



情报杂志
Journal of Intelligence
ISSN 1002-1965, CN 61-1167/G3

《情报杂志》网络首发论文

题目：融合异构图全局结构信息和时间序列的专利价值评估方法
作者：陈晰，程戈，尹智斌
网络首发日期：2025-03-07
引用格式：陈晰，程戈，尹智斌. 融合异构图全局结构信息和时间序列的专利价值评估方法[J/OL]. 情报杂志. <https://link.cnki.net/urlid/61.1167.G3.20250307.0918.002>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

融合异构图全局结构信息和时间序列的 专利价值评估方法*

陈 晰¹ 程 戈^{2,3} 尹智斌³

(1. 湖南工业大学 法学院 株洲 412007; 2. 湖南警察学院 长沙 410138;
3. 湘潭大学计算机学院 湘潭 411105)

摘要: [研究目的] 针对现有专利价值评估方法未能充分利用异构图全局结构信息与专利被引用量时间序列特征的问题, 提出一种融合异构图全局结构信息和时间序列的专利价值评估方法。该方法旨在提高专利价值分类的准确性, 特别是对于高价值专利的识别。 [研究方法] 首先, 使用异构图神经网络模型提取专利背景信息, 加入了节点中心性编码以捕捉图的全局结构信息。然后, 融合专利的被引用量时间序列, 并通过注意力机制预测专利被引用量的变化趋势。最后, 利用这些特征进行专利价值等级分类。 [研究结果/结论] 实验结果表明, HNTSM 模型在美国半导体领域专利数据集上的专利价值分类任务中, 对于高价值专利 (A 等级), 模型的精确率高达 77.37%, F1 值为 76.72%。相比现有方法有显著提升, 特别是通过引入全局结构信息与时间序列模块, 对专利价值评估结果具有积极作用。

关键词: 专利价值评估; 异构图; 全局结构信息; 时间序列; 神经网络模型
中图分类号: G250. 2; G306. 0 **文献标识码:** A

A Patent Value Assessment Method Integrating Global Structural Information of Heterogeneous Graphs and Time Series

Chen Xi¹ Cheng Ge^{2,3} Yin Zhibin³

(1. Law School, Hunan University of Technology, Zhuzhou 412007, China;
2. Hunan Police Academy, Changsha 410138, China;
3. School of Computer Science, Xiangtan University, Xiangtan 411105, China)

Abstract: [Research purpose] In response to the limitations of existing patent value assessment methods, which fail to fully utilize the global structural information of heterogeneous graphs and the time series of patent citation counts, this paper proposes a method that integrates these features. The aim is to improve patent classification accuracy, particularly in identifying high-value patents. [Research method] This paper presents a patent value assessment model, HNTSM (Heterogeneous Networks and Time Series Model), that integrates global structural information from heterogeneous graphs and time series data. First, a heterogeneous graph neural network model is used to extract background information about patents, incorporating centrality encoding to capture the global structure of the graph. Then, the time series of patent citation counts is integrated, and attention mechanisms are applied to capture the trends in citation count changes. Finally, these features are used for patent value classification. [Research result/conclusion] Experimental results show that the HNTSM model achieves a precision of 77.37% and an F1 score of 76.72% in identifying high-value patents (A grade) on a U. S. semiconductor patent dataset. Compared to existing methods, the model demonstrates significant improvements, especially through the introduction of global structure information and time series modules, which have a positive impact on patent value assessment outcomes.

Key words: patent value assessment; heterogeneous graph; global structural information; time series; neural network model

基金项目: 湖南省哲学社会科学项目“人工智能赋能知识产权风险预警研究”(陈晰 编号: 22YBA154)。

作者简介: 陈 晰, 女, 1978 年生, 法学博士, 讲师, 研究方向: 知识产权; 程 戈, 男, 1977 年生, 工学博士, 教授, 博士生导师, 研究方向: 人工智能; 尹智斌, 男, 1998 年生, 工学硕士, 研究方向: 知识表示学习, 专利数据分析。

通信作者: 程 戈

0 引言

专利价值评估是专利分析领域的重要研究内容,在企业专利战略规划和政府知识产权管理中具有核心地位。每年全球专利申请数量持续增长,但其中只有少数专利具有重大社会价值。如何高效识别这些创新性专利,已成为技术管理和创新经济中的关键问题。因此,专利价值评估技术在技术趋势分析、专利组合优化和政策制定等应用场景中具有广泛需求和重要意义。

专利的背景信息(如发明人、专利权人)和被引用量时间序列是评估专利价值的核心因素。背景信息反映了专利的技术来源和持有者能力,而被引用量时间序列揭示了专利在技术领域的传播路径及其重要性动态变化。被高水平发明人持有、引用量呈上升趋势的专利通常更具价值。然而,现有方法在建模这些多维信息时存在局限,难以全面准确地评估专利价值。

近年来,深度学习技术因其在特征自动提取和复杂数据建模中的优势,逐渐成为专利价值评估的主要技术手段。其中,图神经网络^[1](Graph Neural Networks, GNN)以其对结构化数据的强大建模能力,成为专利分析领域的研究热点。GNN 结合异构图^[2]建模技术,可整合专利内容、技术关联性及引证关系等多维度信息,显著提高对专利间复杂关系的捕捉能力。通过嵌入学习,GNN 能够挖掘专利的语义关联和技术传播路径,为价值评估提供深层次支持。

尽管 GNN 取得了显著进展,但现有研究主要关注局部语义聚合,忽视了专利节点在全局图结构中的重要性。此外,专利被引用量的时间序列特征常被忽略,未能充分挖掘其蕴含的专利价值动态变化信息。这些不足限制了现有模型的评估效果。

为解决上述问题,本文提出了一种融合异构图全局结构信息和时间序列特征的专利价值评估模型(Heterogeneous Networks and Time Series Model, HNTSM)。该模型通过在图神经网络中引入节点中心性编码,捕捉专利节点在全局结构中的重要性;同时结合时间序列模块,通过注意力机制提取专利被引用量变化趋势,进一步提升专利价值评估的准确性和鲁棒性。实验结果表明,HNTSM 模型在准确率和 F1 得分方面均显著优于现有方法,充分验证了本文方法的有效性。

1 相关研究

在专利价值分类的研究领域,现有文献可以根据不同的分析方法和技术应用划分为几大类,主要涵盖基于传统方法、机器学习和深度学习的专利价值评估

模型。

传统方法早期在专利价值评估中起着主导作用。研究者通过回归分析、主成分分析、时间序列等传统统计工具,评估专利的技术新颖性、权利要求数量、发明人数量等指标与专利价值的相关性。例如,Hirschey 等^[3]和 Deng 等^[4]的研究揭示了这些指标如何影响专利的技术价值。Boeing 等^[5]利用跨国比较数据来衡量专利质量,为后续研究提供了跨国视角。李娟等^[6]利用层次分析法(AHP)对法律因素进行加权,并通过熵权法调整了权重,展示了法律因素在评估中的重要性。Trajtenberg^[7]的研究采用了专利引文网络,研究专利作为节点、引文作为边的网络结构,通过量化引文网络的指标评估专利的创新性和经济价值。Breschi 等^[8]进一步结合引文网络数据和时间序列分析,观察知识扩散的速度和范围,评估专利质量。Harhoff 等^[9]则采用计量经济学模型,结合专利引文数据及其家族规模、异议情况等特征,通过对引文数量和频率的回归分析,发现高频被引用的专利更有可能在技术市场中产生较高的商业价值。

机器学习技术随着数据处理能力的提升,机器学习逐渐被应用于专利价值评估中。研究者们引入了随机森林、支持向量机等算法,尝试预测专利的潜在价值。王思培等^[10]基于随机森林算法提出了高价值专利预测模型,Ko 等^[11]通过机器学习评估了知识产权的可转移性。Choi 等^[12]应用机器学习模型预测知识产权的商业潜力,评估专利生命周期。此外,Jie 等^[13]结合卷积神经网络和随机森林算法,开发了专利分类模型,进一步推动了机器学习在专利评估领域的应用。

深度学习技术近年来在专利价值评估中的应用显著增加,尤其是在处理专利文本和引文网络时,基于图神经网络的模型展示了更优越的能力。例如,Liu 等^[14]使用卷积张量分解方法预测专利诉讼,并通过深度学习模型优化了专利文本的表示。Pujari 等^[15]^[16]研究了专利景观中神经网络模型的应用,展示了深度学习在大规模专利数据分类和评估中的潜力。Choi 等^[17]结合图嵌入和 Transformer 技术,开发了深度专利景观模型,用于评估专利的商业潜力。

此外,Fang^[18]从多视图专利网络分析的角度开发了一种新的专利分类任务方法,先在单个网络上使用图表示学习,然后,提出了一个视图增强模块和视图融合模块,用以丰富单个视图表示以及将多个单视图表示融合。Jiang 等^[19]则基于专利、发明人、专利权人及其相互关系构建了异构网络,并利用该异构网络学习了网络表示,以更好地捕捉不同实体之间的关联性。在此基础上,D Zhai 等^[20]设计了一种新的专利本体结构,包括专利实体、功效、解决方案和应用领域,结合这

些实体间的技术关联,建立了一个异构专利网络,并提出了一种异构网络嵌入算法,将这些信息嵌入到专利表示中,增强了专利价值评估的效果。Han 等^[21]同样基于异构网络,构建了包含专利和发明人节点的网络,提取了节点的语义特征与局部网络结构特征,从而提高了专利价值评估的准确性。

从上述研究可以看出,深度学习技术,尤其是图神经网络的应用,简化了传统机器学习中复杂的特征提取过程,并且在专利分类与价值评估任务中表现出显著优势。而异构图能够更好地表征专利网络中不同类型实体及其复杂关系,并通过结合图神经网络,异构图能够进一步增强信息的传播与融合能力,利用图神经网络的节点嵌入与图结构学习特性,将异构网络中的多维度信息有效整合,从而在专利价值评估和其他相关任务中发挥重要作用。

2 研究设计与研究方法

2.1 专利价值界定的客观指标

在衡量专利价值时,专利被引用次数是广泛采用的客观指标之一。某一特定专利的被引用次数是指该专利被其他专利引用的数量,被认为是衡量专利在企业并购、技术转让和市场影响中的可靠标准。研究表明,专利的被引用次数与其市场价值、技术影响力密切相关。Trajtenberg 等^[7]通过研究专利引用与创新价值之间的关系,发现被引用次数较多的专利往往具有更高的商业价值和技术含量。这是因为高引用次数通常意味着该专利在技术上对后续发明具有重要影响力。

此外,专利被引用次数仍然是衡量专利价值和技术扩散力的有效指标。Breschi^[8]指出,专利被引用次数不仅反映了该专利的知识扩散能力,还与其在创新领域的应用广度密切相关。他们的研究表明,被引用次数越多的专利,其对其他技术领域的影响越大,从而提升了其整体的市场和技术价值。

此外,Hall 等^[22]进一步证明,专利的被引用次数与其市场潜力存在正相关关系。他们发现,高被引用的专利往往能够获得更大的经济回报,特别是在技术授权和市场转让过程中,这类专利的经济价值更高。Harhoff 等^[9]也指出,高被引用次数的专利不仅具有更高的发明质量,还在专利诉讼和技术转让中占据了更加有利的地位,表明其技术和法律价值的综合优势。

然而,专利的引用次数随着时间的推移会显著增加,因此,仅依靠累积的引用次数可能会对新专利造成偏差。为了更准确地评估专利质量,本研究采用年平均被引用次数作为衡量专利价值的指标。这一方法通过考虑时间因素,能够有效消除发表时间较长的专利被引用次数偏高的问题,避免了由于时间长度差异导

致的估值不均衡。通过年平均被引用次数来衡量专利质量,可以更客观地反映专利的技术影响力与市场潜力,从而为专利价值评估提供更加准确的依据。

2.2 专利价值评估框架

本文针对现有专利价值评估方法中图神经网络未能充分利用全局结构信息,以及未有效引入专利被引用量的时间序列数据等问题,提出了一种融合异构图全局结构信息和时间序列的专利价值评估方法。该方法的框架如图 1 所示,主要包括异构图网络和时间序列两个部分。

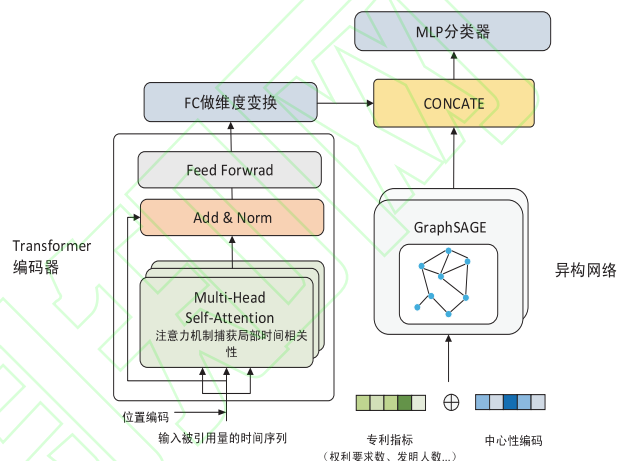


图 1 专利价值评估方法框架

首先,本文构建了异构图网络,并引入节点的中心性编码,将其融入节点的初始嵌入表示中。这一步通过 GraphSAGE^[24]模型实现节点间的消息传递与局部语义的聚合,从而生成专利节点的表示。这种方法弥补了传统图神经网络仅依赖局部结构聚合的不足,通过中心性编码捕获节点在全局网络中的重要性信息,使得专利节点能够更好地反映整体图的结构特征。

与此同时,本文设计了基于 Transformer 架构^[25]的时间序列处理模块,用于处理专利被引用量的时间序列特征。该模块通过多头自注意力机制,捕捉时间序列中的局部和全局相关性,挖掘专利在不同时段的引用变化趋势。这种时间序列的特征提取有助于评估专利的动态影响力,补充了专利价值评估中的时间维度信息。

在整个模型的最后阶段,本文将异构图网络中的专利节点表示与从时间序列模块中提取的特征进行融合,通过全连接层将多维特征向量传递至多层感知机分类器(Multi-Layer Perceptron Classifier, MLP),最终实现对专利价值的预测与分类。通过这一融合过程,不仅能够深度解析专利文本和技术内容,还能够结合历史引用量的变化,全面评估专利的市场表现和未来潜力。

2.3 异构图网络

本文构建的专利异构图网络包含三种主要的节点

类型和三类边,如图 2 所示。节点包括专利节点、发明人节点以及专利权人节点,而边则表示节点之间的三种关系:专利节点之间的引用关系、专利节点与发明人节点之间的发明关系、以及专利节点与专利权人节点之间的专利权关系。这种异构图有效捕捉了专利在技术、法律和人员背景下的复杂关联性。

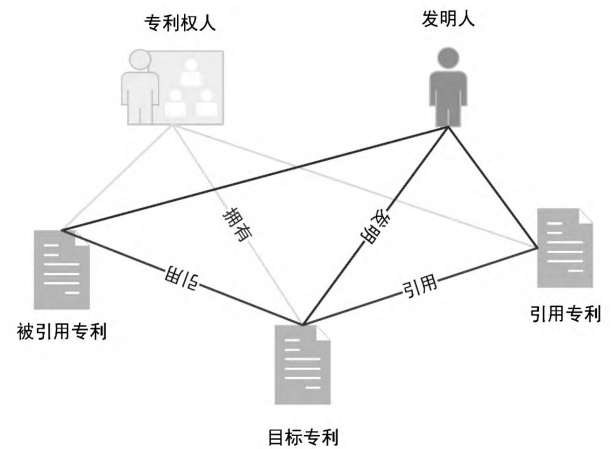


图 2 专利的异构图

2.3.1 异构图节点嵌入

在处理异构图时,传统的图神经网络通过节点之间的消息传递和节点特征聚合,来提取图的空间结构和语义信息。然而,传统图神经网络的深度通常受限于 2 到 3 层,过深会引发节点特征过度平滑的问题。这种限制使得传统图神经网络只能提取到节点附近 2 跳左右的局部结构信息,无法捕捉全局的图结构特征。在专利价值评估任务中,节点的全局中心性是一个至关重要的特征。例如,专利被引用次数较多,往往表明该专利进行了充分的技术调研;发明人数量较多,通常反映出发明团队的规模和潜力。因此,目标专利节点在整个图中的中心性成为评估其价值的一个关键指标。

为了解决传统图神经网络在捕捉全局结构特征方面的不足,本文引入了中心性编码,具体采用了 PageRank 算法进行编码。PageRank 最初用于衡量网页在互联网中的重要性,经过多次迭代为每个节点分配一个正实数,代表该节点的重要性。本文将 PageRank 算法应用于专利异构图中,用以衡量专利节点在全局结构中的重要性。与传统的仅依赖局部连接的图神经网络不同,PageRank 能考虑节点及其邻居的度量信息,从而更准确地捕捉专利的整体结构地位。PageRank 的核心公式为:

$$PR(u) = \frac{1 - d}{N} + d \sum_{v \in B_u} \frac{PR(v)}{L(v)} \tag{1}$$

其中, B_u 是所有直接指向专利 u 的专利的集合, $L(v)$ 是专利 v 指出的链接数, d 是阻尼因子(一般设为 0.85),而 $(1 - d)/N$ 表示随机跳转的概率,即不通

过链接,而是随机打开某个专利的概率。

PageRank 值与如表 1 其他专利指标(如新颖性、科学强度等)一起,作为节点的初始嵌入输入到图神经网络。这种结合全局与局部的多维初始嵌入有助于同时捕捉专利在全局网络中的位置及其重要性。这种方法弥补了图神经网络通常仅限于提取局部结构信息的不足,增强了对全局结构特征的捕捉能力。专利指标(如新颖性、科学强度等)反映了专利技术内容、创新性和法律保护力度等方面的重要属性。通过将 PageRank 值与这些指标融合,图神经网络不仅能够处理专利的定量信息(如引用次数、权利要求数量),还能够捕捉全局图中的结构性信息。这一融合机制使得模型在预测专利价值时,不仅依赖专利的技术特性,还能从专利网络的结构中挖掘更多上下文信息,从而显著提升模型的表达能力。

表 1 专利指标

指标类别	指标计算方式
专利新颖性	参考文献数量
专利科学强度	非专利参考文献数量
专利所在领域的技术周期	引用的专利的平均年龄
专利覆盖范围	专利所属的 IPC 类别数
	专利所属的 CPC 类别数
	覆盖的国家数量
	独立权利要求数量
	从属权利要求数量
专利受让人能力	独立权利要求长度
	从属权利要求长度
	专利权人数量
	发明人数量

此外,结合多维度特征输入,图神经网络能够在更高层次的多维特征空间中进行学习和表征,进而提升专利价值评估模型的预测精度。高价值专利往往不仅具有较强的技术创新性,还在专利网络中占据重要位置。通过融合这些特征,模型能够更精准地识别高价值专利。相比单纯依赖局部信息的模型,若遇到专利特征异常或引用结构偏离的情况,模型的表现可能受到影响。而通过引入全局结构特征,模型能够更加全面地理解专利在整个网络中的相对位置及其整体重要性,进而增强模型的泛化能力。

2.3.2 异构图转换

为了使用图神经网络处理异构图,需要将其转化为同构图。异构图包含多种节点和多种关系类型,而同构图只有一种节点类型和一种边关系,如图 3 所示。在专利领域的异构图中,常见的节点包括专利、发明人和专利权人节点,边表示这些节点之间的关系。为了简化计算和提升模型的可处理性,本文将异构图中的多种节点和边进行转化,具体操作如下:

a. 当专利节点通过专利权人与其他专利节点关联时(即存在“专利-专利权人-专利”的路径),我们消除中间的专利权人节点,在专利节点之间添加一条直接的边,表示这两个专利具有共同的专利权人。

b. 同样,当专利节点通过发明人节点关联时(即存在“专利-发明人-专利”的路径),我们移除发明人节点,并在两个专利节点之间添加一条直接边,表示它们具有共同的发明人。

通过这种转化,原本复杂的异构图被简化为只有专利节点及其直接关联关系的同构图,如图3所示。转换后的同构图中,专利节点的特征将沿着共发明人、共专利权人等关系进行聚合。同一发明人或同一专利权人下的专利具有较高的相似性和价值趋同性,这种同构图能够有效地捕捉这种隐含关系。

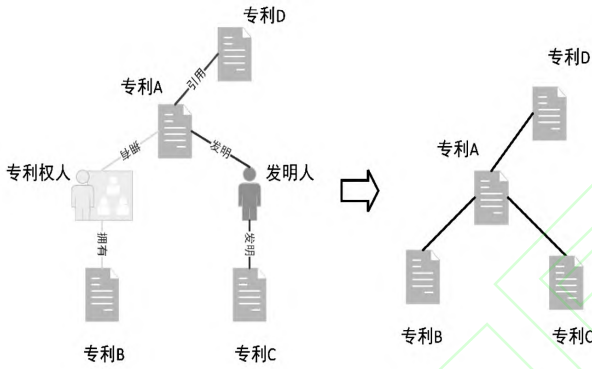


图3 异构图转换

2.3.3 异构图表征

在同构图的基础上,本文采用 GraphSAGE 来学习图中节点的嵌入表示。其的关键思想在于,通过采样和聚合邻居节点的特征,来逐层构建目标节点的嵌入表示。相较于传统的图神经网络,本文通过 GraphSAGE 引入邻居采样策略,能够处理大规模图结构,同时更好地捕捉到节点在图中的局部及全局关系。

首先进行节点采样,异构图在转化为同构图后,每个专利节点可能通过多种边(如共发明人、共专利权人或引用关系)与其他专利节点相连。对于每个目标专利节点,通过随机采样的方式,从邻居集合 $N(v)$ 中选取固定数量的专利节点进行聚合。例如,专利 A 可能通过引用关系连接到专利 D,也可能通过共发明人关系连接到专利 C,将从这些邻居中随机采样固定数量的专利节点来进行下一步的聚合。

接下来进行特征聚合,每个专利节点的初始嵌入包括其各项专利指标(如 PageRank 值、新颖性、科学强度、发明人数量等),这些信息表示专利的局部特征。本文使用聚合函数(如平均聚合、池化或 LSTM 等)来整合这些邻居专利的特征。例如,专利 A 将整合专利 B、C 和 D 的特征,生成新的嵌入表示。

最后,在更新专利节点嵌入时,本文会将目标专利节点自身的特征与从邻居聚合得到的特征进行拼接(CONCAT),形成该专利节点的新嵌入表示。更新公式为:

$$h_v^{(k)} \leftarrow \sigma(W^{(k)} \cdot \text{CONCAT}(h_v^{(k-1)}, h_{N(v)}^{(k-1)})) \quad (2)$$

其中, $h_v^{(k)}$ 表示目标专利节点 v 在第 k 层的嵌入表示, $h_{N(v)}^{(k)}$ 表示从邻居节点聚合得到的嵌入表示, $W^{(k)}$ 是该层的权重矩阵, σ 是非线性激活函数。通过这一更新过程,专利节点的嵌入不仅结合了它的自身特征,还结合了与其具有引用、发明人或专利权人关系的邻居节点的特征信息。

2.3.4 时间序列特征

在专利价值评估中,专利被引用量的时间序列起着至关重要的作用。实践表明,被引用次数较高的专利往往更容易成为高价值专利,且专利的价值与其被引用量之间具有较强的相关性。然而,时间序列不仅反映了被引用量的绝对数值,还包含了变化趋势和增长速度等动态信息。被引用量增长较快的专利,通常表明其技术在市场或学术界具有较强的影响力和应用潜力,因此更有可能成为高价值专利。

对于这类时间序列数据,本文采用 Transformer 模型的自注意力机制聚合时间序列中各时间点的信息,生成整个序列的表征向量。自注意力机制优势在于能够捕捉时间序列中的全局依赖性,不仅能关注单一时间点的被引用量,还能挖掘出专利被引用增长的速度和趋势。

输入的时间序列长度设为 L ,代表专利在 L 个时间单位内的被引用数量。由于专利被引用量随时间递增,故此序列是单调递增的。每个时间点的被引用量视作一个输入 token,该 token 是一个标量,输入特征可表示为 $x \in R^{L \times 1}$ 。

首先,输入特征会经过随机初始化的词嵌入模块,将每个时间点的被引用量转换为相应的嵌入向量 $x_{\text{embedding}} \in R^{L \times d}$,其中 d 为词嵌入的维度。为了让 Transformer 模型能够感知时间序列的位置信息,需要为每个时间点添加位置编码,该位置编码用于表示序列中各 token 的相对位置关系:

$$x_{\text{position}} = [z_{\text{position}}^{(0)}, z_{\text{position}}^{(1)}, \dots, z_{\text{position}}^{(L)}] \quad (3)$$

其中, $z_{\text{position}}^{(i)}$ 是可训练的向量,表示序列中第 i 个 token 的位置嵌入。Transformer 层的输入为:

$$X = x_{\text{embedding}} + x_{\text{position}} \quad (4)$$

该输入经过多层 Transformer 的处理,每个 Transformer 层包含两个子层:自注意力机制层和前馈神经网络层。自注意力机制通过计算序列中所有 token 的自注意力系数矩阵,实现对所有输入向量的聚合。聚合后的向量维度保持不变,随后进入前馈神经网络层

进行进一步的特征提取。最终,通过多层 Transformer 的处理,模型从时间序列中提取出专利被引用量的特征向量。具体而言,取最后一层 Transformer 输出的最后一个时间位置的向量,作为时间序列模块的最终输出特征。

2.3.5 融合特征的专利价值分类

为了全面利用专利的网络结构特征与时间序列特征,本文采用融合特征的专利价值分类方法。首先,专利的全局结构信息通过异构图网络建模,并结合中心性编码作为节点的初始嵌入;然后,使用 GraphSAGE 对这些节点进行信息聚合与特征提取,获得节点的图嵌入表示。与此同时,专利被引用量的时间序列通过 Transformer 模型处理,捕捉时间维度上的引用变化趋势和动态特征。最后,将这两部分特征进行融合,用于专利价值的分类任务。

GraphSAGE 从异构图提取的专利节点嵌入向量,包含了专利在全局网络中的位置与其他专利的关系信息(例如通过共发明人、共专利权人等路径连接的邻居信息)。专利的时间序列特征,则通过 Transformer 捕捉到被引用量随时间变化的趋势和速度,这些特征是预测专利未来市场价值的重要指标。

本文采用特征拼接的方式,将 GraphSAGE 生成的节点嵌入与 Transformer 生成的时间序列特征合并,形成一个多维度的特征表示向量。为确保不同来源的特征能够在模型中得到均衡的考虑,对这些拼接后的特征进行了 Z-score 归一化处理,保证每个特征的均值为 0、标准差为 1,从而避免特征之间因尺度不同产生的不平衡问题。

在分类阶段,本文选择了前馈神经网络(Feedforward Neural Network, FNN)作为分类器。FNN 具有较强的非线性特征学习能力,能够充分利用经过特征融合后的表示向量,进行专利价值的分类任务。通过多层神经网络结构,FNN 可以有效捕捉特征之间的复杂非线性关系,进一步提升专利价值分类的准确性。

通过这种融合异构图全局结构信息与时间序列特征的策略,这种特征融合方法不仅能够捕捉专利间的复杂网络关系,还可以动态跟踪专利被引用量的变化情况,提供了一种信息丰富且具备时序动态感知的专利价值评估方法。

3 实验过程与结果分析

3.1 硬件条件与数据集

本实验在 Ubuntu 18.04 操作系统环境下进行,使用的计算平台配置为 64 位 3.30GHz × 20 Intel® Core™ i9-9820X CPU,128GB 内存,以及两块显存为 20GB 的 NVIDIA GeForce RTX 2080Ti 显卡,保证了数

据处理和模型训练的高效性。

数据集选自 2009 年至 2015 年期间的美国半导体领域专利数据,包含共 102,294 条专利记录。每条记录包含 12 个专利指标,例如独立权利要求数、从属权利要求数、发明人数、专利族专利数、专利族所在国家数等。此外,数据还包括专利摘要、权利要求书文本,以及引用和被引用的专利列表。为了捕捉专利的时间序列特征,本文使用滚动的五年间隔构建引用序列,例如,2009 年专利的被引用量来自 2010 至 2015 年期间的引用,而 2015 年专利的被引用量则来自 2016 至 2021 年期间的引用。

由于发表时间越久的专利往往有更多的引用次数,为消除这一时间积累效应,本文采用了年平均被引用次数作为专利质量的衡量指标。图 4 中展示了截至 2023 年专利的年平均被引用量分布情况。根据年平均被引用次数,将专利划分为三个等级:

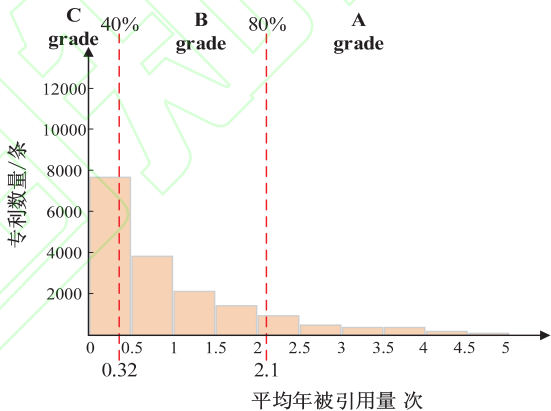


图 4 专利的年平均被引用量

A 等级:年平均被引用次数位于前 20% 的专利,通常代表最高价值的专利;

B 等级:年平均被引用次数位于中间 40% (21% - 60%) 的专利,代表中等价值;

C 等级:年平均被引用次数位于底部 40% 的专利,代表最低价值的专利。

从图 4 可以看出,C 级专利的年平均被引用次数主要集中在 0.32 以下,B 级专利集中在 0.32 至 2.1 之间,而 A 级专利的年平均被引用次数通常超过 2.1。这种基于年平均被引用次数的分类方法,能够较好地反映专利在整个研究时间段内的影响力。

在模型训练时,为了应对不同类别专利数据不平衡的情况,本文使用了带有类别权重的损失函数,确保在类别不平衡的条件下,模型能够对高、中、低价值专利进行更加精确的分类预测。

3.2 模型评价指标

为了验证本文提出 HNTSM 模型的有效性,本文将其与两种现有的专利价值评估模型进行了对比分析,具体对比模型如下:

Dep-Net:该模型通过使用专利摘要、权利要求书文本以及专利指标来评估专利价值。首先,使用卷积神经网络(CNN)和长短期记忆网络(LSTM)提取专利文本的局部和长序列特征。随后,将这些文本特征与专利指标特征通过前馈神经网络进行融合,并最终完成专利价值的分类任务。

GraphSAGE:该模型通过构建包含发明人节点和专利节点的异构图来建模专利的背景信息。专利节点的初始嵌入包含与专利价值相关的指标特征。经过异构图转换为同构图后,使用 GraphSAGE 图神经网络进行节点特征的聚合,最后对专利节点进行分类以判断其是否为高价值专利。

实验的评估指标选用正确率 (Accuracy)、召回率 R (Recall) 和 F1 值 (F1 -Score) 这 3 个指标对实验模型进行评估。通过这些评价指标,能够全面衡量模型的分类性能,尤其是在应对类别不平衡时,F1 分数能够更好地反映模型对正负样本的预测能力。这种对比分析不仅验证了本文所提出模型在专利价值评估中的优越性,还展示了它在处理复杂专利文本与专利网络结构时的鲁棒性。

3.3 实验结果

3.3.1 与主流方法的对比实验

在本实验中,使用不同模型对专利价值进行评估,实验结果如表 2 所示。结果显示,本文提出的 HNTSM 模型在所使用的专利数据集上,相较于其他基线模型在准确率和 F1 得分上均表现出明显优势。这表明,本文提出的两项改进,即引入专利异构图的全局结构中心性编码以及时间序列信息,对最终的预测结果有积极的影响。

表 2 对比实验结果			
	Accuracy/%	召回率 R/%	F1 值/%
Dep-Net	54.67	51.63	51.99
GraphSAGE	61.24	59.35	59.58
HNTSM	67.55	67.10	67.21

通过引入召回率的对比,HNTSM 模型在识别高价值专利的能力上表现更优,召回率达到 67.10%,较 GraphSAGE 提升了 7.75 个百分点。这一结果表明,HNTSM 模型在捕捉高价值专利时有更强的识别能力。

与 Dep-Net 模型相比,HNTSM 模型的准确率提升了 12.88 个百分点。这主要归因在专利价值评估任务中,单靠专利文本(如摘要和权利要求书)难以全面预测专利的真正价值。专利文本往往缺乏对技术应用场景和市场趋势的详细描述,因此模型很难根据这些文本推断出专利是否符合市场需求。特别是在半导体领域,专业术语复杂且相似性较高,专利文本难以清晰

区分技术点的关键性。由于 Dep-Net 模型主要依赖于文本特征和部分专利指标,其评估效果受到限制。

相比 GraphSAGE 模型,HNTSM 模型在准确率上提升了 6.31 个百分点。虽然 GraphSAGE 通过异构图对专利背景信息进行建模,但由于其仅能处理 2 跳邻居节点,缺乏对全局结构信息的处理,无法充分反映专利节点在整个网络中的重要性。HNTSM 通过引入中心性编码,捕捉了专利节点的全局重要性信息,弥补了 GraphSAGE 模型的不足。此外,GraphSAGE 未使用专利的被引用量时间序列,而 HNTSM 通过引入时间序列特征,有效捕捉了专利价值随时间变化的趋势,从而显著提升了预测效果。

综合来看,通过准确率、F1 得分和召回率的提升,本文提出的 HNTSM 模型在专利价值评估任务中展现了更强的整体性能,尤其在识别高价值专利方面表现出色。这表明,HNTSM 模型能够更全面地考虑专利的全局结构信息与时间动态特性,为高价值专利的识别提供了有力支持。

3.3.2 不同专利价值等级实验

表 3 展示了 HNTSM 模型在不同专利价值等级(A、B、C)上的精确率、召回率和 F1 得分。从表中可以看出,模型在 B 类专利上的预测效果最差,其 F1 得分为 47.17%。这一现象可以归因于 B 类专利的年平均被引用次数位于 0.32 ~ 3.1 之间,范围较小且分布较为模糊,使得模型难以有效区分此类专利的价值。这类专利介于高价值与低价值之间,技术或市场表现不够突出,增加了模型的判别难度。

表 3 模型在不同专利价值等级上的实验结果单位:%			
等级	精确度 (P)	召回率 (R)	F1 值
A 等级	77.37	76.09	76.72
B 等级	47.46	46.88	47.17
C 等级	79.19	78.34	78.76

相比之下,模型在 C 类专利上的表现优于 A 类专利,C 类专利的 F1 得分为 78.76%。这表明,识别低质量专利(C 类)相对更为简单,因为低质量专利通常表现出一些共同的、易于捕捉的特征,如技术创新不足、市场表现差等。这些集中的缺点使得模型能够更加准确地学习和识别此类特征。反之,高价值专利(A 类)往往具备独特性和复杂性,这使得模型在此类专利上的预测效果有所挑战,虽然其 F1 得分为 76.72%,但仍低于 C 类。

总体而言,HNTSM 模型在各类别上的表现表明,专利价值越突出或越低,模型的预测效果越好。而对于 B 类专利,由于其价值模糊、特征不明显,模型在该类专利上的表现受到了一定的限制。这也反映了未来模型优化的方向,即如何更好地识别和区分中等价值

的专利。

3.3.3 消融实验

为了探讨 HNTSM 模型中各个模块对专利价值评估任务的贡献,本文进行了消融实验。HNTSM 模型综合了多个方面的特征,包括专利指标、异构图、中心性编码以及时间序列信息。为分析这些模块的独立贡献,本文对 HNTSM 模型进行了逐步拆分,评估在不同特征组合下模型的表现。消融实验结果如表 4 所示。

表 4 消融实验结果

	Accuracy	F1 值
专利指标	47.39	43.44
专利指标 + 异构图	59.64	54.84
专利指标 + 异构图 + 中心性编码	61.32	59.77
专利指标 + 异构图 + 中心性编码 + 时间序列	68.01	67.55

从消融实验的结果可以看出,模型中每个模块对提升专利价值预测的效果都具有重要作用。在仅使用专利指标的情况下,模型的准确率为 47.39%,而当引入异构图信息后,准确率提升至 59.64%。这一提升表明,异构图能够有效捕捉专利背景信息和节点间的关系,为模型提供了额外的结构化特征。

引入中心性编码后,模型的准确率进一步提升了 1.68 个百分点,达到了 61.32%。这一提升符合预期,因为中心性编码并未增加新的特征维度,而是对原有的图结构进行了改进,补充了传统图神经网络无法捕捉的全局结构信息。这一结果表明,中心性编码在捕捉专利节点的全局重要性方面具有显著的效果。

当进一步加入时间序列特征后,模型表现有了显著提升,准确率增加了 6.69 个百分点,达到 68.01%。这一较大的提升可以归因于专利被引用数量的时间序列特征所提供的重要信息。通常情况下,专利的长期被引用数量与短期被引用数量之间有较强的关联性。如果一个专利在短期内的年平均被引用数量较高,通常可以预期该专利在长期内也会表现较好。通过引入时间序列,模型能够捕捉专利引用数量的动态变化趋势,进而提高对专利长期价值的预测能力。此外,使用 Transformer 模型处理时间序列特征,使得模型不仅能够捕捉引用数量的绝对值,还能够识别引用量变化的速度与趋势。通常,年平均被引用量增长较快的专利相比增速缓慢的专利更有可能成为高价值专利。因此,时间序列信息的加入进一步增强了模型的预测能力,特别是在评估专利长期价值方面。

消融实验结果表明,HNTSM 模型中的每个模块都对专利价值评估任务具有积极的贡献,尤其是时间序列和中心性编码模块的引入,为提升模型的预测效果提供了关键支持。

4 结 语

本文提出了一种融合异构图全局结构信息和时间序列的专利价值评估方法,以解决现有专利价值评估模型中对专利全局结构信息和时间序列特征利用不足的问题。模型通过异构图神经网络提取专利的背景信息,并结合中心性编码来捕捉专利节点在全局图结构中的重要性。同时,引入 Transformer 模型处理专利被引用量的时间序列特征,以捕捉专利引用数量的动态变化趋势。研究结果表明,融合异构图全局结构信息和时间序列特征的 HNTSM 模型在专利价值评估任务中表现出显著的优势。

具体而言,本研究通过实验验证了以下几点结论:

a. 本文提出的基于异构图全局结构信息和时间序列特征融合的专利价值评估方法,能够有效提高专利分类的准确性,特别是在识别高价值专利方面。该方法充分利用了专利的结构化信息和时间序列特征,使模型不仅能从专利的技术内容中获取关键信息,还能够分析专利在专利网络中的全局位置及其引用趋势,从而更加全面地评估专利的长期价值。

b. 中心性编码作为全局结构信息的代表,能够有效补充传统图神经网络的局限性。通过引入专利在异构图中的全局中心性,模型能够更加精准地捕捉专利在网络中的影响力与重要性,从而提升对专利价值的预测效果。

c. 时间序列特征对专利价值评估的提升显著,尤其是在结合 Transformer 模型捕捉引用趋势后,模型能够识别出长期被引用数量快速增长的专利。这类专利通常具有较高的创新性和市场潜力,从而使模型能够更准确地判断其长期价值。

本文的贡献在于通过融合异构图全局结构信息和时间序列特征,为专利价值评估提供了一个全新的理论与实践框架。该模型不仅丰富了专利评估的特征维度,还在全局信息感知与动态趋势捕捉方面提供了有效的解决方案,适用于不同领域的专利价值评估任务。然而,本研究也存在一定的局限性。一方面,虽然模型引入了异构图和时间序列特征,但在特征选择的广度上仍存在不足,未来的研究可以探索更多的专利指标以及其他可能影响专利价值的时间序列信息。另一方面,HNTSM 模型的复杂性较高,尤其是在计算时间和资源消耗方面。未来研究可以进一步优化模型,采用轻量化结构,如模型剪枝、量化和知识蒸馏等技术,提升模型的运行效率和实际应用的可扩展性。总之,本文所提出的方法为专利价值评估提供了新的思路和方法基础,未来的研究将进一步完善和扩展该模型,以推动更精准和高效的专利价值识别。

参 考 文 献

- [1] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model [J]. *IEEE Transactions on Neural Networks*, 2008, 20 (1): 61–80.
- [2] Dong Y, Hu Z, Wang K, et al. Heterogeneous network representation learning [C] // *IJCAI*, 2020, 20: 4861–4867.
- [3] Hirschey M, Richardson V J. Are scientific indicators of patent quality useful to investors? [J]. *Journal of Empirical Finance*, 2004, 11 (1): 91–107.
- [4] Deng Y. Private value of European patents [J]. *European Economic Review*, 2007, 51 (7): 1785–1812.
- [5] Boeing P, Mueller E. Measuring patent quality in cross-country comparison [J]. *Economics Letters*, 2016, 149: 145–147.
- [6] 李 娟, 李保安, 方 晗, 等. 基于-熵权法的发明专利价值评估——以丰田开放专利为例 [J]. *情报杂志*, 2020, 39 (5): 59–63.
- [7] Trajtenberg M. A penny for your quotes: Patent citations and the value of innovations [J]. *The Rand Journal of Economics*, 1990, 21 (1): 172–187.
- [8] Breschi S, Lissoni F. Knowledge diffusion and patent citations: A comprehensive review of recent findings [J]. *Research Policy*, 2021, 50 (1): 104–122.
- [9] Sampat B N, Ziedonis R H. The impact of patent citations on litigation outcomes and market value: Evidence from recent cases [J]. *Journal of Technology Transfer*, 2020, 45 (2): 324–346.
- [10] 王思培, 韩 涛. 基于随机森林算法的潜在高价值专利预测方法研究 [J]. *情报科学*, 2020, 38 (5): 120–125.
- [11] Ko N, Jeong B, Seo W, et al. A transferability evaluation model for intellectual property [J]. *Computers & Industrial Engineering*, 2019, 131: 344–355.
- [12] Choi J, Jeong B, Yoon J, et al. A novel approach to evaluating the business potential of intellectual properties: A machine learning-based predictive analysis of patent lifetime [J]. *Computers & Industrial Engineering*, 2020, 145: 106544.
- [13] Jie H, Shaobo L I, Liya Y U, et al. A patent classification model based on convolutional neural networks and rand forest [J]. *Science Technology and Engineering*, 2018, 18 (6): 268–272.
- [14] Liu Q, Wu H, Ye Y, et al. Patent Litigation Prediction: A Convolutional Tensor Factorization Approach [C] // *IJCAI*, 2018: 5052–5059.
- [15] Pujari S, Strötgen J, Giereth M, et al. Three real-world datasets and neural computational models for classification tasks in patent landscape [C] // *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022: 11498–11513.
- [16] Pujari S C, Mantiuk F, Giereth M, et al. Evaluating neural multi-field document representations for patent classification [J]. *Journal of Information Retrieval*, 2022, 25 (4): 410–432.
- [17] Ma S C, Feng L, Yin Y, et al. Research on petroleum patent valuation based on value capture theory [J]. *World Patent Information*, 2019, 56: 29–38.
- [18] Fang L, Zhang L, Wu H, et al. Patent2Vec: Multi-view representation learning on patent-graphs for patent classification [J]. *World Wide Web*, 2021, 24 (5): 1791–1812.
- [19] Jiang H, Fan S, Zhang N, et al. Deep learning for predicting patent application outcome: The fusion of text and network embeddings [J]. *Journal of Informetrics*, 2023, 17 (2): 101402.
- [20] Zhai D, Zhai L, Li M, et al. Patent representation learning with a novel design of patent ontology: Case study on PEM patents [J]. *Technological Forecasting and Social Change*, 2022, 183: 121912.
- [21] Han S, Huang H, Huang X, et al. Core patent forecasting based on graph neural networks with an application in stock markets [J]. *Technology Analysis & Strategic Management*, 2022, 34 (7): 1–15.
- [22] Hall B H, Jaffe A B, Trajtenberg M. Patent citations as a measure of knowledge diffusion and economic value: Revisiting the evidence [J]. *RAND Journal of Economics*, 2021, 52 (3): 486–512.
- [23] Harhoff D, Wagner S. Patent quality and citation impact: The importance of high-quality patents in technology transfer [J]. *Economics of Innovation and New Technology*, 2020, 29 (6): 542–559.
- [24] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs [J]. *Advances in Neural Information Processing Systems*, 2017, 30: 1024–1034.
- [25] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998–6008.