计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(4):52−65 [doi: 10.15888/j.cnki.csa.009030]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于时空图神经网络的商品销量预测①

韦泰丞[1], 刘雁兵[1], 张宓觅[2], 刘慎慎[2], 李　宁[1]

[1](广西中烟工业有限责任公司, 南宁 530001)
[2](中科知道 (北京) 科技有限公司, 北京 100190)
通信作者: 李　宁, E-mail: 276546552@qq.com

摘　要: 精准预测商品的销量是提高商品营销效率的前提和基础. 为了更好地预测商品销量, 现有研究人员提出了基于深度神经网络 (DNNs)、卷积神经网络 (CNNs)、时间序列分析等方法, 但这些方法大多只单方面考虑到商品销售过程中的时间或者空间特征. 同时基于商品销售数据的建模分析发现, 商品的销量和对应的零售商户的空间位置和销售时间有较大的相关性. 为了更加准确地预测某种商品, 在特定商店, 以及在特定时间的销量, 本研究首先构建了以商家为基础的大规模知识图谱系统, 通过一张图的数据模型, 描述商品销售和对应的商圈、商户、用户的相关交互场景. 同时在图模型上增加了商家数据的空间和数据特征, 用于描述商户的时空特性. 最后基于构建的商家知识图谱, 本研究提出了基于图卷积神经网络 (GCN) 聚合信息获取空间特征, 然后使用长短期记忆 (LSTM) 提取时间特征, 并将两种特征进行加权结合, 进行商品销量预测. 初步研究结果表明: 基于图和 LSTM 模型的混合模型的算法预测投放量最为贴近实际销量, 相比于传统的神经网络算法, 该模型预测的平均准确率为 89%. 最后通过构建流水线工作流, 将整个商品销量智能预测系统部署到生产环境中, 为实现商品精准化营销提供了智能化决策.

关键词: 销量预测; 神经网络; 知识图谱; GCN-LSTM; 智能营销

引用格式: 韦泰丞,刘雁兵,张宓觅,刘慎慎,李宁.基于时空图神经网络的商品销量预测.计算机系统应用,2023,32(4):52–65. http://www.c-s-a.org.cn/1003-3254/9030.html

Prediction of Commodity Sales Based on Spatiotemporal Graph Neural Network

WEI Tai-Cheng[1], LIU Yan-Bing[1], ZHANG Mi-Mi[2], LIU Shen-Shen[2], LI Ning[1]

[1](China Tobacco Guangxi Industrial Co. Ltd., Nanning 530001, China)
[2](Zhongke Zhidao Technology Co. Ltd., Beijing 100190, China)

Abstract: Accurate prediction of commodity sales is the premise and basis of improving the efficiency of commodity marketing. In order to better predict commodity sales, existing researchers have proposed methods based on deep neural networks (DNNs), convolutional neural networks (CNNs), time series analysis, etc., nevertheless most of these methods only take into account the temporal or spatial characteristics of the commodity sales process unilaterally. At the same time, based on the modeling analysis of commodity sales data, it is found that there is a great correlation between the commodity sales and the spatial location and sales time of corresponding retail merchants. In order to more accurately predict the sales of a certain commodity in a specific store and at a specific time, this study first constructs a large-scale knowledge graph system based on merchants. Through the data model of a graph, the study describes related interaction scenarios of commodity sales and corresponding business circles, merchants, and users. At the same time, the spatial and data characteristics of merchant data are added to the graph model to describe the spatial and temporal characteristics of merchants. Finally, based on the constructed merchant knowledge graph, the study aggregates information based on a graph convolutional neural network (GCN) to obtain spatial features and then uses long short-term memory (LSTM) to

① 收稿时间: 2022-08-15; 修改时间: 2022-09-15; 采用时间: 2022-10-21; csa 在线出版时间: 2023-03-17
CNKI 网络首发时间: 2023-03-19

extract temporal features. Furthermore, the study performs a weighted combination on the two features to predict commodity sales. Preliminary research results show that the commodity sales predicted by the hybrid model algorithm based on the graph and LSTM model is the closest to the actual sales. In addition, compared with that of traditional neural network algorithms, the average prediction accuracy of the model is 89%. Finally, by constructing an assembly line workflow, the whole intelligent prediction system of commodity sales is deployed in the production environment, which provides intelligent decision-making for realizing accurate commodity marketing.

Key words: sales prediction; neural network; knowledge graph (KG); GCN-LSTM; smart marketing

商品营销和推荐是商业公司的一项核心、基础业务, 也是实现精准预测销量的前提和基础. 然而, 一直以来, 商品营销和推荐由于业务量巨大而繁复则会耗费大量人力、物力、财力, 重复而繁重的业务成为商品营销工作中的痛点. 因此, 用于预测和推荐一部分商品以满足商家进行精准预测销量的智能化方法应运而生. 李贤宗[1] 利用双知识图谱解决商品推荐系统中用户数据缺失与非结构化数据问题, 并结合图卷积神经网络以提高商品推荐的准确度; 和志强等[2] 提出了 WEFCMO 算法, 该算法建立商品关联数据网络, 获取各商品节点关系的集合, 并用该算法进行聚类, 提供引导型商品营销策略; 柯苗等[3] 搭建 LSTM 网络模型, 该模型便于数据输入, 在时间序列预测方面具有很大的准确性; 韩亚娟等[4] 提出的基于随机森林、GBDT、XGBoost 算法的成本厌恶偏向性组合预测模型能够精确地预测销量; 郑琰等[5] 提出的基于多种群遗传算法的时间序列模型对商品的预测精度较高; 黄文明[6] 提出的结合图片等结构化数据的深度学习预测模型提供了更为精确的销量预测方法.

以上研究人员运用了不同的模型与方法对商品销量进行时序预测, 但在商品营销的时空特征结合这一方面, 目前鲜有这方面的系统研究. 针对上述问题, 图卷积网络 (GCN) 和长短期记忆 (LSTM) 时空网络的模型已经被用于滚动频率预测[7], 基于 GCN-LSTM 模型, 本研究首先通过构建大规模商家知识图谱, 用于表示各商店之间的关系; 进而基于 GCN-LSTM 混合模型提取商户的空间特征与时间特征, 并将两种特征相结合进行销量预测.

1 研究内容概述

商品预测营销是商业公司的重要业务之一, 传统的预测方法不可避免会存在数据工作量大、任务繁重, 以及人工在调取、记忆、计算等方面的局限性, 难以实现大规模精确计算, 进而会直接影响商业公司和商户的经济效益. 由此, 如何转型使用人工智能方法进行精准预测营销成为各商业公司的重点研究方向.

深度神经网络具有强大的特征提取能力[8]; 图卷积神经网络 (GCN) 作为一种深度学习表示算法, 已经显示出强大的应用性能, 不仅可以表达复杂的语义关系, 还可以捕获全局的图信息[9]; 被广泛应用于人工智能领域的 LSTM 是一种具有长短期记忆信息能力的神经网络.

伴随以上神经网络的深入应用, 为解决商品精准预测销量的问题, 通过输入商品不同的数据类型, 如: 零售户标签、商店间的距离、品规与销量等, 本研究提出了一种基于图卷积神经网络 (GCN) 和长短期记忆 (LSTM) 神经网络的新框架, 用于时间与空间结合的销量预测, 对此进行了全面综述. 具体来说, 本研究主要工作如下.

(1) 首先构建商店知识图谱, 在商圈数据标签中抽取出结点属性与边的属性. 通过构建全流程的自动化知识图谱系统, 为了更加准确地描述商家知识图谱, 提出了分领域的知识图谱 Schema 构建模型. 同时为了从文本和商户交易账单数据中提取实体, 本文提出了基于 BERT+GRU 针对商家数据的实体自动抽取模型.

(2) 针对图卷积神经网络的商品销量的预测, 构建了 GCN-LSTM 混合模型, LSTM 用于解决无法学习到长期依赖特征关系的问题. 该模型同时融合 GCN 的相关空间特征数据与 LSTM 的时间特征数据, 能够很好地利用商家之间的位置结构, 获得数据中蕴含的空间特征, LSTM 过程可以较好地捕捉到销量随时间动态变化的特征, 从多个方面获得时间相关性, 实现最终对商品销量的预测, 进而提高商品销量预测准确率.

(3) 为了表示商品智能预测销量的准确率, 选取两

种不同品规的商品分别进行不同时间段的销售量预测的相关可视化实验验证, 结果表明: GCN-LSTM 模型预测平均准确率为 89%.

(4) 本研究的数据管理和商品销售预测系统已经发布到多个地级市的商品推荐和智能化推荐平台. 通过大规模知识图谱的构建, 实现了多源数据的高效管理. 同时智能化的商品销售预测系统, 相比传统模式, 对于商品销售量有 6% 的提高指引.

本研究的工作内容能够自动从数据中抽取出重要的特征, 弥补手动寻找特征繁重工作量的不足, 甚至能够学习推理出人工无法提取到的数据之间的潜在特征, 以此来提高工作效率, 实现商品营销智能化与精准化. 但在研究工作中我们存在着收集数据不全面, 以及影响商品投放的综合因素较多等不足, 后续我们也会对相关模型进行因地制宜构建, 在商品智能营销领域, 创造出更加深刻而有效的应用价值.

2 大规模商家知识图谱构建和图神经网络综述

2.1 图

图是一种描述事物与事物之间关系的数据结构, 通常由节点和边构成, 其中节点代表实体, 节点之间有连边表示两个实体之间具有某种关系. 图可以表示为 $G=\{V,E\}$, 其中 $V=\{v_1,v_2,\cdots,v_n\}$ 是图中所有节点的集合; $E=\{e_1,e_2,\cdots,e_m\}$ 是边集, 用于描述两个节点之间的关系. 一张简单的图如图 1 所示.

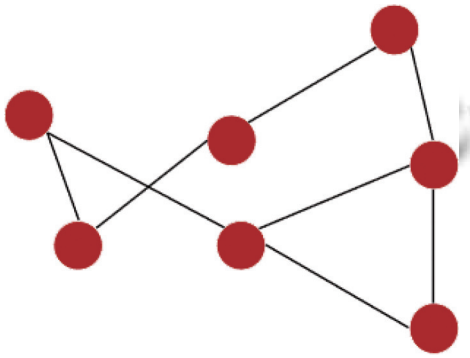


图 1 简单图

图 G 有 n 个节点, $A=(a_{ij})_{n\times n}$ 表示图的邻接矩阵, a_{ij} 的定义如式 (1):

$$a_{ij}=\begin{cases}1, & v_i\text{与}v_j\text{相连}\\ 0, & v_i\text{与}v_j\text{不相连}\end{cases} \tag{1}$$

若 G 是一个无向图, 则 $a_{ij}=a_{ji}$. D 表示 G 的度矩阵, $D=diag(d_1,d_2,\cdots,d_n)$, d_i 表示与 v_i 连接的节点数. 假设图中每个节点有 d 个特征, 将图中所有节点的特征组合成矩阵 $X=(x_{ij})_{n\times d}$.

2.2 商圈信息图谱建设

知识图谱包括: 知识建模、知识抽取、知识融合、知识存储、知识推理、知识应用这 6 个核心步骤. 最后的图结构基于三元组表示出“实体-关系-实体”, 特征信息来自结点的属性与边的属性. 此种形式具有较强的可解释性, 能够直观地反映图谱的结构信息与语义信息. 基于商店信息, 本研究管理的商家知识数据如表 1 所示, 构建的知识图谱语义定义如图 2 所示, 整体知识图谱构建框架如图 3 所示.

表 1 图谱的空间大小

商店数量	商品类别	点数	边数
4 000个	147个	>100 万个	>200 万条

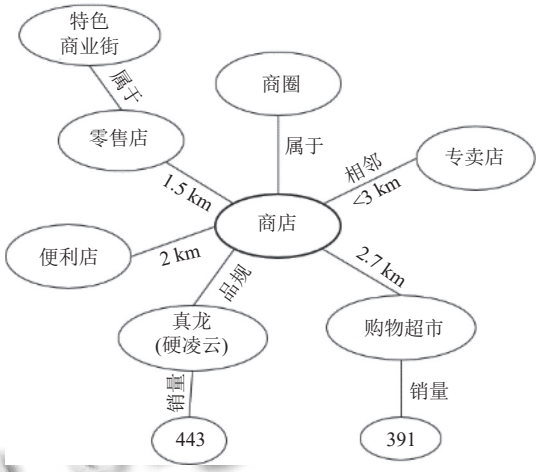


图 2 商店知识图谱

数据获取, 在商店知识图谱构建过程中, 所用的数据是多源异构的. 结构化数据来源于研究团队的整合. 半结构化数据来源于商品投放平台相关数据, 需要做属性归一. 非结构化数据来源于碎片文本内容, 知识加工需要对商圈零售户标签数据进行信息抽取.

知识建模, 在获取到数据源后, 对知识进行建模. 按照“概念域-实体域-事件域”进行定义. 概念域是具体实体的抽象, 两个实体间的边由各商店之间的距离、品规、销量分别表示对应关系. 实体域是业务相关的实例, 如图谱中的不同商店名称、品规与销售量. 事件域指客户的购买行为, 以这些行为背后的事件作为结构化知识的沉淀, 来增强实体域里的一些静态知识. 在第 2.3 节重点介绍知识图谱 Schema 设计的逻辑.

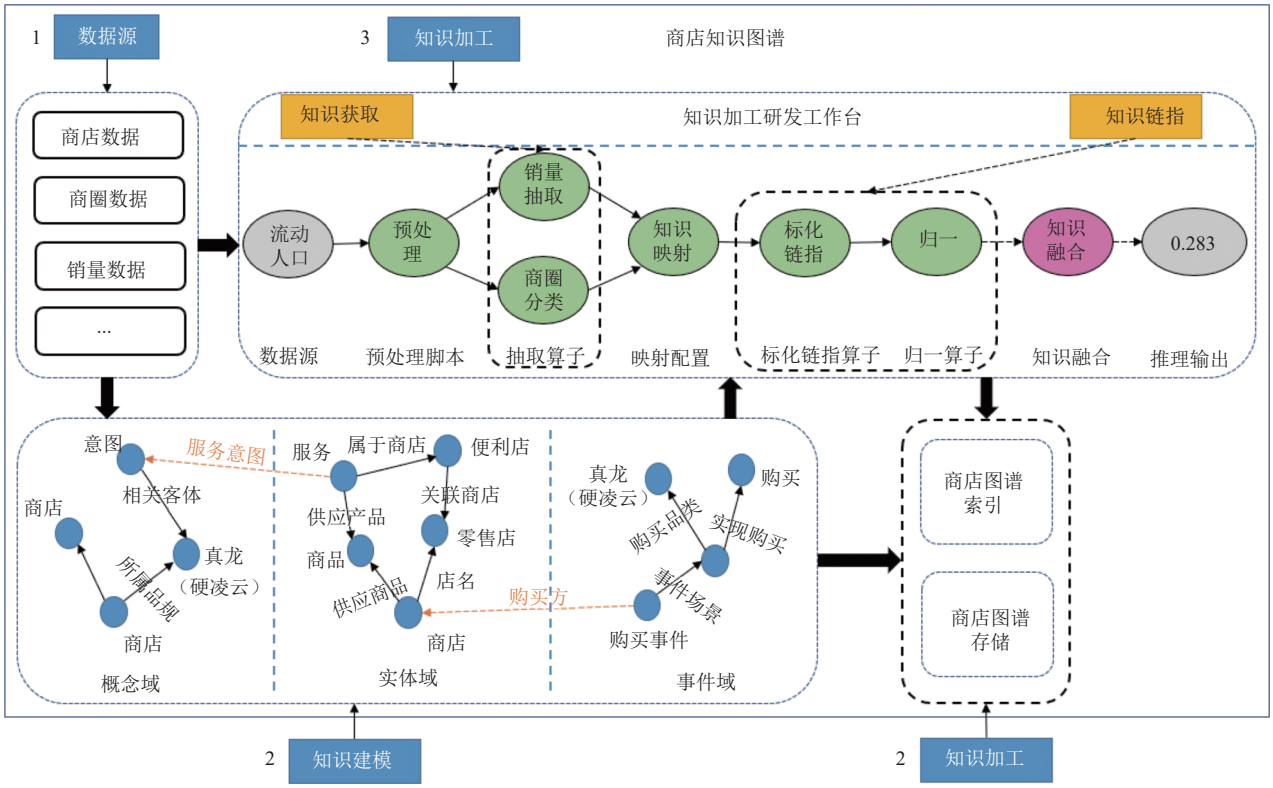


图 3 图谱构建-整体框架

知识抽取和加工, 对于半结构化或者非结构化商圈数据, 在进行获取时, 会涉及知识的分类或抽取, 比如各商店实体之间的关系该如何定义. 在知识获取后, 将知识的结构化数据和知识建模里的架构做映射. 随之做实体链指归一工作, 一是实体链接, 二是实体归一或属性归一. 知识加工是提供了一个自定义 DAG 的图谱构建链路. 这里用户可以根据自己的需要来进行算子的封装, 我们在进行数据获取时, 可能会涉及对知识的分类或者抽取, 如实体抽取、关系抽取等. 在知识获取后, 要做字段的映射, 即将知识的结构化数据和知识建模里的 Schema 做映射. 随后需要实体链指归一的工作, 包含两个部分, 一是实体链接, 另一部分是实体归一或属性归一. 本研究中, 主要基于 BERT+GRU (gate recurrent unit) 的方法来抽取实体.

知识加工后将三元组数据进行知识融合, 对复杂的语义关系做出知识推理, 预测出属性与关系: 小于 3 km 的两个商店之间能够对应; 商店属于商圈这一区域.

为了支持不同的客户业务需求, 设计对知识进行分层的存储架构, 从而能够支撑毫秒级的实时在线查询、秒级别的商家平台知识运营, 以及大规模的图计算数据处理, 最终形成知识图谱的语义网络.

2.3 商家图谱 Schema 设计

商家图谱全局 Schema 架构分为概念域、实体域、事件域 3 个维度. 其中, 实体域表示客观物理世界的真实知识或供息; 概念域是为了业务运营、挖掘用户真实需求, 在精神世界中创造的对事务的分类和特征描述; 事件域则是对特下动态信息的描述.

2.3.1 实体域

在商家认知图谱中, 每个实体域的知识, 在结构上是一个 CVT (compound value type). 即该实体本身是一个包属性信息的核心节点, 如: 名称、地址、id、数据来源、品牌等. 需要注意的是, 这些文本属性, 是非结构化的. 通过关系边指向各个特征语义的节点, 如所类目、意图、与其他供给和用户间的关联. 则该核心节点与语义项实子图拓扑结构, 以及该子图上所有节点和边的属性, 体现了该 CVT 的完整语义. 这种将实体表示为 CVT 结构的 sch 式, 也是为了将具有高度连通性的图结构用于图表示学习, 以获取实体关于领域子图语义的 embedding 表示. 图 4 列举了一个具体的商家实体的定义模型.

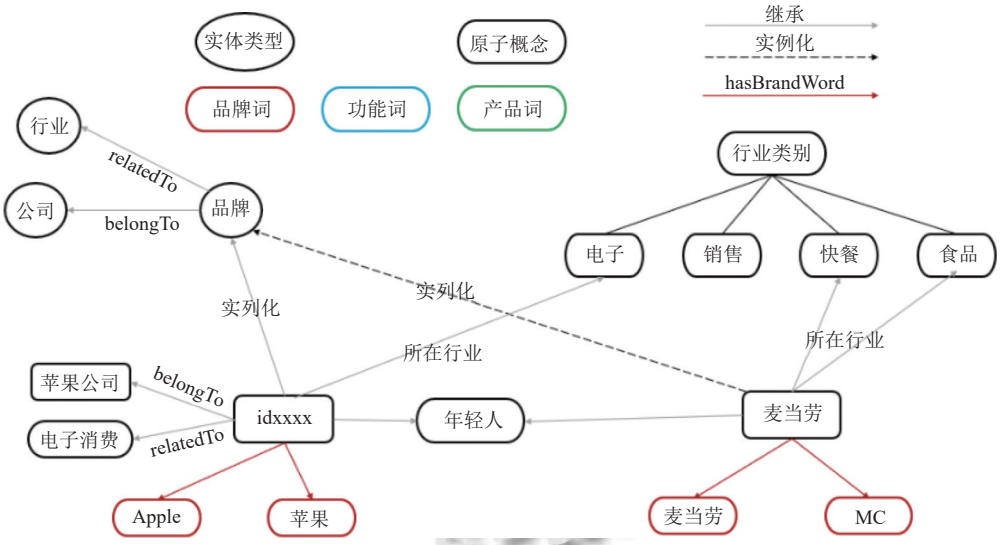


图 4 实体结构定义

2.3.2 概念域

实体域中, 用户、商户、门店都是物理世界或数据层面真实的存在, id、名称、地理位置等属性. 但对于供给的内容、特色、用户的偏好、需求, 却难以用确定的属性或数值来描述. 在此, 对应于第 2 节 Schema 中的“概念层”, 为商家知识图谱构建帮助用户行为和商家内容理解的概念层.

“概念”是对客观事实的抽象, 是对多个实例的所共有的特征属性归纳总结后形成的知识. 在概念域中, 我们重点建设了意图、类目体系. 其中:

意图: 意图是对服务能力提供的抽象, 由意图词 (action) 和实体词 (object) 组合而成, 反映标准化的功能行为及对象. 意图概念作为媒介, 链接了用户需求和攻击内容间的语义 gap. 标签: “标签”是长度为 4–6 个字的短语, 其语义描述和组织 item 或 user 的特征. 为用户和供给挂载上相应特征高效的匹配、召回和输出, 解决搜准场景精准匹配的问题.

类目: 是由业务创建的, 帮助理解各类型供给按照行业或内容分类的体系.

2.3.3 事件域

概念域和实体域是相对静态的知识, 而事件域定义动态和行为, 包括账单交易支付行为、商店内部的销售场景. 本研究中支持对特定场景下用户意图、商家服务的挖掘, 最终获得商户动态理解的结构化知识.

此处以账单支付行为为例, 说明在商家知识图谱中所使用的时间结构化 Schema 表示. 我们把账单中体现的支付事件, 抽象为“履约事件”实体. 履约事件描述一个特定的用户与一个特定的门店之间信息. 而访问/交易的地点、地点标签、时间标签等, 也通过多条带领频次属性的关系链接. 因此, 一个“履约事件”, 能与确定商家间, 在不同时间/空间下的交易画像. 同时, 由于一个门店/商户与其供应的产品、支持的意图、所属行联是存在的, 因此能间接推断出用户在该商户的履约行为的目的 (兴趣产品/意图).

2.4 基于图结构的商家实体抽取

商家知识图谱构建过程中, 涉及从文本数据中抽取出商家和商品实体. 本研究使用 BERT+GRU 的模式来抽取实体. GRU 是循环神经网络 (recurrent neural network, RNN) 的一种. 和 LSTM (long short-term memory) 一样, 也是为了解决长期记忆和反向传播中的梯度等问题而提出来的. 按图 5 这个框架从下往上看, 最下面是用 BERT 获得 token 表征, 上面一层加入了 graph layer, 再上面是双向的 GRU, 最后是 label 预测. 相比直接用 BERT 做实体抽取, 差别在于我们模型加入了 graph layer 这一层. 我们模型是在 graph layer 中通过构建图神经网络来学习词与字之间交互的特征. 获得这个特征之后, 再通过 Bi-GRU 对 BERT 的输出和 graph layer 输出进行融合, 从而进一步预测每一个 token 的 label 信息. 此抽取方法在实验环境和生产环境中, 获得了比传统的基于 BERT 的 NER 模型更高的准确度.

2.5 图神经网络

卷积神经网络利用卷积的方式提取数据的特征.

但传统的卷积神经网络只能处理像图像、文本等这些欧氏空间数据, 且这些数据具有平移不变性. 为了处理像图数据这样的非欧式空间数据, 学者们提出了图神经网络 (GNN)[10] 的方法. GNN 可以将图结构数据进行转化, 然后在输入到各种神经网络中进行训练. 我们可以将图神经网络[11] 划分为以下几类: 图卷积网络, 图注意力网络, 可扩展的图网络等.

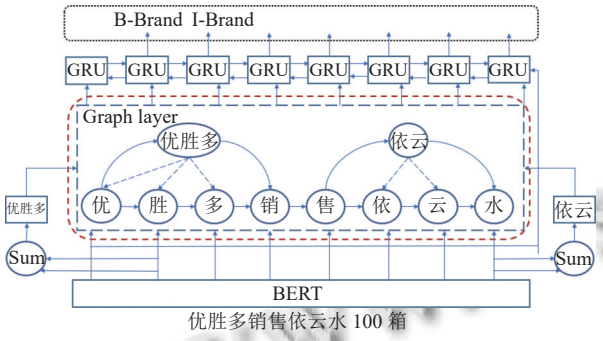


图 5 基于双向 GRU 的商家实体识别

2.5.1 图卷积神经网络

Bruna 等[12] 在 2013 年提出了一种将图和卷积神经网络结合起来的方法即图卷积神经网络 (GCN). GCN 是从图数据中提取特征, 利用获得的特征信息来实现节点预测、节点分类和边预测等任务. GCN 的训练过程如图 6 所示.

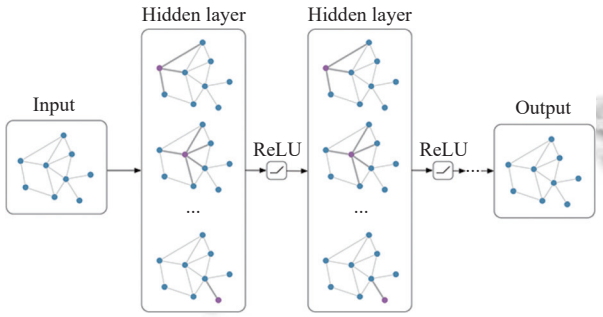


图 6 GCN 的训练过程图

GCN 将图数据作为输入, 对每个结点的邻居都进行一次卷积操作, 并用卷积的结果更新该结点, 实现节点之间信息的聚合; 通过非线性激活函数作用后再将结果作为第 2 层输入进行卷积, 重复上述操作, 直到层数达到预期的深度. 最终图卷积网络也可以将节点状态转换为与任务有关的标签等作为输出.

与传统的卷积神经网络类似, 图卷积神经网络可以从数据中提取特征, 不同的是后者处理的是图数据. 图卷积神经网络的核心也是通过聚合邻居节点的信息进行结构特征的提取, 而这与卷积神经网络的思想不谋而合. 提取图特征的方式有两种, 一种是谱域, 另一种是顶点域. 谱域是利用图的拉普拉斯矩阵特征值和特征向量来研究图的性质; 顶点域是把每个顶点相邻的邻近节点找出来, 定义节点之间的连接关系, 再对邻居节点的信息进行聚合.

(1) 基于谱域的图卷积

图的拉普拉斯矩阵为$L=D-A$, 其中D是节点的度矩阵, A是节点的邻接矩阵. GCN 基于拉普拉斯的谱分解将矩阵分解为特征值和特征向量矩阵之积如式 (2):

$$L=U\Lambda U^{-1}=U\Lambda U^{\mathrm{T}} \tag{2}$$

其中, U是单位特征向量构成的正交矩阵.

Bruna 等[12] 提出的初代图卷积模型为$y_{\text{output}}=\sigma\cdot\left(Ug_{\theta}(\Lambda)U^{\mathrm{T}}x\right)$, 但初代图卷积神经网络计算量很大, 没有空间局部性, 模型性能不好.

为了增加一代图卷积模型的空间局部性, Defferard 等[13] 于 2016 年提出第 2 代图卷积神经网络如式 (3):

$$y_{\text{output}}=\sigma\sum_{j=0}^{K}\alpha_j L^j \tag{3}$$

这不仅对核卷积进行了改进, 还进一步降低了计算的复杂度.

GCN 从提出以来, 经过多次改进, 由 Kipf 等[14] 在 2017 年提出了第 3 代图卷积神经网络. 此次改进加深了网络的深度, 定义了层与层之间的传播方式如式 (4):

$$H^{(l+1)}=\sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{4}$$

其中, $\tilde{A}=A+I$, $\tilde{D}$是$\tilde{A}$的度矩阵; H是每一层的特征, 对于输入层来说$H^{(0)}=X$, W是参数矩阵, σ是非线性激活函数.

(2) 基于顶点域的图卷积

基于顶点域的图卷积可以和欧式空间上的卷积操作进行类比, 从顶点域开始, 通过定义聚合函数聚合每个中心节点与邻居节点. 由此可以看出顶点域图卷积遇到的问题在于图结构中每个节点的邻居节点个数不同. 为了解决此问题, 需要定义可以处理任意长度邻居节点的卷积核, 使得卷积核可以根据不同节点的邻居节点数自适应地进行卷积, 从而进一步提取图中节点

的特征. 对所有邻居节点隐藏状态求和, 更新当前节点隐藏状态, 实现无参数卷积: $x_v^{l+1}=\sum_{u\in N(v)}x_u^l$, x_u^l是第l层特征表示, $N(v)$代表节点v的邻居节点集合. 基于顶点域的图卷积可以处理大规模的图结构, 应用也很广泛.

2.5.2 图注意力网络

GCN 获取图空间特征的方法非常依赖于图的结构, 并且对于邻域中不同节点的权重都是一样的. Veličković等[15] 在 2018 年提出将注意力机制和图卷积神经网络进行结合的图注意力网络 (GAT) 模型. 图注意力网络主要有两个优点, 一个是可以为每个节点分配不同的权重, 其次是引入注意力机制之后节点信息只与其邻居节点有关, 不需要得到整张图的信息.

设图G有n个节点, 节点特征向量集合作为输入, 每个节点有F个特征, $h=\{\vec{h}_1,\vec{h}_2,\cdots,\vec{h}_n\}$, 其中$\vec{h}_i\in\mathbb{R}^F$. 新的节点特征集合, $\vec{h}'_i\in\mathbb{R}^{F'}$, $h'=\{\vec{h}'_1,\vec{h}'_2,\cdots,\vec{h}'_n\}$是输出. 经过处理后, 特征向量的维度可能会改变$F\to F'$. 对于两个节点$i,j$, 使用线性变换$W\in\mathbb{R}^{F'\times F}$将$F$维特征转换为$F'$维特征. 对于节点$i$, 计算它与每个邻居节点之间的相似系数$e_{ij}=a\left(W\vec{h}_i,W\vec{h}_j\right)$, a是一个共享注意力机制, 可以将拼接后的向量映射到实数上. 最后利用 Softmax 进行归一化即可得注意力系数如式 (5) 所示:

$$\alpha_{ij}=\frac{\exp(LeakyReLU(e_{ij}))}{\sum_{k\in N_i}\exp(LeakyReLU(e_{ik}))} \tag{5}$$

第 2 步进行特征的加权求和, 通过非线性激活函数后输出每个节点融合领域信息的特征向量$\vec{h}_i=\sigma\left(\sum_{j\in N_i}\alpha_{ij}W\vec{h}_j\right)$.

如图 7 所示, 对于两个节点i,j, 图注意力网络首先学习它们之间的注意力权重$\alpha_{i,j}$, 之后基于注意力权重$\{a_1,\cdots,a_6\}$对节点$\{1,2,\cdots,6\}$的特征表示$\{\vec{h}_1,\cdots,\vec{h}_6\}$进行加权平均, 从而可以得到节点 1 的特征表示$\vec{h}'_1$.

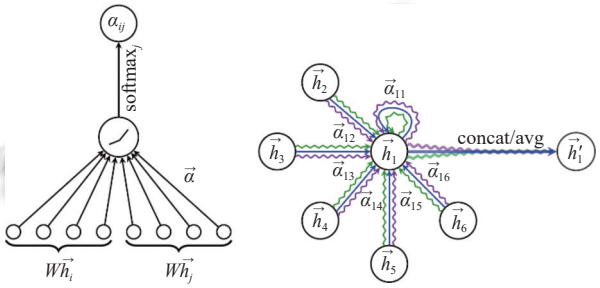
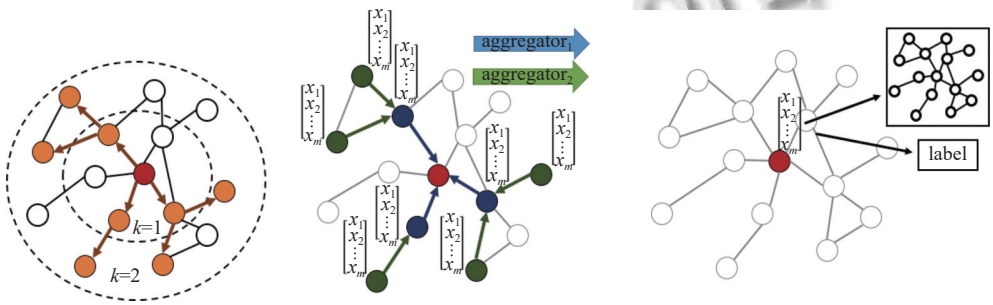


图 7 图注意力网络学习过程

2.5.3 可扩展的图网络

GCN 中所做的卷积融合了全图的信息, 若图结构较大, 节点个数较多, 则 GCN 的效率会很低, 由此出现了可扩展的图网络 Graph-SAGE[16], 学习过程如图 8 所示. 该网络通过随机采子图进行采样, 通过子图更新节点, 以此得出的子图结构本身就是变化的, 从而使模型学到的是一种采样及聚合的参数, 避免了训练过程中需要把整个图的节点特征一起更新的情况, 增加了扩展性.



(a) Sample neighborhood (b) Aggregate feature information from neighbors (c) Predict graph context and label using aggregated information

图 8 Graph-SAGE 学习过程

Graph-SAGE 首先对每个节点采用切子图的方式, 然后随机采样出部分邻居节点作为聚合的特征点. 采出子图之后, 做特征融合, 这一步骤和 GCN 所使用的方式一样能够得到中心节点的特征, 最后对节点做出分类任务等.

3 基于 GCN-LSTM 的商品销量预测模型

商品的销量是一个带有时间特征的数据, 同时销量也与其他空间特征有关. 图卷积神经网络是可以用于处理图数据并提取空间特征, LSTM 能够用于处理时间序列问题并提取数据的时间特征. 因此也有很多

学者提出使用 GCN 与 LSTM 相结合的方法对交通流量, 空气质量等进行预测. Zhao 等[17] 结合 GCN 和 GRU 提出 T-GCN 模型用于预测交通流量, 祁柏林等[18] 使用 GCN-LSTM 模型小微型监测站的空气质量. 本研究中, 主要基于商家知识图谱构建, 把 GCN-LSTM 用于商品的销量预测.

3.1 模型构建

对于商品销售而言, 若两个商户之间的位置距离比较接近, 会对彼此的商品销售有一定影响. 因此将每个商户视为一个节点, 假定商户之间距离小于 3 km 会有连边. 最终构建商户图$G=\{V,E,A\}$, $V=(v_1,v_2,\cdots,v_n)$, n表示节点数, E是边, A是邻接矩阵, 表示商户之间的连边情况, 若商户之间没有连边为 0, 有连边为 1. 各商店之间的简略图结构如图 9 所示.

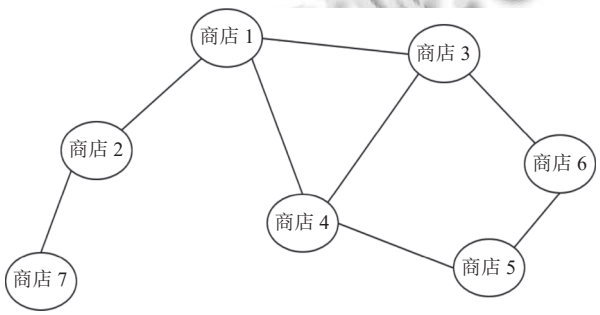


图 9 各零售店之间的结构

节点的特征矩阵: 将商家的一些与商品销量有关的信息看作是节点的属性特征, 形成特征矩阵$X\in\mathbb{R}^{n\times p}$, p表示节点特征的数量. 节点特征包括前文中提到的商店所在商圈基础属性、人群特征、消费能力以及商品销量等. $X_t\in\mathbb{R}^{n\times p}$, 代表$t$时刻的节点属性特征.

因此 GCN-LSTM 模型是通过商店的图结构以及相应的特征矩阵学习如下映射: $f:G;(X_{t-m},\cdots,X_{t-1})\rightarrow X_t$预测未来的销量. 模型分为两个步骤, GCN 和 LSTM, 训练流程图如图 10 所示. 首先使用商家历史m个数据作为输入, 使用 GCN 聚合信息获取空间特征, 再将具有空间特征的时间序列数据输入到 LSTM 模型中, 捕捉时间特征得出最后的预测结果.

使用 GCN-LSTM 模型预测商品销量的流程如图 11 所示. 在输入数据预处理之后, 将数据分为两部分, 一部分用于训练模型, 另一部分用于检验模型的预测效果. 若模型在测试数据上表现不佳, 则再进一步对测试模型进行调整.

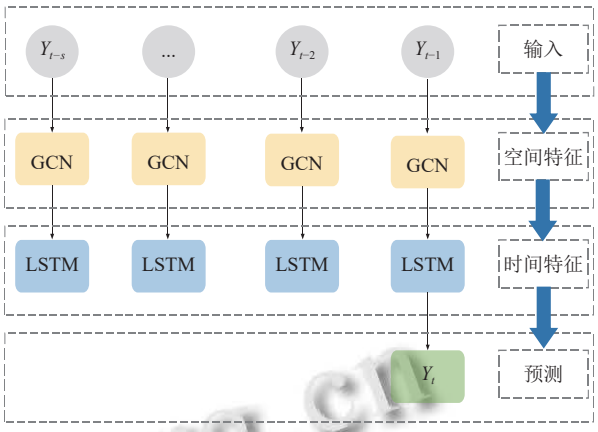


图 10 GCN-LSTM 训练流程图

3.2 空间特征的提取

预测销量的关键是获得商家之间的一个空间结构, 传统的卷积神经网络只能够从欧氏空间数据中提取特征, 无法处理商家之间位置的图结构, 因此我们选择使用图神经网络处理图数据, 提取空间特征.

GCN 可以收集到商家及其周围相连商家之间的拓扑关系, 聚合商家之间的信息, 获得特征的空间相关性. GCN 过程将包含商家信息的特征矩阵、反应位置关系的邻接矩阵作为输入进入到卷积层进行计算, 聚合信息. 每一次聚合的结果都将作为新的输入再次进入卷积层, 直到达到预先设定的卷积层数L, 输出最后的结果即为融合了周围信息以及自身信息的 GCN 输出.

3.3 时间特征的提取

预测商品销量的另一个关键是获得数据之间的时间相关性, 传统的 RNN 对于长期预测具有局限性, 只适合学习短期记忆. LSTM[19] 作为 GNN 的变体, 使用门控机制记忆更多的长期信息, 可以解决 GNN 中存在的问题. 因此我们选择使用 LSTM 模型从销量数据中获取时间特征. h_{t-1}表示$t-1$时刻的隐藏状态, x_t是t时刻的销量信息, c_t是t时刻细胞状态, h_t是t时刻的输出状态. LSTM 通过遗忘门决定从细胞状态中丢弃哪些信息, 再通过输入门决定要放入什么样的信息放入细胞状态中, 更新细胞状态, 最后通过输出门输出最后的值. 我们输入$t-1$时刻的隐藏状态和当前的销量信息来预测下一个时刻的销量. LSTM 结构如图 12 所示.

3.4 GCN-LSTM 模型

为了更好地预测销量, 使用了 GCN-LSTM 模型同时提取数据中的空间特征和时间特征. 具体的计算过程如下.

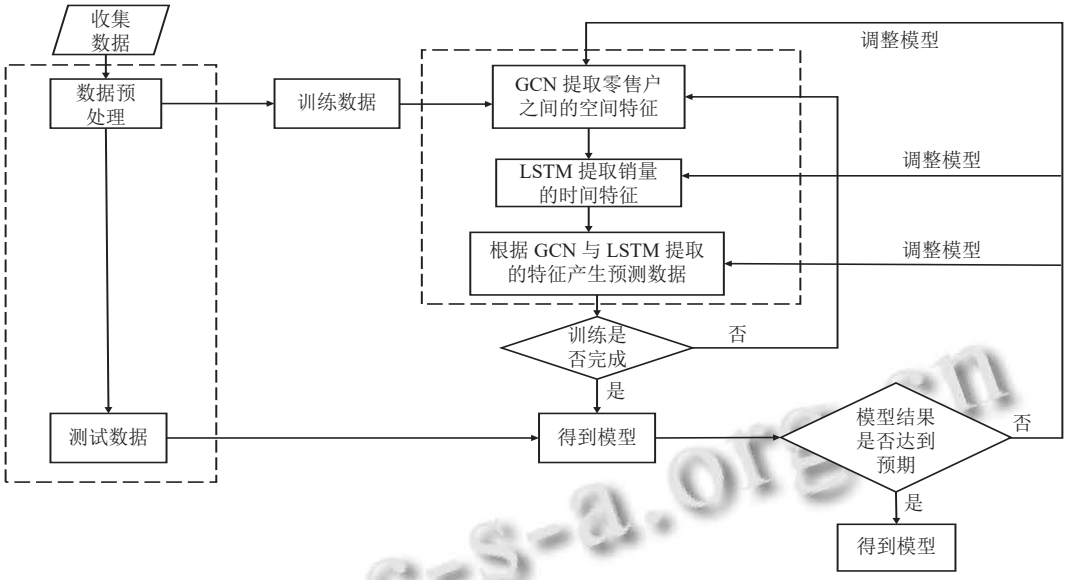


图 11 GCN-LSTM 的销售量预测流程图

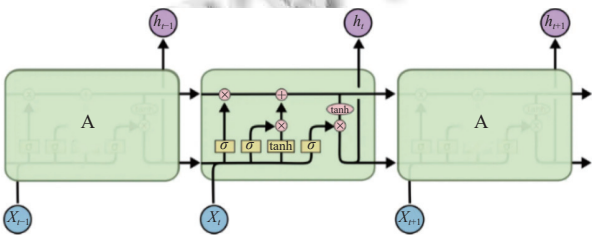


图 12 LSTM 结构

$$\begin{cases} f_t = \sigma\left(W_f \cdot [h_{t-1}, g(A, X_t)] + b_f\right) \\ i_t = \sigma(W_i \cdot [h_{t-1}, g(A, X_t)] + b_i) \\ C_t = f_t \times C_{t-1} + i_t \times \tanh(W_C \cdot [h_{t-1}, g(A, X_t)] + b_C) \\ o_t = \sigma(W_o [h_{t-1}, g(A, X_t)] + b_o) \\ h_t = o_t \times \tanh(C_t) \end{cases}$$

其中, $g(A, X_t)$为进行图卷积过程, W, b是训练过程中的权重和偏差. 本研究对图卷积过程作出的伪代码如算法 1.

算法 1. 图卷积过程代码

Input:

Graph $G(V,E)$; feature matrix X_{t-i}; adjacency matrix A; degree matrix $\tilde{D}$; identity matrix I; layer number L; weight matrix $W^{(l)}$, $l \in \{1,2,\cdots,L$; non-linearity σ

$\tilde{A}=A+I$

$\tilde{D}=\sum_{j=1}^{n} A_{ij}$

for i =1, 2,⋯, s **do**

  for l = 1, 2,⋯, L **do**

    $H_{t-i}^{(0)}=X_{t-s}$

    $H_{t-i}^{(l+1)}=\sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H_{t-i}^{(l)}W^{(l)}$

  end

  $\tilde{X}_{t-i} \leftarrow H_{t-i}^{L}$

end

LSTM prediction: $[\tilde{X}_{t-s},\cdots,\tilde{X}_{t-1}] \rightarrow X_t$

Output:

X_t

GCN-LSTM 模型中图卷积过程可以很好地利用商家之间的位置结构, 获得数据中蕴含的空间特征, LSTM 过程可以较好地捕捉到销量随时间动态变化的特征, 获得时间相关性, 实现最终对商品销量的预测.

4 实验分析与结论

本实验采用卷烟这一商品, 选取真龙 (硬凌云) 和南京 (硬红) 两种品规, 分别进行一周与一个月的销量预测; 对某一零售户真龙 (硬凌云) 这一品规的卷烟作出 2020 年 4 月到 2022 年 5 月的销售时序图; 将 2022 年初的实际销量与 GCN-LSTM 预测销量进行对比; 并作出两种算法的预测准确率.

4.1 数据说明及其预处理

本研究使用到的数据为各零售户所在商圈的基础属性、人群特征、消费能力、市场状态, 近 3 个月不同品牌卷烟单箱平均价格, 不同品牌卷烟销量以及不同品牌卷烟每周的销售情况. 数据类型集一共分为 6 大类, 如表 2 所示.

商圈数据和不同品牌的商品销售数据可能在录入时出现为空的情况, 为了利于后续神经网络的学习, 需要对其进行处理, 将数值型的数据补 0, 类别型的数据补 null.

表 2 零售户标签

类型	标签
商圈基础属性	居民小区数量
	购物中心数量
	写字楼数量
商圈人群特征	居住人口
	流动人口
	工作人口
	常住人口
	性别
	年龄段
	学历
	婚否
商圈消费能力	高档位
	中档位
	低档位
	平均档位
	消费水平
	餐饮水平
商圈市场状态	近3个月不同品牌卷烟单箱平均价格
商圈消费指标	近3个月不同品牌卷烟销量
商圈消费偏好	不同品牌卷烟每周的销售情况

考虑到数据集中存在数值型数据和类别型数据, 不同商圈属性下的数值型数据可能存在量级以上的差距, 可能导致神经网络忽略掉小数量级的特征, 类别型的数据由于不是数值型, 无法直接输入神经网络进行训练, 所以需要对数值型数据和类别型数据进行特征变换.

对于数值型数据, 统一进行 log1p 函数运算, 得到一个较为平滑且服从高斯分布的数据; 对于类别型的数据, 分别进行 LabelEncoder 编码从而得到数值型特征.

4.2 评价指标

为了评估两个算法对于销量的预测性能, 使用绝对平均误差 (*MAE*) 和均方根误差 (*RMSE*) 来评估实际销量与预测销量之间的偏差, 使用准确率评估模型的效果, 如式 (6) 所示:

$$\begin{cases} MAE = \dfrac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \\ RMSE = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \\ Accuracy = 1 - \dfrac{\|Y - \hat{Y}\|_F}{\|Y\|_F} \end{cases} \tag{6}$$

预测效果越好, 代表着预测这与真实值越接近, 偏差越小, *MAE* 和 *RMSE* 的值也就越小.

4.3 实验对照模型

我们基于浅层神经网络构建实验的基准对比算法, 算法流程图如图 13 所示.

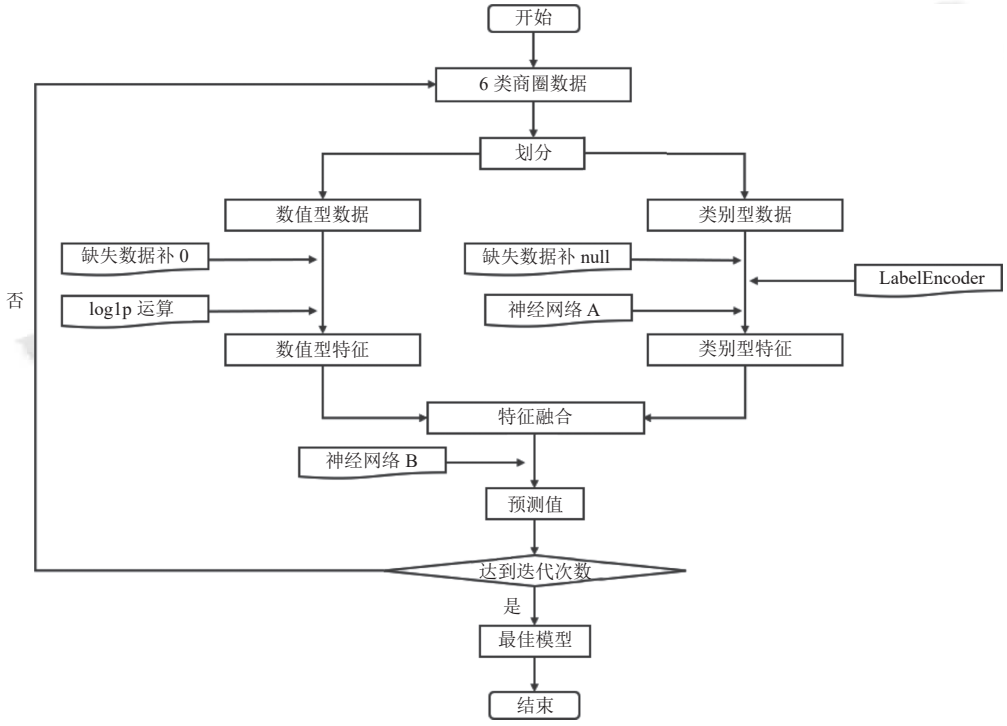


图 13 基于浅层神经网络的商品投放算法流程图

图 13 中, LabelEncoder 将类别型数据变换为数值型, 如存在手机品牌的标签, 该标签下有“苹果”“华为”“小米”, 此时调用 LabelEncoder 会分别将“苹果”“华为”“小米”依次从 0 开始顺序映射, 结果为 0, 1, 2.

神经网络 A 的结构如图 14 所示.

神经网络 A 为两层神经网络, 使用 kaiming_normal_进行权重初始化. 在第 1 层神经网络的输出后使用 batch_normal 函数将输出归一化到标准的分布形态上, 然后使用 ReLU 函数进行激活, 之后输入第 2 层神经网络, 并将第 2 层神经网络的输出作为神经网络通过学习后生成的一组类别特征.

在得到平滑后的数值型特征会和类别特征进行特征融合, 由于数值特征和神经网络特征都为向量, 本算法将这两种特征向量进行拼接, 如式 (7) 所示, 其中 NF_i 代表数值特征, EF_i 代表类别, CF_i 代表融合特征.

$$CF_i = [NF_i, EF_i] \tag{7}$$

在融合特征之后, 会输入到神经网络 B 当中. 该神经网络是一个 3 层结构, 同样使用 kaiming_normal_进行权重初始化. 在第 1 层的输出使用 ReLU 函数激活进行 dropout 操作, 然后送入 batch_normal 层进行归一化, 接着送入第 2 层神经网络, 同样对第 2 层的输出进行一个 dropout 操作, 最后接一层全连接层并将全连接层的输出作为预测值.

整个神经网络的训练使用 MSELoss, 用于衡量目标值与预测值的差异. 如式 (8) 所示, 其中 x 为预测值, y 为真实值.

$$loss(x_i, y_i) = (x_i - y_i)^2 \tag{8}$$

神经网络的优化器采用 Adam 优化器, 学习率为 0.000 1, 每批输入 64 组数据进行算法训练.

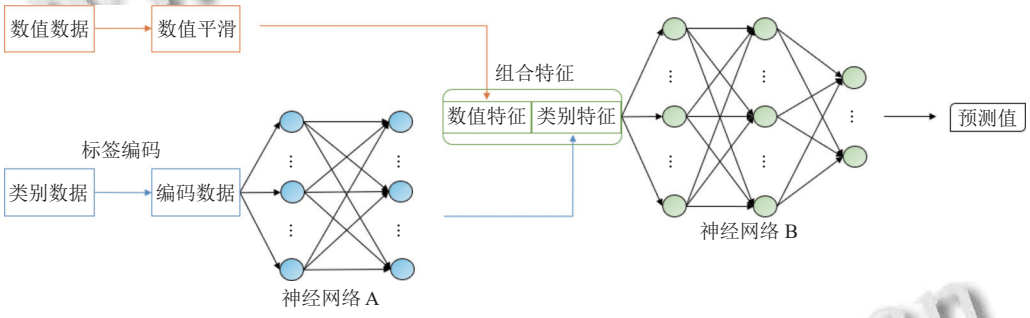


图 14 神经网络架构图

4.4 实验结果

实验中分别使用了浅层神经网络和 GCN-LSTM 两个模型对 10 个零售户的卷烟销售进行了预测, 预测结果对比如表 3 所示.

表 3 不同模型的预测效果对比

零售户ID	神经网络		GCN-LSTM	
	MAE	*RMSE*	*MAE*	*RMSE*
32****5374	9.19	13.305	8.45	12.42
32****5945	9.63	13.94	8.22	12.11
32****5977	8.55	12.87	8.60	12.81
32****0331	8.46	12.20	7.71	11.28
32****3369	7.15	9.93	6.29	9.33
32****0317	10.31	16.06	9.66	15.05
32****7321	8.91	13.08	8.45	12.98
32****7146	12.43	17.95	11.53	17.53
32****5375	10.87	14.81	9.34	12.63
32****5376	9.13	13.63	7.91	11.95

从表 3 中结果可以看出, GCN-LSTM 模型的 *MAE* 和 *RMSE* 评价指标都会比浅层神经网络模型的小, 说明该模型对销量的预测值与真实值之间的偏差更小, 预测效果更佳.

图 15–图 18 为使用神经网络和 GCN-LSTM 算法对 10 个零售户两种品规卷烟一周的销量预测及其真实值对比, 以及使用两种算法预测某一个零售户两种卷烟一个月 (一个月以 28 天计算) 的销量.

上述实验分别列举出给 10 个不同零售商的一周销量预测与一个月的销量预测, 并用两种算法作出预测投放量实验. 结果如以图 15–图 18 所示: 经过不同算法处理后, 可以看出, 两种品规的预测销售量均能达到智能预测的效果, 其中两种品规的 GCN-LSTM 算法预测销售量均更为贴合实际销量.

图 19–图 21 分别为某一零售户真龙 (硬凌云) 这

一品规的卷烟从 2020 年 4 月到 2022 年 5 月近两年的销售时序图, 以及从 2022 年初实际销售量与 GCN-LSTM 预测销量之间的对比, 并对两种算法的预测准确率进行对比. 从图中可以看出, 浅层神经网络的预测平均准确率为 81%, GCN-LSTM 的预测平均准确率为 89%. 由此表明, GCN-LSTM 算法对于销量的预测效果更佳, 预测值更为贴近真实值.

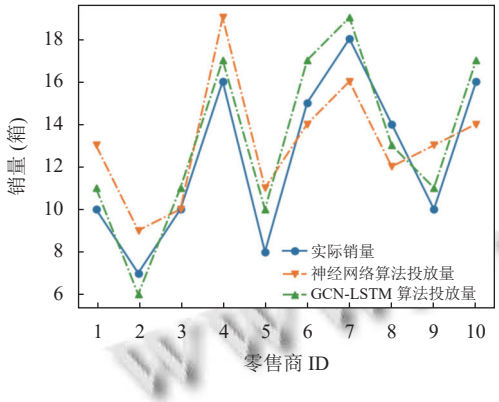


图 15 真龙 (硬凌云) 各零售商一周的销量

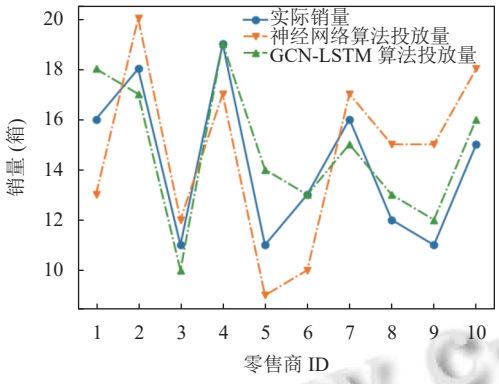


图 16 南京 (硬红) 各零售商一周的销量

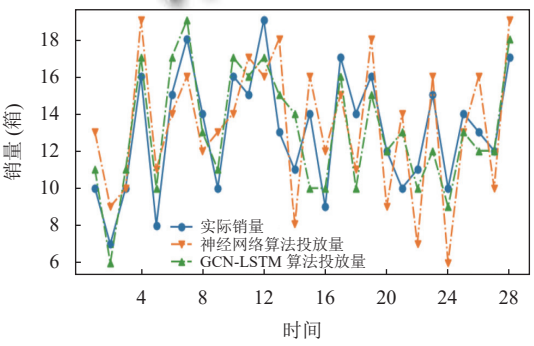


图 17 某零售商一个月销售真龙 (硬凌云) 的销量

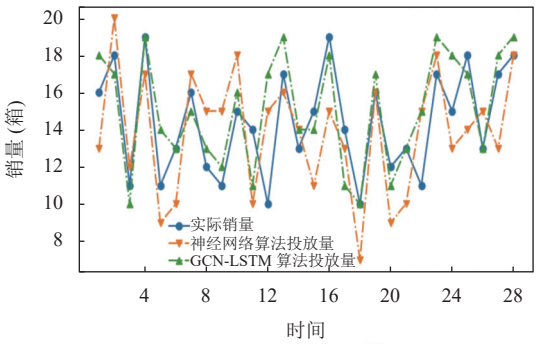


图 18 某零售商一个月销售南京 (硬红) 的销量

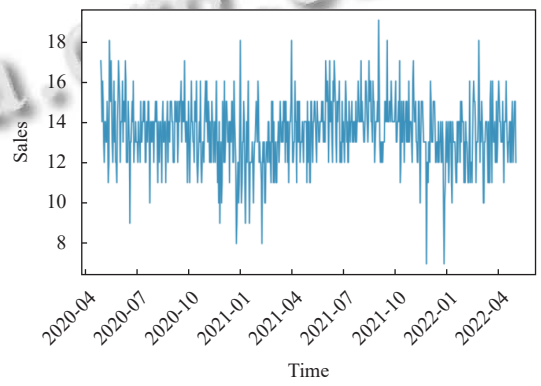


图 19 销售量时序图

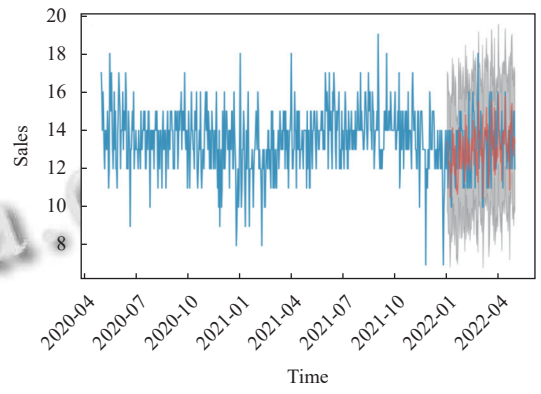


图 20 GCN-LSTM 销售量预测对比图

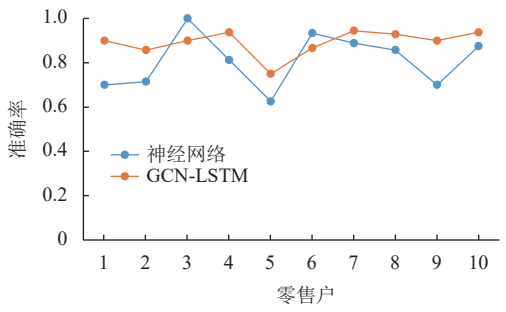


图 21 神经网络与 GCN-LSTM 准确率对比图

4.5 实验结论

综上所述, 由可视化图 15–图 21 得知, 使用特征融合的神经网络算法与 GCN-LSTM 混合模型算法能够弥补传统人工卷烟投放的不足. 其中, GCN-LSTM 算法更加贴近真实场景下的实际销售量, 预测的平均准确率为 89%, 能够更加准确地预测出卷烟营销数量, 从而达到智能化、自动化卷烟营销, 提高商业公司的工作效率.

5 系统原型

5.1 大数据处理工作流

PiFlow 是一个基于分布式计算框架 Spark 开发的大数据流水线系统 [20,21]. 我们把本研究提出的图神经网络算法, 标准化为 PiFlow 的一个可复用的计算算子, 帮助其他有相似场景的用户使用这一算法模型. 在技术角度, 打通 Spark 大数据工作流和深度学习模型 TensorFlow 的计算流. 同时本研究的模型, 已经部署到对应的生产环境.

本研究将浅层神经网络算法与 GCN-LSTM 混合算法分别应用到 PiFlow 大数据流水线系统中, 由此实现数据的采集、清洗、计算、存储. PiFlow 流程截图图 22 所示.

根据商店信息构建图谱, 使用 GCN 提取空间特征数据, 继而使用 LSTM 提取时间特征数据, 将时空特征进行融合, 用于训练模型, 达到预期效果后, 便可得到模型, 以此实现预测.

5.2 业务上线效果

本研究将两种品规的卷烟数据在商圈扩展智能投放平台中进行投放查询, 相关信息截图如图 23 所示.

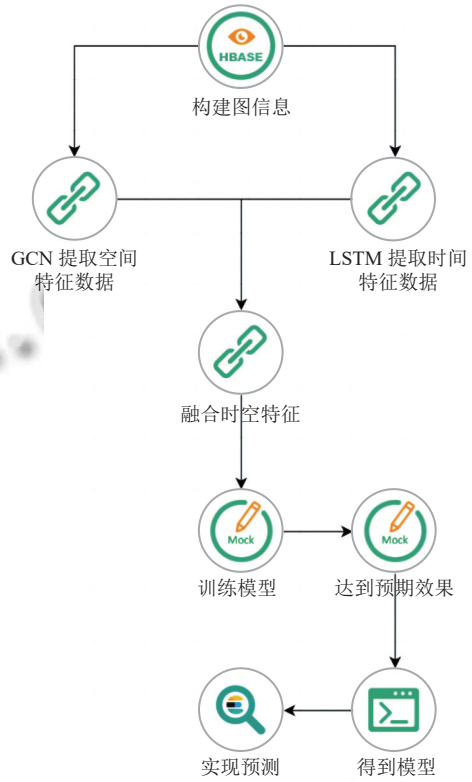


图 22 GCN-LSTM 混合模型 PiFlow 工作流程图

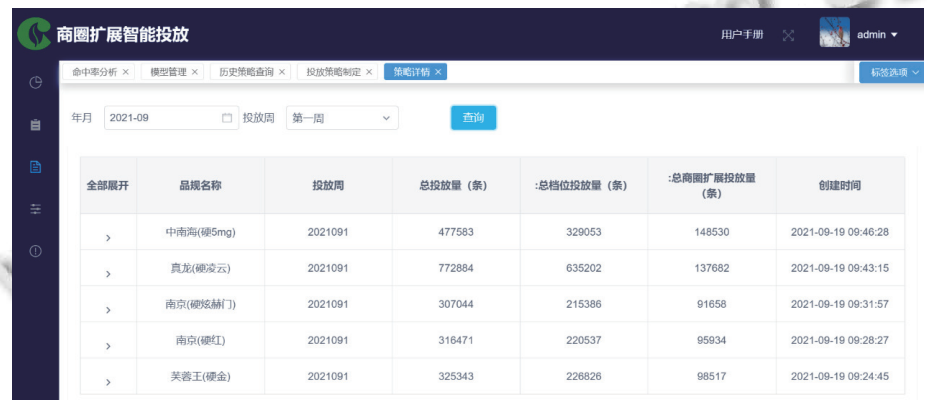


图 23 投放平台查询信息

6 结 语

针对商品智能精准销售问题, 本研究提出了浅层的神经网络算法和 GCN-LSTM 混合模型算法, 两种智能算法分别用于预测销售量. 在预测真龙 (硬凌云) 与南京 (硬红) 两种品规的一周销量、一月销量与不同时间段的销量预测的实验场景下, 验证得出: 浅层神经网络的预测平均准确率为 81%, GCN-LSTM 的预测平均准确率为 89%. 由此可见, GCN-LSTM 混合模型算法的预测结果更为贴合实际销售数据, 表明该算法能够提高工作效率, 并且能够较好地控制对不同零售户的

商品销售, 从而达到精准调控, 实现智能销售.

本研究能够在商品销售领域为商业公司提供参考价值, 未来将收集更多的销售数据, 设计更为有效的算法模型.

参考文献

1 李贤宗. 基于双知识图谱和图卷积神经网络的商品推荐研究 [硕士学位论文]. 长春: 吉林大学, 2022.

2 和志强, 罗长玲, 陈萌, 等. 基于 WEFCMO 算法的引导型商品营销实训辅助策略研究. 河北省科学院学报, 2020, 37(2): 1–7. [doi: 10.16191/j.cnki.hbkx.2020.02.001]

3 柯苗, 黄华国. 基于 LSTM 神经网络的电商商品销售预测方法. 福建师大福清分校学报, 2020, (5): 83–89.

4 韩亚娟, 高欣. 基于机器学习组合模型的电商商品销量预测. 计算机系统应用, 2022, 31(1): 315–321. [doi: 10.15888/j.cnki.csa.008345]

5 郑琰, 黄兴, 肖玉杰. 基于时间序列的商品需求预测模型研究. 重庆理工大学学报 (自然科学), 2019, 33(9): 217–222.

6 黄文明. 基于深度学习的商品销量预测研究 [硕士学位论文]. 南京: 南京理工大学, 2019.

7 Huang DY, Liu H, Bi TS, *et al*. GCN-LSTM spatiotemporal-network-based method for post-disturbance frequency prediction of power systems. Global Energy Interconnection, 2022, 5(1): 96–107. [doi: 10.1016/j.gloei.2022.04.008]

8 吴雨. 深度神经网络的特征融合机制及其应用研究 [博士学位论文]. 成都: 四川大学, 2021.

9 檀莹莹, 王俊丽, 张超波. 基于图卷积神经网络的文本分类方法研究综述. 计算机科学, 2022, 49(8): 205–216. [doi: 10.11896/jsjkx.210800064]

10 Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains. Proceedings of the 2015 IEEE International Joint Conference on Neural Networks. Montreal: IEEE, 2005. 729–734.

11 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述. 计算机学报, 2020, 43(5): 755–780. [doi: 10.11897/SP.J.1016.2020.00755]

12 Bruna J, Zaremba W, Szlam A, *et al*. Spectral networks and locally connected networks on graphs. Proceedings of the 2nd International Conference on Learning Representations. Banff, 2013.

13 Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 3844–3852.

14 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017.

15 Veličković P, Cucurull G, Casanova A, *et al*. Graph attention networks. arXiv:1710.10903, 2018.

16 Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1025–1035.

17 Zhao L, Song YJ, Zhang C, *et al*. T-GCN: A temporal graph convolutional network for traffic prediction. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(9): 3848–3858. [doi: 10.1109/TITS.2019.2935152]

18 祁柏林, 郭昆鹏, 杨彬, 等. 基于 GCN-LSTM 的空气质量预测. 计算机系统应用, 2021, 30(3): 208–213. [doi: 10.15888/j.cnki.csa.007815]

19 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]

20 朱小杰, 赵子豪, 杜一. 模型驱动的大数据流水线框架 PiFlow. 计算机应用, 2020, 40(6): 1638–1647.

21 周园春, 常青玲, 杜一. SKS: 一种科技领域大数据知识图谱平台. 数据与计算发展前沿, 2019, 1(1): 82–93.

(校对责编: 孙君艳)