



计算机科学

COMPUTER SCIENCE

基于全局属性注意力神经过程模型的数据补全研究

程恺, 刘满, 王之腾, 毛绍臣, 申秋慧, 张宏军

引用本文

程恺, 刘满, 王之腾, 毛绍臣, 申秋慧, 张宏军. [基于全局属性注意力神经过程模型的数据补全研究](#)[J]. 计算机科学, 2022, 49(10): 111-117.

CHEN Kai, LIU Man, WANG Zhi-teng, MAO Shao-chen, SHEN Qiu-hui, ZHANG Hong-jun. [Study on Data Filling Based on Global-attributes Attention Neural Process Model](#)[J]. Computer Science, 2022, 49(10): 111-117.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合交叉注意力机制的图像任意风格迁移](#)

Image Arbitrary Style Transfer via Criss-cross Attention

计算机科学, 2022, 49(6A): 345-352. <https://doi.org/10.11896/jsjcx.210700236>

[基于 Transformer 交叉注意力的文本生成图像技术](#)

Text-to-Image Generation Technology Based on Transformer Cross Attention

计算机科学, 2022, 49(2): 107-115. <https://doi.org/10.11896/jsjcx.210600085>

[高斯过程下的 CMA-ES 在医学图像配准中的应用](#)

Gaussian Process Assisted CMA-ES Application in Medical Image Registration

计算机科学, 2018, 45(11A): 234-237.

[PSO-GP 中文文本情感分类方法研究](#)

Research on Chinese Texts Sentiment Classification Approach Based on PSO-GP

计算机科学, 2017, 44(Z6): 446-450. <https://doi.org/10.11896/j.issn.1002-137X.2017.6A.100>

[基于非对称变邻域粗糙集模型的属性约简](#)

Attribute Reduction Based on Asymmetric Variable Neighborhood Rough Set

计算机科学, 2015, 42(6): 282-287. <https://doi.org/10.11896/j.issn.1002-137X.2015.06.059>

基于全局属性注意力神经过程模型的数据补全研究

程 恺¹ 刘 满^{1,2} 王之腾¹ 毛绍臣¹ 申秋慧¹ 张宏军¹

1 中国人民解放军陆军工程大学指挥控制工程学院 南京 210007

2 中国人民解放军 73131 部队 福建 漳州 363000

(chengkai911@126.com)

摘 要 注意力神经过程(Attentive Neural Process, ANP)模型采用生成模型的方法,以样本的任意局部上下文点为输入,输出整个样本的分布函数,从而模仿高斯过程回归完成数据补全任务。样本的属性信息可以为样本的生成提供重要信息,然而 ANP 模型忽略了对属性信息的使用。受条件变分自动编码器(CVAE)模型以标签为条件控制样本生成的启发,文中提出了全局属性注意力神经过程(Global-attribute Attentive Neural Process, GANP),将样本属性嵌入到编码器网络,从而使浅层变量隐含样本属性信息。同时,在解码器网络中加入样本属性作为特征,使模型的生成样本更为准确,特别是当输入上下文点数量稀少时,属性信息能够帮助模型生成更清晰、准确的样本。最后,从定性和定量两个方面证明了 GANP 性能的优越性,可以看出该模型扩展了 NP 家族模型的应用范围,从而更灵活、快速、准确地解决只有部分上下文信息时整个样本的数据补全问题。

关键词: 神经过程;交叉注意力;变分推断;高斯过程;全局属性

中图法分类号 TP391.4

Study on Data Filling Based on Global-attributes Attention Neural Process Model

CHEN Kai¹, LIU Man^{1,2}, WANG Zhi-teng¹, MAO Shao-chen¹, SHEN Qiu-hui¹ and ZHANG Hong-jun¹

1 Institute of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China

2 73131 Unit of PLA, Zhangzhou, Fujian 363000, China

Abstract The attention neural process(ANP) model which adopts the method of generative model, takes any number context points of the sample as input, and outputs the distribution function of the entire sample, so as to approximate the function of Gaussian process regression(GPR) to realize the data fullfilling task. In reality, many scenes or datasets containe the attributes or labels data which are critical for generating the missing data. However, the ANP ignores full use of them. Inspired by CVAE model which control sample generation with lable as condition, this paper proposes global attribute attentional neural process (GANP), which embeds sample attributes or labels into ANP network to make the model generate samples more accurately, especially when the number of input context points are scarce. In detail, the sample attributes are embedded into the encoder network, so that the latent variables contain the sample attribute information. At the same time, the sample attributes are added as features in the decoder network to help generate more accurate samples. Finally, experimental results prove the superiority of GANP in both qualitative and quantitative, and it also reveals that GANP expands the application of NP families which can solve the Gaussian process regression problem more flexibly, quickly and accurately.

Keywords Neural process, Cross attention, Variational inference, Gaussian process, Global attribute

1 引言

高斯过程回归(Gaussian Process Regression, GPR)是使用高斯过程(Gaussian Process, GP)先验对数据进行回归分析的非参数概率模型,在时间序列分析、图像处理和自动控制等领域的问题中得到了广泛应用^[1]。它能够以样本的部分上下文信息为条件,利用贝叶斯推断来估计剩余样本内容,从而完成数据补全的任务。然而, GPR 有两个显著的缺点,限制了

其在大样本和高维情形中的应用:1)高昂的计算代价,经典 GPR 推断的计算复杂度为 $\mathcal{O}(n^3)$,即使当前最优算法的计算复杂度仍为 $\mathcal{O}(n^2)$ ^[2];2)高斯过程中的核函数是固定的,且需要提前指定,导致为特定场景选择合适的核函数较难。

2018 年,DeepMind 公司的 Garnelo 等首次提出了神经过程(Neural Process, NP)模型^[3]。该模型利用神经网络能够处理高维特征和并行计算的优势,模仿 GPR 的功能,实现了不确定性分布函数估计。NP 使用神经网络进行端到端的

到稿日期:2021-08-03 返修日期:2022-01-13

基金项目:国家自然科学基金(61806221)

This work was supported by the National Natural Science Foundation of China(61806221).

通信作者:张宏军(jsnjzhj_lgdx@163.com)

学习和推断,其计算复杂度仅有 $\mathcal{O}(n)$,弥补了 GPR 计算复杂度高的缺点;另外,基于神经网络强大的函数拟合能力,它学到的是隐式核函数(带参数的神经网络),从而避免了 GPR 中使用固定核函数使其适应性有限的问题。

通常利用确定性连接的深度神经网络(Deep Deterministic Neural Networks, DDNNs)拟合一个复杂函数,从而解决回归或分类问题。然而,DDNNs 一旦训练完毕,其模型从输入到输出的映射方式就固定了,当应用场景或任务发生变化时,DDNNs 需要重新训练。NP 采用元学习(Meta Learning)^[4-5]架构,在训练过程中建模随机过程的多种任务,以灵活可变的观察数据为条件,构建编码器网络进行概率推断,通过解码器网络输出随机过程的分布函数,即 NP 家族模型,实现了对随机过程的学习。

受 NP 启发,一系列结构类似于 NP 模型的不确定性函数估计模型被提出,一般都被统称为 NP 家族,如注意力神经网络^[6]、卷积条件神经过程(Convolutional Conditional Neural Process, ConvCNP)^[7]、序列神经过程(Sequential Neural Process, SNP)^[8]等。它们分别从不同的角度改进了 NP 模型的结构,使其不确定性预测的性能更好,并且已经在自动驾驶^[9-10]、视频预测^[11]、机器人路径规划^[12]、数据填充^[13]等领域得到应用。

然而,NP 家族只是以样本中局部上下文点为输入条件,并未关注样本的整体特性,即样本属性或标签。本文借鉴条件变分自动编码器(Conditional Variational Autoencoder, CVAE)的标签数据嵌入原理,将样本属性作为条件嵌入 NP 家族模型,从而提出了一类新的 NP 家族成员-全局属性神经过程。特别地,针对 ANP 模型,我们提出了全局属性注意力神经过程,将样本属性或标签嵌入 ANP 网络中,使模型的生成样本更准确,特别是当输入上下文点对数量稀少时,属性信息帮助模型生成更清晰的样本,即 GANP 能够更出色地完成数据补全任务。本文采用了高斯过程曲线(1-D toy)、MNIST^[14]和 CelebA^[15]这 3 种数据集进行验证,实验结果表明,改进后的 GANP 模型具有更准确的预测或生成性能。为了促进相关领域研究的进步和 GANP 模型的应用,本文开源了相关代码¹⁾。

2 NP 家族模型介绍

用深度神经网络模型进行不确定性估计是一个活跃的研究领域^[16]。近年来,Garnelo 等先后提出了条件神经过程(Conditional Neural Process, CNP)^[17]、神经过程(NP)^[3]和注意力神经过程(ANP)^[6],较好地模仿了 GPR 的灵活预测能力,同时又兼具神经网络高效计算、表示丰富的优点。

对于一个回归任务:假设有样本上下文(Context)数据将集合 $(X_C, Y_C) := \{(x_i, y_i)\}_{i \in C}$ 作为已知观察数据,求目标(Target)数据点集 $X_T := \{x_i\}_{i \in T}$ 对应的点集 $Y_T := \{y_i\}_{i \in T}$ 。GPR 将数据变量建模为高斯过程,利用 (X_C, Y_C) 作为训练数据学习一个特定核函数,而后进行贝叶斯推理,得出 X_T 对应的 Y_T 的高斯分布的均值和方差。假设 Y_T 内变量相互独立,

则该回归模型可以表示为:

$$\log p(Y_T | X_T, X_C, Y_C) = \sum_{i=1}^n \log p(y_i | x_i, X_C, Y_C) \quad (1)$$

CNP^[17]用一个编码器(Encoder)和一个解码器(Decode)神经网络做直接连接来建模 $p(y_i | x_i, X_C, Y_C)$,其推理图模型如图 1(c)所示。编码器对上下文点集内每个数据对 $\{(x_i, y_i)\}_{i \in C}$ 进行编码和聚合运算(一般为求均值)得到 r_C 。解码器 f_{dec} 以 r_C 和 x_i 为输入,输出 y_i 的高斯参数。CNP 的推理过程可以表示为:

$$r_C = f_{enc}(X_C, Y_C) = \frac{1}{|C|} \sum_{i \in C} f_{enc}(x_i, y_i) \quad (2)$$

$$(\mu_i, \sigma_i) = f_{dec}(r_C, x_i) \quad (3)$$

$$p(y_i | x_i, X_C, Y_C) = \mathcal{N}(y_i | \mu_i, \sigma_i^2) \quad (4)$$

CNP 能够宽泛地实现高斯过程的功能,但 r_C 是确定性向量,使解码网络无法输出多样性结果。因此,Garnelo 等又提出了 NP 模型^[3],使网络输出的多样性显著增强。

NP^[3]受变分自动编码器(Variational Autoencoders, VAE)^[18-19]变分思想的启发,将 r_C 映射为高斯分布表示 $\mu(r_C), \sigma(r_C)$,而后使用重参数技巧,从高斯分布 $\mathcal{N}(\mu(r_C), \sigma^2(r_C))$ 中抽样出全局隐变量 z ,最后解码网络以 z 和 x_i 为条件解码出 y_i 的高斯分布参数,其推理图模型如图 1(d)所示。然而,NP 用全局隐变量 z 作为输入条件,模拟随机函数的抽样,该方式存在一个很大的缺陷,即欠拟合。针对这一缺点, Kim 等融合了 CNP 和 NP 模型,并引入交叉注意力机制(cross-attention),提出了 ANP^[6]模型。

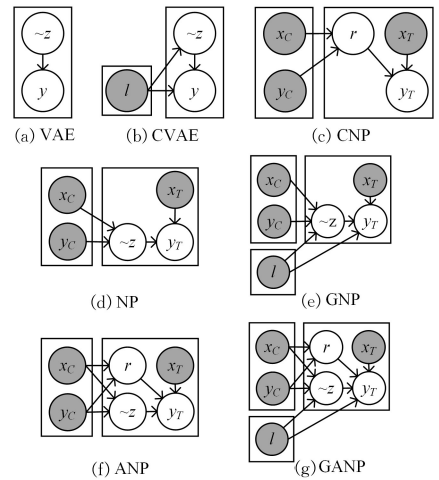


图 1 多种生成网络图模型对比

Fig. 1 Comparison of multi-generative network graph models

ANP 将 CNP 和 NP 进行综合,构建了包含确定性路径和浅层(Latent)路径的双路径条件生成模型,并且将 CNP 中的聚合模块(求均值)用交叉注意力代替,输出与查询 x 。相对应的注意力输出 r^* 。其推理过程可以描述为:

$$P(Y_T | X_T, X_C, Y_C) = \int p(Y_T | X_T, r_C^*, z) q(z | X_C, Y_C) dz \quad (5)$$

ANP 模型将确定性交叉注意力路径 r_C^* 和全局隐层变量 z 共同作为解码网络的输入, r_C^* 关注 X_T 周边的局部信息,而

¹⁾ <https://github.com/daman000/label-NP.git>

z 隐含了样本的整体信息,使该模型同时具有变分网络的多样生成能力和交叉注意力直连模型的预测能力。当上下文点数量较少时, z 发挥主要作用,模型生成能力较强,可以生成符合整体数据分布的样本;当上下文点数量较多时, r_c^* 发挥主要作用,模型的预测能力较强,可以较准确地预测出目标点值。

3 全局属性注意力神经过程

对于生成网络,数据集的属性或者标签能够为样本的生成提供限制条件,从而使生成的样本更加准确,如手写数据集 MNIST^[14] 包含 0-9 共 10 个标签数据,人脸数据集 CelebA^[15] 包含 40 维属性数据。对于 NP 家族模型,它们以组成样本的数据点为研究对象,将样本分解为数据对集合 $(x_i, y_i)_{i \in N}$, N 为单个样本包含的数据对个数。每个属性或是标签与样本对应,如一个手写图像对应一个具体的标签,而样本中的数据点对,如手写图像中的像素点,这些标签或属性对应于单个样本中的所有点是不变的,具有“全局”的特性。因此,在 NP 家族模型中,将样本的属性或标签称为“全局属性”。

变分自动编码器 (VAE)^[18-19] 是经典的无监督生成网络模型,基于这一结构, CVAE^[20] 引入属性或标签数据,使生成样本受属性或标签控制,其改进后的结构如图 1(a)、图 1(b) 所示。受此启发,我们对 ANP 模型嵌入属性数据,提出了全局属性注意力神经过程 (GANP) 模型,其改进后的网络结构如图 1(f)、图 1(g) 所示。

图 2 给出了 GANP 的网络结构。首先, GANP 网络分为编码器和解码器两个部分。编码器包含确定性路径 (实线表示) 和浅层路径 (虚线表示) 两条通路。在确定性路径中,输入的上下文点通过相同的多层感知机 (MLP $_{\phi}$) 编码后得到向量表示 $(r_i)_{i \in C}$, 以 $(r_i)_{i \in C}$ 为值 (Values)、上下文点 $(x_i)_{i \in C}$ 为键 (Keys)、目标点 x_* 为查询 (Queries), 做交叉注意力运算, 这里注意力形式使用多头注意力 (Multi-head Attention)^[21], 将得到的结果 r_* 作为解码器的输入。在解码器的浅层路径中, 在网络得到上下文内容的聚合 s_c 后, 与全局属性 l 进行连接, 而后经过多层感知机 (MLP $_{\omega}$) 得到均值和方差表示的概率分布, 使用重参数抽样技巧得到浅层变量 z 。可见 z 中隐含了上下文内容和属性的全局信息。在解码器中, 以 x_* , r_* , z 和 l 为条件并将其加入到解码网络, 得到 y_* 的高斯参数表示。其生成或推理过程可以表示为:

$$P(Y_T | X_T, X_C, Y_C, l) = \int p(Y_T | X_T, r_c^*, z, l) q(z | X_C, Y_C, l) dz \quad (6)$$

其中, $q(z | X_C, Y_C, l)$ 为编码器浅层网络, $p(Y_T | X_T, r_c^*, z, l)$ 为解码器网络。

其证据下限 (Evidence Lower-bound, ELBO) 优化函数可以表示为:

$$\log p(y_{m+1:n} | x_{1:n}, y_{1:n}, l) \geq E_{q(z | x_{1:n}, y_{1:n})} \left[\sum_{i=m+1}^n \log p(y_i | z, r_c, x_i, l) + \log \frac{q(z | x_{1:m}, y_{1:m}, l)}{q(z | x_{1:n}, y_{1:n}, l)} \right] \quad (7)$$

其中, $[1, m]$ 为上下文点的索引序号, $[n - m, n]$ 为目标点的

索引序号, $p(y_i | z, r_c, x_i, l)$ 为解码器的概率似然, $q(z | x_{1:m}, y_{1:m}, l)$ 和 $q(z | x_{1:n}, y_{1:n}, l)$ 为用神经网络近似的 z 的变分后验概率, 其对于任意大小序列 $x_{1:n}, y_{1:n}$ 具有移位不变性。利用估计次数为 1 的 EM (Expectation Maximization) 算法即可训练优化 ANP 模型网络。

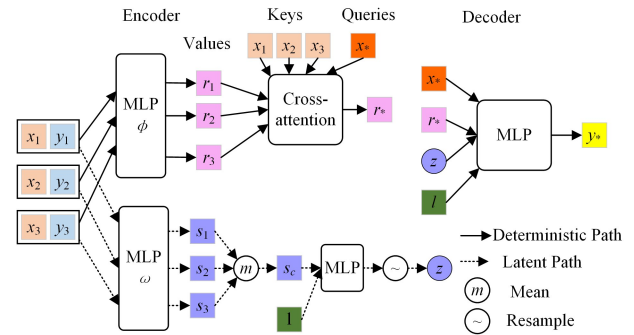


图 2 GANP 网络结构图

Fig. 2 Network structure of GANP

针对解码网络的浅层路径输出 $q(z | x_{1:m}, y_{1:m})$, 当上下文点稀少时, 浅层变量 z 难以从 $(x_{1:m}, y_{1:m})$ 中获取样本的全局信息, 进而导致解码器难以根据 z 生成新的样本。嵌入全局属性变量 l 后, 在上下文点输入较少时, 全局属性为编码器和解码器提供了较为丰富的全局信息, 从而使生成样本质量更好。

同理, 针对 NP 模型, 我们可以做相似的嵌入, 即在编码器中将标签或属性数据嵌入到上下文内容聚合 r_c 中, 生成新的内容表示, 在解码器网络中增加标签或属性数据并将其作为输入, 从而得到全局属性神经过程 (GNP)。因此将全局属性加入 NP 家族模型, 即可形成相应的全局属性神经过程模型。

4 实验分析

4.1 模型结构

参照 ANP^[6] 的模型结构, 我们搭建了 GANP 网络, 模型结构如图 3 所示。编码器网络分为确定性路径和浅层路径。确定性路径对上下文点和目标点做交叉注意力运算, 交叉注意力模块使用 8 头的多头注意力层, 输出目标集合 $(x_j)_{j \in T}$ 的注意力表示 $(r_j)_{j \in T}$ 。浅层路径以 (X_C, Y_C) 和属性 l 为输入, 经过 MLP 变换后, 将聚合表示 s_c 和属性表示 r_c 连接并做线性变换, 输出 $\mu_c, \sigma_c \in \mathbb{R}^{128}$, 将它参数化为独立同分布的高斯分布 $q(z | s_c) = N(z | \mu_c, 0.1 + 0.9 * \omega(\sigma_c))$, ω 是 sigmoid 函数, 按重参数技巧抽样出 z 。解码器以 x_j, r_j, l, z 为输入, 经 MLP 变换后, 输出为 $\mu_j, \sigma_j \in \mathbb{R}^{d_y}$, 将它参数化为独立同分布的高斯分布 $p(y_j | z, x_c, y_c, x_j) = N(y_j | \mu_j, 0.1 + 0.9 * f(\sigma_j))$, f 是 softplus 函数。所有的 MLP 使用 relu 非线性层, 最后一层不使用非线性层。MLP $[128] * 4, [d_y * 2]$ 表示多层感知机共有 5 层: 4 层维度为 128, 1 层维度为 $d_y * 2$, $d_y * 2$ 表示输出的 μ_j, σ_j 的维度都为 d_y 。其他 MLP 层数的表示方法与此相似。

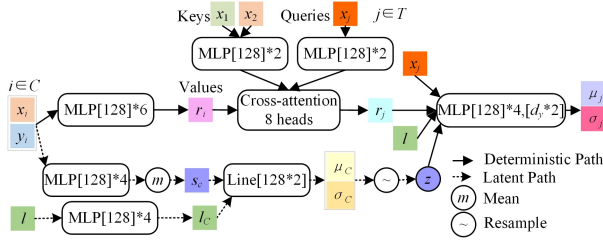


图3 GANP网络详细结构图

Fig. 3 Detailed architecture of GANP

根据 NP^[3] 的结构,我们搭建了 GNP 模型,其网络结构与图 3 类似,但去除了确定性路径。另外,我们复现了 NP 和 ANP 模型,将其作为比较的基准模型。

4.2 实验设置

本实验的计算机主要配置如下:CPU 为 Intel Xeon(R) E5-2630 v4 @ 2.20GHz * 40;内存为 64GB;显卡为 GeForce RTX 2080 Ti * 4。编程语言为 python,使用的神经网络框架为基于 PyTorch 的支持并行训练的集成框架 Pytorch Lightning¹⁾,训练时使用 ADAM^[22] 的优化方式,学习率的调整方式使用自适应调整学习率 ReduceLROnPlateau(mode='min', factor=0.2, patience=30, threshold=5×10⁻³, min_lr=5×10⁻⁵),初始学习率设为 4×10⁻⁵。参考基于 tensorflow 的开源程序²⁾,将神经网络改写到 PyTorch 框架之上,并增加了图像数据的实验。

本文使用了 3 个数据集,分别是合成高斯曲线 (toy 1d)^[3]、MNIST^[14] 和 CelebA^[15]。

(1) 1D 合成高斯曲线 (toy 1d)

我们使用了一个平方指数 (RBF) 核函数的 GP 生成器产生高斯过程数据, RBF 的超参数 (横尺度超参数 l , 纵尺度超参数 σ) 采用均匀分布随机数, $l \sim U[0.1, 0.6]$, $\sigma \sim U[0.1, 1]$, 与 NP^[3] 中设置一致。另外,我们加入了似然噪声, $\sigma_n = 0.02$ 。按照随机抽样的 RBF 和矩阵分解的方法生成 1 维的 GP 曲线, $x \in [-2, 2]$, 均匀生成总数为 100 个点对 (x, y) , 上下文点个数 $m \sim U[3, 50]$, 其他目标点个数 $extra \sim U[0, 50 - m]$, 目标点个数 $n = m + extra$ 。我们从 100 个总点中抽出 m 个作为上下文 (context) 点集, 从剩下的集合中抽出 $extra$ 个其他目标点, 这些点与 context 点组合后形成目标 (target) 点集。在测试或验证时, 目标 (target) 点集为全部 100 个点对 (x, y) 。

同时,我们也会生成超参数 l 和 σ 作为 GNP 和 GANP 的输入。

本文中, batch 大小为 16, 训练集的迭代次数为 1600, 验证集和测试集的迭代次数为 160, 使用单张显卡训练, 总训练 epoch 为 200。

(2) 图像数据集

可以把图像像素值的变化看作随机过程, 对图像像素值的预测过程可以建模为高斯过程回归, x 为 2 维行列坐标, y 为行列坐标对应的像素幅度 (灰度图像 $\in \mathbb{R}^1$, RGB 图像 $\in \mathbb{R}^3$)。本文与文献^[3, 6]使用相同的图片数据集, 手写数字图片 MNIST^[14] 为 28×28 的黑白灰度图像, 标签为 0-9。人脸数据集 CelebA^[15] 为 32×32 大小的 RGB 彩色图像, 每幅图像都有 40 个属性注释。

对于这两个图像数据集, 首先需要对图片进行规范化预处理, 使所有维度 $x \in [-1, 1]$, $y \in [-0.5, 0.5]$ 。在训练时, 上下文点个数 $m \sim U[3, 200]$, 其他目标点个数 $extra \sim U[0, 200 - m]$, 目标点个数 $n = m + extra$ 。我们从整张图片中抽出 m 个上下文 (context) 点集, 从剩下的集合中抽出 $extra$ 个其他目标点, 这些点与 context 点组合后形成目标 (target) 点集。本文中, batch 大小为 32, 使用 4 张显卡进行并行训练, 总训练 epoch 如下: MNIST 为 1000, CelebA 为 300。

4.3 定量分析

本文主要根据模型的损失函数大小来评价模型的量化性能。我们记录了训练总损失、训练 KL 散度损失和验证总损失。完成训练后, 使用模型最终的参数进行测试, 记录测试总损失。

图 4 给出了 GP 数据训练定量指标的性能。从图 4(a) 可以看出, 非注意力模型有更高的 KL 散度, 即非注意力模型有更强的生成能力, 当上下文点较少时, 仍然能生成符合上下文点分布的样本; 从图 4(b) 所示的训练总损失可以看出, 4 种网络基本达到收敛, 其中含有注意力的模型的损失更小, 说明注意力模型的预测性能更好, 与真实数据的差距更小; 从图 4(c) 所示的验证总损失可以看出, 训练基本没有过拟合。在训练的过程中, 全局属性模型比未嵌入属性模型的总损失更低, 这说明全局属性模型的预测准确度更高。另外, 训练总损失和验证总损失没有可比性, 在验证时上下文点数量随机分布的方式和训练时相同, 但是目标点数量为样本的全体, 而不是随机选取的。

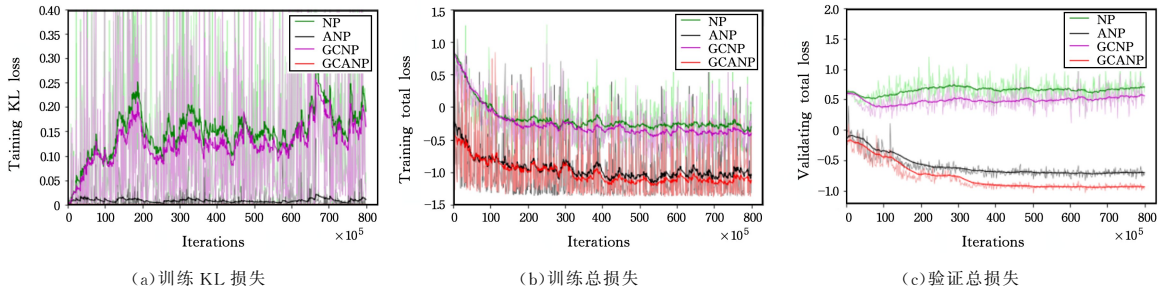


图4 GP数据训练过程定量分析

Fig. 4 GP data quantitative analysis in training

¹⁾ <https://github.com/PyTorchLightning/pytorch-lightning>

²⁾ <https://github.com/deepmind/neural-processes>

为了进一步证明全局属性神经过程的性能,我们设置了3组上下文点数量,分别为样本总点数的0~10%,10%~50%,50%~100%,分别记录了测试总损失,如表1所列。对于GP数据,GNP和GANP模型的总损失一直明显小于NP和ANP的总损失,说明全局属性数据对于预测的准确性有帮助;

表1 不同数量上下文点条件下的测试总损失

Table 1 Total loss of test with different context points

模型名称	GP 数据			MNIST			CelebA		
	0%~10%	10%~50%	50%~100%	0%~10%	10%~50%	50%~100%	0%~10%	10%~50%	50%~100%
NP	4.91	0.01	-0.38	0.11	-0.84	-0.97	-0.77	-2.38	-2.48
GNP	4.64	-0.11	-0.47	-0.49	-0.73	-0.75	-0.85	-2.36	-2.46
ANP	0.52	-0.92	-1.28	-0.54	-1.07	-1.30	-2.08	-3.61	-4.01
GANP	-0.09	-1.10	-1.31	-0.57	-1.10	-1.30	-2.08	-3.58	-3.99

分析以上测试结果可以知,在不同数据集中,属性或标签对预测性能的影响不同。GP数据是随机核函数的高斯过程数据,其动态变化性非常强,少数上下文点很难抓住数据的分布特性,如果已知高斯核的超参数(全局属性),则能较好地帮助预测。MNIST手写数据图像数据集,像素点亮度值规律性很强(10个数字),当上下文点少量时,标签数据可以很好地帮助预测,但是当上下文点数量增加时,上下文点已经能反映数据的整体分布,此时标签数据显得不重要,但标签数据的存在会控制解码网络,产生负面作用。CelebA人脸图像数据集中像素点RGB变化多样,单个像素点的数据分布和40维的属性标签相关性不明显,因此属性数据对预测性能的提升不明显。

总体上看,当上下文点数量较少时,GNP,GANP相比NP,ANP模型得到的损失值更小,体现了更大的优势,但当上下文点数量增多时,上下文点已经隐含了属性信息,全局属性的加入不能明显降低模型的损失值。

4.4 定性分析

神经过程模型家族可以输出待求目标点的均值和方差,

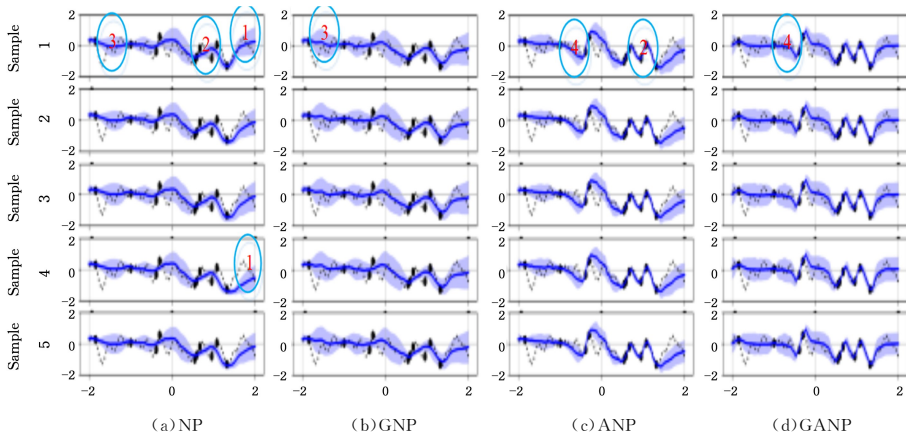


图5 GP数据预测结果(电子版为彩图)

Fig. 5 GP data prediction results

为了研究模型生成结果的多样性,我们将解码器的输出改为待预测值 y^* ,训练时损失值中第一项求似然改为求均方差。图6给出了GP数据在不同上下文点数量($n=4, n=40$)、不同核函数参数($l=0.6, \sigma=0.1; l=0.6, \sigma=1, l$ 决定高斯曲线的震动幅度, σ 决定高斯曲线的震动频率)的情况下,模型生成数据多样性的对比结果,黑色虚线为待预测的真实

对于MNIST数据,当上下文点数量较少时(即0~10%),全局属性模型的测试总损失更小,但是当上下文点数量增大后,全局属性发挥的作用越来越小,甚至可能产生负面作用(GNP模型的总损失更大);对于CelebA人脸数据集,全局属性的加入对模型的总损失影响不大,正面作用和负面作用都不明显。

均值反映了待预测值的大小,方差反映了预测值的波动范围,同时也反映了预测值的可信度。图5给出了神经过程模型对GP数据的预测输出,黑色虚线为待预测的真实高斯曲线,黑色点为上下文点,蓝色实线为模型预测的均值,淡蓝色背景为预测的方差范围。从图中可以看出:1)NP模型的预测输出具有多样性,观察图中标示的1号区域可以看出,抽样1和4预测输出的均值和方差有变化,但从整体上看,均值方差输出的变化性不大;2)NP的欠拟合较为明显,而ANP避免了欠拟合,观察图中2号区域,NP模型上下文点的预测值与真实值不一致,但ANP模型上下文点的预测值与真实值很接近;3)GNP模型和NP模型都有较为光滑的预测均值,但GNP模型输出的方差更准确,观察图中3号区域可以看出,GNP模型的方差更能反映预测均值与真实值的差距;4)GANP模型和ANP模型都有较好的上下文点预测准确度,但GANP模型输出的方差更准确,观察图中4号区域可以看出,GANP模型的方差更能反映预测均值与真实值的差距。

高斯曲线,黑色点为上下文点,蓝色实线为模型预测值,预测值总共抽样5次。可以看出,NP和GNP模型生成的样本变化明显,同时也能看出,相比NP模型,GNP多次抽样的样本方差变化更小,更符合真实GP数据的变化。ANP和GANP模型集成了直连的交叉注意力和隐变量抽样,但交叉注意力对解码器起主要作用,多次抽样的结果变化很小。另外,ANP和

GANP 模型对上下文点的预测更准确, 弥补了 NP 和 GNP 模型欠拟合的缺陷。

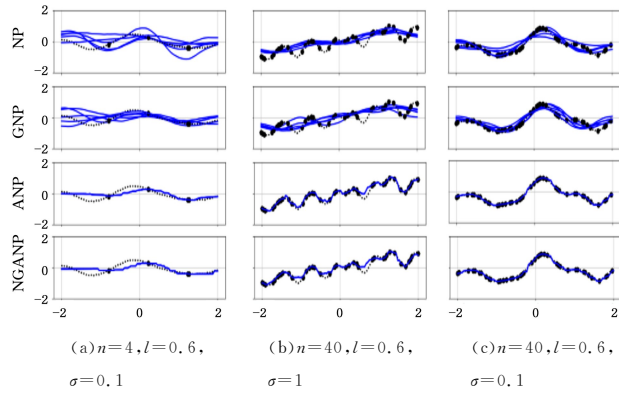


图 6 GP 数据生成多样性对比(抽样次数为 5)(电子版为彩图)

Fig. 6 Comparison of GP data generation diversity(5 sampling)

针对图像数据, 通过不同的上下文点分布和数量来检验神经过程模型对整体的预测或生成性能。图 7 给出了 MNIST 数据集已知图像的上半部分内容时预测的整个图像像素点亮度的均值和方差。图 8 给出了已知随机分布的上下文点的个数分别为 10, 50, 200 和全部时, 预测的整个图像的均值和方差。图 9 给出了 CelebA 图像已知左半部分时, 预测的整个图像的均值和方差。

一致性, 且弥补了 NP 模型预测能力不足的缺点。在相同的上下文点输入情况下, 图 7 和图 8 中 GNP 模型准确地生成了与原图一致的数字, 图 9 中 GNP 模型更好地预测了性别、微笑、眼镜、胡须等特征; 3) ANP 模型能够较好预测上下文点的信息, 对于目标点的预测(生成)也较为准确, 但是也能看出注意力机制和浅层变量很难融合, 图像整体一致性较差; 4) 在上下文数量较少时, 相比 ANP 模型的结果, GANP 模型生成的图像准确性更强, 并且整体的一致性也更好。图 7 和图 8 中 GANP 模型生成了更准确且一致性更好的数字图像, 图 9 中 GANP 模型增强了预测图像的整体一致性, 更好地预测了眼镜、胡须等特征。

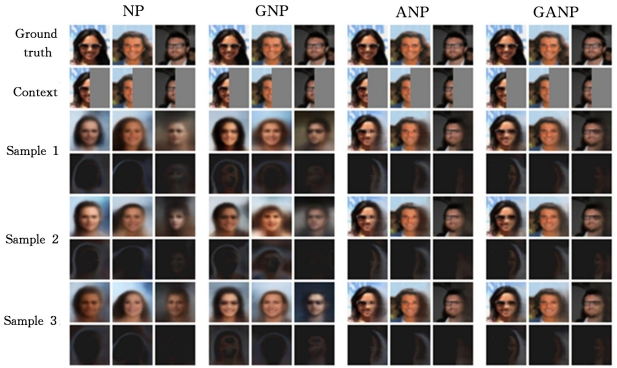


图 9 CelebA 图像已知左边部分预测整体

Fig. 9 Reconstruction of full CelebA image from left half

通过图像实验的定性观察, 进一步验证了定量实验的结论, 即本文提出的全局属性神经模型 GNP 和 GANP 在上下文点数量较少时, 生成的样本整体一致性更好, 预测的结果更准确。特别是 GANP 模型兼顾预测和生成的能力, 在已知部分观察点的情况下, 能够较好完成数据补全任务, 在不确定预测领域有着广泛的应用。

结束语 NP 家族是一种以少量上下文内容为条件的条件生成模型, 能够输出高斯过程中的随机分布函数, 具有灵活性、可伸缩性、计算效率高的优点。本文将全局属性或标签变量嵌入 NP 家族模型, 提出了一种新的 NP 家族成员, 即全局属性条件神经过程, 特别是提出了 GANP 模型, 改善了 ANP 模型在上下文点数量较少时模型预测(生成)准确性低的问题, 其在继承 NP 家族优点的同时, 能够更好地处理数据补全问题, 具有更广泛的应用价值。GANP 以少量上下文信息为条件预测全体的能力, 非常契合战场不完全信息预测的问题, 下一步将用 GANP 解决兵棋不完全信息处理中敌方位置、战术意图的预测问题。

参 考 文 献

[1] RASMUSSEN C E. Gaussian processes in machine learning [C]// Summer School on Machine Learning. Berlin: Springer, 2003:63-71.

[2] QUINONERO-CANDELA J, RASMUSSEN C E. A unifying view of sparse approximate Gaussian process regression[J]. The Journal of Machine Learning Research, 2005, 6(65):1939-1959.

[3] GARNELO M, SCHWARZ J, ROSENBAUM D, et al. Neural processes[J]. arXiv:1807.01622, 2018.

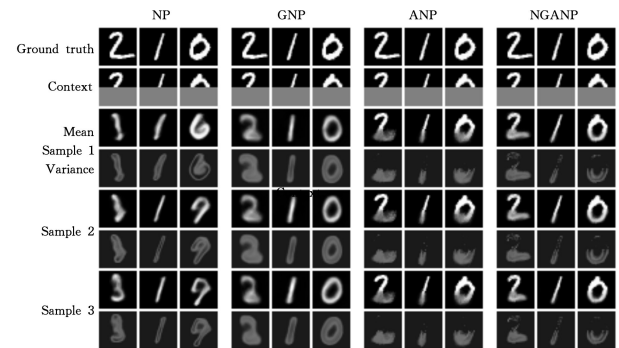


图 7 MNIST 图像已知上半部分预测整体

Fig. 7 Reconstruction of full MNIST image from top half

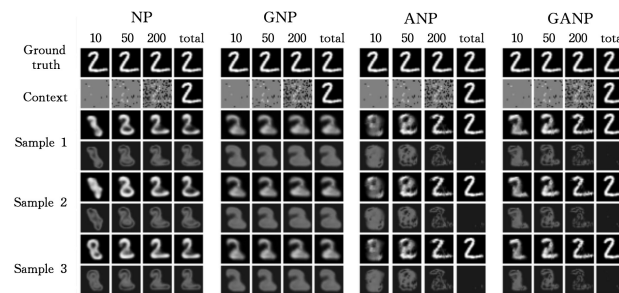


图 8 MNIST 图像已知部分散点预测整体

Fig. 8 Reconstruction of full MNIST image from random context points

本文对模型输出的结果抽样 3 次并进行对比, 通过对预测均值和方差的定性观察, 可以得出以下几点结论: 1) NP 模型的生成多样性能力强, 且生成的图像整体一致性, 但是生成的图像和原图像差别较大, 即预测能力不足(欠拟合现象严重); 2) GNP 模型提升了 NP 模型的性能, 生成的图像整体

- [4] WANG J X, KURTH-NELSON Z, TIRUMALA D, et al. Learning to reinforcement learn[J]. arXiv:1611.05763, 2016.
- [5] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]// International Conference on Machine Learning. PMLR, 2017:1126-1135.
- [6] KIM H, MNIH A, SCHWARZ J, et al. Attentive neural processes[J]. arXiv:1901.05761, 2019.
- [7] GORDON J, BRUINSMA W P, FOONG A Y K, et al. Convolutional conditional neural processes[J]. arXiv:1910.13556, 2019.
- [8] SINGH G, YOON J, SON Y, et al. Sequential neural processes[J]. arXiv:1906.10264, 2019.
- [9] KUMAR S. Spatiotemporal Modeling using Recurrent Neural Processes[D]. Pittsburgh, PA: Carnegie Mellon University, 2019.
- [10] ZHU J, QIN S, WANG W, et al. Probabilistic trajectory prediction for autonomous vehicles with attentive recurrent neural process[J]. arXiv:1910.08102, 2019.
- [11] KUMAR A, ESLAMI S M, REZENDE D J, et al. Consistent generative query networks[J]. arXiv:1807.02033, 2018.
- [12] MA Y, WANG S Q, MAO Y X. Path Planning of Mobile Robot Based on Neural Process-particle Swarm Optimization[J]. Journal of Hubei University of Technology, 2020, 35(1):17-20.
- [13] SUN X L, GUO Y, LI N, et al. Missing data imputing algorithm based on modified neural process[J]. Journal of University of Chinese Academy of Sciences, 2021, 38(2):280-287.
- [14] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [15] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015:3730-3738.
- [16] BLUNDELL C, CORNEBISE J, KAVUKCUOGLU K, et al. Weight uncertainty in neural network[C]// International Conference on Machine Learning. PMLR, 2015:1613-1622.
- [17] GARNELO M, ROSENBAUM D, MADDISON C, et al. Conditional neural processes[C]// International Conference on Machine Learning. PMLR, 2018:1704-1713.
- [18] KINGMA D P, WELING M. Auto-encoding variational bayes[J]. arXiv:1312.6114, 2013.
- [19] DOERSCH C. Tutorial on variational autoencoders[J]. arXiv:1606.05908, 2016.
- [20] SOHN K, LEE H, YAN X. Learning structured output representation using deep conditional generative models[C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. 2015:3483-3491.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017:5998-6008.
- [22] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980, 2014.



CHENG Kai, born in 1983, Ph.D, associate professor. His main research interests include data analysis and data mining, artificial intelligence and planning techniques.



ZHANG Hong-jun, born in 1963, Ph.D, professor, Ph. D supervisor. His main research interests include data & knowledge engineering and computer simulation theory.

(责任编辑:喻黎)