

基于 LSTM-Attention 的 RFID 移动对象位置预测



刘嘉琛 秦小麟 朱润泽

南京航空航天大学计算机科学与技术学院 南京 211106

(liujiachen@nuaa.edu.cn)

摘要 随着射频识别(RFID)技术的不断发展,其相比全球定位系统(GPS)具有高精度、数据信息量大的优势,将其应用于智能交通以预测移动对象位置受到广泛关注。然而,由于其定位基站分布离散,并且不同基站对位置预测的影响权重不同,以及长期的历史信息会带来维数灾难等,移动对象的位置预测面临着严峻的挑战。针对这些挑战,在分析现有预测算法的不足的基础上,提出了一种长短期记忆网络(LSTM)和注意力(Attention)机制相结合的机器学习模型(LSTM-Attention)。该算法将 one-hot 编码后的输入向量通过神经网络进行降维处理后,利用注意力机制来发掘不同的定位基站对位置预测的权重影响,最后进行位置预测。在南京交管局提供的 RFID 数据集上进行的对比实验表明,与现有算法相比,所提算法在预测准确性上有明显的提升。

关键词:RFID;降维;位置预测;长短期记忆网络;注意力机制

中图分类号 TP311

Prediction of RFID Mobile Object Location Based on LSTM-Attention

LIU Jia-chen, QIN Xiao-lin and ZHU Run-ze

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract With the continuous development of radio frequency identification (RFID) technology, due to its advantages of high accuracy and large amount of data information compared to global positioning system (GPS), the application of RFID to intelligent transportation to predict the location of moving objects attracts widespread attention. However, due to the discrete distribution of its positioning base stations, the different influences weight of different base stations on position prediction, and the long-term historical information will bring dimensional disasters and other issues, and the position prediction of mobile objects is facing severe challenges. In response to these challenges, based on the analysis of the shortcomings of existing prediction algorithms, a machine learning model combining long short-term memory (LSTM) and attention mechanism is proposed. This algorithm reduces the dimension of the input vector encoded by one-hot through the neural network, and uses the attention mechanism to explore the weighting effect of different positioning base stations on position prediction, and finally performs position prediction. Comparative experiment on the RFID data set provided by Nanjing Traffic Management Bureau shows that compared with the existing algorithms, the LSTM-Attention algorithm has a significant improvement in prediction accuracy.

Keywords Radio frequency identification, Dimension reduction, Location prediction, Long short-term memory, Attention mechanism

1 引言

移动对象位置预测是位置服务(Location Based Services, LBS)的重要技术部分,使用时间复杂度低、准确率高的算法来处理海量的移动对象数据并进行高效的预测,具有重要的价值和意义。

现今,随着第五代通信技术的飞速发展,高度智能化的移动设备影响着人们的衣食住行,并且由此产生了海量的数据,在这些数据中快速挖掘到移动对象的运动信息,成为我们研究的重点。轨迹信息是移动对象最为重要的上下文信息,历

史的轨迹信息对移动对象未来的轨迹预测具有极其重要的意义,可为城市生活中的交通管理、出行规划等提供技术支持,如智能导航系统(Intelligent navigation system)、推荐系统^[1](Recommended system)、轨迹简化^[2](Trajectory simplified)、城市大数据^[3]等。由此产生的需求以及可用性高、信息量大的数据使移动对象的轨迹位置预测逐渐成为研究热点。

射频识别技术(Radio Frequency Identification, RFID)的数据通信方式为阅读器与标签的非接触式识别,因此已被广泛应用于车辆识别定位。RFID 卡片定位与全球定位系统(Global Position System, GPS)是两种不同类型的定位方式。

收稿日期:2020-06-22 返修日期:2020-09-17 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61728204)

This work was supported by the National Natural Science Foundation of China(61728204).

通信作者:秦小麟(qinxcs@nuaa.edu.cn)

RFID 通过无线电天线向特定目标频繁发送信号来进行定位,在天线基站密集的环境下,其区域定位具有较高的精确度^[4]。RFID 利用离散分布的基站监测到移动对象通过基站时的信息特征,其中部分信息特征存储在电子卡片上,如车辆的车牌、车辆颜色。另外的移动对象数据,如通过时间、通过的车道号、通过速度等信息也可以快速、准确地监测到。由于精确性高、利用率高、信息量大的特点,RFID 数据成为了研究的重点。RFID 技术相比全球定位系统具有独特的优势,在智能交通方面应用广泛。

自 2011 年起,南京市 RFID 检测基站已经覆盖了主城区以及周边区县,检测点数量达到 518 处,南京市安装有 RFID 电子标签的车辆达到上百万辆,因此可以获取海量的数据。对如此大规模的 RFID 数据进行分析与发掘以及应用推广成为了目前亟需解决的问题。

移动对象轨迹位置预测方法的一般步骤为:首先将获取到的移动对象数据抽象化,筛选出需要的数据信息,并将获取到的时间信息和空间信息进行预处理,其中包括对缺失数据、噪声数据,以及冗余数据的处理;然后,选择合适的计算机模型对预处理的数据进行分析和处理,挖掘其运动模式^[5];最后,对移动对象进行轨迹位置预测。

随着智能交通和 5G 的发展,车辆、用户甚至一些具有时空特性的移动对象的位置预测已经得到了广泛的研究,许多与轨迹研究相关的技术被提出并应用。其核心思想是利用移动对象的历史信息数据挖掘出一般的运动模式,计算出下一个到达概率最高的轨迹位置。一般常用的方法有 3 种。1) 通过在数据集中发现项与项之间的关系来发现移动对象的运动规律,构建关联规则,然后将当前移动对象轨迹与之匹配,最后通过序列的匹配进行预测^[6]。2) 马尔可夫链(Markov Chain, MC)模型。马尔可夫模型是一种统计模型,每一个位置状态转移只依赖于前一个状态,与再之前的状态无直接关系。模型通过此性质来建立各位置状态转移矩阵^[7],并从矩阵中找到移动对象的最大转移概率来预测移动对象下一步的位置^[8]。3) 机器学习方法,如循环神经网络(Recurrent Neural Network, RNN)等机器学习模型。大量的神经元非线性地连接在一起,各节点之间的权重随着数据和时间进行调整,从而学习出特定数据集下的运动模型来对移动对象进行预测^[9]。

虽然以上 3 类方法在交通预测、图像识别、轨迹数据分析等场景取得了很好的效果,但是它们均是针对时间跨度短、信息量小的轨迹数据序列,并没有考虑到将长期历史信息作为输入会导致模型出现维数灾难的问题,以及某些关键的地理位置对未来位置预测的影响。同时,通过 GPS 以及一些定位设施获取到的轨迹数据只是单一经纬度信息或者地理位置信息,挖掘出的轨迹信息采用连续的数字以及符号表示,而 RFID 数据包含丰富的移动对象上下文语义信息,在特定的语义环境下,轨迹数据上下文信息对位置预测也有着重要的影响。若忽略长期的历史信息,则不确定因素增多,对未来轨迹位置的预测会产生较大的影响,预测的准确性不高。

随着人工智能和深度学习研究的深入,研究者开始使用神经网络来预测轨迹位置。RFID 数据中包括电子车牌、车道号、时间等信息,这些信息具有多维度和时效性的特点,RNN

可能造成梯度爆炸或梯度消失的问题^[10],并且在处理长距离依赖的问题上,RNN 无法连接相关的信息。长短期记忆网络(Long Short-Term Memory, LSTM)^[11]是一种特殊的 RNN,它能够学习长期依赖,因此被用来处理长期时序数据。但是该网络模型中隐层之间神经元全相连,导致对输入神经网络的向量维度有所限制,若输入模型的向量维度过大,则会因各层神经元信息传播时的矩阵运算而出现维数灾难问题,进而大大降低模型的学习效率。此外,在交通轨迹预测问题中,不同的位置信息对预测的影响力和重要程度不同,简单的 LSTM 也无法解决这个问题。

近年来,注意力(Attention)机制被提出,将其应用于神经网络尤其是分类预测问题效果显著。在神经网络中加入注意力机制,通过对神经网络中间层神经元分配不同的概率权重,使得神经网络模型能够更加关注对预测结果影响较大的数据信息,降低甚至忽略对分类预测影响较小的一些信息的关注。因此,在特定的移动对象位置预测问题上,加入注意力机制的隐层能够进一步提高预测的准确率,这在句法分析、分类预测等任务中已经得到了很好的证明。

针对上述问题,基于机器学习的方法,本文提出了一种基于 LSTM 并且加入了注意力机制(Attention Mechanism)的算法。该模型将位置基站通过 one-hot 编码后生成的输入向量进行神经网络降维处理后,同时在 LSTM 隐层中引入注意力机制来发掘不同的定位基站对位置预测的权重影响,最后在南京市交管局提供的 RFID 数据集上进行位置预测。实验结果表明,该方法考虑了数据具有多维性和时效性的特点,明显提高了预测模型的效率和准确度。

2 相关工作

近年来,随着智能设备的发展,获取到海量的移动对象数据变得更加容易,因此对移动对象位置预测的研究不断深入。移动对象的轨迹位置预测通常是根据历史轨迹数据来进行预测。首先对历史数据进行预处理,处理噪声数据、冗余数据,以及将数据转化为神经网络可以处理的输入向量,其次选择和构建合适的模型来进行学习和对移动对象进行预测。

轨迹数据中往往存在时间和空间维度,呈现出多维、语义、时空动态关联的复杂性,通常使用启发式搜索固定阈值和基于聚类的提取方法^[12]寻找和筛选轨迹的常驻点。

基于启发式搜索固定阈值的方法利用的是数据的时空信息。Li 等^[13]通过划分停留点,在移动对象运动轨迹范围中找到超出特定时间的驻留点,其所处的区域就构成了停留区域。对于静止的停留点,通过距离阈值、速度即可识别。对于徘徊型停留点,则需要判别时间跨度是否超过阈值。基于聚类的停留点提取方法依赖于 K-Means、均值漂移聚类、具噪声基于密度的空间聚类算法(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)。Birant 等^[14]在 DBSCAN 算法的基础上增加了时间信息,提出了 ST-DBSCAN 算法,该算法更加适用于具有时空信息的数据。Palma 等^[15]提出了 CB-SMoT 算法,从移动对象的其他运动特性出发,根据其速度识别出停留区域。

Koren 等^[16]将包含空间地理位置信息的移动对象行为通过概率矩阵分解,来预测该移动对象最有可能到达的下一

地理位置。Xiong 等^[17]优化了文献[16]的方法,并使用张量分解的方法,将历史轨迹信息中的时空信息考虑在内,提高了预测精确度。上述算法都没有考虑移动对象当前信息以及长期历史轨迹信息对位置预测的影响,很难学习到序列中长距离的依赖关系,故模型的预测准确率不高。

Ying 等^[18]分析了轨迹和地理的语义特征,通过分析不同语义环境下轨迹的特点来预测移动对象下一时刻的位置。Monreale 等^[19]通过从移动对象历史轨迹位置数据中挖掘出移动对象的运动模式,来预测该用户接下来的移动位置,并在前人的基础上提出了 WhereNext。Morzy^[20]通过改进 Apriori 算法来生成特定规则,使用改进的 PrefixSpan 算法从大量的偏序现象中挖掘经常出现的偏序规律,来预测移动对象轨迹位置。但是,由于 PrefixSpan 算法的局限性,其每次都要重新从大量的偏序现象中挖掘移动对象的频繁轨迹模式,使得该算法效率低下。

马尔可夫链模型利用了移动对象轨迹序列的状态特征,其状态分布只取决于现在,与过去无关,因此可以使用马尔可夫链描述移动对象轨迹序列在各个位置之间的状态转移概率,进而进行位置预测。在交通问题中,可通过查看隐藏层转换矩阵的转换概率,根据最后一个结点的隐藏状态预测未来移动对象可能到达的位置。Rendle 等^[21]通过因式分解概率转移矩阵来优化马尔可夫模型,在处理轨迹数据、预测轨迹位置方面取得了良好的效果。Mathew^[22]在马尔可夫模型的基础上,聚焦只依赖于当前时刻的马尔可夫链的状态,并利用隐马尔可夫模型(HMM)来预测移动对象轨迹位置。但是对于多个位置状态的情况,当数据量增加时,状态信息会呈现指数式增长,从而增大了预测的时间复杂度^[23]。

随着神经网络的发展,其在数据挖掘和处理时序信息方面的应用逐渐广泛。而 RNN 的提出解决了大量的时序数据问题。此外,一些 RNN 的变体也逐渐被提出以提高预测精度。Liu 等^[9]改进了 RNN 算法,利用具有时空上下文的数据来训练 RNN 模型,预测移动对象在某个时空的位置。但是传统的 RNN 并不能解决长期历史轨迹信息所造成的信息遗失,以及训练时出现的梯度消失以及梯度爆炸问题。随着 LSTM 的提出,这些问题均得以解决。LSTM 由于其特殊的结构,在机器翻译、信息检索^[24]等方面已经取得了很好的成就,但其仍然没有考虑到历史数据的重要程度对预测的影响。

为了更好地捕捉神经网络模型在学习中的不同关注点,近年来一些学者提出了注意力(Attention)机制,如 SoftAM、GlobalAM^[25]和层次化 AM 等,并且将其成功地运用于机器翻译^[25]、文本摘要^[26]、阅读理解^[27]等任务中。注意力机制使得 Decoder 端在生成目标序列时可以将 Encoder 端的输出序列考虑在内,生成一个固定大小的中间向量。这不仅可以体现输入源中各个基站对目标结果的不同影响力,还可以缓解由于输入向量维度过高所导致的细节信息丢失的问题。

3 RFID 原始数据的预处理与降维

3.1 数据预处理

为了便于说明,本文所用的轨迹信息数据符号表示以及其含义如表 1 所列。

表 1 符号定义说明

Table 1 Description of symbol definition

Symbol	Description
$A = \{A_1, A_2, \dots, A_j, \dots, A_{m-1}, A_m\}$	Base station dataset
$H = \{T_1, T_2, \dots, T_h\}$	Base station location sequence
l_i	l_i 'one-hot vector
e_i	e_i ' position embedding vector
W_i	W_i ' Weight matrix

RFID 原始数据集是基站监测到的通过该基站的所有车辆的数据,数据以基站为主体。由于 RFID 基站所采集到的每一条数据为某一时刻通过定位基站的车辆信息,在时空上难以直观地表示车辆的运动轨迹,不能从中挖掘出其运动模式,因此无法对车辆轨迹进行预测。针对海量的原始交通数据,需要通过数据预处理将其转化为以车辆为主体的轨迹数据结构,如图 1 所示。

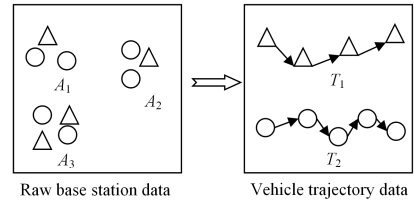


图 1 将原始 RFID 数据转化为轨迹数据

Fig. 1 Convert original RFID data to track data

通过 Python 中的数据分析模块 pandas 库将原始 RFID 数据中以基站为主体的移动对象数据转化为以车辆为数据主体的轨迹信息,以便于我们对移动对象进行轨迹预测。通过以上转换,将 RFID 原始数据转化为移动对象通过 h 个基站的轨迹序列集 $H = \{T_1, T_2, \dots, T_h\}$,这样就可以将离散的基站信息转化为连续的轨迹信息,便于进行轨迹位置预测。故移动对象位置轨迹预测问题为:设移动对象通过的 h 个基站组成的轨迹集为 $H = \{T_1, T_2, \dots, T_h\}$,假设当前的轨迹位置为 $T = \{\langle L_i, t_i \rangle\}, 0 \leq i \leq t$,则预测移动对象在 $t+1$ 时刻所在的位置为 T_{i+1} 。

3.2 对基站数据进行 one-hot 编码

将离散基站信息转化为连续的轨迹数据位置序列后,该序列是由基站编号组成的轨迹数据,蕴涵了时空信息,但是由于每个位置是离散的基站序号,各位置之间无关联性,无法以向量的形式输入到神经网络 LSTM 进行训练,需要将基站编码转化为神经网络可以处理的输入向量。通常数据预处理使用独热(one-hot)编码的方法将离散值进行编码,采用 N 位状态寄存器来对 N 个状态进行独热编码。具有 n 个位置的 RFID 基站集合 $A = \{A_1, A_2, A_3, \dots, A_j, \dots, A_{n-1}, A_n\}$ 。对于每一个基站 A_i ,我们将分类变量用二进制表示,将分类值映射到整数,对于编码后的向量,除了索引第 i 位标记为 1 以外,其他位都是 0。如位置 A_2 可以表示为: $(0, 1, 0, 0, 0, \dots)$,其维数即为基站数量。这种编码的优点就是将离散的基站编码后转化为机器学习可以处理的输入向量并将向量输入到 LSTM 中,在一定程度上起到了扩充特征的作用。但是由于基站数目多,one-hot 编码后的向量维度较大,通过全连接神经网络训练后会导致维数灾难的问题。

3.3 利用神经网络对输入向量降维

针对以上问题,本文采用神经网络对输入向量进行降维。

神经网络将通过 one-hot 编码后的基站位置信息输入到降维模型中,将其转化为低维度的位置嵌入向量(Location Embedding Vector, LEV),从而解决了维数爆炸增长的问题。神经网络的结构如图 2 所示。

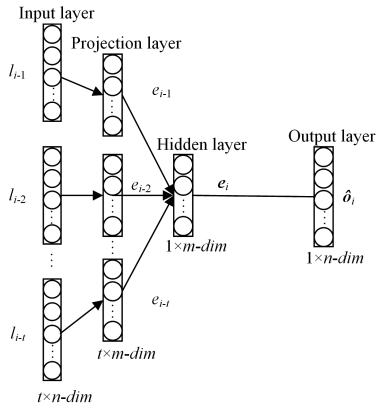


图 2 神经网络结构

Fig. 2 Structure of neural network

假设一个长度为 N 的位置序列出现的概率为 $P(L_1, L_2, L_3, \dots, L_n)$, 在实际路况中,移动对象当前的位置与历史轨迹是密切相关的,历史的位置信息影响着当前的轨迹位置。假设在一段轨迹位置序列中,该车辆当前经过的位置和历史经过的前 t 个基站有关,概率计算如式(1)所示:

$$P(L_1, L_2, \dots, L_n) = \prod_{i=1}^N P(L_i | L_{i-t}, L_{i-t+1}, \dots, L_{i-1}) \quad (1)$$

该神经网络的核心思想是,在对 one-hot 向量进行降维处理的同时,使得 $P(L_i | L_{i-t}, L_{i-t+1}, \dots, L_{i-1})$ 最大化。假设 RFID 数据集中共有 n 个基站,那么通过独热编码后,每个位置构成一个输入向量 l ,该向量是 n 维的,通过降维神经网络的第一层后,每个 one-hot 向量转化为一个 m 维的位置嵌入向量。设该基站编号经过 one-hot 编码后的输入向量为 l_i ,那么与之相关的前 t 个历史轨迹位置分别为 $\{l_{i-t}, l_{i-t+1}, \dots, l_{i-1}\}$ 。假设由编码后的向量所构成的输入矩阵为 $V \in R^{m \times n}$,那么由式(2)可以计算得出映射层每个基站位置所对应的中间向量为 $\{e_{i-t}, e_{i-t+1}, \dots, e_{i-1}\}$ 。

$$e_r = V l_r, i-t \leq r \leq i-1 \quad (2)$$

设该模型的输出矩阵为 $U \in R^{n \times m}$,中间向量 e 通过式(3)输出为 n 维的向量 \hat{O}_i 。

$$\hat{O}_i = \text{softmax}(U e_i) \quad (3)$$

其中, \hat{O}_i 为 n 维输出向量, softmax 为归一化函数,将输入的向量映射为 $0 \sim 1$ 的实数。 \hat{o}_i 表示输出向量中概率最大的位置对应的 one-hot 向量。由分析可知,该降维模型的输出向量,即预测结果,应该与当前基站所对应的独热编码向量相同,即 $\hat{o}_i = l_i$ 。因此该降维神经网络模型的损失函数的计算公式为:

$$\text{loss}(\hat{O}_i, o_i) = - \sum_{j=1}^n o_{ij} \ln(\hat{O}_{ij}) \quad (4)$$

为了避免模型的泛化性能下降,使用梯度下降的方法让模型优化自身参数,得到最小的损失函数和参数模型值,通过计算输出向量 \hat{u}_i 和中间嵌入向量 e_i 得到矩阵 V 和 U 。然后通过公式 $e = V l$ 即可得出每个基站 one-hot 向量所对应的降维处理后的位置向量。其中向量 e 为 LSTM-Attention 模型的

输入向量。上述计算流程如算法 1 所示。

算法 1 数据预处理降维算法

输入:每个基站的车辆信息集合 A ,相关位置数 t

输出:每个位置对应的位置嵌入向量构成的集合 S_{LEV}

Step1 $A \rightarrow H = \{T_1, T_2, \dots, T_h\}$ // 数据转化

Step2 $H \rightarrow S_{\text{onehot}}$ // 轨迹数据 one-hot 编码

Step3 For S_i in all H // 遍历基站轨迹序列

Step4 $\text{data} = \{l_{i-t}, l_{i-t+1}, \dots, l_{i-1}\}, \text{label} = l_i$ // 构建数据集

Step5 Put $\{\text{data}, \text{label}\}$ into Trainset // 构建训练集

Step6 Put Trainset into model // 将训练集合输入模型

Step7 Solve Objective function // 数据通过神经网络并求解目标函数

Step8 Adjustment parameter // 反向传播并调整神经元各节点间的参数

Step9 重复 Step6—Step8,直到神经网络学习完全部的轨迹序列并且学习次数达到 epoch,得到矩阵 V 和 U

Step10 For S_i in all H // 遍历轨迹序列

Step11 For l_i in S_i // 遍历轨迹中的基站

Step12 $e_i = U \cdot l_i$ // 得出每个基站的嵌入向量

Step13 put each e_i into S_{LEV} // 保存结果

上述算法首先将以基站为主体的原始 RFID 数据转化为连续的以车辆为主体的基站轨迹数据;其次对基站进行 one-hot 编码,将离散的基站编号编码为神经网络可以识别的输入向量;最后,将编码后的向量输入到神经网络模型中进行学习和降维处理,通过计算得到一个合适维度的中间嵌入向量,作为 LSTM-Attention 模型的输入。

由算法 1 得知,通过该神经网络可将每个位置的 one-hot 向量转化为低维度位置嵌入向量,有效地解决了 one-hot 向量带来的数据稀疏与信息丢失问题。

4 LSTM-Attention 预测模型

在对移动对象位置进行预测时,LSTM 通常是自动识别历史轨迹数据与预测轨迹的关系,而且输入轨迹序列的每一项对于预测输出结果的影响权重一样,因此输入和输出的关联性较小,不仅预测效率低下,而且准确度也较低。而 Attention 机制是计算当前隐藏层的状态输出值与自身匹配度的机制,故本文设计了一个基于 LSTM 网络构建的注意力机制神经网络,分为 LSTM 层和注意力层。LSTM-Attention 模型如图 3 所示。

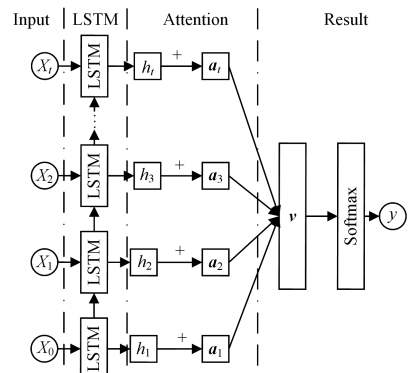


图 3 LSTM-Attention 模型

Fig. 3 LSTM-Attention model

4.1 LSTM 层

LSTM 在 RNN 的基础上引入了细胞状态,根据细胞的

状态可以判断哪些状态应该保留下来,哪些状态应该被遗忘,在一定程度上解决了梯度消失问题,在处理长期历史问题上比 RNN 有更好的表现。LSTM 单元由 4 个不同的组件构成,分别是遗忘门、输出门、输入门和单元状态。LSTM 利用输出门来选择信息,选择的信息表示为 $y(x) = \sigma(Wx + b)$, W 为权重矩阵, b 为偏移量。LSTM 的单元结构如图 4 所示。

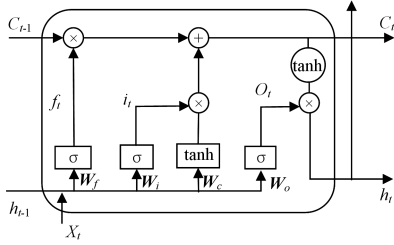


图 4 LSTM 单元示意图

Fig. 4 Structure of LSTM network

图 4 中, x_t 和 h_t 为 t 时刻的输入和输出, h_{t-1} 为前一个 LSTM 单元的输出, C_{t-1} 为前一个单元的记忆。 C_t 为 t 时刻当前 LSTM 单元的状态值。LSTM 网络的各单元计算如下。

LSTM 记忆单元对上一 LSTM 单元传来的数据信息进行处理,决定了遗忘的信息;遗忘门接收一个长期的记忆 C_{t-1} 并决定要保留和遗忘 C_{t-1} 的哪个部分,故通过遗忘门来计算遗忘因子 f_t , 即 $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$ 。其中, W_f 是输入到该 LSTM 单元的权重矩阵; b_f 为其所对应的偏移量; σ 是 *sigmoid* 激活函数,可以产生 0~1 之间的值。

输入门通过计算出 i_t 来控制当前的输入对接下来的记忆单元的影响。根据 RNN 的激活函数(见式(5))来计算当前时刻的候选记忆单元值 \tilde{C}_t , 其中激活函数 \tanh 的取值在 (0, 1) 之间。

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \end{cases} \quad (5)$$

如式(6)所示,计算得出新的单元状态,单元状态逐个点乘遗忘向量。如果它乘以接近 0 的值,则有可能在单元状态中丢弃。然后从输入门获取输出并进行逐点加法,并将单元状态更新,计算当前状态的状态值,这就得到了新的单元状态。

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (6)$$

最后确定输出值,输出门通过 *sigmoid* 函数计算决定输出的部分,具体如式(7)所示:

$$\begin{cases} o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \quad (7)$$

经过大量的数据训练后,LSTM 各层各神经元之间的权重发生改变,误差函数缩小到一个可接受的范围,从而完成了神经网络对训练数据的学习。

4.2 注意力层

目前的 LSTM 机制通常将输入神经网络的向量设为固定的维度,在处理低维度输入向量时,模型效果比较好。然而,当输入序列非常长时,由于模型的全相连特性,会造成维度爆炸问题,模型难以学习到合理的向量表示,模型的性能也会变得较差。为了更好地捕捉学习过程中的不同关注点,本文采用 attention 机制解决了传统编码器-解码器(Encoder-

Decoder)结构在处理数据时将输入序列转化为一个固定长度的向量然后再经过神经网络输出的问题,即不论输入序列和输出序列长度如何,中间的向量长度都是固定的。Attention 机制保留了以上结构对输入向量的中间输出结果,并在此基础上添加了注意力机制,更加关注到不同信息对预测结果的影响权重,对不同的输入进行更加智能化的选择性学习,将模型训练成一个适合当前数据集的预测模型。本文基于 LSTM 模型来预测移动对象位置,同时引入 Attention 机制来解决该体系结构在长序列上的局限性,丰富了上下文关键信息。

Attention 层的基本结构如图 5 所示。

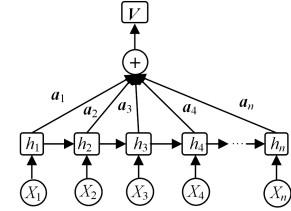


图 5 Attention 机制示意图

Fig. 5 Attention mechanism diagram

首先将 LSTM 隐藏层的输出 h 经过一层非线性转化得到 u_i , 如式(8)所示:

$$u_i = \tanh(w_i H_i + b) \quad (8)$$

其中, w_i 为权重矩阵。在路网交通中,一些重要交通位置的车流量显然较大,故在移动对象轨迹序列中,这些关键位置点出现的次数也是最多的。那么根据其在轨迹中出现次数所占的比例,给每一个位置赋予初始重要性权值,计算方式如式(9)所示:

$$W_i = \beta \frac{\text{count}(l_i)}{\sum_{j=1}^n \text{count}(l_j)} \quad (9)$$

其中, $\text{count}(l_i)$ 为位置 l_i 在整个序列中出现的次数, $\sum_{j=1}^n \text{count}(l_j)$ 为每一个位置在所有序列中出现的次数, β 为比例系数且 $\beta > 0$ 。模型训练不断更新,然后通过一层 *softmax* 进行归一化操作得到注意力权重矩阵 a_i , 如式(10)所示:

$$a_i = \text{softmax}(u_i^T, u_w) \quad (9)$$

将 a_i 和 H_i 经过注意力机制得到最终向量 v_i , 如式(11)所示:

$$v_i = \sum a_i H_i \quad (11)$$

通过概率权重的分配,对不同的基站分配不同的权重,使得一些关键路口基站能够得到更多的关注,从而提高预测的准确率。最后 Attention 机制的输出 V_i , 通过 *softmax* 函数进行归一化处理,输出 y 如式(12)所示:

$$y = \text{softmax}(w_y V_i + b) \quad (12)$$

以上公式中, u_i 为隐藏层单元, a_i 为注意力向量, H_i 为 LSTM 的输出, v_i 为 Attention 机制的输出向量。

LSTM-Attention 算法如算法 2 所示。

算法 2 LSTM-Attention 算法

输入:降维后的输入集合 S_{LEV} , 轨序列集合 H , 待预测位置序列 T_{test}

输出:one-hot 编码的位置预测结果 l_{result}

Step 1 Input = S_{LEV} // 由算法 1 得到输入向量集合

Step 2 For T_j in H // 遍历基站轨迹序列

Step 3 For I_i in T_j & I_i in S_{LEV} //遍历轨迹中的基站并且基站已通过降维处理

Step 4 $data = \{I_{i-1}, I_i, I_{i+1}, \dots, I_{i-1}\}$, $label = I_i$ //构建数据集

Step 5 Put $\{data, label\}$ into Trainset//构建训练集

Step 6 Put Trainset into LSTM-Attention //将训练集合输入 LSTM-Attention

Step 7 Solve Objective function //数据通过神经网络并求解目标函数

Step 8 Calculate error function //通过输出向量和期待结果向量计算误差函数

Step 9 Adjustment parameters //反向调整网络各神经元节点间参数

Step 10 重复 Step 6—Step 9,直到神经网络学习完全部的轨迹序列并且学习次数达到 epoch

Step 11 Put T_{test} into LSTM-Attention and get e_{output} //输入模型得到输出向量

Step 12 Output corresponding I_{result} of e_{output} //转化为 one-hot 向量并输出

在算法 2 中,将经过神经网络降维处理后的向量作为 LSTM-Attention 模型的输入。该算法在 LSTM 的基础上引入了 Attention 机制,并关注不同地理位置对移动对象位置预测的不同影响程度。

5 实验结果与分析

5.1 数据集与实验环境

为了验证本文所提算法的有效性实用性,我们采用南京交管局提供的 RFID 车辆数据集进行测试。该数据集包括 518 个基站检测到的车辆信息,每个基站都与南京市地理经纬度相映射,包括具体的地理位置信息,时间覆盖了 2014 年 9 月 1 日 0 时到 9 月 31 日 24 时上百万辆车的轨迹信息。数据包括通过时间、基站编号、车牌号、车道号等信息。本文关注时间位置信息,并通过 pandas 库将这些信息转化为轨迹信息,然后利用机器学习的方法进行预测。

实验在 PC 机上进行,模型基于 pytorch 深度学习框架。本次实验的运行环境为:ubuntu16.04 (64 bit), IntelCorei7-8750 CPU, 16 GB RAM, NVIDIA GTX1060Ti GPU。

5.2 预测评估标准

在路网交通位置预测中,常用准确率(Accuracy)和召回率(Recall)来衡量模型的性能。

在多元分类任务中,将一类样本设置为正例,即正样本,其余样本则定义为负例,即负样本。模型预测结果可以用正类和负类来表示,被正确分类的正类和负类分别称为真正类(TruePositive, TP)和真负类(TrueNegative, TN)。将负类样本预测为正类,则称其为假正类(FalsePositive, FP),而将正类样本预测为负类,称其为假负类(FalseNegative, FN)。由此,在路网交通语义空间中,我们定义准确率为正确预测的基站占测试样本总基站的比例,以此作为衡量模型性能指标之一,如式(13)所示:

$$pre = \frac{TP + TN}{(TP + TN + FN + FP)} = \frac{rightquantity}{total} \quad (13)$$

本文还以召回率作为衡量指标,其表示为正确预测正类的样本数与正样本的比例,即正确预测为该基站的个数与所有预测为该基站的样本数之比,如式(14)所示:

$$recall = \frac{TP}{TP + FN} \quad (14)$$

5.3 LSTM-Attention 算法内部调参的结果比较

5.3.1 降维模型在不同中间维度下的准确率比较

该数据集中位置基站共有 518 个,因此通过 one-hot 编码后的向量共有 518 维。通过数据的预处理和 one-hot 编码后,原始 RFID 数据转化为车辆轨迹信息数据,然后将其按比例划分为轨迹位置序列训练集与测试集。在将数据输入到 LSTM-Attention 中进行预测前,要对基站 one-hot 编码后的向量进行降维处理,并选取合适的中间嵌入向量维度。假设神经网络降维得到的嵌入向量维度为 m ,那么 m 的取值也会对预测的结果有较大的影响, m 太小会造成大量的信息丢失, m 太大则仍然存在维数灾难的问题。因此我们需要对 m 取值的合理性进行判断,通过交叉验证计算预测准确率。实验结果如图 6 所示。

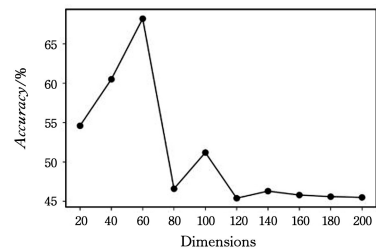


图 6 在不同维度下的预测精度

Fig. 6 Prediction accuracy with different dimensions

由图 6 可知,降维位置嵌入向量在 40~80 维之间时,模型的预测精度最高,因此选择中间值 60 维作为降维模型的嵌入向量。

5.3.2 LSTM-Attention 在不同 epoch 下的比较

完整的数据集通过了神经网络并且反向传播一次,这个过程称为一次 Epoch,所有的训练样本在神经网络中都进行了一次正向与反向传播。为了更好地学习数据间的关系,在神经网络中传递完整的数据集仅一次是不够的,往往需要训练多次才能学习出数据中的关系。虽然增加训练次数会提高准确率,但会造成过拟合的情况。

图 7 所示为 LSTM-Attention 模型在不同 epoch 下的预测准确度。

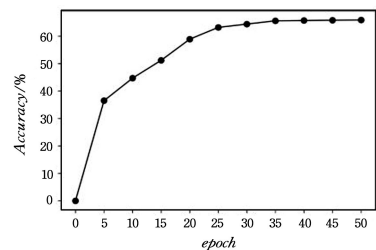


图 7 在不同训练轮次下模型的预测精度

Fig. 7 Prediction accuracy of model with different epoches

由图 7 可知,LSTM-Attention 模型的准确率随着训练轮次 epoch 的增加而提高。在 20 次之前随着 epoch 的增加,模型预测准确率明显提高,后期虽然准确率继续提高,但是幅度明显减小,趋于平稳。为了防止出现过拟合现象,即模型对训练数据的预测准确度提高,但是处理数据的能力下降,模型的

泛化能力较弱,故选取 epoch 为 25 进行训练,在保证预测准确率的前提下,提高模型处理一般数据的预测能力,提高模型的综合性能。

5.3.3 LSTM-Attention 在不同梯度更新块下的比较

梯度更新块 Batch_Size 是指 LSTM-Attention 模型一次训练所选取样本的数量,其大小对模型的优化程度和速度均有影响。神经网络在训练时不会将数据全部输入神经网络,而是分成几个 batch 输入到神经网络模型,选择合适的 Batch_Size 可以提高模型的效率。随着数据集的增大和内存的限制,一次载入所有的数据(即 Batch_Size 最大),选取一个全局的学习率并不可行。若 Batch_Size 太小则模型难以收敛,时间开销增大,模型性能变差。图 8 给出了在不同梯度更新块大小下,LSTM-Attention 模型在测试集上的预测准确率以及花费时间的比较。

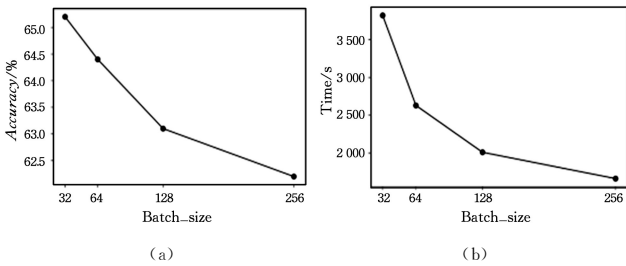


图 8 模型在不同梯度更新块下的预测准确率

Fig. 8 Prediction accuracy of model with different Batch_Size

从图 8 可以看出,随着梯度更新块的增大,模型的训练时间和预测准确率都有所下降。因此本文综合考虑选取 Batch_Size 为 32。

5.4 不同算法对比结果的分析

经过以上分析,选取维度 60 作为降维神经网络模型的中间嵌入向量,选取 epoch 为 25, Batch_Size 为 32,作为 LSTM-Attention 模型的参数进行预测。本文以 Pytorch 框架进行网络模型结构的搭建,对 RFID 数据中移动对象的轨迹位置进行预测。本文将传统的轨迹位置预测方法 2-MC^[8], HMM^[22], XGBoost^[28],以及神经网络 LSTM, STS-LSTM^[29]与本文模型进行对比实验。

图 9 和图 10 分别以准确率和召回率作为评价模型指标,对比了 5 不同方法得出的预测结果。

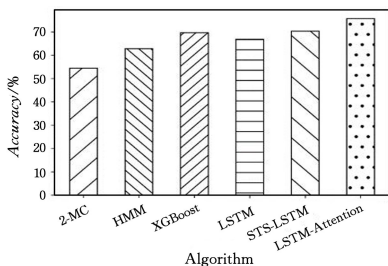


图 9 位置预测准确率

Fig. 9 Location prediction accuracy

由图 9 可以看出, LSTM-Attention 的准确率较传统的 LSTM 和 STS-LSTM 分别高出了 8.8% 和 5.4%,取得了显著的提升。在 LSTM 预测方法的基础上加入 Attention 机制后,预测的准确率比未加入 Attention 机制有了显著提高,说

明该算法更能关注到路网数据中不同基站的不同影响力,从而对模型不同节点分配不同的权重,提高了模型的预测率。

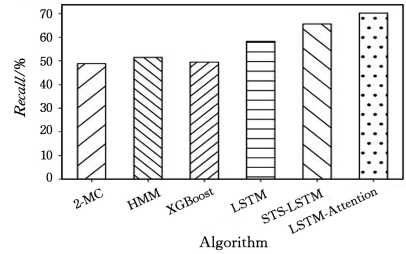


图 10 位置预测召回率

Fig. 10 Recall rate of location prediction

从图 10 可以看出,在召回率方面 LSTM-Attention 算法有了大幅提升,相比 LSTM 和 STS-LSTM 分别提高了 10.9% 和 4.6%,说明该算法预测分类正确的样本数在测试集中的比例有明显的提高,从另一方面反映了模型性能的提升。

通过比较图 9 和图 10 的实验结果可以发现:

1)神经网络的预测准确率以及召回率明显优于传统的位置预测算法。2-MC, HMM 以及 XGBoost 方法都是针对短时间、短信息的轨迹序列, LSTM 没有考虑到过长的历史信息对预测性能的影响,这正说明了神经网络在对路网位置预测以及处理历史时序问题上的优越性。

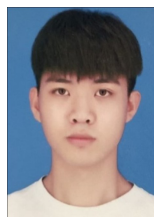
2)本文提出的基于注意力机制的神经网络模型相比普通的 LSTM 以及 STS-LSTM 在性能上有了明显的提升。这是因为本文模型考虑到历史过多信息带来的维数灾难问题,对输入向量进行了降维处理并且关注到在路网空间中不同的位置对预测的影响不同,这与实验结果相吻合。

结束语 移动对象轨迹位置预测一直是计算机和交通领域研究人员的研究热点,随着智能手机以及 5G 技术的成熟,可为各类基于位置的服务和应用提供重要支持。RFID 数据相比 GPS 数据具有更高的精确性和更大信息量,拥有较高的研究价值。针对传统预测算法处理长序列问题存在的不足,本文提出了 LSTM-Attention 算法,在考虑不同位置权重对预测精度影响的同时,解决了长序列输入易造成的维数灾难问题。在真实的 RFID 数据集上的实验表明,该算法的预测精度高于传统的预测算法,具有一定的实用价值。本文对原始数据 RFID 的信息挖掘较浅,接下来的研究工作将重点关注特殊语义环境下的移动对象位置预测。

参考文献

- [1] BAO J, ZHENG Y, MOKBEL M F. Location-based and preference-aware recommendation using sparse geo-social networking data[C]// Proceedings of the 20th International Conference on Advances in Geographic Information Systems. New York: ACM, 2012: 199-208.
- [2] PARK H, LEE Y J, CHAE J, et al. Online Approach for Spatio-Temporal Trajectory Data Reduction for Portable Devices[J]. Computer Science and Technology, 2013, 28(4): 597-604.
- [3] HSIEH H P, LIN S D, ZHENG Y. Inferring Air Quality for Station Location Recommendation Based on Urban Big Data

- [C]//the 21th ACM SIGKDD International Conference. ACM, 2015.
- [4] YANG H,ZHANG Y H, GUO P. Vehicle trajectory recognition based on RFID data and GPS data [C]// Proceedings of 2017 China Urban Transportation Planning Annual Conference. 2017.
- [5] GIANNOTTI F,NANNI M,PINELLI F, et al. Trajectory pattern mining [C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York; ACM, 2007;330-339.
- [6] MIKOŁA J,MORZY A. Prediction of Moving Object Location Based on Frequent Trajectories[J]. Lecture Notes in Computer Science, 2006, 4263:583-592.
- [7] KRUMM J. A Markov model for driver turn prediction[C]// Proceedings of the Society of Automotive Engineers World Congress. 2016;1-7.
- [8] KILLIJIAN M O. Next place prediction using mobility Markov chains[C]//Proceedings of the Workshop on Measurement, Privacy, and Mobility, Helsinki, New York: ACM, 2012;3.
- [9] LIU Q, WU S, WANG L, et al. Predicting the next location: a recurrent model with spatial and temporal contexts[C]// Proceedings of the 30th AAAI Conference on Artificial Intelligence. Menlo Park; AAAI, 2016;194-200.
- [10] BENGIO Y, SIMARD P Y, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2):157-166.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [12] SU L M, LI L. Research on trajectory prediction method based on machine learning[C]// IOP Conference Series Materials Science and Engineering, 2019.
- [13] LI Q, ZHENG Y, XIE X, et al. Mining user similarity based on location history[C]// ACM Sigspatial International Symposium on Advances in Geographic Information Systems. ACM, 2008.
- [14] BIRANT D, KUT A. ST-DBSCAN: An algorithm for clustering spatial-temporal data [J]. Data & Knowledge Engineering, 2007, 60(1):208-221.
- [15] PALMA, ANDREY T, BOGORN Y, et al. A clustering-based approach for discovering interesting places in trajectories[C]// Acm Symposium on Applied Computing. DBLP, 2008.
- [16] KOREN Y, BELL R M, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. IEEE Computer, 2009, 42(8):30-37.
- [17] XIONG L, CHEN X, HUANG T K, et al. Temporal collaborative filtering with Bayesian probabilistic tensor factorization [C]//Proceedings of the SIAM International Conference on Data Mining. Philadelphia; SIAM, 2010;211-222.
- [18] YING J J C, LEE W C. Semantic trajectory mining for location prediction[C]//Proceedings of the 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems. New York; ACM, 2011;34-43.
- [19] MONREALE A, PINELLI F, TRASARTI R, et al. WhereNext: a location predictor on trajectory pattern mining[C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York; ACM, 2009; 637-646.
- [20] MORZY M. Mining frequent trajectories of moving objects for location prediction [C]// Proceedings of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition. Berlin, Heidelberg; Springer, 2007;667-680.
- [21] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing personalized Markov chains for next-basket recommendation[C]// Proceedings of the International Conference on World Wide Web. New York ; ACM, 2010;811-820.
- [22] MATHEW W, RAPOSO R, MARTINS B. Predicting future locations with hidden Markov models [C]// Proceedings of the 2012 ACM conference on ubiquitous computing. ACM, 2012; 911-918.
- [23] GAO Y, JIANG G H, QIN X L, et al. Position prediction algorithm of moving objects based on LSTM[J]. Journal of Computer Science and Exploration, 2019, 13(1):23-34.
- [24] PALANGI H, DENG L, SHEN Y L, et al. Deep sentence embedding using long short-term memory networks; analysis and application to information retrieval [J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2016, 24(4): 694-707.
- [25] LUONG M T , PHAM H , MANNING C D. Effective Approaches to Attention-based Neural Machine Translation[J]. arXiv: 1508.04025v5, 2015.
- [26] PAULUS R , XIONG C , SOCHER R . A Deep Reinforced Model for Abstractive Summarization[J]. arXiv: 1705.04304v3, 2017.
- [27] CHENG J P, LI D, LAPATA M. Long Short-Term Memory Networks for Machine Reading[J]. arXiv:1601.06733v7, 2016.
- [28] CHEN T Q, GUESTRIN C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining. New York; ACM, 2016;785-794.
- [29] FAN W, KUN F, YANG W, et al. A Spatial-Temporal-Semantic Neural Network Algorithm for Location Prediction on Moving Objects[J]. Algorithms, 2017, 10(2):37.



LIU Jia-chen, born in 1996, postgraduate. His main research interests include location prediction of moving object and data mining, etc.



QIN Xiao-lin, born in 1953, professor, Ph.D, is a senior member of China Computer Federation. His main research interests include spatial and spatio-temporal databases, data management and security in distributed environment, etc.