**一种嵌入式GPS轨迹表示用RNN在基于点的分类**

摘要：

明白和发现知识从人类活动的GPS轨迹是一个基本的话题在基于流动性城市计算中。我们提议轨迹网----一个神经网络体系结构用于基于点的轨迹分类去预测真实世界人类的出行方式从GPS跟踪中。为了克服计算在低维和各种各样的特征空间由GPS数据强加的潜在的因素，我们开发了一个新奇的表示：将传统的特征空间嵌入到另一个空间中，它可以被理解为一个基础上扩展的表格。我们也使特征空间更丰富通过基于段的信息以及用Maxout激活去提高RNNs的预测能力。我们实现了超过98%的分类精度当发现四种交通出行的方式，不通过额外的传感器数据或者基于位置优先知识，胜过了现有的模型。

关键字：

GPS，轨迹分类，递归神经网络，嵌入

## 1 介绍

无处不在的定位获取技术的出现，例如GPS and AIS（自动识别系统），已启用海量收集时空轨迹数据。理解和从GPS和AIS数据发现知识，使我们可以绘制全球人类活动的图景和改善我们的关系与行星地球。在他们的许多应用程序，轨迹数据挖掘算法搜索图片集群，预测或分类各种移动物体，包括动物，人类，汽车，和船只[2,12,15,27,31,33,49,50]。这样的应用包括时间序列预测任务，比如预测人群的流行[47,48]和时间序列分类任务，如检测人的运输模式[51]和捕捞活动[9,23,25]。这些应用程序使我们能够改善交通管理，公共安全，和环境可持续性。在本文中，我们调查交通模式检测任务使用Recurrent Neural网络（RNN）将GPS跟踪分为四类（即，自行车，汽车，步行和巴士）。

神经网络[6,16]在这个任务上未能实现高精度模型的原因是由于低维和异构特征空间中特征构成层次的不同发展。 为了解决这些问题，我们将以前的RNNs [25]扩展到海量数据以外的地方，从GPS跟踪中检测人类交通模式。 在[25]中有一个重要的新颖性：对嵌入方法的深入分析和对基本扩展，分段函数和离散化的联系的深入分析。 另外，我们首次提出了Maxout在GRU中作为通用逼近器的积极用法。

我们提出了在现实世界GPS运输模式分类任务中实现最先进性能的TrajectoryNet方法。 所提出的TrajectoryNet与现有的方法不同之处在于，它首次使用GPS数据的嵌入，将低维和异质特征空间映射为分布向量表示，以捕获高层语义。 嵌入可以被看作是一种基本扩展的形式，其改进了特征表示，即使嵌入空间中的线性决策边界可以被映射到原始特征空间中的高度非线性函数。 我们也使用细分市场信息和Maxout激活[17]提高RNNs的预测能力。 TrajectoryNet达到在检测4种和7种类型的运输方式时，分类精度分别达到98％和97％。

本文的其余部分结构如下：在第二部分我们提供关于轨迹数据挖掘的定义。 我们还介绍了RNNs，特别是在本文中将使用的门控循环单元（GRU）。 在第3节中，我们介绍了提出的模型框架 - TrajectoryNet。 我们详细介绍了定义邻域的分割方法，嵌入方法和Maxout GRU分类模型。 我们重点介绍神经网络中的嵌入和离散化之间的关系提供关于嵌入的需要的直观证明连续的功能。 在第4节中，我们详细介绍了实验装置并讨论实验结果。 在第5节中，我们总结一下结论和未来的工作。

## 2 初步措施

2.1 定义

定义2.1。 轨迹[11]是一个运动的一部分由给定时间间隔[t Begin，t End]分隔的对象。 它是从时间到空间的连续功能。

定义2.2。 一个轨迹的离散表示[11,15]由三元组S = <（x 0，y 0，t 0），...，（x k，y k，t k）>表示轨迹的时空位置，但不是提供物体运动的连续性。 这里（x i，y i）表示时间t i处的空间坐标。

离散表示是由于轨迹数据在离散时间戳采样的位置采集技术的采样本质。

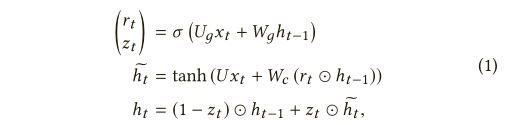
定义2.3。 轨迹的基于点的分类是学习一对一映射S→M的实践，其将离散轨迹数据S = <（x 0，y 0，t 0），...，（xk ，yk，tk）>到相应的标签序列M = <m 0，...，mk>其中mi表示三元组（xi，yi，ti）的类标签。

定义2.4。 轨迹的分割[15]是将轨迹划分为具有诸如时间间隔，轨迹形状或语义的一些准则的不相交的分段，其可以从轨迹数据提供更丰富的知识。

定义2.5。 不连续特征的离散化[1，13，14，28]使用由C =（c 1，c 2，...，cn）表示的n个切点将连续属性D∈[l，u]的区域划分为一组区间，其中c 1 <c 2 < ·<cn。 域D分为不相交的区间[1，c 1）∪[c 1，c 2）∪...∪[c n，u]其中l，u是该属性的上下界。

2.2 递归神经网络

RNN是一个从顺序数据学习的强大模型。 GPS轨迹是一种自然适合于RNN框架的时空数据。 与标准前馈神经网络不同，RNN使用循环连接来保持不同时间步骤之间的状态信息。 为了克服RNNs中的优化挑战，引入了长时间的短期记忆网络（LSTMs）[5,21,22]，并采用了一个复杂的网络结构，在不同的时间步骤选择性地传递信息。 LSTM体系结构丰富[8,19]，我们最近的工作[25]表明门控重复单元（GRU）[7]适用于基于点的轨迹分类。 等式（1）定义的门控重复单元[7]是LSTM的变体。



其中r t，z t是学习短期和长期记忆的复位和更新门，e h t和h t是t处候选和最初的单元状态，U和W是输入到隐藏和循环连接。 表示单元乘法。 与LSTM相比，这样可以得到一个简化的架构，其中包含的参数较少，易于培训。 本文通过引入Maxout激活函数来扩展GRU，以学习更多表达性的记忆状态。

# 3 方法：TrajectoryNet

图1显示了建议的TrajectoryNet 2的框架。 GPS记录被分成几段，然后提取基于点和段的特征。 然后离散化连续的特征，将它们嵌入到另一个空间中，然后在3.3节中描述的Maxout GRU中进行分类。

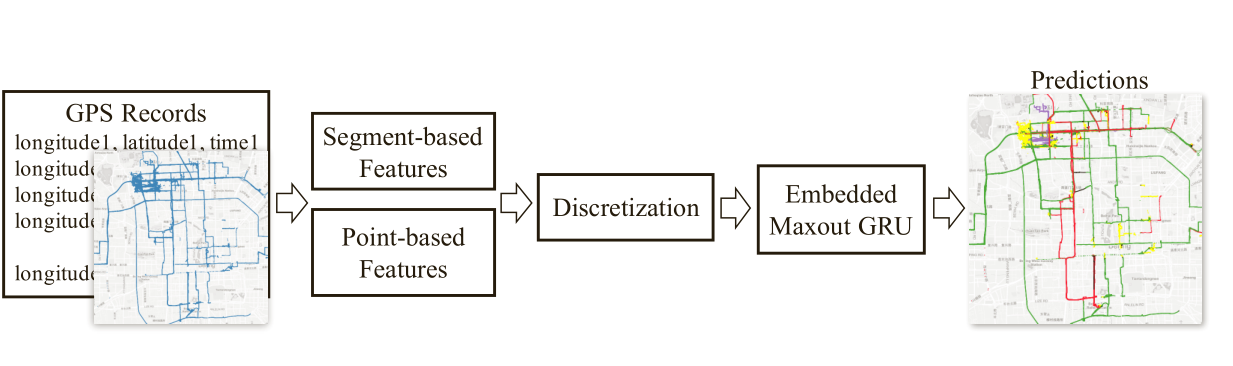


图1：拟议的轨迹网的框架

#### 3.1分割：从基于点的特征到分割特征

许多运输模式检测算法将轨迹划分为单模跳闸段，并将单一运输模式分配给每个段。 郑等人 [51]以步行为分界标准，假设“人在换乘时必须停下来走路”[51]。 然而，这个假设有一个主要的实际限制：由于采样过程中的不确定性导致步行分段的缺失可以将具有两种不同运输模式的轨迹合并成单个分段。 通过将单个标签分配给包含两个类的段，可能会损害模型性能。为了解决这个问题，我们根据每个离散的GPS样本而不是单独的分段来分配运输模式。 该方法利用基于片段的特征来丰富基于点的特征，同时防止由不精确的分割引起的误分类。

各种方法可以用于分割GPS轨迹，包括基于过渡的方法[51]，使用步行划分轨迹，基于聚类的方法[29,41]，测量子轨迹的相似性，基于时间的方法[42] 在每个段中使用相等的时间间隔，基于距离的方法[30]使用在每个段中行程相等的距离，基于轴承的方法[10]，测量轴承方向的变化和基于窗口的方法[32] 每段中GPS采样的数量。 这篇论文对实施过程中的简单性进行了实证评估。

#### 3.2嵌入：从特征空间到语义空间

3.2.1动机。 在自然语言处理中，嵌入是将名义特征（即，单词）转换成连续和分布的向量表示的过程。 与局部表示相比，分布式表示具有非局部的泛化能力[3]，更加高效，可以编码语言规则和语义。

在交通模式检测的环境中，连续的特征（如速度）受限于其在不同应用中捕捉各种语义的能力。 连续的特征可以被看作是从不同应用中带有不同语义的潜在因素中得到的观察结果。 以速度为例，10km / h对于跑步者来说是“快”，而对于“慢”来说则是“快”驾驶者 - 相同的速度可以取决于上下文而采取相反的频谱端。 这不是连续性特征的价值，更重要的是我们如何解释这个意义。 为此，希望开发可以解释连续特征的语义的表示。 其他单词，我们感兴趣的是将连续的特征转换成对应于它们意义的向量表示。

3.2.2平滑先验。 嵌入开发面临的一个挑战是，与名词性特征（如单词）不同，对于连续性特征而言，有无限的可能值。 为了解决这个NP难题[37]，我们使用平滑先验的概念来帮助确定我们嵌入连续特征。 交通模式检测的任务表现出物理属性（例如，速度和加速度）连续表现的性质，即它们在空间或时间邻域中通常不突然改变，并呈现出一些连贯性。 我们引入一个光滑度先验假设：围绕特定连续属性的值，例如， 速度，它的语义或多或少是连贯的，交通方式的变化不会突然发生。

定理3.1。 假设F是连续的随机变量X∈[l，u]的离散（分类）变量Y的连续条件累积分布函数，其中l和u是X的下限和上限。 对于每一个ε> 0，对于i = 0,1，...，n-1，存在对[l，u]的一个有限的分割l≤c1 <c 2 <... <c n≤u，F（Y | c -i + 1）-F（Y | c i）≤ε。

证明: 令ε> 0，c 0 = l并且i≥0定义

c j+1 = sup{z : F(Y|z) ≤ F(Y|cj ) +ϵ}

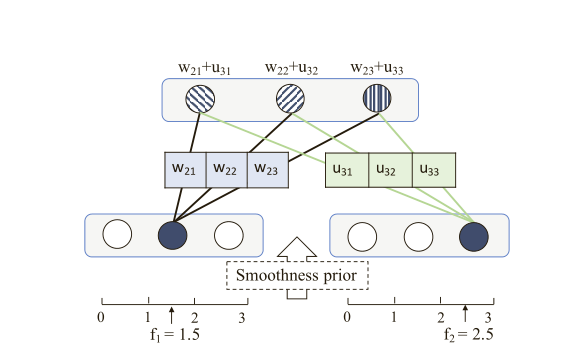
我们首先证明F（Y | c j + 1）= F（Y | c j）+ε。 我们可以证明F（Y | c j + 1）≥F（Y | cj）+ε由定理F（Y | cj + 1）≤F（Y | cj）+ε构成，因此F（Y | cj + 1）= F（Y | cj）+ ε。 为了通过矛盾来证明F（Y | c j + 1）≥F（Y | cj）+ε，假设F（Y | cj + 1） 分布函数F，在δδ> 0的半径c j + 1附近存在F（Y | c j + 1 +δ）<F（Y | cj）+ε，这与c j + 1。 因此，F（Y | c j + 1）= F（Y | c j）+ε。 接下来我们证明F（Y | c -i + 1）-F（Y | c i）≤ε。 由定义F（Y | c -j + 1）≤F（Y | c j + 1-δ）对于δ> 0来说。定义cj + 1 = sup {z：F（Y | z）≤F Y | cj）+ε}，我们有F（Y | c-j + 1）≤F（Y | c j + 1-δ）≤F（Y | cj）的F（Y | c j + 1-δ） ）+ε。 这完成了我们的证明F（Y | c -i + 1） - F（Y | c i）≤ε.

定理3.1证明了先验光滑性证明存在对随机变量X的特征空间的划分或离散化，使得条件累积分布F（Y | c -i + 1）-F（Y | ci ）在每个区间内是任意小的。更具体地说，在交通方式检测领域，如果给定速度v的步行y的累积概率是F（y | v），则存在这样的离散化，即在由该离散化定义的速度区间v∈[ci，c i + 1）内，步行的累积概率更多或不太一致，交通方式的变化不会突然发生。这个定理的证明基于引理1.1 [39]证明了Glivenko-Cantelli定理[44]。

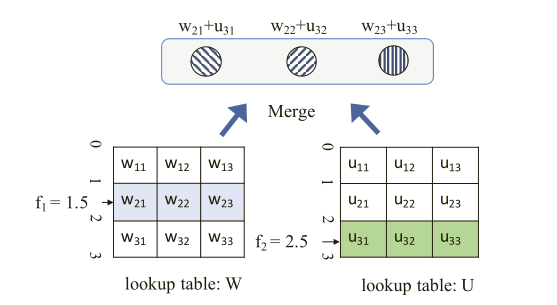
先验平滑使得我们可以通过离散化嵌入连续的特征，嵌入离散的属性。我们只是区分不同的时间间隔，并没有限制，即在相同的时间间隔内需要不同的参数化。离散化在密度估计中已经被广泛使用，如密度估计树[35]，它使用分段常数函数来估计概率分布。它也被用于数据挖掘算法，如C4.5 [34]和朴素贝叶斯[46] 。从贝叶斯的观点来看，离散化允许我们使用P（Y = y | X = x \*）来估计P（Y = y | X = x）其中x \*是输入x的离散版本，y是标签。Yang 等人 [46]表明，“离散化等同于使用真正的概率密度函数”在朴素贝叶斯框架，它是经验地使用离散化，而不是不安全的参数假设分布。此外，与连续特征的有限值相比，离散化通过大数定律来加强参数估计，其中每个区间有更多样本可用。在基于点的轨迹分类的背景下，通过开发特征组合来解决数据变化的因素是困难的[3]低维密集的特征空间。因此，将连续特征离散化并将其嵌入矢量表示中以更好地进行分类是合理的。

3.2.3离散化技巧。我们介绍通过矩阵乘法将离散特征映射到嵌入空间的“离散化技巧”，如图2a所示。 Te算法首先将连续特征转换为一个单向热矢量，该向量表示特征所在的区间：第一特征f 1 = 1。 5被转换成i = [0,1,0]并且第二特征f 2 = 2。 5被转换成j = [0,0,1]。然后通过矩阵乘法[i，j]×将单热矢量转换成嵌入空间

 ×=iW + jU = +,其中w e = [W 21，W 22，W 23]和u e = [U 31，U 32，U 33]。该方法与图2b所示的嵌入查找方法是等价的，它首先定义嵌入查找表W和U，根据输入值选择行向量（w e和u e），并通过元素相加来组合它们。我们在实验中使用前一种方法，因为它提供了一种通过直接矩阵乘法嵌入的简单方法。这种嵌入是通过反向传播学习的。



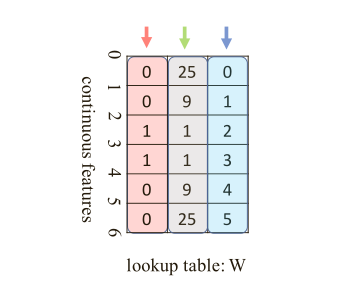
1. 通过矩阵运算嵌入



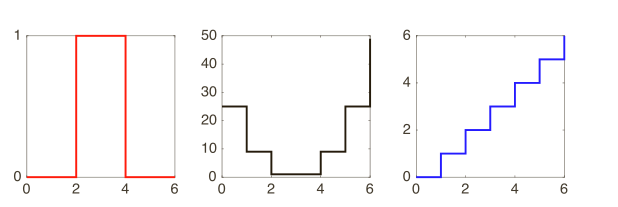
1. 通过查找表进行嵌入

3.2.4离散化策略。有意义的分割需要为语义嵌入提供足够的预测能力。本文研究了三种离散化策略，即等宽化，递归最小熵分割（RMEP）和模糊离散化。等宽合并确定每个特征的范围，然后用相等宽度的间隔分割这个范围。这种方法是无人监督的，直接执行。 RMEP [13]使用Shannon熵[40]来测量每个分区内标签的杂质，并利用信息增益递归地划分每个特征。 RMEP改进了监督预期的间隔的预测能力。当达到特定数量的分箱时，或者按照最小描述长度原则[36]，递归终止。模糊离散化旨在通过允许连续值以某种（梯形）隶属函数软化的方式属于不同的间隔来改进重叠数据的结果[38]。

3.2.5直观的理由。总结关于特征空间嵌入的讨论，重点介绍其特征。嵌入通过将原始特征映射到使用不同嵌入向量表示不同值的另一个空间，开发了更多的特征表示，而不是直接在低维和异构特征空间上学习神经网络层的每个功能。这与内核技巧有一些相似之处。但是，适应性分段的基础在这里没有明确的内核函数设计就实现了特征空间的扩展。而嵌入矩阵的行图像表示将连续值转换成矢量的过程，嵌入表的列图像可以理解为一种基本扩展。表3a示出了将属性D∈[0,6]映射到三维嵌入空间中的示例嵌入矩阵。如图3b所示，每个嵌入维度可以被看作是一个基本扩展，它检测输入空间中的不同模式 - 第一维检测[2，4]中的值，第二维是一个二次变换而最后一个维度是一个身份近似。每个基础扩展可以被视为专家，专门检测不同的投入产出关系。模型的输出可以被解释为专家的混合[24]，它涵盖了具有不同非线性函数的决策边界的不同区域。



（a）嵌入

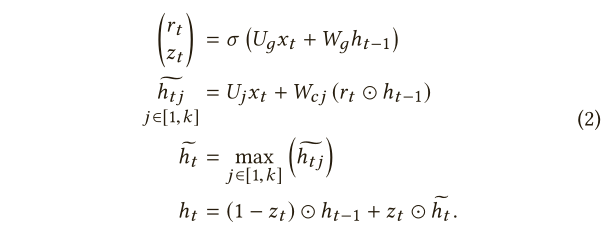


（b）基础扩展的嵌入

图3：基础扩展：嵌入查询表的列图像

#### 3.3Maxout门控递归单元

在计算候选状态e h t时，我们建议Maxout GRU应用Maxout激活[17]而不是双曲正切（tanh）。 最大激活取最大值的一组线性变换导致自适应凸函数。 它比tanh更有表现力，因为只有两个最大隐藏单元可以任意地逼近任何连续函数[17]。 它也有更好的梯度性能。 我们将（2）中的Maxout GRU定义为“表达”到下一个层次 - 除了学习的嵌入之外，我们还获得更多的灵活的记忆状态。 在等式2中，j表示分段线性变换的指数，k是分段变换的总数，其余的记号与公式1相同。



我们将这个模型与3.2节中定义的离散化方法结合使用，将输入x t映射到嵌入，从而可以学习独立嵌入来计算门和候选存储器状态。 除了学习语义空间，这种嵌入通过减少门单元和候选状态之间的相关性来改善偏差; 这是通过解耦两个查找矩阵U g和U j来完成的。

#### 3.4网络架构

图4显示了架构TrajectoryNet。 第一层学习嵌入空间，第二层学习特征构成。 我们使用双向GRU而不是单向GRU来考虑双向信息，以获得更好的预测能力[18]。

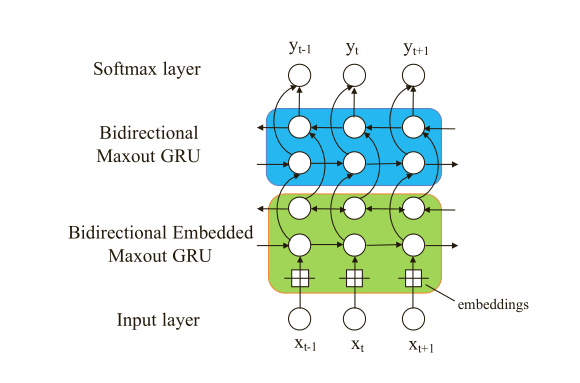


图4：轨迹网络体系结构

## 4 实验

#### 4.1实验设置

4.1.1数据。 我们使用Zheng等人收集的数据集。 [51]和随机选择23个人的轨迹与2,194,281个GPS记录。 这个数据集提供了大量的数据和各种运输方式。 我们关注四类分类任务，最后谈到七类结果。

4.1.2特性 使用两种主要类型的特征来检测交通模式：基于点的和基于段的特征。 基于点的特征与每个单独的GPS记录相关联，并且基于段的特征来自旨在提供更高的信息顺序的段并且规则化噪声 - 敏感的基于点的功能。 更具体地说，我们使用原始GPS记录计算以下位置和用户不可知特征：基于点的速度Vp，每段平均速度Vavg和每段速度的标准偏差Vsd。 但是，从GPS记录计算的特征的质量可能是不准确的由于传感器相关的原因[11]。 为了减轻这种不确定性，我们使用Hampel滤波器[20]来识别特征空间中的异常值并对其进行滤波。

4.1.3网络培训 我们使用时间截断的反向传播[45]，用Adam优化器[26]以小批量优化交叉熵损失。 学习率的阈值为0.01，我们使用基于验证的早期停止来改进泛化[4]。 我们在[0，0.001]范围内使用统一的初始化。 [43]。 我们对不同的网络结构进行了实验，发现每层有50个隐藏节点的两层结构效果最好。 每个功能分为20个区间，嵌入维度为50，Maxout激活由5个变换组成。

4.1.4评估。 我们使用分层留批次法（SLOBO）来评估学习模型，将23人的数据分为三组：分别包含16人，1人和6人的轨迹的训练，验证和测试集。 由于不同人的轨迹可能包含不同比例的交通方式，我们选择三个小组，目的是使代表人口的交通方式的比例相近。 分层可减少训练过程中的差异，并防止基于验证的早期停止不良泛化。 由于每条轨迹都有其自身的内在特征，因此SLOBO还具有泛化超越新个人轨迹的能力。

选择四种方法来衡量模型的性能：基于点的分类精度一个点，基于距离的精度一个距离，交叉熵损失E H和平均F1得分A F1。 表1显示了这些说明。

表1：评估方法的描述

|  |  |
| --- | --- |
| 方法 | 描述 |
| Apoint | 基于GPS样本数量的精度 |
| Adistance | 基于行驶距离的精确度与[51] |
| EH | 洞察模型训练过程（学习曲线） |
| AF1 | 在不同阶层失衡的情况下进行绩效评估 |

#### 4.2结果和讨论

4.2.1与基准方法的比较。 表2显示了基于距离的评估测量的测试数据的F1分数和准确性。 最常见的分类错误是汽车和公共汽车之间以及步行和公共汽车。 总的来说，TrajectoryNet的准确率高达98％，并且在相同的数据集上大大超越了现有的基准方法，包括决策树，支持向量机，朴素贝叶斯和条件随机场。 Zheng等人提出的基于决策树的框架提高了F1评分。 [51，52]每班有31％，16％，22％和22％。 这表明了所提出的TrajectoryNet的总体有效性。 我们在本节的其余部分进一步分析拟议的TrajectoryNet的各个组件的影响。

表2：轨迹网性能在测试数据中

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | F1 分数 | | | | | Accuracy |
| Bike | car | Walk | Bus | Average |
| 轨迹网（本方法） | 0.988 | 0.980 | 0.972 | 0.980 | 0.980 | 0.979 |
| 决策树[51] | 0.675 | 0.814 | 0.757 | 0.748 | 0.749 | 0.762 |
| 支持向量机[52] | - | - | - | - | - | 0.462 |
| 朴素贝叶斯[52] | - | - | - | - | - | 0.523 |
| 条件随机场[52] | - | - | - | - | - | 0.544 |

4.2.2嵌入的效果 图5显示了有和没有嵌入的GRU的学习曲线。 所有三个嵌入式GRU都具有更低的交叉熵损失E H，并且比常规GRU更快地收敛：E H嵌入式GRU仅在两个时期内达到比GRU在70多个时期内达到的最好的E H。 这种加速是因为嵌入使不同输入值之间的依赖关系解耦，使得优化变得更直接。 与GRU相比，嵌入式GRU更稳定，因为GRU受到爆炸梯度的影响在循环20（ReLU）和55（tanh）中。 这种改进是由于这样一个事实，即嵌入空间中的线性决策边界甚至可以被映射到原始空间中的高度非线性函数，同时更容易优化。 它强化了我们声称通过嵌入学习的表示提高了预测性RNN的力量。

4.2.3不同激活函数的作用 我们还强调了图5中不同激活函数的作用。我们发现Maxout激活函数收敛不管是否使用嵌入。 这表明Maxout激活可以在GRU中学习更多灵活的内存状态。 请注意，学习曲线中的符号是亚当随机梯度下降与小批量的结果。 鉴于Tanh和ReLU从爆炸梯度中获益，而不使用嵌入，Maxout激活并不能从爆炸梯度中获得，因为它具有更好的梯度特性 - 分段线性。

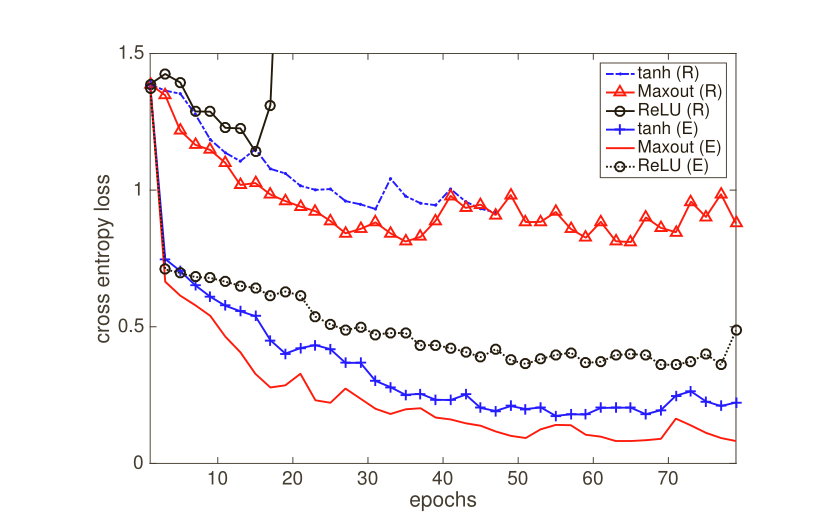


图5：具有不同激活函数的GRU方法的学习曲线。 （R）表示没有嵌入的GRU，（E）表示嵌入的GRU

4.2.4需要基于细分的功能。 表3显示了使用不同特征组合的分类结果，并且证明了基于细分的特征可以提高模型的性能。 这与我们关于概率密度函数的结论是一致的，不同的特征相互补充以提供各种运输方式之间的分离。 这也表明，基于细分的功能可以丰富基于点的功能。 请注意，我们在整个实验中使用了相同的网络架构。 还要注意标记的方法★在表中是明显好于没有方法★在对E H进行t检验时，95％的置信度水平，所有方法中差异不显着。 在95％的信任水平

表3：前项特征选择

|  |  |  |  |
| --- | --- | --- | --- |
|  | EH | Apoint | AF1 |
| Vp | 0.24 | 0.91 | 0.92 |
| Vp,Vavg | 0.11 | 0.97 | 0.97 |
| Vp,Vavg,Vsd | 0.08 | 0.98 | 0.98 |

4.2.5分割的效果。 选择分割策略可以预期会影响模型的性能。 表4显示了不同分割方法的结果。 基于距离的方法执行最差，因为不同的运输模式以不同的速度行进，导致每个段内的样本数量变化。 这进一步影响了质量由于在不同的运输模式下不同的采样复杂性导致的基于细分的特征。 其余的分割方法没有统计学差异，并且基于方位的分割方法具有最好的平均性能。 请参考[10]进一步比较和讨论。

表4：分割的效果

|  |  |  |  |
| --- | --- | --- | --- |
|  | EH | Apoint | AF1 |
| Time | 0.11 | 0.95 | 0.95 |
| Distance | 0.25 | 0.86 | 0.86 |
| Bearing | 0.08 | 0.98 | 0.98 |
| Window | 0.09 | 0.97 | 0.97 |

4.2.6离散化的效果。 表5显示了不同离散化方法的结果。 等宽分箱与基于熵的方法相比可能看起来令人惊讶，但我们发现等宽分箱强健，易于实现且易于训练。 模糊编码[38]不像以前那样工作两种方法。 由于本实验中使用的梯形模糊函数，模型不得不同时学习两个嵌入向量的加权和，这使得优化变得困难。这意味着给定分区的先验和适当粒度的平滑性，重叠区间不是 学习好模型的必要性。 如表6所示，我们对各种离散粒度进行了实验，发现当将间隔数从10增加到20时，模型得到了显着改善，但是当每个特征的间隔数在20到50之间时。

表5：离散化策略的影响

|  |  |  |  |
| --- | --- | --- | --- |
|  | EH | Apoint | AF1 |
| Width | 0.08 | 0.98 | 0.98 |
| Entropy | 0.19 | 0.95 | 0.95 |
| Fuzzy | 0.14 | 0.98 | 0.96 |

表6：离散化粒度的影响

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Intervals | 10 | 20 | 30 | 40 | 50 |  |
| Eh | 0.168 | 0.076 | 0.062 | 0.068 | 0.070 |  |
| Standard error | 0.004 | 0.013 | 0.005 | 0.007 | 0.012 |  |

4.2.7七级分类。 为了进一步验证TrajectoryNet的有效性，我们承担了更具挑战性的任务：将GPS记录分为七类，即火车，汽车，公共汽车，地铁，飞机和自行车。 这是一个更具挑战性的任务，我们实现了97.3％的分类分类精确度平均F1得分为93.0％，如表7所示。与四级分类任务相比，模型性能的下降主要是由于地铁和飞机两类训练案例数量不足造成的。

表7：混淆矩阵检测7种运输方式（Apoint = 97.3%）

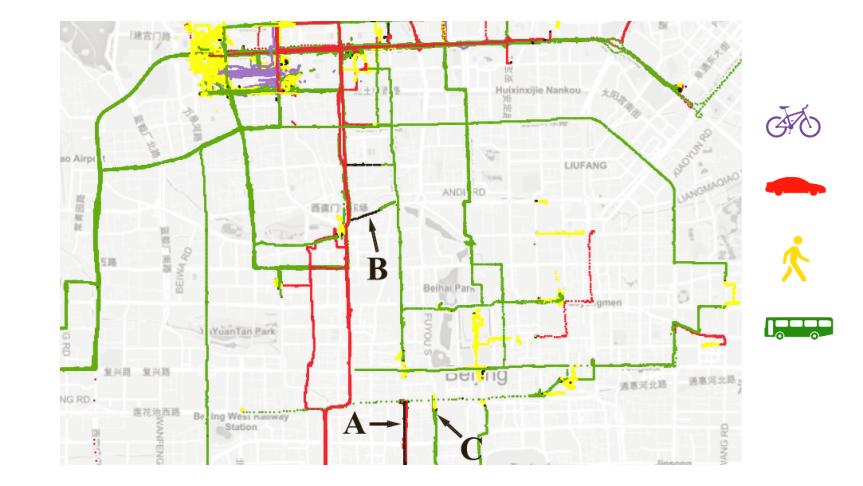
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 目标 | 预测 | | | | | | | Recall |
| Train | Car | Walk | Bus | Subway | Airplane | Bike |
| Train | 380563 | 889 | 853 | 67 | 514 | 1 | 0 | 0.994 |
| Car | 83 | 66309 | 261 | 938 | 852 | 0 | 300 | 0.965 |
| Walk | 82 | 260 | 139026 | 2397 | 784 | 26 | 1217 | 0.967 |
| Bus | 29 | 298 | 2920 | 121853 | 76 | 12 | 421 | 0.970 |
| Subway | 1394 | 1405 | 6083 | 142 | 24875 | 10 | 45 | 0.733 |
| Airplane | 4 | 0 | 43 | 23 | 0 | 1979 | 13 | 0.960 |
| Bike | 0 | 0 | 95 | 26 | 14 | 0 | 64194 | 0.998 |
| 精确度 | 0.996 | 0.959 | 0.931 | 0.971 | 0.917 | 0.976 | 0.970 |  |
| F1分数 | 0.995 | 0.962 | 0.949 | 0.971 | 0.815 | 0.837 | 0.984 |  |

4.2.8可视化分类结果。图6显示了测试数据中的预测。自行车，汽车，步行和公共汽车的交通方式分别是紫色，红色，黄色和绿色。错误分类的GPS记录是黑色的。该图显示了TrajectoryNet的总体有效性 - 只有很小一部分数据被错误分类。为了进一步调查TrajectoryNet引起的分类错误，我们在数据中突出显示了三种情景，在图中标记为A，B和C.情景错误分类的汽车进入公交车和情景B错误分类的公交车进入汽车。这些是TrajectoryNet最常见的分类错误;区分这两个类别是有挑战性的，因为它们有时表现出类似的特征。情景C很难从可视化中注意到，因为它们只发生在两种运输方式之间的转换过程中。这是错误分类的另一种常见类型，但仅在极少数情况下才会发生，这意味着所提出的TrajectoryNet在检测不同模式之间的转换时非常敏感。

## 5结论和未来的工作

我们提出了一个用于点对点网络的神经网络架构，从GPS轨迹推断真实世界的人类交通模式。为了克服捕捉GPS数据所强加的低维和异构特征空间的语义的挑战，我们开发了一种将原始特征空间嵌入到另一个空间的新颖表示，这个空间可以被理解为一种基本扩展的形式。嵌入可以被看作是一种基础扩展的形式，提高了原始特征空间的预测能力。嵌入也可以被看作是专门研究具有不同非线性函数的决策边界的不同区域的专家混合。我们还使用基于片段的特征来丰富特征空间，并使用Maxout激活来提高RNN的记忆状态的表现力。我们的实验证明，在检测4种和7种类型的交通模式时，所提出的模型相对于基线结果实现了显着的改善，分类精确度超过98％和97％。

对于今后的工作，我们考虑结合GIS信息等基于位置的先验知识，开发在线分类系统和建立用户依赖分类表，进一步完善该系统。 我们还考虑将所提出的嵌入方法应用于其他类型的低维和异构时间序列数据，例如， 物联网，进一步探讨了提出的嵌入方法的有效性。



图六: 可视化分类结果