

DOI:10.13196/j.cims.2023.0639

基于 BERT 和 TextCNN 的智能制造成熟度评估方法

张 淦¹, 袁堂晓^{1,2}, 汪惠芬¹⁺, 柳林燕¹

(1. 南京理工大学 机械工程学院, 江苏 南京 210000;

2. 洛林大学 LCOMS, 梅斯 法国 57000)

摘 要:随着智能制造 2025 目标的临近,企业为了解自身能力水平纷纷加入到智能制造成熟度评估的行列中。然而,由于智能制造成熟度评估标准的复杂性,企业缺乏其对行业水平的了解,导致企业贸然申请,浪费自身时间的同时又占用大量评估资源。鉴于此,设计了一种新的评估流程,采用文本处理算法对整个评估过程进行了重构,通过利用国标文件中智能制造成熟度评估标准,将其作为训练集,采用基于预训练语言模型与文本神经网络(BERT+TextCNN)相结合的智能评估算法代替人工评估。在真实的企业智能制造数据集上的验证表明,当 BERT+TextCNN 评估模型在卷积核为[2,3,4]、迭代次数为 6 次、学习率为 $3e-5$ 时,对智能制造成熟度进行评估,准确率达到 85.32%。这表明所设计的评估方法能够较准确地帮助企业完成智能制造成熟度自评估,有助于企业了解自身智能制造能力水平,制定正确的发展方向。

关键词:智能制造成熟度模型;BERT 预训练语言模型;文本卷积神经网络;评估过程重构

中图分类号:TP18;TP29

文献标识码:A

Smart manufacturing maturity assessment method based on BERT and TextCNN

ZHANG Gan¹, YUAN Tangxiao^{1,2}, WANG Hui fen¹⁺, LIU Linyan¹

(1. School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;

2. LCOMS, University of Lorraine, Metz 57000, France,)

Abstract: As the goal of Smart Manufacturing 2025 approaches, enterprises are joining the ranks of Smart Manufacturing Maturity Assessment to understand their own capability level. However, due to the complexity of the smart manufacturing maturity assessment criteria, enterprises lack their understanding of their level in the industry, which leads them to apply hastily, wasting their own time while taking up a lot of assessment resources. A new assessment process was designed, and the whole assessment process was reconstructed using text processing algorithms. By utilizing the intelligent manufacturing maturity assessment standards in the national standard documents as a training set, an intelligent assessment algorithm based on the combination of pre-trained language model BERT and Text Convolutional Neural Networks (BERT+TextCNN) was used instead of manual assessment. Validation on a real enterprise smart manufacturing dataset showed that the BERT+TextCNN assessment model achieved an accuracy of 85.32% when it assessed smart manufacturing maturity with the convolution kernel of [2,3,4], six iterations and the learning rate of $3e-5$. This indicated that the proposed method could help enterprises complete the self-assessment of smart manufacturing maturity more accurately, which helped enterprises understand their own smart manufacturing capability level and formulate the correct development direction.

Keywords: capability maturity smart manufacturing model; BERT pre-trained language model; text convolutional neural networks; reconfiguration of assessment process

收稿日期:2023-10-19;**修订日期:**2024-01-09。Received 19 Oct. 2023;accepted 09 Jan. 2024.

基金项目:中华人民共和国科技部高端外国专家引进计划资助项目(G2022182015L);中华人民共和国工信部 2020 年工业互联网创新发展工程资助项目(TC200802F/001)。**Foundation items:** Project supported by the Ministry of Science and Technology of the People's Republic of China, High-end Foreign Experts Recruitment Plan, China (No. G2022182015L), and the Ministry of Industry and Information Technology of the People's Republic of China, 2020 Industrial Internet Innovation and Development Project, China(TC200802F/001).

0 引言

近年来世界上各工业大国为加快制造业转型先后提出符合自己国情的升级转型计划,引领制造业向数字化和智能化发展。美国和德国分别提出了“工业互联网”和“工业 4.0”战略,推动了全球制造业的变革。中国随后提出了“智能制造 2025”战略,明确数字化和智能化也必将成为我国制造业的归宿^[1]。

随着我国智能制造转型计划的推进,为了指导企业更好地进行智能制造改革,帮助他们了解自身情况和明确改进方向,工信部在 2020 年底发布了智能制造能力成熟度模型^[2],该模型结合了“中国制造 2025”战略计划和国内实际情况。然而,在申请智能制造成熟度等级认证时,尽管企业对自身能力有一定的认识,但由于智能制造成熟度标准的复杂性,企业缺乏对其行业水平的全面了解。如果企业盲目申请认证,一方面可能会浪费自身时间,另一方面授权机构面临着大量的企业申请,而评估专家资源有限,无法在短时间内对所有企业进行评估。因此,有必要在企业申请智能制造成熟度等级认证之前引入预处理机制,帮助企业了解当前的能力水平,从而缓解当下的矛盾。

与此同时,现行的智能制造成熟度评估方法在实际应用中也面临一些挑战。首先,企业需要对评估标准十分熟悉;其次,评估人员需要充分了解被评估企业的特点,这使得评估工作对评估专家依赖较

大;最后,采用专家打分的方法需要对数百条成熟度标准逐一进行打分评估,工作量巨大。这些问题导致现行评估方法不太适用于企业自评估,主要原因体现在以下 3 方面:①企业缺少专业评估人员;②工作量巨大导致评估周期过长;③评估结果受主观影响过大。

企业的智能制造成熟度评估主要依据生产经营现状描述,因此本文选用自然语言处理方法构建智能评估算法,与传统成熟度评估方法相比,其具备更贴合企业生产现状以及受主观影响小的特点与优势。为解决企业在进行智能制造成熟度评估过程面临的几方面问题,本文将重新构建智能制造评估过程,引入智能评估方法代替传统的人工评估。这种方法不仅可以改善现行评估方法中存在的问题,还可以帮助企业更快速地了解自身的能力水平,实现自评估。

1 智能制造成熟度评估现状分析

1.1 智能制造能力成熟度框架

本文采用我国工业和信息化部于 2020 年底发布的《智能制造能力成熟度模型》(GB/T 39116—2020)。该模型由成熟度等级、能力要素和成熟度要求构成,具体描述如图 1 所示。成熟度等级反映了企业在不同阶段应达到的智能制造能力水平,能力要素则突出了成熟度模型对企业关键方向的考察,同时也是企业在提升自身智能制造能力时需要改进的重要方面。成熟度要求则指明了在不同成熟度等级下,各个能力要素甚至能力子域需要满足的条件。

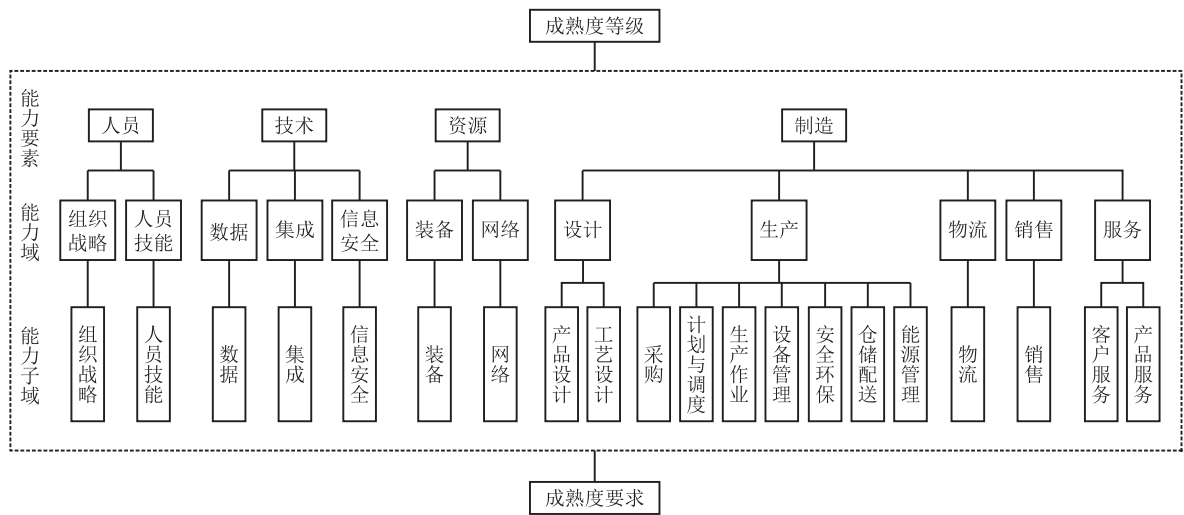


图 1 智能制造成熟度模型构成

该成熟度模型由 5 个等级组成:第一级是规划级,第二级是规范级,第三级是集成为级,第四级是优化级,第五级是引领级。高等级成熟度的要求包含低等级成熟度的所有要求。智能制造的能力要素包括制造、人员、技术和资源 4 个方面。这些能力要素反映了通过人员的应用、资源的调配以及技术的运用来提升智能制造能力的过程。该智能制造能力成熟度模型涵盖了 12 个能力域,并进一步细分为 20 个能力子域。

1.2 评估过程与评估难点

当前,电子技术标准化研究院采用的评估方法,无论是企业自评还是专家评估,均使用专家打分法。评估过程涉及 8 张表和 229 条标准,并根据一定的权重进行加权求和,通过得分来判断企业是否符合当前的成熟度等级,然后对企业进行逐级判断从而得到企业最终智能制成熟度等级。本文称上述评估流程为传统纸质评估流程,具体评估流程如图 2 所示。

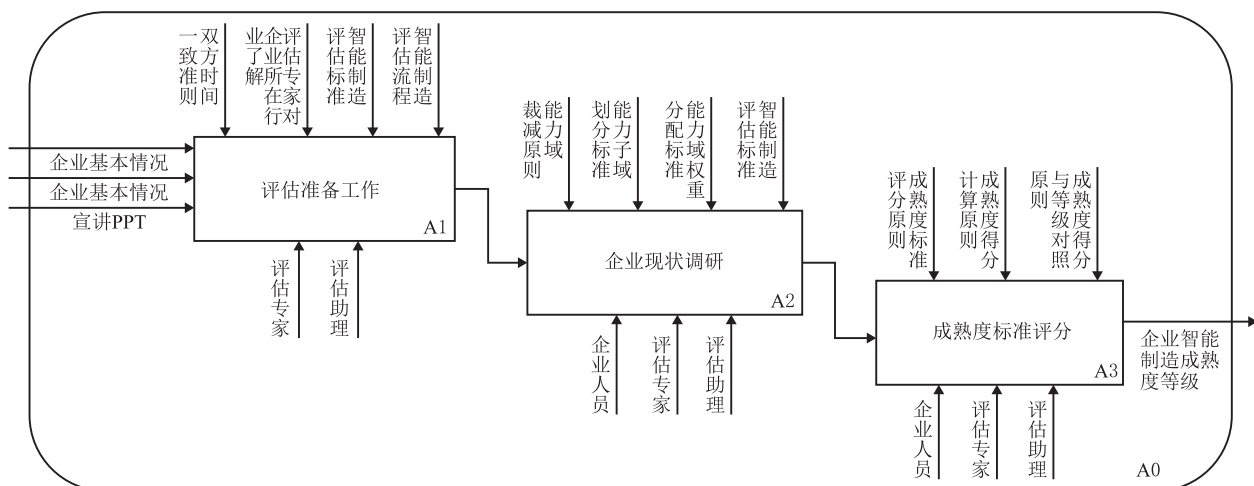


图 2 智能制造成熟度评估流程

以无锡某公司的评估过程为例,整个评估过程耗时 6 天,提前 3 天预约评估专家,在所有评估人员到齐后,还需花费 3 天时间完成企业评估。首先,需花费 1 天时间与企业沟通挑选合适的时间,到达约定的日期时,评估人员进入企业进行评估内容和评估流程介绍。此后,请企业人员针对企业现状进行能力域裁剪,该企业目前不包含设计、物流、限售和服务能力域。

随后,评估工作进入企业现场调研阶段。因为评估标准按照能力子域划分且数量众多,所以对企业现场调研需要按照能力子域分别进行,每部分的调研完成后,评估人员会结合调研情况和企业方的回答为评估标准逐条打出 1、0.8、0.5 或者 0 分,完成所有的打分工作需耗时 1.5 天的时间。

最后,根据重新调整的权重进行加权求和得到该企业的智能制成熟度最终得分为 0.57,根据分数与等级的对应关系,该公司的成熟度等级未达到规划级。

通过上述事例和图 2 可以看出现行评估方法中评估专家是评估资源中的关键,但是对于评估机构,

有限的评估资源无法满足大量的企业的评估需求,若企业未明确自身能力水平即贸然申请成熟度等级判定,会造成达不到要求的企业占用资源,能够达到要求的企业无法及时得到评估资源。

与此同时不难发现传统纸质评估流程存在 3 个挑战导致现行的评估方法不适用于企业进行自评:①企业缺少专业评估人员:企业内部缺少专业评估人员,难以理解评估标准和灵活运用评估方法,致使评估工作难度增加。②评估周期方面:由于需要结合企业调研现状对企业众多能力子域的评估标准进行评估打分,且对每个成熟度等级进行逐级评估,导致评估工作量大、评估周期过长。③评估的主观性方面:评估过程中尽管评估人员和企业方共同参与对评估标准进行评估打分,但评估人员的主观性评价无法避免,影响了评估结果的客观性。

对于制造行业而言,这些问题严重影响了企业智能制成熟度评估的工作进展,本文将对评估过程进行改进,引入智能评估算法帮助企业进行自评。这种方法既能够保证评估结果公正准确同时又可以企业快速了解自身能力水平。

1.3 相关研究现状

智能制造能力成熟水平的评估可以分为两个层次。首先,需要构建一个合适的成熟度评价指标体系,目的是对评估对象的能力要素进行综合评估。其次,选用科学合理的成熟度指标量化评估方法,根据已建立的成熟度评价指标体系,实现对企业智能制造能力水平的评估。这两个层次相互关联,通过评价指标体系和评估方法的结合,能够全面准确地评估企业的智能制造能力成熟水平。

GB/T 39116—2020 智能制造能力成熟度模型是一项通过文本描述即自然语言来评估企业的智能制造成熟度的方法。本研究采用自然语言处理领域中的文本分类技术,以改进成熟度评估过程。

预训练语言模型(Bidirectional Encoder Representations from Transformers, BERT)^[3]模型是由谷歌提出的一种无监督预训练语言模型,专注于自然语言处理领域。该模型通过采用“预训练+微调”的方法,在多个自然语言处理任务中取得了显著的成果。

在制造业领域,徐美娇等^[4]为了提高复杂产品装配制造成熟度等级评估的效率以及客观性,利用成熟度等级评价实例数据,提出基于反向传播神经网络(Back Propagation neural networks, BP)和适应增强算法(Adaptive Boosting, Adaboost)的制造成熟度等级评估方法,并通过实验结果表明基于BP-Adaboost 算法的复杂产品装配制造成熟度等级评估方法具有较好的可靠性与准确度。阴艳超等^[5]针对流程生产由于变量间耦合复杂、时序特征显著而导致工艺质量精准预测困难问题,提出一种融合卷积神经网络与双向门控循环单元(Convolutional Neural Networks - Bidirectional Recurrent Neural Network, CNN-BiGRU)的多工序工艺质量预测方法,并通过实验验证其有效提高了预测精度,为流程生产多工序质量精准预测提供实现方法和途径。为解决人工评估复杂工艺表格的相似性用于工艺重用设计存在效率低、精度差等问题,花豹等^[6]提出一种图神经网络组合算法,来有效提取工艺表格的结构、文字等特征以度量语义相似性,经实验分析表明了所提方法的有效性,并以工艺重用实例验证了方法的可行性。何彦等^[7]针对汽车组合仪表组装过程质检时间长、效率低的现状,提出卷积神经网络与支持向量回归相结合的汽车组合仪表组装质量预测方法,并使用车间生产数据进行实验,证明了所提方法

能够准确有效地预测汽车组合仪表的组装质量。

在其他领域,CHEN 等^[8]使用改进 BERT 实现对长期临床文本的表示,降低时间复杂度和内存消耗,使用 CNN 对肺癌实现更高水平的可解释性预测。LI 等^[9]使用 BERT 进行了医学过程实体归一化任务的实验和分析,WEN 等^[10]讨论了在酒店在线评论中使用 BERT 进行客户情感分析。这都证明了 BERT 在服务行业提高客户理解、推荐和服务智能性方面的有效性。

在深度学习算法的短文本分类研究中,淦亚婷等^[11]全面分析了 CNN、循环神经网络(Recurrent Neural Network, RNN)、CNN-RNN、图卷积神经网络(Graph Convolutional Network, GCN)等不同深度学习方法,并对它们的优缺点进行了比较。这有助于研究者更好地选择适合自己研究的算法。BAO 等^[12]提出一种短文本分类模型,LIU 等^[13]提出了基于 BERT-CNN 的多标签文本分类方法,JOLOUDARI 等^[14]提出了将预训练语言模型与深度卷积神经网络(BERT-Deep Convolutional Neural Networks, BERT-DeepCNN)相结合用于新冠推文情感分析。这些研究表明,在不同领域中,深度学习方法都具有广泛的应用前景。

综上所述,各领域的学者积极应用自然语言处理技术,尤其是基于 BERT 的深度学习方法,为解决各自领域的问题提供了新的思路和方法。在智能制造成熟度评估领域,采用智能评估模型能够解决传统评估过程中的一系列问题,包括评估人员预约、评估周期长和主观性等问题。与此同时,TextCNN 模型在短文本分类方面具有独特的优势,包括并行计算和局部特征提取等特点。而 BERT+TextCNN 模型则充分结合了 BERT 的语义表示和 TextCNN 的特征提取,有效提升了模型性能。最后,智能评估模型的建立不仅有助于更好地利用有限的评估标准数据,还提高了成熟度评估的客观性和可靠性。

2 智能模型驱动的评价过程设计

2.1 成熟度评估过程重构

为克服上述评估过程中所面临的难点与挑战,本文将智能制造成熟度评估过程进行重构。通过深入研究现有评估流程,并根据企业实践过程中的经验和教训,开展有针对性的改进,旨在优化、升级和提升目前的成熟度评估过程。

在企业的自评估过程中,因为缺少类似于评估

专家的专业评估人员,所以评估过程不能过分地依赖评估专家。为了解决这一问题,本文计划采用泛化且全面覆盖的智能评估模型。这一举措可以有效减少对相关经验的评估人员的依赖,帮助企业顺利完成企业自评估,快速了解当前的能力水平。

评估周期过长主要原因是逐级评估和评估打分的方法。经过重构的评估过程放弃传统的专家打分法,采用智能评估模型,通过对企业现状的描述来确定成熟度等级。这一改进可以显著缩短评估周期,以完成对更多企业的成熟度评估。

此外,传统的评估打分法容易受到评估人员主观因素的影响。一些学者采用集体决策、决策试验和评价实验法(Decision making trial and evaluation laboratory, DEMATEL)^[15]等理论降低影响,但模型构建本身复杂,不适用于本场景。本文则通过采用智能评估的方法由智能评估模型确定最终的成熟度等级,可以有效地减轻主观因素对评估结果的影响问题,确保评估结果更加客观和可靠。重构前后的具体评估流程对比如图3所示。

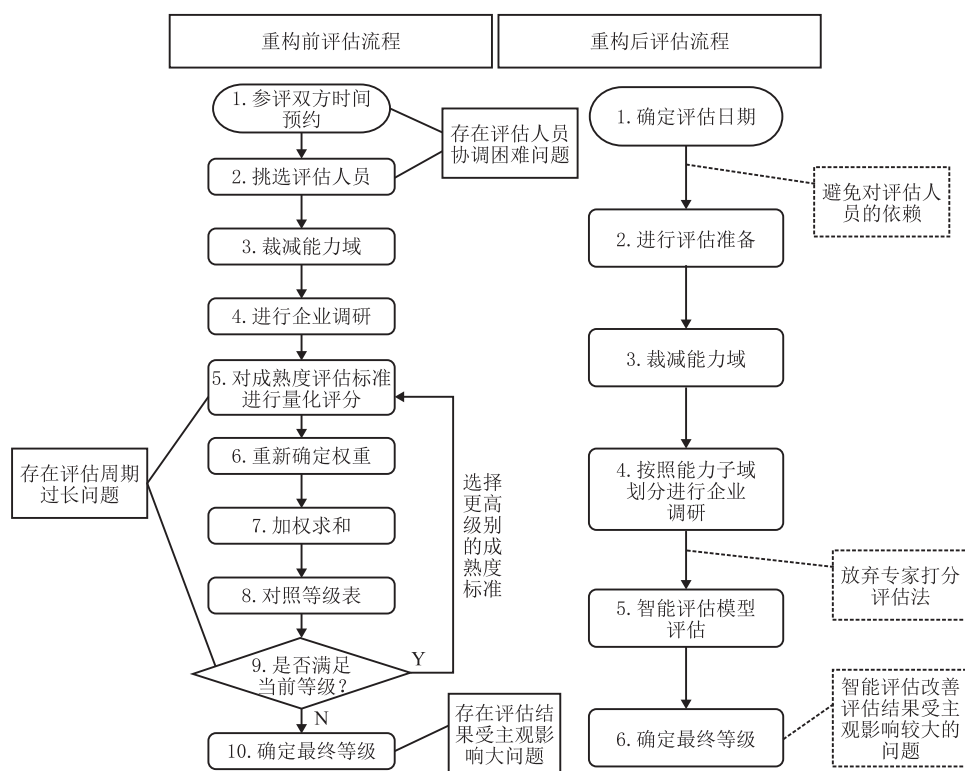


图3 重构前后的具体评估流程对比图

在现行评估流程中,评估人员的选择需要满足两个关键点:①必须具备被评估企业特点和所属行业的相关经验;②时间安排必须与企业确定的进场时间相匹配。步骤5~步骤9中逐条对评估标准进行打分,按照权重进行加权求和,并逐级进行评估,过程持续时间太长;步骤4中采用的专家打分法,虽做到了权威但受主观影响较大。

重构后,因为建立了泛化的成熟度评估模型,一方面不过分依赖于评估专家来进行工作;另一方面企业在申请成熟度等级认证前可以充分地了解自身能力水平。智能评估模型不采用打分加权求和的方式确定成熟度等级。通过对企业现状的描述便可得

到最终的成熟度等级,极大地缩短了评估周期。同时采用智能评估模型对企业进行智能评估,可以改善评估结果受主观影响大的问题。

2.2 文本数据集构建

本文使用 GB/T 39116—2020 智能制造能力成熟度模型文件作为训练数据集的来源。该文件包含了达到不同成熟度级别所需满足的标准,在进行训练数据集的预处理过程中,对评估标准中的每个句子进行加工和转换,将其转换为陈述句的形式,同时去除了空格、无效符号和无效字段,得到了经过处理的文本数据集。训练数据集中按照成熟度等级分为不足规划级、规划级、规范级、集成级、优化级、引领

级 6 个类别,分别使用 0~5 表示,训练数据集示例如表 1 所示。验证数据集和测试数据集是以 106 家企业的智能制成熟度评定结果为基础,将每家企业的评定情况按照能力子域进行划分并相互打乱顺

序,然后将细化和打乱后的企业评定情况以 5:5 的比例进行划分。模型建立过程中,使用训练数据集建立评估模型,验证数据集调试评估模型,测试数据集验证评估模型效果,数据集构成如表 2 所示。

表 1 训练数据集示例

能力子域	内容	对应等级
组织战略	企业没有对发展智能制造的资源进行投资	0
组织战略	企业具有明确的智能制造发展规划	1
人员技能	企业目前具有智能制造统筹规划能力的个人与团队	2
数据	企业目前实现了数据及分析结果的跨部门在线共享	3
装备	关键工序设备机油预测维护功能	4
生产作业	生产现场能够应用人工智能、大数据等技术,实现生产过程非预见性异常的自动调整	5

表 2 数据集构成

数据来源	数据集类型	数量/条
国标文件	训练数据集	3 332
106 家企业数据	验证数据集	1 897
106 家企业数据	测试数据集	1 751

训练、验证和测试数据集中的每条文本长度均控制在 256 字以内。训练数据集中包括整理后的每条成熟度评定标准以及相应的成熟度等级。

而验证和测试数据集则涵盖了真实企业智能制成熟度评估现状及对应的成熟度等级。在整理这些数据集的过程中,得到了本领域专家和学者的积极协助。这些专家和学者拥有多年从事智能制造相关工作的经验,他们为整个过程提供了宝贵的意见和见解。他们的参与和见证是本研究工作不可或缺的重要组成部分。具体参与人员的详细信息可参见表 3。

表 3 数据集的整理人员

姓名	职称	从业年份	姓名	职称	从业年份
徐 X	高级咨询师	24	汪 X 芬	教授	33
李 X 萍	高级咨询师	25	徐 X 善	副教授	32
崔 X	高级咨询师	26	柳 X 燕	副教授	10
秦 X	咨询师	8	吴 X 顺	高级工程师	11

2.3 智能制成熟度智能评估算法设计

2.3.1 BERT 预训练模型

BERT 是基于 Transformer 架构的预训练语言模型,并将多头注意力机制应用于其编码器部分,多头注意力机制的引入赋予了 BERT 强大的语义理解能力,使其能够捕捉不同层次和不同角度的语义信息。每个注意力头可以关注不同的语义特征,从而提供了更全面的表示能力。BERT 通过预训练过程中的掩码任务(Masked Language Model, MLM)和下一句预测任务(Next Sentence Prediction, NSP)来学习词汇和句子之间的关系。这种结合了 MLM 和 NSP 任务的 BERT 模型在处理词义歧义问题时表现卓越,能够根据上下文准确理解词汇的含义,并将输入的文本转化为带有特征信息的字向量矩阵。具体的模型示意图如图 4 所示。

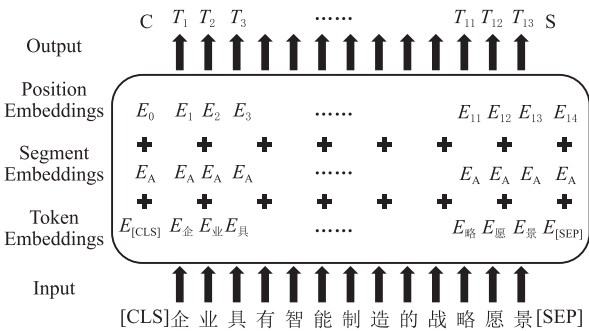


图 4 BERT 预训练模型

2.3.2 TextCNN 文本神经网络

TextCNN 作为卷积神经网络的一种,通过多种尺度的卷积核对文本进行滑动窗口取样,从而获取不同大小的局部特征,并能捕捉到文本中不同层次的语义信息。每个卷积核都拥有单独的滑动窗口尺

寸和数量,使得模型能够在不同尺度上提取特征。此外,TextCNN中的卷积核可以在整个输入上共享,从而降低了参数的数量,简化了模型复杂度,提高了模型的泛化性能。模型中的卷积层可以并行处理不同位置的输入,极大地加快了计算速度。正因如此,TextCNN模型既展现出较高的计算效率,又能够有效捕捉文本数据中不同尺度和层次的语义特征。使其非常适合处理大规模的文本数据和长文本数据。具体模型结构图如图5所示。

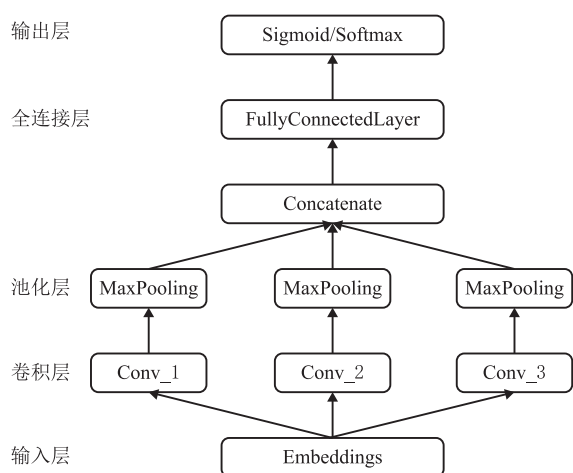


图5 TextCNN模型

2.3.3 基于预训练模型和文本神经网络的成熟度评估模型设计

因此,本研究采用基于中文预训练模型和文本神经网络的BERT+TextCNN成熟度评估模型代替原有的评估方法帮助企业进行智能制造成熟度评估,该模型总体分为3个阶段:第一阶段,通过Tokenizer将输入的中文文本转化中文预训练模型bert-base-chinese模型可接受的输入格式。Tokenizer负责将句子分割成单词或者子词,并为每个单词或子词分配唯一的编号。第二阶段,将Tokenizer的输出作为bert-base-chinese模型的输入,经过bert-base-chinese模型中多层Transformer Encoder的编码。每一层都包含多头自注意力机制和前馈神经网络,使得模型逐渐提取抽象的特征表示,并逐步实现文本的深层次理解,最终将输入模型的单词序列转换成等长的词向量矩阵,其中包括了单词的语义信息和位置信息。第三阶段,采用TextCNN模型对BERT模型的输出进行分类,得出具体成熟度评估结果。TextCNN模型作为一种文本卷积神经网络,能够有效地学习文本数据的特征表达,从而对不同的文本进行分类。在此,TextCNN模型的主要任务是提取和分类BERT模型输出的文本特征,最终生成相应的成熟度评估结果。具体的模型结构如图6所示。

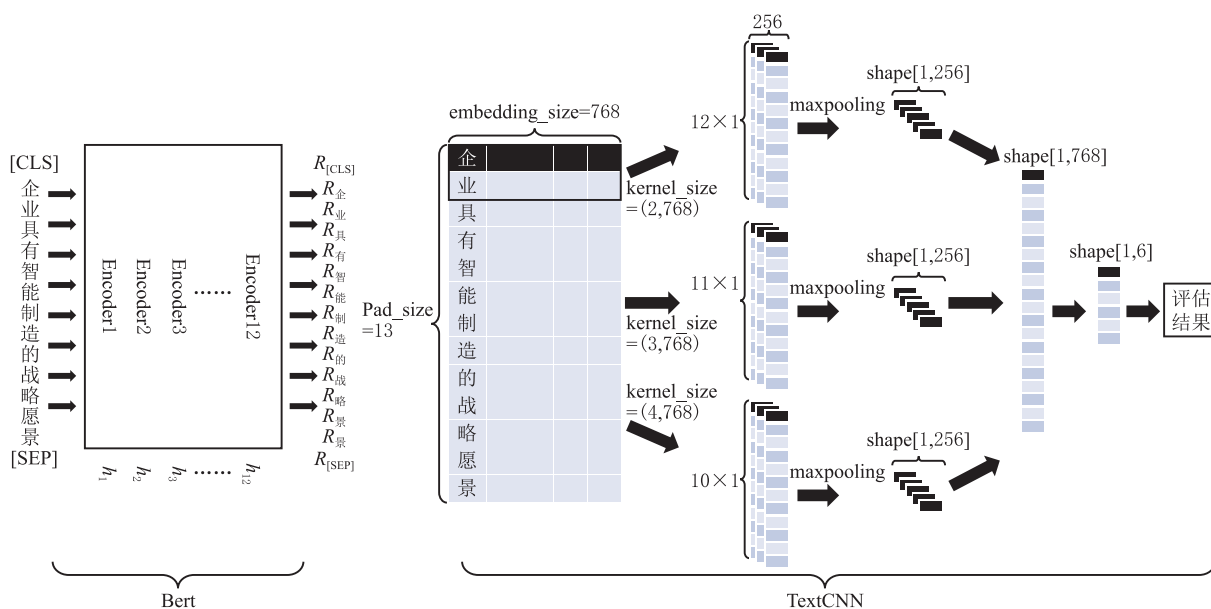


图6 BERT+TextCNN成熟度评估模型

对于整体的智能评估流程,输入模型的文本描述会经过由词嵌入层、卷积层、池化层和全连接层最终得到成熟度判定等级。当评估人员使

用时可直接以成熟度能力子域为单位以企业现状的文本描述作为评估模型的输入,经过Tokenizer的分词处理转换为input_ids和attention_

mask 输入进 BERT 模型。然后, BERT 模型对输入进行嵌入得到 bert_embedding。接下来, bert_embedding 作为 TextCNN 的输入, 经过一系列卷积、激活函数、池化和全连接层操作后得到输出层的 cnn_output, 最终由 cnn_output 转变为成熟度等级。基于 BERT+TextCNN 成熟度评估算法如下:

算法 1 BERT+TextCNN 成熟度评估算法。

```
Input: input_text # 评估文本
Output: Lable # 等级
1. input_text = input_tensor() # 输入文本
2. tokenized_input = tokenizer.tokenize(input_text) # 使用 Tokenizer 对进行分词
3. input_ids, attention_mask = tokenizer.convert_tokens_to_ids(tokenized_input) # 将分词结果转换为 BERT 模型可接受的输入格式
4. bert_embedding = bert_embedding_layer(input_ids, attention_mask) # 使用 BERT 模型进行嵌入
5. cnn_input = bert_embedding # 将 BERT 的嵌入结果作为 TextCNN 的输入
6. for filter_size in filter_sizes: # 遍历不同尺寸的卷积核
7. conv_output = convolution(cnn_input, filter_size) # 卷积操作
8. activation_output = activation(conv_output) # 激活函数
9. maxpool_output = max_pooling(activation_output) # 最大池化
10. cnn_outputs.append(maxpool_output) # 保存每个卷积核的输出
11. concat_output = concatenate(cnn_outputs) # 将所有卷积核的输出连接起来
12. dropout_output = dropout(concat_output) # Dropout 操作
13. fully_connected_output = fully_connected_layers(dropout_output) # 全连接层
14. cnn_output = output_projection(fully_connected_output) # 输出层
15. cnn_output → Lable # 转化为等级
16. end
```

3 实验与分析

3.1 实验环境配置与评价指标选取

本实验所采用的计算机处理器为 Intel(R) Xeon(R) Platinum 8375C CPU, 显卡为 NVIDIA GTX4090(24 GB), 基于 Python 3.8 进行实验, 主要使用的深度学习框架是 PyTorch 1.10.0 版本, 运行内存为 80 GB。本次实验选择的是中文版本的 BERT(bert-base-Chinese), 将经过预处理的文本数据输入到经过参数微调的 BERT 模型中, 以提升 BERT 在下游任务中的效果。

在训练、测试和验证模型的过程中, 模型会记录损失函数的值。若在一段时间内损失值没有发生变化, 则可以提前结束训练过程。为了比较不同模型之间的性能, 本文选择精确率 *Precision*、召回率 *Recall* 以及调和平均值 *F1Score* 作为评估指标。这些指标的计算公式如下所示:

(1) 精确率 (*Precision*): 即正确预测为正例的样本数占有所有预测为正例的样本数的比例:

$$Precision = \frac{TP}{TP + FP}。 \quad (1)$$

(2) 召回率 (*Recall*): 即正确预测为正例的样本数占有所有正例样本数的比例:

$$Recall = \frac{TP}{TP + FN}。 \quad (2)$$

(3) 调和平均值 (*F1Score*): 是精确率和召回率的调和平均数:

$$F1Score = \frac{2 \times R \times P}{R + P}。 \quad (3)$$

真正类 (True Positive, TP) 指的是实际为正例且被模型正确预测为正例的样本数; 假正类 (False Positive, FP) 指的是实际为负例但被模型错误地预测为正例的样本数; 假负类 (False Negative, FN) 则表示实际为正例但被模型错误地预测为负例的样本数。通过计算这些评估指标, 可以更好地比较不同模型之间的性能, 并找到最优的模型。同时, 这些指标也有助于了解模型的性能特点, 并做出相应的改进。

3.2 评估模型参数分析

在参数设置过程中, 本文尝试了不同的卷积核大小、迭代次数和学习率等参数, 以评估其对结果的影响, 通过系统地调整不同的参数组合, 可以进一步优化评估模型的性能。

首先, 卷积核的大小对卷积神经网络在特征提取方面的效果具有影响。一般而言, 较小的卷积核能够捕捉更细微的特征细节, 但也容易导致过拟合问题。相反, 较大的卷积核能够捕捉到更高层次的特征, 但随着卷积核宽度的增加, 计算成本也随之增加。因此, 需要通过比较不同大小的卷积核来确定最合适的尺寸。根据表 4 的结果可以看出, 当卷积核的大小设置为 [2, 3, 4] 时, 模型评估效果最佳; 而

随着卷积核的增大,评估效果却逐渐下降。基于此,本文将卷积核的大小设定为[2,3,4]。

表 4 卷积核大小对评估结果的影响

卷积核 大小	精确率 <i>Precision</i> / %	召回率 <i>Recall</i> / %	调和平均值 <i>F₁</i> / %
[2,3,4]	80.50	80.13	80.09
[3,4,5]	74.60	73.29	73.17
[4,5,6]	71.54	71.10	71.10
[5,6,7]	69.89	69.07	68.96
[2,3,4,5]	79.57	79.04	79.11
[3,4,5,6]	73.55	73.11	73.08

其次,迭代次数也是影响模型性能的一个重要参数。通常情况下,增加迭代次数可以帮助模型更好地学习数据中的特征,从而提高模型的性能。然而,过多的迭代次数可能会导致过拟合问题,并且增加了训练时间的消耗。因此,需要通过比较不同的迭代次数来确定最佳取值。根据表 5 的结果可以看出,随着迭代次数的增加,评估模型的效果也逐渐提升。在迭代次数达到 6 次时,评估模型的效果达到最佳,而超过 6 次后,评估效果开始下降。基于此,本文中迭代次数设定为 6 次。

表 5 迭代次数对评估结果的影响

迭代次数	精确率 <i>Precision</i> / %	召回率 <i>Recall</i> / %	调和平均值 <i>F₁</i> / %
3	77.94	75.07	76.04
4	80.5	80.13	80.09
5	83.30	82.75	82.70
6	84.56	83.44	83.52
7	83.82	82.98	82.94
8	83.15	82.18	82.14

最后,学习率作为控制模型在训练过程中权重更新速度的重要参数,决定了每次参数更新的步长。较小的学习率有助于维持模型的稳定性,而较大的学习率则能够加速模型的收敛速度。因此,需要通过比较不同学习率的效果,以找到最佳的学习率值。如表 6 所示,当学习率设置为 $1e-4$ 时,评估模型未能收敛;然而,当将学习率调整为 $3e-5$ 时,训练时间未显著增加,同时评估模型达到最佳效果。基于以上分析,本文中迭代次数设定为 $3e-5$ 。

表 6 学习率对评估结果的影响

学习率大小	精确率 <i>Precision</i> / %	召回率 <i>Recall</i> / %	调和平均值 <i>F₁</i> / %
$1e-4$	0	0	0
$5e-5$	81.99	81.21	81.17
$3e-5$	85.32	84.05	86.67
$1e-5$	84.56	83.44	83.52
$5e-6$	79.46	79.10	79.13
$3e-6$	80.19	80.01	80.02

在参数设置过程中,关键是进行系统化的实验和评估。通过尝试不同的参数组合,并结合验证数据集的表现进行评估,以获取最佳性能的参数组合。通过采用这种方法,进一步优化 BERT-TextCNN 评估模型的性能。最终,当卷积核大小设置为[2,3,4]、迭代次数为 6 次、学习率为 $3e-5$ 时,BERT-TextCNN 评估模型展现出最优的性能。

3.3 评估模型结果验证

为了研究不同模型在评估过程中的性能表现,将 BERT 的输出作为嵌入输入到 TextCNN、RNN、RCNN 和 DPCNN 模型中。然后,对 BERT、BERT + TextCNN、BERT + RNN、BERT + RCNN 和 BERT + DPCNN 进行了效果对比分析。通过参数调整和多轮试验,获取了各评估模型评价指标的最优值,并对它们进行了对比。具体对比结果详见表 7。此外,还得到了各评估模型的混淆矩阵,并对各混淆矩阵也进行了对比,更直观地观察模型的性能差异。具体信息见图 7,通过对比分析,从中选择出最优的评估模型。

表 7 各评估模型实验对比表

评估模型	参数设置 (卷积核+ 迭代次数+ 学习率)	精确率 <i>Precision</i> / %	召回率 <i>Recall</i> / %	调和平均值/ <i>F₁</i> / %
BERT	$5+2e-5$	84.18	83.04	83.13
BERT+ TextCNN	[2,3,4]+ $6+3e-5$	85.32	84.05	86.67
BERT+DPCNN	$6+4e-5$	70.99	72.89	68.89
BERT+RNN	$5+2e-5$	78.36	74.42	75.53
BERT+RCNN	$5+2e-5$	84.98	83.33	85.43

根据以上实验对比表,可以明显看出,当仅利用 BERT 预训练模型进行评估时,3 项评价指标基本维持在 83%~84%之间,取得了不错的效果。这表明 BERT 预训练模型在成熟度评估任务中具有出色的性能,为后续模型提供了强大的基础。进一步,在其后面接适合的变体神经网络后,效果会再次得到提升。在对比中可以看出循环神经网络的评估模型在评价指标方面低于其他模型,这表明循环神经网络的短时记忆优势在此未能充分发挥,导致其评估结果较为分散且评估效果不佳。BERT+TextCNN 评估模型评估等级时达到了 85.32%的准确率,显著高于其他评估模型。通过混淆矩阵的对比可以看出, BERT+TextNN 模型的错误评估结果通常维持在与正确等级相邻的两个等级之间,不存在所有等级都可能被预测到的现象,证明了该评估模型在智能制造评估方面的可行性。此外,就混淆矩阵而言, BERT+TextCNN 评估模型在各个等级的评估中取得了较为均衡和良好的性能表现,这证明了该评估模型在智能制造评估方面的稳定性。因此,不论是从评价指标还是从总体效果来看, BERT+TextCNN 评估模型都表现出卓越的性能。

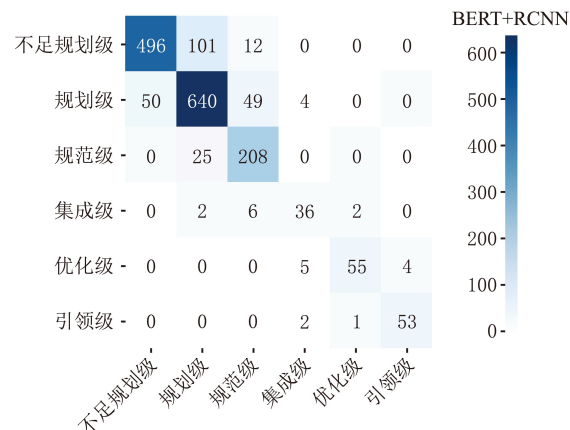
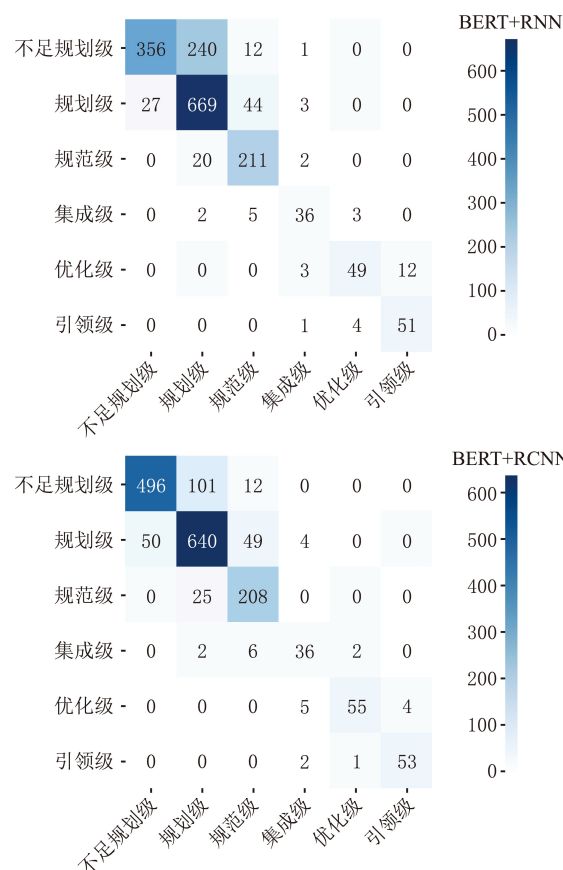
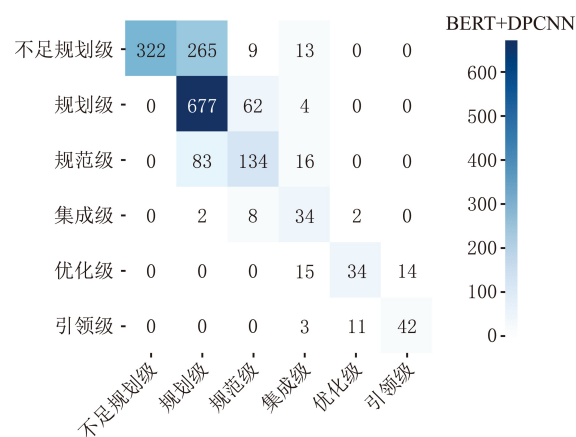
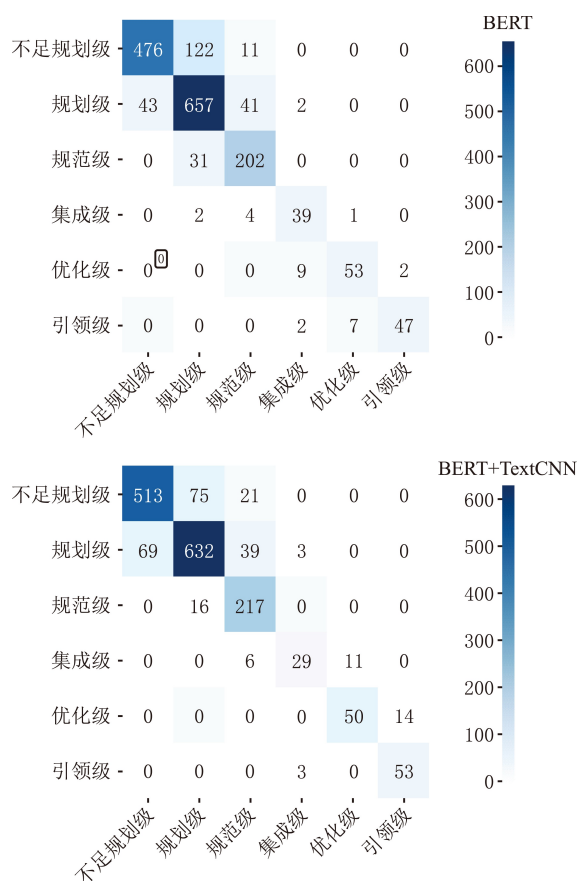


图7 评估模型混淆矩阵对比图

3.4 讨论

本文为企业提供了一种新的智能制成熟度自评估方法,以解决现行评估方法不适用于企业自评估的 3 个挑战:①企业缺少专业评估人员;②评估周期过长;③评估结果受主观影响大。采用智能评估方法后,企业的智能制成熟度自评估工作可以不过度依赖评估专家。以一家企业为例,无论从评估人员还是评估周期来讲,相较于以往需要 2 名专家 3 名助理投入 $n+3$ 天才可完成的任务, n 为预约评估专家时间一般约为 2~7 天。现在经过培训后,可以做到 3 名

企业人员的参与下,评估周期缩短至 1 天或 1.5 天即可完成,同企业同等级评估情况下完成评估所需时间与人员对比如图 8 所示。在工作量方面,重构前成熟度等级需要企业人员配合评估专家对每一等级进行逐级评估,重构后采用智能评估的方法可以直接得到企业的成熟度等级,从而大幅减少了企业人员的工作

量。重构前的评估结果由专家打分得到,过程中极易受到评估人员的主观影响,重构后由智能评估模型得出评估结果,经过对同一家企业的多次实验,证明其结果准确可靠,这就避免了评估结果受评估人员主观影响的问题。

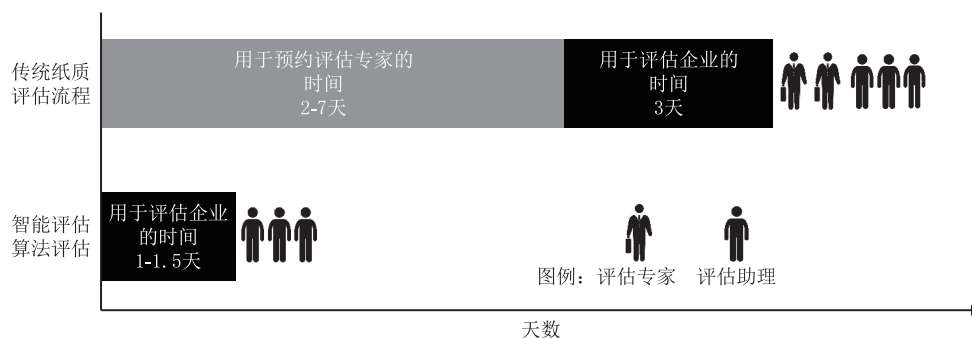


图 8 同企业同等级评估情况下完成评估所需时间与人员对比

通过对建立的 BERT+TextCNN 评估模型的验证,目前其准确率维持在 85% 以上。然而,该方法的运用仍然面临一些挑战。一方面,当前建立的评估模型仍处于初步阶段,需要不断优化改进;另一方面,参与评估的企业人员经验参差不齐,同时训练数据集中涉及的企业领域较多,特定行业的企业数据集样本不足,导致评估模型的泛化能力尚待提升。因此,还需要从多个角度持续改进和完善智能评估模型,以提高其准确率和适用性。

4 结束语

为支持国内众多企业快速完成智能造成熟度评估工作,解决评估资源有限以及评估过程中低效率、长周期和评估结果受主观因素影响较大等问题。本文以智能制造能力成熟度模型文件中不同成熟度等级需要满足的标准作为训练数据集,通过对成熟度等级评定标准的学习,发现成熟度等级之间存在明显的差异。基于这一发现,本文提出一种基于预训练模型 BERT 和文本神经网络 TextCNN 的成熟度评估模型,在与各主流深度学习模型比较中,各项评估指标显示 BERT+TextCNN 评估模型明显优于其他主流深度学习模型。此外,通过对各评估模型的混淆矩阵进行分析,结果显示深度学习模型在智能造成熟度评估方面具备可行性。在众多评估模型中,BERT+TextCNN 评估模型在成熟度标准的特征提取和语义挖掘方面表现出卓越能力。本文提供的方法为企业进行智能造成熟度自评估提供了一种有效的途径,并证实了深度学习

模型在该领域的潜力和竞争优势。这对于国内企业快速、准确地完成智能造成熟度评估工作具有重要意义。

尽管本研究取得了一定的成果,但仍需要进一步改进。首先,需要解决数据集的限制性问题。随着更多企业进行智能造成熟度评估或更新自身的成熟度水平,将有更详细的评估标准可用于进一步完善评估模型。其次,随着数据集的增加,可以对数据集进行更精细的行业分析,以创建更精确的行业智能评估模型,从而提高评估的准确性,以满足企业和用户更多的需求。这些改进将有助于进一步提高评估模型的效能和可用性。

参考文献:

- [1] YAO Xifan, JING Xuan, ZHANG Jianming et al. Toward intelligent manufacturing in the new industrial revolution[J]. Computer Integrated Manufacturing Systems, 2020, 26(9): 2299-2320(in Chinese). [姚锡凡,景轩,张剑铭,等. 走向新工业革命的智能制造[J]. 计算机集成制造系统, 2020, 26(9): 2299-2320.]
- [2] MIIT. GB/T39116-2020, Intelligent manufacturing capability maturity model[S]. Beijing: Standards Press of China(in Chinese). [中华人民共和国工业和信息化部. GB/T39116-2020, 智能制造能力成熟度模型[S]. 北京: 中国标准出版社]
- [3] DEVLIN J, CHANG M W, LEE K et al. BERT: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. (2018-10-11) [2023-05-26]. //Proceedings of NAACL-HLT. 2019: 4171-4186.
- [4] XU Meijiao, XUE Shanliang, ZHANG Hui, et al. Manufacturing

- Readiness Level Assessment Method of Complex Product Assembly Based on BP-AdaBoost Algorithm[J]. China Mechanical Engineering, 2023, 34(20): 2513-2519 (in Chinese). [徐美姣, 薛善良, 张 惠, 等. 基于 BP-Adaboost 算法的复杂产品装配制成熟度等级评估方法[J]. 中国机械工程, 2023, 34(20): 2513-2519.]
- [5] YIN Yanchao, HONG Zhimin, GU Wenjuan, et al. Multi-process quality prediction method integrating multi-channel CNN-BiGRU with temporal pattern attention mechanism[J/OL]. Computer Integrated Manufacturing Systems; 1-19 [2023-09-25]. <http://kns.cnki.net/kcms/detail/11.5946.TP.20230802.0948.002.html> (in Chinese). [阴艳超, 洪志敏, 顾文娟, 等. 融合多通道 CNN-BiGRU 与时间模式注意力机制的多工序工艺质量预测方法[J/OL]. 计算机集成制造系统; 1-19 [2023-09-25]. <http://kns.cnki.net/kcms/detail/11.5946.TP.20230802.0948.002.html>.]
- [6] HUA Bao, ZHOU Bin, GU Xinghai et al. Semantic similarity metrics for process forms based on graph neural networks[J]. Computer Integrated Manufacturing Systems, 2022, 28(12): 3805-3821. (in Chinese). [花 豹, 周 彬, 顾星海, 等. 基于图神经网络的工艺表格语义相似性度量[J]. 计算机集成制造系统, 2022, 28(12): 3805-3821.]
- [7] HE Yan, XIAO Zhen, LI Yufeng, et al. Quality prediction method for automobile combination instrument assembly using CNN-SVR[J]. China Mechanical Engineering, 2022, 33(7): 825-833 (in Chinese). [何 彦, 肖 圳, 李育锋, 等. 使用 CNN-SVR 的汽车组合仪表组装质量预测方法[J]. 中国机械工程, 2022, 33(7): 825-833.]
- [8] CHEN Z Y, XIAO F, WANG X K, et al. An interpretable diagnostic approach for lung cancer: Combining maximal clique and improved BERT [J] Expert Systems, 2023, 40(8): e13310.
- [9] LI L Q, ZHAI Y K, GAO J H, et al. Stacking-BERT model for Chinese medical procedure entity normalization[J]. Mathematical Biosciences and Engineering, 2023, 20(1): 1018-1036.
- [10] WEN Y, LIANG Y Z, ZHU X H. Sentiment analysis of hotel online reviews using the BERT model and ERNIE model—Data from China[J]. Plos one, 2023, 18(3): e0275382.
- [11] GAN Yating, AN Jianye, XU Xue. A research review of short text categorization methods based on deep learning[J]. Computer Engineering and Applications, 2023, 59(4): 43-53 (in Chinese). [淦亚婷, 安建业, 徐 雪. 基于深度学习的短文本分类方法研究综述[J]. 计算机工程与应用, 2023, 59(4): 43-53.]
- [12] BAO T, REN N, LUO R, et al. A BERT-based hybrid short text classification model incorporating CNN and attention-based BiGRU[J]. Journal of Organizational and End User Computing, 2021, 33(6): 1-21. DOI: 10.4018/JOEUC.294580.
- [13] LIU W F, PANG J M, LI N, et al. Research on multi-label text classification method based on tALBERT-CNN[J]. International Journal of Computational Intelligence Systems, 2021, 14(1): 201. DOI: 10.1007/s44196-021-00055-4.
- [14] JOLOUDARI J H, HUSSAIN S, NEMATOLLAHI M A, et al. BERT-deep CNN: State of the art for sentiment analysis of COVID-19 tweets[J]. Social Network Analysis and Mining, 2023, 13(1): 99. DOI: 10.1007/s13278-023-01102-y.
- [15] SI S L, YOU X Y, LIU H C, et al. DEMATEL technique: A systematic review of the state-of-the-art literature on methodologies and applications[J]. Mathematical Problems in Engineering, 2018, 2018: 3696457.

作者简介:

张 淦(1997—),男,河南南阳人,硕士研究生,研究方向:智能制造、数据分析挖掘、数字化设计制造等,E-mail: zhanggan@njust.edu.cn;
袁堂晓(1993—),男,江苏连云港人,博士研究生,研究方向:智能制造理论、数据采集、算法应用风险评估等,E-mail: yuantx@njust.edu.cn;

+汪惠芬(1965—),女,江苏苏州人,教授,博士,研究方向:智能制造、数字化设计制造、数字孪生、知识库与知识管理等,通讯作者,E-mail: wanghf@njust.edu.cn;

柳林燕(1985—),女,江苏苏州人,副教授,博士,研究方向:数字化设计制造、智能制造、数字孪生、知识库与知识管理等,E-mail: liulinyan@njust.edu.cn。