摘要

破产风险预测在金融机构的贷款决策中起着关键作用。针对中小企业的信用评估，有效的信用评级模型可以帮助金融机构做出贷款发放的决策，并限制风险暴露。然而，现有的研究在中小企业破产风险预测领域存在一些问题。首先，数据的获取困难是一个主要问题。中小企业往往无法公开获得财务报表等重要数据，这限制了现有研究的可行性和准确性。为了解决这个问题，我们着手收集真实世界中中小企业的多源数据，并构建了一个新的数据集，其中包括企业的基本业务信息、诉讼信息以及其他相关数据。这样的数据集为我们的研究提供了更广泛和全面的信息基础。其次，现有模型无法很好地建模中小企业之间复杂的关联和相互依赖关系。中小企业之间存在着复杂的供应商关系、客户关系、合作伙伴关系等，这些关系对于评估企业破产风险至关重要。为了克服这个问题，我们采用了异质图网络和超图网络的方法。异质图网络能够处理不同类型的节点和边，更好地捕捉企业之间的异质性关系。而超图网络则能够更全面地建模企业之间的复杂关联，包括基于超边的共同风险因素和来自邻居的直接扩散风险。这两种网络结构的结合为我们的研究提供了更强大的建模能力。通过利用我们构建的新数据集和异质图网络、超图网络的方法，我们的研究旨在提供一种综合考虑多源数据和企业间关系的破产风险预测模型。我们的模型将能够更准确地评估中小企业的破产风险，帮助金融机构做出更明智的贷款决策，并降低其风险暴露。

此外，最近的研究表明，在金融领域中应用人工智能技术时，可解释性变得越来越重要。传统的学习方法常常是黑箱模型，因此可解释的人工智能成为一个研究焦点。虽然已经有人在研究图神经网络的可解释性，但对于异质图的可解释性研究还较为有限。为此，我们提出了一种名为CF3的方法。该方法专门针对异质图，能够捕捉到异质边的信息，提供对模型预测的解释。CF3的核心思想是通过对异质图进行解释，揭示模型的决策依据。它通过学习每个节点特征和边的重要性，以及它们在预测中的贡献，从而生成解释结果。这样，用户可以理解模型为什么做出某个预测。通过CF3，我们可以获得对于异质图中的预测结果更深入的理解和解释。这有助于用户对对破产风险进行评估，使得预测模型的结果更加可信。同时，该方法的应用还可以提高模型在实践中的可解释性和可靠性。

综上所述，本文的研究在中小企业破产风险预测领域的贡献是多方面的。通过收集真实世界中的多源数据、构建新的数据集，并应用异质图网络和超图网络的方法，综合考虑多种数据和企业间关系，更准确地评估中小企业的破产风险。此外，通过引入CF3方法，为解释异质图中的预测结果提供了一种可行的方法，增强了模型的可解释性。这些研究成果对金融机构的贷款决策和风险管理具有重要意义，并有助于推动金融领域人工智能技术的发展。

**关键词：**破产预测，图神经网络，图神经网络的可解释性

**Abstract**

Bankruptcy risk prediction plays a key role in the lending decisions of financial institutions. Effective credit rating models for credit assessment of SMEs can help financial institutions make decisions on loan disbursement and limit risk exposure. However, existing research has some problems in the area of SME bankruptcy risk prediction. First, the difficulty in obtaining data is a major problem. SMEs often do not have public access to important data such as financial statements, which limits the feasibility and accuracy of existing research. To address this problem, we set out to collect multi-source data on real-world SMEs and constructed a new dataset that includes basic business information, litigation information, and other relevant data. Such a dataset provides a broader and more comprehensive information base for our study. Second, the existing models cannot well model the complex associations and interdependencies among SMEs. There are complex supplier relationships, customer relationships, partner relationships, etc. among SMEs, and these relationships are crucial for assessing the risk of bankruptcy. In order to overcome this problem, we adopt the approach of heterogeneous graph networks and hypergraph networks. Heterogeneous graph networks are able to handle different types of nodes and edges to better capture the heterogeneous relationships between firms. Hypergraph networks, on the other hand, are able to more comprehensively model complex associations between firms, including common risk factors based on hyperedges and direct diffusion risks from neighbors. The combination of these two network structures provides more powerful modeling capabilities for our study. By utilizing the new dataset we constructed and the methods of heterogeneous graph networks and hypergraph networks, our study aims to provide a bankruptcy risk prediction model that integrates multiple sources of data and inter-firm relationships. Our model will be able to more accurately assess the bankruptcy risk of SMEs and help financial institutions make more informed lending decisions and reduce their risk exposure.

In addition, recent studies have shown that interpretability is becoming increasingly important when applying AI techniques in finance. Traditional learning methods are often black-box models, making interpretable AI a research focus. Although the interpretability of graph neural networks has been studied, the research on the interpretability of heterogeneous graphs has been more limited. To this end, we propose a method called CF3. The method is specialized for heterogeneous graphs and is able to capture the information of heterogeneous edges to provide an explanation of model predictions. CF3’s core idea is to reveal the decision basis of the model by interpreting the heterogeneous graph. It generates explanations by learning the importance of each node feature and edge, and their contribution to the prediction. In this way, the user can understand why the model made a certain prediction. With CF3, we can gain a deeper understanding and explanation of the predictions in heterogeneous graphs. This helps the user to assess the risk of bankruptcy and makes the results of the predictive model more credible. At the same time, the application of this method can also improve the interpretability and reliability of the model in practice.

In summary, the contribution of the research in this paper in the field of SME bankruptcy risk prediction is manifold. By collecting multi-source data from the real world, constructing a new dataset, and applying the methods of heterogeneous graph networks and hypergraph networks, the bankruptcy risk of SMEs is assessed more accurately by integrating multiple data and inter-firm relationships. In addition, by introducing the hetegnnexplainer method, it provides a feasible method for interpreting the prediction results in heterogeneous graphs and enhances the interpretability of the model. These research results are of great significance to financial institutions' lending decisions and risk management, and help to promote the development of artificial intelligence technology in the financial field.

**Key words:** Bankruptcy Prediction, Graph Neural Networks, Interpretability of Graph Neural Networks.

**目录**

[一、绪论 1](#_Toc156429398)

[1.1研究背景及意义 1](#_Toc156429399)

[1.2国内外研究现状 2](#_Toc156429400)

[1.2.1破产预测领域研究现状 2](#_Toc156429401)

[1.2.2金融领域可解释性研究现状 4](#_Toc156429402)

[1.3研究内容及创新点 6](#_Toc156429403)

[1.4论文组织结构 6](#_Toc156429404)

[二、理论基础和技术介绍 7](#_Toc156429405)

[2.1图神经网络 7](#_Toc156429406)

[2.1.1同质图 7](#_Toc156429407)

[2.1.2异质图 8](#_Toc156429408)

[2.1.3超图 11](#_Toc156429409)

[2.2图神经网络的可解释性 12](#_Toc156429410)

[2.3反事实推理 19](#_Toc156429411)

[三、面向企业破产风险预测的图神经网络模型的设计与实现 21](#_Toc156429412)

[3.1问题定义 21](#_Toc156429413)

[3.2图神经网络模型 21](#_Toc156429414)

[3.2.1异质图模块 21](#_Toc156429415)

[3.2.2超图模块 21](#_Toc156429416)

[3.3实验设置与结果分析 21](#_Toc156429417)

[3.3.1数据集 21](#_Toc156429418)

[3.3.2实验环境 21](#_Toc156429419)

[3.3.3评估指标 21](#_Toc156429420)

[3.3.4实验结果 21](#_Toc156429421)

[3.4本章小结 21](#_Toc156429422)

[四、面向企业破产风险预测的可解释性图神经网络模型的设计与实现 22](#_Toc156429423)

[4.1问题描述 22](#_Toc156429424)

[4.2可解释性图神经网络模型 22](#_Toc156429425)

[4.2.1边信息处理 22](#_Toc156429426)

[4.2.2反事实 22](#_Toc156429427)

[4.3实验设置与结果分析 22](#_Toc156429428)

[4.3.1数据集 22](#_Toc156429429)

[4.3.2实验环境 22](#_Toc156429430)

[4.3.3评估指标 22](#_Toc156429431)

[4.3.4定量研究 22](#_Toc156429432)

[4.3.5定性研究 22](#_Toc156429433)

[4.4本章小结 23](#_Toc156429434)

[五、总结与展望 24](#_Toc156429435)

[5.1工作总结 24](#_Toc156429436)

[5.2未来展望 24](#_Toc156429437)

[参考文献 25](#_Toc156429438)

# 一、绪论

## 1.1研究背景及意义

随着经济的长期发展，中小企业已经成为国民经济中的重要组成部分。根据数据显示，他们不仅对国内生产总值（GDP）的贡献超过40%，还提供了全球50%以上的就业机会。所以，对中小企业的财务风险进行准确预测对于政府决策者和金融机构具有重要意义[1][2]。、中小企业的破产可能会对整个经济系统产生连锁效应，影响就业、经济增长和金融稳定。近年来，互联网的迅猛发展为中小企业带来了新的财务风险，因此，有效的金融风险管理成为当务之急。中小企业的特殊性使其在财务管理方面面临一些独特的挑战。为了应对这些挑战，政府决策者需要了解中小企业的财务状况，以采取适当的政策措施来促进其可持续发展和创造就业机会。同时，金融机构需要评估中小企业的破产风险，以决定是否为其提供贷款和融资支持。有效的破产风险预测模型可以帮助金融机构降低风险暴露并做出明智的贷款决策。因此，致力于构建准确且可靠的财破产风险预测模型对于促进中小企业的稳健发展和维护经济体系的稳定至关重要。

模型的可解释性指的是对于模型的预测结果和决策过程能够进行理解和解释的能力。在机器学习和人工智能领域，可解释性是指能够揭示模型如何根据输入数据做出决策或预测的能力，以及模型对不同特征的依赖程度和影响力。可解释性在各个应用场景中都具有重要意义。首先，在可信度和信任方面，可解释性可以帮助用户或利益相关方理解模型的决策过程和预测结果的依据，从而增强对模型的信任和可信度。这对于涉及重要决策的场景尤为重要，例如医疗诊断、金融风险评估等。其次，在合规和监管要求方面，一些领域和行业对模型的决策过程和预测结果有着严格的合规和监管要求，可解释的模型可以提供对预测结果的解释和验证，满足监管机构的要求。例如，在金融领域的信贷审批决策中，需要解释预测结果的原因；在医疗领域的患者诊断中，需要解释模型依据的医学知识。此外，在错误排查和改进方面，可解释性可以帮助定位模型出现错误或不准确的根源，并进行改进。通过理解模型的决策过程，我们可以识别出可能存在的偏差、缺陷或不完备的特征，从而提高模型的性能和可靠性。最后，在社会影响和伦理考虑方面，一些决策和预测结果可能对个人、社会和环境产生重大影响。可解释性可以揭示模型对不同特征的依赖程度，帮助发现潜在的偏见、歧视或不公平，并采取相应的措施进行纠正和改进。

中小企业在经济发展中起着重要的作用，然而，由于其规模较小、经营环境复杂等因素，中小企业面临着较高的破产风险。因此，准确预测中小企业的破产风险对于金融机构的风险管理和决策具有重要意义。尽管已经有许多研究在这一领域取得了一定的成果，但仍存在一些问题需要解决。

首先，数据的获取和质量是一个挑战。中小企业的财务数据通常不易获得，且可能存在数据不完整、不准确的情况，这会影响预测模型的准确性和稳定性。

其次，模型的建立需要考虑到中小企业之间复杂的关联关系。中小企业之间存在着供应链关系、客户关系、合作伙伴关系等，这些关系对企业的财务状况和破产风险有重要影响。因此，如何有效地将这些关联关系纳入模型是一个问题。

此外，模型的可解释性也是一个重要考虑因素。金融机构需要能够理解模型的预测结果，并对其进行解释，以便做出相应的贷款决策。然而，目前许多预测模型都是黑盒模型，难以解释其预测结果的原因。因此，研究者们需要在提高模型预测能力的同时，注重模型的可解释性。

综上所述，解决中小企业破产风险预测中的数据获取问题、关联关系建模问题以及模型可解释性问题，将进一步提升中小企业破产风险预测的准确性和实用性，为金融机构提供更可靠的决策支持。

## 1.2国内外研究现状

破产预测一直是比较热门的研究领域，发展过程可以追溯至较早的时期。随着时间的推移，对其的研究也在不断深入。同时，在金融领域方面，近期的研究也表明了对于可解释性的重视。以下是对这些发展过程的简要介绍。

### 1.2.1破产预测领域研究现状

破产预测可以通过分析企业的财务状况、市场环境、行业竞争情况等多个因素来进行。利用大数据和人工智能技术，通过构建复杂的模型来预测企业的破产概率。这些模型可以考虑企业的历史财务数据、经营指标、行业发展趋势等因素，从而提供更加准确的预测结果。

国内外的研究者们在中小企业破产预测领域进行了广泛的研究。他们使用各种方法和技术，包括传统的统计模型，如线性判别分析（Linear Discriminant Analysis，LDA）[3]，多变量判别分析（Multiple Discriminant Analysis，MDA）和逻辑回归（Logistic Regression，LR），机器学习算法，如支持向量机（Support Vector Machine，SVM）和决策树，以及深度学习模型[4][5][6]，如深度信念网络（DBN）和卷积神经网络（CNN）等。

随着时间的发展，他们也在中小企业破产预测方面采用了多种机器学习算法和技术。Shetty等人[7]的研究利用了XGBoost（eXtreme Gradient Boosting）和SVM等方法，对比利时2002年到2012年期间的3728家中小企业的财务数据进行了破产风险预测。Min等人[8]提出了一种基于遗传算法的新的二元分类方法来预测企业破产，并通过实证分析验证了其预测能力。du Jardin [9]的研究基于破产的支配知识，引入了"财务状况"的概念，并运用了基于分类器组合的集成技术进行预测。Barboza等人[10]利用了1985年到2013年间的北美公司数据，整合了来自所罗门中心数据库和Compustat的信息，并加入了六个额外的财务指标。在模型选择方面，研究者采用了SVM、随机森林、bagging、boosting等模型和技术。实验结果表明，bagging、boosting和随机森林模型在预测中小企业破产风险方面的表现优于其他技术。另外，当额外变量被引入时，所有测试样本的预测准确率都得到了提高，这进一步验证了额外变量对预测模型的增益效果。除了之前提到的研究，还有[11][12][13][14]等文献也对中小企业财务风险预测进行了相关研究。

随着深度学习技术的兴起，越来越多的研究者开始探索将其应用于中小企业破产风险预测领域。深度学习作为一种强大的机器学习方法，具有对大规模数据进行高效处理和学习复杂特征的能力，因此在破产风险预测方面具有巨大的潜力。在深度学习方面的相关研究中，研究者们通常会利用深度神经网络来构建预测模型。这些模型可以自动从大量的财务数据中学习关键特征和模式，以预测中小企业的破产风险。其中一些研究使用人工神经网络（Artificial Neural Networks，ANN）作为预测模型，并与传统的方法进行比较。例如，Odom等人[15]开发了一个神经网络模型用于破产预测，并与传统的多变量判别分析方法进行比较，结果显示神经网络在这个问题上可能是有效的。Kovacova等人[16]则比较了ANN和多变量判别分析法在预测破产方面的能力，结果表明神经网络可能适用于这个问题。Hosaka等人[17]的研究将卷积神经网络（Convolutional Neural Networks，CNN）应用于企业破产预测，并表明相较于传统方法，基于CNN的预测性能更好[17]。此外，一些研究尝试将深度学习模型与其他机器学习方法相结合，以进一步提升预测能力。例如， Korol等人[18]将模糊集、循环神经网络和决策树等方法结合起来，开发了预测欧洲企业破产的动态模型，并进行了评估[18]。而Kim等人[19]使用了循环神经网络（Recurrent Neural Networks，RNN）和长短期记忆算法（Long Short Term Memory，LSTM），并与LR、SVM和随机森林等方法进行比较，结果显示RNN和LSTM方法可以改善破产预测性能。杨钟瑾[20]的研究将粒子群算法、遗传算法和SVM相结合，提出了一种优化的破产预测模型，该模型利用粒子群算法和遗传算法对SVM模型进行参数优化，成功地提高了破产预测的准确率。顾啸天[21]关注传统的数据增强算法（如SMOTE）和高斯混合模型，并将它们与集成分类算法相结合，提出了一种名为GSmoteABNN的新算法，用于企业破产预测。这样的方法提高了破产预测的性能，使其优于传统方法的效果。

在企业互连方面，企业之间的投资、合作、供应链关系等可以被表示为图中的节点和边。这些节点和边之间的连接关系可以反映企业之间的依赖、交易和风险传播。通过构建企业互连图，可以更好地理解和分析企业之间的相互作用，以及它们对整个经济系统的影响。因此，研究人员开始使用图神经网络（Graph Neural Networks，GNN）来模拟和分析企业互连的图结构，以处理各种风险预测任务。例如，Wang等人[22]提出了SemiGNN，利用多视角的标记和未标记数据进行欺诈检测，提高了欺诈检测的准确性。Hu等人[23]对企业互连图中的关系、对象、节点和边的丰富属性进行建模，使用GNN来学习图中节点和边之间的复杂关系，以提高贷款违约的预测性能。Feng等人[24]结合传统的信用评级方法和GNN模型，提出CCR-GNN方法，用于学习和预测企业的信用等级。Yang等人[25]的研究中关注于中小企业与信用相关的拓扑结构和时间变化，通过在中小企业图上挖掘供应链关系来分析财务风险。Kosasih等人[26]将供应链可见性问题视为链接预测问题，并使用GNN模型来预测供应链中节点之间的连接和可见性。Wang等人[27]建立了动态图上的信用风险预测模型TemGNN，以挖掘短期和长期的时间-结构信息，从而预测用户的信用风险。Zheng等人[28]提出了一种基于三层关注网络的破产预测方法，考虑了基于元路径的不同邻居节点之间的关系，使用GNN来学习和捕捉图中不同层次的关注关系，从而提高破产预测的准确性。

### 1.2.2金融领域可解释性研究现状

在金融领域，了解模型预测结果的原因和依据至关重要，因为金融决策可能对个人、企业和整个经济系统产生深远影响。随着技术的发展和金融产品的复杂性增加，投资者和监管机构需要理解金融模型和算法背后的决策过程。可解释性可以帮助我们了解金融决策的基础和逻辑，增加对风险的认识，以及监测和预测市场的变化。首先，可解释性可以提高金融产品的透明度。金融产品的复杂性可能使投资者难以理解其内部运作方式。通过提供可解释的模型和算法，投资者可以更好地了解金融产品的特征和风险，从而做出更明智的投资决策。其次，可解释性对于监管机构来说也非常重要。监管机构需要对金融市场进行监督和管理，以确保市场的稳定和公平性。可解释的金融模型和算法可以帮助监管机构更好地理解市场的运作，及时发现潜在的风险，并采取适当的监管措施。另外，可解释性还有助于金融机构自身的风险管理。金融机构需要评估和管理各种风险，包括市场风险、信用风险和操作风险等。通过可解释的模型和算法，金融机构可以更好地识别和量化风险因素，并制定相应的风险管理策略。最后，可解释性帮助决策者做出决策。金融决策通常涉及重要的资产配置、投资组合和风险管理。了解模型预测结果的原因和依据可以帮助决策者做出明智的决策，并对决策结果负责。

为了满足上述需求，一些研究人员进行了相应的研究。其中，Tsogoo [29]开发了一种基于49000家英国公司数据集的破产预测可解释型机器学习方法，通过该方法，研究人员能够比较不同模型的解释性能，从而帮助理解模型预测的原因和依据。Cho等人[30]提出了一种基于遗传算法的反事实生成算法，用于破产预测。该算法可以生成模型的反事实解释，为用户提供备选案例，使用户能够从模型中获取他们期望的输出，从而增强了模型的可解释性。

除了破产预测领域外，欺诈检测领域的可解释性也越来越受到关注。Rao等人[31]提出了一个名为xFraud的可解释欺诈交易预测框架。该框架旨在预测传入交易的合法性，并提供合理的解释。通过使用xFraud，用户能够有效地了解模型对于交易的预测结果，并获得解释来支持他们的决策。Qin等人[32]提出了NGS（神经元图搜索）来用于欺诈检测。它将GNN的消息传递过程形式化为一个元图，并设计了一个可区分的神经架构搜索来确定优化的消息传递图结构。通过聚合多个搜索到的元图，研究人员进一步增强了模型的性能和可解释性。Wu等人深耕欺诈检测的可解释性领域，首创了DualFraud[33]，专注于检测具有可解释性的欺诈企业和交易。该框架通过促进两个领域的信息共享，提升学习和建模能力，从而显著改进欺诈预测。紧接着，他们进一步提出了多任务学习框架MultiFraud[34]，充分利用实体周围多视角的异构信息，实现多领域共享嵌入信息，进一步增强了欺诈检测的建模能力。

## 1.3研究内容及创新点

引入

本文的具体研究工作如下：

1. 提出了一个破产风险预测模型。基于实验室的数据集，建立适用于中小企业的破产风险预测模型。因为企业之间的关系是复杂而多样的，涉及到不同企业之间的合作、竞争、交易等多种互动方式。传统的建模方法难以完全捕捉和表达这些复杂的企业关系，故采用图神经网络模型进行建模，能更好的学习企业之间的复杂关系。创新点：（1）同时考虑企业的内部风险和传染风险，将异质图和超图融合在一起进行建模；（2）超图建模时考虑了残差，保留了一些信息

2. 在破产风险预测模型中，可解释性是一个重要的考量因素。需要探索模型的可解释性，即通过解释模型的预测结果，揭示模型如何基于输入特征做出决策，以便从业者能够理解模型背后的逻辑和规律。针对于此，提出了一个异质图神经网络的可解释性模型。创新点：（1）用不同的metapath得到不同的子图，再将子图拼接起来；（2）将反事实和事实同时应用于模型。（3）引入GCN，按照关系子图关系进行分类，以此来获取不同子图的子图的特征表示

## 1.4论文组织结构

根据本文的研究内容，对应的章节安排如下：

第一章为绪论，首先介绍了企业破产风险预测的研究背景及其对社会经济的重要意义。随后，按照时间发展，系统性阐述了其国内外研究领域的现状。为了更好地契合现实需求，进一步引入了金融领域可解释性的来源及当前发展趋势。然后，简述了企业破产风险预测和可解释性方面所面临的挑战和问题，最后，具体阐述了本文所提出的研究方法和创新点。

第二章为理论基础和技术介绍，深入探讨了本文提出的模型所涉及的关键技术。其中包括对图神经网络中同质图、异质图和超图等概念的详细解释及其在模型中的应用。然后进一步介绍了图神经网络可解释性的细致分类，提供对该领域的清晰认识。此外，章节还对反事实推理的定义和在模型中的具体运用进行了详细阐述。

第三章介绍了提出的面向企业破产风险预测的图神经网络模型ComHHGNN，

第四章介绍了提出的面向企业破产风险预测的可解释性图神经网络模型CF3，

第五章为总结与展望，对本文提出的模型进行了一个概括性总结，并对未来研究的方向和可能的拓展提出了一些展望。

# 二、理论基础和技术介绍

本章主要是对后续研究可能会使用的关键技术和理论进行归总。通过对这些内容的整理，将为后续的研究提供坚实的基础。

## 2.1图神经网络

图神经网络（GNN）使用深度神经网络来对图数据进行表示学习，已经在许多图上的任务中取得了成功应用，例如节点分类[22][23]、链接预测[33][36]和社区检测[37][38]。除了图上的任务，GNN也在传统的领域中发挥了积极作用。在推荐系统[39][40]中，GNN可以利用用户-物品交互图来提供个性化的推荐结果。在自然语言处理[41][42]中，GNN可以处理基于语义和语法结构的文本数据。在计算机视觉[43][44]中，GNN可以处理图像和视频数据，并进行对象识别、目标跟踪等任务。文献[45]里详细介绍了关于GNN的更多信息。

### 2.1.1同质图

同质图是指节点和边的类型都相同的图，其特点在于节点和边使用相同的特征进行表示。同质图广泛应用于各种领域，如社交网络、引文网络、蛋白质相互作用网络等。GNN已经在同质图的表示学习和任务解决中展现出了良好的性能和潜力。在图1-1(a)中，可以看到一个由节点和边组成的网络结构。这个网络中每个节点都代表了一个特定的文章，这些文章可能来自不同的作者、领域或主题。同时，网络中每条边都表示了一篇文章对另一篇文章的引用，这种引用关系可以是直接的，也可以是间接的。



图1-1 同质图与异质图

Kipf等人[46]首次提出了一种在图结构数据上进行半监督学习的可扩展方法GCN（Graphical Convolutional Neural Networks），通过局部一阶近似的谱图卷积来激励卷积结构的选择，并学习编码节点特征和局部图结构的隐藏层表征。GCN通过将节点的特征与其邻居节点的特征进行聚合和更新，从而逐步传播和融合全局和局部信息。这种聚合过程可以通过邻接矩阵的乘法来实现，从而捕捉节点之间的关系和图的拓扑结构，在节点分类、节点嵌入、链接预测等任务中取得了良好的性能。Hamilton等人[47]在GCN模型的基础上进行了改进，提出了GraphSAGE（Graph SAmple and aggreGatE），其核心思想是通过采样和聚合的方式来学习节点的表示。通过多次迭代采样和聚合的过程，GraphSAGE能够逐渐扩展节点的邻域范围，从而捕捉到更广泛的上下文信息。最终，GraphSAGE能够学习到具有丰富语义表示的节点嵌入，可用于各种图相关的任务，如节点分类、链接预测等。Veličković等人[48]则将注意力机制引入GNN当中，提出了GAT（Graph Attention Networks），其主要思想是为每个节点和其邻居节点分配不同的注意力权重，从而对不同节点之间的关系进行建模。相比于传统的GNN模型，GAT能够更加灵活地学习节点之间的关联性，使得模型能够更好地适应复杂的图拓扑结构。通过多层的注意力计算和节点表示更新，GAT模型逐渐融合全局和局部信息，捕捉节点之间的复杂关系。GAT具有高度的灵活性和表达能力，在图节点分类、链接预测等任务中取得了优秀的性能。

### 2.1.2异质图

异质图指的是节点和边的类型可以不同的图，其特点在于节点和边具有不同的属性和特征。异质图广泛应用于许多实际场景中，例如，在在线交易网络中，欺诈者更有可能与客户建立联系，而不是与其他欺诈者建立联系；在约会网络中，大多数人更喜欢与异性约会。异质图具有更复杂的结构和丰富的信息，因此需要特殊的方法来进行表示学习和任务解决，由此诞生了异质图神经网络。异质图神经网络的目标是学习异质图中节点的表示向量，以便于进行节点分类、链接预测、图生成等任务。它通过结合不同类型节点和边的信息，利用GNN的方式进行特征传播和聚合，从而捕捉异质图中的结构和语义信息。同质图中的节点和边具有相同的类型，而异质图中的节点和边可以具有不同的属性和连接方式，这使得异质图能够更好地描述复杂的关系和依赖关系，如图1-1(b)。在该图中，我们可以观察到三个不同类型的节点：作者、文章和会议。每个节点都代表了不同的实体或对象，具有独特的属性和特征。此外，图中还存在四种不同类型的边：参加、出版、撰写和引用。这些边表示了节点之间的各种关系，如作者参加会议、文章被出版、作者撰写文章以及文章被其他文章引用。通过分析作者节点与会议节点之间的参加边，我们可以了解哪些作者参加了哪些会议，从而研究作者与会议之间的合作关系和互动情况。这有助于揭示作者在不同会议中的活动范围和研究领域。另一方面，通过分析文章节点与文章节点之间的引用边，我们可以探索文章之间的引用关系和知识传播。这可以帮助我们了解哪些文章被其他文章引用，进而评估文章的影响力和重要性。文献[83]里对异质图神经网络进行了全面的回顾。

Schlichtkrull等人[50]将GCN框架应用于关系数据建模，以解决在图结构中不同边关系对节点的影响。他们在具有大量关系的多图中应用了参数共享和稀疏约束技术，采用了一个执行多步信息传播的模型，以加强因子分解模型的效果。Zhang等人[51]则引入了随机行走与重启策略，用于为每个节点抽取固定规模的强相关异质邻居。然后，根据节点的类型将这些邻居节点分组，并设计了一个包含两个模块的神经网络架构，以汇总这些采样的邻居节点的特征信息。Wang等人[52]意识到了注意力机制在深度学习中的巨大潜力，并将其引入了异质图神经网络中。他们提出了一种基于分层注意力的新型异质图神经网络，该模型包括节点级注意力和语义级注意力。节点级注意力的目标是学习节点与基于元路径的邻居之间的重要性，而语义级注意力则用于学习不同元路径的重要性。模型通过从节点级和语义级注意力中学到的重要性，可以分层地聚合基于元路径的邻居特征，能够捕捉到不同级别的信息，从而生成节点的嵌入表示。Yang等人[53]的研究专注于解决异质信息网络中GCN方法存在的缺陷，由此提出了一种高效的异质图神经网络用于学习异质信息网络中对象的表示。通过理论分析，研究人员表明所提出的模型具备评估所有可能的元路径有用性的能力，并与基于谱的图卷积方法在异质信息网络上的联系密切。

在现实世界中，同质图和异质图都广泛应用于不同的领域和问题。同质图通常用于描述具有相同类型的节点和边的关系网络，如引文网络、传播网络等。而异质图则用于描述具有不同类型的节点和边，节点和边可能具有不同的属性和特征，如推荐系统中的用户-物品关系、知识图谱中的实体关系等。由于同质图和异质图的特点和复杂性不同，研究人员开始思考如何设计一个模型，能够同时在这两种类型的图上表现出较好的效果。这种模型将具有广泛的应用前景，能够更全面地处理不同类型的图结构，提取丰富的信息并应用于多样的任务。

Wang等人[54]提出了一个模型，能够根据节点对之间的同质性或异质性自动调整传播和聚合过程。为了实现自适应学习传播过程，研究人员引入了两个度量来衡量节点对之间的同质性程度，这些度量基于拓扑信息和属性信息进行学习。通过学习这些度量，研究人员可以确定节点对之间的同质性程度，并将可学习的同质性程度整合到图卷积网络框架中。这样做的好处是，传播和聚合过程能够根据节点对之间的同质性程度进行灵活调整，从而更好地捕捉到图中的异质性信息。此外，研究人员在理论上证明了该模型的有效性，验证了根据节点之间的同质性程度约束节点表示的合理性。这进一步支持了该模型在处理图数据中的异质性问题上的有效性和可行性，使得模型能够更好地捕捉到图中的异质性，并提供更准确和可解释的节点表示。

He等人[55]的研究旨在解决GCN在处理同质性和异质性网络时的限制。为了应对这一挑战，研究人员引入了块建模的概念，并将其融入到GCN的框架中。通过块建模，GCN可以根据邻居的亲和度有选择地聚合来自同类邻居和异类邻居的信息。这种自适应的聚合机制使得GCN能够更好地处理异质性网络，并根据不同邻居之间的关系进行有针对性的信息传播和聚合。通过根据邻居的亲和度来调整聚合过程，GCN能够更准确地捕捉到网络中的异质性信息，从而提高节点表示的质量。一系列实验表明，该模型在处理异质性数据集时相对于现有方法具有明显的优势，并且在同质性数据集上表现出有竞争力的性能。这意味着它不仅能够有效地处理具有异质性的网络，而且在处理同质性网络时仍然能够保持较高的性能水平。

由于企业风险涉及多个因素和关联关系，这些因素和关联关系涵盖了不同类型的数据，例如企业基本信息、财务数据、市场数据、员工信息等。这些数据之间存在着多样性和复杂性，不适合简单地使用单一的图结构来表示。因此，采用异质图进行建模是更为合适的选择。异质图能够灵活地处理不同类型的节点和边，从而更准确地反映出这些数据之间的关联关系。通过利用异质图进行企业风险建模，我们能够更全面、详细地分析和理解不同类型数据之间的相互作用，进而提供更精准、全面的风险分析和管理。

### 2.1.3超图

传统的图神经网络中，图被表示为由节点和边构成的网络结构。每个节点代表一个实体或对象，边则表示节点之间的关系或连接。然而，有时节点之间的关系可能涉及多个节点的关联，而不仅仅是一对一的连接。为了处理这种复杂情况，研究人员引入了超图（Hypergraph）的概念。超图是图的一种扩展形式，允许一条边连接多个节点，而不仅限于两个节点之间的连接。这样的边被称为超边（Hyperedge），超边可以连接任意数量的节点。通过这种连接方式，超图能够更好地表达多个节点之间的复杂关系和依赖关系。根据图1-2，可以观察到左侧(a)图展示了一个普通图的示例。在普通图中，每条边仅连接两个节点。而右侧(b)图呈现了一个不同的情况，其中包含了5条超边，分别标记为。需要特别注意的是，边连接了三个节点，这展示了一种更为复杂的连接方式。



图1-2 普通图与超图

在超图神经网络中，传统的图神经网络模型被扩展以处理超图数据。超图神经网络考虑了超边的信息以及节点之间的关联来学习节点的表示。超边的信息可以包括超边本身的特征，也可以包括连接的节点的特征，从而丰富了节点的表示能力。文献[56]里系统回顾了有关超图的文章，为超图神经网络的研究提供了参考。

Sun等人[57]提出了一种基于图神经网络的异质超图表示学习框架，旨在有效地捕捉多种非成对关系。这个框架首先将异质超图映射到一系列快照中，然后采用小波基进行局部超图卷积。其核心思想是利用图神经网络的能力来处理异质超图数据，并通过局部超图卷积来捕捉节点之间的复杂关系。这种方法不仅可以很好地表征多种非成对关系，还在实验中展现出了出色的性能。Feng等人[58]的研究专注于数据表示学习，并提出了一种超图神经网络框架，旨在能够在超图结构中捕捉高阶数据的相关性。为了处理复杂数据，他们设计了一种超图卷积操作，以处理表示学习过程中的数据相关性。这种设计使得传统的超图学习方法能够有效地利用超边缘卷积操作。通过对引证网络分类和视觉物体识别任务进行实验，研究结果表明该框架具有出色的性能。Feng的研究团队[59]意识到，现有的图神经网络框架在处理复杂的多模态/多类型数据相关性方面存在局限性，因为它们通常是基于简单图的部署。而最近一些基于超图的方法被提出来解决多模态/多类型数据的相关性问题，这些方法通常采用直接连接每个单独模态/类型构建的超图，很难为每个模态/类型学习到自适应的权重。为了解决这个问题，他们扩展了之前的工作，并引入了一个通用的高阶多模态/多类型数据关联建模框架[59]。该框架采用了基于超图的方法，以学习一个最优的数据表示。在这个框架中，首先构建了超图组，用于表示每个特定模态/类型中的潜在高阶关联，并使用显式或隐式图结构。然后，采用自适应的超边缘组融合策略，有效地将来自不同模态/类型的相关性融合成一个统一的超图。接下来，他们提出了一种新的超图卷积方案，基于空间域进行学习，用于学习各种任务的通用数据表示。通过在几个流行的数据集上进行实验证明，这个框架能够持续超越现有方法，特别是在建模隐含数据相关性方面具有显著优势。

在企业风险建模中，涉及的数据往往包含了大量的复杂关系和依赖。这些关系不仅存在于企业内部的各个部门和岗位之间，还涉及到与企业相关的外部实体和人员。因此，使用传统的图结构来表示和分析这些关系可能会遇到一些限制。

超图作为一种扩展形式的图结构，可以更好地处理企业风险建模中的复杂关系。超图允许一条边连接多个节点，这在企业风险建模中非常适用，因为一个风险事件通常涉及多个企业实体和相关人员之间的关联。通过使用超图，我们可以更全面地描述和分析企业风险的复杂性。然而，目前在企业风险建模领域应用超图神经网络的研究还比较有限。

## 2.2图神经网络的可解释性

可解释性在各个应用场景中都具有重要意义。模型的可解释性对于用户、利益相关方、合规要求、错误排查和改进、以及社会影响和伦理考虑都具有重要意义。通过提高模型的可解释性，可以增强对模型决策和预测结果的理解，提升模型的可信度、合规性和性能，从而更好地应用于实际场景中。

随着对模型可解释性需求的扩大，研究人员和从业者已经提出了许多方法和技术，包括特征重要性分析、局部可解释性方法、规则提取等，以增强机器学习模型的可解释性。特征重要性分析，指通过分析模型中特征的权重、特征的贡献或特征选择的结果，了解哪些特征对于模型的预测结果起到了重要作用。例如，对于线性模型，可以通过查看特征的系数来了解其对预测的影响程度；局部可解释性方法，可以显示出模型对于某个特定样本中每个特征的重要性，或者通过可视化模型的决策边界来解释为什么模型对某个样本做出了特定的预测。常见的局部可解释性方法包括局部可解释性模型（Local Interpretable Model-agnostic Explanations，LIME）[60]和SHapley Additive exPlanations（SHAP）[61]等；规则提取，指某些模型（如关联规则、关联规则学习、关联规则挖掘）可以从数据中提取出一组规则，这些规则描述了输入特征与输出之间的关系。这样的规则可以帮助理解模型中的模式和逻辑；决策树解释，通过分析决策树的分支和节点，可以理解模型是如何基于输入特征进行决策的；简化模型，使用简化模型来近似复杂模型，因为简化模型通常具有更高的可解释性，例如线性模型、决策树等。通过使用简化模型，可以更容易地理解和解释模型的预测结果；证明和证据，一些方法试图通过形式化逻辑或概率推理的方式提供模型决策的证明或证据。这种方法通常与符号推理、贝叶斯网络等技术结合使用，以提供更可靠的解释结果。

尽管图神经网络（GNN）在处理图结构数据和捕捉节点之间的复杂关系方面具有优势，但它们的预测过程通常缺乏直观的可解释性。然而，理解GNN预测的能力对于多个方面都非常重要。首先，可解释性有助于用户、利益相关方或部署GNN模型的人员更好地理解模型的预测结果和决策过程。通过了解模型如何基于输入图结构进行决策，人们可以更容易地相信模型的准确性和可靠性，从而增加对模型的信任。其次，对于决策关键的应用领域（如社交网络分析、金融风险评估和医疗诊断），要求模型的决策过程和预测结果能够进行解释和验证。通过理解GNN的预测过程，可以检查模型是否存在潜在的偏见、歧视或不公平，并采取相应的措施进行纠正和改进。此外，对于涉及隐私数据的应用，如个人用户数据或敏感信息，模型的透明度可以帮助验证模型是否合规，并保护用户隐私。此外，理解GNN预测的过程有助于从业者识别模型所犯错误的系统模式。通过观察模型在不同图结构上的表现和决策依据，可以发现模型的弱点、偏差或不准确性。这些洞察可以指导改进模型的方法，并在将模型部署到现实世界之前进行必要的纠正。因此，透明化GNN的预测过程是促进其在现实世界中更广泛应用和提高可靠性的关键。通过增强对GNN模型预测的可解释性，可以提高模型的信任度，处理公平性、隐私保护和安全挑战，并帮助从业者识别和改进模型的系统错误。对此，有一些研究者进行了研究。范钰[62]将图神经网络的可解释性应用于推荐算法中，提出了一种名为GLAM的模型。GLAM基于待推荐候选之间的多种关系设计了异构图，并利用图卷积和匹配机制来提取重要信息，从而实现解释性推荐。董彬[63]将图神经网络的可解释性应用于“初等数学类人解题系统”，主要分为两个类别：输入无关的模型级别可解释性和输入相关的实例级别可解释性。这些可解释性方法被用于知识表示和实例化定理选取。在初等数学知识的表示学习中，首先进行实体关系的嵌入学习和表达式的嵌入学习。对于实体关系三元组的嵌入表示学习，进行了对TransE方法的计算可解释性的研究分析，并将其应用于嵌入表示学习。在数学表达式的表示学习中，提出了符合数学知识的表达式置换等价任务，并对BERT结构的可解释性进行探讨，将其应用于数学表达式置换等价任务的训练中。

一般来说，可以将GNN的可解释性分为事前可解释性（Intrinsic Explanations），也称为自解释性，和事后可解释性（Post-hoc Explanations）[62]。事前可解释性指的是在进行预测之前，通过解释模型的结构、参数或学习过程，以提供对模型行为的理解。这种解释通常与GNN的架构、图卷积层、节点嵌入方法等相关。通过事前可解释性，人们可以了解模型如何利用图结构信息进行学习和推理，从而对模型的预测过程有更深入的理解。因为本文所采用的方法是事后可解释，故对于事前可解释这里就一笔带过，不进行深入讨论。

事后可解释性指的是在模型完成预测之后，通过解释预测结果和模型决策过程，以提供对模型行为的解释。这种解释通常与特定样本的预测结果、节点重要性、图结构中的关键路径等相关。通过事后可解释性，人们可以理解模型为什么对特定样本做出了某种预测，并且可以通过可视化、特征重要性分析等方式来解释模型的决策依据。事后可解释又可以分为实例级（instance-level）事后解释和模型级（model-level）事后解释[62][65]。具体分类见图1-3。



图1-3 图神经网络的可解释性分类

实例级事后解释主要分为模型已知（Model-aware）与模型不可知（Model-agnostic）。这些方法旨在揭示模型预测的原理和决策过程，以便更好地理解和解释模型的行为。在模型已知方法中，研究人员利用对模型内部结构和参数的了解，使用基于梯度/特征的方法或者使用基于分解的方法来解释模型的预测。在模型不可知的方法中，研究人员尝试通过不依赖于具体模型结构和参数的方式来解释模型的预测。这些方法主要关注于输入和输出之间的关系，而不涉及模型内部的推理过程，包括基于扰动的方法、基于代理的方法和基于反事实的方法。

在使用基于梯度的方法解释深度模型时，关键思想是将梯度视为输入特征的重要性度量。较大的梯度值表示该特征对于模型的预测结果具有更大的影响。通过计算模型输出相对于输入的梯度，可以评估每个输入特征对预测结果的重要性。这种方法常用于图像和文本任务的解释，例如图像中的像素重要性分析或文本中的单词重要性分析。Baldassarre等人[66]的研究采用了梯度的平方值作为不同输入特征的重要性分数。这种方法可以直接通过反向传播计算，类似于训练模型时计算梯度，但目标是输入特征而不是模型参数，这些输入特征可以是图的节点、边或节点特征。该方法基于一个假设，即绝对梯度值越高，相应的输入特征越重要。因此，通过计算梯度的平方值，可以得到不同特征的重要性分数，用于解释模型预测的结果。Pope等人[67]的研究关注于图分类模型，将CAM[67]方法扩展到一般的图分类模型。传统的CAM方法使用全局平均池化层输出和全连接层输出之间的权重来生成特征映射。但在该研究中，他们放宽了GAP层的约束，引入了梯度作为权重，以组合不同的特征映射。具体而言，他们首先计算目标预测相对于最终节点嵌入的梯度。然后对这些梯度进行平均，得到每个特征映射的权重。这样，通过将权重与特征映射相乘，可以获得组合后的特征映射，用于解释模型的预测结果。

基于分解的方法是一种常见的用于解释深度图像分类器的模方法。它通过将原始模型的预测分解为几个术语来量化输入特征的重要性，并将这些项作为相应输入特征的重要性分数。这些分解方法直接关注模型参数，以揭示输入空间中的特征与输出预测之间的关系。它们通常基于一些假设或方法来分解预测结果，以便确定每个输入特征对于最终预测结果的贡献程度。具体的分解方法可以根据问题和模型的特点而有所不同。Schnake等人[68]提出了GNN-LRP模型，将输出的预测分数分解为不同的节点重要性分数，以解释分类器的预测结果。这种方法基于隐藏特征和权重，并提出了一种分数分解规则。根据该规则，对于目标神经元，它的分数被表示为前一层神经元分数的线性近似值。换句话说，目标神经元的激活程度与贡献度较高的神经元的分数有较大的关联。通过计算前一层神经元分数的线性近似，可以推断出哪些神经元对目标神经元的评分贡献较大。通过对模型进行分解，可以识别出对于特定预测结果起关键作用的节点，并为这些节点分配相应的重要性分数。Pope等人[67]同时还提出了Excitation BP模型，该模型的核心思想是定义当前层神经元的概率等于其输出到下一层所有连接神经元的总概率。根据模型，目标神经元的概率可以被分解为若干个条件概率项，通过分解目标神经元的概率，可以推断出对目标神经元激活程度有重要贡献的神经元。这个模型的提出为分数分解规则提供了解释。

基于扰动的方法是一种常用的图神经网络解释方法之一，该方法通过对输入图进行微小的扰动或修改，来观察模型预测结果的变化，从而推断模型对不同图结构的敏感性和重要性。Ying等人[69]提出了GNNExplainer，旨在提供对任何基于图的机器学习任务中GNN模型预测的可解释性解释。它的目标是帮助理解GNN模型的预测过程，揭示模型如何利用图结构和节点特征进行推理和决策，其核心思想是通过解释图神经网络中节点的重要性来解释模型的预测。它学习边缘和节点特征的软掩码，通过掩码优化来解释预测。为了获得掩码，它随机初始化软掩码，并将其视为可训练的变量。然后通过元素乘法将掩码与原始图结合起来。接下来，通过最大化原始图的预测和新获得的图的预测之间的互信息来优化掩码。Lin等人[70]提出将GNN的决策解释问题转化为一个因果学习任务，并训练了一个因果解释模型，该模型采用基于格兰杰因果关系的损失函数。Gem与现有的GNN解释器不同，它从因果的角度对图结构数据上的GNN进行解释。Gem具有更好的泛化能力，因为它不对GNN的内部结构和图形学习任务的先验知识有要求。通过将因果学习与图神经网络相结合，Gem为GNN的决策提供了更具解释性的视角。Yuan等人[71]探讨了深度图模型的子图级解释，提出了SubgraphX。它利用蒙特卡罗树搜索（Monte Carlo Tree Search, MCTS）算法[72]来有效地探索不同的子图，并从搜索树的叶子节点中选择最重要的子图作为预测的解释。然后，通过节点修剪有效地探索不同的子图，并从搜索树的叶子中选择最重要的子图作为预测的解释。此外，它还采用Shapley值[73]作为MCTS的奖励来衡量子图的重要性，并提出了仅考虑消息传递范围内交互作用的Shapley值的有效逼近。总体而言，SubgraphX利用MCTS算法和Shapley值来生成和评估不同的子图，并选择最重要的子图作为模型预测的解释。通过节点修剪和仅考虑消息传递范围内的交互作用，能够提供更通顺和可解释的子图级解释。

在图像模型中，由于输入空间和输出预测之间的复杂非线性关系，深度模型的解释性常常面临挑战。代理方法是一种流行的方法，用于为图像模型提供实例级解释。其基本思想是使用一个简单且可解释的代理模型来近似复杂深度模型在输入示例相邻区域的预测。这些方法假设输入示例相邻区域中的关系相对简单，可以通过更简单的代理模型有效捕捉。然后，利用可解释的代理模型的解释来解释原始模型的预测。然而，将代理方法应用于图领域是具有挑战性的，因为图数据是离散的，并且包含拓扑信息。因此，在图领域中，研究人员需要探索和发展适用于图数据的代理方法。这可能包括定义图的邻近性，考虑节点之间的关系和拓扑结构，并选择合适的代理模型来近似图神经网络模型的预测。Huang等人[74]针对节点分类任务研究了不同节点特征的重要性。他们提出了一种名为GraphLime的方法，用于解释预测结果。该方法将目标节点及其预测所涉及的N跳相邻节点视为本地数据集，其中N的选择基于训练过的图神经网络模型的层数。然后，他们使用非线性代理模型对本地数据集进行拟合。根据模型中不同特征的权重，选择重要的特征来解释预测结果。这些被选中的特征被视为原始图神经网络模型预测结果的解释。Vu等人[75]构建了一个概率图模型，用于为图神经网络提供实例级解释。该方法通过对节点特征进行随机扰动来获取局部数据集。具体而言，对于给定的输入图，每次模型会随机选择几个节点并对其节点特征进行扰动。然后，针对图中的每个节点，记录一个随机变量，表示该节点的特征是否受到扰动以及对GNN预测的影响程度。通过多次重复这个过程，可以得到一个局部数据集。最后，研究人员使用可解释的贝叶斯网络对局部数据集进行拟合，并解释原始GNN模型的预测结果。这样可以通过分析节点特征的变化和它们对预测的贡献来解释GNN的决策过程。

与实例级解释方法不同，模型级解释方法旨在提供对深度图模型的整体性见解和高层次理解。这些方法研究输入图形模式与特定GNN行为之间的关系，例如如何最大化目标预测。然而，由于图形拓扑信息的离散性，很多应用于图像的方法无法直接应用于图模型，从而使得在模型级别解释GNN变得更具挑战性。Yuan等人[76]提出了XGNN，一种用于解释GNN的图生成方法。在XGNN中，图的生成被看作是一个强化学习问题。生成器在每一步决定如何向当前图添加一条边。然后，将生成的图输入经过训练的GNN中，利用策略梯度来训练生成器，获得反馈信息。此外，XGNN结合了一些图形规则，以鼓励生成的解释既有效又易于理解。Wang等人[77]提出了一种名为GNNInterpreter的方法，旨在解释遵循消息传递方案的不同GNN模型的高层决策过程。GNNInterpreter的主要目标是学习一个概率生成图分布，并通过优化专门为GNN模型级解释设计的新型目标函数来生成最具鉴别力的图样，以反映GNN在进行特定预测任务时的决策过程。

总的来说，近些年来图神经网络的可解释性研究备受关注，并且在各个领域都取得了显著进展。特别是将可解释性技术应用于金融领域，引起了广泛的期待和兴趣。金融领域的决策过程往往牵涉到重要的财务和风险考量，因此对模型的解释和理解具有至关重要的意义。通过图神经网络的可解释性研究，我们可以深入了解金融领域中的各种决策和预测模型是如何运作的，并能够提供详尽的解释和理由。这不仅可以增强模型的可信度，还可以帮助金融从业者更好地理解和应用这些模型。

## 2.3反事实推理

为了解释黑箱模型所做的预测，一种被称为基于特征的解释器的模型通过分析导致结果的特征来提供解释。然而，仅仅揭示预测原因的解释还不足以完全理解特定模型的结果。因此，为了描述输入数据和结果之间的因果关系，解释者必须提供改变哪些输入特征以预测不同结果的示例。换句话说，反事实方法在可解释性中越来越受到重视。反事实推理是一种基本的推理方式，有助于人们深入理解自身行为和揭示世界运作的规则。在推理的过程中，通过考虑与实际情况相反的条件和情境，能够更全面地认识事件或决策的可能性和影响。反事实解释作为一种有效的基于实例的局部解释方法，在特定实例中提供了产生与模型不同结果所需的变化。这种方法的关键在于创建与原始数据集相似但在特征上有微小变化的反事实示例。这样的示例有助于更好地理解模型对于不同输入的响应，为决策提供更全面的考虑。

经典的反事实方法的目标函数可定义为：

其中，𝑥 表示目标实例，𝑥′ 是反事实，𝑦′ 表示期望结果。𝐿 是目标函数，𝑓 代表分类器，𝑑𝑖𝑠 表示距离度量函数，𝜆 是平衡两个距离贡献的超参数。

反事实方法首先被应用于机器学习的可解释性。如[78]所述，反事实可以分为三类：利用模型的内部结构、利用模型的梯度以及直接利用模型进行预测。利用模型的内部结构也被称之为白盒，通常是通过优化器或者是决策树等的已经清楚内部运算过程的模型来生成反事实解释。梯度方法则是通过优化目标函数或变种损失函数来生成反事实解释。直接利用模型预测也被称之为黑盒，通常是通过一些启发式的算法来进行优化。后来反事实也被应用于图神经网络的可解释性。Bajaj等人[79]通过明确模拟 GNN 在相似输入图上的共同决策逻辑，提出了一种生成稳健反事实解释的方法：RCExplainer。该方法通过考虑GNN的共同决策边界，使得生成的解释对噪声具有天然的鲁棒性。Lucic等人[80]提出了一种方法，通过对输入数据（图）进行最小扰动来生成GNN的反事实（CF）解释，他们称之为CF-GNNExplainer。实验表明，CF-GNNExplainer主要删除了对原始预测至关重要的边，以生成最小的CF解释。Tan等人[81]基于因果推理理论的观点，提出了反事实和事实（CF^2）推理的方法，以解决可解释的GNN的学习和评估问题。为了生成解释，他们提出了一个模型无关的框架，该框架利用了这两种推理视角并制定了一个优化问题。这使得CF^2与之前只考虑其中一种推理的可解释GNN方法有所区别。

总体而言，反事实方法在模型可解释性方面具有广泛的应用前景，为深入理解模型决策提供了多层次、多维度的解释途径。反事实方法可以帮助我们发现模型在不同情境下的行为规律，揭示模型的优缺点以及潜在的偏见。特别是在图神经网络等复杂模型中，反事实方法可以有效地生成稳健且具有鲁棒性的解释，有助于提高模型的可信度和泛化能力。值得注意的是，在企业破产风险检测领域，并没有关于图结构数据反事实解释应用的研究。

# 三、面向企业破产风险预测的图神经网络模型的设计与实现

## 3.1问题定义

发广告

## 3.2图神经网络模型

发广告

### 3.2.1异质图模块

### 3.2.2超图模块

## 3.3实验设置与结果分析

### 3.3.1数据集

### 3.3.2实验环境

### 3.3.3评估指标

### 3.3.4实验结果

主实验、节点聚类、消融实验、变体分析、参数敏感性分析

## 3.4本章小结

# 四、面向企业破产风险预测的可解释性图神经网络模型的设计与实现

## 4.1问题描述

## 4.2可解释性图神经网络模型

### 4.2.1边信息处理

### 4.2.2反事实

## 4.3实验设置与结果分析

### 4.3.1数据集

方法

### 4.3.2实验环境

### 4.3.3评估指标

### 4.3.4定量研究

结果、消融实验、参数敏感性分析

### 4.3.5定性研究

画图

## 4.4本章小结

# 五、总结与展望

## 5.1工作总结

## 5.2未来展望

# 参考文献

1. Dvorsky J, Belas J, Gavurova B, et al. Business risk management in the context of small and medium-sized enterprises[J]. Economic Research-Ekonomska Istraživanja, 2021, 34(1): 1690-1708.
2. Tan Y. Financial Risk Management of Small and Medium Sized Enterprises in the Internet Environment[C]//2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City: Volume 2. Springer Singapore, 2022: 215-223.
3. Ptak-Chmielewska A. Bankruptcy prediction of small-and medium-sized enterprises in Poland based on the LDA and SVM methods[J]. Statistics in Transition. New Series, 2021, 22(1): 179-195.
4. Devi S S, Radhika Y. A survey on machine learning and statistical techniques in bankruptcy prediction[J]. International Journal of Machine Learning and Computing, 2018, 8(2): 133-139.
5. Qu Y, Quan P, Lei M, et al. Review of bankruptcy prediction using machine learning and deep learning techniques[J]. Procedia Computer Science, 2019, 162: 895-899.
6. Pena T, Martínez S, Abudu B. Bankruptcy prediction: A comparison of some statistical and machine learning techniques[M]. Springer Berlin Heidelberg, 2011.
7. Shetty S, Musa M, Brédart X. Bankruptcy Prediction Using Machine Learning Techniques[J]. Journal of Risk and Financial Management, 2022, 15(1): 35.
8. Min J H, Jeong C. A binary classification method for bankruptcy prediction[J]. Expert Systems with Applications, 2009, 36(3): 5256-5263.
9. du Jardin P. A two-stage classification technique for bankruptcy prediction[J]. European Journal of Operational Research, 2016, 254(1): 236-252.
10. Barboza F, Kimura H, Altman E. Machine learning models and bankruptcy prediction[J]. Expert Systems with Applications, 2017, 83: 405-417.
11. Bragoli D, Ferretti C, Ganugi P, et al. Machine-learning models for bankruptcy prediction: do industrial variables matter?[J]. Spatial Economic Analysis, 2022, 17(2): 156-177.
12. Feng X, Xiao Z, Zhong B, et al. Dynamic ensemble classification for credit scoring using soft probability[J]. Applied Soft Computing, 2018, 65: 139-151.
13. He H, Zhang W, Zhang S. A novel ensemble method for credit scoring: Adaption of different imbalance ratios[J]. Expert Systems with Applications, 2018, 98: 105-117.
14. Moscatelli M, Parlapiano F, Narizzano S, et al. Corporate default forecasting with machine learning[J]. Expert Systems with Applications, 2020, 161: 113567.
15. Odom M D, Sharda R. A neural network model for bankruptcy prediction[C]//1990 IJCNN International Joint Conference on neural networks. IEEE, 1990: 163-168.
16. Kovacova M, Kliestik T, Valaskova K, et al. Systematic review of variables applied in bankruptcy prediction models of Visegrad group countries[J]. Oeconomia Copernicana, 2019, 10(4): 743-772.
17. Hosaka T. Bankruptcy prediction using imaged financial ratios and convolutional neural networks[J]. Expert systems with applications, 2019, 117: 287-299.
18. Korol T. Dynamic bankruptcy prediction models for European enterprises[J]. Journal of Risk and Financial Management, 2019, 12(4): 185.
19. Kim H, Cho H, Ryu D. Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data[J]. Computational Economics, 2022, 59(3): 1231-1249.
20. 杨钟瑾.粒子群和遗传算法优化支持向量机的破产预测[J].计算机工程与应用,2013,49(18):265-270.
21. 顾啸天. 一种新的基于高斯混合模型的企业破产预测系统[D].上海交通大学,2020.DOI:10.27307/d.cnki.gsjtu.2020.002652.
22. Wang D, Lin J, Cui P, et al. A semi-supervised graph attentive network for financial fraud detection[C]//2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019: 598-607.
23. Hu B, Zhang Z, Zhou J, et al. Loan default analysis with multiplex graph learning[C]//Proceedings of the 29th ACM international conference on information & knowledge management. 2020: 2525-2532.
24. Feng B, Xu H, Xue W, et al. Every Corporation Owns Its Structure: Corporate Credit Rating via Graph Neural Networks[C]//Pattern Recognition and Computer Vision: 5th Chinese Conference, PRCV 2022, Shenzhen, China, November 4–7, 2022, Proceedings, Part I. Cham: Springer International Publishing, 2022: 688-699.
25. Yang S, Zhang Z, Zhou J, et al. Financial risk analysis for SMEs with graph-based supply chain mining[C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021: 4661-4667.
26. Kosasih E E, Brintrup A. A machine learning approach for predicting hidden links in supply chain with graph neural networks[J]. International Journal of Production Research, 2022, 60(17): 5380-5393.
27. Wang D, Zhang Z, Zhou J, et al. Temporal-aware graph neural network for credit risk prediction[C]//Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). Society for Industrial and Applied Mathematics, 2021: 702-710.
28. Zheng Y, Lee V C S, Wu Z, et al. Heterogeneous graph attention network for small and medium-sized enterprises bankruptcy prediction[C]//Advances in Knowledge Discovery and Data Mining: 25th Pacific-Asia Conference, PAKDD 2021, Virtual Event, May 11–14, 2021, Proceedings, Part I. Cham: Springer International Publishing, 2021: 140-151.
29. Tsogoo M. Bankruptcy prediction with explainable machine learning methods[D]. Masarykova univerzita, Ekonomicko-správní fakulta, 2021.
30. Cho S H, Shin K. Feature-Weighted Counterfactual-Based Explanation for Bankruptcy Prediction[J]. Expert Systems with Applications, 2023, 216: 119390.
31. Rao S X, Zhang S, Han Z, et al. xFraud: explainable fraud transaction detection[J]. Proceedings of the VLDB Endowment, 2021 (3): 427-436.
32. Qin Z, Liu Y, He Q, et al. Explainable Graph-based Fraud Detection via Neural Meta-graph Search[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022: 4414-4418.
33. Wu B, Chao K M, Li Y. DualFraud: Dual-Target Fraud Detection and Explanation in Supply Chain Finance Across Heterogeneous Graphs[C]//International Conference on Database Systems for Advanced Applications. Cham: Springer Nature Switzerland, 2023: 370-379.
34. Wu B, Chao K M, Li Y. Heterogeneous graph neural networks for fraud detection and explanation in supply chain finance[J]. Information Systems, 2024, 121: 102335.
35. Abu-El-Haija S, Perozzi B, Al-Rfou R, et al. Watch your step: Learning node embeddings via graph attention[J]. Advances in neural information processing systems, 2018, 31.
36. Cai L, Li J, Wang J, et al. Line graph neural networks for link prediction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 5103-5113.
37. You J, Ying R, Leskovec J. Position-aware graph neural networks[C]//International conference on machine learning. PMLR, 2019: 7134-7143.
38. Luo L, Fang Y, Cao X, et al. Detecting communities from heterogeneous graphs: A context path-based graph neural network model[C]//Proceedings of the 30th ACM international conference on information & knowledge management. 2021: 1170-1180.
39. Chang J, Gao C, Zheng Y, et al. Sequential recommendation with graph neural networks[C]//Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. 2021: 378-387.
40. Guo W, Su R, Tan R, et al. Dual graph enhanced embedding neural network for CTR prediction[C]//Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. 2021: 496-504.
41. Zhang Y, Yu X, Cui Z, et al. Every Document Owns Its Structure: Inductive Text Classification via Graph Neural Networks[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 334-339.
42. Fei Z, Zhang Q, Zhou Y. Iterative GNN-based decoder for question generation[C]//Proceedings of the 2021 conference on empirical methods in natural language processing. 2021: 2573-2582.
43. Zhou S, Zhang J, Zuo W, et al. Cross-scale internal graph neural network for image super-resolution[J]. Advances in neural information processing systems, 2020, 33: 3499-3509.
44. Xie G S, Liu J, Xiong H, et al. Scale-aware graph neural network for few-shot semantic segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 5475-5484.
45. Chen F, Wang Y C, Wang B, et al. Graph representation learning: a survey[J]. APSIPA Transactions on Signal and Information Processing, 2020, 9: e15.
46. Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[C]//International Conference on Learning Representations.2017.
47. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs[J]. Advances in neural information processing systems, 2017, 30.
48. Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[C]//International Conference on Learning Representations.2018.
49. Brody S, Alon U, Yahav E. How Attentive are Graph Attention Networks?[C]//International Conference on Learning Representations.2022.
50. Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15. Springer International Publishing, 2018: 593-607.
51. Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 793-803.
52. Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network[C]//The world wide web conference. 2019: 2022-2032.
53. Yang Y, Guan Z, Li J, et al. Interpretable and efficient heterogeneous graph convolutional network[J]. IEEE Transactions on Knowledge and Data Engineering, 2021.
54. Wang T, Jin D, Wang R, et al. Powerful graph convolutional networks with adaptive propagation mechanism for homophily and heterophily[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(4): 4210-4218.
55. He D, Liang C, Liu H, et al. Block modeling-guided graph convolutional neural networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(4): 4022-4029.
56. Gao Y, Zhang Z, Lin H, et al. Hypergraph learning: Methods and practices[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(5): 2548-2566.
57. Sun X, Yin H, Liu B, et al. Heterogeneous hypergraph embedding for graph classification[C]//Proceedings of the 14th ACM international conference on web search and data mining. 2021: 725-733.
58. Feng Y, You H, Zhang Z, et al. Hypergraph neural networks[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 3558-3565.
59. Gao Y, Feng Y, Ji S, et al. HGNN $^+ $: General Hypergraph Neural Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
60. Ribeiro M T, Singh S, Guestrin C. " Why should i trust you?" Explaining the predictions of any classifier[C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016: 1135-1144.
61. Lundberg S M, Lee S I. A unified approach to interpreting model predictions[J]. Advances in neural information processing systems, 2017, 30
62. 范钰. 基于图网络的可解释性算法研究与分析[D].北京邮电大学,2021.DOI:10.26969/d.cnki.gbydu.2021.000597.
63. 董彬. 图神经网络可解释性的研究与应用[D].电子科技大学,2021.DOI:10.27005/d.cnki.gdzku.2021.001574.
64. Dai E, Zhao T, Zhu H, et al. A comprehensive survey on trustworthy graph neural networks: Privacy, robustness, fairness, and explainability[J]. arXiv preprint arXiv:2204.08570, 2022.
65. Yuan H, Yu H, Gui S, et al. Explainability in graph neural networks: A taxonomic survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
66. Baldassarre F, Azizpour H. Explainability techniques for graph convolutional networks[J]. arXiv preprint arXiv:1905.13686, 2019.
67. Pope P E, Kolouri S, Rostami M, et al. Explainability methods for graph convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 10772-10781.
68. Schnake T, Eberle O, Lederer J, et al. Higher-order explanations of graph neural networks via relevant walks[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(11): 7581-7596.
69. Ying Z, Bourgeois D, You J, et al. Gnnexplainer: Generating explanations for graph neural networks[J]. Advances in neural information processing systems, 2019, 32.
70. Lin W, Lan H, Li B. Generative causal explanations for graph neural networks[C]//International Conference on Machine Learning. PMLR, 2021: 6666-6679.
71. Yuan H, Yu H, Wang J, et al. On explainability of graph neural networks via subgraph explorations[C]//International Conference on Machine Learning. PMLR, 2021: 12241-12252.
72. Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of go without human knowledge[J]. nature, 2017, 550(7676): 354-359.
73. Contributions to the Theory of Games[M]. Princeton University Press, 1953.
74. Huang Q, Yamada M, Tian Y, et al. Graphlime: Local interpretable model explanations for graph neural networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2022.
75. Vu M, Thai M T. Pgm-explainer: Probabilistic graphical model explanations for graph neural networks[J]. Advances in neural information processing systems, 2020, 33: 12225-12235.
76. Yuan H, Tang J, Hu X, et al. Xgnn: Towards model-level explanations of graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 430-438.
77. Wang X, Shen H W. GNNInterpreter: A Probabilistic Generative Model-Level Explanation for Graph Neural Networks[J]. arXiv preprint arXiv:2209.07924, 2022.
78. Verma S, Boonsanong V, Hoang M, et al. Counterfactual explanations and algorithmic recourses for machine learning: A review[J]. arXiv preprint arXiv:2010.10596, 2020.
79. Bajaj M, Chu L, Xue Z Y, et al. Robust counterfactual explanations on graph neural networks[J]. Advances in Neural Information Processing Systems, 2021, 34: 5644-5655.
80. Lucic A, Ter Hoeve M A, Tolomei G, et al. Cf-gnnexplainer: Counterfactual explanations for graph neural networks[C]//International Conference on Artificial Intelligence and Statistics. PMLR, 2022: 4499-4511.
81. Tan J, Geng S, Fu Z, et al. Learning and evaluating graph neural network explanations based on counterfactual and factual reasoning[C]//Proceedings of the ACM Web Conference 2022. 2022: 1018-1027.
82. Chen M, Wei Z, Huang Z, et al. Simple and deep graph convolutional networks[C]//International conference on machine learning. PMLR, 2020: 1725-1735.
83. Zheng X, Liu Y, Pan S, et al. Graph neural networks for graphs with heterophily: A survey[J]. arXiv preprint arXiv:2202.07082, 2022.
84. Zhang S, Zhang J, Song X, et al. PaGE-Link: Path-based graph neural network explanation for heterogeneous link prediction[C]//Proceedings of the ACM Web Conference 2023. 2023: 3784-3793.