning for engineering and environmental systems[J]. *ACM Computing Surveys*, 2022, 55(4): 1–37
- [100] Luo Xing, Zhang Dongxiao, Zhu Xu. Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge[J]. *Energy*, 2021, 225: 120240
- [101] Park S, Wang A Y, Kawas B, et al. Facilitating knowledge sharing from domain experts to data scientists for building NLP models[C]//Proc of the 26th Int Conf on Intelligent User Interfaces. New York: ACM, 2021: 585–596
- [102] Xie Xiaozheng, Niu Jianwei, Liu Xuefeng, et al. A survey on incorporating domain knowledge into deep learning for medical image analysis[J]. *Medical Image Analysis*, 2021, 69: 101985
- [103] Van Oort C M. Leveraging domain knowledge in deep learning systems[D]. Burlington: The University of Vermont and State Agricultural College, 2021
- [104] Chitchyan R, Bird C. Theory as a source of software requirements[C]//Proc of the 28th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2020: 227–237
- [105] Vázquez-Ingelmo A, García Peñalvo F J, Therón R. Advances in the use of domain engineering to support feature identification and generation of information visualizations[C]//Proc of the 8th Int Conf on Technological Ecosystems for Enhancing Multiculturality. New York: ACM, 2020: 1053–1056
- [106] Chai Yidong, Liu Hongyan, Xu Jie. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models[J]. *Knowledge-Based Systems*, 2018, 161: 147–156
- [107] Song Yangqiu, Roth D. Machine learning with world knowledge: The position and survey[J]. *arXiv preprint*, arXiv: 1705.02908, 2017
- [108] Haidry S Z, Falkner K, Szabo C. Identifying domain-specific cognitive strategies for software engineering[C]//Proc of the 2017 ACM Conf on Innovation and Technology in Computer Science Education. New York: ACM, 2017: 206–211
- [109] Laato S, Mäntymäki M, Minkinen M, et al. Integrating machine learning with software development lifecycles: Insights from experts[C/OL]//Proc of the 30th European Conf on Information Systems. Association for Information Systems, 2022[2022-05-30]. https://aisel.aisnet.org/ecis2022_rp/118
- [110] Washizaki H, Uchida H, Khomh F, et al. Machine learning architecture and design patterns[J/OL]. *IEEE Software*, 2020[2022-04-01]. http://g7.washi.cs.waseda.ac.jp/wp-content/uploads/2019/12/IEEE_Software_19__ML_Patterns.pdf
- [111] Lewis G A, Ozkaya I, Xu Xiwei. Software architecture challenges for ML systems[C]//Proc of the 37th IEEE Int Conf on Software Maintenance and Evolution (ICSME). Piscataway, NJ: IEEE, 2021: 634–638
- [112] Serban A, Visser J. An empirical study of software architecture for machine learning[J]. *arXiv preprint*, arXiv: 2105.12422, 2021
- [113] Chen Zhenpeng, Yao Huihan, Lou Yiling, et al. An empirical study on deployment faults of deep learning based mobile applications[C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2021: 674–685
- [114] Serban A, van der Blom K, Hoos H, et al. Adoption and effects of software engineering best practices in machine learning[C]//Proc of the 14th ACM/IEEE Int Symp on Empirical Software Engineering and Measurement (ESEM). New York: ACM, 2020: 3
- [115] John M M, Olsson H H, Bosch J. Developing ML/DL models: A design framework[C/OL]//Proc of the 14th Int Conf on Software and System Processes. New York: ACM, 2020[2022-05-30]. <https://>

- www.diva-portal.org/smash/get/diva2:1553907/FULLTEXT01.pdf
- [116] Correia J L, Pereira J A, Mello R, et al. Brazilian data scientists: Revealing their challenges and practices on machine learning model development[C]//Proc of the 19th Brazilian Symp on Software Quality. New York: ACM, 2020: 10
- [117] Reimann L, Kniesel-Wünsche G. Achieving guidance in applied machine learning through software engineering techniques[C]//Proc of the 4th Int Conf on Art, Science, and Engineering of Programming. New York: ACM, 2020: 7–12
- [118] Kourouklidis P, Kolovos D, Matragkas N, et al. Towards a low-code solution for monitoring machine learning model performance [C]//Proc of the 23rd ACM/IEEE Int Conf on Model Driven Engineering Languages and Systems: Companion Conf. Piscataway, NJ: IEEE, 2020: 62
- [119] Prado F F, Digiampietri L A. A systematic review of automated feature engineering solutions in machine learning problems[C]//Proc of the 16th Brazilian Symp on Information Systems. New York: ACM, 2020: 12
- [120] Oppold S, Herschel M. A system framework for personalized and transparent data-driven decisions[C]//Proc of the 32nd Int Conf on Advanced Information Systems Engineering. Berlin: Springer, 2020: 153–168
- [121] Ole M, Volker G. Towards concept based software engineering for intelligent agents[C]//Proc of the 7th IEEE/ACM Int Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE). Piscataway, NJ: IEEE, 2019: 42–48
- [122] Hesenius M, Schwenzfeier N, Meyer O, et al. Towards a software engineering process for developing data-driven applications[C]//Proc of the 7th IEEE/ACM Int Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE). Piscataway, NJ: IEEE, 2019: 35–41
- [123] Mucha T M, Ma Sijia, Abhari K. Sustainability of machine learning-based solutions: A lifecycle perspective[C/OL]//Proc of the Pacific Asia Conf on Information Systems. Association for Information Systems, 2022[2022-05-30]. <https://aisel.aisnet.org/pacis2022/262/>
- [124] Yu Kui, Guo Xianjie, Liu Lin, et al. Causality-based feature selection: Methods and evaluations[J]. ACM Computing Surveys, 2020, 53(5): 1–36
- [125] Cabrera Á A, Ribeiro M T, Lee B, et al. What did my AI learn? How data scientists make sense of model behavior[J]. ACM Transactions on Computer-Human Interaction, 2022, 30(1): 1–27
- [126] Sun Jiao, Liao Q V, Muller M, et al. Investigating explainability of generative AI for code through scenario-based design[C]//Proc of the 27th Int Conf on Intelligent User Interfaces. New York: ACM, 2022: 212–228
- [127] Shen M W. Trust in AI: Interpretability is not necessary or sufficient, while black-box interaction is necessary and sufficient[J]. arXiv preprint, arXiv: 2202.05302, 2022
- [128] Balasubramaniam N, Kauppinen M, Hiekkänen K, et al. Transparency and explainability of AI systems: Ethical guidelines in practice[C]//Proc of the 28th Int Working Conf on Requirements Engineering: Foundation for Software Quality. Berlin: Springer, 2022: 3–18
- [129] Piorkowski D, Richards J, Hind M. Evaluating a methodology for increasing AI transparency: A case study[J]. arXiv preprint, arXiv: 2201.13224, 2022
- [130] Sadeghi M, Klös V, Vogelsang A. Cases for explainable software systems: Characteristics and examples[C]//Proc of the 29th IEEE Int Requirements Engineering Conf Workshops (REW). Piscataway, NJ: IEEE, 2021: 181–187
- [131] Langer M, Baum K, Hartmann K, et al. Explainability auditing for intelligent systems: A rationale for multi-disciplinary perspectives [C]//Proc of the 29th IEEE Int Requirements Engineering Conf Workshops (REW). Piscataway, NJ: IEEE, 2021: 164–168
- [132] Chazette L, Brunotte W, Speith T. Exploring explainability: A definition, a model, and a knowledge catalogue[C]//Proc of the 29th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2021: 197–208
- [133] Velez M, Jamshidi P, Siegmund N, et al. White-box analysis over machine learning: Modeling performance of configurable systems [C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2021: 1072–1084
- [134] Narteni S, Ferretti M, Orani V, et al. From explainable to reliable artificial intelligence[C]//Proc of the 5th Int Cross-Domain Conf for Machine Learning and Knowledge Extraction. Berlin: Springer, 2021: 255–273
- [135] Zohdinasab T, Riccio V, Gambi A, et al. Deephyperion: Exploring the feature space of deep learning-based systems through illumination search[C]//Proc of the 30th ACM SIGSOFT Int Symp on Software Testing and Analysis. New York: ACM, 2021: 79–90
- [136] Liao Q V, Gruen D, Miller S. Questioning the AI: Informing design practices for explainable AI user experiences[C]//Proc of the 2020 CHI Conf on Human Factors in Computing Systems. New York: ACM, 2020: 1–15
- [137] Köhl M A, Baum K, Langer M, et al. Explainability as a non-functional requirement[C]//Proc of the 27th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2019: 363–368
- [138] Yang Zijiang, Al-Bahrani R, Reid A C E, et al. Deep learning based domain knowledge integration for small datasets: Illustrative applications in materials informatics[C]//Proc of the Int Joint Conf on Neural Networks (IJCNN). Piscataway, NJ: IEEE, 2019: 19028509
- [139] Wolf C T. Explainability scenarios: Towards scenario-based XAI design[C]//Proc of the 24th Int Conf on Intelligent User Interfaces. New York: ACM, 2019: 252–257
- [140] Eiband M, Schneider H, Bilandzic M, et al. Bringing transparency design into practice[C]//Proc of the 23rd Int Conf on Intelligent User Interfaces. New York: ACM, 2018: 211–223
- [141] Thakkar D, Ismail A, Kumar P, et al. When is machine learning data good?: Valuing in public health datafication[C]//Proc of the CHI Conf on Human Factors in Computing Systems. New York: ACM, 2022: 322
- [142] Jung J Y, Steinberger T, King J L, et al. How domain experts work with data: Situating data science in the practices and settings of craftwork[J]. Proceedings of the ACM on Human-Computer Interaction, 2022, 6(CSCW1): 58

- [143] Lwakatare L E, Ränge E, Crnkovic I, et al. On the experiences of adopting automated data validation in an industrial machine learning project[C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering: Software Engineering in Practice (ICSE-SEIP). Piscataway, NJ: IEEE, 2021: 248–257
- [144] Hutchinson B, Smart A, Hanna A, et al. Towards accountability for machine learning datasets: Practices from software engineering and infrastructure[C]//Proc of the 2021 ACM Conf on Fairness, Accountability, and Transparency. New York: ACM, 2021: 560–575
- [145] Shao Zhijun, Wu Ji, Zhao Wenxiao, et al. How data plays in the requirements of face recognition system: A concern driven systematic literature review[C]//Proc of the 28th Asia-Pacific Software Engineering Conf Workshops (APSEC Workshops). Piscataway, NJ: IEEE, 2021: 9–12
- [146] Astegher M, Busetta P, Perini A, et al. Specifying requirements for data collection and analysis in data-driven RE. A research preview[C]//Proc of the 27th Int Working Conf on Requirements Engineering: Foundation for Software Quality. Berlin: Springer, 2021: 182–188
- [147] Hayes J H, Payne J, Leppelmeier M. Toward improved artificial intelligence in requirements engineering: Metadata for tracing datasets[C]//Proc of the 27th IEEE Int Requirements Engineering Conf Workshops (REW). Piscataway, NJ: IEEE, 2019: 256–262
- [148] Biffl S, Lüder A, Rinker F, et al. Efficient engineering data exchange in multi-disciplinary systems engineering[C]//Proc of the 29th Int Conf on Advanced Information Systems Engineering. Berlin: Springer, 2019: 17–31
- [149] Mei Songzhu, Liu Cong, Wang Qinglin, et al. Model provenance management in MLOps pipeline[C]//Proc of the 8th Int Conf on Computing and Data Engineering. New York: ACM, 2022: 45–50
- [150] Kaminwar S R, Goschenhofer J, Thomas J, et al. Structured verification of machine learning models in industrial settings[J/OL]. Big Data, 2021[2022-04-01]. <https://www.liebertpub.com/doi/10.1089/big.2021.0112>
- [151] Xiao Yan, Beschastnikh I, Rosenblum D S, et al. Self-checking deep neural networks in deployment[C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2021: 372–384
- [152] Wang Song, Shrestha N, Subburaman A K, et al. Automatic unit test generation for machine learning libraries: How far are we?[C]//Proc of the 43rd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2021: 1548–1560
- [153] Hestness J, Ardalani N, Damos G. Beyond human-level accuracy: Computational challenges in deep learning[C/OL]//Proc of the 24th Symp on Principles and Practice of Parallel Programming. New York: ACM, 2019[2022-05-30]. <https://dl.acm.org/doi/10.1145/3293883.3295710>
- [154] Gharibi G, Walunj V, Rella S, et al. Modelkb: Towards automated management of the modeling lifecycle in deep learning[C]//Proc of the 7th IEEE/ACM Int Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE). Piscataway, NJ: IEEE, 2019: 28–34
- [155] Borg M, Englund C, Wnuk K, et al. Safely entering the deep: A review of verification and validation for machine learning and a challenge elicitation in the automotive industry[J]. arXiv preprint, arXiv: 1812.05389, 2018
- [156] Czvetkó T, Kummer A, Ruppert T, et al. Data-driven business process management-based development of Industry 4.0 solutions [J]. *CIRP Journal of Manufacturing Science and Technology*, 2022, 36: 117–132
- [157] Merkelbach S, Von Enzberg S, Kühn A, et al. Towards a process model to enable domain experts to become citizen data scientists for industrial applications[C]//Proc of the 5th IEEE Int Conf on Industrial Cyber-Physical Systems (ICPS). Piscataway, NJ: IEEE, 2022
- [158] Heyn H M, Subbiah P, Linder J, et al. Setting AI in context: A case study on defining the context and operational design domain for automated driving[C]//Proc of the 28th Int Working Conf on Requirements Engineering: Foundation for Software Quality. Berlin: Springer, 2022: 199–215
- [159] Zhang Ran, Albrecht A, Kausch J, et al. DDE process: A requirements engineering approach for machine learning in automated driving[C]//Proc of the 29th IEEE Int Requirements Engineering Conf (RE). Piscataway, NJ: IEEE, 2021: 269–279
- [160] Borg M, Bronson J, Christensson L, et al. Exploring the assessment list for trustworthy AI in the context of advanced driver-assistance systems[C]//Proc of the 2nd IEEE/ACM Int Workshop on Ethics in Software Engineering Research and Practice (SEthics). Piscataway, NJ: IEEE, 2021: 5–12
- [161] Martínez-Fernández S, Franch X, Jedlitschka A, et al. Developing and operating artificial intelligence models in trustworthy autonomous systems[C]//Proc of the 15th Int Conf on Research Challenges in Information Science. Berlin: Springer, 2021: 221–229
- [162] Kolyshkina I, Simoff S. Interpretability of machine learning solutions in public healthcare: The CRISP-ML approach[J]. *Frontiers in Big Data*, 2021, 4: 660206
- [163] Pinto A. Requirement specification, analysis and verification for autonomous systems[C]//Proc of the 58th ACM/IEEE Design Automation Conf (DAC). Piscataway, NJ: IEEE, 2021: 1315–1318
- [164] Leung C K, Fung D L, Mai D, et al. Explainable data analytics for disease and healthcare informatics[C]//Proc of the 25th Int Database Engineering & Applications Symp. New York: ACM, 2021: 65–74
- [165] León A, García S A, Costa M, et al. Evolution of an adaptive information system for precision medicine[C]//Proc of the 33rd Int Conf on Advanced Information Systems Engineering. Berlin: Springer, 2021: 3–10
- [166] Pereira A, Thomas C. Challenges of machine learning applied to safety-critical cyber-physical systems[J]. *Machine Learning and Knowledge Extraction*, 2020, 2(4): 579–602
- [167] Zhang Ru, Xiao Wencong, Zhang Hongyu, et al. An empirical study on program failures of deep learning jobs[C]//Proc of the 42nd IEEE/ACM Int Conf on Software Engineering (ICSE). Piscataway, NJ: IEEE, 2020: 1159–1170
- [168] du Preez A, Bekker J. A machine learning decision support framework for industrial engineering purposes[C]//Proc of the 1st Int Conf on Industrial Engineering and Industrial Management. New York: ACM, 2020: 9–14

- [169] Leung C K, Fung D L X, Mushtaq S B, et al. Data science for healthcare predictive analytics[C]//Proc of the 24th Symp on Int Database Engineering & Applications. New York: ACM, 2020: 8
- [170] Loucopoulos P, Kavakli E, Chechina N. Requirements engineering for cyber physical production systems[C]//Proc of the 31st Int Conf on Advanced Information Systems Engineering. Berlin: Springer, 2019: 276–291
- [171] Bao N, Chung S T. A rule-based smart thermostat[C]//Proc of the 1st Int Conf on Computational Intelligence and Intelligent Systems. New York: ACM, 2018: 20–25
- [172] Ma J, Park S C, Shin J H, et al. AI based intelligent system on the EDISON platform[C]//Proc of the 2018 Artificial Intelligence and Cloud Computing Conf. New York: ACM, 2018: 106–114
- [173] Gupta P, Suryavanshi A, Maheshwari S, et al. Human-machine interface system for pre-diagnosis of diseases using machine learning [C]//Proc of the 18th Int Conf on Machine Vision and Applications. New York: ACM, 2018: 71–75
- [174] Sothilingam R, Eric S K. Modeling agents, roles, and positions in machine learning project organizations[C]//Proc of the 15th Int iStar Workshop. Berlin: Springer, 2020: 61–66
- [175] Lim S, Henriksson A, Zdravkovic J. Data-driven requirements elicitation: A systematic literature review[J]. *SN Computer Science*, 2021, 2(1): 1–35
- [176] Kirikova M. Continuous requirements engineering[C/OL]//Proc of the 18th Int Conf on Computer Systems and Technologies. New York: ACM, 2017[2022-05-31]. <https://dl.acm.org/doi/10.1145/3134302.3134304>
- [177] Hartmann T, Moawad A, Fouquet F, et al. The next evolution of MDE: A seamless integration of machine learning into domain modeling[J]. *Software & Systems Modeling*, 2019, 18(2): 1285–1304
- [178] Arruda D, Laigner R. Requirements engineering practices and challenges in the context of big data software development projects: Early insights from a case study[C]//Proc of the IEEE Int Conf on Big Data (Big Data). Piscataway, NJ: IEEE, 2020: 2012–2019
- [179] Alrajeh D, Van Lamsweerde A, Kramer J, et al. Risk-driven revision of requirements models[C]//Proc of the 38th Int Conf on Software Engineering. New York: ACM, 2016: 855–865
- [180] Ishikawa F, Yoshioka N. How do engineers perceive difficulties in engineering of machine-learning systems?—Questionnaire survey [C]//Proc of the 7th IEEE/ACM Joint Int Workshop on Conducting Empirical Studies in Industry (CESI) and 6th Int Workshop on Software Engineering Research and Industrial Practice (SER&IP). Piscataway, NJ: IEEE, 2019: 2–9
- [181] Muccini A, Vaidhyanathan K. Software architecture for ML-based systems: What exists and what lies ahead[C]// Proc of 2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN). Piscataway, NJ: IEEE, 2021:121-128
- [182] Murdock R J, Kauwe S K, Wang A Y T, et al. Is domain knowledge necessary for machine learning materials properties?[J]. *Integrating Materials and Manufacturing Innovation*, 2020, 9(3): 221–227
- [183] Childs C M, Washburn N R. Embedding domain knowledge for machine learning of complex material systems[J]. *MRS Communications*, 2019, 9(3): 806–820
- [184] Latef M, Aslam T, Sehar P. Impact of domain knowledge in phase of requirement engineering[J]. *International Journal of Advanced Research and Development*, 2018, 3(6): 54–57
- [185] Fu Wei, Menzies T. Easy over hard: A case study on deep learning[C]//Proc of the 11th Joint Meeting on Foundations of Software Engineering. New York: ACM, 2017: 49–60
- [186] Caramujo J, Rodrigues A, Monfared S, et al. RSL-IL4Privacy: A domain-specific language for the rigorous specification of privacy policies[J]. *Requirements Engineering*, 2019, 24(1): 1–26
- [187] Darimont R, Delor E, Massonet P, et al. GRAIL/KAOS: An environment for goal-driven requirements engineering[C]//Proc of the 19th Int Conf on Software Engineering. New York: ACM, 1997: 612–613
- [188] Doerr J, Kerkow D, Koenig T, et al. Non-functional requirements in industry-three case studies adopting an experience-based NFR method[C]//Proc of the 13th IEEE Int Conf on Requirements Engineering (RE'05). Piscataway, NJ: IEEE, 2005: 373–382
- [189] Mesquita R, Jaqueira A, Lucena M, et al. US2StarTool: Generating *i** models from user stories[C]//Proc of the 10th Int iStar Workshop. Berlin: Springer, 2015: 103–108
- [190] Robertson S, Robertson J. Mastering the Requirements Process: Getting Requirements Right[M]. Reading, MA: Addison-Wesley, 2012
- [191] Saeed W, Omlin C. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities[J]. *arXiv preprint, arXiv: 2111.06420*, 2021
- [192] Nazir R. Studying software architecture design challenges, best practices and main decisions for machine learning systems[D]. Gothenburg, Sweden: Department of Computer Science and Engineering, Chalmers University of Technology, University of Gothenburg, 2021
- [193] Di Stefano J S, Menzies T. Machine learning for software engineering: Case studies in software reuse[C]// Proc of the 14th IEEE Int Conf on Tools with Artificial Intelligence (ICTAI 2002). Piscataway, NJ: IEEE, 2002: 246–251
- [194] Crawley F, Tyler B. HAZOP: Guide to Best Practice[M]. Amsterdam: Elsevier, 2015
- [195] Chiozza M L, Ponzetti C. FMEA: A model for reducing medical errors[J]. *Clinica Chimica Acta*, 2009, 404(1): 75–78
- [196] Luthra P. FMECA: An integrated approach[C]//Proc of the 37th Annual Reliability and Maintainability Symp. Piscataway, NJ: IEEE, 1991: 235–241
- [197] Willey R J. Layer of protection analysis[J]. *Procedia Engineering*, 2014, 84: 12–22
- [198] Ericson C A, Li C. Fault tree analysis[C/OL]//Proc of the 17th System Safety Conf. Saint Paul, USA: The International System Safety Society, 1999[2022-05-31]. <https://ftaassociates.files.wordpress.com/2018/12/C.-Ericson-Fault-Tree-Analysis-A-History-Proceedings-of-the-17th-International-System-Safety-Conference-1999.pdf>
- [199] Andrews J D, Dunnett S J. Event-tree analysis using binary decision diagrams[J]. *IEEE Transactions on Reliability*, 2000, 49(2): 230–238



Pei Zhongyi, born in 1987. PhD, assistant professor. Member of CCF. His main research interests include software engineering for machine learning, industrial intelligence.

裴忠一, 1987 年生. 博士, 助理研究员. CCF 会员. 主要研究方向为机器学习软件工程、工业智能化.



Liu Lin, born in 1973. PhD, professor. Her main research interests include requirements engineering, domain data and knowledge engineering.

刘 璘, 1973 年生. 博士, 研究员. 主要研究方向为需求工程、领域数据与知识工程.



Wang Chen, born in 1981. Associate professor. His main research interests include database and industrial big data.

王 晨, 1981 年生. 副研究员. 主要研究方向为数据库、工业大数据.



Wang Jianmin, born in 1968. PhD, professor. His main research interests include database, and industrial big data.

王建民, 1968 年生. 博士, 教授. 主要研究方向为数据库、工业大数据.