软件学报 ISSN 1000-9825, CODEN RUXUEW
2025,36(2):608−624 [doi: 10.13328/j.cnki.jos.007151] [CSTR: 32375.14.jos.007151]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于预训练模型的用户评分预测*

强敏杰, 王中卿, 周国栋

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

通信作者: 王中卿, E-mail: wangzq@suda.edu.cn

摘　要: 随着商家评论网站的快速发展, 推荐系统所带来的效率提升使得评分预测成为近年来新兴研究任务之一. 现有的评分预测方法通常局限于协同过滤算法以及各类神经网络模型, 并没有充分利用目前预训练模型提前学习的丰富的语义知识. 针对此问题, 提出一种基于预训练语言模型的个性化评分预测方法, 其通过分析用户和商家的历史评论, 为用户在消费前提供评分预测作为参考. 该方法首先设计一项预训练任务, 让模型学习捕捉文本中的关键信息. 其次, 通过细粒度情感分析方法对评论文本进行处理, 从而获取评论文本中的属性词. 接下来, 设计一个属性词嵌入层将上述外部领域知识融入模型中. 最后, 采用基于注意力机制的信息融合策略, 将输入文本的全局和局部语义信息进行融合. 实验结果表明, 该方法相较于基准模型, 在两个自动评价指标上均取得显著的提升.

关键词: 推荐系统; 评分预测; 预训练模型; 注意力机制

中图法分类号: TP18

中文引用格式: 强敏杰, 王中卿, 周国栋. 基于预训练模型的用户评分预测. 软件学报, 2025, 36(2): 608–624. http://www.jos.org.cn/1000-9825/7151.htm

英文引用格式: Qiang MJ, Wang ZQ, Zhou GD. Prediction of User Rating Based on Pre-trained Model. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 608–624 (in Chinese). http://www.jos.org.cn/1000-9825/7151.htm

Prediction of User Rating Based on Pre-trained Model

QIANG Min-Jie, WANG Zhong-Qing, ZHOU Guo-Dong

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: As merchant review websites develop rapidly, the efficiency improvement brought by recommender systems makes rating prediction one of the emerging research tasks in recent years. Existing rating prediction methods are usually limited to collaborative filtering algorithms and various types of neural network models, and do not take full advantage of the rich semantic knowledge learned in advance by the current pre-trained models. To address this problem, this study proposes a personalized rating prediction method based on pre-trained language models. The method analyzes the historical reviews of users and merchants to provide users with rating predictions as a reference before consumption. It first designs a pre-training task for the model to learn to capture key information in the text. Next, the review text is processed by a fine-grained sentiment analysis method to obtain aspect terms in the review text. Subsequently, the method designs an aspect term embedding layer to incorporate the aforementioned external domain knowledge into the model. Finally, it utilizes an information fusion strategy based on the attention mechanism to fuse the global and local semantic information of the input text. The experimental results show that the method achieves significant improvement in both automatic evaluation metrics compared to the benchmark models.

Key words: recommendation system; rating prediction; pre-trained model; attention mechanism

随着互联网的发展和生活水平的提高, 人们对于商家评论网站的需求不断增长. 这些评论网站 (如: Yelp, Amazon) 内部通常会建立一个评分预测系统, 该系统将会预测用户对商家可能的评分, 从而为用户提供有价值的推荐. 显然, 建立一个能较为准确地预测用户评分的系统是一项具有现实意义的任务. 为了模拟用户消费前的任务

* 收稿时间: 2023-07-23; 修改时间: 2023-09-05, 2023-11-16; 采用时间: 2023-12-07; jos 在线出版时间: 2024-07-17
CNKI 网络首发时间: 2024-07-20

场景, 本文的任务设定是用户没有评论过目标商家. 其中, 任务的输入是用户的历史评论文本和商家的历史评论文本, 输出是用户对目标商家的个性化评分. 用户历史评论指的是用户对其他商家的评论, 商家历史评论指的是其他顾客对商家的评论. 值得注意的是, 尽管本文的任务在某种程度上与情感分析和推荐系统类似, 但它们之间仍有一定的差异. 情感分析通常是根据用户对商家的已有评论分析出用户的情感倾向或打分, 而推荐系统则通常基于用户的历史交互以及消费行为进行评分预测和推荐. 此外, 在推荐系统中较为经典的评分预测任务也与我们十分类似, 任务目标都是输出用户的评分, 但该项任务通常是基于用户对其他商家的评分来进行预测的. 在实际应用中, 由于无法提前获得用户对目标商家的实际评论, 也无法获得用户的历史交互等信息, 因此, 本文提出的任务是符合实际且具有挑战的.

如图 1 所示, 展示了一个利用历史评论来预测用户个性化评分的示例. 从商家的历史评论中可以提取到一些关键信息, 商家烹饪的海鲜十分美味, 同时装修也很精美, 但消费的顾客较多、环境嘈杂. 从用户的历史评论中可以得知, 该用户喜欢吃海鲜, 但对嘈杂的环境十分敏感, 会让他感到不适. 因此, 尽管商家的菜肴非常符合用户的口味, 但是在环境方面并不能给用户一个满意的体验. 综合以上分析出的信息, 模型预测用户的评分为 1 分.

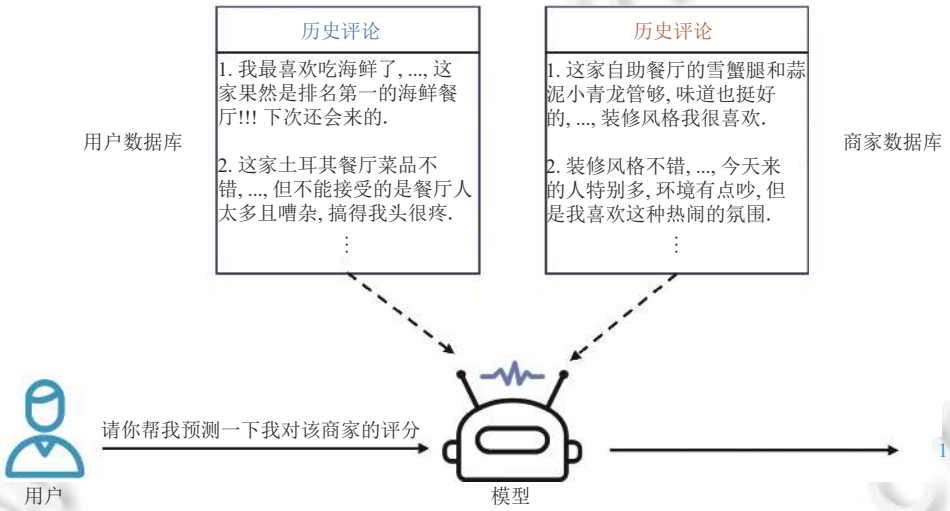


图 1 用户个性化评分生成示例

在过去的几十年里, 评分预测的研究取得了显著进展. 早期的评分预测系统主要使用协同过滤技术, 其主要的思想是通过计算用户/项目之间的相似度来进行预测. 此外, 矩阵分解也是协同过滤中重要的方法之一. Koren 等人[1]通过将用户-项目评分矩阵分解为低维潜在特征向量矩阵, 从而提高了预测性能. 但这些方法都存在一定的局限性, 如“冷启动”、数据稀疏等问题. 为了缓解这些问题, 基于内容的评分预测系统[2]开始得到关注, 其主要思想是利用关于用户和商家的文本信息进行推荐. McAuley 等人[3]发现评论文本中蕴含着丰富的信息, 利用 LFM[4]、LDA[5]等模型来提取评论文本的潜在的特征与主题可以有效提高预测性能. 尽管基于评论文本的评分预测系统已经取得了重大的进展, 但利用预训练的生成式模型的研究较为有限, 本文的研究填补了这一空缺.

近年来, 预训练模型在多项 NLP 任务上展现出较强的竞争力, 其在无监督方式下使用大量无标注文本进行预训练任务的学习, 从而让模型提前学习了丰富的语义知识, 并且能够更好地捕捉文本特征与隐含的语义信息. 此外, 生成式模型相较于分类模型 (如: BERT[6]), 能够生成连续的文本来充分地理解和利用上下文信息, 同时, 生成式模型的应用前景也更为广泛, 在未来的研究中可以额外生成决策原因以增加模型的可解释性. 因此, 本文采用了预训练的生成式模型 T5[7]作为基础模型, 整体流程如下: 首先设计了一项预训练任务, 该任务旨在让模型学习如何根据用户和商家的历史评论来预测两者的基本信息, 从而提升模型对文本中关键信息的提取能力. 其次, 通过细粒度情感分析 (aspect-based sentiment analysis, ABSA) 技术对评论文本进行预处理, 从而获取到评论文本中所有的属性词. 接着, 设计了一个属性词嵌入层, 并将属性词的位置信息提供给该嵌入层. 随后, 为了把外部领域知识融入输

入文本的每个词的语义信息中, 将属性词嵌入层的输出向量与 T5 模型自身的嵌入层的输出向量相加并送入 T5 编码器中, 从而获取到输入文本的全局语义表示. 然后, 采用基于注意力机制的信息融合策略, 将模型的输入文本进行拆分并分别编码, 之后将编码后的表示向量进行拼接作为输入文本的局部语义表示, 随后, 利用多头注意力机制将输入文本的全局与局部语义表示进行融合. 最后, 使用 T5 解码器结合信息融合层输出的隐藏向量进行解码, 生成用户的个性化评分.

本文的主要贡献如下.

● 本文提出了一种新的用户个性化评分预测任务, 该任务仅依赖用户和商家的历史评论, 为用户生成一个个性化的评分.

● 本文提出了一个基于预训练模型的用户个性化评分预测框架, 其中包括预训练任务、设计属性词嵌入层融合外部邻域知识和采用基于注意力机制的信息融合策略.

● 为了评估模型的有效性, 本文进行了一系列对比与消融实验, 研究了模型各组件的性能与作用. 实验结果表明, 相较于其他基准模型, 本文在 *ACC* 和 *MSE* 两个评价指标上均有着较大的提升.

1 相关工作

1.1 推荐系统

(1) 传统推荐系统: 早期的推荐系统研究主要集中在协同过滤[8–10]、矩阵分解等方法. 这些方法本质上依赖用户和项目自身的属性与特征, 并通过计算特征之间的相似度来寻找相似用户, 从而实现个性化推荐. Ganu 等人[11]首次考虑用户评论信息并配合人工标注的领域知识, 提高了餐厅场景中推荐的准确性, 这为后续工作开辟了新的道路. McAuley 等人[3]提出 HFT (hidden factors as topics) 模型, 使用潜在因素模型 (latent factor model, LFM) 和潜在狄利克雷分配模型 (latent Dirichlet allocation, LDA) 分别获取潜在评级维度与文本评论的主题, 将两者相结合来提高推荐系统中预测评级的准确性和可解释性. Zhang 等人[12]提出 CMLE (collaborative multi-level embedding) 模型, 通过投影级别将词嵌入 (word embedding) 模型[13]和标准矩阵分解 (matrix factorization, MF) 模型[1]进行集成, 从而获得捕捉单词局部上下文信息的能力, 并将评论嵌入投影到用户和项目嵌入上, 以此来应对用户、项目和评论之间的直接等价性问题.

(2) 结合评论文本的推荐系统: 随着深度学习的发展, 神经网络模型已经被证明比传统的推荐系统方法性能更好, 特别是在评论文本可用时. Zheng 等人[14]首次使用神经网络对用户和商家的评论联合建模, 他们提出了 DeepCoNN 模型, 由两个并行网络分别学习用户行为和商家属性, 最后通过顶部的共享层让两个网络相互交互. 用户和商家的评论相互交互有效地提升了预测性能, 缓解了数据稀疏问题. 但 DeepCoNN 模型的性能取决于模型是否能够访问用户对商家的评论, 而这在实际应用中是无法实现的. 为了解决上述问题, Catherine 等人[15]提出了 TransNets 模型, 该模型是在 DeepCoNN 模型的基础上扩展了一个 Transformer[16]层, 该附加层专门用于学习将用户和商家的潜在表示转换成对应评论的潜在表示, 从而在测试阶段可以生成评论的近似表示, 并用于模型的评分预测. 尽管基于评论文本的推荐系统已经取得了重大的进展, 但是对于自然形成的用户-商家二分图以及评论感知图的利用研究目前是缺乏的, Shuai 等人[17]提出 RGCL (review-aware graph contrastive learning) 框架, 首先构建一个含有评论特征增强边的评论感知的用户-商家图, 其次还设计了两个对比学习任务 (即节点和边的判别任务) 为推荐过程中的嵌入学习和交互建模提供自监督信号. 此前, 基于评论文本的推荐系统大多数只是简单地将评论表示和其他特征连接在一起, 而没有考虑到文本表示包含大量噪声信息. Yang 等人[18] 提出一种基于评论的多意图对比学习方法, 该方法采用了多意图对比策略, 在用户评论和项目评论之间建立了细粒度的联系, 进而提高了预测性能.

1.2 情感分析

(1) 情感分类: 情感分类的早期研究主要集中在文档级情感分类上[19,20]. 但随着用户对于商家评论网站需求的不断提高, 属性级情感分类任务开始受到关注[21–23]. 该任务的早期研究主要依赖于特征工程和一系列规则进行分

类. 受益于深度学习的发展, Tang 等人[24]基于 LSTM 提出了 TD-LSTM 和 TC-LSTM 模型, 在生成句子表示时捕捉目标词与其上下文之间的联系. 随着预训练模型的发展, 研究人员[25–27]使用预训练模型 BERT 进行属性级情感分析. 此后, 情感三元组和四元组抽取任务被研究人员[28–31]所提出并进行研究, 这类任务可以为用户提供更多有用的信息, 但同时更具有挑战性. 最近, 利用预训练的编码器-解码器模型在上述任务表现出了优秀的潜力[32]. Bao 等人[33]提出了一种观点树生成模型, 该模型旨在联合发现树中的所有情感元素. 本文复现了他们的工作, 并使用他们的模型从评论文本中抽取属性词. 最近, 为了解决多重重叠三元组抽取问题, Zhao 等人[34]在编码器-解码器模型的基础上, 设计了一种结合解码机制, 同时, 他们还构建了一个相关增强网络, 加强相关方面与意见之间的相互作用以进行情感预测.

(2) 情感打分: 情感打分是情感分析中的一项基本任务, 其旨在预测给定评论的评分. Pang 等人[35]首次对该任务进行了研究, 并将情感打分问题看作是一个分类/回归任务. 受到他们工作的启发后, 许多研究都采用了有监督的机器学习模型来进行评分预测. 为了利用文本中的细粒度特征来提升模型的预测性能, Qu 等人[36]提出了一种观点表示包, 其中每一个观点都由词根、一组修饰词和一个或多个否定词组成, 并使用脊回归算法学习每个观点的分数并预测评分. 类似的, Lu 等人[37]提出了一种计算用户观点中表达的形容词、副词强度的方法, 从而估计用户评论的情感强度, 即评分. 近年来 Transformer 的出现鼓励了注意力机制 (attention mechanism) 在情感打分模型中的进一步应用, Li 等人[38]提出了 HUARN 模型, 该模型使用分层架构来编码单词、句子和文档级别的信息. 之后, 利用用户注意力和方面注意力将文档表示与用户信息相结合, 以预测评论的方面评分. 现有的评论评级预测方法忽略了用户对方面的意见, 针对上述问题, Wu 等人[39]提出了一个协作学习框架来联合训练评论级别的评级预测器和多个方面级别的评级预测器, 从而进行评分预测. 受到近年来许多工作利用方面信息来提高评论评级预测性能的启发, Bu 等人[40]认为方面类别情感分析 (ACSA) 和评论评级预测这两个任务高度相关和互补, 因此建立了新的数据集 ASAP, 并提出了一个联合模型以此来综合解决两个任务并实现了更好的性能.

和之前的工作不同的是, 本文提出了一种新的用户个性化评分预测任务, 输入是用户和商家的历史评论, 输出是用户对商家的评分. 在表 1 中展示了与相关任务的对比, 本文提出的任务与情感分析和推荐系统类似, 但都存在一定的区别. 与情感分析类似的是, 都基于文本进行预测, 但需要注意的是, 情感分析是根据用户对目标商家已有的评论文本分析用户可能的情感倾向或打分, 而本文任务的设定是在用户对目标商家的评论文本不存在的前提下进行预测的. 与推荐系统类似的是, 都是为用户提前预测对商家的评分, 但不同的是, 推荐系统主要根据用户的历史交互以及消费行为进行预测, 而本文任务的设定是在不获取用户历史消费行为以及交互的前提下进行预测的. 此外, 在推荐系统中有一项与我们类似的评分预测任务, 都是预测用户对某个商家的评分, 但与本文任务不同的是, 该项任务通常基于用户对其他商家的评分进行预测的. 在实际应用中, 在消费前针对用户进行对商家的评分预测很有意义. 但这面临了一个问题, 即用户历史行为、与商家的交互和用户在消费后给出的评论在消费前是无法获取的, 因此本文仅依赖历史评论来进行评分预测, 这是符合实际且具有挑战的.

表 1　相关任务的对比

任务	输入	输出
推荐系统	用户&商家ID	商家评分
情感分析	已有评论文本	情感倾向/评分
本文任务	历史评论文本	商家评分

2　基于预训练模型的评论打分预测

在本文中研究了一个特定任务, 即在用户去商家消费之前, 基于用户和商家的历史评论预测用户可能对商家的评分.

具体来说, 本文的任务可以定义为: 给定用户历史评论集合 $r_u = \{r_u^1, r_u^2, \ldots, r_u^{n-1}, r_u^n\}$, 其中每条评论 r_u^i 都是用户过去对其他商家的评论文本; 商家历史评论集合 $r_b = \{r_b^1, r_b^2, \ldots, r_b^{n-1}, r_b^n\}$, 其中每条评论 r_b^i 都是其他用户过去对商家

的评论文本; 本文将这些数据组成模型的输入, 形式为“r_u < /s > r_b”, 其中每条单独评论中间都添加分隔符“ < /s >”. 最终, 模型将生成一个用户对商家的个性化评分, 形式是 1–5 的整数, 如图 2 所示.

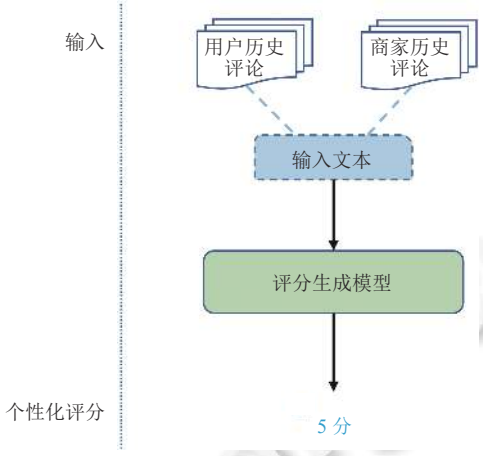


图 2 任务定义

如图 3 所示, 本文模型以预训练模型 T5 为基础, 模型结构可大致分为以下 3 个部分: 融入外部领域知识的 T5 编码层, 信息融合层以及 T5 解码层.

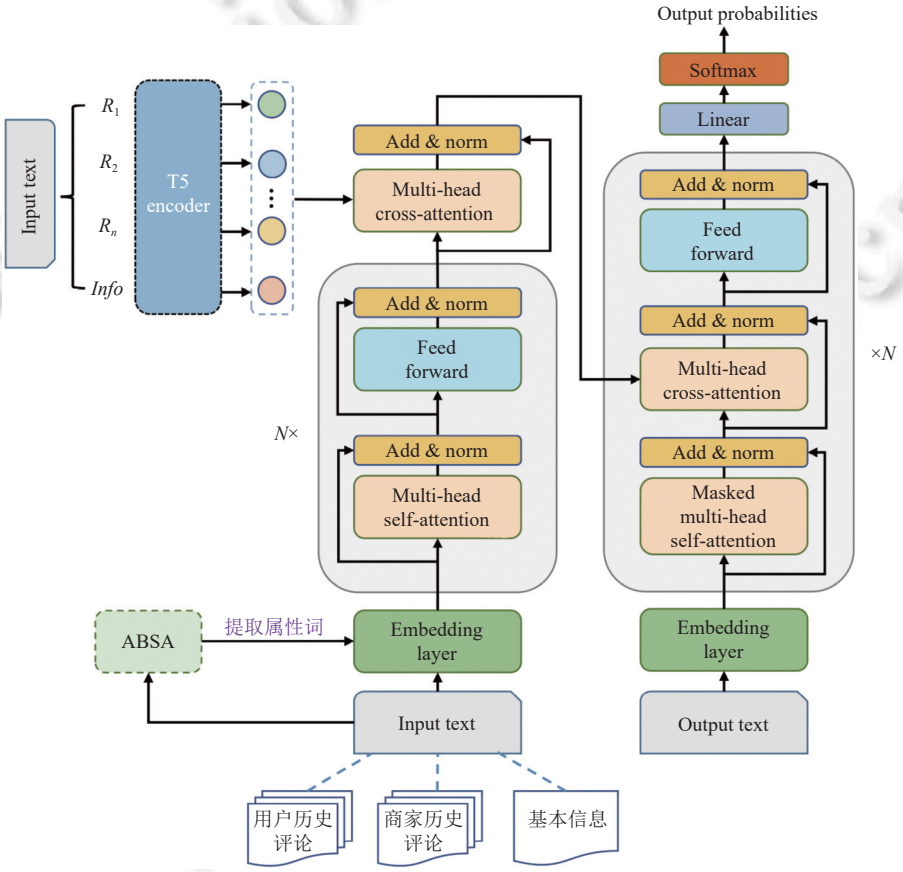


图 3 模型流程图

整体流程如图 4 所示, 首先, 设计了一个预训练任务, 该任务旨在让模型学习如何根据用户和商家的历史评论预测两者的基本信息, 从而提升模型捕捉评论文本中关键信息的能力. 其次, 复现了一个情感四元组抽取模型, 并利用该模型对用户和商家的历史评论进行细粒度情感分析, 从而获取评论文本中的所有属性词. 随后, 设计了一个属性词嵌入层, 并将属性词的位置信息送入到该嵌入层中, 同时将历史评论和基本信息送入到 T5 模型自带的词嵌入层中. 为了把外部领域知识融入每个词的语义信息中, 将几个嵌入层输出的词向量相加, 并送入 T5 编码器中获取输入文本的全局语义表示. 接下来, 采用基于注意力机制的信息融合策略, 将输入文本进行拆分后, 使用 T5 编码器分别对它们进行编码, 并将编码后的表示向量拼接作为输入文本的局部语义表示, 随后利用多头注意力机制对输入文本的全局和局部语义表示进行信息融合. 最后, 使用 T5 解码器结合信息融合层输出的隐藏向量进行解码, 生成用户的个性化评分.

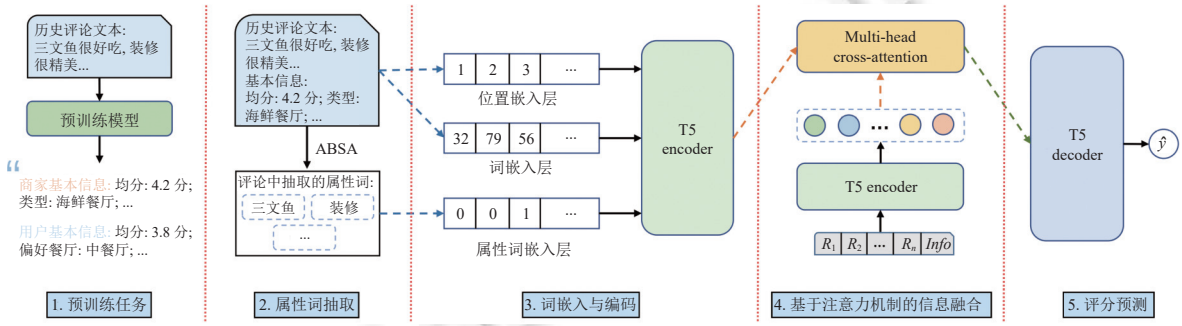


图 4 模型图

2.1 外部领域知识的表示

2.1.1 属性词抽取

在过去的 10 年里, 为了更深入了解用户对商家各方面的评价, 越来越多的学者开始关注细粒度情感分析任务[41,42]. 本文复现了 Bao 等人[33]的工作, 根据他们的方法, 本文选择了预训练模型 T5 作为基础模型. 此外, 本文使用了 ACOS[30]、Yelp 和 Amazon 数据集对模型进行训练. 经过联合学习多个预训练任务后, 模型能够较为准确地提取属性词-情感词四元组. 由于任务输入中包含与目标商家无关的评论, 即用户的历史评论, 这些无关的历史评论中的情感词可能会误导模型产生错误的判断. 因此, 本文选择仅将属性词作为外部领域知识融入模型, 这样既能使模型识别到评论文本中的关键属性, 又避免了引入无关情感词所造成的干扰.

2.1.2 属性词嵌入

这些抽取的属性词反映了用户和商家的多个方面, 如食物、环境、服务等. 在获取到属性词之后, 本文构造了一个属性词提示序列 $S=\{s_1,s_2,\ldots,s_{n-1},s_n\}$, 对于每个单词 w_i, 如果该单词是属性词, 则在提示序列中相应位置标记为 1, 否则标记为 0, 即:

$$s_i=\begin{cases}1, & \text{如果}w_i\text{是一个属性词}\\ 0, & \text{其他}\end{cases} \tag{1}$$

例如, 输入文本为: “The Chinese food is very tasty (中国食物非常美味)”, 该文本中的属性词为 Chinese food, 则属性词提示序列为: [0, 1, 1, 0, 0, 0]. 为了让这些外部领域知识融入模型中, 本文对基础模型进行了扩展. 如后文图 5 所示, 本文在原有的词嵌入层之后添加了一个属性词嵌入 (aspect embedding) 层, 用于处理属性词提示序列.

2.2 用户、商家和基本信息的表示

2.2.1 用户信息的表示

本文获取了用户过去对其他不同商家的评论文本, 这些评论信息将作为用户特征向量的基础. 具体地, 本文将用户历史评论词序列 $W_u=\{w_1,w_2,\ldots,w_{n-1},w_n\}$ 送入到词嵌入层中, 该层的输出是一个词向量序列 $E_{\text{word}}=\{e_{w_1},e_{w_2},\ldots,e_{w_{n-1}},e_{w_n}\}$, 其中 e_{w_i} 是词 w_i 的词向量表示. 之后, 将属性词提示序列 $S_u=\{s_1,s_2,\ldots,s_{n-1},s_n\}$ 送入到属性词嵌入

层中, 该层的输出是一个属性词向量序列 $E_{\text{aspect}}=\{e_{s_1},e_{s_2},\ldots,e_{s_{n-1}},e_{s_n}\}$, 其中 e_{s_i} 是属性词 s_i 的词向量表示. 最后, 将两个嵌入层的输出向量按元素相加, 得到了一个融合了属性词信息的用户嵌入向量序列 $E_u=\{e_{u_1},e_{u_2},\ldots,e_{u_{n-1}},e_{u_n}\}$, 其中计算公式如下:

$$e_{u_i}=e_{w_i}+e_{s_i} \tag{2}$$

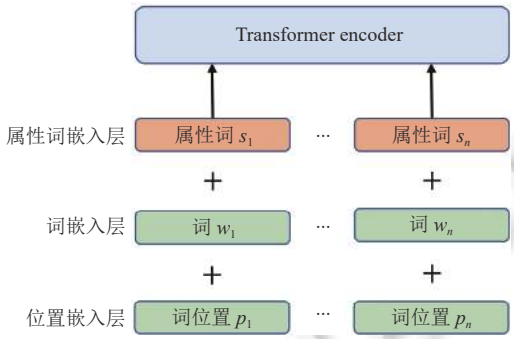


图 5 模型嵌入层

这个嵌入向量序列 E_u 不仅包含了原始的用户评论信息, 还融合了用户历史评论中的外部领域知识. 用户历史评论中的属性词通常会反映用户的习惯、服务品质的要求、商家类型偏好等信息. 通过将这些属性词融入表示向量序列中, 使模型更好捕捉和理解用户的需求与偏好.

2.2.2 商家信息的表示

本文获取了其他用户过去对目标商家的评论文本, 这些评论信息将作为商家特征向量的基础. 具体地, 本文将商家历史评论词序列 $W_b=\{w_1,w_2,\ldots,w_{n-1},w_n\}$ 送入到词嵌入层中, 该层的输出是一个词向量序列 $E_{\text{word}}=\{e_{w_1},e_{w_2},\ldots,e_{w_{n-1}},e_{w_n}\}$, 其中 e_{w_i} 是词 w_i 的词向量表示. 然后, 将属性词提示序列 $S_b=\{s_1,s_2,\ldots,s_{n-1},s_n\}$ 送入到属性词嵌入层中, 该层的输出是一个属性词向量序列 $E_{\text{aspect}}=\{e_{s_1},e_{s_2},\ldots,e_{s_{n-1}},e_{s_n}\}$, 其中 e_{s_i} 是属性词 s_i 的词向量表示. 最后, 将两个嵌入层的输出向量按元素相加, 得到了一个融合了属性词信息的商家嵌入向量序列 $E_b=\{e_{b_1},e_{b_2},\ldots,e_{b_{n-1}},e_{b_n}\}$, 其中计算公式如下:

$$e_{b_i}=e_{w_i}+e_{s_i} \tag{3}$$

这个嵌入向量序列 E_b 不仅包含了原始的商家评论信息, 还融合了商家历史评论中的外部领域知识. 商家历史评论中的属性词通常会反映商家的环境、特色美食、服务质量等信息. 通过将这些属性词融入表示向量序列中, 让模型在预测评分时能充分考虑商家的优点与特色.

2.2.3 基本信息的表示

为了让模型更为准确地预测评分, 本文从用户的基本信息中选取关键信息, 其中包含用户的历史平均打分、常去的商家类型、获得的点赞数量等信息. 此外, 还从商家的基本信息中选取关键信息, 其中包含商家的平均打分, 商家的类型 (如中餐厅, 海鲜餐厅等), 地理位置等信息. 类似地, 本文将用户和商家的基本信息词序列 $W_c=\{w_1,w_2,\ldots,w_{n-1},w_n\}$ 送入到词嵌入层中, 该层的输出是一个词向量序列 $E_{\text{word}}=\{e_{w_1},e_{w_2},\ldots,e_{w_{n-1}},e_{w_n}\}$, 其中 e_{w_i} 是词 w_i 的词向量表示. 需要注意的是, 由于仅从历史评论中提取属性词, 因此基本信息的属性词提示序列 S_i 由全 0 组成, 输出是一个属性词向量序列 $E_{\text{aspect}}=\{e_{s_1},e_{s_2},\ldots,e_{s_{n-1}},e_{s_n}\}$. 最后, 将两个嵌入层的输出向量按元素相加, 得到了一个基本信息的嵌入向量序列 $E_c=\{e_{c_1},e_{c_2},\ldots,e_{c_{n-1}},e_{c_n}\}$, 其中计算公式如下:

$$e_{c_i}=e_{w_i}+e_{s_i} \tag{4}$$

2.2.4 信息的编码

在获取到用户、商家以及基本信息的嵌入向量序列后, 本文对这些信息进行编码. 具体地, 首先沿着序列长度这一维进行拼接, 形成一个新的嵌入向量序列 $E=\{e_1,\ldots,e_n,e_{n+1},\ldots,e_{m+n},e_{m+n+1},\ldots,e_{m+n+l}\}$, 其中, n 是用户评论的长

度, m 是商家评论的长度, l 是基本信息的长度, 即:

$$E = Concat(E_u, E_B, E_C) \tag{5}$$

下面将拼接后的嵌入向量 E 与位置向量 E_{pos} 相加, 之后送入 T5 的编码器中进行特征提取抽象表示, 从而生成一个全局语义表示向量 H_{global}, 即:

$$H_{\text{global}} = Encoder(E + E_{\text{pos}}) \tag{6}$$

其中, E_{pos} 是 T5 模型自带的词的位置向量表示. *Encoder* 是标准 Transformer 中的一个模块, 里面包含一个多头注意力层 (multi-head attention layer) 和一个前馈神经网络层 (feed forward network), 在这些层之后都添加了残差连接 (residual connection) 与层归一化 (LayerNorm). 通过编码器的处理, 全局语义表示向量 H_{global} 不仅捕捉到了用户和商家的历史评论和基本信息, 还获取了用户和商家之间的关联信息. 之后, 本文将对该语义表示向量进行基于注意力机制的信息融合, 进一步强化上述捕获的信息.

2.3 基于注意力机制的信息融合

本文提出了一种基于注意力机制的信息融合策略. 如图 6 所示, 本文在原模型的 encoder 和 decoder 之间添加了一个信息融合层, 从而实现对输入文本的全局和局部语义信息的有效融合.

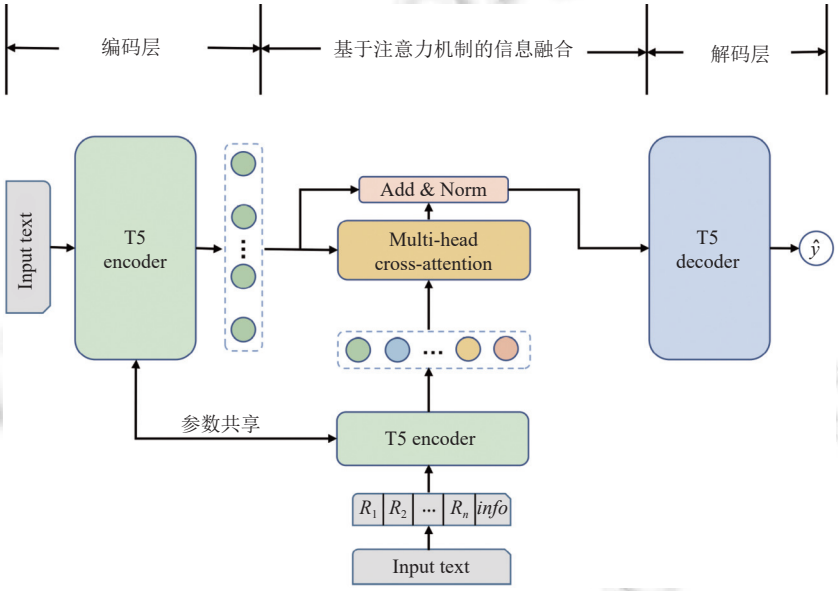


图 6 基于注意力机制的信息融合示意图

具体地, 首先将输入文本按照分隔符拆分出多条单独评论以及基本信息 $\{r_1, r_2, \ldots, r_{n-1}, r_n, info\}$, 随后使用 T5 的编码器分别对每条单独评论进行编码, 得到相应的评论表示向量 h_i. 此外, 本文对基本信息也进行了编码, 进而得到了由评论以及基本信息组成的表示向量序列 $\{h_1, h_2, \ldots, h_n, h_{info}\}$, 其中 h_{info} 为基本信息的表示向量, n 为评论个数. 最后沿着序列长度这一维度对表示向量序列进行拼接, 获得了输入文本的局部语义表示向量 H_{local}. 为了融合输入文本的全局与局部语义信息, 本文利用拥有 k 个注意力头的多头交叉注意力层来学习两种语义表示向量之间的交互.

下面介绍多头交叉注意力层的处理流程, 对于每个注意力头, 首先将全局语义表示投影到 Query 空间, 将局部语义表示投影到 Key、Value 空间中:

$$Q_i^{\text{global}} = H_{\text{global}} W_i^q, \ [K_i^{\text{local}}; V_i^{\text{local}}] = H_{\text{local}} [W_i^k; W_i^v] \tag{7}$$

其中, $\{W_i^q, W_i^k, W_i^v\} \in \mathbb{R}^{d \times d}$ 是可训练的参数, $Q_i^{\text{global}} \in \mathbb{R}^{l_1 \times d}$, $\{K_i^{\text{local}}, V_i^{\text{local}}\} \in \mathbb{R}^{l_2 \times d}$ 分别是 Query、Key 和 Value 的表示,

l_1 和 l_2 表示序列长度. 然后基于 Query 和 Key 的表示计算注意力分数 W_i^{att}, 之后与 Value 的表示相乘得到单个注意力头的输出.

$$W_i^{\text{att}} = \sigma\left(\frac{Q_i^{\text{global}}(K_i^{\text{local}})^{\text{T}}}{\sqrt{d_k}}\right) \tag{8}$$

$$H_i^{\text{head}} = W_i^{\text{att}} V_i^{\text{local}} \tag{9}$$

其中, $\sigma(\cdot)$ 是一个 Softmax 函数, d_k 是键-值向量的维度. 最后使用输出权重矩阵 W^o 来融合每个注意力头的输出, 最终的多头交叉注意力函数如下:

$$Cross\text{-}attention(H_{\text{global}}, H_{\text{local}}) = [H_1^{\text{head}}; \ldots; H_k^{\text{head}}] W^o \tag{10}$$

与标准的 Transformer 结构类似, 在该层后进行了残差链接和层归一化, 即:

$$Z = LN(H_{\text{global}} + Cross\text{-}attention(H_{\text{global}}, H_{\text{local}})) \tag{11}$$

其中, $LN(\cdot)$ 表示层归一化. 最后将生成的新语义表示向量 Z 送入到 *Decoder* 中进行解码.

2.4 基于生成模型的解码

相较于传统的分类模型 (仅具有 encoder), 本文使用生成模型 (Encoder-Decoder 结构) 将语义表示向量解码成用户对商家的打分, 利用解码器的注意力机制捕捉复杂的文本特征. 在获取到信息融合策略处理后的语义表示向量 Z 后, 将向量 Z 作为输入, 送入到 T5 的解码器中进行解码. 解码器的输出是一个概率分布, 该分布对应着预测评分的概率, 公式如下:

$$P(rating \mid Z) = \sigma(Decoder(Z)) \tag{12}$$

其中, *Decoder* 是标准 Transformer 中的一个模块. $P(rating \mid Z)$ 表示在给定语义向量 Z 的条件下, 用户对商家打分的概率分布. $\sigma(\cdot)$ 为 Softmax 函数, 其将解码器的输出转成概率分布. 为了获得模型最终的预测分数, 本文选择概率最大的评分作为预测结果:

$$\hat{y} = \text{argmax}(P(rating \mid Z)) \tag{13}$$

其中, $\hat{y}$ 是模型的预测结果. 通过上述步骤预测了用户对商家的打分, 虽然相比于分类模型, 本文模型解码过程更为复杂, 但能够更好地捕捉与评分相关的特征信息, 通过注意力机制更好地理解文本中的语义信息, 从而提升评分预测的准确性.

2.5 预训练任务

在训练模型之前, 本文为模型设计了一项预训练任务. 该预训练任务旨在让模型根据用户和商家的历史评论生成对评分预测有帮助的基本信息, 其中包含用户的平均打分、商家的类型、名字、城市以及地址. 在该预训练中, 为了生成用户和商家的基本信息, 模型需要学习如何获取评论文本中的关键信息, 提高对文本的理解能力.

具体地, 本文将随机选取用户和商家的各两条历史评论作为输入, 其对应的基本信息作为输出. 本文设定的预训练目标可以让模型学习如何从历史评论中提取到与基本信息相关的信息, 同时捕捉到它们之间的关联性, 这对本文任务很有帮助. 在经过预训练之后, 模型能够关注到更多历史评论和基本信息之间的上下文关系, 这样即使遇到陌生数据样本, 模型也能拥有不错的预测性能. 此外, 在进行预训练之后, 模型会拥有一个较好的初始参数, 从而可以更快收敛.

为了预训练模型, 本文使用交叉熵函数作为训练损失. 目标是给定评论集 R 最大化用户和商家的基本信息 I 的概率. 因此, 训练损失计算公式如下:

$$L = -\frac{1}{|\tau|}\sum_{(R,I)\in\tau} \log p(I \mid R, \theta) \tag{14}$$

其中, (R, I) 是训练集合 τ 中的一对, θ 是模型的参数, 并且:

$$\log p(I \mid R,\theta) = \sum_{k=1}^{m} \log p(i_k \mid i_1, i_2, \ldots, i_{k-1}, R;\theta) \tag{15}$$

其中, $\log p(i_k \mid i_1, i_2, \ldots, i_{k-1}, R;\theta)$ 被解码器所计算.

3 实验分析

3.1 数据集

Yelp 是美国非常受欢迎的评论网站, 涵盖了各种类型的商家, 如餐厅、酒店等. 用户可以在该网站上对商家进行评论和打分. 本文从该网站上获得了他们提供的最新数据集, 其中包含了 3 种类型的数据: Business (商家)、User (用户)、Review (评论). Business 数据集包含了商家的基本信息, 如名字、地址、类型等. User 数据集包含了用户的统计数据, 如注册时间, 被赞次数等. Review 数据集包含了评论文本和对应的评分, 其中评分是 1–5 的整数.

为了构建所需的数据集, 本文从数据集中提取商家-用户对, 并制定了以下提取规则: 每个商家-用户对, 其中商家至少拥有 80 条历史评论, 用户至少拥有 30 条历史评论. 此外, 为了确保用于训练的历史评论的质量, 每一条历史评论至少获得其他 6 位以上的其他用户的赞同. 之后, 挑选符合条件的商家-用户对制作本文的任务数据集. 对于数据集中每一条数据样本, 本文将从商家-用户对中随机选取两条商家的历史评论和两条用户的历史评论作为模型的输入, 用户对该商家的评分作为模型输出的标准答案. 此外, 本文还从用户和商家的基本信息中选取了平均打分、类型、地址等重要信息作为任务的额外信息. 考虑到模型输入上限问题, 本文对每条历史评论文本进行了截断, 将评论文本的长度限制在 120 词. 此外, 为了防止模型在训练过程中提前学习到用户的相关信息, 本文确保了训练集和测试集中需要预测的对象不是同一用户. 需要注意的是, 为了避免数据不平衡带来的模型性能下降, 因此还确保了每种评分的样本数量一致. 最终, 本文将制作好的数据集划分成 3 个部分, 分别为训练集、开发集和测试集, 其大小分别为 10 000、500 和 500. 表 2 展示了数据集的统计数据.

表 2 数据集的统计数据

类别	数量
商家数	18 514
用户数	307 842
评论数	279 476
平均用户历史评论数	49.1
平均商家历史评论数	239.4
数据样本总数	11 000

3.2 实验设置及评价指标

本文使用 T5-Base 作为基础模型进行实验. 为了微调超参数采用了 Adam[43]优化器, 设置动量 β 为 0.1, 权重衰退为 0.01. 本文的学习率设置为 0.000 1, batch 大小设置为 4, 共进行 20 轮训练.

本文在实验中采用了两种自动评价指标, 均方误差 (*MSE*) 和准确率 (*ACC*). 准确率通常用于计算模型完全预测正确的比例, 准确率的值越高代表模型性能越好. 然而, 这一指标并不能充分反映模型的预测性能, 为了更加全面地评估模型的预测性能, 本文还采用了均方误差作为评价指标. 均方误差常被用于衡量两个值之间的差异程度, 均方误差的值越小代表模型性能越好, 其计算公式如下:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{16}$$

其中, N 是样本数量, $\hat{y}_i$ 是模型预测值, y_i 是真实评分.

3.3 基准模型

为了评估在评分预测任务上的表现, 本文选取了几种基准模型进行了对比, 其中包含随机评分、平均分、

HabNet[44]、BiLSTM+Attn[45]、BART-Base[46]、SentiLARE[47]和 SentiWSP[48]等. 下面是这些基准模型的简要介绍.

● 随机评分: 随机选取 1–5 的整数作为最终的预测.

● 平均分: 使用商家和用户的平均打分直接作为最终的预测.

● UCF: 它是一种利用评分计算用户相似度, 并利用相似用户对商家的评分进行评分预测的算法.

● ICF: 它是一种利用评分计算商家相似度, 并利用用户对相似商家的评分进行评分预测的算法.

● SVD: 它是一种推荐系统中经典的基于矩阵分解的评分预测算法, 通过对评分矩阵进行矩阵分解降低数据噪声、缓解数据稀疏问题.

● HabNet: 它是一种用于评论评分预测和推荐的分层双向自注意力网络框架, 为了利用评论文本的分层结构, 设计了 3 种级别的编码器对文本进行编码, 之后利用注意力机制对文本语义信息进行整合, 最后进行评分预测.

● BiLSTM+Attn: 它是由前向 LSTM 和后向 LSTM 组成的, 可以更好捕捉到双向的语义依赖. BiLSTM-Attn 在 LSTM 输出向量的合成方面做出创新, 不再简单相加或者取平均, 而是使用注意力机制将他们融合, 提高了模型的预测性能.

● BERT: 它是由 Google 发布的一种预训练模型, 在无监督方式下使用大量无标注文本, 并经过 MLM 任务和 NSP 任务的预训练, BERT 获得了强大的语言表征和特征提取能力, 在文本分类、命名实体识别、问答系统等任务上取得了不错的效果.

● BART: 它是一种基于 Transformer 架构的神经网络模型, 它被广泛应用于多种自然语言处理任务中, 如文本生成、翻译和摘要等.

● T5: 它是一种多任务语言模型. 使用大规模的文本数据进行预训练, 学习通用的语言表示, 希望将所有的自然语言处理任务都转化为文本到文本的任务. 因此, 在多个领域和任务中取得优秀的性能.

● SentiLARE: 它是一种预训练语言模型, 使用 SentiWordNet 为每个词导出上下文感知的情感极性, 并采用一种标签感知的掩码预训练任务来学习构建情感感知的语言表示.

● SentiWSP: 它是一种情感感知的预训练语言模型, 通过词级和句子级的预训练, 增强模型捕获情感信息的能力. 在句子级和方面级情感分类任务中取得了最好的结果.

● LLaMA: 在 LLaMA[49]模型的基础上进行微调, 利用低秩自适应 (low-rank adaption) 技术, 冻结预训练好的模型权重参数, 仅训练低秩矩阵, 让微调质量与全模型微调相当.

3.4 实验结果与分析

从表 3 中的实验结果可以发现 (加粗数据为最优结果), 基于统计的方法在 *ACC* 指标上的表现普遍较差. 随机评分作为最基本的基准模型, 性能最差, 这说明随机评分在该任务上并不可靠. 与随机评分相比, 基于平均分的方法表现更为出色, 它们在 *MSE* 指标上达到了基准模型中最好的表现. Adomavicius 等人[50]发现, 面对相同品质的商家, 顾客们倾向于给平均打分更高的商家一个较好的评分. 如果商家缺乏特别之处, 那么用户会参考商家的平均打分或者给出一个自己经常打的分数. 因此, 基于平均分的方法简单且有效, 但在预测准确性方面还有改进的空间. 由于存在数据稀疏问题, 传统的推荐系统方法在该任务上的表现也不理想, 评分矩阵中有很多缺失的评分, 因此基于用户/商家的协同过滤方法在两个指标上的表现普遍较差. SVD 算法将评分矩阵进行分解, 转化为更低维的矩阵表示. 一方面, 低维表示可以更好地捕捉用户和商家的潜在特征, 另一方面, 评分矩阵中缺失的评分会被近似地预测出来, 从而缓解数据稀疏问题, 因此 SVD 算法在预测准确度上有显著提升.

相较于仅基于评分的方法, 神经网络模型 (例如: BiLSTM-Attn 和 HabNet) 结合了注意力机制, 能够更好地理解用户和商家历史评论之间的语义信息和上下文关系, 在预测准确性方面有了进一步的提升, 这证明了神经网络在处理序列数据时具有一定的优势.

预训练模型 (例如: BERT、T5 和 SentiLARE 等) 在两个评价指标上均高于 BiLSTM-Attn 模型. 一方面, 尽管它们在模型中都应用了注意力机制, 但预训练模型利用大量无标注数据进行预训练, 提前学习了丰富的语义知识, 从而可以更好地捕捉文本中的语义信息. 另一方面, 本文任务的输入内容较多, 而 Transformer 结构完全基于注意

力机制, 因此它可以进一步缓解长距离依赖问题, 更好地处理长文本内容, 进而增强模型在评分生成任务上的效果.

值得注意的是, 拥有 70 亿参数量的大预训练模型 LLaMA 在 *MSE* 和 *ACC* 指标上均取得了较好的性能. 这是由于大模型在大规模数据集上可以学习到广泛的语义知识和语言模式, 这将有助于提升模型的上下文理解能力. 此外, 更多的参数意味着模型拥有更好的表示能力, 可以进一步拟合数据的复杂模式与特征, 从而提升模型的性能.

表 3 对比实验结果

模型	对比项	*MSE*	*ACC* (%)
统计方法	随机评分	3.79	21.4
	用户平均分	1.84	22.0
	商家平均分	1.80	25.2
推荐系统方法	UCF	3.02	20.0
	ICF	3.01	20.6
	SVD	3.12	28.4
	HabNet	2.85	34.4
神经网络	BiLSTM+Attn	3.10	30.2
预训练模型	BART-Base	2.60	34.6
	BERT-Base	2.43	31.0
	T5-Base	2.17	32.0
	SentiLARE	2.52	34.4
	SentiWSP	2.10	31.8
大模型	LLaMA	2.41	36.6
	Ours	**1.52**	**46.2**

本文的模型的表现超越所有基准模型, 这表明模型中引入的属性词嵌入层、基于注意力的信息融合策略以及设计的预训练任务使得模型能够更为准确地预测用户的评分.

3.5 不同影响因素比较

为了进一步了解模型中各模块的重要性以及作用, 本文对属性词嵌入层、基本信息、信息融合和预训练任务进行了消融实验. 从表 4 中可以发现, 每一个模块在模型中都扮演了重要的角色.

表 4 消融实验结果

属性词嵌入层	基本信息	信息融合	预训练任务	*MSE*	*ACC* (%)
×	×	×	×	2.17	32.0
√	—	—	—	2.20	34.6
√	√	—	—	1.62	42.6
√	√	√	—	1.65	44.0
√	√	√	√	**1.52**	**46.2**

根据表 4 中的实验结果, 可以得出以下结论.

第一, 添加属性词嵌入层能够提高模型预测精确度方面的性能, 这说明本文将外部领域知识融入模型中的思路是正确的. 通过将属性词的判别信息融入每个词的语义表示中去, 能够帮助模型识别出输入文本中的属性词, 并提高这些涉及用户和商家关键方面的词的关注度. 此外, 在面对陌生属性词时, 外部领域知识也能帮助模型识别, 从而进行较为准确的评分预测.

第二, 在输入文本中加入用户和商家的基本信息带来的提升最大, 这说明基本信息对于模型预测评分是很有帮助的. 通过基本信息能够让模型了解到用户的偏好、打分习惯以及商家的整体情况, 这些信息在一定程度上可以帮助模型计算用户与商家之间的匹配程度, 并预测用户可能的用餐体验, 显然, 这可以有效地帮助模型生成出较为精确的用户评分.

第三, 采用基于注意力机制的信息融合策略可以取得进一步的提升, 这说明本文设计的信息融合方法是有意

义的. 在输入文本内容较多的情况下, 初始位置的内容影响力有限, 模型容易聚焦于当前位置的信息, 而遗忘远端的内容, 难以建立长距离依赖关系. 本文将输入文本拆分成多条单独文本进行编码, 从而获取到输入文本的局部语义表示, 这一过程显著降低了输入序列的长度, 能有效缓解长距离依赖问题. 此外, 为了让局部信息与全局信息进行交互, 本文设计了一个多头注意力层来用于输入文本的全局与局部语义信息的交互. 这种策略能够提供输入文本的多维度语义信息, 从而有效地增强了模型预测评分的效果.

第四, 模型的预训练也提高了预测性能, 这说明本文设计的预训练任务是有效的. 通过学习由用户和商家的历史评论生成用户和商家基本信息, 这能鼓励模型关注评论文本中的关键信息, 提高对文本中关键信息的捕捉能力, 在之后的评分预测任务中, 能够更好地捕获到有用信息, 从而准确地预测用户评分.

3.6 历史评论的影响

为了充分探究历史评论对模型性能的影响, 本文对多种用户和商家历史评论组合进行了性能评测, 实验结果可见表 5.

表 5 历史评论条数的影响

商家历史评论 (条)	用户历史评论 (条)	*MSE*	*ACC* (%)
1	0	3.11	29.6
0	1	3.19	27.6
2	0	2.83	30.8
0	2	3.02	31.0
1	1	3.64	31.0
1	2	2.93	31.2
2	1	**2.13**	31.0
2	2	2.17	**32.0**

从表 5 的实验结果中可以得到如下结论: 第一, 通过前 4 个实验结果可以发现, 仅利用单一类型的历史评论并不能达到较好的效果, 它们拥有一个较高的准确率, 但是对于错误样例, 模型的预测与真实评分之间的差异较大, 即 *MSE* 值较大. 第二, 通过观察多个保持一类历史评论数量不变, 增加另一类历史评论的数量的实验, 从而发现商家历史评论相较于用户历史评论能够带来更多的性能提升. 这可能是因为商家历史评论能够提供更多商家的细节, 例如装修风格、菜品质量、服务态度等, 然而用户历史评论提供的信息更多集中于用户本身的饮食习惯等, 显然前者能更好地帮助模型预测评分. 第三, 通过对比仅包含单一类型的历史评论和包含两种类型的历史评论的实验结果可知, 同时利用两种类型的历史评论能够进一步提升模型的性能. 这归因于用户历史评论能够提供个人偏好、口味等信息, 商家历史评论则能提供商家的特点、质量等方面的信息, 两种类型的历史评论可以让模型多角度地分析与预测用户的评分.

3.7 实例分析

本文选择了在 *MSE* 评价指标上表现较好的基准模型 BERT 作为对比模型, 对图 7 中的实例进行分析.

● 如实例 1 所示, 在商家的历史评论中, 用户普遍给出了正面的评价, 这表明商家为用户提供了优质的服务. 在用户的历史评论中, 都是一些负面的评论. 但需要注意的是, 这些是用户对其他商家的评论. 然而, BERT 模型受到用户历史评论的影响, 错误地认为用户对该商家不满意, 最终给出的评分是 2 分. 本文的模型从基本信息中发现用户和商家的平均打分都相对较高, 同时商家类型符合用户的偏好, 因此可以初步判断用户对商家的评价将是积极的. 通过结合其他用户的评论, 最终本文的模型没有受到无关负面评论的影响, 准确预测了用户的评分.

● 如实例 2 所示, 从商家历史评论中可以发现, 用户对商家的评价褒贬不一. 如果无法从用户历史评论中挖掘出用户的习惯, 将会很难做出正确的预测, 因此 BERT 模型给出了错误的评分. 本文的模型通过分析用户的历史评论发现, 用户缺乏足够的耐心, 不喜欢等待. 恰好在商家历史评论中, 有人提到上菜速度较慢, 需要用户等待较长时间. 再结合基本信息, 用户的平均打分仅为 3 分, 这表明用户的要求相当高. 因此, 模型判断该商家无法满足用户的要求, 最终成功预测出了正确的评分.

本文通过引入外部领域知识, 让模型更精确地关注到用户和商家的各方面特征信息. 同时, 基于注意力机制的信息融合策略缓解了长距离依赖问题, 并提供了多维度的语义信息. 此外, 基本信息也为模型提供了必要的帮助. 上述这些改进让模型能够综合各种信息进行评分预测, 进而本文模型的性能普遍优于基准模型.

图 7　本文模型与基线模型 BERT 的实例分析图

4 总　结

本文提出了一个基于预训练模型的个性化评分预测方法, 该方法采用预训练模型 T5 作为基础模型, 并设计了一项预训练任务, 从而提升模型对文本中关键信息的提取能力. 在此基础上通过细粒度情感分析以及设计的属性词嵌入层, 将外部领域知识融入模型中, 进而使模型能有效地捕捉到评论文本中的属性词. 接下来, 采用了一种基于注意力机制的信息融合策略, 将模型的输入文本进行拆分并分别编码, 随后将编码后的表示向量进行拼接作为输入文本的局部语义表示, 最后利用注意力机制有效地融合输入文本的全局和局部语义表示. 在自制数据集上, 本文进行了对比实验与消融实验. 实验的结果表明, 相较于基准模型, 本文提出的模型在两个评价指标上均实现了显著提升, 模型各组件都扮演了重要的角色, 提供了正向的改进. 本文未来的工作方向是, 目前获取到了评论中的关键方面, 但是默认了每个方面的重要性是一致的, 但事实上, 用户对于某些方面十分看重. 所以下一步希望让模型学会合理分配各个方面的注意力, 从而更加准确地生成个性化评分. 此外, 近年来大模型通过对巨大语料库的学习, 已经展现出其对语言结构和含义的深度理解能力. 这使得它们能够更好地理解用户和商家的评论内容, 从而更准确地预测出用户对商家的态度和可能给出的评分. 因此, 未来将致力于研究大模型在该任务上的表现, 尝试通过最新的微调技术 (如: CoT) 来提高大模型的性能.

References:

[1] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30–37. [doi: 10.1109/MC.2009.263]

[2] Pazzani MJ, Billsus D. Content-based recommendation systems. In: Brusilovsky P, Kobsa A, Nejdl W, eds. The Adaptive Web. Berlin: Springer, 2007. 325–341. [doi: 10.1007/978-3-540-72079-9_10]

[3] McAuley J, Leskovec J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In: Proc. of the 7th ACM Conf. on Recommender Systems. Hong Kong: ACM, 2013. 165–172. [doi: 10.1145/2507157.2507163]

[4] Koren Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Las Vegas: ACM, 2008. 426–434. [doi: 10.1145/1401890.1401944]

[5] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. The Journal of Machine Learning Research, 2003, 3: 993–1022.

[6] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2018. 4171–4186. [doi: 10.18653/v1/N19-1423]

[7] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou YQ, Li W, Liu PJ. Exploring the limits of transfer learning with a unified text-to-text Transformer. The Journal of Machine Learning Research, 2020, 21(1): 140.

[8] Zhu GH, Lu W, Yuan CF, Huang YH. AdaMCL: Adaptive fusion multi-view contrastive learning for collaborative filtering. In: Proc. of the 46th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Taipei: ACM, 2023. 1076–1085. [doi: 10.1145/3539618.3591632]

[9] Choi J, Hong S, Park N, Cho SB. Blurring-sharpening process models for collaborative filtering. In: Proc. of the 46th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Taipei: ACM, 2023. 1096–1106. [doi: 10.1145/3539618.3591645]

[10] Ren XB, Xia LH, Zhao JS, Yin DW, Huang C. Disentangled contrastive collaborative filtering. In: Proc. of the 46th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Taipei: ACM, 2023. 1137–1146. [doi: 10.1145/3539618.3591665]

[11] Ganu G, Elhadad N, Marian A. Beyond the stars: Improving rating predictions using review text content. In: Proc. of the 12th Int'l Workshop on the Web and Databases. Providence: WebDB, 2009. 1–6.

[12] Zhang W, Yuan Q, Han JW, Wang JY. Collaborative multi-level embedding learning from reviews for rating prediction. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence. New York: AAAI, 2016. 2986–2992.

[13] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv:1301.3781, 2013.

[14] Zheng L, Noroozi V, Yu PS. Joint deep modeling of users and items using reviews for recommendation. In: Proc. of the 10th ACM Int'l Conf. on Web Search and Data Mining. Cambridge: ACM, 2017. 425–434. [doi: 10.1145/3018661.3018665]

[15] Catherine R, Cohen W. TransNets: Learning to transform for recommendation. In: Proc. of the 11th ACM Conf. on Recommender Systems. Como: ACM, 2017. 288–296. [doi: 10.1145/3109859.3109878]

[16] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

[17] Shuai J, Zhang K, Wu L, Sun PJ, Hong RC, Wang M, Li Y. A review-aware graph contrastive learning framework for recommendation. In: Proc. of the 45th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Madrid: ACM, 2022. 1283–1293. [doi: 10.1145/3477495.3531927]

[18] Yang W, Huo TF, Liu ZQ, Lu C. Review-based multi-intention contrastive learning for recommendation. In: Proc. of the 46th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Taipei: ACM, 2023. 2339–2343. [doi: 10.1145/3539618.3592053]

[19] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: Proc. of the 2002 Conf. on Empirical Methods in Natural Language Processing. ACL, 2002. 79–86. [doi: 10.3115/1118693.1118704]

[20] Yang ZC, Yang DY, Dyer C, He XD, Smola A, Hovy E. Hierarchical attention networks for document classification. In: Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: ACL, 2016. 1480–1489. [doi: 10.18653/v1/N16-1174]

[21] Wang JQ, Gong ZH, Xue Y, Pang SG, Gu DH. Aspect-based sentiment analysis based on hybrid multi-head attention and capsule networks. Journal of Chinese Information Processing, 2020, 34(5): 100–110 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2020.05.014]

[22] Zeng F, Zeng BQ, Han XL, Zhang M, Shang Q. Double attention neural network for aspect-based sentiment analysis. Journal of Chinese Information Processing, 2019, 33(6): 108–115 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2019.06.016]

[23] Feng C, Li HH, Zhao HY, Xue Y, Tang JY. Aspect-level sentiment analysis based on hierarchical attention and gate networks. Journal of Chinese Information Processing, 2021, 35(10): 128–136 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2021.10.015]

[24] Tang DY, Qin B, Feng XC, Liu T. Effective LSTMs for target-dependent sentiment classification. In: Proc. of the 26th Int'l Conf. on Computational Linguistics: Technical Papers. Osaka: ACL, 2015. 3298–3307.

[25] Song YW, Wang JH, Jiang T, Liu ZY, Rao YH. Targeted sentiment classification with attentional encoder network. In: Proc. of the 28th Int'l Conf. on Artificial Neural Networks. Munich: Springer, 2019. 93–103. [doi: 10.1007/978-3-030-30490-4_9]

[26] Mao Y, Shen Y, Yu C, Cai LJ. A joint training dual-MRC framework for aspect based sentiment analysis. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. AAAI, 2021. 13543–13551. [doi: 10.1609/aaai.v35i15.17597]

[27] Zeng BQ, Yang H, Xu RY, Zhou W, Han XL. LCF: A local context focus mechanism for aspect-based sentiment classification. Applied Sciences, 2019, 9(16): 3389. [doi: 10.3390/app9163389]

[28] Peng HY, Xu L, Bing LD, Huang F, Lu W, Si L. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 8600–8607. [doi: 10.1609/aaai.v34i05.6383]

[29] Wan H, Yang YF, Du JF, Liu YN, Qi KX, Pan JZ. Target-aspect-sentiment joint detection for aspect-based sentiment analysis. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 9122–9129. [doi: 10.1609/aaai.v34i05.6447]

[30] Cai HJ, Xia R, Yu JF. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). ACL, 2021. 340–350. [doi: 10.18653/v1/2021.acl-long.29]

[31] Chen H, Zhai ZP, Feng FX, Li RF, Wang XJ. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Dublin: ACL, 2022. 2974–2985. [doi: 10.18653/v1/2022.acl-long.212]

[32] Zhang WX, Deng Y, Li X, Yuan YF, Bing LD, Lam W. Aspect sentiment quad prediction as paraphrase generation. In: Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing. ACL, 2021. 9209–9219. [doi: 10.18653/v1/2021.emnlp-main.726]

[33] Bao XY, Wang ZQ, Jiang XT, Xiao R, Li SS. Aspect-based sentiment analysis with opinion tree generation. In: Proc. of the 31st Int'l Joint Conf. on Artificial Intelligence. Vienna: IJCAI, 2022. 4044–4050.

[34] Zhao SM, Chen W, Wang TJ. Learning cooperative interactions for multi-overlap aspect sentiment triplet extraction. In: Proc. of the 2022 Findings of the Association for Computational Linguistics (EMNLP 2022). Abu Dhabi: ACL, 2022. 3337–3347. [doi: 10.18653/v1/2022.findings-emnlp.243]

[35] Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics. Ann Arbor: ACL, 2005. 115–124. [doi: 10.3115/1219840.1219855]

[36] Qu LZ, Ifrim G, Weikum G. The bag-of-opinions method for review rating prediction from sparse text patterns. In: Proc. of the 23rd Int'l Conf. on Computational Linguistics (Coling 2010). Beijing: ACL, 2010. 913–921.

[37] Lu Y, Kong XF, Quan XJ, Liu WY, Xu YL. Exploring the sentiment strength of user reviews. In: Proc. of the 11th Int'l Conf. on Web-age Information Management. Jiuzhaigou: Springer, 2010. 471–482. [doi: 10.1007/978-3-642-14246-8_46]

[38] Li JJ, Yang HT, Zong CQ. Document-level multi-aspect sentiment classification by jointly modeling users, aspects, and overall ratings. In: Proc. of the 27th Int'l Conf. on Computational Linguistics. Santa Fe: ACL, 2018. 925–936.

[39] Wu CH, Wu FZ, Liu JX, Huang YF, Xie X. ARP: Aspect-aware neural review rating prediction. In: Proc. of the 28th ACM Int'l Conf. on Information and Knowledge Management. Beijing: ACM, 2019. 2169–2172. [doi: 10.1145/3357384.3358086]

[40] Bu JH, Ren L, Zheng S, Yang Y, Wang JG, Zhang FZ, Wu W. ASAP: A Chinese review dataset towards aspect category sentiment analysis and rating prediction. In: Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, 2021. 2069–2079. [doi: 10.18653/v1/2021.naacl-main.167]

[41] Hu MQ, Liu B. Mining opinion features in customer reviews. In: Proc. of the 19th National Conf. on Artificial Intelligence. San Jose: AAAI. 2004. 755–760.

[42] Dai HL, Song YQ. Neural aspect and opinion term extraction with mined rules as weak supervision. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 5268–5277. [doi: 10.18653/v1/P19-1520]

[43] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Proc. of the 3rd Int'l Conf. on Learning Representations. 2015.

[44] Deng ZF, Peng H, Xia CY, Li JX, He LF, Yu P. Hierarchical bi-directional self-attention networks for paper review rating recommendation. In: Proc. of the 28th Int'l Conf. on Computational Linguistics. Barcelona: ACL, 2020. 6302–6314. [doi: 10.18653/v1/2020.coling-main.555]

[45] Zhou P, Shi W, Tian J, Qi ZY, Li BC, Hao HW, Xu B. Attention-based bidirectional long short-term memory networks for relation classification. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 2: Short Papers). Berlin: ACL, 2016. 207–212. [doi: 10.18653/v1/P16-2034]

[46] Lewis M, Liu YH, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 7871–7880. [doi: 10.18653/v1/2020.acl-main.703]

[47] Ke P, Ji HZ, Liu SY, Zhu XY, Huang ML. SentiLARE: Sentiment-aware language representation learning with linguistic knowledge. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). ACL, 2020. 6975–6988. [doi: 10.18653/v1/2020.emnlp-main.567]

[48] Fan S, Lin C, Li HN, Lin ZH, Su JS, Zhang H, Gong YY, Guo J, Duan N. Sentiment-aware word and sentence level pre-training for sentiment analysis. In: Proc. of the 2022 Conf. on Empirical Methods in Natural Language Processing. Abu Dhabi: ACL, 2022. 4984–4994. [doi: 10.18653/v1/2022.emnlp-main.332]

[49] Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, Rodriguez A, Joulin A, Grave E, Lample G. LLaMA: Open and efficient foundation language models. arXiv:2302.13971, 2023.

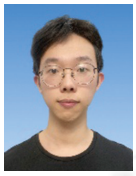
[50] Adomavicius G, Bockstedt JC, Curley SP, Zhang JJ. Understanding effects of personalized vs. aggregate ratings on user preferences. In: Proc. of the 2016 Joint Workshop on Interfaces and Human Decision Making for Recommender Systems Co-located with ACM Conf. on Recommender Systems. Boston: IntRS@RecSys. 2016. 14–21.

附中文参考文献:

[21] 王家乾, 龚子寒, 薛云, 庞士冠, 古东宏. 基于混合多头注意力和胶囊网络的特定目标情感分析. 中文信息学报, 2020, 34(5): 100–110. [doi: 10.3969/j.issn.1003-0077.2020.05.014]

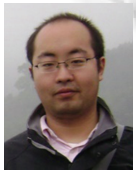
[22] 曾锋, 曾碧卿, 韩旭丽, 张敏, 商齐. 基于双层注意力循环神经网络的方面级情感分析. 中文信息学报, 2019, 33(6): 108–115. [doi: 10.3969/j.issn.1003-0077.2019.06.016]

[23] 冯超, 黎海辉, 赵洪雅, 薛云, 唐婧尧. 基于层次注意力机制和门机制的属性级别情感分析. 中文信息学报, 2021, 35(10): 128–136. [doi: 10.3969/j.issn.1003-0077.2021.10.015]



强敏杰(2000－), 男, 硕士生, CCF 学生会员, 主要研究领域为自然语言处理.

周国栋(1967－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为自然语言处理.



王中卿(1987－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为自然语言处理.