用于 Airbnb 房价预测的多源信息学习框架

Lu Jiang[1], Yuanhan Li[1],Na Luo[1],JiananWang[2], *, Qiao Ning[3], *[1]
东北师范大学信息科学与技术学院，长春[2]东北师范大学物理学院，长春[3]大连海事大学信息科学与技术学院，大连{*jiangl761, liyh447, luon110, wangjn*}*@nenu.edu.cn, ningq669@dlmu.edu.cn* 通讯作者*

arXiv:2301.01222v1 [cs.LG] 1 Jan 2023

摘要——**随着技术的发展和共享经济的兴起，Airbnb 作为著名的短租平台，已成为许多年轻人的首选。Airbnb 的定价问题一直以来都是值得研究的课题。尽管以往的研究取得了令人鼓舞的成果，但仍存在一些有待解决的不足。例如：（1）出租房源的特征属性不够丰富；（2）对出租房源文本信息的研究不够深入；（3）很少有研究结合房屋周边的兴趣点（POI）来预测租赁价格。为应对上述挑战，我们提出一种多源信息嵌入（MSIE）模型，用于预测 Airbnb 的租赁价格。具体而言，首先，我们选取统计特征，对原始出租房源数据进行嵌入。其次，我们生成三种不同文本信息的词特征向量和情感得分组合，形成文本特征嵌入。第三，我们利用出租房屋周边的兴趣点（POI）信息生成多种空间网络图，并学习网络的嵌入表示，以获得空间特征嵌入。最后，本文将这三个模块结合起来，形成多源出租房源表示，并使用构建的全连接神经网络预测价格。对实验结果的分析表明，我们提出的模型是有效的。**

I. 引言

近年来，住宿共享系统已在越来越多的城市中兴起，因此产生了海量数据。Airbnb 是一个分享住房和体验的在线交易平台，但面临着定价混乱的问题。租客需要了解出租房屋的合理价格，以免受骗。房主则需要为短期出租房屋制定合理价格，以吸引更多客户。因此，Airbnb 房价预测在住宿共享系统中发挥着关键作用。然而，租客和房主数量的迅速增长，使传统的人工方法变得 [1]耗时且效率低下。计算方法在准确的 Airbnb 房价预测方面 [2]日益受到关注。

用于价格预测的计算方法主要分为两类：(1) 基于特征的方法 [3], 和 (2) 深度学习方法 [4, 5]。在基于特征的方法中，采用各种类型的特征提取策略，以提取与价格相关的租户和房主特征。基于特征的方法将价格预测转化为机器学习方法，例如支持向量机(SVM)和随机森林。例如，为了区别于传统的定价方法，Li 等选择粗糙集 (RS) 和 SVM 算法，在享乐价格 [6]的基础上建立了新的定价数学模型。PR Kalehbasti等提出了一种价格预测模型，利用机器学习、深度学习和自然语言处理技术，对出租房源的特征、房主特征以及客户评论 [7]进行嵌入。深度学习方法使用多层神经网络来映射输入特征与输出结果之间的相关性。例如，Chen 等应用自回归积分滑动平均模型生成基线，同时利用长短期记忆网络构建预测模型 [8]。

然而，基于特征的方法在研究 Airbnb 房价预测时，大多只考虑单一特征。随着表示学习的发展， [9, 10],空间嵌入 [11, 12] 日益受到关注。已有研究对与房价相关的统计特征、文本特征和空间特征进行建模，但尚无统一框架整合上述特征。针对上述不足，我们提出一种基于多源信息嵌入的预测模型，以研究 Airbnb 价格问题。主要贡献总结如下。

- 首先，为了获得最佳特征集，本文选取房屋自身的特征，以获取统计信息特征。
- 其次，本文将文本信息分为三类，并将房屋描述和房东介绍转换为特征矩阵。随后，将租客评论转换为每套房屋的情感得分。
- 然后，我们使用不同类型的兴趣点（POI）数据和房屋构建各种空间网络图，并学习其网络嵌入，以获取空间信息特征。
- 最后，我们将这三类特征嵌入整合为多资源住房特征作为输入，并使用本文构建的神经网络进行价格预测。我们通过两组真实数据验证了模型的有效性。

II. 预备知识

我们首先介绍一些关键定义和问题定义，然后概述所提出的方法。

A. 定义与问题陈述

定义 1. 统计特征 我们的模型构建的统计特征为 $\mathbf{S} = (s_1, s_2, ..., s_n)$，其中 s_i是经过预处理的房源特征，包括“房东加入时间”“房东是否为超赞房东”“验证信息”等。

定义 2. 文本特征 文本特征分为三类：房源描述、房东介绍和租客评价。我们将房源描述和房东介绍转换为特征向量 $\mathbf{L} = (l_1, l_2, ..., l_n)$ 和 $\mathbf{H} = (h_1, h_2, ..., h_n)$，并将租客评价转换为情感得分 $\mathbf{R} = (r_1, r_2, ..., r_n)$。因此，我们将文本特征定义为 $\mathbf{T} = (\mathbf{L}, \mathbf{H}, \mathbf{R})$。

定义 3. 空间特征 我们首先将每个出租房屋及其周围 *1,000* 米范围内的兴趣点组合成一个空间网络 $G = (V, E, W)$。然后，我们通过 *SDNE [13]*，学习网络嵌入，并得到空间特征矩阵 $\mathbf{P} = (p_1, p_2, ..., p_n)$。

定义 4. 问题陈述 本文研究 *Airbnb* 房价预测问题。我们将该问题表述为多源特征嵌入任务。形式化地说，我们旨在寻找一个映射函数 $f : (S, T, P) \rightarrow V$，该函数以统计特征 S、文本特征 T和空间特征 P 为输入，并输出统一向量化表示 V，用于预测具体的房源价格。

B. 框架概览

图 1 展示了多源特征嵌入的整体框架。具体而言，我们从三个方面对原始数据进行嵌入。(1) 对于统计特征嵌入，我们使用 Lasso交叉验证结合出租房屋特征来选择特征集。(2) 对于文本特征嵌入，我们将文本特征分为三类，包括房屋描述、房东介绍和租客评论。通过负采样连续词袋模型，将房屋描述和房东介绍转换为词特征向量；再基于朴素贝叶斯原理，使用贝叶斯模型将租客评论转换为情感得分，并将这些内容组合为文本特征。(3) 对于空间特征嵌入，我们收集不同类型的兴趣点，并将每套房屋及其周围 1,000 米范围内的兴趣点组合为空间网络。通过 SDNE模型学习空间特征。(4) 将三种不同特征组合为多源特征，并输入神经网络，以获得最终的租赁价格。

III. 多源信息学习

本节将介绍我们框架的核心架构，具体包括：（1）统计特征嵌入；（2）文本特征嵌入；（3）空间特征嵌入。

A. 统计特征嵌入

每套房屋的统计特征由一个 245 维向量表示，用于描述该房屋的房源信息，包括房源编号、房东编号、房东加入时间、房东回复率、房东是否为超赞房东、房东是否有头像、房东身份已验证、浴室数量、卧室、纬度、经度、可容纳人数、押金、包含的客人数、验证信息等。

我们使用 Lasso交叉验证进行特征集选择。损失函数定义如下:

$$\text{obj} = \frac{1}{2}\sum_{i=1}^{n}\left(\boldsymbol{y_i} - \boldsymbol{w}^T\boldsymbol{x_i}\right)^2 + \alpha\sum_{j=1}^{m}|w_i| \tag{1}$$

其中，n 为房屋数量，m 为参数数量，α为正则化系数，$\alpha\sum_{j=1}^{m}|w_i|$ 为 L1正则化项，y_i 为出租住房的租赁价格，x_i 为出租住房的统计特征，w 为出租住房特征的系数矩阵，$x_i = s_i$。统计特征矩阵$\mathbf{S} = (s_1, s_2, ..., s_n)$。Lasso交叉验证可以将不重要特征的系数压缩至 0，从而实现特征选择的目的，最终保留本文所需的重要统计特征集。

B. 文本特征嵌入

我们从原始数据中提取了三类文本数据，分别是房源描述、房东介绍和租客评价。房源描述主要介绍房屋的位置、周边环境、室内布局和住房规定等。房东介绍主要介绍房东的年龄、身高、职业、爱好和个性，而租客评价则表达租客对出租住房的感受以及对房东态度的评价。由于租客评价包含情感价值，我们采用两种不同的方法对文本特征进行建模。首先，我们使用 CBOW [14]模型嵌入房源描述和房东介绍的文本特征。我们选择预处理后的维基百科中文词库作为训练语料库 W，目标函数定义如下:

$$\mathcal{L} = \sum_{c\in W}\left\{\log\left[\sigma\left(x_c^T\theta^c\right)\right] + \sum_{u\in NEG(c)}\log\left[\sigma\left(-x_c^T\theta^u\right)\right]\right\} \tag{2}$$

然后，使用随机梯度上升法优化上述目标函数，得到:

$$\mathcal{L}(c,u) = L^c(u)\log\left[\sigma\left(x_c^T\theta^u\right)\right] + [1 - L^c(u)]\log\left[1 - \sigma\left(x_c^T\theta^u\right)\right] \tag{3}$$

然后计算 $\mathcal{L}(c, u)$ 的梯度，得到:

$$\boldsymbol{v}(\tilde{c}) := \boldsymbol{v}(\tilde{c}) + \eta\sum_{u\in\{c\}\cup NEG(c)}\frac{\partial\mathcal{L}(c,u)}{\partial x_c} \tag{4}$$

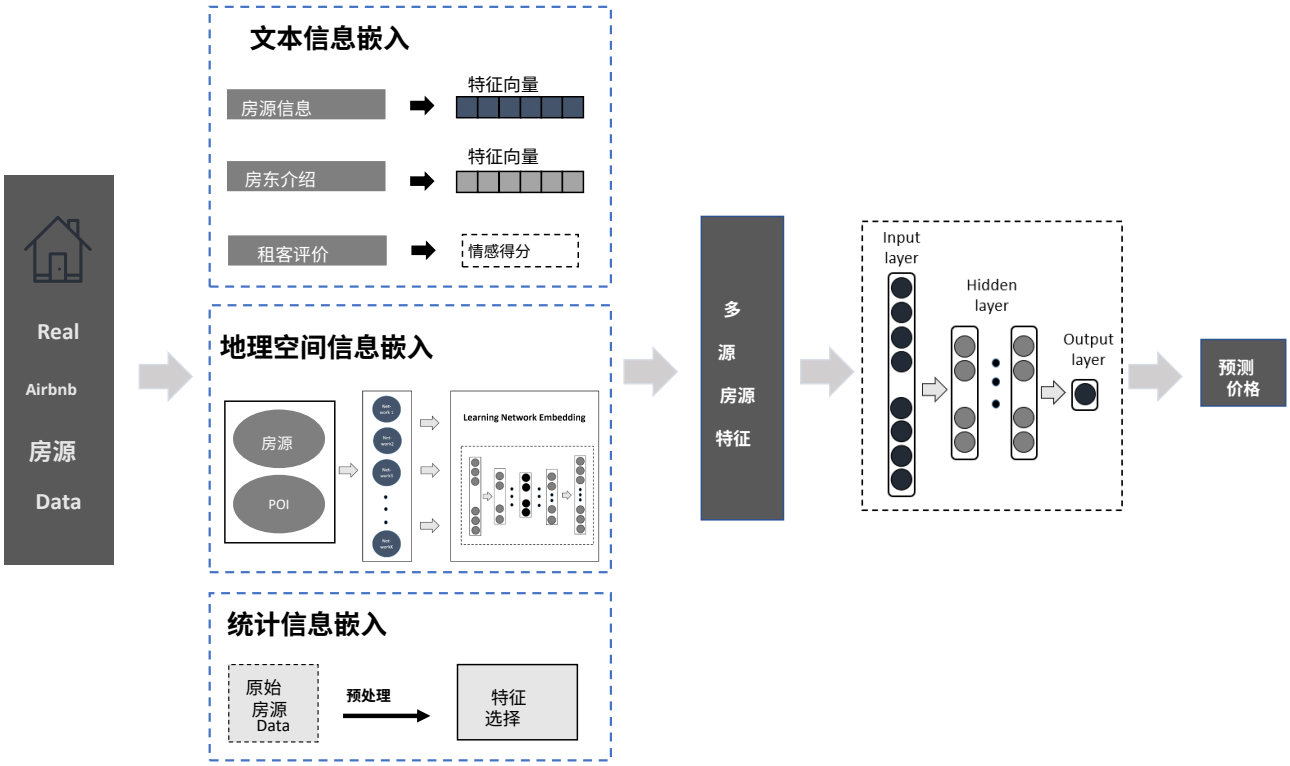


图1：框架概览。

本文将词向量的维度设为 100。 l 和 h 分别表示房源描述和房东介绍的嵌入。

$$l = \frac{1}{Z}\sum_{i=1}^{Z} v\left(\tilde{c}_l\right) \tag{5}$$

$$h = \frac{1}{z}\sum_{i=1}^{z} v\left(\tilde{c}_h\right) \tag{6}$$

因此，我们得到房源描述和房东介绍的文本信息特征矩阵：$\mathbf{L}$ =$(l_1, l_2, ..., l_n)$ 和$\mathbf{H}$ =$(h_1, h_2, ..., h_n)$。

对于租客评论嵌入，由于其包含强烈的情感表达，为了体现租客对房屋的评价是正面还是负面，我们使用朴素贝叶斯方法为每套房屋生成相应的情感得分 $r \in [0,1]$，其中 0 表示负面，1 表示正面。具体而言，租客评论文本属于正类的概率可表示为：

$$P\left(pos \mid c_1, \ldots c_d\right) = \frac{P\left(c_1, \ldots c_d \mid pos\right) P(pos)}{P\left(c_1, \ldots c_d\right)} \tag{7}$$

化简上述公式后，我们可以得到：

$$P\left(pos \mid c_1, \ldots c_d\right) = \frac{1}{1+\gamma} \tag{8}$$

在本研究中，一段文本代表租户的一条评论，而一位租户有多条评论，因此租户的情感得分可以表示为：

$$r = \frac{1}{q}\sum_{i=1}^{q} P\left(\text{pos} \mid c_1, \ldots c_d\right) \tag{9}$$

其中， q 表示某个出租房源的评论数量，d 表示一条评论中的词语总数，P(pos $\mid c_1, \ldots c_d$) 表示该评论属于正面情感类别的概率。因此，租客评论的情感得分向量可表示为

$\mathbf{R}$ =$(r_1, r_2, ..., r_n)$.

C. 空间特征嵌入

我们提出了一种学习空间嵌入的方法。首先，将兴趣点划分为 8 种不同类型，出租房屋及其周边不同类型的兴趣点共同构成一个空间网络。然后，通过 SDNE模型学习这些空间图的网络嵌入。该方法能够准确捕捉与景点、火车站等重要兴趣点相关的空间特征。

我们使用欧氏距离计算房屋与兴趣点之间的权重 W，具体如下：

$$W = R \cdot \arccos(\text{ dis }) \cdot \pi/180 \tag{10}$$

其中 dis = sin(Lat_A) sin(Lat_B) cos($Lon_A - Lon_B$) + cos(Lat_A) cos(Lat_B)表示两类节点 A 和 B,

分别代表出租房和兴趣点。 Lon_A 和Lon_B 是它们的经度，Lat_A 和 Lat_B 是它们的纬度， R 是地球的平均半径，取值为 6371.004 km。

在SDNE模型中，编码器的映射范围是从 x_i 到 $y_i^{(k)}$，解码器的映射范围是从 $y_i^{(k)}$ 到 $\widehat{x}_i$， $y_i^{(k)}$ 是 v_i的节点嵌入，在本文中，$y_i^{(k)} = p_i$。编码器的公式为：

$$\boldsymbol{y_i^{(k)}} = \sigma\left(\boldsymbol{W^{(k)}}\boldsymbol{y_i^{(k-1)}} + b^{(k)}\right) \tag{11}$$

因此，空间嵌入可表示为**P**=(p_1， p_2，……， p_n)。获得统计嵌入、文本嵌入和空间嵌入后，我们将上述特征组合成多源特征 M =(S， T， P)，并使用全连接神经网络预测出租房源的租赁价格。我们将多源特征矩阵**M** =(m_1，m_2，……， m_n)作为神经网络的输入，得到如下结果：

$$\boldsymbol{y} = w^T\boldsymbol{m} + b, \quad A = \sigma(\boldsymbol{y}) \tag{12}$$

其中， y 是实际租赁价格， m 是多源特征， w 是参数矩阵， b 是偏置项。A 是激活函数，我们使用ReLU函数作为激活函数。最后，输出层使用一个神经元进行输出，得到预测价格 $\widehat{y}_i$。

IV. 实验

本节首先介绍两个真实数据集和评估指标。随后，我们设计实验来回答以下三个问题：

- **Q1.** 我们提出的**MSIE** 在 Airbnb 房价预测任务中的表现如何?
- **Q2.** 特征组合如何影响价格预测性能?
- **Q3.** 哪些因素对 Airbnb 房价影响最大?

A. 数据集

我们从一个开放的在线 Airbnb 网站收集数据集。表 I 展示了预处理后来自两个城市——北京和上海的两个真实 Airbnb 数据集的统计信息。

表 I：数据统计信息

City	# 房屋	# 评论	时间段
北京	10779	191876	01/2017-06/2019
上海	8638	159069	01/2020-07/2021

此外，我们还在表II中收集了北京和上海的兴趣点。我们将其分为8个类别，包括教育、娱乐、食品、饮品购物、旅游、交通、医疗服务和公共服务。

B. 评估指标

我们采用以下指标评估模型性能。

表II：兴趣点类别

数量	兴趣点类别名称	#北京	#上海
1	教育	8711	2635
2	娱乐	6501	2607
3	Food	5744	6301
4	饮品购物	6601	5632
5	旅游	6713	4176
6	交通	4322	1753
7	医疗服务	3660	2862
8	公共服务	5976	3699

（1）平均绝对误差（**平均绝对误差**）表示预测值与真实值之间误差绝对值的平均值。

$$\textbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y} - y| \tag{13}$$

(2) 均方误差（**均方误差**）用于衡量预测值与实际值之间的接近程度。

$$\textbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2 \tag{14}$$

(3) 均方根误差(**RMSE**) 定义如下：

$$\textbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2} \tag{15}$$

其中，$\hat{}y$ 是回归模型预测的价格， y 是实际价格。RMSE 越低，方法越好。

(4) 决定系数(R^2) 将预测结果转换为准确度，结果介于 [0,1] 之间。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y}_l - y_i)^2} \tag{16}$$

R^2 的值越高，估算方法越准确。

C. 基线算法

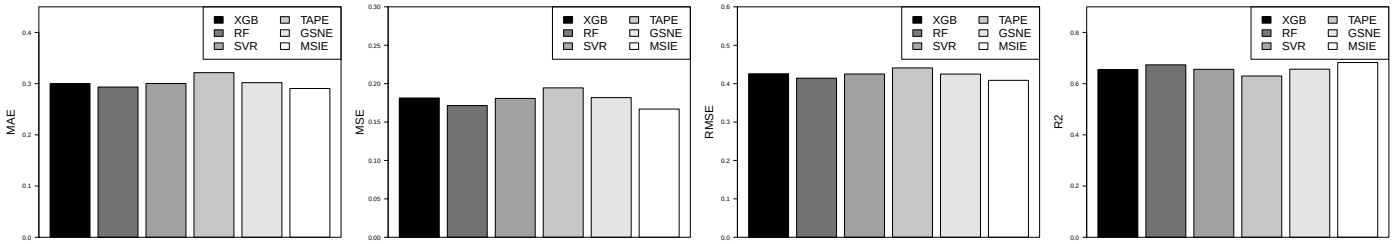
为了证明我们模型的有效性，我们将我们的方法与以下算法进行比较。

(1)极端梯度提升 XGBOOST [15]是基于梯度提升决策树的提升算法的改进版本，旨在提高速度和效率。

(2)随机森林 随机森林 [16] 是一种通过集成学习思想整合多棵树的算法，其基本单元是决策树。

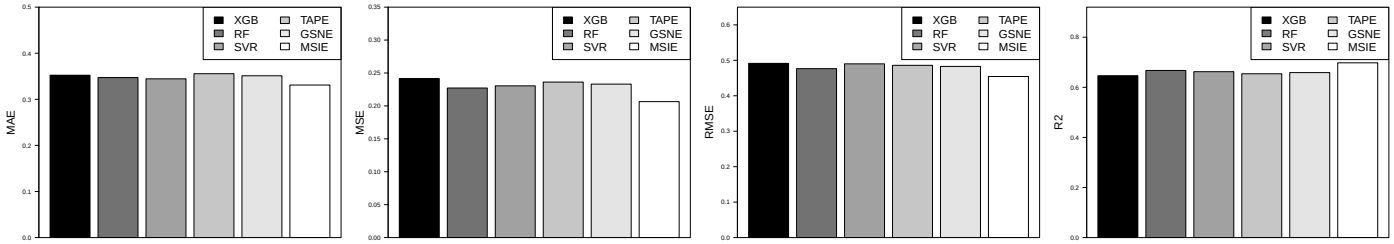
(3)支持向量回归 支持向量回归 [17] 是一种将支持向量机应用于回归问题的算法。

(4)TAPE TAPE [18] 分析了每套出租房源的描述与价格之间的关系，并加入地理因素组件，为房东的每套新出租房源推荐合理的价格。



(a) 平均绝对误差　(b) 均方误差　(c) 均方根误差　(d) R^2

图 2：北京数据集上的整体比较。



(a) 平均绝对误差　(b) 均方误差　(c) 均方根误差　(d) R^2

图 3：上海数据集上的整体比较。

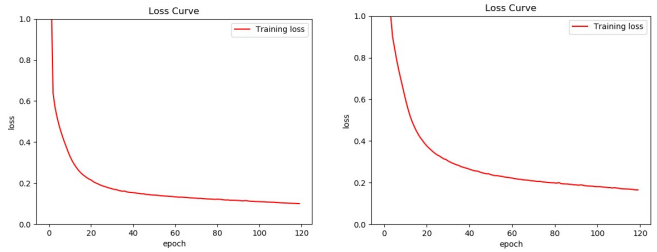
(5)GSNE GSNE [19] 是一种地理空间嵌入框架，能够准确捕捉房屋与周边兴趣点之间的地理空间邻域关系。本质上，它旨在学习地理空间网络节点上的低维高斯嵌入，并可与回归方法结合，对房价预测有一定效果。

此外，我们提出的模型有三种特征集组合变体：**(1) MSIE-S**，该模型使用统计特征；**(2) MSIE-ST**，该模型使用统计特征和文本特征；**(3) MSIE-STP**，该模型使用统计特征、文本特征和空间特征。我们使用模型对这三种变体进行了评估。

在实验中，我们将数据集划分为两个不重叠的集合：对于所有记录，最早的 80% 记录作为训练集，其余 20% 作为测试集。我们使用 PyTorch 实现该模型，并在 Windows10、Inter(R)Core(TM) i7-7700HQ@2.80GHZ、内存为 8G 的环境下运行代码。

D. 总体表现

我们展示了“平均绝对误差”“均方误差”“均方根误差”和“R^2”的结果，并将其与基线算法进行了比较。图 2 和图 3 显示，在北京和上海的数据集上，我们提出的“MSIE”方法均优于基线方法。“平均绝对误差”“均方误差”“均方根误差”的值越低，“R^2”的值越高，表示性能越好。在所有情况下，我们观察到该方法相较于基线算法均有所提升，尤其是在“均方误差”和“R^2”方面。一个有趣的发现是，传统机器学习方法（如“极端梯度提升”“随机森林”和“支持向量回归”）的表现优于“TAPE”。我们分析



（a）北京　（b）上海

图4：两个数据集上的损失曲线。

其原因在于，本文提出的算法特征工程做得很好，并且具有普适性。

此外，我们使用全连接神经网络作为预测模型。为了防止过拟合，我们在输入层设置了128个神经元，使用2个隐藏层，并将训练轮数设置为=120，批次大小设置为256。图4分别展示了神经网络在两个数据集上进行神经网络训练时的损失曲线。

E. 鲁棒性检验

我们评估特征嵌入对建模表示的贡献。为设置对照组，我们构建了所提出的“MSIE”的变体，分别为“MSIE-S”、“MSIE-ST”和“MSIE-STP”。“MSIE-S”、“MSIE-ST”和“MSIE-STP”以不同的特征集组合为输入，而其他组成部分保持不变。表 III 和表 IV 展示了对比结果。我们可以观察到

表 III：北京数据集上的特征组合。

特征集	MAE	MSE	RMSE	R^2
MSIE-S	0.3652	0.2341	0.4839	0.5545
MSIE-ST	0.2941	0.1688	0.4109	0.6786
MSIE-STP	0.2905	0.1669	0.4086	0.6824

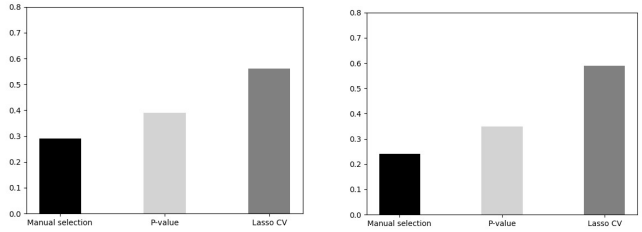
表 IV：上海数据集上的特征组合。

特征集	MAE	MSE	RMSE	R^2
MSIE-S	0.4003	0.2852	0.5340	0.5824
MSIE-ST	0.3512	0.2371	0.4869	0.6527
MSIE-STP	0.3310	0.2065	0.4544	0.6977

在两个数据集上，就四项指标而言，“MSIE-STP”的表现均优于“MSIE-S”和“MSIE-ST”。结果验证了文本特征与空间特征的结合确实增强了价格预测的建模效果。

F. 关键影响分析

为了分析价格的关键特征影响，根据以往研究，我们采用了三种特征选择方法，包括人工选择、P值 [20] 和Lasso交叉验证 [21]。我们使用 R^2 作为分析指标，结果如图5所示。最佳结果是使用 Lasso交叉验证从原始数据中选择特征。



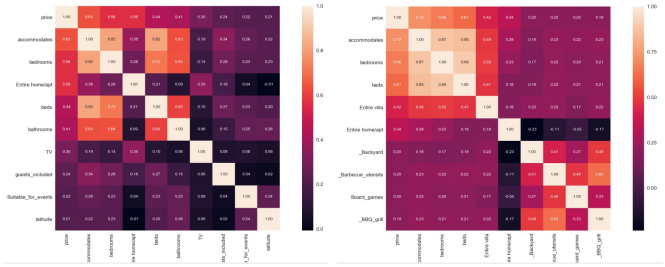
(a) 北京　　(b) 上海

图 5：使用不同特征选择方法的 R^2 。

然后，我们根据 P 值选择与租赁价格相关性最高的 20 个特征，并使用 Lasso交叉验证进行特征选择，获取统计信息作为预测的输入。图 6 展示了两个数据集中与租赁价格相关性最高的 10 个特征。

V. 相关工作

在这项工作中，我们提出了一种用于预测 Airbnb 房源价格的模型。一些研究使用情感分析来研究 Airbnb 问题。Martinez R D 等人 通过情感分析研究了 Airbnb 房东的房源描述与入住率之间的关系 [22]。Zhang 等人提出了一个文本分析框架，用于研究自我描述、信任感知与购买行为之间的关系



(a) 北京　　(b) 上海

图 6：与价格最相关的特征。

（延续上文）在 Airbnb 上。他们使用文本挖掘提取情感强度，并采用回归方法识别语言特征和语义特征对信任感知的影响 [23]。Kalehbast P R 等人 使用情感分析和机器学习来预测 Airbnb 房源的价格 [24]。

大多数研究都提出了许多利用房源信息研究价格的工作。Wang 等人利用各种房源信息（例如房东身份已验证、可容纳人数、无线网络、配套设施与服务以及免费停车）研究了价格与其决定因素之间的关系 [25]。P.Choudhary等人 分析了旧金山的 Airbnb 房源，以更好地了解如何利用不同属性（例如卧室、位置和房源类型）准确预测新房源的价格，使其既能为房东带来最佳盈利能力，又能让房客负担得起 [26]。Shen 等人 分析了每条房源描述与其价格之间的关系，并提出了一种基于文本的模型，为新发布的房源推荐合理的价格 [27]。Tang 等人 将文本信息标注为九个精选类别，提取图像相关特征，最终使用所有特征预测房源所在的街区及其价格 [28]。

VI. 结论

本文提出了一种基于多源信息嵌入的预测模型，用于研究 Airbnb 价格问题。具体而言，为了获得最佳特征集，我们首先选取房屋本身的特征，以提取统计信息特征。其次，本文将文本信息分为三类，并将房屋描述和房东介绍转换为特征矩阵。随后，将租客评论转换为关于各房屋的情感得分。然后，我们使用不同类型的兴趣点（POI）数据和房屋构建多种空间网络图，并学习其网络嵌入，以提取空间信息特征。最后，我们将这三类特征嵌入整合为多资源住房特征，作为输入，并使用本文构建的神经网络进行价格预测。我们通过两份真实数据验证了模型的有效性。未来，我们计划结合一些启发式方法 [29, 30] 以进一步提升性能。

致谢

本研究得到中国吉林省自然科学基金（项目编号：YDZJ202201ZYTS415）、中央高校基本科研业务费（项目编号：2412019ZD013）以及国家自然科学基金委员会（项目编号：61976050 和 61972384）的资助。

参考文献

[1] S. Rosen，“享乐价格与隐含市场：完全竞争中的产品差异化”，政治经济学杂志，第82卷，第1期，第34–55页，1974年1月至2月。[2] P. Morano 和 F. Tajani，“裸所有权评估：享乐价格模型与人工神经网络”，国际商业智能与数据挖掘杂志，第8卷，第4期，第340–362页，2013年。[3]X. Xu、Z. Huang、J. Wu、Y. Fu、N. Luo、W. Chen、J. Wang 和 M. Yin，“基于长春房地产数据的有限混合模型房价关键影响因素分析”，载于*DASFAA*，第11448卷，2019年，第378–382页。[4]X. Xu、Y. Fu、J. Wu、Y. Wang、Z. Huang、Z. Fu 和 M. Yin，“自适应加权有限混合模型：房地产特征影响识别”，数据科学汇刊，第1卷，第3期，第20:1–20:16页，2020年。[5] Y. Fu、Y. Ge、Y. Zheng、Z. Yao、Y. Liu、H. Xiong 和 J. Yuan，“结合在线用户评论与线下移动行为的稀疏房地产排名”，载于*ICDM*，2014年，第120–129页。[6] Y. qing Li、T. Wang 和 S. fei Zhao，“基于粗糙集的SVM在房地产价格预测中的应用”，载于*WiCOM*，2008年，第1–4页。[7] P. R. Kalehbasti、L. Nikolenko 和 H. Rezaei，“使用机器学习和情感分析进行Airbnb 房价预测”，载于计算机科学讲义，2021年，第173–184页。[8] X. Chen、L. Wei 和 J. Xu，“使用长短期记忆网络进行房价预测”，*CoRR*，第abs/1709.08432卷，2017年。[9] P. Wang、Y. Fu、J. Zhang、X. Li 和 D. Lin，“学习城市社区结构：基于周期性时空移动图的集体嵌入视角”，*ACM*智能系统与技术汇刊，第9卷，第6期，第63:1–63:28页，2018年。[10]P. Wang、Y. Fu、H. Xiong 和 X. Li，“面向移动用户画像的对抗性子结构表示学习”，载于*SIGKDD*。ACM，2019年，第130–138页。[11] P. Wang、K. Liu、L. Jiang、X. Li 和 Y. Fu，“增量式移动用户画像：利用空间知识图谱对事件流进行建模的强化学习”，载于*KDD*，2020年。[12] P. Wang、Y. Fu、J. Zhang、P. Wang、Y. Zheng 和 C. C. Aggarwal，“你的驾驶方式定义了你：用于驾驶行为分析的同伴感知与时间感知表示学习”，载于*SIGKDD*，2018年，第2457–2466页。[13] D. Wang、P. Cui 和 W. Zhu，“结构化深度网络嵌入”，载于*SIGKDD*，2016年，第1225–1234页。[14] Q. Luo、W. Xu 和 J. Guo，“关于CBOW模型过拟合与稳定性的研究”，载于*Web-KR@CIKM*。ACM，2014年，第9–12页。[15]T. Chen 和 C. Guestrin，“XGBOOST：一种可扩展的树提升系统”，载于*ACM*，2016年。[16] L. Breiman，“随机森林”，机器学习，第45卷，第1期，第5–32页，2001年。[17]M. Awad 和 R. Khanna，支持向量回归。伯克利，加利福尼亚州：Apress，2015年，第67–80页。[18] L. Shen、Q. Liu、G. Chen 和 S. Ji，“面向在线租赁房屋的基于文本的价格推荐系统”，大数据挖掘与分析，第3卷，第2期，第143–152页，2020年。[19] S. S. S. Das、M. E. Ali、Y. Li、Y. Kang 和 T. Sellis，“利用地理空间网络嵌入提升房价预测效果”，数据挖掘与知识发现，第35卷，第6期，第2221–2250页，2021年。[20] R. Feise，“多个结局指标是否需要进行P值调整？”，*BMC*医学研究方法学，第2卷，第8期，2002年。[21] H. Zou，“自适应Lasso及其预言机性质”，美国统计学会会刊，第101卷，第476期，第1418–1429页，2006年。[22] R. D. Martinez、A. Carrington、T. Kuo、L. Tarhuni 和 N. Abdel-Motaal，“Airbnb房东房源描述的‘情感’与长度对入住率的影响”，2017年。[23] L. Zhang、Q. Yan 和 L. Zhang，“用于理解房东自我描述、信任感知与Airbnb购买行为之间关系的文本分析框架”，决策支持系统，第133卷，文章113288，2020年。[24] P. R. Kalehbasti、L. Nikolenko 和 H. Rezaei，“使用机器学习和情感分析进行Airbnb 房价预测”，2019年。[25]D. Wang 和 J. L. Nicolau，“共享经济住宿租赁的价格决定因素：基于Airbnb.com上33座城市房源的研究”，国际酒店管理杂志，第62卷，第120–131页，2017年。[26] P. Chou D Hary、A. Jain 和 R. Baijal，“揭秘Airbnb：预测新房源价格”，*Papers*，2018年。[27] L. Shen、Q. Liu、G. Chen 和 S. Ji，“面向在线租赁房屋的基于文本的价格推荐系统”，大数据挖掘与分析，第3卷，第2期，第143–152页，2020年。[28] E. Tang 和 K. Sangani，“旧金山Airbnb房源的街区与价格预测”，2015年。[29] Y. Wang、S. Cai、J. Chen 和 M. Yin，“SCCWalk：一种高效的最大权团问题局部搜索算法及其改进”，人工智能，第280卷，文章103230，2020年。[30]S. Pan、Y. Ma、Y. Wang、Z. Zhou、J. Ji、M. Yin 和 S. Hu，“最小独立支配集问题的改进主从式进化算法”，计算机科学前沿，第17卷，第4期，第1–14页，2023年。