

基于多源异构数据融合的移动群智感知任务推荐算法

周易歆, 何杏宇, 李金慧

上海理工大学, 出版印刷与艺术设计学院, 上海

收稿日期: 2024年9月24日; 录用日期: 2024年10月22日; 发布日期: 2024年10月30日

摘要

现有的移动群智感知任务推荐方法主要包括基于时空信息和社交关系两大类, 分别受到时空信息时变性和数据稀疏性问题的限制。为了突破上述限制并综合多源工人行为数据对任务推荐准确度的影响, 本文提出了一种基于多源异构数据融合的移动群智感知任务推荐算法。该算法综合考虑了工人地理位置、时间可用性以及社交网络关系等多方面的因素, 利用隐语义模型、双向长短时记忆网络模型分别生成空间、时间维度的工人-任务匹配度矩阵实现对工人偏好时空变化的挖掘, 并利用Node2Vec图嵌入模型生成社交维度的工人-任务匹配度矩阵, 缓解数据稀疏性问题。最后利用BP神经网络将上述三个矩阵进行融合以实现工人行为的多源异构数据融合, 并最终进行任务推荐。实验结果表明, 本文方法比单一数据源的任务推荐方法具有更好的推荐准确度。

关键词

移动群智感知, 工人行为分析, 多源异构融合, 任务推荐

Task Recommendation Based on Multi-Source Heterogeneous Data Fusion in Mobile Crowd Sensing

Yixin Zhou, Xingyu He, Jinhui Li

College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai

Received: Sep. 24th, 2024; accepted: Oct. 22nd, 2024; published: Oct. 30th, 2024

Abstract

Existing task recommendation methods for mobile crowd sensing primarily fall into two categories:

文章引用: 周易歆, 何杏宇, 李金慧. 基于多源异构数据融合的移动群智感知任务推荐算法[J]. 运筹与模糊学, 2024, 14(5): 671-684. DOI: 10.12677/orf.2024.145505

those based on spatial-temporal information and those based on social relationships, both constrained by the variability of spatial-temporal information and data sparsity issues. To overcome these limitations and enhance task recommendation accuracy by integrating multi-source worker behavior data, this paper proposes a task recommendation method based on multi-source heterogeneous data fusion for mobile crowd sensing. The proposed method comprehensively considers multiple factors, including workers' geographic locations, time availability, and social network relationships. It employs a latent semantic model and a bi-directional long short-term memory model to generate worker-task matching matrices in the spatial and temporal dimensions to capture temporal-spatial variations in worker preferences, while the Node2Vec graph embedding model generates the worker-task matching matrix in the social dimension to alleviate data sparsity issues. Finally, a BP neural network is utilized to fuse these three matrices, achieving multi-source heterogeneous data fusion of worker behaviors and producing task recommendations. Experimental results demonstrate that the proposed method achieves better recommendation accuracy than task recommendation methods based on a single data source of worker behaviors.

Keywords

Mobile Crowd Sensing, Worker Behavior Analysis, Multi-Source Heterogeneous Fusion, Task Recommendation

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

移动群智感知(Mobile Crowd Sensing, MCS)作为一种新兴的信息收集和处理方法[1], 可以实现大规模数据收集。为了提高数据的质量和参与者的积极性, 对参与者进行合理的任务推荐是重要的技术措施。现有的移动群智感知任务推荐方法的主要原理是从不同维度对工人行为数据进行分析, 预测出工人的未来偏好并进行任务推荐, 主要包括基于时空信息的任务推荐方法[2]与基于社交网络信息的任务推荐方法[3]。其中, 基于时空信息的任务推荐方法主要是通过结合工人历史执行任务的记录中的时间与空间信息来更好地进行任务推荐[4], 而基于社交网络信息的任务推荐方法则主要利用社交网络中的用户行为和关系信息来提高任务推荐的效果[5]。然而, 第一类方法没有考虑到工人偏好的时变性, 影响任务推荐的准确度。第二类方法的任务推荐精确性则受到社交网络数据稀疏性问题的限制。此外, 现有任务推荐方法大多基于单一维度的工人行为数据分析进行任务推荐, 如工人相似度、任务相似度、工人移动序列等维度, 限制了推荐准确度的提高[6]。而基于多个维度的工人行为数据分析进行任务推荐存在异构数据融合的挑战。

为此, 本文将提出一种基于多源异构数据融合的移动群智感知任务推荐方法, 具体地, 该方法首先定义时变性的工人-任务兴趣因子, 基于此构造初始的时间维度工人-任务匹配度矩阵, 并通过隐语义模型对初始的时间维度工人任务匹配度矩阵进行补全优化; 接着通过双向长短期记忆网络预测工人的位置, 并计算预测位置与任务位置的相似度以构造空间维度的工人-任务匹配度矩阵; 然后根据工人社交信息生成社交网络图, 使用 Node2Vec 图嵌入技术得到工人的社交维度特征向量, 以计算工人之间的社交相似度来构造工人亲密度矩阵, 并结合工人亲密度矩阵与工人的历史任务参与度计算社交维度的工人-任务匹配度矩阵; 最后, 使用 BP 神经网络将上述三个维度的工人-任务匹配度矩阵进行融合, 得到最终的多维度工人-任务匹配度矩阵, 并根据此实现任务推荐。

2. 相关工作

2.1. 基于时空信息的任务推荐方法

手机等移动定位设备的快速普及推动了时空众包技术的发展, 各种基于空间位置的服务与人们的生活联系越来越密切。从工人的历史执行任务信息中提取时空信息进行任务推荐, 能够有效提高任务推荐的准确率。文献[7]基于用户移动轨迹预测对用户进行任务推荐, 通过递归神经网络预测用户下一目的地, 并通过注意力机制捕获与候选任务相似的记录, 获取任务选择偏好进行任务推荐。文献[8]通过构建时空知识图谱, 利用 TransE 模型进行知识图谱嵌入, 使用图卷积神经网络和循环神经网络学习工人与任务的密集表示, 结合时空信息与工人偏好进行任务推荐。文献[9]提出了一种基于指数分布的时长敏感任务推荐模型, 通过构建效用函数, 结合贪心算法, 综合考虑工人的历史执行记录和任务时长进行任务推荐; 文献[10]提出一种动态延迟决策任务推荐算法, 通过移动预测方法计算用户在感知任务结束前到达目的地的概率, 采用双深度 Q 网络, 优化使得平台利润最大化, 工人移动距离与响应时间最小化来进行任务推荐。文献[11]通过考虑用户参与下一任务时的出行方式变化, 构建变速多任务分配的约束多目标优化模型, 通过最大化用户奖励和最小化任务完成时间来进行任务推荐。

然而, 大多基于时空信息的任务推荐方法没有考虑到工人偏好的时空变化, 影响任务推荐的准确度, 因此, 本文采取结合了兴趣因子的隐语义模型与双向长短时记忆网络模型来生成时间维度与空间维度的工人-任务匹配度矩阵, 考虑工人历史行为的时序性, 挖掘工人偏好的动态变化以进行任务推荐。

2.2. 基于社交网络信息的任务推荐方法

人的行为往往受到从众心理的影响, 工人参与任务的意愿往往受到周围工人的影响。因此在任务推荐中利用社交网络来分析工人的影响力和工人之间的社交关系, 据此提高任务推荐的精度[12]。文献[13]通过矩阵分解模型, 引入工人在社交网络中的外向性与亲密度计算工人对任务的评分, 生成每个工人的任务推荐列表; 文献[14]在基于位置的社交网络中, 结合用户偏好和社交影响力进行任务推荐。文献[15]通过构建社交网络, 分析工人的社交关系信息, 与工人的偏好信息相结合, 为不活跃的工人进行任务推荐。文献[16]通过对工人偏好、历史记录与社交网络信息进行综合分析, 为工人进行任务推荐。

但是, 在实际任务推荐过程中考虑社交网络信息往往存在着数据稀疏性问题, 导致任务推荐准确率较低。因此, 本文通过构建社交网络图, 使用 Node2vec 图嵌入技术, 学习工人的特征向量表示, 计算工人之间的社交相似度来帮助生成社交维度工人-任务匹配度矩阵, 在一定程度上缓解数据稀疏性问题, 提高了任务推荐的准确率。

3. 模型与方案陈述

本文提出基于多源异构数据融合的移动群智感知任务推荐算法, 算法的整体架构如图 1 所示, 由时间维度工人行为分析模块、空间维度工人行为分析模块、社交维度工人行为分析模块和多维度工人行为分析融合模块共四个模块组成。分别为基于隐语义模型的时间维度工人-任务匹配度矩阵生成、基于 Bi-LSTM 模型的空间维度工人-任务匹配度矩阵生成、基于 Node2Vec 图嵌入技术的社交维度工人-任务匹配度矩阵生成与基于 BP 神经网络的多维度工人-任务匹配度矩阵生成。

3.1. 基于隐语义模型的时间维度工人-任务匹配度矩阵生成

在时间维度工人行为分析模块中, 构造时间维度的工人-任务匹配度矩阵的算法流程图如图 2 所示, 首先构造初始时间维度工人任务匹配度矩阵, 然后利用隐语义模型对其进行迭代优化。

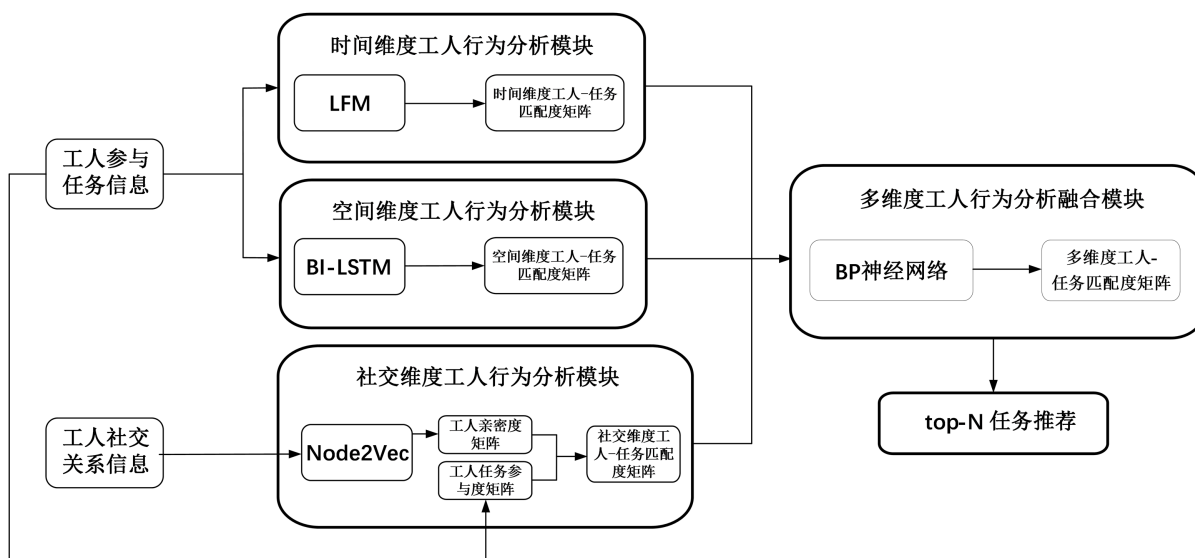


Figure 1. Task recommendation model structure diagram based on multi-source heterogeneous data fusion

图 1. 基于多源异构数据融合的任务推荐模型结构图

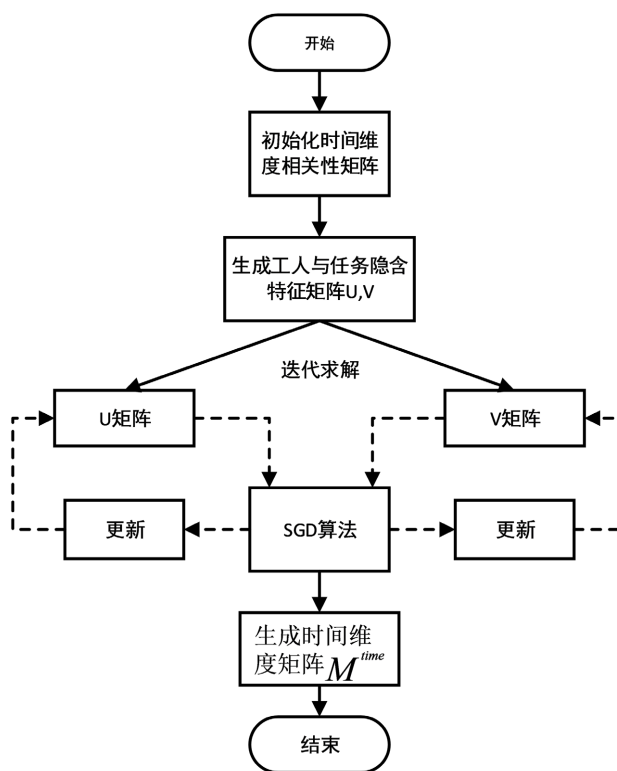


Figure 2. Algorithm flowchart for constructing worker-task matching matrix from temporal dimension

图 2. 构造时间维度的工人 - 任务匹配度矩阵的算法流程图

1) 构造初始时间维度工人任务匹配度矩阵 M^{time} : m_{ij}^{time} 为矩阵 M^{time} 中第 i 行第 j 列的元素, 表示工人 i 对任务 j 的时间维度匹配度, 具体计算方法如公式(1)和(2)所示:

$$m_{ij}^{time} = \sum_{l=1}^k N_{ij}^{(l)} \quad (1)$$

$$N_{ij}^{(l)} = \exp\left(-\beta \cdot \left|t_{ij}^{(l)} - t_{rec}\right|\right) \quad (2)$$

其中, $N_{ij}^{(l)}$ 表示工人 i 对任务 j 的第 l 次参与时的兴趣因子, k 为工人 i 对任务 j 的参与次数。兴趣因子 N_{ij} 的计算使用指数衰减函数, 其中 β 为兴趣衰减因子, $t_{ij}^{(l)}$ 为工人 i 第 l 次参与任务 j 的时间, t_{rec} 为进行推荐的时间。

2) 优化时间维度工人 - 任务匹配度矩阵 M^{time} :

在优化时间维度工人 - 任务匹配度矩阵的过程中, 本文首先通过公式(3)将初始的时间维度工人 - 任务匹配度矩阵 M^{time} 分解为工人隐含特征矩阵 U 和任务隐含特征矩阵 V , 然后分别对工人隐含特征矩阵 U 与任务隐含特征矩阵 V 之中的参数 u 和 v 进行迭代优化, 提取出最优的隐含特征矩阵 U 和 V , 并据此计算得到最终的时间维度工人 - 任务匹配度矩阵 M^{time} 。在矩阵分解过程中使用余弦相似度预测时间维度工人 - 任务匹配度。具体地, 工人 i 与任务 j 的时间维度匹配度 $\widehat{m}_{ij}^{time}$ 计算方式如公式(4)所示。

$$M^{time} = U \times V \quad (3)$$

$$\widehat{m}_{ij}^{time} = \frac{u_i \cdot v_j}{\|u_i\| \|v_j\|} \quad (4)$$

其中 u_i 为 U 的第 i 行, 代表工人 i 的隐含特征向量, v_j 为 V 的第 j 行, 表示任务 j 的隐含特征向量。采用随机梯度下降法(SGD)计算迭代求解过程中的损失函数全局最小值, 计算方法如公式(5)所示。为了避免过拟合现象, 在损失函数的迭代优化过程中加入正则化因子。

$$\min_{u,v} = \min_{u,v} \left\{ \sum_{i,j} \left(m_{ij}^{time} - \widehat{m}_{ij}^{time} \right)^2 + \lambda \|u_i\|^2 + \lambda \|v_j\|^2 \right\} \quad (5)$$

其中, m_{ij}^{time} 为工人 i 与任务 j 的时间维度实际匹配度, $\widehat{m}_{ij}^{time}$ 为工人 i 与任务 j 的时间维度预测匹配度, λ 为防止过拟合的正则化参数。

3.2. 基于 Bi-LSTM 模型的空间维度工人 - 任务匹配度矩阵生成

在空间维度工人行为数据分析模块中, Bi-LSTM 模型结构如图 3 所示, 使用 Bi-LSTM 模型预测工人可能到达的下一个位置, 基于预测的位置和各个任务的位置进行相似度计算, 得到空间维度的工人 - 任务匹配度矩阵。

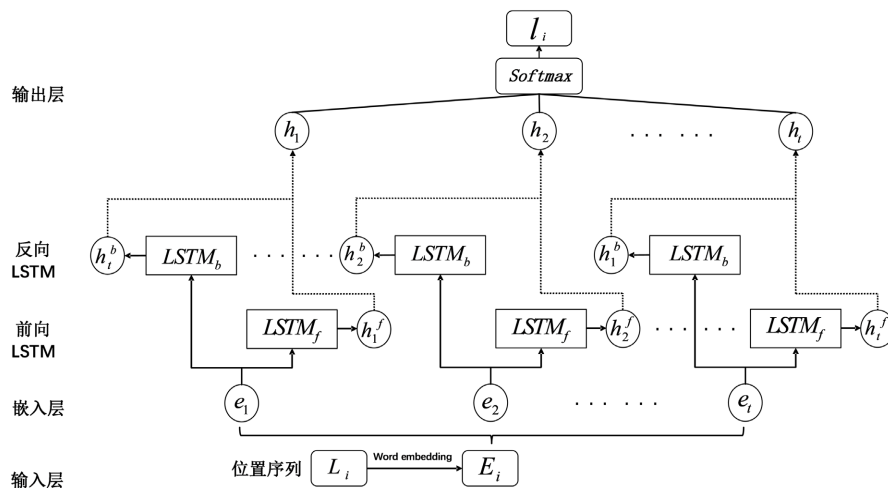


Figure 3. The structure of Bi-LSTM model

图 3. Bi-LSTM 模型结构图

1) 位置序列嵌入：在输入层输入每个工人的位置序列 $L_i = [(t_1, l_1), (t_2, l_2), \dots]$ ，其中 l 为工人 i 在 t 时刻所处的位置。在嵌入层通过 word embedding 技术对工人位置序列信息进行嵌入，得到序列嵌入向量 E_i 。

2) Bi-LSTM：将通过嵌入层转化而得的嵌入向量序列 E_i 输入到 Bi-LSTM 层，采用正向 LSTM 为每个时间点生成隐藏状态 h_t^f ，采用反向 LSTM 为每个时间点生成隐藏状态 h_t^b ，最后合并正反向 LSTM 的输出得到 h_t ，公式如(6)、(7)、(8)所示：

前向 LSTM 为每个时间点生成隐藏状态：

$$h_t^f = \text{LSTM}_f(e_t, h_{t-1}^f) \quad (6)$$

后向 LSTM 为每个时间点生成隐藏状态：

$$h_t^b = \text{LSTM}_b(e_t, h_{t-1}^b) \quad (7)$$

合并前向和后向 LSTM 的输出：

$$h_t = [h_t^f, h_t^b] \quad (8)$$

其中， h_t^f 为前向 LSTM 的输出， h_t^b 为后向 LSTM 的输出， h_t 为 LSTM 层的输出。

3) 预测位置输出：在全连接层对工人 i 可能到达的下一个位置 l_i 进行预测，公式如(9)所示：

$$l_i = \text{softmax}(W_i h_t + b_i) \quad (9)$$

其中 W_i 为权重矩阵， b_i 为偏置项， l_i 为工人 i 下一可能到达位置的位置 ID。

4) 生成空间维度工人 - 任务匹配度矩阵 M^{space} ： m_{ij}^{space} 为空间维度工人 - 任务匹配度矩阵 M^{space} 中的元素，表示工人 i 和任务 j 的空间维度匹配度，公式如(10)所示：

$$m_{ij}^{space} = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{|lat_i - lat_j|}{2} \right) + \cos(lat_i) \cdot \cos(lat_j) \cdot \sin^2 \left(\frac{|lan_i - lan_j|}{2} \right)} \right) \quad (10)$$

其中 (lan_i, lat_i) 为对应的工人位置 ID 信息得到的下一可能到达位置 l_i 的经纬度， r 为地球半径。

3.3. 基于 Node2Vec 图嵌入技术的社交维度工人-任务匹配度矩阵生成

3.3.1. 基于 Node2Vec 图嵌入提取工人社交维度特征向量

在社交维度工人行为数据分析模块中，首先，根据工人社交信息生成工人社交网络图 $Graph G$ ，使用 Node2Vec 图嵌入技术学习工人特征向量，具体算法过程如表 1 所示。

1) $LearnFeatures(Graph G = (V, E, W))$ 函数：用于提取图特征，其中 $Graph G$ 表示社交网络图， V 为节点，表示工人； E 为边，表示二者有朋友关系， W 为权重，设为 1。

2) $\pi = PreprocessModifiedWeights(G, p, q)$ 函数：用于生成随机游走采样策略，其中 G 为社交网络图， p, q 为超参数。每个节点生成 r 个随机游走序列，对于图中的每个节点，生成一个随机游走序列 $walk$ 。

3) $walk = Node2vecWalk(G', u, l)$ 函数：用于生成随机游走序列 $walk$ ，其中 G' 为经过权重调整后的新社交网络图， l 为随机游走序列个数。初始化 $walk$ 为包含起始节点 u 的序列。对于每个游走迭代步骤，从当前节点开始，向邻居节点进行随机游走。调用 $GetNeighbors(curr, G')$ 函数，获取当前节点的邻居节点，其中 $curr$ 为当前节点， V_{curr} 为当前节点的邻居节点。根据采样策略 $AliasSample$ ，选择下一个节点 u_i 。将节点 u_i 添加到游走序列 $walk$ 中，并返回随机游走序列 $walk$ 。

4) $f(u) = StochasticGradientDescent(k, d, walks)$ 函数：用于基于 skip-gram 模型训练生成工人节点 u 的特征向量。

Table 1. Node2Vec graph embedding technique algorithm
表 1. Node2Vec 图嵌入技术算法

算法 1: Node2Vec 算法

```

LearnFeatures(Graph  $G = (V, E, W)$ )
     $\pi = \text{PreprocessModifiedWeights}(G, p, q)$ 
     $G' = (V, E, \pi)$ 
    Initialize walks to Empty
    for iter = 1 to r do
        for all nodes  $u \in V$  do
             $walk = \text{Node2vecWalk}(G', u, l)$ 
            Append walk to walks
     $f(u) = \text{StochasticGradientDescent}(k, d, \text{walks})$ 
return  $f(u)$ 

node2vecWalks(Graph  $G = (V, E, \pi)$ )
    Initialize walks to [u]
    for walk_iter = 1 to l do
         $curr = walk[-1]$ 
         $V_{curr} = \text{GetNeighbors}(curr, G')$ 
         $s = \text{AliasSample}(V_{curr}, \pi)$ 
        Append s to walk
    return walk
  
```

3.3.2. 构造社交维度工人 - 任务匹配度矩阵

1) 构建工人亲密度矩阵 S : s_{ik} 为工人亲密度矩阵 S 中的元素, 表示工人 i 与工人 k 之间的社交相似度, 计算方法如公式(11)所示:

$$s_{ik} = \frac{f(i) \cdot f(k)}{\|f(i)\| \|f(k)\|} \quad (11)$$

其中, $f(i)$ 和 $f(k)$ 分别为两个工人的特征向量。

2) 构建工人任务参与度矩阵 P : p_{kj} 为工人任务参与度矩阵 P 中的元素, 代表工人 k 对任务 j 的参与总次数。计算方法如公式(12)所示:

$$p_{kj} = \sum_{l=1}^n t_{kj}^{(l)} \quad (12)$$

其中, $t_{kj}^{(l)}$ 的值为 1, 代表了工人 k 对任务 j 的一次参与, l 为工人 k 对任务 j 的第 l 次参与, n 为参与的总数。

3) 生成社交维度工人 - 任务匹配度矩阵 M^{social} : m_{ij}^{social} 为社交维度工人 - 任务匹配度矩阵的元素, 表示工人 i 对任务 j 的匹配度。具体计算方法如公式(13)所示:

$$m_{ij}^{social} = \sum_k s_{ik} \cdot p_{kj} \quad (13)$$

其中, p_{kj} 为工人任务参与度矩阵 P 中的元素, 表示工人 k 对任务 j 的参与度。

3.4. 基于 BP 神经网络的多维度工人 - 任务匹配度矩阵生成

在多维度工人行为分析融合模块中, 利用 BP 神经网络对上述三个维度的工人 - 任务匹配度矩阵进

行融合，生成多维度工人-任务匹配度矩阵，BP神经网络结构如图4所示。

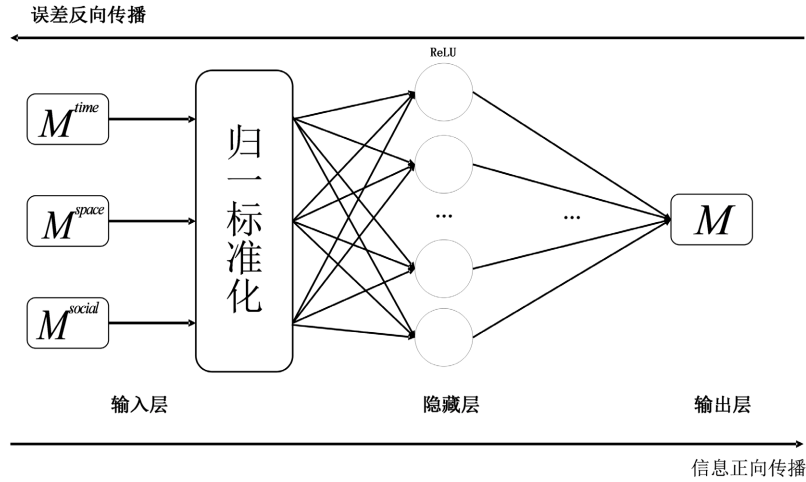


Figure 4. The structure of BP neural network

图4. BP神经网络结构示意图

1) 归一化：将时间维度工人-任务匹配度矩阵、空间维度工人-任务匹配度矩阵、社交维度工人-任务匹配度矩阵三个矩阵进行归一化， m_{ij}^M 为各个维度矩阵的元素，代表对应维度工人 i 与任务 j 的匹配度，计算方法如公式(14)所示：

$$m_{ij}^M = \frac{m_{ij}^M - m_{\min}^M}{m_{\max}^M - m_{\min}^M} \quad (14)$$

其中， $M \in \{time, space, social\}$ ， $m_{\min}^M$ 为输入各个维度矩阵中的最小值， $m_{\max}^M$ 为输入各个维度矩阵中的最大值。

2) 输入层：将归一化后的矩阵用于神经网络的输入和目标，输入层包括归一化后的时间维度工人-任务匹配度矩阵 M^{time} ，空间维度工人-任务匹配度矩阵 M^{space} ，社交维度工人-任务匹配度矩阵 M^{social} 。

3) 隐藏层：使用 $ReLU$ 激活函数，公式如(15)所示：

$$H = ReLU(W \cdot X + b) \quad (15)$$

其中 H 为隐藏层输出， W 为输入层到隐藏层的权重矩阵， b 为隐藏层的偏置向量， X 为训练的批量样本。

4) 输出层：输出层根据隐藏层的特征表示，将其映射到最终的输出空间，得到最终的工人-任务匹配度矩阵 M ，公式如(16)所示：

$$M = W' \cdot H + b' \quad (16)$$

其中 M 即为最终生成的多维度工人-任务匹配度矩阵， H 为隐藏层的输出， W' 为输出层的权重矩阵， b' 为输出层的偏置向量。

5) 损失函数：以工人对任务的签到次数作为多维度工人-任务匹配度矩阵中的实际匹配度，损失函数如公式(17)所示：

$$LOSS = \frac{1}{N} \sum_{i=1}^N (m_{ij} - \widehat{m}_{ij})^2 \quad (17)$$

其中， N 为评分矩阵元素数量， m_{ij} 为工人 i 和任务 j 的多维度实际匹配度， $\widehat{m}_{ij}$ 为工人 i 和任务 j 的多维度预测匹配度。

训练好的 BP 神经网络可以用来计算工人和任务之间的多维度匹配度，生成最终的多维度工人 - 任务匹配度矩阵。根据生成的最终多维度工人 - 任务匹配度矩阵，可以得到每个工人对每个任务的评分，然后得到工人推荐任务列表，最后选取评分最高的 N 个任务对工人进行 $top-N$ 任务推荐。

4. 实验结果与分析

4.1. 数据集

本实验采用的数据集 Gowalla，该数据集包含了 196,591 个用户从 2009 年 8 月至 2010 年 9 月的签到记录。每条签到记录包括用户 ID，签到时间，签到地点的纬度，签到地点的经度，以及每个地点的经纬度所唯一对应的地点 ID。鉴于数据中用户的签到地点较为广泛，所以本文筛选了位于纽约市的数据，并且通过对工人的签到地点以及签到次数进行筛选，共筛选出 12,319 条数据，包含 161 个工人和 179 个地点，试验数据以 8:2 的比例随机划为训练集和测试集，从而更准确地评估模型的性能。

4.2. 评价指标

本实验推荐结果采用常用的 $top-N$ 列表的形式，评价指标选取准确率 precision、召回率 recall 和 F1 Score。其中 precision 表示推荐列表中的物品属于测试集的比例，recall 描述测试集中的物品占推荐列表的比例，F1 Score 表示准确率和召回率的调和平均值，综合考虑了模型的准确率和召回率。F1 Score 越高，表示模型在准确率和召回率之间取得了更好的平衡。公式如下：

$$\text{precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (18)$$

$$\text{recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (19)$$

$$\text{F1 Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (20)$$

其中 $|R(u) \cap T(u)|$ 为正确推荐的数目， $|R(u)|$ 为推荐的任务数目， $|T(u)|$ 为工人实际参与的任务数目。

4.3. 参数设置

4.3.1. LFM 模型

主从给定的文件中读取数据，这些数据包含了用户的签到记录和签到时间，将签到时间转换为时间戳，并以参考时间为基准计算时间衰减因子，用于衡量签到时间的影响。根据用户和任务的签到记录，构建初始的工人 - 任务匹配度矩阵。使用隐语义模型进行矩阵分解，通过学习用户和任务的隐含特征向量来表示它们之间的关系，利用随机梯度下降算法(SGD)更新模型参数，以最小化预测评分与实际评分之间的差异。采用余弦相似度作为评分预测的指标，并根据差异调整模型参数，以提高评分预测的准确性，通过计算最终预测矩阵与初始矩阵非零部分的均方误差(MSE)，对模型的性能进行评估，保存最终时间维度工人 - 任务匹配度矩阵 M^{time} ，参数设置如表 2 所示。

4.3.2. Bi-LSTM 模型

使用 Bi-LSTM 模型进行用户位置预测任务。该模型包括一个嵌入层、一个双向 LSTM 层和一个线性输出层，定义模型的结构和超参数，包括输入维度、隐藏维度、LSTM 层数、输出维度等。使用交叉熵损失函数和随机梯度下降(SGD)优化器进行模型训练，在训练过程中，定期输出当前的训练损失和准确率，并在每个周期结束后对测试集进行评估，计算测试误差和准确率。保存取得最佳准确率的模型参数，加

载训练好的模型，并利用该模型对测试集中的数据进行推断，以预测用户的下一个位置。根据用户位置和任务位置之间的距离计算余弦相似度，构建并保存空间维度工人 - 任务匹配度矩阵 M^{space} ，参数如表 3 所示。

Table 2. LFM parameter setting
表 2. LFM 参数设置

参数名称	参数值
时间衰减因子	0.01
隐含特征数量	15
正则化参数	0.02
学习轮数	30
学习率	0.1

Table 3. Bi-LSTM parameter setting
表 3. Bi-LSTM 参数设置

参数名称	参数值
学习轮数	500
批处理大小	32
学习率	0.005
传播层数	3
Dropout	0.2

4.3.3. Node2Vec 图嵌入技术

使用 NetworkX 库构建一个无向图，其中每个节点代表一个工人，边表示工人之间的社交关系，遍历社交网络边数据，将边添加到图中，使用 Node2Vec 算法对构建的社交网络图进行训练，以学习节点的嵌入表示，进行多次随机游走，从每个节点开始，沿着边随机游走一定步数，并记录下每次游走的路径，使用 Word2Vec 算法将这些随机游走路径转换为节点的嵌入向量。定义一个用于计算余弦相似度的函数，用于衡量两个节点嵌入向量之间的相似度。使用定义的余弦相似度函数，计算工人之间的社交匹配度矩阵，其中每个元素表示两个工人之间的社交相似度，计算工人和任务之间的参与度矩阵，其中每个元素表示一个工人参与一个任务的次数，将社交相似度矩阵和参与度矩阵相乘，得到了社交维度下的工人 - 任务匹配度矩阵 M^{social} ，参数设置如表 4 所示。

4.3.4. BP 神经网络

读取三个匹配度矩阵： M^{time} 、 M^{space} 和 M^{social} ，对这三个矩阵进行归一化处理，使用 MinMaxScaler 进行归一化。定义一个多层感知器模型，包含一个输入层、一个隐藏层和一个输出层。输入层的大小为三个归一化矩阵的特征数之和，隐藏层大小为 50，输出层大小为工人 - 任务矩阵的列数。使用均方误差 (MSE) 作为损失函数，使用随机梯度下降 (SGD) 作为优化器，学习率设置为 0.1，将归一化后的矩阵转换为 PyTorch 张量，用于神经网络的输入和目标。在循环中进行模型的训练，迭代次数为 20,000 次，在每

个 epoch 中，执行前向传播、计算损失、反向传播和优化，计算模型的均方根误差(RMSE)和平均绝对误差(MAE)，使用训练好的模型对测试集进行预测，得到最终的多维度工人 - 任务匹配度矩阵 M ，参数设置如表 5 所示。

Table 4. Node2Vec parameter setting
表 4. Node2Vec 图参数设置

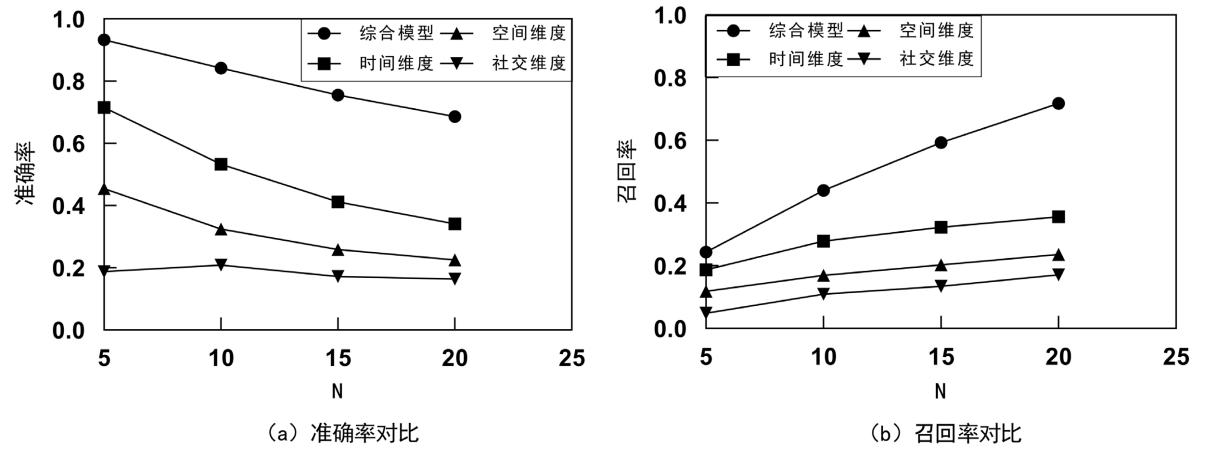
参数名称	参数值
窗口大小	5
随机游走长度	10
返回概率参数	1
离开概率参数	1
学习轮数	10
随机游走序列个数	10
特征向量维度	64

Table 5. BP neural network parameter setting
表 5. BP 神经网络参数设置

参数名称	参数值
学习轮数	2×10^4
隐含层大小	50
学习率	0.1

4.4. 性能对比

为了验证该综合模型推荐方法的有效性，本文对上述的时间维度、空间维度、社交维度和综合模型 4 种推荐方式进行比较，根据 4 种方式推荐的任务数目，分别计算 N ($N = 5, 10, 15, 20$)时的准确率 precision@N、召回率 recall@N 和 F1 Score，绘制图像后结果如图 5 所示。



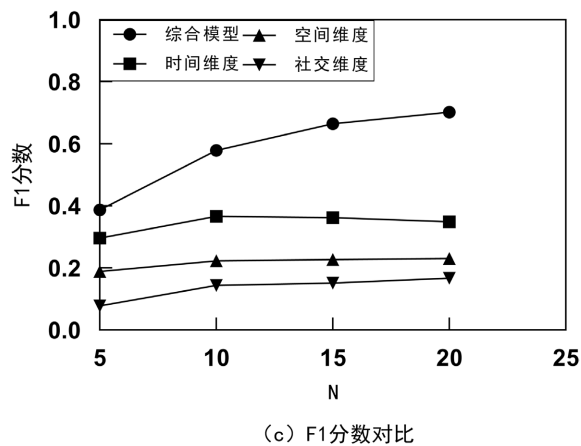


Figure 5. (a) Comparison of Top-N precision for four models; (b) Comparison of Top-N recall for four models; (c) Comparison of Top-N F1 score for four models
图 5. (a) 四种模型 Top-N 准确率对比; (b) 四种模型 Top-N 召回率对比; (c) 四种模型 Top-N 分数对比

由图 5 可知,随着推荐任务数目 N 的增大,准确率(precision@N)整体呈下降趋势,召回率(recall@N)逐渐上升, F1 Score 也逐渐上升且趋于平稳,且从高到低的顺序为综合模型 - 时间维度模型 - 空间维度模型 - 社交维度模型。在不同的推荐任务数目 N 下,综合模型的 precision@N 、 recall@N 和 F1 Score 均优于单独的时间维度、空间维度和社交维度模型。由此可说明在推荐工人任务时,本文的综合模型更为准确,能为工人提供更好的服务体验。

由表 6 可见,由于数据集经过筛选后,每个工人参与的任务较少,随着推荐任务数目的增加,真正匹配工人兴趣的任务比例会相应减少,从而导致准确率下降。在推荐任务数目为 5 时,综合模型的准确率能够达到 0.9333,而当推荐任务数目增至 20 时,准确率下降至 0.6863,但仍保持较高的推荐精度。从整体来看,综合模型在不同推荐任务数下的准确率始终高于其他三个单一维度模型;由表 7 可见,随着推荐任务数量的增加,任务覆盖率随之提升,因此召回率也逐渐上升。在推荐数目为 20 时,综合模型的召回率达到 0.7179,在不同推荐任务数目下,综合模型的召回率均优于单维度模型;由表 8 可见,随着推荐任务数量的增加,召回率的提升速度通常快于准确率的下降,特别是在数据稀疏的情况下,召回率的提升显著推动了 F1 分数的上升。综合模型能够从多个维度捕捉工人的行为偏好,即使准确率有所下降,模型仍能依靠较高的召回率保持较高的 F1 分数。在推荐任务数目为 20 时,综合模型的 F1 分数达到 0.7018,展现出在准确率和召回率之间的良好平衡,整体推荐效果优异。

综上所述,本文提出的综合模型在三个评价指标(准确率、召回率、F1 分数)下均优于单一维度模型,能够深入挖掘工人的时空偏好变化,提升任务推荐的准确性与覆盖率,并有效缓解了社交网络数据稀疏性问题,进而使任务推荐更加精准。

Table 6. Accuracy of different models at different recommendation lengths N
表 6. 各模型在不同推荐长度 N 的准确率

模型	5	10	15	20
综合模型	0.9333	0.8424	0.7555	0.6863
时间维度模型	0.7151	0.5333	0.4121	0.3409
空间维度模型	0.4545	0.3242	0.2586	0.2258
社交维度模型	0.1879	0.2090	0.1717	0.1636

Table 7. Recall of different models at different recommendation lengths N
表 7. 各模型在不同推荐长度 N 的召回率

模型	5	10	15	20
综合模型	0.2441	0.4406	0.5927	0.7179
时间维度模型	0.1870	0.2789	0.3233	0.3566
空间维度模型	0.1189	0.1696	0.2029	0.2361
社交维度模型	0.0491	0.1094	0.1347	0.1711

Table 8. F1 scores of different dimensions at different recommendation lengths N
表 8. 各维度在不同推荐长度 N 的 F1 分数

模型	5	10	15	20
综合模型	0.3869	0.5786	0.6643	0.7018
时间维度模型	0.2965	0.3663	0.3623	0.3486
空间维度模型	0.1884	0.2227	0.2274	0.2308
社交维度模型	0.0779	0.1436	0.1510	0.1673

5. 结束语

目前移动群智感知领域两大主要的任务推荐方法：基于时空信息的任务推荐方法与基于社交网络信息的任务推荐方法，分别存在着忽略了工人偏好的时空变化与存在数据稀疏性问题。此外，还存在着多源异构数据融合问题。针对这些问题，本文综合考虑了工人地理位置、时间可用性以及社交网络关系多种因素，提出了基于多源异构数据融合的移动群智感知任务推荐算法。基于隐语义模型、双向长短时记忆网络模型分别生成时间维度工人-任务匹配度矩阵、空间维度工人-任务匹配度矩阵来挖掘工人偏好的动态变化以进行任务推荐，基于 Node2Vec 图嵌入技术模型生成社交维度工人-任务匹配度矩阵环节数据稀疏性问题，然后基于 BP 神经网络将以上三种工人-任务匹配度矩阵进行融合生成多维度工人-任务匹配度矩阵，实现工人行为多源异构数据的融合，并进行任务推荐。最后通过计算模型本身与模型变体的 precision、recall 和 F1 Score 性能指标来对比推荐效果，证明了本文算法能够在提高任务推荐准确率的同时实现多维度覆盖。

在今后的工作中，将考虑更多可能影响移动群智感知任务推荐效果的因素与多源异构数据融合方法，进一步提高任务推荐的精确度，为工人提供更好的服务体验。

参考文献

- [1] 江海峰, 商景杰, 王树豪, 等. 移动群智感知中发掘潜在高质量用户的激励机制[J/OL]. 小型微型计算机系统: 1-8. <http://kns.cnki.net/kcms/detail/21.1106.TP.20240205.1821.002.html>, 2024-06-06.
- [2] 童咏昕, 袁野, 成雨蓉, 等. 时空众包数据管理技术研究综述[J]. 软件学报, 2017, 28(1): 35-58.
- [3] Li, N., Mo, W. and Shen, B. (2016) Task Recommendation with Developer Social Network in Software Crowdsourcing. 2016 23rd Asia-Pacific Software Engineering Conference (APSEC), Hamilton, 6-9 December 2016, 9-16. <https://doi.org/10.1109/apsec.2016.013>
- [4] Yang, G., Li, Y., He, X., Song, Y., Wang, J. and Liu, M. (2021) Profile-Free and Real-Time Task Recommendation in Mobile Crowdsensing. *IEEE Transactions on Computational Social Systems*, 8, 1311-1322. <https://doi.org/10.1109/tcss.2021.3073031>

- [5] Wang, S., Zhang, L., Yu, M., Wang, Y., Ma, Z. and Zhao, Y. (2020) Attribute-Aware Multi-Task Recommendation. *The Journal of Supercomputing*, **77**, 4419-4437. <https://doi.org/10.1007/s11227-020-03440-6>
- [6] 王晓. 基于多模态信息融合的移动群智感知任务推荐方法[D]: [硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2023.
- [7] Ji, Y., Mu, C., Qiu, X. and Chen, Y. (2022) A Task Recommendation Model in Mobile Crowdsourcing. *Wireless Communications and Mobile Computing*, **2022**, Article ID: 9191605. <https://doi.org/10.1155/2022/9191605>
- [8] Zhao, B., Dong, H. and Yang, D. (2023) A Spatio-Temporal Task Allocation Model in Mobile Crowdsensing Based on Knowledge Graph. *Smart Cities*, **6**, 1937-1957. <https://doi.org/10.3390/smartcities6040090>
- [9] Lai, C. and Zhang, X. (2020) Duration-Sensitive Task Allocation for Mobile Crowd Sensing. *IEEE Systems Journal*, **14**, 4430-4441. <https://doi.org/10.1109/jsyst.2020.2967847>
- [10] Ding, Y., Zhang, L. and Guo, L. (2022) Dynamic Delayed-Decision Task Assignment under Spatial-Temporal Constraints in Mobile Crowdsensing. *IEEE Transactions on Network Science and Engineering*, **9**, 2418-2431. <https://doi.org/10.1109/tnse.2022.3163925>
- [11] Shen, X., Chen, Q., Pan, H., Song, L. and Guo, Y. (2022) Variable Speed Multi-Task Allocation for Mobile Crowdsensing Based on a Multi-Objective Shuffled Frog Leaping Algorithm. *Applied Soft Computing*, **127**, Article ID: 109330. <https://doi.org/10.1016/j.asoc.2022.109330>
- [12] 乐梦云. 基于知识图谱和社会网络的推荐算法研究及应用[D]: [硕士学位论文]. 北京: 北京邮电大学, 2021.
- [13] Chen, S., Zhao, X., Liu, J., Gao, G. and Du, Y. (2022) Social-network-assisted Task Recommendation Algorithm in Mobile Crowd Sensing. 2022 the 7th International Conference on Information and Education Innovations (ICIEI), Belgrade, 14-16 April 2022, 136-142. <https://doi.org/10.1145/3535735.3535751>
- [14] Wang, S., Wang, X., Yang, Y. and Cai, H. (2016) Based Point of Interest and Experience to Task Assignment on Location-Based Social Networks. 2016 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), Hefei, 16-18 December 2016, 81-85. <https://doi.org/10.1109/msn.2016.021>
- [15] Li, Y., Hsieh, C., Lin, L. and Wei, C. (2021) A Social Mechanism for Task-Oriented Crowdsourcing Recommendations. *Decision Support Systems*, **141**, Article ID: 113449. <https://doi.org/10.1016/j.dss.2020.113449>
- [16] Li, Z., Cheng, B., Gao, X., Chen, H. and Chen, G. (2021) A Unified Task Recommendation Strategy for Realistic Mobile Crowdsourcing System. *Theoretical Computer Science*, **857**, 43-58. <https://doi.org/10.1016/j.tcs.2020.12.034>