

# 训练前的上下文和时间感知的位置嵌入从时空轨迹的用户下一个位置预测

林燕,<sup>1, 2</sup> Wan怀玉<sup>1, 2, 3</sup> 郭胜男<sup>1, 2, 4</sup> 林友芳<sup>1, 2, 3, 4</sup> \*

<sup>1</sup> 北京交通大学计算机与信息技术学院, 中国北京

<sup>2</sup> 北京晶体管数据分析与挖掘关键实验室, 北京, 中国

<sup>3</sup> 民航智能客运服务重点实验室, 北京

<sup>4</sup> 综合交通运输大数据应用技术重点实验室, 中国北京. fiedu.cn

## 摘要

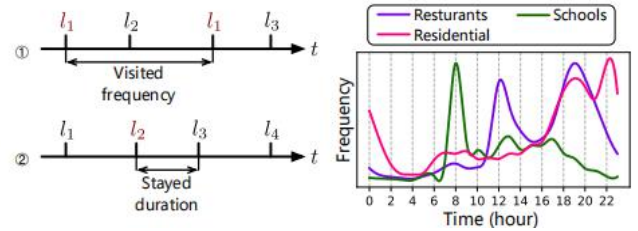
来自时空轨迹的训练前位置嵌入是一个基本的过程, 对用户的下一个位置预测非常有利。在现实世界中, 一个位置在不同的环境环境下通常具有不同的功能。如果在其嵌入过程中可以保留轨迹中的一个位置的精确函数, 则应提高用户下一个位置预测的精度。然而, 现有的对轨迹进行预先训练的位置嵌入大多基于分布式词表示, 它将一个位置的各种功能混合到一个潜在的表示向量中。为了解决这个问题, 我们提出了一个上下文和时间感知的位置嵌入 (CTLE) 模型, 该模型计算一个位置的表示向量, 并考虑其轨迹中的特定上下文邻居。这样, 就可以正确地解决位置的多功能特性。此外, 为了将轨迹中的时间信息整合到位置嵌入中, 我们提出了一个微妙的时间编码模块和一个新的预训练目标, 从而进一步提高了位置嵌入的质量。我们在两个真实世界的移动用户轨迹数据集上评估了我们提出的模型。实验结果表明, 与现有的嵌入方法相比, 我们的CTLE模型可以预先训练更高质量的位置嵌入, 并显著提高了下游用户位置预测模型的性能。

## 介绍

近年来, 基于位置的服务 (LBS) 数据的日益普及, 如GPS轨迹、兴趣点签到 (POIs) 和蜂窝信号记录, 导致了大量专注于挖掘时空数据的研究。其中, 用户位置预测受到了广泛的关注 (Liu等。2016; 肖等人。2017; Kong和Wu 2018), 并可用于多个应用程序, 包括建模用户的移动行为 (Fu和Lee 2020), 或为用户推荐位置 (poi) (Zhao等。2019)。在这些研究中, 学习位置嵌入向量是一个非常基本和关键的问题, 因为它具有准确的用户位置预测能力

\* 通讯作者: 林友芳

版权所有, 艺术智能促进协会 (www.aaai.org). 保留所有权利。



(a) 访问时间差异 (b) 访问时间分布

图1: 人类运动轨迹中所包含的时间信息的两个方面。(a) 说明, 在一个轨迹中, 访问同一位置之间的访问时间差可以显示其访问频率, 两个连续访问记录之间的差可以视为前一个位置的停留时间。(b) 显示了具有不同功能的位置的变量访问时间分布。

需要高质量的嵌入向量来更好地区分位置。与大多数预测模型中广泛使用的全连接嵌入层相比, 使用无监督或自监督目标的训练前位置嵌入可以包含更一般和全面的位置信息。通过这种方式, 它们可以在广泛的下游模型中共享, 以提高计算效率, 同时也提高整体预测性能。

从序列轨迹中捕获位置特征的想法类似于自然语言处理 (NLP) 中的语言建模。事实上, 现有的预训练位置嵌入方法都是基于分布式词表示, 在自然语言处理系统中得到了广泛的应用。例如, DeepMove (Zhou和Huang 2018) 实现了 Word2Vec (Mikolov等人。2013年) 对轨迹数据进行提取位置函数信息。然而, 除了与句子有许多相似的特征外, 轨迹还包含了独特的不可忽视的时间信息。如图1(a)所示, 在一个轨迹中, 不同位置之间的相对访问时间差异可以恢复它们的属性, 如访问频率和停留时间。另一方面, 时间会影响人们对位置的偏好, 因此一个位置的访问时间可以作为其功能的指示, 如图1(b)所示。Geo-



图2：一个位置可能在不同的上下文邻居中承担不同的功能。在轨迹-中，用户离开家后访问位置1吃早餐，然后去上班；在轨迹-，中，用户在两餐之间呆在同一位置1进行商务会议。

预告片(Zhao等。2017年)，TALE (Wan等人。和HIER (清水、Yabe和子内2020) 在建模序列相关性的基础上，从位置的绝对访问时间中提取时间信息。然而，他们并没有明确地考虑到相对访问时间的时间差异。

另一方面，多功能的位置在现实世界中是很常见的。例如，购物中心可能包含餐馆或电影院，而冰建筑可能包含娱乐场所或健身房。人们在不同的上下文邻居中为了不同的目的访问相同的位置，如图2所示。将一个位置的特定上下文合并到其嵌入向量中可能有利于位置预测，因为所提供的信息将更准确。然而，目前基于分布式表示的位置嵌入方法为每个位置分配了一个单一的潜在向量。这种方法可以被视为各种上下文环境的简单混合，并且不能处理不同上下文下的可变位置功能。

为了解决现有的位置嵌入模型中存在的上述问题，我们希望建立一个预训练模型，该模型能够基于位置的特定上下文邻居自适应地生成位置的嵌入向量。在本文中，我们提出了一种新的上下文和时间感知的位置嵌入（CTLE）模型。目标位置不指定一个嵌入向量作为每个位置的初始表示，而是一个目标位置的嵌入向量是通过其上下文位置编码向量的参数化映射函数来计算的。通过这种方式，我们能够将位置的上下文特殊功能合并到它们的嵌入向量中。我们使用双向变压器编码器实现了映射函数(Vaswani等。)，并采用在BERT中引入的蒙面语言模型（MLM）预训练目标(Devlin等人。来模拟用户轨迹中的顺序相关性。此外，为了在轨迹中加入独特的时间信息，我们设计了一个微妙的时间编码模块来建模位置之间的相对访问时间差异。我们还提出了一个新的蒙面小时（MH）预训练目标，从他们访问的时间分布中提取位置的功能。我们的模型在两个真实的移动用户轨迹数据集和三个下游模型上验证了模型的有效性。实验结果表明，我们的CTLE模型可以明显地帮助提高下游模型对密集轨迹的预测精度。

本文的主要贡献总结如下：

- o我们提出了一种上下文位置嵌入模型CTLE，它能够自适应地生成嵌入

基于其特定上下文的目标位置的向量。因此，可以区分一个位置在不同背景环境下的可变功能。

- o提出了一个微妙的时间编码模块和一个新的预训练目标来建模轨迹中的时间信息的两个方面。这将提高我们的模型对位置特征的理解。
- o我们将我们的CTLE模型应用到三个下游模型中，以进行用户的下一个位置预测，并在两个真实世界的移动用户轨迹数据集上进行了广泛的实验。实验结果表明，预测性能明显提高，证明了该模型的优越性。

## 相关工作

大多数用户位置预测模型都是基于特征的，需要位置由潜在的嵌入向量来表示。最直接和最广泛应用的策略之一是使用全连接的嵌入层。这一层为每个位置随机初始化一个潜在向量，然后用任务特定目标进行训练(Kong和Wu 2018；赵等人。2019)。然而，以这种方式生成的嵌入向量很难迁移到其他模型和任务中。嵌入层也存在过点蚀问题，无法整合全面的位置信息，这会影响预测性能。

具有无监督或自监督目标的预训练嵌入向量是自然语言处理中的一种常见做法(Mikolov等。2013；Devlin等人。和计算机视觉(Szegedy等人。2015)。近年来，这一课题也引起了时空数据挖掘的广泛关注。实际上，作为一种顺序数据，用户的移动轨迹与自然语言中的句子有许多共同的特征。因此，通过建模目标位置及其轨迹上下文的共现，我们可以提取位置的功能信息。受此思想的启发，大多数现有的位置嵌入模型都是基于自然语言处理中的单词嵌入模型。例如，DeepMove (Zhou和Huang 2018) 实现了跳跃-gram(Mikolov等人。2013年)来模拟人类在不同位置之间的运动，和(Yao等人。通过N-gram模型学习潜在表示 (Pauls和Klein 2011)。这些训练前的模型只需要未标记的轨迹数据，而这些数据通常是丰富的。通过将结果嵌入向量合并到下游任务中来验证其质量，如用户位置预测(Zhao等。和土地利用分类(Zhang等。2020)。

然而，用户的发展轨迹也包括一些独特的属性，而时间相关性就是其中之一。从时间信息中可以更准确地提取出位置的功能，因此应该通过嵌入模型来考虑。地理预告片(赵等。通过在跳过图中扩展输出嵌入，来区分在工作日和周末访问的位置。TALE (Wan等。通过设计CBOW中一种新的树结构来计算访问的时间信息(Mikolov等。2013)。希尔

(清水、Yabe和子内2020)通过向基于RNN的N-gram模型提供额外的嵌入向量,结合访问时间和停留持续时间信息。实验结果表明,时间信息确实有助于预训练更高质量的位置嵌入向量。

一个位置通常是多功能的,人们可以在不同的上下文环境下出于不同的目的访问同一位置。相应地,也可以从它的上下文*i*来指示一个位置的特殊功能。e.,附近的访问地点的轨迹。然而,现有的位置嵌入方法只为每个位置分配一个潜在向量,这意味着它们无法区分一个位置的可变功能。受自然语言处理中上下文嵌入的最新进展的启发(Devlin等。2019;杨等人。2019年),我们提出了一个上下文感知的位置嵌入模型,该模型基于其上下文邻居动态计算位置的潜在嵌入。通过这种方式,一个位置的特殊功能就能够被很好地建模。

### 前期准备工作

**决定。时空轨迹。**在基于位置的服务中,用户在某一时期内的移动可以用一个由连续的访问记录组成的轨迹来表示。访问记录(*u*, *l*, *t*)表示用户*u*在时间*t*访问了位置*l*。我们将用户轨迹集表示为*S*,数据集中所有位置的集合表示为*L*,所有用户的集合表示为*U*。

**问题声明。训练前的上下文嵌入向量。**给定一组用户时空轨迹*S*,我们的目标是预先训练一个参数化映射函数*f*,为给定上下文*C*(*l*)的目标位置*l*生成一个上下文嵌入向量*z*(*l*)。该功能应该以自我监督的方式进行培训,不需要任务的特殊目标。

### CTLE模型

图3说明了我们的CTLE模型的架构。它主要由三个部分组成:1)编码层,它将时间编码向量与位置编码向量融合,从相对访问的时间差异中提取信息;2)双向变压器编码器,给定其特定的上下文邻居,计算目标位置的嵌入;3)预训练目标,通过建模目标位置与上下文的共现概率,以及对目标的绝对访问时间,将位置特征信息纳入嵌入中。本节将详细解释我们的CTLE模型的构造。

#### 训练前的双向变压器

##### 上下文感知位置嵌入

多功能的位置在现实世界中是非常常见的。通常,用户访问一个位置,但是如果该位置是多功能的,单个访问记录(*u*, *l*, *t*)不能直接指示该位置的功能-

在这种特殊情况下的合规性。上下文环境,我.e.,访问轨迹中目标位置附近的记录,可以更明确地指示目标位置的某种功能,如图2所示。

基于上述观察结果,我们提出了一种上下文软件位置嵌入模型,该模型在考虑其上下文邻居的情况下生成一个位置的潜在表示。具体地说,给定一个目标位置*l*及其上下文*C*(*l*),我们通过一个参数化的映射函数*f*来计算它的嵌入向量*z*(*l*),表示为以下抽象形式:

$$z(l) = f(l, C(l)). \quad (1)$$

这样,位置的嵌入向量就依赖于它们的上下文环境,这使得区分一个位置的特定功能成为可能。为了实现函数*f*,我们使用了一个双向变压器编码器(Vaswani等人。2017年),原因有二。首先,轨迹中的位置与左右两侧的上下文相关。双向变压器能够从两边捕获上下文信息,同时也能考虑它们之间的相互作用。第二,变压器架构已经被许多研究证明(Dai等。比传统的序列模型,如LSTM(霍克雷特和施米德胡伯,1997年)具有更高的表达能力。

现在我们介绍了上下文嵌入-的计算详细的线程。给定一个轨迹*s* = {(*u*, *l*<sub>1</sub>, *t*<sub>1</sub>), (*u*, *l*<sub>2</sub>, *t*<sub>2</sub>), ..., (*u*, *l*<sub>*n*</sub>, *t*<sub>*n*</sub>)},我们获取一个输入的潜在向量*z*<sup>(*l*)</sup>对于位置编码模块中的每个位置*l*,记为:

$$z^{(l)} = \quad (1), \quad (2)$$

从而形成了一个输入序列{*z*<sup>(*l*<sub>1</sub>)</sup>, *z*<sup>(*l*<sub>2</sub>)</sup>, ..., *z*<sup>(*l*<sub>*n*</sub>)</sup>}. 表示位置编码模块,它是使用一个全连接的嵌入层来实现的。然后,输入序列被输入到变压器编码器层的堆栈中,其中每一层都是由一个多头自编码器层组成注意模块和一个位置级全连接的前馈网络。正式来说,我们有:

$$\begin{aligned} & \{h_1^{(k)}, h_2^{(k)}, \dots, h_n^{(k)}\} \\ & = \text{TransEncLayer}(\{h_1^{(k-1)}, h_2^{(k-1)}, \dots, h_n^{(k-1)}\}), \quad (3) \\ & \{h_1^{(0)}, h_2^{(0)}, \dots, h_n^{(0)}\} = \{z'(l_1), z'(l_2), \dots, z'(l_n)\}, \end{aligned}$$

其中,变压器编码器层表示一个变压器编码器层。

{*h*<sub>1</sub><sup>(*k*)</sup>, *h*<sub>2</sub><sup>(*k*)</sup>, ..., *h*<sub>*n*</sub><sup>(*k*)</sup>}是第*k*层的输出内存序列,以及第*k*+1)层的输入序列。假设该模型总共包含*N*个变压器编码器层。我们考虑最后一层的输出内存序列的第*i*项, *h*<sub>*i*</sub><sup>(*N*)</sup>,作为位置*l*的上下文嵌入向量*i*在轨迹。如果我们表示*s*中除*l*外的所有位置*i*作为其上下文*C*<sub>*i*</sub>,计算方程为*z*(*l*<sub>*i*</sub>),可以写成:

$$\begin{aligned} z(l_i) &= h_i^{(N)} = \text{TransEnc}(\{z'(l_1), z'(l_2), \dots, z'(l_n)\})_i \\ &= \text{TransEnc}(\Omega(\{l_1, l_2, \dots, l_n\}))_i \\ &= f(\{l_1, l_2, \dots, l_n\})_i \\ &= f(l_i, C(l_i))_i, \end{aligned} \quad (4)$$



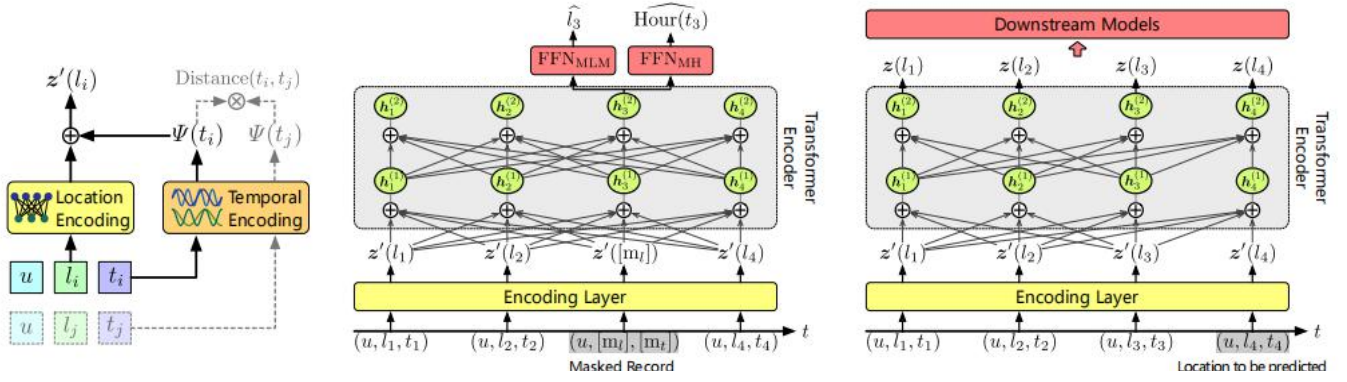


图3: CTLE的模型架构。(a)说明了CTLE的编码层。(b)显示了CTLE的预训练过程,其中两个全连接的前馈网络旨在分别预测一个掩码记录的原始位置和访问的小时索引。(c)表示利用CTLE来计算用户下一个位置预测的上下文嵌入。以全轨迹作为输入,并采用特殊注意掩模,防止信息泄漏。请注意,(b)和(c)都只是演示了一个具有两层的变压器编码器。实际上,层数是可调节的。

对应于式(1)中的抽象形式。TransEnc表示变压器编码器层的堆栈。

映射函数 $f$ 应该学习理解目标位置与其对应的上下文之间的相关性,以便将这些位置的特征信息合并到它们的嵌入向量中。受NLP模型中语言理解进展的启发,我们实现了BERT提出的蒙面语言模型(MLM)(Devlin等。构建一个自我监督的培训目标。给定一个轨迹 $s$ ,我们随机选择20%的记录作为掩码记录,并将它们对应的位置替换为特殊的标记和访问时间 $[m_l]$ 和 $[m_t]$ 。结果轨迹被视为函数 $f$ 的输入。 $\tilde{s}$ 对于一个蒙面的位置 $l_m$ ,我们的目标是预测它的原始标记使用 $f$ 的输出:

$$\hat{l}_m = \text{FFN}_{\text{MLM}}(h_m^{(N)}) = \text{FFN}_{\text{MLM}}(f(\tilde{s})_m), \quad (5)$$

其中 $\text{FFN}_{\text{MLM}}$ 是一个全连接的前馈网络,用于将变压器编码器的输出存储器向量转换到位置令牌的预测中。然后,可以通过最大化预测精度来构建MLM目标。如果我们将被掩蔽的位置视为目标,而将未被掩蔽的位置视为它们的上下文,那么MLM的训练前目标可以描述为:

$$\begin{aligned} O_{\text{MLM}} &= \arg \max_{\theta} \sum_{l_m \in \Gamma} p(l_m | \text{FFN}_{\text{MLM}}(f(\tilde{s})_m)) \\ &= \arg \max_{\theta} \sum_{l_m \in \Gamma} p(l_m | \text{FFN}_{\text{MLM}}(f(C(l_m))_m)), \end{aligned} \quad (6)$$

其中,  $\Gamma$ 为训练集中所有掩码记录的集合,为嵌入模型中所有可学习参数的集合。通过最大化目标及其上下文的共现概率,具有相似上下文环境的位置将具有更紧密的嵌入向量,从而纳入功能信息。这个想法类似于Word2Vec(Mikolov等人。2013)。然而,而不是直接优化一组静态嵌入-

我们的模型优化了一个映射函数的参数。通过这种方式,结果嵌入向量将是与上下文相关的,并且能够在不同的上下文邻居中合并位置的特定特征信息。

### 合并时间信息

时间信息是时空数据挖掘的关键。对于轨迹数据,时间信息主要由位置的访问时间携带,可以从相对访问时间差和绝对访问时间两个方面来看。位置的访问频率或停留时间可以从相对访问的时间差异中显示出来,这意味着功能信息,如图1(a).所示另一方面,人类的活动通常是受时间限制的,而一个位置所携带的功能可以决定用户到达它的时间,如图1(b).所示相应地,我们可以从位置的绝对访问时间中提取特征信息。

基于上述分析,我们通过一个微妙的时间编码模块和一个新的训练前目标,将时间信息的两个方面合并到我们的模型中。

编码层中的时间编码为了明确地模拟轨迹中不同位置之间的相对访问时间差异,我们提出了时间编码模块。在最初的变压器中(Vaswani等人。2017),引入位置编码的模型来区分输入序列中标记的顺序。其计算方程可写为:

$$\begin{aligned} \Phi(o) &= [\cos(\omega_1 o), \sin(\omega_1 o), \dots, \cos(\omega_d o), \sin(\omega_d o)], \\ \omega_k &= 1/10000^{2k/d}, \end{aligned} \quad (7)$$

其中,函数表示位置编码模块, $o$ 为一个位置的索引, $2d$ 为位置编码向量的维数。

位置编码只能表示简单的顺序顺序。然而,对于运动轨迹,访问记录并不均匀

在时间轴上的范围，以及位置之间的相对访问时间差异包含了重要的信息，如位置的访问频率或停留时间，如图1(a). 所示灵感来自于(Xu等人。2020年)，我们提出的时间编码通过对位置编码进行两个简单的调整来解决这个问题：用绝对访问时间戳 $t$ 替换位置 $o$ 索引，设置参数 $\{\omega_1, \omega_2, \dots, \omega_d\}$ 可训练。正式来说，我们有：

$(t) = [\cos(\omega_1 t), \sin(\omega_1 t), \dots, \cos(\omega_d t), \sin(\omega_d t)]$ , (8), 其中是我们的时间编码模块。 $(t)$ 和 $(t + 6)$ 的点积的计算方法为：

$$(t) \cdot (t + 6) = \cos(\omega_1 6) + \cos(\omega_2 6) + \dots + \cos(\omega_d 6), \quad (9)$$

这就是说 $(t)$ 和 $(t + 6)$ 之间的距离（用点乘积测量）是可学习的，并且独立于 $t$ 无关。在变压器编码器中并不缺少点积操作，所以这个距离信息注定要被模型捕获。

然后，我们将时间编码与位置编码结合，将时间编码引入模型中。具体地说，我们修改方程2来获取记录 $(u, l, t)$ 的输入潜在向量如下：

$$z^{\setminus}(1) = (1) + (t). \quad (10)$$

这样，我们的模型就能够将嵌入在相对访问时间差异中的时间信息整合到训练前和嵌入生成过程中，从而得到更高质量的嵌入向量。CTLE的编码层与时间编码相结合的总体架构如图3(a). 所示

为了进一步从绝对访问时间中提取信息，我们提出了蒙面小时（MH）训练前目标。为一个蒙面访问的时间 $t_m$ 在一个蒙面轨迹中，我们的目标是预测其原始访问小时指数： $\tilde{s}$

$$\text{Hour} \rightarrow \text{ur}(t_m) = \text{FFN}_{\text{马绍尔群岛}}(f(\tilde{s}_m)), \quad (11)$$

其中 $\text{hour}_t(t)$ 表示时间 $t$ 的小时指数。芬尼马绍尔群岛是一个全连接的网络，用于将变压器编码器的输出内存向量转换为小时索引的预测。在训练前的过程中，MH与MLM结合，形成一个多任务的学习目标。MH目标和总体培训前目标可描述为：

$$O_{\text{MH}} = \arg \max_{\theta} \sum_{t_m \in \Gamma} p(\text{Hour}(t_m) | \text{FFN}_{\text{MH}}(f(\tilde{s}_m))); \quad (12)$$

$$O = O_{\text{MLM}} + O_{\text{MH}}, \quad (13)$$

$\theta$ 式中为嵌入模型中所有可学习参数的集合。我们可以给出上述训练前目标的一个抽象形式为 $\arg \max_{\theta} p(l_m, t_m | C(l_m))$ . 通过最大化一个目标位置在特定的上下文范围内被访问的概率，我们将绝对访问时间中隐含的信息纳入到模型中。这使得模型能够从时间信息中提取位置的功能，从而提高生成的嵌入向量的质量。CTLE在训练前阶段的整体模型架构如图3(b). 所示

## 适用范围

为了使CTLE的上下文嵌入思想有效，轨迹需要相对密集，以便目标位置与其上下文环境有足够的相关性。换句话说，上下文需要足够完整，以揭示目标的功能。这就是大多数签到记录等稀疏轨迹所缺乏的。此外，签到数据中的位置经常被手动划分为具有特定目的的poi，这意味着它们的多功能特性并不明显。总之，当应用于非常稀疏的轨迹时，我们的CTLE将失去它比现有的基于分布式嵌入的方法的优势。

## 实验

为了评估我们的模型生成的上下文嵌入向量的质量，我们将这些向量纳入三个下游用户位置预测模型，并将结果与其他位置嵌入方法进行比较。

## 数据集

实验在包含中国北京和沈阳移动信号数据的两个真实数据集上进行，分别表示为Mobile-PEK和MobileSHE。它们记录了匿名移动用户在电信基站之间的切换事件。每个基站负责向其周边区域提供信号，用户将记录进入该区域的情况。在本文中，我们将基站视为位置。与一些公共可用的签到数据集不同，移动信号数据集包含密集的轨迹，使它们更适合评估有效的-

CTLE的内径。

从技术上讲，信号数据可以代表用户的完整移动轨迹。然而，并不是所有的轨迹点都是用户访问的位置。例如，人们从一个地方旅行到另一个地方也会产生一系列的轨迹点，但这些点缺乏语义信息。为了过滤出用户刚刚经过的点，我们将删除基站和相关记录的平均持续时间低于以下的用户

30分钟，因为他们产生的大多数信令记录只是传递点。然后我们去除所有持续时间小于5分钟的轨迹点。在此处理之后，所有剩余的信令记录都可以被视为停留点。用户按上升时间戳排序的一天内的停留点序列被视为一个轨迹。最后，我们在我们的数据集中包含了超过5个停留点的轨迹和超过4个轨迹的用户。数据集的统计数据如表1所示。

数据集	#Users	#Locations	#Records	时间跨度
移动-PEK移	12,691	7,279	1,383,422	5天
动-SHE	10,564	7,201	607,581	11天

表1：数据集的统计数据。

## 基线位置嵌入方法

为了证明我们的模型的优越性，我们包括了两个经典的分布式嵌入模型，它们是由许多位置嵌入方法实现的，以及一些最先进的位置嵌入模型进行比较。

- o FC层：一种广泛使用的潜在表示策略，它为每个位置分配一个随机初始化的嵌入向量。然后将这些向量与整个模型的其他参数一起进行训练。
- o 跳过-格兰姆(Mikolov等人。2013)：Word2Vec模型的变体，已被(Liu, Liu和Li 2016)用于建立移动轨迹建模。
- o POI2Vec (Feng等。2017年)：一种基于Word2Vec的嵌入方法，它通过将位置分布到地理二叉树中来建模空间相关性。

**地理(赵等。2017年)**：地理时间序列嵌入秩通过扩展输出向量，修改跳跃图的负采样策略，融合时空信息。

- o TALE (Wan等。2019年)：时间感知位置嵌入通过设计一种新的时间树结构进行分层软件计算，整合时间信息。
- o HIER (清水、Yabe和子内2020)：分层细粒度位置嵌入采用了基于LSTM (霍克雷特和施米杜伯1997年)的n克架构 (保罗斯和克莱因2011)。

## 用户的下一个位置预测模型

下游预测模型可以用来评价嵌入方法的有效性，因为预测性能依赖于位置嵌入向量的质量。在本文中，我们采用了三种流行的用户下一个位置预测方法作为下游模型：

- o ST-RNN (Liu等。2016年)：一个用于用户下一个位置预测的时空递归神经网络模型。它通过循环传播中的时间特定和距离特定转移矩阵结合了空间和时间相关性。
- o ERPP (Xiao等人。事件循环点过程将时间事件序列视为点过程。在本文中，我们将访问一个地点作为一个事件，并将到达时间作为其时间戳。
- o ST-LSTM (Kong和Wu 2018)：一个用于用户下一个位置预测的时空长-短期记忆模型。它通过将时间和距离值嵌入到潜在向量中来扩展LSTM的输入。

给出一个序列轨迹  $s = \{(u; l_1; t_1), (u; l_2; t_2), \dots, (u; l_n; t_n)$ ，我们向预测模型提供一系列相应的嵌入向量  $\{z(l_1); z(l_2); \dots; z(l_i)\}$ 。值得注意的是，当将CTLE应用于位置预测模型时，我们使用了一个特殊的注意掩码来防止被预测的位置信息泄漏到历史序列中，如图3(c)。所示除FC层外，所有嵌入方法在位置预测过程中都不参与反向传播。

## 镶嵌

对于Mobile-PEK数据集，我们选择前3天的轨迹进行训练，第4天进行评估，最后一天进行测试。对于Mobile-SHE数据集，我们选择前7天的轨迹进行训练，最后2天进行测试，其他几天进行评估。在位置预测任务中，我们选择每个轨迹中的最后三个点作为预测目标，其余的点作为源序列。

对预训练后的位置嵌入模型和下游预测模型进行训练，并在评价集上进行验证，得到最优超参数。请注意，训练前的嵌入模型本身并没有一个可靠的性能评估度量标准，所以我们在ST-LSTM模型的帮助下调整了超参数。位置预测模型采用交叉熵损失进行训练，并采用精度、宏观召回率和宏观f1评分进行评估。

我们在PyTorch中实现了所有的基线模型和我们的CTLE模型 (Paszke等人。2019)。所有嵌入模型的结果嵌入向量的维数均设为128。我们实现了基于隐藏大小为512的2层LSTM网络的ERPP和ST-LSTM，以及基于隐藏大小为512的单层RNN的ST-RNN。对于我们的CTLE模型，我们堆叠了4层变压器编码器，有8个注意力头，前馈层的隐藏大小设置为512。CTLE在训练集上进行200个时的预训练，所有的预测模型都采用早期停止机制进行训练，以获得在评价集上表现最好的时。我们选择了Adam优化器，并且有一个全面的初始学习率为0.0001。

## 实验结果

表2显示了不同模型对用户下一个位置预测的性能比较。将所有的基线嵌入方法与多个下游预测模型进行融合，计算结果。除了ST-RNN预测模型、Mobile-SHE数据集外，我们的CTLE始终优于其他位置嵌入方法。

随机初始化全连接的嵌入层，只在特定的预测目标上，并且不能合并一般的信息，如位置的功能。Skip-gram只利用了目标位置及其上下文的共现概率，而忽略了时空轨迹的一些独特特性，如被访问的时间和位置的地理位置。从表2中可以看出，这些方法通常比较新的、更全面的位置嵌入方法的预测性能更差。

Geo-Teaser和POI2Vec通过不同的方法将空间信息引入到嵌入过程中。然而，一个城市中不同地点之间的地理相关性并不是很明显。一方面，人们的访问偏好并不局限在城市的小区域内；另一方面，地理位置的属性与其地理位置没有很强的相关性，因为特写位置也可能有不同的功能。

地理预告还通过将绝对访问时间考虑到位置，将时间信息整合到位置嵌入中。然而，它只歧视访问

预测模型		ST-RNN			埃尔普			ST-LSTM		
指标		精度 (%)	宏召回 (%)	大的 F1 (%)	精度 (%)	宏召回 (%)	大的 F1 (%)	精度 (%)	宏召回 (%)	大的 F1 (%)
数据集	嵌入									
移动 PEK	FC层*跳转图	3. $\pm$ 7440.10	.739 $\pm$ 10.06	.449 $\pm$ 10.19	4.3730.14	2. $\pm$ 0170.04	.595 $\pm$ 10.05	4.5420.15	2. $\pm$ 0920.09	.689 $\pm$ 10.07
	POI2Vec	3.6710.11	1.7770.11	1.4230.05	5.0240.08	2.3680.07	1.7790.04	5.1630.10	2.5860.06	1.9470.04
	地理定位器	3.9920.08	2.2810.08	1.8380.06	5.1590.05	2.5950.07	2.0350.06	5.3050.05	2.6820.08	2.0770.10
	故事	3.9980.13	2.1660.07	1.7960.07	5.4570.03	2.6710.08	2.0390.03	5.5110.05	2.7390.04	2.0840.02
	希尔	4.1990.05	2.2400.07	1.8150.06	5.5555.55	3.237 $\pm$ 5.55	2.5555.55	5.5555.55	3.152 $\pm$ 5.55	2.5555.55
	CTLE (我们的)	4.4444.3	2.5555.4	1.5555.8	6.07 $\pm$ 0.09	$\pm$ 0.07	$\pm$ 0.03	5.89 $\pm$ 0.15	$\pm$ 0.10	$\pm$ 0.13
		39 $\pm$ 0.04	40 $\pm$ 0.07	62 $\pm$ 0.08	6.4810.05	2.8700.08	2.1760.04	6.4730.09	2.8390.11	2.1650.02
		5.0680.05	2.8900.11	2.3120.02	$\pm$	4.0020.04	3.0660.06	$\pm$	4.0720.13	3.0970.13
Mobile SHE	FC层*跳转图	3. $\pm$ 6740.07	2. $\pm$ 4080.07	.946 $\pm$ 10.05	4.3430.18	2. $\pm$ 4540.10	2. $\pm$ 0370.09	4.4160.20	2. $\pm$ 4500.14	2. $\pm$ 0050.11
	POI2Vec	3.6460.05	2.2780.08	1.8090.05	4.4050.06	2.4590.06	1.9740.05	4.5080.05	2.5070.07	1.9980.07
	地理定位器	3.9360.04	2.6050.04	2.0840.03	4.9230.06	2.9920.02	2.4080.02	4.9300.07	2.8900.10	2.3050.08
	故事	4.0060.05	2.4550.03	1.8970.02	4.9320.12	2.8950.02	2.4100.07	5.1300.15	2.7540.07	2.2450.06
	希尔	4.4444.689	3.4440.09	2.7610.08	5.5555.55	3.5555.446	2.5555.883	5.5555.55	3.5555.399	2.5555.787
	CTLE (我们的)	$\pm$ 0.10	3.1170.15	2.5210.09	6.24 $\pm$ 0.16	$\pm$ 0.06	$\pm$ 0.04	6.72 $\pm$ 0.09	$\pm$ 0.11	$\pm$ 0.11
		4.5390.22	3.3920.11	2.7200.07	6.3110.04	3.2730.17	2.7080.18	6.3250.08	3.2520.07	2.6800.05
		5.1240.20	5.5555.55	5.5555.55	$\pm$	3.9840.05	3.3400.07	$\pm$	3.9500.11	3.2910.06

\*FC层是下游预测模型中最初使用的位置嵌入方法，因此对应行的结果是这些模型的原始预测性能。

表2：不同方法的用户下一步位置预测性能比较。

记录发生在工作日和周末，忽略了线粒度的时间信息。TALE和HIER都利用了访问时间指数，这可能意味着用户对地点的偏好。因此，它们可以生成对预测任务更有利的位置嵌入。然而，上述位置嵌入方法都没有考虑相对访问时间差异，这能够指示位置之间的关系。更重要的是，它们只分配一个潜在向量来表示每个位置，这使得它们无法在结果嵌入向量中恢复位置的可变功能属性。这可能会严重损害下游模型的性能，因为移动信令数据中位置的多功能特性尤为明显。

我们的CTLE模型根据特定上下文动态生成位置的嵌入，从而可以区分特定上下文环境中一个位置的确切功能。此外，CTLE还考虑了位置之间的相对访问时间差异和不粒度绝对访问时间。如表2所示，这些设计产生了更高质量的位置嵌入，并有助于下游的位置预测模型获得更好的性能。

## 组件分析

为了进一步研究CTLE各组成部分的有效性，我们设计了CTLE模型的三种变体，如

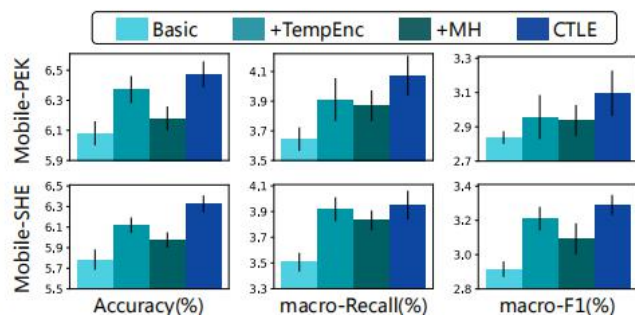


图4：CTLE的成分分析。

下列

- 基本原理：此模型使用了原始的位置编码来自变压器 (Vaswani等。)，并且仅利用MLM目标进行预训练。
- +TempEnc：此模型取代了位置编码 在基本模型中与我们提出的时间编码。
- +MH：这个模型仍然使用位置编码，但

在训练前，将蒙面小时目标与MLM目标结合起来。

我们将这三种变体与ST-LSTM下游模型上的CTLE模型进行了比较。图4显示了结果。与基于分布式词嵌入的位置嵌入方法相比，该基本模型已经具有一定的性能优势。这表明，学习上下文感知的位置嵌入向量可以有利于完成位置预测任务。

时间编码和屏蔽小时目标将嵌入在轨迹中的时间信息整合到模型中。如图4所示，它们都可以比基本模型提高预测性能。由于这两个模块关注的是时间差异的不同方面，我们的inal模型将它们结合起来，得到了最佳的结果。fl

## 结论

本文提出了一种新的上下文和时间感知的位置嵌入方法CTLE。它根据给定目标位置的上下文动态生成目标位置的嵌入向量，因此可以合并特定上下文环境下的位置的特殊功能。一个微妙的时间编码模块和一个新的蒙面小时的训练前目标进一步将轨迹中的时间信息的多个方面合并到嵌入中。在两个真实的移动用户轨迹数据集和三个流行的下游模型上对用户下一个位置预测下游任务进行了实验，结果证明了我们的模型在学习密集轨迹上准确、全面的位置嵌入方面的有效性。



## 确认信息

中央高校基本科研业务费基金(No. 2019JBM024).

## 参考文献

戴Z.; 杨, Z.; 杨, Y.; 卡博内尔, J. G.; 勒, 问。V.; 萨拉库蒂诺夫, R. 2019. 转换器-xl: 超越固定长度上下文的注意语言模型。《计算语言学协会第57届会议》, 2978-2988。

DevlinJ.; 李; 和图塔努瓦, K. 2019. 对语言理解的深度双向转换器的预培训。《2019年计算语言学协会北美分会会议: 人类语言技术》, 4171-4186。

冯年代; 丛G.; 安B; CheeY. M. 2017. POI2Vec: 预测未来游客的地理潜在表示。《第31届AAAI人工智能会议》, 102-108。

.-傅, 泰伊。; 和李, WC. .-2020. Trembr: 探索轨迹表示学习的道路网络。《ACM智能系统与技术交易记录11(1): 1-25。

霍克雷特, s. ; 和施米德胡伯, J. 1997. 长期内存。《神经计算方法9(8): 1735-1780年。

孔, D.; 和吴, F. 2018. HST-LSTM: 一种用于位置预测的分层时空长-短期记忆网络。《第27届国际艺术情报联席会议》, 第18卷, 第2341-2347卷。

刘Q.; 吴s; 王l. 和Tan, T. 2016. 预测下一个位置: 一个具有空间和时间上下文的循环模型。《在第30届AAAI人工智能会议上》, 194-200人。

刘, X.; 刘, Y.; 和李, X. 2016. 探索个性化位置推荐的位置上下文。《第25届国际艺术情报联席会议》, 1188-1194。

Mikolov, T.; 陈, K.; 科拉多, G.; 和院长, J. 2013. 向量空间中词表示的有效估计。《在第一届国际学习代表会议上。

Paszke, A.; 格罗斯, 美国; Massa, F.; Lerer, A.; 布拉德伯里, J.; Chanang; 基林T.; 林z; 他, N.; Antiga, L.; 等人。2019. PyTorch: 一个命令风格、高性能的深度学习库。《在第32神经信息处理系统的研究进展》, 8024-8035。

保尔斯 A.; 和克莱因国际石油工业股份有限公司, D. 2011. 更快和更小的NGram语言模型。《在《计算语言学协会第49届年会上: 人类语言技术》》, 258-267。

清水, T.; YabeT.; 和他, K. 2020. 从人类移动轨迹中学习具有空间层次的细粒度位置嵌入。《arXiv预印本arXiv: 2002.02058。

有, C.; 刘, W.; 贾, Y.; Sermanet, P.; 里德, S.; 安格洛夫, D.; 厄, D.; 范胡克, 诉; 和拉比诺维奇, A. 2015. 更深入地讨论卷积。《在第25届IEEE计算机视觉和模式识别会议》, 1-9。

瓦斯瓦尼, 一个; N; Uszkoreit, J.; 琼斯, L.; 戈麦斯, A. LN.; 凯撒的话, 。; 和波洛苏欣, 我。2017. 你所需要的就是注意力。《在神经信息处理系统的第30个进展中》, 5998-6008。

万H. 李F. 郭S. 曹Z. 和林Y. 2019. 从时空轨迹中学习位置的时间感知分布表示。《在第24届高级应用程序数据库系统国际会议上》, 268-272。

肖S.; 燕J.; 杨X.; 查h. 和Chu, S. M. 2017. 利用递归神经网络建模点过程的强度函数。《1597-1603年, 在第31届AAAI人工智能会议上。

徐勇明, D.; 鲁安 C.; Krpeoglu, E.; 库马尔, S.; 和Achan, K. 2020. 在时间图上的归纳表示学习。《在第八届国际学习表征会议上。

杨z; 戴z; 杨Y.; 卡博内尔, J. G.; 萨拉库蒂诺夫, R.; 和勒, Q. V. 2019. 语言理解的广义自回归预训练。《在神经信息处理系统的第32个进展中》, 5754-5764。

姚Z.; 傅Y.; 刘B.; 胡W. 和熊, H. 2018. 通过与人类流动模式相结合的区域嵌入来代表城市功能。《第27届国际艺术情报联席会议》, 3919-3925。

张, M.; 李, T.; 李, Y.; 和回族, P. 2020. 城市区域嵌入中的多视图联合图表示学习。《第29届国际艺术情报联席会议。

赵P.; 朱H.; 刘Y.; 徐J.; 李Z.; 庄, F.; 盛, 五。S.; 周, X. 2019. 下一步该去哪里: 为下一个POI推荐提供的时空门控网络。《第33届AAAI人工智能会议》, 第33, 5877-5884卷。

赵S.; 赵, T.; 国王, 我; 和Lyu, M. R. 2017. 地理提示: 为感兴趣点推荐的地理时间序列嵌入排名。《在第26届国际米兰万维网伙伴组织全国会议》, 153-162。

周; 和黄。2018. 深度移动: 通过大规模的移动数据来学习位置表示。《第六届IEEE大数据国际会议》, 2403-2412。