(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114757975 B
(45)授权公告日 2024.04.16

(21)申请号 202210464974.X
(22)申请日 2022.04.29
(65)同一申请的已公布的文献号
申请公布号 CN 114757975 A
(43)申请公布日 2022.07.15
(73)专利权人 华南理工大学
地址 510640 广东省广州市天河区五山路381号
(72)发明人 徐红云 邝涛杰 姚楷曦 李怡泽 罗咫酉 张静怡 屈一伟 苏怡
(74)专利代理机构 广州市华学知识产权代理有限公司 44245
专利代理师 冯炳辉
(51)Int.Cl.
G06T 7/246(2017.01)
G06N 3/0464(2023.01)
G06N 3/08(2023.01)
(56)对比文件
CN 110781838 A,2020.02.11
CN 112347923 A,2021.02.09
CN 112766561 A,2021.05.07
CN 113269114 A,2021.08.17
CN 113269115 A,2021.08.17
审查员 何诚

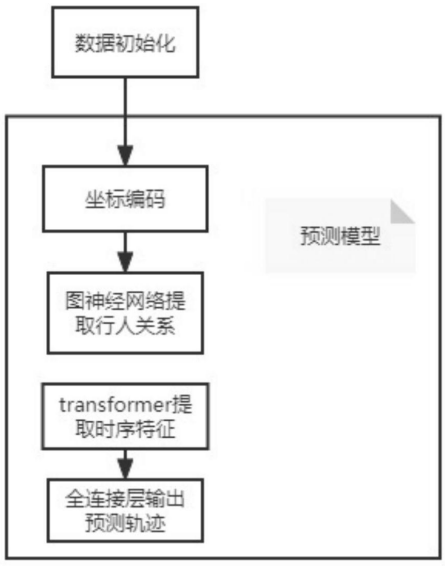
权利要求书6页 说明书12页 附图2页

(54)发明名称

基于transformer与图卷积网络的行人轨迹预测方法

(57)摘要

本发明公开了一种基于transformer与图卷积网络的行人轨迹预测方法,提取出若干个时间戳内所有包含坐标的行人轨迹数据;对每一个样本的每个时间戳做一个行人关系图;把总样本集分为测试集,训练集与验证集;再用行人关系图对时序数据进行图卷积学习,让轨迹数据附有行人关系;transformer用时序数据提取出每个时间戳的时序特征向量,利用每个时间戳的时序特征向量来生成遵循双变量高斯分布的具体轨迹分布;用损失函数对预测轨迹与真实未来轨迹作对比,得出损失值,再用损失值对预测模型优化,取出最优预测模型,把测试集输入到最优预测模型,得出预测轨迹数据。本发明可准确预测出未来行人的轨迹。



CN 114757975 B

1.基于transformer与图卷积网络的行人轨迹预测方法,其特征在于,包括以下步骤:

1)提取出若干个时间戳内所有的行人轨迹数据,包含x、y坐标的行人信息;前T_{obs}个时间戳为历史轨迹数据$V_{1:T_{obs}}=\{V_1,V_2,...,V_{T_{obs}}\}$,$T_{obs}$为历史轨迹时间戳长度,$V_{1:T_{obs}}\in R^{T_{obs}\times n\times axis}$,R表示是属于实数域,n为行人个数,axis表示坐标维度,$V_{1:T_{obs}}$简称为V;后T_{pred}个时间戳为未来轨迹数据$V_{T_{obs}+1:T_{obs}+T_{pred}}=\{V_{T_{obs}+1},V_{T_{obs}+2},...,V_{T_{pred}+T_{obs}}\}$,$T_{pred}$为预测轨迹时间戳长度,$V_{T_{obs}+1:T_{pred}+T_{obs}}\in R^{T_{pred}\times n\times axis}$;对每一个样本的每个时间戳做一个行人关系图G;V、$V_{T_{obs}+1:T_{pred}+T_{obs}}$、G的集合为一个样本;以若干个样本为一个批进行并行处理;把总样本集分为训练集、验证集和测试集;将预测模型f()形式化为:

$$\widehat{V}_{T_{obs}+1:T_{pred}+T_{obs}}=f(V,\phi)$$

式中,$\widehat{V}_{T_{obs}+1:T_{pred}+T_{obs}}$是预测轨迹数据,φ是预测模型f()中可学习的参数;

2)先用全连接网络对V进行坐标编码,提取V的坐标特征表示V_{emb},$V_{emb}\in R^{T_{obs}\times n\times d_{model}}$,编码空间维度大小为$d_{model}$;再用行人关系图G对$V_{emb}$进行图卷积学习,提取附有行人关系信息的行人坐标编码V_g;

3)采用transformer的编码器将附有行人关系信息的行人坐标编码V_g提取出每个时间戳的时序特征向量,用transformer的解码器以每个时间戳的时序特征向量为输入来生成具体行人轨迹分布,该行人轨迹分布遵循双变量高斯分布;

4)采用损失函数把预测轨迹数据与未来轨迹数据作对比生成损失值,再用反向传播损失值优化预测模型;在优化预测模型时,用训练集对预测模型进行训练,用验证集挑选最优预测模型,把测试集输入到最优预测模型,得出预测轨迹数据。

2.根据权利要求1所述的基于transformer与图卷积网络的行人轨迹预测方法,其特征在于,在步骤1)中,Vp_i^j表示在第i秒内第j个行人的坐标,每个样本至少有两条行人轨迹;

$$V=Vp_i^j,i\le T_{obs}$$

$$V_{T_{obs}+1:T_{obs}+T_{pred}}=Vp_i^j,T_{obs}<i\le T_{obs}+T_{pred}$$

把每个样本分为历史轨迹数据V与未来轨迹数据$V_{T_{obs}+1:T_{obs}+T_{pred}}$;

$$G_{i,j}^t=1/\|S_i^t+S_j^t\|_2$$

$$S_i^t=Xrel_i^t+Yrel_i^t$$

式中,G为行人关系图,$G\in R^{T_{obs}\times n\times n}$,$S_i^t$、$S_j^t$、$Xrel_i^t$和$Yrel_i^t$为在一个样本中第i个行人在第t个时间戳的合速度向量、横坐标分量速度向量和纵坐标分量速度向量;$G_{i,j}^t$表示第t个时间戳中,第i个行人与第j个行人的相互关系;

一个批次中包含若干个样本,便于预测模型的并行运行,再把若干个批次分为训练集、验证集与测试集,分别用来训练预测模型、取最优预测模型、测试预测模型。

3.根据权利要求1所述的基于transformer与图卷积网络的行人轨迹预测方法,其特征在于,在步骤2)中,把V中的x、y坐标信息进行编码与图卷积操作;

2.1)首先用n_{emb_axis}层全连接层对x、y坐标做编码操作,公式如下:

$$V_{emb}^{i,t}=V_{emb}^{i-1,t}*W_f^i$$

式中,$V_{emb}^{i,t}$表示在全连接层第i层,第t个时间戳的行人集坐标编码;$V_{emb}^{i-1,t}$表示在全连接层第i-1层,第t个时间戳的行人集坐标编码;W_f^i表示在第i层的全连接层可学习矩阵参数,*表示矩阵乘法;第一层全连接层把x、y坐标维度axis维扩展为d_{model}维,当i=2,3,...,n_{axis_emb}时,第i层全连接层的输入坐标编码维度和输出坐标编码维度都维持d_{model}维;

2.2)采用图卷积神经网络利用行人关系图,把行人集坐标编码进行空间卷积运算;邻接性的归一化是图卷积正常工作的必要条件,先对每一个时间戳的行人关系图进行标准化处理,行人关系图$G=\{G_1,G_2,...,G_t,...,G_{T_{obs}}\}$,$G_t$表示第t个时间戳的行人关系图,公式如下:

$$G_t^{'}=D_t^{-1/2}(G_t+I)D_t^{-1/2}$$

式中,G'_t为第t个时间戳的标准化行人关系图,I为单位矩阵,D_t为对角节点度矩阵;

用n_{gcn}层图卷积层赋予轨迹信息以行人关系信息,公式如下:

$$F_{gcn}(V_{emb}^i,G^{'})=\tanh(G^{'}*V_{emb}^i*W_g^i)$$

式中,$V_{emb}^i=\{V_{emb}^{i,1},V_{emb}^{i,2},...,V_{emb}^{i,t},...,V_{emb}^{i,T_{obs}}\}$, $V_{emb}^i\in R^{T_{obs}\times n\times d_{model}}$表示在图卷积层第i层的行人集坐标编码,$V_{emb}^{i,t}$示在图卷积层第i层、第t个时刻的行人集坐标编码,$F_{gcn}$表示图卷积操作,$W_g^i$表示在第i层的图卷积层可学习矩阵参数;*表示矩阵乘法;tanh为图卷积层的激活函数;最后一层图卷积完成的输出为附有行人关系信息的行人坐标编码V_g,$V_g\in R^{T_{obs}\times n\times d_{model}}$。

4.根据权利要求1所述的基于transformer与图卷积网络的行人轨迹预测方法,其特征在于,在步骤3)中,把transformer作为预测模型基底,进行轨迹预测:

用transformer以图卷积的输出为输入,考虑历史轨迹数据V对预测第i个时间戳的行人位置分布的影响;

a、位置编码,公式如下:

$$\tau_n^t(k)=\begin{cases}\sin(10000^{-k/d_{model}}),k\bmod\equiv 0\\ \cos(10000^{-k/d_{model}}),k\bmod\equiv 1\end{cases}$$

$V_{gt}=V_g+\tau$

式中,$\tau_n^t(k)$为第t个时间戳,第n个行人的第k个特征值,τ为行人位置编码;把行人位置编码与V_g相加得到附有行人位置编码的V_{gt};

b、编码器:

编码器以经过位置编码处理的V_{gt}为输入,经过注意力机制,提取特征A_{f3};Transformer的编码器有6层子编码器,每一层的子编码器结构相同,当u等于1时,下面的子编码器代表第u个子编码器;当u=2,3,…,6重复以上操作;

b1、子编码器的第一部分是多头注意力机制,公式如下:

$$q_i=V_{gt}*W_{q_i}$$

$$k_i=V_{gt}*W_{k_i}$$

$$v_i = V_{gt} * W_{v_i}$$

$$A_i(q_i, k_i, v_i) = soft\max(\frac{q_i * k_i}{\sqrt{d_{model}}} v_i)$$

$A = F_{cat}(A_i) * W_a$

式中，q_i、k_i、v_i表示子编码器第i个头的查询、键、值；*为矩阵运算，q_i、k_i、$v_i \in R^{T_{obs} \times n \times d_{model}}$；$W_{q_i}$、$W_{k_i}$、$W_{v_i}$为子编码器的第i头注意力把$V_{gt}$生成$q_i$、$k_i$、$v_i$的可学习矩阵参数；$A_i$表示子编码器第i个头注意力机制的注意力，$A_i \in R^{T_{obs} \times n \times d_{model}}$，$F_{cat}$函数的作用是把多个头的注意力拼接起来，$W_a$为把拼接起来的$A_i$变化成A的可学习矩阵参数；A为附有注意力机制的行人位置编码，$A \in R^{T_{obs} \times n \times d_{model}}$；

用残差网络和标准化函数对行人特征表示进行处理，公式如下：

$A_{f1} = F_{norm}(V_{gt} + F_{drop}(A))$

式中，A_{f1}为经过残差网络和标准化函数处理的注意力，$A_{f1} \in R^{T_{obs} \times n \times d_{model}}$；$F_{norm}()$为标准化函数；

b2、子编码器的第二部分是前向反馈网络，公式如下：

$$A_{f2} = F_{drop}(\tanh(\tanh(A_{f1} * W_{fa}^1) * W_{fa}^2))$$

式中，A_{f2}为经过前向反馈网络的子编码器注意力，$A_{f2} \in R^{T_{obs} \times n \times d_{model}}$；$F_{drop}$为随机参数不更新函数；tanh为激活函数，$W_{fa}^1$、$W_{fa}^2$为第一、二层全连接层可学习矩阵参数；再用一遍残差网络及标准化函数对行人特征表示进行处理，公式如下：$A_{f3} = F_{norm}(A_{f1} + F_{drop}(A_{f2}))$，$A_{f3}$为经过残差网络和标准化函数处理的注意力；

以上的步骤为一个子编码器的过程，而transformer的编码器部分为6个这样的子编码器串行拼接；当u＝1，2，...，5时，A_{f3}为下一层子编码器的输入，当u＝6时，A_{f3}会作为整个transformer的编码器的输出，$A_{f3} \in R^{T_{obs} \times n \times d_{model}}$；

c、解码器：

解码器的工作目标是预测第t个时间戳的行人位置$\hat{V}_t$，解码器以编码器的输出A_{f3}和经过了位置编码的已预测时间戳的行人位置$\hat{V}_{T_{obs}+1:t}^{gt}$为输入，输出为$\hat{V}_t$；

与编码器一样，Transformer的解码器有6层子解码器，每一层的子解码器结构相同，当we等于1时，下面的子解码器代表第we个子解码器，当we＝u＝2，3，…，6重复以上操作；

c1、子解码器的第一部分是掩码多头注意力机制：

因为不能泄露行人未来轨迹信息，这里进行掩码操作，公式如下：

$$V_t^{dgt} = \begin{cases} \hat{V}_t^{gt}, t > T_{obs} + y \\ 0, otherwise \end{cases}$$

式中，$\hat{V}_t^{gt}$为第t个时间戳的经过了位置编码的已预测时间戳的行人位置，V_t^{dgt}表示子解码器输入在第t个时间戳的行人轨迹信息编码，y为已预测未来轨迹步长，

$V_{T_{obs}+1:T_{obs}+T_{red}}^{dgt}=\{V_{T_{obs}+1}^{dgt},V_{T_{obs}+2}^{dgt},...,V_{t}^{dgt},...,V_{T_{obs}+T_{pred}}^{dgt}\}$是在解码器中行人轨迹信息编码集合,这里把$V_{T_{obs}+1:T_{obs}+T_{red}}^{dgt}$简称为$V_{dgt}$;

注意力操作公式如下:

$$q_i^{dec}=V_{dgt}*W_{q_i^{dec}}^{dec}$$

$$k_i^{dec}=V_{dgt}*W_{k_i^{dec}}^{dec}$$

$$v_i^{dec}=V_{dgt}*W_{v_i^{dec}}^{dec}$$

$$A_i^{dec}(q_i^{dec},k_i^{dec},v_i^{dec})=soft\max(\frac{q_i^{dec}*k_i^{dec}}{\sqrt{d_{model}}}v_i^{dec})$$

$$A^{dec}=F_{cat}(A_i^{dec})*W_a^{dec}$$

式中,q_i^{dec}、k_i^{dec}、v_i^{dec}表示子解码器第i个掩码多头注意力机制的查询、键、值,q_i^{dec}、k_i^{dec}、$v_i^{dec}\in R^{T_{pred}\times n\times d_{model}}$;$V_{dgt}$为在编码器中的行人轨迹信息编码;$W_{q_i^{dec}}^{dec}$、$W_{k_i^{dec}}^{dec}$、$W_{v_i^{dec}}^{dec}$为子解码器第i个掩码多头注意力机制把$V_{dgt}$生成为$q_i^{dec}$、$k_i^{dec}$、$v_i^{dec}$的可学习矩阵参数;$A_i^{dec}$表示子解码器第i个掩码多头注意力机制的注意力,$A_i^{dec}\in R^{T_{pred}\times n\times d_{model}}$,$W_a^{dec}$为把拼接起来的$A_i^{dec}$变化成$A^{dec}$的可学习矩阵参数;$A^{dec}$为附有掩码注意力机制的行人位置编码,$A^{dec}\in R^{T_{pred}\times n\times d_{model}}$;

用残差网络和标准化函数对行人特征表示进行处理;

$A_{f4}=F_{norm}(V_{dgt}+F_{norm}(A^{dec}))$

式中,A_{f4}为经过残差网络和标准化函数处理的解码器掩码多头注意力,$A^{dec}\in R^{T_{pred}\times n\times d_{model}}$;

c2、子解码器的第二部分是多头注意力机制,公式如下:

$$q_i^{dec}=V_{dgt}*W_{q_i^{dec2}}^{dec2}$$

$$k_i^{dec}=V_{dgt}*W_{k_i^{dec2}}^{dec2}$$

$$v_i^{dec}=V_{dgt}*W_{v_i^{dec2}}^{dec2}$$

$$A_i^{dec2}(q_i^{dec2},k_i^{dec2},v_i^{dec2})=soft\max(\frac{q_i^{dec2}*k_i^{dec2}}{\sqrt{d_{model}}}v_i^{dec2})$$

$$A^{dec2}=F_{cat}(A_i^{dec2})*W_a^{dec2}$$

式中,q_i^{dec2}、k_i^{dec2}、v_i^{dec2}表示子解码器第i个多头注意力机制的查询、键、值;这里的q_i^{dec2}由子解码器的掩码多头注意力模块的输出生成,而k_i^{dec2}与v_i^{dec2}由编码器的输出生成,*为矩阵运算,q_i^{dec2}、k_i^{dec2}、$v_i^{dec2}\in R^{T_{pred}\times n\times d_{model}}$;$W_{q_i^{dec2}}^{dec2}$、$W_{k_i^{dec2}}^{dec2}$、$W_{v_i^{dec2}}^{dec2}$为子解码器第i头多头注意力机制把$V_{dgt}$生成为$q_i^{dec2}$、$k_i^{dec2}$、$v_i^{dec2}$的可学习矩阵参数;$A_i^{dec2}$表示子解码器第i个多头注意力机制的注意力,$A_i^{dec2}\in R^{T_{pred}\times n\times d_{model}}$,$W_a^{dec2}$为把拼接起来的$A_i^{dec2}$变化成$A^{dec2}$的可学习矩阵

参数;A^{dec2}为结合历史轨迹数据和已预测轨迹数据的注意力,$A^{dec2} \in R^{T_{pred} \times n \times d_{model}}$;

用残差网络和标准化函数对A^{dec2}进行处理,公式如下:

$A_{f5}=F_{norm}(A_{f4}+F_{drop}(A^{dec2}))$

式中,A_{f5}为经过残差网络和标准化函数处理的子解码器注意力,$A_{f4} \in R^{T_{pred} \times n \times d_{model}}$;

c3、子解码器的第三部分是前向反馈网络,公式如下:

$$A_{f6} = F_{drop}(\tanh(\tanh(A_{f5} * W_{fda}^{1}) * W_{fda}^{2})$$

式中,W_{fda}^{1}、W_{fda}^{2}为可学习矩阵参数,再用一遍残差网络及标准化函数对行人特征表示进行处理,公式如下:

$A_{f7}=F_{norm}(A_{f5}+F_{drop}(A_{f6}))$

式中,A_{f7}为行人预测轨迹的特征,以上的步骤为一个子解码器的过程,而transformer的解码器部分为6个这样的子解码器串行拼接;当we=1,2,...,5时,这里的A_{f7}为下一层子解码器的输入,当we=6时,A_{f7}会作为整个transformer的解码器的输出,$A_{f7} \in R^{T_{pred} \times n \times d_{model}}$;

以时间维度聚合每一个子解码器的输出A_{f7}得到$A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7}$,$A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7}$为每一个预测时间戳的$A_{f7}$聚合;用全连接层对$A_{T_{obs}+1:T_{obs}:T_{pred}}^{f7}$进行处理,生成高斯分布:

$$Tr = soft\max(A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7} * W_{gass})$$

式中,Tr表示行人预测轨迹的高斯分布参数,$Tr \in R^{T_{pred} \times n \times d_{model}}$,$W_{gass}$为可学习变量;

d、全连接网络与双变量高斯分布

全连接网络以transformer的解码器的输出Tr为输入,输出第i个时间戳的行人位置分布,这里的行人位置分布为双变量高斯分布,公式为:

$$P(Tr_{i,n}) = p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$$

式中,$P(Tr_{i,n})$为第n个行人,第i个时间戳双变量高斯分布;p()为双变量高斯分布函数,$p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$为第i个时间戳,第n个行人的位置分布;$\mu_{i,n}^{x}$、$\mu_{i,n}^{y}$、$\sigma_{i,n}^{x}$、$\sigma_{i,n}^{y}$、$corr_{i,n}$分别表示第i个时间戳的第n个行人位置分布的x坐标的均值、y坐标的均值、x坐标的标准差、y坐标的标准差、x坐标与y坐标的相关性,$Tr_{i,n}$表示第i个时间戳第n个行人的高斯轨迹,所以第i个时间戳的一个行人位置分布需要五个参数,全连接层就是把transformer的解码器输出变成第i个时间戳的双变量(x,y)高斯分布。

5.根据权利要求1所述的基于transformer与图卷积网络的行人轨迹预测方法,其特征在于,在步骤4)中,用损失函数把所得到的双变量高斯分布与未来轨迹数据$V_{T_{obs}+1:T_{pred}+T_{obs}}$做差值;

损失函数L(W)为:

$$L(W) = -\sum_{n=1}^{n\_ped} \sum_{i=T_{obs}+1}^{T_{pred}} \log(p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n}))$$

第n个行人的损失函数为L,W为预测模型参数;$p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$为第i个时间

戳,第n个行人的位置分布;$\mu_{i,n}^{x}$、$\mu_{i,n}^{y}$、$\sigma_{i,n}^{x}$、$\sigma_{i,n}^{y}$、$\mathrm{corr}_{i,n}$分别表示第i个时间戳的第n个行人位置分布的x坐标的均值、y坐标的均值、x坐标的标准差、y坐标的标准差、x坐标与y坐标的相关性,n_ped为样本中行人个数;

使用时间反向传播算法和梯度优化方法ADAM训练预测模型,取最优预测模型;把行人的历史轨迹数据输入最优预测模型,就能生成行人预测轨迹。

基于transformer与图卷积网络的行人轨迹预测方法

技术领域

[0001] 本发明涉及时序数据预测的技术领域,尤其是指一种基于transformer与图卷积网络的行人轨迹预测方法。

背景技术

[0002] 目前基于深度学习的行人轨迹预测的研究有很多,Social-LSTM是最早专注于行人轨迹预测的深度模型之一。Social-LSTM使用一个RNN网络来模拟每个行人的运动轨迹特征,然后使用池化机制来聚合RNN的输出,也就是把行人周围的物体轨迹特征聚合在一起,以此为辅助信息,并且与需预测行人的轨迹特征相结合,从而预测之后的轨迹。Social-LSTM假设行人轨迹遵循双变量高斯分布,预测的轨迹不是一个确定的值,而是一个高斯分布,以此来模拟行人轨迹的不确定性。该工作是同时行人关系和时序关,并且使用神经网络进行模型训练的开山之作。但Social-LSTM在考虑行人关系时只考虑了距离较近的行人,不考虑距离较远的行人,这其实不符合真实情况,并且Social-LSTM使用lstm提取时序特征,效率和效果都太查了。后来的工作,如窥视未来轨迹(PIF)和轨迹状态细化(SR-LSTM),通过视觉特征和新的池化机制扩展了Social-LSTM,以提高预测精度,但他们仍然使用了lstm这种低级的时序提取方法。基于行人轨迹遵循多模态分布的假设,Social-GAN将Social-LSTM扩展为基于递归神经网络(RNN)的生成模型,利用对抗生成网络来生成更具有鲁棒性的轨迹,Social-GAN使用了GAN作为生成模型基底,但这种方法需要生成器和判别器的完美协调,所以结果是比不上端到端的神经网络模型。Sophie使用中枢神经网络从整个场景中提取特征,然后对每个行人使用双向注意机制。随后,Sophie将注意力输出与视觉CNN输出连接起来,然后使用一个基于长短期记忆(LSTM)自动编码器的生成模型来生成未来的轨迹,该方法考虑到了场景图和视觉图的重要性,但这也只是考虑到了辅助信息,其提取行人关系和时序特征的方法并没有改变。我注意到,以前的大多数工作都围绕着两个问题来建立深度学习网络,一是如何对行人的时序特征进行提取,常见的方法使用RNN网络来模拟每个行人运动,常见的RNN网络有LSTM、GRU等,也有的学者提出RNN的参数利用效率与时间效率很低,TCN这种方法又不时被人使用。二是如何提取行人之间的互动关系,很多基于Social-LSTM方法使用池化机制组合循环网络来提取行人之间的关系,也有工作使用图卷积网络方法来表示行人关系。最近的研究表明,Social-BiGAT依赖于图形注意网络来模拟行人之间的社会互动,LSTM的输出被输入到Social-BiGAT中的图中。

发明内容

[0003] 本发明的目的在于克服现有技术的缺点与不足,提出了一种基于transformer与图卷积网络的行人轨迹预测方法,使用在自然语言处理中取得出色表现的transformer来提取行人轨迹的时序信息,使用图卷积网络提取行人间的关系,从而预测未来行人的轨迹。

[0004] 为实现上述目的,本发明所提供的技术方案为:基于transformer与图卷积网络的行人轨迹预测方法,包括以下步骤:

[0005] 1)提取出若干个时间戳内所有的行人轨迹数据,包含x、y坐标的行人信息;前T_{obs}个时间戳为历史轨迹数据$V_{1:T_{obs}}=\{V_1,V_2,...,V_{T_{obs}}\}$,$T_{obs}$为历史轨迹时间戳长度,$V_{1:T_{obs}}\in R^{T_{obs}\times n\times axis}$,R表示是属于实数域,n为行人个数,axis表示坐标维度,$V_{1:T_{obs}}$简称为V;后T_{pred}个时间戳为未来轨迹数据$V_{T_{obs}+1:T_{obs}+T_{pred}}=\{V_{T_{obs}+1},V_{T_{obs}+2},...,V_{T_{pred}+T_{obs}}\}$,$T_{pred}$为预测轨迹时间戳长度,$V_{T_{obs}+1:T_{pred}+T_{obs}}\in R^{T_{pred}\times n\times axis}$;对每一个样本的每个时间戳做一个行人关系图G;V、$V_{T_{obs}+1:T_{pred}+T_{obs}}$、G的集合为一个样本;以若干个样本为一个批进行并行处理;把总样本集分为训练集、验证集和测试集;将预测模型f()形式化为:

[0006] $$\widehat{V}_{T_{obs}+1:T_{pred}+T_{obs}}=f(V,\phi)$$

[0007] 式中,$\widehat{V}_{T_{obs}+1:T_{pred}+T_{obs}}$是预测轨迹数据,φ是预测模型f()中可学习的参数;

[0008] 2)先用全连接网络对V进行坐标编码,提取V的坐标特征表示V_{emb},$V_{emb}\in R^{T_{obs}\times n\times d_{model}}$,编码空间维度大小为$d_{model}$;再用行人关系图G对$V_{emb}$进行图卷积学习,提取附有行人关系信息的行人坐标编码V_g;

[0009] 3)采用transformer的编码器将附有行人关系信息的行人坐标编码V_g提取出每个时间戳的时序特征向量,用transformer的解码器以每个时间戳的时序特征向量为输入来生成具体行人轨迹分布,该行人轨迹分布遵循双变量高斯分布;

[0010] 4)采用损失函数把预测轨迹数据与未来轨迹数据作对比生成损失值,再用反向传播损失值优化预测模型;在优化预测模型时,用训练集对预测模型进行训练,用验证集挑选最优预测模型,把测试集输入到最优预测模型,得出预测轨迹数据。

[0011] 进一步,在步骤1)中,Vp_i^j表示在第i秒内第j个行人的坐标,每个样本至少有两条行人轨迹;

[0012] $$V=Vp_i^j,i\le T_{obs}$$

[0013] $$V_{T_{obs}+1:T_{obs}+T_{pred}}=Vp_i^j,T_{obs}<i\le T_{obs}+T_{pred}$$

[0014] 把每个样本分为历史轨迹数据V与未来轨迹数据$V_{T_{obs}+1:T_{obs}+T_{pred}}$;

[0015] $$G_{i,j}^t=1/\|S_i^t+S_j^t\|_2$$

[0016] $$S_i^t=Xrel_i^t+Yrel_i^t$$

[0017] 式中,G为行人关系图,$G\in R^{T_{obs}\times n\times n}$,$S_i^t$、$S_j^t$、$Xrel_i^t$和$Yrel_i^t$为在一个样本中第i个行人在第t个时间戳的合速度向量、横坐标分量速度向量和纵坐标分量速度向量;$G_{i,j}^t$表示第t个时间戳中,第i个行人与第j个行人的相互关系;

[0018] 一个批次中包含若干个样本,便于预测模型的并行运行,再把若干个批次分为训练集、验证集与测试集,分别用来训练预测模型、取最优预测模型、测试预测模型。

[0019] 进一步,在步骤2)中,把V中的x、y坐标信息进行编码与图卷积操作;

[0020] 2.1)首先用n_{emb_axis}层全连接层对x、y坐标做编码操作,公式如下:

[0021] $$V_{emb}^{i,t}=V_{emb}^{i-1,t}*W_f^i$$

[0022] 式中,$V_{emb}^{i,t}$表示在全连接层第i层,第t个时间戳的行人集坐标编码;$V_{emb}^{i-1,t}$表示在全

连接层第i-1层,第t个时间戳的行人集坐标编码;W_f^i表示在第i层的全连接层可学习矩阵参数,*表示矩阵乘法;第一层全连接层把x、y坐标维度axis维扩展为d_{model}维,当i＝2,3,...,n_{axis_emb}时,第i层全连接层的输入坐标编码维度和输出坐标编码维度都维持d_{model}维;

[0023] 2.2)采用图卷积神经网络利用行人关系图,把行人集坐标编码进行空间卷积运算;邻接性的归一化是图卷积正常工作的必要条件,先对每一个时间戳的行人关系图进行标准化处理,行人关系图G＝{G_1,G_2,...,G_t,...,G_{Tobs}},G_t表示第t个时间戳的行人关系图,公式如下:

[0024] $$G_t^{'} = D_t^{-1/2}(G_t + I)D_t^{-1/2}$$

[0025] 式中,G_t'为第t个时间戳的标准化行人关系图,I为单位矩阵,D_t为对角节点度矩阵;

[0026] 用n_{gcn}层图卷积层赋予轨迹信息以行人关系信息,公式如下:

[0027] $$F_{gcn}(V_{emb}^i, G^{'}) = \tanh(G^{'} * V_{emb}^i * W_g^i)$$

[0028] 式中,$V_{emb}^i = \{V_{emb}^{i,1}, V_{emb}^{i,2}, ..., V_{emb}^{i,t}, ..., V_{emb}^{i,T_{obs}}\}$, $V_{emb}^i \in R^{T_{obs} \times n \times d_{model}}$表示在图卷积层第i层的行人集坐标编码,$V_{emb}^{i,t}$示在图卷积层第i层、第t个时刻的行人集坐标编码,$F_{gcn}$表示图卷积操作,$W_g^i$表示表示在第i层的图卷积层可学习矩阵参数;*表示矩阵乘法;tanh为图卷积层的激活函数;最后一层图卷积完成的输出为附有行人关系信息的行人坐标编码V_g,$V_g \in R^{T_{obs} \times n \times d_{model}}$。

[0029] 进一步,在步骤3)中,把transformer作为预测模型基底,进行轨迹预测:

[0030] 用transformer以图卷积的输出为输入,考虑历史轨迹数据V对预测第i个时间戳的行人位置分布的影响;

[0031] a、位置编码,公式如下:

[0032] $$\tau_n^t(k) = \begin{cases} \sin(10000^{-k/d_{model}}), k \bmod \equiv 0 \\ \cos(10000^{-k/d_{model}}), k \bmod \equiv 1 \end{cases}$$

[0033] V_{gt}＝V_g+τ

[0034] 式中,$\tau_n^t(k)$为第t个时间戳,第n个行人的第k个特征值,τ为行人位置编码;把行人位置编码与V_g相加得到附有行人位置编码的V_{gt};

[0035] b、编码器:

[0036] 编码器以经过位置编码处理的V_{gt}为输入,经过注意力机制,提取特征A_{f3};Transformer的编码器有6层子编码器,每一层的子编码器结构相同,当u等于1时,下面的子编码器代表第u个子编码器;当u＝2,3,…,6重复以上操作;

[0037] b1、子编码器的第一部分是多头注意力机制,公式如下:

[0038] $$q_i = V_{gt} * W_{q_i}$$

[0039] $$k_i = V_{gt} * W_{k_i}$$

[0040] $$v_i = V_{gt} * W_{v_i}$$

[0041] $$A_i(q_i,k_i,v_i)=soft\max(\frac{q_i*k_i}{\sqrt{d_{model}}}v_i)$$

[0042] $A=F_{cat}(A_i)*W_a$

[0043] 式中，q_i、k_i、v_i表示子编码器第i个头的查询、键、值；*为矩阵运算，q_i、k_i、$v_i\in R^{T_{obs}\times n\times d_{model}}$；$W_{q_i}$、$W_{k_i}$、$W_{v_i}$为子编码器的第i头注意力把$V_{gt}$生成$q_i$、$k_i$、$v_i$的可学习矩阵参数；$A_i$表示子编码器第i个头注意力机制的注意力，$A_i\in R^{T_{obs}\times n\times d_{model}}$，$F_{cat}$函数的作用是把多个头的注意力拼接起来，$W_a$为把拼接起来的$A_i$变化成A的可学习矩阵参数；A为附有注意力机制的行人位置编码，$A\in R^{T_{obs}\times n\times d_{model}}$；

[0044] 用残差网络和标准化函数对行人特征表示进行处理，公式如下：

[0045] $A_{f1}=F_{norm}(V_{gt}+F_{drop}(A))$

[0046] 式中，A_{f1}为经过残差网络和标准化函数处理的注意力，$A_{f1}\in R^{T_{obs}\times n\times d_{model}}$；$F_{norm}()$为标准化函数；

[0047] b2、子编码器的第二部分是前向反馈网络，公式如下：

[0048] $$A_{f2}=F_{drop}(\tanh(\tanh(A_{f1}*W_{fa}^1)*W_{fa}^2))$$

[0049] 式中，A_{f2}为经过前向反馈网络的子编码器注意力，$A_{f2}\in R^{T_{obs}\times n\times d_{model}}$；$F_{drop}$为随机参数不更新函数；tanh为激活函数，$W_{fa}^1$、$W_{fa}^2$为第一、二层全连接层可学习矩阵参数；再用一遍残差网络及标准化函数对行人特征表示进行处理，公式如下：$A_{f3}=F_{norm}(A_{f1}+F_{drop}(A_{f2}))$，$A_{f3}$为经过残差网络和标准化函数处理的注意力；

[0050] 以上的步骤为一个子编码器的过程，而transformer的编码器部分为6个这样的子编码器串行拼接；当u＝1,2,...,5时，A_{f3}为下一层子编码器的输入，当u＝6时，A_{f3}会作为整个transformer的编码器的输出，$A_{f3}\in R^{T_{obs}\times n\times d_{model}}$；

[0051] c、解码器：

[0052] 解码器的工作目标是预测第t个时间戳的行人位置$\hat{V}_t$，解码器以编码器的输出A_{f3}和经过了位置编码的已预测时间戳的行人位置$\hat{V}_{T_{obs}+1:t}^{gt}$为输入，输出为$\hat{V}_t$；

[0053] 与编码器一样，Transformer的解码器有6层子解码器，每一层的子解码器结构相同，当we等于1时，下面的子解码器代表第we个子解码器，当we＝u＝2,3,…,6重复以上操作；

[0054] c1、子解码器的第一部分是掩码多头注意力机制：

[0055] 因为不能泄露行人未来轨迹信息，这里进行掩码操作，公式如下：

[0056] $$V_t^{dgt}=\begin{cases}\hat{V}_t^{gt}, t>T_{obs}+y\\ 0, otherwise\end{cases}$$

[0057] 式中，$\hat{V}_t^{gt}$为第t个时间戳的经过了位置编码的已预测时间戳的行人位置，V_t^{dgt}表示子解码器输入在第t个时间戳的行人轨迹信息编码，y为已预测未来轨迹步长，$V_{T_{obs}+1:T_{obs}+T_{red}}^{dgt}=\{V_{T_{obs}+1}^{dgt},V_{T_{obs}+2}^{dgt},...,V_t^{dgt},...,V_{T_{obs}+T_{pred}}^{dgt}\}$是在解码器中行人轨迹信息编码集合，这里把

$V_{T_{obs}+1:T_{obs}+T_{red}}^{dgt}$简称为$V_{dgt}$;

[0058] 注意力操作公式如下:

[0059] $$q_i^{dec}=V_{dgt}*W_{q_i^{dec}}^{dec}$$

[0060] $$k_i^{dec}=V_{dgt}*W_{k_i^{dec}}^{dec}$$

[0061] $$v_i^{dec}=V_{dgt}*W_{v_i^{dec}}^{dec}$$

[0062] $$A_i^{dec}(q_i^{dec},k_i^{dec},v_i^{dec})=soft\max(\frac{q_i^{dec}*k_i^{dec}}{\sqrt{d_{model}}}v_i^{dec})$$

[0063] $$A^{dec}=F_{cat}(A_i^{dec})*W_a^{dec}$$

[0064] 式中,q_i^{dec}、k_i^{dec}、v_i^{dec}表示子解码器第i个掩码多头注意力机制的查询、键、值,q_i^{dec}、k_i^{dec}、$v_i^{dec}\in R^{T_{pred}\times n\times d_{model}}$;$V_{dgt}$为在编码器中的行人轨迹信息编码;$W_{q_i^{dec}}^{dec}$、$W_{k_i^{dec}}^{dec}$、$W_{v_i^{dec}}^{dec}$为子解码器第i个掩码多头注意力机制把$V_{dgt}$生成为$q_i^{dec}$、$k_i^{dec}$、$v_i^{dec}$的可学习矩阵参数;$A_i^{dec}$表示子解码器第i个掩码多头注意力机制的注意力,$A_i^{dec}\in R^{T_{pred}\times n\times d_{model}}$,$W_a^{dec}$为把拼接起来的$A_i^{dec}$变化成$A^{dec}$的可学习矩阵参数;$A^{dec}$为附有掩码注意力机制的行人位置编码,$A^{dec}\in R^{T_{pred}\times n\times d_{model}}$;

[0065] 用残差网络和标准化函数对行人特征表示进行处理;

[0066] $$A_{f4}=F_{norm}(V_{dgt}+F_{norm}(A^{dec}))$$

[0067] 式中,A_{f4}为经过残差网络和标准化函数处理的解码器掩码多头注意力,$A^{dec}\in R^{T_{pred}\times n\times d_{model}}$;

[0068] c2、子解码器的第二部分是多头注意力机制,公式如下:

[0069] $$q_i^{dec}=V_{dgt}*W_{q_i^{dec2}}^{dec2}$$

[0070] $$k_i^{dec}=V_{dgt}*W_{k_i^{dec2}}^{dec2}$$

[0071] $$v_i^{dec}=V_{dgt}*W_{v_i^{dec2}}^{dec2}$$

[0072] $$A_i^{dec2}(q_i^{dec2},k_i^{dec2},v_i^{dec2})=soft\max(\frac{q_i^{dec2}*k_i^{dec2}}{\sqrt{d_{model}}}v_i^{dec2})$$

[0073] $$A^{dec2}=F_{cat}(A_i^{dec2})*W_a^{dec2}$$

[0074] 式中,q_i^{dec2}、k_i^{dec2}、v_i^{dec2}表示子解码器第i个多头注意力机制的查询、键、值;这里的q_i^{dec2}由子解码器的掩码多头注意力模块的输出生成,而k_i^{dec2}与v_i^{dec2}由编码器的输出生成,*为矩阵运算,q_i^{dec2}、k_i^{dec2}、$v_i^{dec2}\in R^{T_{pred}\times n\times d_{model}}$;$W_{q_i^{dec2}}^{dec2}$、$W_{k_i^{dec2}}^{dec2}$、$W_{v_i^{dec2}}^{dec2}$为子解码器第i头多头注意力机制把$V_{dgt}$生成为$q_i^{dec2}$、$k_i^{dec2}$、$v_i^{dec2}$的可学习矩阵参数;$A_i^{dec2}$表示子解码器第i个多头注意力机制的注意力,$A_i^{dec2}\in R^{T_{pred}\times n\times d_{model}}$,$W_a^{dec2}$为把拼接起来的$A_i^{dec2}$变化成$A^{dec2}$的可学

习矩阵参数;A^{dec2}为结合历史轨迹数据和已预测轨迹数据的注意力,$A^{dec2} \in R^{T_{pred} \times n \times d_{model}}$;

[0075] 用残差网络和标准化函数对A^{dec2}进行处理,公式如下:

[0076] $A_{f5}=F_{norm}(A_{f4}+F_{drop}(A^{dec2}))$

[0077] 式中,A_{f5}为经过残差网络和标准化函数处理的子解码器注意力,$A_{f4} \in R^{T_{pred} \times n \times d_{model}}$;

[0078] c3、子解码器的第三部分是前向反馈网络,公式如下:

[0079] $$A_{f6} = F_{drop}(\tanh(\tanh(A_{f5} * W_{fda}^{1}) * W_{fda}^{2})$$

[0080] 式中,W_{fda}^{1}、W_{fda}^{2}为可学习矩阵参数,再用一遍残差网络及标准化函数对行人特征表示进行处理,公式如下:

[0081] $A_{f7}=F_{norm}(A_{f5}+F_{drop}(A_{f6}))$

[0082] 式中,A_{f7}为行人预测轨迹的特征,以上的步骤为一个子解码器的过程,而transformer的解码器部分为6个这样的子解码器串行拼接;当we=1,2,...,5时,这里的A_{f7}为下一层子解码器的输入,当we=6时,A_{f7}会作为整个transformer的解码器的输出,$A_{f7} \in R^{T_{pred} \times n \times d_{model}}$;

[0083] 以时间维度聚合每一个子解码器的输出A_{f7}得到$A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7}$,$A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7}$为每一个预测时间戳的$A_{f7}$聚合;用全连接层对$A_{T_{obs}+1:T_{obs}:T_{pred}}^{f7}$进行处理,生成高斯分布:

[0084] $$Tr = soft\max(A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7} * W_{gass})$$

[0085] 式中,Tr表示行人预测轨迹的高斯分布参数,$Tr \in R^{T_{pred} \times n \times d_{model}}$,$W_{gass}$为可学习变量;

[0086] d、全连接网络与双变量高斯分布

[0087] 全连接网络以transformer的解码器的输出Tr为输入,输出第i个时间戳的行人位置分布,这里的行人位置分布为双变量高斯分布,公式为:

[0088] $$P(Tr_{i,n}) = p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$$

[0089] 式中,$P(Tr_{i,n})$为第n个行人,第i个时间戳双变量高斯分布;p()为双变量高斯分布函数,$p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$为第i个时间戳,第n个行人的位置分布;$\mu_{i,n}^{x}$、$\mu_{i,n}^{y}$、$\sigma_{i,n}^{x}$、$\sigma_{i,n}^{y}$、$corr_{i,n}$分别表示第i个时间戳的第n个行人位置分布的x坐标的均值、y坐标的均值、x坐标的标准差、y坐标的标准差、x坐标与y坐标的相关性,$Tr_{i,n}$表示第i个时间戳第n个行人的高斯轨迹,所以第i个时间戳的一个行人位置分布需要五个参数,全连接层就是把transformer的解码器输出变成第i个时间戳的双变量(x,y)高斯分布。

[0090] 进一步,在步骤4)中,用损失函数把所得到的双变量高斯分布与未来轨迹数据$V_{T_{obs}+1:T_{pred}+T_{obs}}$做差值;

[0091] 损失函数L(W)为:

[0092] $$L(W) = -\sum_{n=1}^{n}\sum_{i=T_{obs}+1}^{T_{pred}} \log(p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n}))$$

[0093] 第n个行人的损失函数为L,W为预测模型参数;$p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$为第i个

时间戳,第n个行人的位置分布;$\mu_{i,n}^{x}$、$\mu_{i,n}^{y}$、$\sigma_{i,n}^{x}$、$\sigma_{i,n}^{y}$、$\mathrm{corr}_{i,n}$分别表示第i个时间戳的第n个行人位置分布的x坐标的均值、y坐标的均值、x坐标的标准差、y坐标的标准差、x坐标与y坐标的相关性,n_ped为样本中行人个数;

[0094] 使用时间反向传播算法和梯度优化方法ADAM训练预测模型,取最优预测模型;把行人的历史轨迹数据输入最优预测模型,就能生成行人预测轨迹。

[0095] 本发明与现有技术相比,具有如下优点与有益效果:

[0096] 1、使用了在自然语言处理中取得出色表现的transformer来提取行人轨迹的时序信息,相比于其它行人轨迹预测工作使用lstm或者lstm的变种来提取行人轨迹时序信息,transformer使用的注意力机制能更好地提取每个时间戳位置信息对未来轨迹的影响,能够比lstm模型更好地预测行人未来轨迹。

[0097] 2、使用了图卷积网络来考虑同一样本的行人集的关系,利用行人的速度向量来衡量行人间的关系,速度向量越相似,行人间的联系越大。想象一下如果两个人并排走,那这两个人有强烈的联系,而图卷积网络能很好地反映出这一点。

附图说明

[0098] 图1是本发明方法的框架图。

[0099] 图2是预测模型示意图。

具体实施方式

[0100] 下面结合实施例及附图对本发明作进一步详细的描述,但本发明的实施方式不限于此。

[0101] 参见图1和图2所示,本实施例提供了一种基于transformer与图卷积网络的行人轨迹预测方法,其具体情况如下:

[0102] 1)提取出若干个时间戳内所有的行人轨迹数据,包含x、y坐标的行人信息;前T_{obs}个时间戳为历史轨迹数据$V_{1:T_{obs}}=\{V_1,V_2,...,V_{T_{obs}}\}$,$T_{obs}=8$,$T_{obs}$为历史轨迹时间戳长度,$V_{1:T_{obs}}\in R^{T_{obs}\times n\times axis}$,R表示是属于实数域,n为行人个数,axis表示坐标维度,axis=8,$V_{1:T_{obs}}$简称为V;后T_{pred}个时间戳为未来轨迹数据$V_{T_{obs}+1:T_{obs}+T_{pred}}=\{V_{T_{obs}+1},V_{T_{obs}+2},...,V_{T_{pred}+T_{obs}}\}$,$T_{pred}$为预测轨迹时间戳长度,$T_{pred}=12$,$V_{T_{obs}+1:T_{pred}+T_{obs}}\in R^{T_{pred}\times n\times axis}$;对每一个样本的每个时间戳做一个行人关系图G;V、$V_{T_{obs}+1:T_{pred}+T_{obs}}$、G的集合为一个样本;以若干个样本为一个批进行并行处理;把总样本集分为训练集、验证集和测试集;将预测模型f()形式化为:

[0103] $$\widehat{V}_{T_{obs}+1:T_{pred}+T_{obs}}=f(V,\phi)$$

[0104] 式中,$\widehat{V}_{T_{obs}+1:T_{pred}+T_{obs}}$是预测轨迹数据,φ是预测模型f()中可学习的参数;

[0105] Vp_i^j表示在第i秒内第j个行人的坐标,每个样本至少有两条行人轨迹;

[0106] $$V=Vp_i^j,i\le T_{obs}$$

[0107] $$V_{T_{obs}+1:T_{obs}+T_{pred}}=Vp_i^j,T_{obs}<i\le T_{obs}+T_{pred}$$

[0108] 把每个样本分为历史轨迹数据V与未来轨迹数据$V_{T_{obs}+1:T_{obs}+T_{pred}}$；

[0109] $G_{i,j}^{t}=1/\|S_i^t+S_j^t\|_2$

[0110] $S_i^t=Xrel_i^t+Yrel_i^t$

[0111] 式中,G为行人关系图,$G\in R^{T_{obs}\times n\times n}$， S_i^t、 S_j^t、 $Xrel_i^t$和$Yrel_i^t$为在一个样本中第i个行人在第t个时间戳的合速度向量、横坐标分量速度向量和纵坐标分量速度向量;$G_{i,j}^{t}$表示第t个时间戳中,第i个行人与第j个行人的相互关系;

[0112] 一个批次中包含若干个样本,便于预测模型的并行运行,再把若干个批次分为训练集、验证集与测试集,分别用来训练预测模型、取最优预测模型、测试预测模型。

[0113] 2)先用全连接网络对V进行坐标编码,提取V的坐标特征表示V_{emb},$V_{emb}\in R^{T_{obs}\times n\times d_{model}}$，编码空间维度大小为$d_{model}$;$d_{model}$=64,再用行人关系图G对$V_{emb}$进行图卷积学习,提取附有行人关系信息的行人坐标编码V_g,具体步骤如下:

[0114] 2.1)首先用3层全连接层对x、y坐标做编码操作,公式如下:

[0115] $V_{emb}^{i,t}=V_{emb}^{i-1,t}*W_f^i$

[0116] 式中,$V_{emb}^{i,t}$表示在全连接层第i层,第t个时间戳的行人集坐标编码;$V_{emb}^{i-1,t}$表示在全连接层第i-1层,第t个时间戳的行人集坐标编码;W_f^i表示在第i层的全连接层可学习矩阵参数,*表示矩阵乘法;第一层全连接层把x、y坐标维度axis维扩展为d_{model}维,当i=2,3时,第i层全连接层的输入坐标编码维度和输出坐标编码维度都维持d_{model}维;

[0117] 2.2)采用图卷积神经网络利用行人关系图,把行人集坐标编码进行空间卷积运算;邻接性的归一化是图卷积正常工作的必要条件,先对每一个时间戳的行人关系图进行标准化处理,行人关系图$G=\{G_1,G_2,...,G_t,...,G_{T_{obs}}\}$,$G_t$表示第t个时间戳的行人关系图,公式如下:

[0118] $G_t^{'}=D_t^{-1/2}(G_t+I)D_t^{-1/2}$

[0119] 式中,G_t'为第t个时间戳的标准化行人关系图,I为单位矩阵,D_t为对角节点度矩阵;

[0120] 用n_{gcn}层图卷积层赋予轨迹信息以行人关系信息,公式如下:

[0121] $F_{gcn}(V_{emb}^{i},G^{'})=\tanh(G^{'}*V_{emb}^{i}*W_g^i)$

[0122] 式中,$V_{emb}^{i}=\{V_{emb}^{i,1},V_{emb}^{i,2},...,V_{emb}^{i,t},...,V_{emb}^{i,T_{obs}}\}$， $V_{emb}^{i}\in R^{T_{obs}\times n\times d_{model}}$表示在图卷积层第i层的行人集坐标编码,$V_{emb}^{i,t}$示在图卷积层第i层、第t个时刻的行人集坐标编码,$F_{gcn}$表示图卷积操作,$W_g^i$表示表示在第i层的图卷积层可学习矩阵参数;*表示矩阵乘法;tanh为图卷积层的激活函数;最后一层图卷积完成的输出为附有行人关系信息的行人坐标编码V_g,$V_g\in R^{T_{obs}\times n\times d_{model}}$。

[0123] 3)采用transformer的编码器将附有行人关系信息的行人坐标编码V_g提取出每个时间戳的时序特征向量,用transformer的解码器以每个时间戳的时序特征向量为输入来生成具体行人轨迹分布,该行人轨迹分布遵循双变量高斯分布,具体如下:

[0124] 把transformer作为预测模型基底,进行轨迹预测。用transformer以图卷积的输

出为输入,考虑历史轨迹数据V对预测第i个时间戳的行人位置分布的影响;

[0125] a、位置编码,公式如下:

[0126] $$\tau_n^t(k)=\begin{cases}\sin(10000^{-k/d_{model}}), k \bmod \equiv 0\\ \cos(10000^{-k/d_{model}}), k \bmod \equiv 1\end{cases}$$

[0127] $V_{gt}=V_g+\tau$

[0128] 式中,$\tau_n^t(k)$为第t个时间戳,第n个行人的第k个特征值,τ为行人位置编码;把行人位置编码与V_g相加得到附有行人位置编码的V_{gt};

[0129] b、编码器:

[0130] 编码器以经过位置编码处理的V_{gt}为输入,经过注意力机制,提取特征A_{f3};Transformer的编码器有6层子编码器,每一层的子编码器结构相同,当u等于1时,下面的子编码器代表第u个子编码器;当u=2,3,…,6重复以上操作;

[0131] b1、子编码器的第一部分是多头注意力机制,公式如下:

[0132] $$q_i=V_{gt}*W_{q_i}$$

[0133] $$k_i=V_{gt}*W_{k_i}$$

[0134] $$v_i=V_{gt}*W_{v_i}$$

[0135] $$A_i(q_i,k_i,v_i)=soft\max(\frac{q_i*k_i}{\sqrt{d_{model}}}v_i)$$

[0136] $A=F_{cat}(A_i)*W_a$

[0137] 式中,q_i、k_i、v_i表示子编码器第i个头的查询、键、值;*为矩阵运算,q_i、k_i、$v_i\in R^{T_{obs}\times n\times d_{model}}$;$W_{q_i}$、$W_{k_i}$、$W_{v_i}$为子编码器的第i头注意力把$V_{gt}$生成$q_i$、$k_i$、$v_i$的可学习矩阵参数;$A_i$表示子编码器第i个头注意力机制的注意力,$A_i\in R^{T_{obs}\times n\times d_{model}}$,$F_{cat}$函数的作用是把多个头的注意力拼接起来,$W_a$为把拼接起来的$A_i$变化成A的可学习矩阵参数;A为附有注意力机制的行人位置编码,$A\in R^{T_{obs}\times n\times d_{model}}$;

[0138] 用残差网络和标准化函数对行人特征表示进行处理,公式如下:

[0139] $A_{f1}=F_{norm}(V_{gt}+F_{drop}(A))$

[0140] 式中,A_{f1}为经过残差网络和标准化函数处理的注意力,$A_{f1}\in R^{T_{obs}\times n\times d_{model}}$;$F_{norm}()$为标准化函数;

[0141] b2、子编码器的第二部分是前向反馈网络,公式如下:

[0142] $$A_{f2}=F_{drop}(\tanh(\tanh(A_{f1}*W_{fa}^1)*W_{fa}^2))$$

[0143] 式中,A_{f2}为经过前向反馈网络的子编码器注意力,$A_{f2}\in R^{T_{obs}\times n\times d_{model}}$;$F_{drop}$为随机参数不更新函数;tanh为激活函数,$W_{fa}^1$、$W_{fa}^2$为第一、二层全连接层可学习矩阵参数;再用一遍残差网络及标准化函数对行人特征表示进行处理,公式如下:$A_{f3}=F_{norm}(A_{f1}+F_{drop}(A_{f2}))$,$A_{f3}$为经过残差网络和标准化函数处理的注意力;

[0144] 以上的步骤为一个子编码器的过程,而transformer的编码器部分为6个这样的子编码器串行拼接;当u=1,2,...,5时,A_{f3}为下一层子编码器的输入,当u=6时,A_{f3}会作为整

个transformer的编码器的输出，$A_{f3} \in R^{T_{obs} \times n \times d_{model}}$；

[0145] c、解码器：

[0146] 解码器的工作目标是预测第t个时间戳的行人位置$\hat{V}_t$，解码器以编码器的输出A_{f3}和经过了位置编码的已预测时间戳的行人位置$\hat{V}^{gt}_{T_{obs}+1:t}$为输入，输出为$\hat{V}_t$；

[0147] 与编码器一样，Transformer的解码器有6层子解码器，每一层的子解码器结构相同，当we等于1时，下面的子解码器代表第we个子解码器，当we＝u＝2,3,…,6重复以上操作；

[0148] c1、子解码器的第一部分是掩码多头注意力机制：

[0149] 因为不能泄露行人未来轨迹信息，这里进行掩码操作，公式如下：

[0150] $$V_t^{dgt} = \begin{cases} \hat{V}_t^{gt}, t > T_{obs} + y \\ 0, otherwise \end{cases}$$

[0151] 式中，$\hat{V}_t^{gt}$为第t个时间戳的经过了位置编码的已预测时间戳的行人位置，V_t^{dgt}表示子解码器输入在第t个时间戳的行人轨迹信息编码，y为已预测未来轨迹步长，$V^{dgt}_{T_{obs}+1:T_{obs}+T_{red}} = \{V^{dgt}_{T_{obs}+1}, V^{dgt}_{T_{obs}+2}, ..., V^{dgt}_{t}, ..., V^{dgt}_{T_{obs}+T_{pred}}\}$是在解码器中行人轨迹信息编码集合，这里把$V^{dgt}_{T_{obs}+1:T_{obs}+T_{red}}$简称为$V_{dgt}$；

[0152] 注意力操作公式如下：

[0153] $$q_i^{dec} = V_{dgt} * W^{dec}_{q_i^{dec}}$$

[0154] $$k_i^{dec} = V_{dgt} * W^{dec}_{k_i^{dec}}$$

[0155] $$v_i^{dec} = V_{dgt} * W^{dec}_{v_i^{dec}}$$

[0156] $$A_i^{dec}(q_i^{dec}, k_i^{dec}, v_i^{dec}) = soft\max(\frac{q_i^{dec} * k_i^{dec}}{\sqrt{d_{model}}} v_i^{dec})$$

[0157] $$A^{dec} = F_{cat}(A_i^{dec}) * W_a^{dec}$$

[0158] 式中，q_i^{dec}、k_i^{dec}、v_i^{dec}表示子解码器第i个掩码多头注意力机制的查询、键、值，q_i^{dec}、k_i^{dec}、$v_i^{dec} \in R^{T_{pred} \times n \times d_{model}}$；$V_{dgt}$为在编码器中的行人轨迹信息编码；$W^{dec}_{q_i^{dec}}$、$W^{dec}_{k_i^{dec}}$、$W^{dec}_{v_i^{dec}}$为子解码器第i个掩码多头注意力机制把$V_{dgt}$生成为$q_i^{dec}$、$k_i^{dec}$、$v_i^{dec}$的可学习矩阵参数；$A_i^{dec}$表示子解码器第i个掩码多头注意力机制的注意力，$A_i^{dec} \in R^{T_{pred} \times n \times d_{model}}$，$W_a^{dec}$为把拼接起来的$A_i^{dec}$变化成$A^{dec}$的可学习矩阵参数；$A^{dec}$为附有掩码注意力机制的行人位置编码，$A^{dec} \in R^{T_{pred} \times n \times d_{model}}$；

[0159] 用残差网络和标准化函数对行人特征表示进行处理；

[0160] $A_{f4} = F_{norm}(V_{dgt} + F_{norm}(A^{dec}))$

[0161] 式中，A_{f4}为经过残差网络和标准化函数处理的解码器掩码多头注意力，$A^{dec} \in R^{T_{pred} \times n \times d_{model}}$；

[0162] c2、子解码器的第二部分是多头注意力机制，公式如下：

[0163] $q_i^{dec}=V_{dgt}*W_{q_i^{dec2}}^{dec2}$

[0164] $k_i^{dec}=V_{dgt}*W_{k_i^{dec2}}^{dec2}$

[0165] $v_i^{dec}=V_{dgt}*W_{v_i^{dec2}}^{dec2}$

[0166] $A_i^{dec2}(q_i^{dec2},k_i^{dec2},v_i^{dec2})=soft\max(\frac{q_i^{dec2}*k_i^{dec2}}{\sqrt{d_{model}}}v_i^{dec2})$

[0167] $A^{dec2}=F_{cat}(A_i^{dec2})*W_a^{dec2}$

[0168] 式中，q_i^{dec2}、k_i^{dec2}、v_i^{dec2}表示子解码器第i个多头注意力机制的查询、键、值；这里的q_i^{dec2}由子解码器的掩码多头注意力模块的输出生成，而k_i^{dec2}与v_i^{dec2}由编码器的输出生成，*为矩阵运算，q_i^{dec2}、k_i^{dec2}、$v_i^{dec2}\in R^{T_{pred}\times n\times d_{model}}$；$W_{q_i^{dec2}}^{dec2}$、$W_{k_i^{dec2}}^{dec2}$、$W_{v_i^{dec2}}^{dec2}$为子解码器第i头多头注意力机制把$V_{dgt}$生成为$q_i^{dec2}$、$k_i^{dec2}$、$v_i^{dec2}$的可学习矩阵参数；$A_i^{dec2}$表示子解码器第i个多头注意力机制的注意力，$A_i^{dec2}\in R^{T_{pred}\times n\times d_{model}}$，$W_a^{dec2}$为把拼接起来的$A_i^{dec2}$变化成$A^{dec2}$的可学习矩阵参数；$A^{dec2}$为结合历史轨迹数据和已预测轨迹数据的注意力，$A^{dec2}\in R^{T_{pred}\times n\times d_{model}}$；

[0169] 用残差网络和标准化函数对A^{dec2}进行处理，公式如下：

[0170] $A_{f5}=F_{norm}(A_{f4}+F_{drop}(A^{dec2}))$

[0171] 式中，A_{f5}为经过残差网络和标准化函数处理的子解码器注意力，$A_{f4}\in R^{T_{pred}\times n\times d_{model}}$；

[0172] c3、子解码器的第三部分是前向反馈网络，公式如下：

[0173] $A_{f6}=F_{drop}(\tanh(\tanh(A_{f5}*W_{fda}^1)*W_{fda}^2)$

[0174] 式中，W_{fda}^1、W_{fda}^2为可学习矩阵参数，再用一遍残差网络及标准化函数对行人特征表示进行处理，公式如下：

[0175] $A_{f7}=F_{norm}(A_{f5}+F_{drop}(A_{f6}))$

[0176] 式中，A_{f7}为行人预测轨迹的特征，以上的步骤为一个子解码器的过程，而transformer的解码器部分为6个这样的子解码器串行拼接；当we＝1,2,...,5时，这里的A_{f7}为下一层子解码器的输入，当we＝6时，A_{f7}会作为整个transformer的解码器的输出，$A_{f7}\in R^{T_{pred}\times n\times d_{model}}$；

[0177] 以时间维度聚合每一个子解码器的输出A_{f7}得到$A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7}$，$A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7}$为每一个预测时间戳的$A_{f7}$聚合；用全连接层对$A_{T_{obs}+1:T_{obs}:T_{pred}}^{f7}$进行处理，生成高斯分布：

[0178] $Tr=soft\max(A_{T_{obs}+1:T_{obs}+T_{pred}}^{f7}*W_{gass})$

[0179] 式中，Tr表示行人预测轨迹的高斯分布参数，$Tr\in R^{T_{pred}\times n\times d_{model}}$，$W_{gass}$为可学习变量；

[0180] d、全连接网络与双变量高斯分布

[0181] 全连接网络以transformer的解码器的输出Tr为输入，输出第i个时间戳的行人位置分布，这里的行人位置分布为双变量高斯分布，公式为：

[0182] $$P(Tr_{i,n}) = p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$$

[0183] 式中,P($Tr_{i,n}$)为第n个行人,第i个时间戳双变量高斯分布;p()为双变量高斯分布函数,$p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$为第i个时间戳,第n个行人的位置分布;$\mu_{i,n}^{x}$、$\mu_{i,n}^{y}$、$\sigma_{i,n}^{x}$、$\sigma_{i,n}^{y}$、$corr_{i,n}$分别表示第i个时间戳的第n个行人位置分布的x坐标的均值、y坐标的均值、x坐标的标准差、y坐标的标准差、x坐标与y坐标的相关性,$Tr_{i,n}$表示第i个时间戳第n个行人的高斯轨迹,所以第i个时间戳的一个行人位置分布需要五个参数,全连接层就是把transformer的解码器输出变成第i个时间戳的双变量(x,y)高斯分布。

[0184] 4)采用损失函数把预测轨迹数据与未来轨迹数据作对比生成损失值,再用反向传播损失值优化预测模型;在优化预测模型时,用训练集对预测模型进行训练,用验证集挑选最优预测模型,把测试集输入到最优预测模型,得出预测轨迹数据。

[0185] 其中,用损失函数把所得到的双变量高斯分布与未来轨迹数据$V_{T_{obs}+1:T_{pred}+T_{obs}}$做差值,损失函数L(W)为:

[0186] $$L(W) = -\sum_{n=1}^{n\_ped}\sum_{i=T_{obs}+1}^{T_{pred}} \log(p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n}))$$

[0187] 第n个行人的损失函数为L,W为预测模型参数;$p(\mu_{i,n}^{x}, \mu_{i,n}^{y}, \sigma_{i,n}^{x}, \sigma_{i,n}^{y}, corr_{i,n})$为第i个时间戳,第n个行人的位置分布;$\mu_{i,n}^{x}$、$\mu_{i,n}^{y}$、$\sigma_{i,n}^{x}$、$\sigma_{i,n}^{y}$、$corr_{i,n}$分别表示第i个时间戳的第n个行人位置分布的x坐标的均值、y坐标的均值、x坐标的标准差、y坐标的标准差、x坐标与y坐标的相关性,n_ped为样本中行人个数;

[0188] 使用时间反向传播算法和梯度优化方法ADAM训练预测模型,取最优预测模型;把行人的历史轨迹数据输入最优预测模型,就能生成行人预测轨迹。

[0189] 上述实施例为本发明较佳的实施方式,但本发明的实施方式并不受上述实施例的限制,其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化,均应为等效的置换方式,都包含在本发明的保护范围之内。

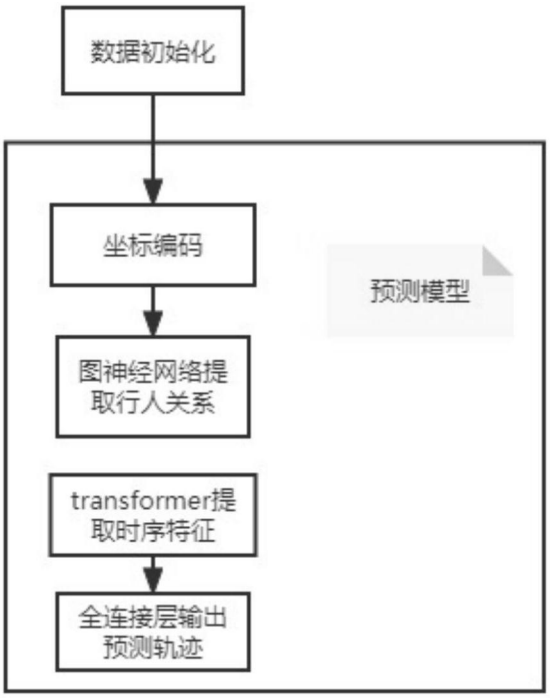

数据初始化
坐标编码
预测模型
图神经网络提取行人关系
transformer提取时序特征
全连接层输出预测轨迹

图1

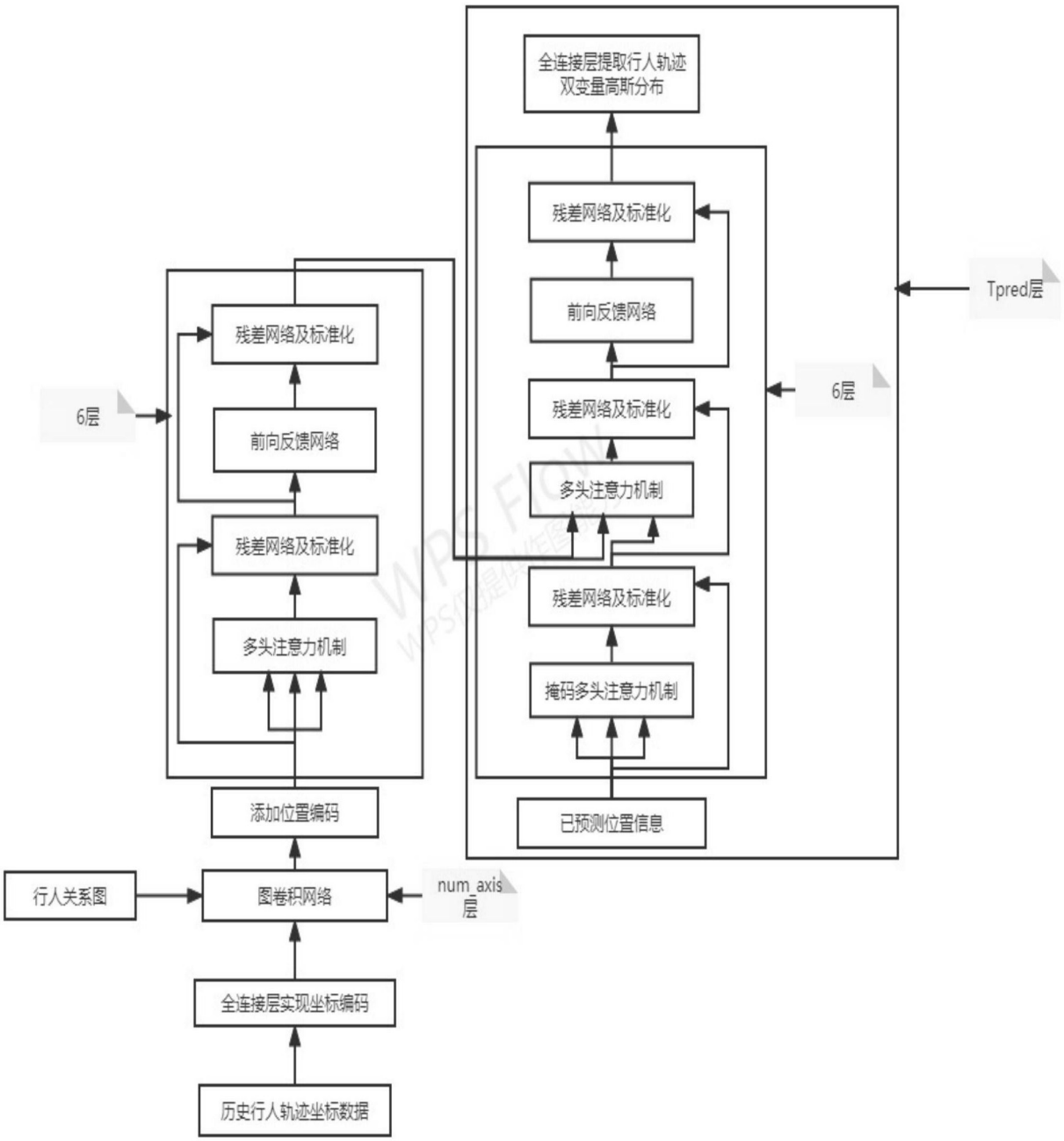

全连接层提取行人轨迹
双变量高斯分布
残差网络及标准化
前向反馈网络
残差网络及标准化
多头注意力机制
残差网络及标准化
掩码多头注意力机制
已预测位置信息
Tpred层
6层
残差网络及标准化
前向反馈网络
残差网络及标准化
多头注意力机制
6层
添加位置编码
行人关系图
图卷积网络
num_axis
层
全连接层实现坐标编码
历史行人轨迹坐标数据

图2