

基于异质图神经网络的服装兼容性预测

鲁鸣鸣¹⁾, 郭清明^{1)*}, 张亚^{1,2)}, 易贤康¹⁾

¹⁾ (中南大学计算机学院 长沙 410083)

²⁾ (国网长沙供电公司 长沙 410035)

(guoqmer@csu.edu.cn)

摘 要: 针对目前服装兼容性预测模型效果虚高, 且对服装单品关系的建模没有考虑单品的类型, 以及样本不均衡导致模型训练不充分的问题, 提出一个基于异质图神经网络的服装兼容性预测模型. 首先采用一种负样本采样策略, 使用同类替换的原则重新构造负样本集, 解决了现有模型效果虚高的问题; 然后采用异质图神经网络, 结合图全局池化技术实现了对同一套装内不同类型单品之间复杂关系的建模; 最后通过难例学习, 使得样本量较少的类型单品有更加均衡的训练机会. 在 Polyvore 数据集上的实验结果表明, 所提模型的 AUC 值达到 0.838, 综合兼容性预测性能优于对比方法.

关键词: 服装兼容性; 采样策略; 图神经网络; 时尚推荐

中图分类号: TP391

DOI: 10.3724/SP.J.1089.2024.19974

Heterogeneous Graph Neural Network for Fashion Outfit Compatibility Prediction

Lu Mingming¹⁾, Guo Qingming^{1)*}, Zhang Ya^{1,2)}, and Yi Xiankang¹⁾

¹⁾ (School of Computer Science and Engineering, Central South University, Changsha 410083)

²⁾ (State Grid Changsha Power Supply Company, Changsha 410035)

Abstract: To solve the problem of the inflated effectiveness of current clothing compatibility prediction models, as well as the lack of consideration for the type of clothing when modeling clothing relationships and insufficient model training because of imbalanced samples, we propose a clothing compatibility prediction model based on heterogeneous graph neural networks. The model first adopts a new negative sample sampling strategy, which reconstructs the negative sample set based on the principle of using similar items for replacement to cope with the problem of the inflated model's effectiveness. Then, a heterogeneous graph neural network is used, together with graph global pooling techniques, to model the complex relationships between different types of items within the same outfit. Finally, through hard example learning, clothing items with fewer samples have more balanced training opportunities. Experimental results on the Polyvore dataset show that the model achieves an AUC value of 0.838, with significant advantages in comprehensive performance compared to other methods.

Key words: outfit compatibility; sampling strategy; graph neural networks; fashion recommendation

随着时尚市场的快速发展, 人们对时尚服饰之间相互搭配是否美观(即时尚服饰套装是否兼

容)的需求越来越高, 套装兼容性的相关工作已成为该领域的热点, 具有十分重要的研究意义. 然

收稿日期: 2022-09-19; 修回日期: 2023-03-31. 基金项目: 国家自然科学基金联合基金重点项目(U20A20182). 鲁鸣鸣(1978—), 男, 博士, 副教授, 博士生导师, CCF 会员, 主要研究方向为深度学习、类脑计算、计算机视觉、物联网、计算材料; 郭清明(1998—), 男, 硕士研究生, 论文通信作者, 主要研究方向为深度学习、图像分割、联邦学习; 张亚(1996—), 男, 硕士, 工程师, 主要研究方向为图神经网络、图生成对抗和攻击、交通流预测; 易贤康(1995—), 男, 硕士, 工程师, 主要研究方向为深度学习、图神经网络.

而,现有的套装兼容性建模方法都存在一些问题.已有的套装兼容性预测模型可以分为成对单品的兼容性模型和套装整体的兼容性模型.前者将同一套装中的单品两两配对,并对每对单品之间的兼容性进行建模^[1-4];但这种方式不能全面地反映套装中多个单品之间的复杂关系;后者中早期的工作将套装看成是多个单品组成的序列^[5],通过双向长短期记忆(bi-directional long short-term memory, Bi-LSTM)对该序列进行兼容性预测,但假定套装中单品存在固定的先后顺序的做法忽视了多个单品相互的复杂关系,因此仍然不够全面.

为了解决上述问题, Cui 等^[6]提出节点式图神经网络(node-wise graph neural networks, NGNN)模型,将套装构建成为时尚图,图中每个节点表示一种类型的单品,边表示 2 个不同类型单品之间的相互关系.但该模型中基于随机替换的负样本采样策略会以较大的概率导致负样本套装中出现同类型的单品,而有同类型单品的负样本较容易被模型判定为不兼容,导致兼容性预测的准确率出现虚高的情况.为了解决该问题,本文使用同类替换原则对 NGNN 的负样本采样策略进行改进,确保负样本套装不会重复出现同类型单品.此外, NGNN 构造的时尚图虽然使用有向边表示不同类型单品之间的连接,但没有对这些有向边的类型进行区分,无法对不同类型(如“上衣”“外套”“裤子”等)的单品之间的连接进行区别对待,无论是上衣与帽子之间的连接,还是上衣与上衣之间的连接,都将其看作是同一种类型的连接.为了解决这一问题,本文使用一个有向带权的异质图^[7]表示套装,将单品之间的连接根据其类型进行区别对待,实现时尚服装兼容性的细粒度建模.最后针对数据集中不同类型单品之间的数量分布不均而导致模型性能不佳的问题,通过难例学习对数据集中类别不平衡的样本进行调整,使模型对所有类型的样本都可以进行均衡的训练.

基于以上思路,本文提出基于异质图神经网络(relational graph convolutional networks, RGCN)^[8]的服装兼容性预测模型.为了评估本文模型的性能,将其与目前普遍使用的时装兼容性预测模型进行比较,并在 Polyvore 数据集上进行了广泛的实验验证.实验结果表明,本文模型的性能比目前性能最好的模型提升 4.5%.

1 相关工作

计算机视觉领域中,套装兼容性指将套装中

不同类型的服饰单品作为一个整体进行和谐的、美的搭配.服装的兼容性预测则是判定套装中的所有单品的组合是否符合上述服装兼容性的要求.

早期,成对服装的兼容性工作是将服装映射到一个潜在的兼容空间,通过估计服装的风格向量之间的距离刻画服装的兼容性. Chen 等^[9]引入一个混合类别度量学习模型,将成对的服装映射到相同的嵌入空间中,通过度量学习使得在该嵌入空间中兼容性高的服装单品彼此接近,而兼容性低的服装单品相互远离.然而,将相似的服装单品映射到同一个嵌入空间,使得在嵌入空间中的服装单品距离都比较接近,会导致服装单品之间可能出现间接的兼容性,违背了兼容性不具有传递性这一事实.例如,在嵌入空间的相邻区域,如果一双鞋与一顶帽子相匹配,而这顶帽子又与一件上衣相匹配,那么一个自然的后果是鞋也被迫与上衣相匹配,但事实却并不一定如此.因此, Vassileva 等^[10]在单一嵌入空间的基础上,将服装单品根据其类别信息的不同映射到不同的子嵌入空间,进而改进服装单品的嵌入表示.

从领域知识的角度看,大多数的兼容性建模都依赖于数据驱动的方法,忽视了时尚领域的知识.为了解决这一不足, Song 等^[11]应用知识蒸馏的教师-学生网络架构,其中,教师网络通过蒸馏损失将时尚领域的知识传递给学生网络,指导学生网络训练,提升其兼容性模型的性能; Yin 等^[12]使用外部专家时尚收藏数据集,并将视觉兼容关系和风格信息纳入时尚兼容学习中.然而,这种将时尚套装的兼容性单独表示为成对服装之间兼容性的做法,并不能反映多个服装单品之间的间接关系.

进一步,研究人员将整套服装视为具有固定顺序的序列对服装的兼容性进行建模. Li 等^[13]利用循环神经网络预测时尚套装的整体兼容性; Han 等^[5]和 Pang 等^[14]引入 Bi-LSTM 计算套装整体的兼容性分数,这 2 个工作都是将时尚套装中的服装单品排列成一个事先确定好的次序,作为时序序列数据输入循环神经网络,并以此对整个套装的兼容性进行建模.然而,由于时尚套装中的单品实际上不存在固定的先后顺序,同时该模型也不能反映任意 2 件服装单品之间的直接交互关系,因此研究人员将一套服装中的服装单品视为一个集合,并使用图来对服装的兼容性进行建模.即将一套服装表示为一个时装图来模拟服装的兼容性,其中,每个节点代表一个服装单品类别,每条边代表 2 个类别之间的相互关系. Li 等^[15]使用一个 3 层的

分层图神经网络对服装的兼容性进行建模,并对用户的服装进行个性化搭配,但由于样本采样策略及节点与边类型单一等原因,这类方法的鲁棒性不强。

现有的成对服装的兼容性模型或套装的兼容性模型都存在缺陷,针对这些服装兼容性建模方法的特点及存在的问题,本文提出基于 RGCN 的服装兼容性预测模型。采用同类替换原则的负样本采样策略,解决了负样本套装中同类型单品导致模型性能虚高的问题;使用基于 RGCN 的兼容性建模方法,能够更细致地建模不同类型节点之间的交互关系;通过难例学习,使得不同类型的数据样本获得更加均衡的训练机会。最后在 Polyvore 数据集上进行了充分的实验。实验结果表明,本文模型比其他方法具有明显优势。

2 本文模型

2.1 负样本采样策略

由于 NGNN 使用随机替换策略构造的负样本较容易被模型识别,因此导致模型性能的虚高。下面详细阐述其产生的问题,并提出一种解决该问题的负样本构造方法。

2.1.1 NGNN 数据集简介

NGNN 使用的时尚服装数据集来自时尚网站 Polyvore.com 提供的服装数据。时尚造型师可以在该网站上向公众分享他们设计的服装,网站也允许用户自行创建时尚套装,每套时尚套装中包含丰富的多模态信息,如服装单品的图片、文本描述、标签、流行度评分和类型等信息。在 NGNN 使用的数据集中,共有 21 889 套服装,其中,训练集 17 316 套,验证集 1 497 套,测试集 3 076 套。

2.1.2 NGNN 的负样本采样策略

因为 Polyvore 数据集中只有正样本,而模型的训练需要负样本,因此 NGNN 采用随机替换的原则,基于正样本来构造负样本。随机替换是选

定一个正样本套装后,替换掉套装中的一件单品(如鞋子),并从所有类型单品中随机采样一件单品进行替代。这种随机替换的方法有较大的概率出现同类型单品重复出现在同一负样本套装中的情况。由于这些出现同类单品重复的负样本较容易被 NGNN 模型识别出来,因此难以验证 NGNN 模型辨别时尚套装兼容性的真实能力,会导致模型预测效果的虚高,本文将通过实验验证这一点。

2.1.3 本文改进的负样本采样策略

为了解决 NGNN 模型负样本采样策略带来的问题,本文改进了该模型的负样本采样策略,使用同类替换的原则重新构造负样本集,使得每个负样本套装中不存在类型相同的 2 个单品,构造出更合理的负样本数据。同类替换指在构造负样本时,从一个套装内选定一个需要被替换的单品(如鞋子)后,不是随机地从所有类型中选择单品替换,而是从数据集中选择同一类型的单品(仍然是鞋子)进行替换。通过同类型的单品替换,可以确保同一套装中不会出现 2 种同类型单品,以确保负样本的合理性。

在构造负样本时,为了避免出现正负样本不均衡的问题,本文用同类替换原则构造的负样本在数量上与正样本相当,即负样本占比为 50%。

2.2 本文模型建模

已有的套装兼容性建模方法,无论是成对单品之间的兼容性进行建模,还是将套装看成序列进行建模,或是将套装建模成一个带有节点和边的图,都没有将服装之间的连接根据服装类型的不同进行区别对待,因此无法更具细粒度地对时尚套装的兼容性进行建模。为此,本文使用一个有向带权的异质图建模时尚套装,同时结合图全局池化技术提出基于 RGCN 的服装兼容性预测模型,其架构如图 1 所示。首先使用预训练模型将服装单品图片转化成对应的特征向量,然后构造由服装单品特征向量组成的图,并将其输入 RGCN 中得到套装的兼容性分数。

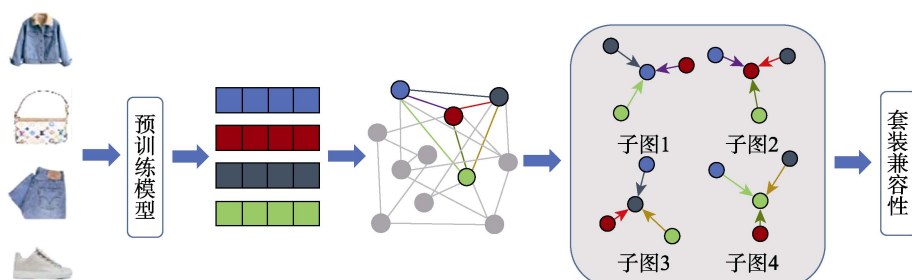


图1 本文模型架构

不失一般性, 本文用集合 $S = \{s_1, s_2, s_3, \dots\}$ 表示训练集, 大小为 $|S|$; 其中, 每个元素 s_k ($k \in [1, |S|]$) 表示一套时尚套装, 该套装由 $|s_k|$ 个服装单品组成. $s_k = \{v_1, v_2, v_3, \dots\}$, 其中的每个元素 v_i ($i \in [1, |s_k|]$) 表示该套装中的一个单品.

2.2.1 构造异质图

现有的套装兼容性建模方法没有根据单品类型对单品之间的连接进行区分, 导致无法实现更细粒度的兼容性建模. 为此, 本文以异质图的形式对套装进行建模, 并根据数据集中所有可能的单品类型构造了一个服饰异质图, 该图的子图表示具体的一套时尚服饰, 每个节点代表一个对应类型的单品. 如果 2 个节点对应的单品类型实例在训练数据集中共同出现在同一个套装中, 则在服饰异质图中对应的 2 个节点之间有 2 条反向的有向边. 之所以把边定义成有向的, 是因为有向边在 2 个方向上具有不同的意义. 如根据上衣或外套来搭配合适的配饰, 与根据配饰来搭配上衣或外套的含义是不相同的. 从模型的角度出发, 当使用 RGCN 进行节点间的信息传播时, 是通过节点之间的连接关系(即边)进行信息传播的. 如果节点 v_i 与 v_j 之间通过 2 条反向的有向边进行连接, 则从 v_i 传播到 v_j 的信息代表 v_i 对 v_j 的重要程度, 从 v_j 传播到 v_i 的信息代表 v_j 对 v_i 的重要程度; 同时, 基于训练用到的数据集, 本文为每条有向边赋予特定的权重. v_i 到 v_j 的有向边对应的权重计算公式为

$$\omega(i, j) = C_{i,j} / C_i.$$

其中, $\omega(i, j)$ 表示 v_i 到 v_j 的有向边的权值; $C_{i,j}$ 表示 v_i 与 v_j 对应单品类型在训练集的同一套装中共同出现的次数; C_i 表示 v_i 对应的单品类型在训练集中出现的次数.

本文还对服饰异质图中边的类型进行分类, 根据节点所属单品类型的不同, 节点之间的有向边所属的类型也不同. 这样, RGCN 就能根据边的不同类型, 基于消息传递机制学习特定于类型的节点表征.

2.2.2 套装兼容性建模

基于 RGCN 的套装兼容性预测模型的实施步骤中, 包括节点属性的初始化、基于 RGCN 的节点表达计算和图全局池化.

为了得到节点的初始特征, 利用预训练模型

ResNet18^[16]对输入的服装图像进行特征提取. 因为只需要利用预训练好的 ResNet18 提取服装图像的特征, 就可以去掉最后一层中用于分类的 Softmax, 并提取该层全连接层神经元的值. 这样, 每幅服装单品的图像都可以经过预训练模型, 用一个 d 维的向量表示.

实际上, 输入模型的每套服饰对应服饰异质图中的一个子图. 如图 2 所示, 模型在训练时, 将含有 4 件单品的套装映射到时尚异质图中的 4 个节点, 构成表示该套装的子图; 该子图经过激活后, 输入 RGCN 中进行服装的兼容性预测. 模型在每次读取一套服装进行兼容性预测时, 只有该套装中包含的单品所对应类型的节点处于激活状态, 其余节点处于非激活状态. 图 2 中, 灰色节点表示处于非激活状态的节点, 其余 4 个彩色节点表示处于激活状态的节点. 处于激活状态的节点的特征是非 0 的, 而处于非激活状态的节点对应特征为全 0 状态. 尽管非激活状态的节点状态为全 0, 但其在图中的结构信息和对应边的类型信息仍得以保留. 图 2 中, 任意 2 个节点 v_i 与 v_j 之间的关系可表示为 $\{(v_i, v_j), \omega(i, j) | v_i, v_j \in s_k\}$.

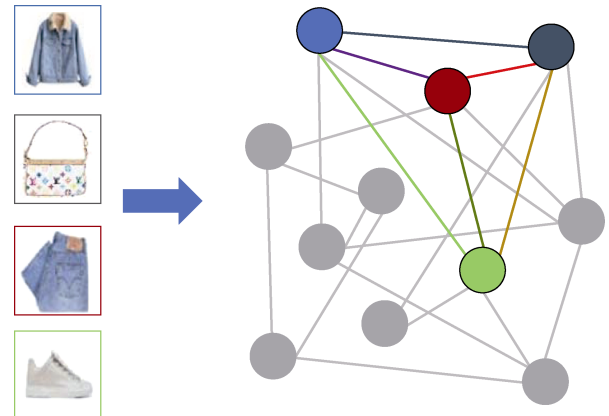


图 2 节点激活过程示意图

通过预训练模型 ResNet18 得到每幅单品图像的特征向量, 以及服饰异质图整体与每套套装所对应的激活子图后, 即可采用 RGCN, 通过挖掘邻居节点中有价值的信息学习每个节点的向量表达. RGCN 中, 节点在第 l 层的更新公式为

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{C_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)} \right) \quad (1)$$

其中, r 表示边的类型, 而 \mathcal{R} 表示边类型的集合; $j \in \mathcal{N}_i^r$ 表示边类型为 r 的前提下, 节点 v_i 的邻居节点的集合; $W_r^{(l)}$ 和 $W_0^{(l)}$ 表示神经网络需要学习的

参数; $h_i^{(l)}$ 表示神经网络中节点 v_i 在第 l 层的隐藏状态; $c_{i,r}$ 表示一个提前定义好或训练好的正则化参数. 直观上, 式(1)将变换后的相邻节点特征向量进行累加和归一化, 与常规的图卷积神经网络(graph convolutional network, GCN)^[17]不同, 它引入特定于关系的转换, 即依赖边的类型和方向, 由此可以得到每个节点汇集了邻居节点的信息之后的向量表达.

2.2.3 图全局池化

在由 RGCN 得到每个服装单品节点的向量表达的基础上, 可通过池化技术得到整个套装图的向量表达, 并根据此向量预测最终的套装兼容性. 本文采用图全局池化技术, 将所有节点向量汇聚成一个套装图的向量.

经过 RGCN 的信息传播, 可以得到一个大小为 $t \times d$ 的节点向量表达, 其中, t 表示节点数, d 表示每个向量的长度. 通过图全局池化技术, 可将该向量矩阵压缩为一个 d 维向量, 表示该套装图.

图全局池化分为最大池化、求和池化和平均池化. 图全局最大池化通过比较 t 个向量中的元素在每个位置上的最大值, 取该值作为最后套装图的向量表达; 图全局求和池化对 t 个节点向量按位置进行求和, 最后套装图的向量表达中, 每个位置的值就是 t 个向量在对应位置的值之和; 同理, 在图全局平均池化中, 池化后套装图的向量表达中, 每个位置的值就是 t 个向量在对应位置的值的均值. 后续的实验结果表明, 使用全局求和池化的实验效果是最好的.

2.3 难例学习

上述操作构造了服饰异质图, 使用 ResNet18 为预训练模型得到服装图像的特征表达, 将其作为服饰异质图的节点特征, 并利用 RGCN 结合图全局池化技术, 完成服饰套装兼容性预测模型的搭建. 然而, 如图 3 所示, 经过对数据集的统计发现, 不同类型的单品在数量上的分布严重不均衡. 可以看出, 数量最多的类别有 10 430 个, 最少的类别数量仅有 4 个, 这种样本数量的不平衡导致训练模型侧重样本数目较多的类别, 而“轻视”样本数目较少类别, 影响模型在测试数据上的泛化能力, 进而影响模型的性能.

本文使用难例采样的数据增强的方法缓解类别数量不均的问题. 当模型每次读取一套套装时, 不会将套装中所有单品作为输入, 而是只采样部分服装单品, 同时对数量不同的服装单品设置不

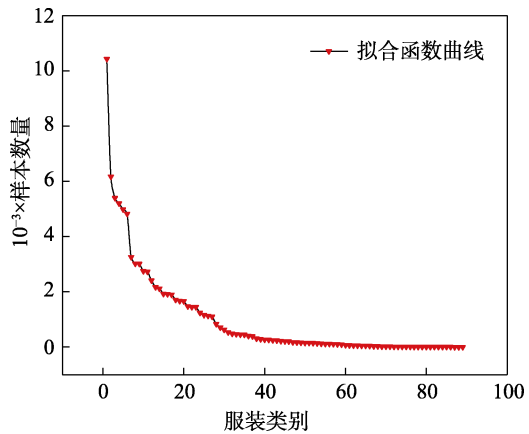


图3 数据集中各服装类别数量分布图

同的采样率. 本文考虑在模型训练阶段使用难例采样的方法, 对模型的输入数据做相应的预处理. 从输入的一个套装中采样一部分单品构成一个新的套装, 并将这个新的套装输入模型进行兼容性判断. 如果输入的套装为兼容的套装, 那么采样之后构成的新套装样本也被认为是兼容的.

在单品采样之前, 难例采样的方法首先统计每种类型的服装在数据集中出现的次数, 然后计算每个类型的出现频率. 每件单品的采样概率与单品所属类型的出现频率成反比, 频率越低, 该单品的采样率就越高. 通过这种方法, 可以使得数量分布不均类型的样本的训练更加均衡.

本文还考虑了另外 2 种形式的采样方式: 重要性采样和均匀采样, 并在实验中与难例采样进行对比. 在重要性采样中, 每件单品的采样概率由单品所属类别的出现频率决定, 频率越高, 该单品的采样概率越高; 均匀采样则是套装中所有单品的采样概率保持一致.

3 实验及结果分析

3.1 数据集简介与数据预处理

本文使用的数据集基于 Vasileva 等^[10]从著名时尚网站 Polyvore 采集到的服装数据(采集的数据已经进行了数据清洗和筛选等预处理操作), 使用其中的单品图片和类别信息.

原始的 Polyvore 数据集中共有 223 个服装类别, 由于类别数量的过多和过于细化, 导致某些类别的服装数量特别稀少, 而且存在部分类别与类别名不能一一对应的问题. 为此, 本文将某些特别细化的类别进行合并, 并且对类别与类别名没有一一对应的类别进行统一. 表 1 所示为数据集中部分类别重命名的案例, 其中, 前 3 行所示为类别名

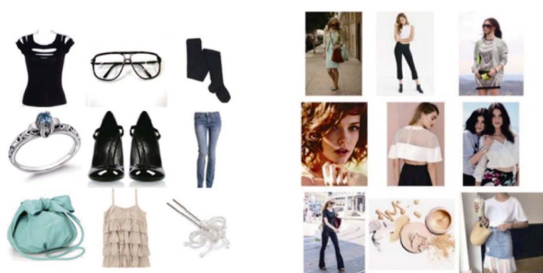
与类别不一一对应的案例, 它们的类别分别为 3, 4 和 5, 小类名都是 dress, 此时可以对这 3 个类别进行合并, 将它们的类别都重命名为 0; 最后 5 行所示为将过于细分的类别进行合并的案例, 他们对应的类别都为 11, 而它们的小类名却细分为 sweater, sleeveless top, shirt 和 top, 过于细化, 本文将其小类名统一为 tops_Comprehensive. 经过对服装小类的合并和统一, 最后共得到 89 种服装小类.

表 1 数据集部分类别重命名案例

类别	小类名	大类名	重命名类别	重命名小类名
3	dress	all-body	0	dress
4	dress	all-body	0	dress
5	dress	all-body	0	dress
6	gown	all-body	1	gown
7	skirt	bottoms	7	skirt
8	skirt	bottoms	7	skirt
9	skirt	bottoms	7	skirt
9	skirt	bottoms	7	skirt
9	skirt	bottoms	7	skirt
10	long skirt	bottoms	8	long skirt
11	sweater	tops	20	tops_Comprehensive
11	sleeveless top	tops	20	tops_Comprehensive
11	shirt	tops	20	tops_Comprehensive
11	sleeveless top	tops	20	tops_Comprehensive
11	top	tops	20	tops_Comprehensive

3.2 实验设置

对服装图片进行统一筛选, 剔除掉不合格的服装图片. 标准的服装图片只包含服装本身, 无任何其余物品的信息, 并且以白板为背景(如图 4a 所示). 本文对那些包含人物全身照、背景为风景照而不是白板, 或者根本不是服装内容的图片(如图 4b 所示)进行剔除.



a. 标准服装图片

b. 不合格服装图片

图 4 服装图片对比

经过对数据集中服装类别的统一和合并, 以及服饰图片的清洗, 最后生成的服饰数据集中训练集有 33 990 套套装, 测试集有 30 290 套套装.

(1) 评价指标

本文的任务是预测任何给定的时尚套装的兼容性, 这是一个二分类任务. 为了评估模型的性能, 采用广泛使用的度量指标曲线下面积(area under curve, AUC), 计算公式为

$$AUC = \frac{1}{|D_{\text{test}}|} \sum_{(s, s^{-1}) \in D_{\text{test}}} \delta(x_s > x_{s^{-1}}).$$

其中, D_{test} 表示测试集; s 和 s^{-1} 分别代表正样本和负样本; x_s 和 $x_{s^{-1}}$ 则分别表示模型对正样本和负样本预测为正样本的分数; $\delta(\cdot)$ 是一个指示函数, 若括号内的内容为真, 则输出 1, 反之则输出 0.

(2) 基线模型

为了评估本文模型, 将其与下列有代表性的时装兼容性预测模型进行比较:

RAND. 为套装兼容性分配一个随机数值.

SiameseNet^[11]. 将一对服装单品输入孪生卷积网络中, 并将相似度的平均值作为套装整体的兼容性得分.

CSN^[9]. 先把所有的服装嵌入同一个向量空间中, 然后将共有向量空间映射到根据类别分割的子空间中, 得到成对类别的服装向量之间的相似度.

FHN^[18]. 首先使用卷积神经网络提取服装特征, 然后通过 2 层全连接网络和 1 个符号函数构建哈希模块, 最后通过打分函数计算 2 个服装间的兼容性得分.

Bi-LSTM^[5]. 将套装看成一个具有固定次序的序列, 并将其输入 Bi-LSTM 预测时尚套装的兼容性分数.

HFGN^[14]. 使用一个 3 层的分层图神经网络对服装的兼容性进行建模, 并对用户的服装进行个性化搭配, 其中, 从底层到高层的节点分别表示服装单品、服装套装和用户.

NGNN^[6]. 将时尚服装构建成一个时尚图, 进而建模套装整体的兼容性.

其中, SiameseNet, CSN 和 FHN 关注的是成对的服装单品之间的兼容性, Bi-LSTM 是将时尚套装当成一个序列来对套装的兼容性进行建模, 而 HFGN 和 NGNN 是将套装建模成一个图来预测套装的兼容性.

(3) 模型设置

以深度学习框架 PyTorch^[19]和 GNN 扩展库 PYG^[20]为实验平台, 实现了本文模型. 参数设置如下: 学习率为 0.001, 批次大小为 32, 权重衰减系数为 0.005; 网络的最后一层采用 Sigmoid 函数, 并使用均方差(mean square error, MSE)作为损失函数.

3.3 负样本采样策略验证

NGNN 的负样本采样策略使得构造的负样本中可能出现多个同类型的单品,这种负样本极易被模型判别为不兼容,因而不合理的.本文采用同类替换方式,在原数据集和重新构造的数据集(简称新数据)上进行实验,结果如表2所示.其中,AUC 用于衡量模型的整体性能; R 表示实际为正的样本中预测为正的样本所占的比例; P 表示模型预测为正的样本中实际为正的样本所占的比例;3个指标的取值范围均为[0,1],值越大,表示模型的性能越好.可以看出,当改变负样本的采样策略之后,NGNN 的 R 指标基本不变,AUC 与 P 指标均出现明显下降,且 P 下降得尤为明显.

表2 不同采样策略下 NGNN 模型效果对比

数据集	AUC	R	P
原数据集	0.931	0.962	0.860
新数据集	0.807	0.960	0.660

为了探究 NGNN 模型性能下降的原因,对评价指标在负样本采样策略改变前后的变化进行分析.兼容性预测是二分类任务,可用1代表正例,0代表负例; R 和 P 定义所需的真正例(true positive, TP)、假正例(false positive, FP)、假负例(false negative, FN)和真负例(true negative, TN)与真实类别和预测类别之间的对应关系如表3所示.

表3 正负例对比

预测类别	真实类别	
	1	0
1	TP	FP
0	FN	TN

根据表3的定义, R 和 P 分别定义为

$$R = \frac{TP}{TP + FN};$$

$$P = \frac{TP}{TP + FP}.$$

经过分析,可以得出如下结论:因为本文采用同类替换的负样本采样策略仅对负样本的预测产生影响,即采用同类替换原则的采样得到的负样本比随机替换的负样本更容易被预测为正样本,因此 FP 会增大;而正样本的预测情况并没有发生改变,即 TP 和 FN 保持不变.所以 R 基本保持不变,而 P 则会降低.

在二分类任务中,分类器可以得到将一个样本预测为正样本的概率,此概率称为这个样本的得分.指标 AUC 的作用可以描述为:随机给定一

个正样本和一个负样本,用一个分类器进行分类和预测,该正样本的得分的概率比该负样本得分的概率高.然而,由于同类替换的负样本采样策略导致构造的负样本更易被分类器预测为正样本,因此 AUC 会降低.

3.4 兼容性预测性能评估

在处理后的 Polyvore 数据集上,评估本文模型和7个基线模型对套装兼容性预测的性能,实验结果如表4所示.

表4 8个模型在兼容性预测任务上的 AUC 对比

模型	AUC
RAND	0.503
SiameseNet ^[1]	0.532
Bi-LSTM ^[5]	0.629
CSN ^[9]	0.759
FHN ^[18]	0.781
HFGN ^[14]	0.785
NGNN ^[6]	0.793
本文	0.838

注.加粗表示最优值.

从表4可以得到以下结论:

(1) 与其他模型相比, SiameseNet 的性能最差,仅好于 RAND 的随机打分,说明由于忽视了套装的整体性,仅用单品对的兼容性取平均值的做法不足以学习到套装整体的兼容性特征.

(2) Bi-LSTM 模型的性能比 SiameseNet 的性能要好,说明前者可以更好地学习到有关套装兼容性的潜在知识.然而, Bi-LSTM 将整个套装看成一个固定的序列,虽然这种方式可以实现对时尚套装的整体兼容性进行建模,但会漏掉时尚套装中不同单品之间的直接关系.

(3) CSN 与 FHN 的性能比前2个模型好得多,尽管它们也是通过衡量成对服装之间的兼容性来预测套装整体的兼容性,但其中引入了服装的类别信息,说明类别信息在建模时尚套装兼容性任务中的重要性.但是, CSN 过于重视服装颜色对套装兼容性的影响, FHN 只考虑了3个类别标签,导致无法建模更复杂的兼容性关系.

(4) 基于图的套装兼容性建模 HFGN 和 NGNN 比前面的模型取得了更好的性能,说明与成对建模和序列化建模相比,图结构可以更好地建模服装之间复杂的交互,方便进一步推断套装的兼容性信息.

(5) 本文模型取得了最好的性能,除了受益于图结构的建模方法外,与 HFGN 和 NGNN 相比,本文将套装表示成异质图,对不同类型节点之间

的边的类型进行区分,并通过引入难例采样的方法使得模型能更好地建模套装的兼容性.

3.5 单品推荐

为了验证本文模型在 FITB(fill in the blank)任务中的有效性,将其与基线模型在 FITB 任务中的结果进行比较. FITB 任务中,给定一套不完整的套装,目标是从候选的单品集中找到与不完整套装最兼容的一件单品. 表 5 和图 5 分别从定量和定性的角度展示 FITB 任务的结果. 从表 5 可以看出,本文模型在准确率上有一定的优势. 从图 5 可以看出,图 5a 中,本文模型、NGNN 和 FHN 这 3 个模型都判断出此套衣服缺少一双鞋,并选择了正确的答案;图 5b 中, FHN 虽然正确地选择了与原套装互补的单品类别(鞋子),但是由于该模型只考虑 3 个比较粗粒度的单品类型,无法利用更加细粒度的类型信息对套装兼容性进行判断,因此没有选出正确的答案;图 5c 中,虽然本文模型与 NGNN 都推断出原套装缺少一件下装,但本文模型选择的下装显然比 NGNN 选择的下装更能与原套装形成合理的搭配. 实验结果证明了本文模型在兼容性匹配任务上的合理性和有效性.

表 5 8 个模型在 FITB 任务上的准确率

模型	准确率
RAND	0.341
SiameseNet ^[1]	0.563
Bi-LSTM ^[5]	0.708
CSN ^[9]	0.769
FHN ^[18]	0.751
HFGN ^[14]	0.804
NGNN ^[6]	0.832
本文	0.870

注. 加粗表示最优值.



图 5 3 个模型在 FITB 任务上的结果比较

3.6 服饰搭配性能评估

为了对实验结果有更直观的认识,本文提取一些套装实例,包括 3 套兼容的套装(Ground truth 为 1)和 2 套不兼容的套装(Ground truth 为 0),5 个模型对这些套装预测的兼容性分数如表 6 所示. 其中,兼容性分数的分布在 0~1,大于等于 0.5 为兼容套装,小于 0.5 为不兼容套装;分数越大,说明模型判断它属于兼容套装的可能性越高. 可以看出: (1) 第 1 套套装中,本文模型取得了较高的兼容性得分,CSN 与 FHN 输出的分数比本文模型略低; NGNN 的分数最高,说明该模型较好地捕捉到

表 6 5 个模型在不同套装上兼容性得分比较

套装编号	套装	Ground truth	模型				
			本文	CSN ^[9]	FHN ^[18]	Bi-LSTM ^[5]	NGNN ^[6]
1		1	0.84	0.83	0.87	0.39	0.92
2		1	0.93	0.45	0.93	0.42	0.90
3		1	0.96	0.86	0.86	0.74	0.85
4		0	0.28	0.44	0.35	0.68	0.61
5		0	0.13	0.32	0.42	0.46	0.40

了这类服装的兼容性特征; Bi-LSTM 对这套服装的兼容性打分为 0.39, 说明该模型将它判定为不兼容套装, 这可能是因为其采用的序列表示方法没有捕捉到第 1 件衣服与最后一个包之间的兼容性. (2) 第 2 套套装中, Bi-LSTM 同样出现了误判的情况; CSN 给出了 0.45 的兼容性得分, 可能是在该模型的衣服-帽子或衣服-鞋子类别子空间中, 不同颜色服装之间的兼容性偏低, 其确实倾向于将同一种色系的服装判别为兼容套装, 而对于不同颜色的服装, 很有可能判定为不兼容. (3) 第 3 套套装中, 所有模型都有良好的表现, 本文模型给出了最高的兼容性分数. (4) 第 4 套套装中, 本文模型、CSN 和 FHN 均准确地判断出该套装不兼容, 而 Bi-LSTM 和 NGNN 出现判断错误, 说明 Bi-LSTM 和 NGNN 并没有充分地捕捉到这种风格的服装的兼容性特征. (5) 第 5 套套装中, 所有模型给出的兼容性分数都小于 0.5, 本文模型给出了最低的 0.13 分.

3.7 参数敏感性分析

进一步, 在保持其他模块的参数设置不变的情况下进行消融实验, 评估不同参数与实验设置对模型性能的影响.

3.7.1 二分类任务设置对模型性能的影响

兼容性预测本质上可以看作二分类任务, 一般有 2 种做法: (1) 将神经网络的输出通道设为 1, 因为神经网络最后得到的输出可能是任意的数, 所以首先让输出经过一个 Sigmoid 函数, 使其数值归一化到[0,1]; 然后将这个值和标签放入损失函数进行计算, 本文使用二元交叉熵损失(binary cross entropy, BCE)作为损失函数. (2) 将神经网络的输出通道设为 2, 分别表示兼容与不兼容, 首先让输出的 2 个数经过一个 Softmax 函数将其数值归一化到[0,1], 2 个值加起来等于 1; 然后使用 BCE 作为损失函数, 取值更大的那个数对应的通道就是模型预测的结果.

如果把兼容性预测看成一个回归任务来处理, 那么神经网络在最后一层通过 Sigmoid 函数得到一个归一化的值后, 可以用 MSE 作为损失函数来训练模型.

表 7 所示为 3 种不同设置下模型的结果. 可以看出, 使用 Sigmoid 函数结合 MSE 时, 模型取得了最佳的性能. 因此, 本文实验中都是采用 Sigmoid 函数结合 MSE 的方式.

3.7.2 不同嵌入维度对模型性能的影响

本文使用 RGCN 将服装嵌入到不同的维度,

以此判断不同的嵌入维度对模型预测套装兼容性的影响, 结果如表 8 所示. 可以看出, 在维度为 16 时性能达到峰值, 太大或太小的嵌入尺寸都会影响模型的性能. 为了实现更好的模型性能, 本文将嵌入维度设置为 16.

表 7 二分类任务不同设置对模型性能的影响

设置方法	AUC
Sigmoid+BCE	0.821
Softmax+BCE	0.802
Sigmoid+MSE	0.838

表 8 不同嵌入维度对模型性能的影响

嵌入维度	AUC
1	0.759
2	0.792
4	0.813
8	0.820
16	0.838
32	0.821
64	0.818
128	0.823

3.7.3 不同图全局池化方法对模型性能的影响

为了从所有节点的向量表达得到整个图的向量表达, 采用作用于图神经网络的全局池化技术. 全局池化技术分为全局最大池化、全局求和池化和全局平均池化 3 种, 分别是对图中所有节点的向量表达进行逐个元素的求最大值、求和以及求均值操作. 3 种全局池化技术对模型性能的影响如表 9 所示. 可以看出, 使用全局求和池化能达到最佳的性能, 因此本文在最终实验中采用该方法进行兼容性预测.

表 9 不同图全局池化方法下模型的性能表现

设置方法	AUC
全局最大池化	0.825
全局求和池化	0.838
全局平均池化	0.821

3.7.4 不同预训练模型对模型性能的影响

为了找到最合适数据集的预训练模型, 使用 ResNet18^[16], ResNet50^[16]和 RexnetV1^[21]这 3 种卷积神经网络. 其中, ResNet18 和 ResNet50 是 He 等^[16]于 2016 年提出的残差网络模型, RexnetV1 是 Han 等^[21]于 2021 年提出的轻量级卷积网络, 旨在消除表达瓶颈. 使用不同的预训练模型得到服装图片的特征后, 对本文模型性能的影响如表 10 所示.

可以看出,对于本文使用的数据集,ResNet18 表现的实验效果好于其他预训练模型。

表 10 不同预训练模型下模型的性能表现

预训练模型	AUC
ResNet18 ^[16]	0.838
ResNet50 ^[16]	0.805
RexnetV1 ^[21]	0.761

由于各服装类型的数量分布严重不均衡,因此本文使用难例采样的数据增强方法缓解类别比例不均衡的问题。对比难例采样、重要性采样和均匀采样 3 种策略下模型的性能,结果如表 11 所示。可以看出,实施重要性采样的效果是最差的,实施难例采样的效果最好,原因是当采用重要性采样策略时,出现频率高的类别的单品将有更多的被采样的机会,采样策略是在模型的训练阶段进行的步骤,那些出现频率高的服装类别对应的节点本就能得到更多的训练机会,而恰恰是那些出现频率少的节点需要更多的训练机会。因此本文采用难例采样策略,使得出现频率不同的节点得到更均衡的训练机会,最后的效果自然比重要性采样策略好。

表 11 不同采样策略下模型的性能表现

采样策略	AUC
重要性采样	0.758
均匀采样	0.791
难例采样	0.838

3.8 负样本合理性及配饰兼容性分析

(1) 由于人的服饰风格具有多样性,不一定构造的负样本就是不合理的,因此为了验证本文负样本的合理性,本文增加一个用户调查。首先随机地从数据集中选取 20 个正样本,然后用本文提出的同类替换原则构造相应的负样本,并要求用户对这 40 个样本进行兼容性打分;为了避免用户主观印象对样本兼容性判断的影响,在用户进行打分时打乱这些负样本和正样本,并将所有样本经过模型推理后的 Softmax 值作为用户评分的参考,最后调查 10 个用户对样本的兼容性打分。经过数据统计分析发现,由于每个人的服饰风格的不同,确实存在构造的负样本也是合理的情况,即用户对负样本的兼容性打分高于相应的正样本;但是这种情况相对较少,大约只占 10%。因此,从概率的角度考虑,可以认为本文构造负样本的方法是

有意义的。

(2) 为了进一步解释 RGCN 的有效性和除衣服外的饰品与整体服装风格的兼容性,对套装内的单品进行多次替换。通过观察模型对替换不同服装单品后套装的兼容性评分发现,当服装单品被同类型的单品替换掉后,兼容性分数虽然有所下降,但仍然高于模型判定服装是否兼容的阈值 0.5,且当被不同类型的单品替换后,模型能够有效地对这些负样本套装的兼容性做出准确的判断,表明对于套装兼容性分析而言,本文采用的异质图模型是有效且合理的方法;当除衣服外的饰品被不同类型的单品替换掉时,服装整体的兼容性会显著降低,且套装删除该饰品后的兼容性比有该饰品时的兼容性低,表明饰品的存在对整体服装的兼容性是有意义的。

4 结 语

本文对现有的套装兼容性建模方法的特点以及存在的问题进行分析,详细阐述了 NGNN 模型数据集的负样本采样策略存在的问题,提出基于 RGCN 的套装兼容性预测模型。本文模型将时尚套装建模成一个异质服饰搭配图,使用 RGCN 进行图卷积得到节点的嵌入向量;再使用图全局求和池化技术得到图的向量表达,进行套装的兼容性预测;采用难例采样的策略,有效地缓解了服装类别数量不均衡的问题。对本文模型进行广泛而深入的实验评估,证明了该模型的有效性。但是,本文也存在着一定的不足和限制:(1) 在针对多模态信息的处理上,一件时尚单品往往涉及图片和文字等多模态信息,如颜色、图案等视觉信息可以通过图片进行直观的描述,其材质、样式等信息可以通过文本数据用较短的文字简洁地总结。通过学习时尚单品的多模态特征有助于对服装兼容性的联合建模,提升单品搭配效果。因此,针对服装兼容性的多模态联合建模是本文未来进一步的研究方向。(2) 由于各种服饰风格具有多样性,每个人对服装兼容与否的判断也具有一定的主观性,负样本不一定就是不合理的,因此如何构造更加合理的负样本,也是下一步可以探究的方向。

参考文献(References):

- [1] Veit A, Kovacs B, Bell S, *et al.* Learning visual clothing style with heterogeneous dyadic co-occurrences[C] //Proceedings of the IEEE International Conference on Computer Vision. Los

- Alamitos: IEEE Computer Society Press, 2015: 4642-4650
- [2] Shih Y S, Chang K Y, Lin H T, *et al.* Compatibility family learning for item recommendation and generation[C] //Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 2403-2410
- [3] McAuley J, Targett C, Shi Q F, *et al.* Image-based recommendations on styles and substitutes[C] //Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2015: 43-52
- [4] He R N, Packer C, McAuley J, *et al.* Learning compatibility across categories for heterogeneous item recommendation[C] //Proceedings of the 16th IEEE International Conference on Data Mining. Los Alamitos: IEEE Computer Society Press, 2016: 937-942
- [5] Han X T, Wu Z X, Jiang Y G, *et al.* Learning fashion compatibility with bidirectional LSTMs[C] //Proceedings of the 25th ACM International Conference on Multimedia. New York: ACM Press, 2017: 1078-1086
- [6] Cui Z Y, Li Z K, Wu S, *et al.* Dressing as a whole: outfit compatibility learning based on node-wise graph neural networks[C] //Proceedings of the World Wide Web Conference. New York: ACM Press, 2019: 307-317
- [7] Shi C, Li Y T, Zhang J W, *et al.* A survey of heterogeneous information network analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(1): 17-37
- [8] Schlichtkrull M, Kipf T N, Bloem P, *et al.* Modeling relational data with graph convolutional networks[C] //Proceedings of the 15th International Conference on the Semantic Web. Heidelberg: Springer, 2018: 593-607
- [9] Chen L, He Y H. Dress fashionably: learn fashion collocation with deep mixed-category metric learning[C] //Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 2103-2110
- [10] Vasileva M I, Plummer B A, Dusad K, *et al.* Learning type-aware embeddings for fashion compatibility[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 405-421
- [11] Song X M, Feng F L, Han X J, *et al.* Neural compatibility modeling with attentive knowledge distillation[C] //Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York: ACM Press, 2018: 5-14
- [12] Yin R P, Li K, Lu J, *et al.* Enhancing fashion recommendation with visual compatibility relationship[C] //Proceedings of the World Wide Web Conference. New York: ACM Press, 2019: 3434-3440
- [13] Li Y C, Cao L L, Zhu J, *et al.* Mining fashion outfit composition using an end-to-end deep learning approach on set data[J]. IEEE Transactions on Multimedia, 2017, 19(8): 1946-1955
- [14] Pang K C, Zou X X, Wong W K. Modeling fashion compatibility with explanation by using bidirectional LSTM[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 3889-3893
- [15] Li X C, Wang X, He X N, *et al.* Hierarchical fashion graph network for personalized outfit recommendation[C] //Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2020: 159-168
- [16] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [17] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[OL]. [2022-09-19]. <https://arxiv.org/abs/1609.02907>
- [18] Lu Z, Hu Y, Jiang Y C, *et al.* Learning binary code for personalized fashion recommendation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 10554-10562
- [19] Paszke A, Gross S, Massa F, *et al.* PyTorch: an imperative style, high-performance deep learning library[C] //Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019: Article No.721
- [20] Fey M, Lenssen J E. Fast graph representation learning with PyTorch geometric[OL]. [2022-09-19]. <https://arxiv.org/abs/1903.02428>
- [21] Han D, Yun S, Heo B, *et al.* Rethinking channel dimensions for efficient model design[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamito: IEEE Computer Society Press, 2021: 732-741