



(12) 发明专利申请

(10) 申请公布号 CN 113238197 A

(43) 申请公布日 2021.08.10

(21) 申请号 202011600963.7

(22) 申请日 2020.12.29

(71) 申请人 杭州电子科技大学

地址 310018 浙江省杭州市下沙高教园区
二号路

(72) 发明人 吕帅帅 潘勉 唐三鼎 李训根
陈晴 方笑海 张杰

(74) 专利代理机构 浙江永鼎律师事务所 33233
代理人 陆永强

(51) Int.Cl.

G01S 7/41 (2006.01)

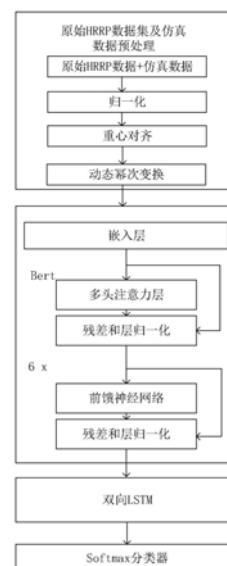
权利要求书3页 说明书7页 附图1页

(54) 发明名称

基于Bert和BiLSTM的雷达目标识别及据判方法

(57) 摘要

本发明公开了一种基于Bert和BiLSTM的雷达目标识别及据判方法,包括以下步骤:S1,收集样本并划分训练集和测试集,进行数据预处理,包括强调归一化和重心对齐;S2,使用带注意力机制的幂次变换;S3,用Bert处理幂次变换的输出,提取更加深层的特征;S4,通过双向LSTM,加强特征提取;S5,使用softmax分类器分类;S6,将经过S1处理后的HRRP测试集送入S2、S3、S4、S5步骤中已训练完成的模型中进行测试。



1. 一种基于Bert和BiLSTM的雷达目标识别及据判方法,其特征在于,包括以下步骤:

S1,收集样本并划分训练集和测试集,通过雷达采集的HRRP数据,总共c类,对每个类别进行采样,分别选出训练集和测试集,之后进行合并,保证每个类别的数据形式都有被采样到,训练集和测试集比例保持在7:3,并且将仿真数据作为负样本加入到训练集中,作为第c+1类,以提高模型的据判能力,将训练数据集记为 $D = \{(x_i, y_k)\}_{i \in [1, n], k \in [1, c]}$,其中 x_i 表示第i个样本, y_k 表示样本属于第k类,一共采集了c类目标,n表示样本总数;进行数据预处理,包括强调归一化和重心对齐;

S2,使用带注意力机制的幂次变换;

S3,用Bert处理幂次变换的输出,提取更加深层的特征;

S4,通过双向LSTM,加强特征提取;

S5,使用softmax分类器分类;

S6,将经过S1处理后的HRRP测试集送入S2、S3、S4、S5步骤中已训练完成的模型中进行测试。

2. 如权利要求1所述的基于Bert和BiLSTM的雷达目标识别及据判方法,其特征在于,S1中,进行数据预处理进一步包括:

S101,强度归一化,假设原始HRRP数据表示为 $x_{\text{raw}} = [x_1, x_2, x_3 \dots, x_M]$,其中M表示HRRP数据中距离单元总数,那么归一化之后的HRRP表示为:

$$x_{\text{normalization}} = \frac{x_{\text{raw}}}{\sqrt{\sum_{m=1}^M x_m^2}}$$

S102,重心对齐,为了使训练和测试具备统一标准,采用重心对齐法消除平移敏感性,重心对齐法通过平移HRRP使其重心位于HRRP中心点附近,重心g的计算方法如下:

$$g_{\text{HRRP}} = \frac{\sum_{m=1}^M m x_m}{\sum_{m=1}^M x_m}。$$

3. 如权利要求1所述的基于Bert和BiLSTM的雷达目标识别及据判方法,其特征在于,S2,使用带注意力机制的幂次变换,具体步骤如下:

对HRRP数据进行幂次变换,并使用注意力机制,对经过上述预处理的HRRP数据进行多次幂次变换,幂次变换层输出表示为:

$$\text{Output} = \alpha_i \sum_{k=0}^K X^k$$

其中 α_i 为每种幂次变换的权重,k表示幂次,通过学习,学出不同的权重,使Bert模块可以得到更合适的输入,其中每一条样本最终会呈现如下形式:

$$\mathbf{x} = \begin{bmatrix} x_1^{0.1} & x_2^{0.1} & \dots & x_M^{0.1} \\ \vdots & \vdots & & \vdots \\ x_1^{k-1} & x_2^{k-1} & \dots & x_M^{k-1} \\ x_1^k & x_2^k & \dots & x_M^k \end{bmatrix}。$$

4. 如权利要求1所述的基于Bert和BiLSTM的雷达目标识别及据判方法,其特征在于,S3,用Bert处理幂次变换的输出,提取更加深层的特征,具体步骤如下:

S301,确定使用Bert的层数,构建Bert模型;

S302,假设输入为X的形状为[batch_size,N],batch_size为每个批次的数目,N为各时间点维数,先对数据进行嵌入式操作,并且为它加入时序信息,即位置信息,Bert中按照特征长度,会随机初始化同样形状的数据,作为位置信息,在训练中进行学习;

S303,多头注意力机制,提取有效特征,多头注意力机制将通过嵌入层中得到的M维特征,分成12组,在每一组中进行self-attention,在将12组的结果进行合并,X中每一组分别乘以三个随机初始化的矩阵 W_q, W_k, W_v 得到Q,K,V三个矩阵,进行如下计算得到 Z_i :

$$Z_i = \text{soft max}\left(\frac{Q_i \times K_i^T}{\sqrt{d_k}}\right)$$

Z_i 表示每一组的计算结果。再将 $[Z_0 \dots Z_{11}]$ 按列堆叠起来,乘以一个随机初始化的W,即可得到最终此层的输出Z;

S304,前馈层中,将N维特征,映射到高维,再从高维映射到低维,进一步保留有用的信息;

S305,残差和层归一化,残差层就是残差连接模块,因为要叠加深层神经网络,残差结构可以很好地消除层数加深所带来的信息损失问题,也可以将损失快速回传,避免梯度消失;与CNN中批归一化一样,在这里加入了层归一化,使训练过程更加稳定:

$$LN(x_i) = \alpha \times \frac{x_i - \mu_L}{\sqrt{\sigma_L^2 + \varepsilon}} + \beta$$

式中 $x \in \mathbb{R}^{N \times C \times H \times W}$,层归一化对每个样本C,H,W维度上的数据求均值和标准差,保留N维度, μ 为均值, σ 为方差, α 和 β 为参数, ε 为一个非常小的值;

S306,将S302到S305组在一起为一层Bert,总共堆叠6层。

5.如权利要求1所述的基于Bert和BiLSTM的雷达目标识别及据判方法,其特征在于,S4,通过双向LSTM,加强特征提取,具体步骤如下:

S401,Bert模块之后,输出的形状为[batch_size,M,N],为了加强特征提取能力和据判能力,应用了双向LSTM,在Bert中仅仅用位置嵌入式操作来告诉模型输入向量token的位置信息,但是在序列任务中位置信息和方向信息都很必要,使用双向LSTM习得观测序列的依赖关系,

假设输入是特征 $F_{LSTM} = [F_{LSTM}^1, F_{LSTM}^2, \dots, F_{LSTM}^M] \in \mathbb{R}^{M \times N_i}$,其中i表示第i个BiLSTM的各时间点维数,M表示输入序列长度,假设输出为 $F_{output} = [F_{output}^1, F_{output}^2, \dots, F_{output}^M] \in \mathbb{R}^{M \times H}$,其中H为BiLSTM里的隐单元节点数。

6.如权利要求1所述的基于Bert和BiLSTM的雷达目标识别及据判方法,其特征在于,S5进一步包括:

S501,通过softmax层对特征 F_{output} 进行分类,若训练集中所包含的目标总数为C,测试HRRP样本 X_{test} 对应于目标集中第i类目标的概率表示为:

$$P(i|X_{test}) = \frac{\exp(F_s(i))}{\sum_{c=1}^C \exp(F_s(c))}$$

其中, $\exp(\cdot)$ 表示取指数运算, $F_s(i)$ 指的是向量 F_s 中的第i个元素, $F_s = W_s F_{ATT}$, W_s 是向量 F_s 的权重矩阵,通过最大后验概率将测试HRRP样本 X_{test} 分类到最大目标概率 c_0 中:

$$c_0 = \arg \max_i P(i | \mathbf{x}_{\text{test}})$$

S502,在所提模型的参数训练过程中,采用交叉熵作为代价函数,数据中,我们有三类实测飞机,和一类仿真数据,作为库外目标一起训练,每个训练样本 $\mathbf{x}_{\text{train}}$ 的代价可以表示为:

$$L = -\sum_{i=1}^C z(i) \log P(i | \mathbf{x}_{\text{train}}) - 2 \sum_{i=1}^{\bar{C}} z(i) \log P(i | \mathbf{x}_{\text{train}})$$

其中,C为库内互相分错的目标, \bar{C} 为将三类实测目标分到第四类或者将第四类分为前三类的目标,在这个loss function中使用了比较简单的方法,给予式中第二部分两倍的损失, $z(i)$ 表示训练样本 $\mathbf{x}_{\text{train}}$ 的真实标签 z 中的第 i 个元素, z 是一个one-hot vector, $P(i | \mathbf{x}_{\text{train}})$ 表示训练样本 $\mathbf{x}_{\text{train}}$ 对应于第 i 个目标的概率。

基于Bert和BiLSTM的雷达目标识别及据判方法

技术领域

[0001] 本发明属于雷达目标识别技术领域,具体涉及一种基于Bert和BiLSTM的雷达目标识别及据判方法。

背景技术

[0002] 雷达目标识别的技术,在战争和生活中的应用逐渐广泛。它是利用雷达和计算机对目标进行识别分类的技术,通过对雷达回波中的幅度、相位、频谱和极化等目标特征信息的分析,使用模型计算目标的大小、形状、重量和表面层的物理特性参数,最后依据大量训练数据确定模型,在分类器中进行识别分类。

[0003] 目标识别的关键步骤是特征提取,随着深度学习的发展,目标识别的中传统的特征提取方法逐渐被深度学习方法所替代。基于监督学习的方法可以自动的提取训练数据的可分性特征。

[0004] 在雷达目标识别中,要求模型不仅需要可靠的识别能力,也要求模型应该具有可靠的据判能力。在传统HRRP自动目标识别系统的研究背景中,待识别目标通常具有完备的训练数据库,并且该数据库涵盖了测试HRRP出现的所有可能性。然而在实际战场环境下,识别对象通常是非合作目标和敌对目标,这些目标所能构建的目标库往往不完备,有可能接收到的HRRP来自于之前从未观测到的目标。并且,雷达收到的HRRP回波有可能源于敌方的无源干扰和有源干扰,当有别于构建的目标库内所有目标的新目标的测试样本出现时,将其简单粗暴地直接判别为目标库内的任一类样本是不正确的,有可能会造成判断决策失误,带来灾难性的严重后果。此时最合理地做法是将该库外目标的测试样本拒判,以备后续处理。许多传统方法在解决库外目标拒判问题时,通常直接将这一问题作为识别模型的延伸,直接测试拒判能力。然而拒判和识别所对应的优化目标不一致,一个识别能力很强的模型其拒判效果并不一定好,反之亦然。在雷达HRRP目标识别领域,库外目标拒判能力是评价分类器性能的一项重要指标。

发明内容

[0005] 鉴于以上存在的技术问题,本发明用于提供一种基于Bert和BiLSTM的雷达目标识别及据判方法,提出了一种基于Bert和BiLSTM的雷达HRRP目标识别及据判框架。先对数据进行预处理,降低敏感性,通过幂次变化调整HRRP数据中蕴含的多种形式的特征,之后再通过Bert模块对数据进行深层特征的提取,最后通过双向LSTM,并且使用softmax进行目标识别分类。

[0006] 为解决上述技术问题,本发明采用如下的技术方案:

[0007] 一种基于Bert和BiLSTM的雷达目标识别及据判方法,包括以下步骤:

[0008] S1,收集样本并划分训练集和测试集,通过雷达采集的HRRP数据,总共c类,对每个类别进行采样,分别选出训练集和测试集,之后进行合并,保证每个类别的数据形式都有被采样到,训练集和测试集比例保持在7:3,并且将仿真数据作为负样本加入到训练集中,作

为第 $c+1$ 类,以提高模型的据判能力,将训练数据集记为 $D = \{(x_i, y_k)\}_{i \in [1, n], k \in [1, c]}$,其中 x_i 表示第 i 个样本, y_k 表示样本属于第 k 类,一共采集了 c 类目标, n 表示样本总数;进行数据预处理,包括强调归一化和重心对齐;

[0009] S2,使用带注意力机制的幂次变换;

[0010] S3,用Bert处理幂次变换的输出,提取更加深层的特征;

[0011] S4,通过双向LSTM,加强特征提取;

[0012] S5,使用softmax分类器分类;

[0013] S6,将经过S1处理后的HRRP测试集送入S2、S3、S4、S5步骤中已训练完成的模型中进行测试。

[0014] 优选地,S1中,进行数据预处理进一步包括:

[0015] S101,强度归一化,假设原始HRRP数据表示为 $x_{\text{raw}} = [x_1, x_2, x_3, \dots, x_M]$,其中 M 表示HRRP数据中距离单元总数,那么归一化之后的HRRP表示为:

$$[0016] \quad x_{\text{normalization}} = \frac{x_{\text{raw}}}{\sqrt{\sum_{m=1}^M x_m^2}}$$

[0017] S102,重心对齐,为了使训练和测试具备统一标准,采用重心对齐法消除平移敏感性,重心对齐法通过平移HRRP使其重心位于HRRP中心点附近,重心 g 的计算方法如下:

$$[0018] \quad g_{\text{HRRP}} = \frac{\sum_{m=1}^M m x_m}{\sum_{m=1}^M x_m}。$$

[0019] 优选地,S2,使用带注意力机制的幂次变换,具体步骤如下:

[0020] 对HRRP数据进行幂次变换,并使用注意力机制,对经过上述预处理的HRRP数据进行多次幂次变换,幂次变换层输出表示为:

$$[0021] \quad \text{Output} = \alpha_i \sum_{k=0}^K X^k$$

[0022] 其中 α_i 为每种幂次换的权重, k 表示幂次,通过学习,学出不同的权重,使Bert模块可以得到更合适的输入,其中每一条样本最终会呈现如下形式:

$$[0023] \quad \mathbf{X} = \begin{bmatrix} x_1^{0.1} & x_2^{0.1} & \dots & x_M^{0.1} \\ \vdots & \vdots & & \vdots \\ x_1^{k-1} & x_2^{k-1} & \dots & x_M^{k-1} \\ x_1^k & x_2^k & \dots & x_M^k \end{bmatrix}。$$

[0024] 优选地,S3,用Bert处理幂次变换的输出,提取更加深层的特征,具体步骤如下:

[0025] S301,确定使用Bert的层数,构建Bert模型;

[0026] S302,假设输入为 X 的形状为 $[\text{batch_size}, N]$, batch_size 为每个批次的数目, N 为各时间点维数,先对数据进行嵌入式操作,并且为它加入时序信息,即位置信息,Bert中按照特征长度,会随机初始化同样形状的数据,作为位置信息,在训练中进行学习;

[0027] S303,多头注意力机制,提取有效特征,多头注意力机制将通过嵌入层中得到的 M 维特征,分成12组,在每一组中进行self-attention,在将12组的结果进行合并, X 中每一组分别乘以三个随机初始化的矩阵 W_q, W_k, W_v 得到 Q, K, V 三个矩阵,进行如下计算得到 Z_1 :

$$[0028] \quad Z_i = \text{soft max}(\frac{Q_i \times K_i^T}{\sqrt{d_k}})$$

[0029] Z_i 表示每一组的计算结果。再将 $[Z_0 \dots Z_{11}]$ 按列堆叠起来,乘以一个随机初始化的 W ,即可得到最终此层的输出 Z ;

[0030] S304,前馈层中,将 N 维特征,映射到高维,再从高维映射到低维,进一步保留有用的信息;

[0031] S305,残差和层归一化,残差层就是残差连接模块,因为要叠加深层神经网络,残差结构可以很好地消除层数加深所带来的信息损失问题,也可以将损失快速回传,避免梯度消失;与CNN中批归一化一样,在这里加入了层归一化,使训练过程更加稳定:

$$[0032] \quad LN(x_i) = \alpha \times \frac{x_i - \mu_L}{\sqrt{\hat{\sigma}_L^2 + \varepsilon}} + \beta$$

[0033] 式中 $x \in \mathbb{R}^{N \times C \times H \times W}$,层归一化对每个样本 C, H, W 维度上的数据求均值和标准差,保留 N 维度, μ 为均值, $\hat{\sigma}$ 为方差, α 和 β 为参数, ε 为一个非常小的值;

[0034] S306,将S302到S305组在一起为一层Bert,总共堆叠6层。

[0035] 优选地,S4,通过双向LSTM,加强特征提取,具体步骤如下:

[0036] S401,Bert模块之后,输出的形状为 $[\text{batch_size}, M, N]$,为了加强特征提取能力和据判能力,应用了双向LSTM,在Bert中仅仅用位置嵌入式操作来告诉模型输入向量token的位置信息,但是在序列任务中位置信息和方向信息都很必要,使用双向LSTM习得观测序列的依赖关系,

[0037] 假设输入是特征 $F_{LSTM} = [F_{LSTM}^1, F_{LSTM}^2, \dots, F_{LSTM}^M] \in \mathbb{R}^{M \times N_i}$,其中 i 表示第 i 个BiLSTM的各时间点维数, M 表示输入序列长度,假设输出为 $F_{output} = [F_{output}^1, F_{output}^2, \dots, F_{output}^M] \in \mathbb{R}^{M \times H}$,其中 H 为BiLSTM里的隐单元节点数。

[0038] 优选地,S5进一步包括:

[0039] S501,通过softmax层对特征 F_{output} 进行分类,若训练集中所包含的目标总数为 C ,测试HRRP样本 X_{test} 对应于目标集中第 i 类目标的概率表示为:

$$[0040] \quad P(i | x_{\text{test}}) = \frac{\exp(F_s(i))}{\sum_{c=1}^C \exp(F_s(c))}$$

[0041] 其中, $\exp(\cdot)$ 表示取指数运算, $F_s(i)$ 指的是向量 F_s 中的第 i 个元素, $F_s = W_s F_{\text{ATT}}$, W_s 是向量 F_s 的权重矩阵,通过最大后验概率将测试HRRP样本 X_{test} 分类到最大目标概率 c_0 中:

$$[0042] \quad c_0 = \arg \max_i P(i | x_{\text{test}})$$

[0043] S502,在所提模型的参数训练过程中,采用交叉熵作为代价函数,数据中,我们有三类实测飞机,和一类仿真数据,作为库外目标一起训练,每个训练样本 x_{train} 的代价可以表示为:

$$[0044] \quad L = -\sum_{i=1}^C z(i) \log P(i | x_{\text{train}}) - 2 \sum_i \bar{z}(i) \log P(i | x_{\text{train}})$$

[0045] 其中, C 为库内互相分错的目标, \bar{C} 为将三类实测目标分到第四类或者将第四类分

为前三类的目标,在这个loss function中使用了比较简单的方法,给予式中第二部分两倍的损失, $z(i)$ 表示训练样本 x_{train} 的真实标签 z 中的第 i 个元素, z 是一个one-hot vector, $P(i|x_{\text{train}})$ 表示训练样本 x_{train} 对应于第 i 个目标的概率。

[0046] 采用本发明具有如下的有益效果:

[0047] (1) 本发明实施例中应用了带注意力机制的幂次变换模块,幂次变换可以很好地提高识别的性能,加入了注意力机制,可以学习每种幂次变换的不同权重,从而为下一层网络提供更加合适的输入。

[0048] (2) 本发明实施例中应用了Bert模块。Bert模型可以更好地利用雷达目标的整体固定的物理结构特性这一先验,而且Bert的网络结构可以使模型层数更深,更好地抽象HRRP数据的特征,进行更好地识别。

[0049] (3) 使用双向LSTM,在强化位置信息的基础上,同时也加强了方向信息。更有利于特征提取和目标分类。

附图说明

[0050] 图1为本发明实施例的基于Bert和BiLSTM的雷达目标识别及据判方法的步骤流程图。

具体实施方式

[0051] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0052] 参见图1,所示为本发明实施例的基于Bert和BiLSTM的雷达目标识别及据判方法的步骤流程图,包括以下步骤:

[0053] S1,原始的HRRP数据主要存在强度敏感性和平移敏感性问题,需要对其进行预处理。使用 L_2 强度归一化,改善HRRP强度敏感性的问题。原始HRRP数据是通过距离窗从雷达回波数据中截取的,截取时,录取的距离像在距离波门中的位置不固定,从而导致HRRP的平移敏感性。重心对齐对HRRP进行处理,降低平移敏感性。收集样本并划分训练集和测试集,通过雷达采集的HRRP数据,总共 c 类,对每个类别进行采样,分别选出训练集和测试集,之后进行合并,保证每个类别的数据形式都有被采样到,训练集和测试集比例保持在7:3,并且将仿真数据作为负样本加入到训练集中,作为第 $c+1$ 类,以提高模型的据判能力,将训练数据集记为 $D = \{(x_i, y_k)\}_{i \in [1, n], k \in [1, c]}$,其中 x_i 表示第 i 个样本, y_k 表示样本属于第 k 类,一共采集了 c 类目标, n 表示样本总数;进行数据预处理,包括强调归一化和重心对齐。

[0054] S2,使用带注意力机制的幂次变换。直接将HRRP数据送入Bert模型中,由于HRRP数据中各个距离单元内的回波幅度较大,会导致Bert训练不稳定。幂次变换对提高识别性能有很大的帮助下。因此在将数据输入Bert之前,先对数据进行幂次变换。为了使幂次变化的适应性更好,在幂次变化模块使用了注意力机制,用来给予不同幂次变化后数据不同的权重,以使网络获得最有利于网络学习的输入。

[0055] S3,用Bert处理幂次变换的输出,提取更加深层的特征。构建Bert模型,优化模型

超参数和结构。在传统模型中,循环神经网络,只能依靠当前时刻,和之前时刻的信息。但是HRRP数据中包含了目标样本的整体结构先验,如果只是利用HRRP结构的单向信息,是不利于对HRRP数据进行学习并且识别的。

[0056] Bert模型用于抽象并描述HRRP所蕴含的目标物理结构特征,并融合多层次的物理结构特征进行识别。Bert结构中的注意力机制,可以更好的抑制相关性较弱的特征,给予重要特征更大的权重,降低两边噪声的干扰。

[0057] S4,通过双向LSTM,更好地利用HRRP数据的方向信息。

[0058] S5,搭建分类器,对HRRP目标分类。采用softmax对网络的输出进行分类。

[0059] 测试阶段:

[0060] S6,将经过S1处理后的HRRP测试集送入S2、S3、S4、S5步骤中已训练完成的模型中进行测试。

[0061] 具体应用实例中,S1详细步骤进一步包括:

[0062] S101,强度归一化。假设原始HRRP数据表示为 $x_{\text{raw}} = [x_1, x_2, x_3 \cdots, x_M]$,其中M表示HRRP数据中距离单元总数,那么归一化之后的HRRP可以表示为:

$$[0063] \quad x_{\text{normalization}} = \frac{x_{\text{raw}}}{\sqrt{\sum_{m=1}^M x_m^2}}。$$

[0064] S102,重心对齐。为了使训练和测试具备统一标准,采用重心对齐法消除平移敏感性。重心对齐法通过平移HRRP使其重心位于HRRP中心点附近,重心g的计算方法如下:

$$[0065] \quad g_{\text{HRRP}} = \frac{\sum_{m=1}^M m x_m}{\sum_{m=1}^M x_m}$$

[0066] 具体应用实例中,S2具体包括,对HRRP数据进行幂次变换,并使用注意力机制。对经过上述预处理的HRRP数据进行多次幂次变换,幂次变换层输出可以表示为:

$$[0067] \quad \text{Output} = \alpha_i \sum_{k=0}^K X^k$$

[0068] 其中 α_i 为每种幂次换的权重,k表示幂次。通过学习,学出不同的权重,使Bert模块可以得到更合适的输入。其中每一条样本最终会呈现如下形式:

$$[0069] \quad \mathbf{X} = \begin{bmatrix} x_1^{0.1} & x_2^{0.1} & \cdots & x_M^{0.1} \\ \vdots & \vdots & & \vdots \\ x_1^{k-1} & x_2^{k-1} & \cdots & x_M^{k-1} \\ x_1^k & x_2^k & \cdots & x_M^k \end{bmatrix}。$$

[0070] 具体应用实例中,S3详细步骤为:

[0071] S301,对输入的数据X进行嵌入操作,并且为数据加入时序信息,即位置信息。采用随机初始化位置嵌入的方式,与数据嵌入之后的X相加。

[0072] S302,多头注意力机制,提取有效特征。可以有效利用其它维的特征,从而使得以往模型的远距离依赖特征之间的距离被极大的缩短。嵌入层的N维特征,分成12组,在每一组中进行self-attention,再将12组的结果进行合并。X中每一维特征分别乘以三个随机初始化的矩阵 W_q, W_k, W_v 得到Q,K,V三个矩阵。进行如下计算得到Z:

$$[0073] \quad Z = \text{soft max}(\frac{Q \times K^T}{\sqrt{d_k}})$$

[0074] 再将 Z_0 到 Z_{11} 按列相连,乘以随机初始化的矩阵 W ,即可得到最终的 Z 。

[0075] S303,前馈层中,将 N 维特征,映射到高维,再从高维映射到低维,进一步保留有效特征。

[0076] S304,残差&层归一化层,残差层就是残差连接模块,因为要叠加深层神经网络,残差结构可以很好地消除层数加深所带来的信息损失问题,也可以将损失快速回传,避免梯度消失。

[0077] 与CNN中批归一化一样,在这里加入了层归一化,使训练过程更加稳定。

$$[0078] \quad LN(x_i) = \alpha \times \frac{x_i - \mu_L}{\sqrt{\sigma_L^2 + \epsilon}} + \beta$$

[0079] 式中 $x \in \mathbb{R}^{N \times C \times H \times W}$,层归一化对每个样本 C, H, W 维度上的数据求均值和标准差,保留 N 维度。 μ 为均值, σ 为方差, α 和 β 为参数, ϵ 为一个非常小的值。

[0080] 具体实施例中,S4详细步骤为:Bert模块之后,输出的形状为 $[\text{batch_size}, M, N]$,为了加强特征提取能力和据判能力,应用了双向LSTM,在Bert中仅仅用位置嵌入式操作来告诉模型输入token的位置信息,但是在序列任务中位置信息和方向信息都是很必要的,所以使用双向LSTM习得观测序列的依赖关系。

[0081] 假设输入是特征 $F_{LSTM} = [F_{LSTM}^1, F_{LSTM}^2, \dots, F_{LSTM}^M] \in \mathbb{R}^{M \times N_i}$,其中 i 表示第 i 个BiLSTM的各时间点维数, M 表示输入序列长度,假设输出为 $F_{output} = [F_{output}^1, F_{output}^2, \dots, F_{output}^M] \in \mathbb{R}^{M \times H}$,其中 H 为BiLSTM里的隐单元节点数。

[0082] 具体实施例中,S5详细步骤为:

[0083] S501,通过softmax层对特征 F_{output} 进行分类。若训练集中所包含的目标总数为 C ,测试HRRP样本 X_{test} 对应于目标集中第 i 类目标的概率可表示为:

$$[0084] \quad P(i | x_{test}) = \frac{\exp(F_s(i))}{\sum_{c=1}^C \exp(F_s(c))}$$

[0085] 其中, $\exp(\cdot)$ 表示取指数运算, $F_s(i)$ 指的是向量 F_s 中的第 i 个元素, $F_s = W_s F_{ATT}$, W_s 是向量 F_s 的权重矩阵。我们通过最大后验概率将测试HRRP样本 X_{test} 分类到最大目标概率 c_0 中:

$$[0086] \quad c_0 = \arg \max_i P(i | x_{test})$$

[0087] S502,在所提模型的参数训练过程中,采用交叉熵作为代价函数。在数据中有三类实测飞机和一类仿真数据,作为库外目标一起训练,每个训练样本 x_{train} 的代价可以表示为:

$$[0088] \quad L = -\sum_{i=1}^C z(i) \log P(i | x_{train}) - 2 \sum_{i=1}^{\bar{C}} z(i) \log P(i | x_{train})$$

[0089] 其中, C 为库内互相分错的目标, \bar{C} 为将三类实测目标分到第四类或者将第四类分为前三类的目标,在这个loss function中使用了比较简单的方法,给予式中第二部分两倍的损失, $z(i)$ 表示训练样本 x_{train} 的真实标签 z 中的第 i 个元素, z 是一个one-hot vector, $P(i | x_{train})$ 表示训练样本 x_{train} 对应于第 i 个目标的概率。

[0090] 应当理解,本文所述的示例性实施例是说明性的而非限制性的。尽管结合附图描述了本发明的一个或多个实施例,本领域普通技术人员应当理解,在不脱离通过所附权利要求所限定的本发明的精神和范围的情况下,可以做出各种形式和细节的改变。

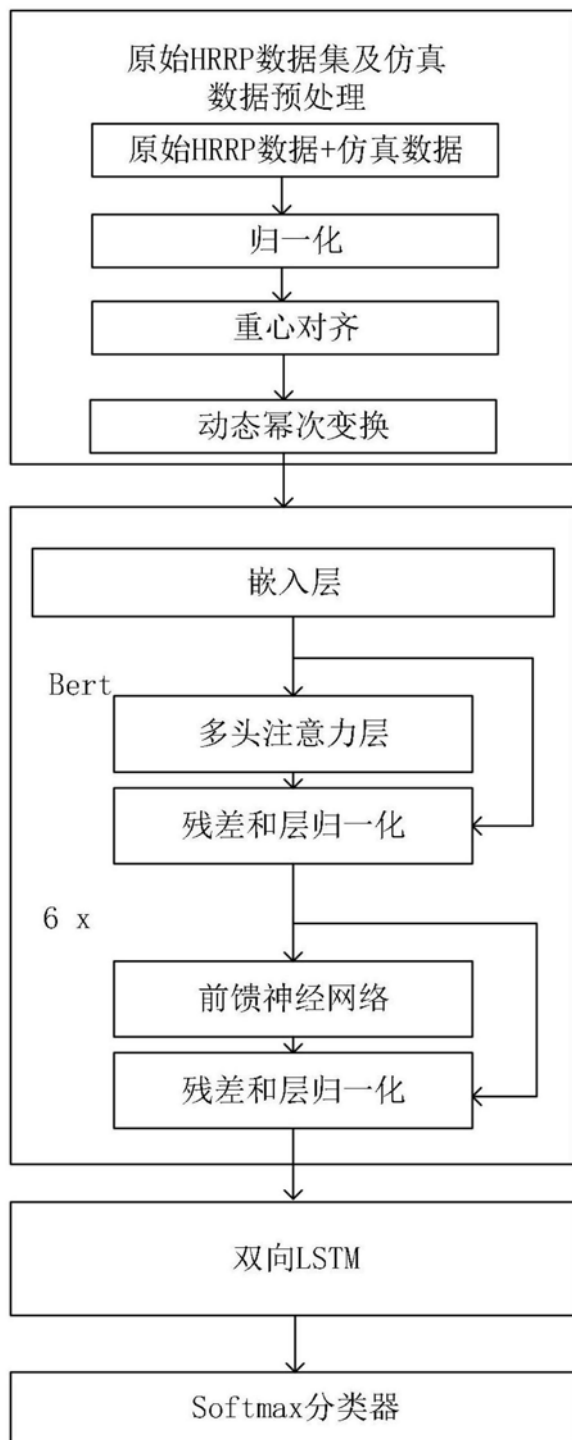


图1