

Time Series Change Point Detection with Self-Supervised Contrastive Predictive Coding(2021)

2022年10月1日

笔记

一.摘要

1.变化点检测 (Change Point Detection、CPD) 方法识别与时间序列数据的趋势和属性变化相关的时间, 以描述系统的基本行为。

个人理解: 比如纵轴是一个与时间有关的属性, 横轴是时间, 需要检测该属性变化的时间节点

2.提出一种基于对比预测编码 (TS - CP2) 的自我监督的时间序列变化点检测方法的新方法。

3.TS - CP2 是第一个采用对比性学习的方法, TS - CP2 优于五个最先进的CPD方法。

二.CCS概念

正文

一.简介

1.我们有强大的访问和分析数据的能力, 但是缺乏资源来利用传统的数据整理过程。

有监督的机器学习数据注释昂贵、笨重、不准确

所以自我监督和无监督学习方法很热门

2.变化点检测是一种分析方法, 用于识别与序列数据突然过渡相关的时间, 可以从非注释数据当中提取意义

变化点代表系统状态的改变, 需要人类的关注

CPD还可以用于相关的时间分割、事件检测和时间异常检测问题

3.CPD技术被广泛应用于多变量时间序列数据, 变化点通常是从时间序列的不同属性中的一个来估计的

无监督的CPD方法通常是为了识别基于一个特定属性的变化而开发的

4.目前的CPD未能有效推广的原因 (不足) :

不同的语义边界与不同的属性相关, 可能互相矛盾

针对缓慢变化的时间形状和统计属性, 也有些麻烦

5.提出TS-CP2: 基于对比预测编码的自监督时间序列变化点检测方法

连续的时间间隔之间的共享信息最大化, 时间上分离的时间间隔对之间的共享信息最小化

学习到的表征再相邻的时间间隔之间有明显的不同, 那么变化点就更有可能出现

目标是证明这种自我监督的表征能检测出比以前更广泛的变化点

二.相关工作和背景

1.时间序列变化点检测

1) 目前的CPD方法大多是基于非深度学习方法

统计学方法通常在确定时间序列中相邻短间隔的统计差异上计算变化点

重点是统计差异：参数方法和非参数方法

将变化点确定为分段边界，使整个分段时间序列的统计学成本函数最优化
(类似于动态规划的分段最小二乘法)

自上而下和动态编程的方法估计变化点

搜索使分段时间序列的信息增益最大化的边界 (和上一种比较像)

用贪婪的搜索来确定使分割的高斯模型的正则化似然估计最大化边界

另一大类利用时间序列的时间形状模式，FLOSS：通过识别时间序列中与形状模式突出变化相关的位置来检测变化点

ESPRESSO提出了一种混合CPD方法，利用时间序列的时间形状模式和统计分布，比纯粹的方法要好

2) 基于深度学习的CPD方法最近也提出来了

利用编码表示的重建误差的峰值，使用自动编码器进行CPD

kernelLearning变化点检测，是一种最先进的端到端CPD方法，自动学习kernel的参数核结合多个kernel捕获不同类型的变化点

KL-CPD被证明优于其他深度学习和非深度学CPD方法

CPD可以用于视频处理应用

现有的视频分割方法是有监督的 (对动作顺序有了解)

自动回归模型：根据最近看到的帧来预测下一个视频帧，预测误差的增加被用来检测分割的边界

2.表征学习 (representation learning)

自监督表征学习已被用于捕捉视频、图像、文本和时间序列数据的信息和紧凑表征。

1) 对比学习

对比学习用于制定数据集中样本相似或不相似的原因 (应该就是作为表征)

用一组由正面样本对 (某种意义上认为相似的样本) 和负面样本对 (认为不同的样本) 组成的训练实例

使正面样本对更接近、在嵌入空间内进一步分离负样本对。

常用的损失函数：对比性损失、三重损失 (更优、考虑正负样本对之间的关系)

Multiple Negative Learning loss functions, 同时考虑多个负样本对, 采样方法很昂贵, 被硬负数实例挖掘 (Hard Negative Instance Mining)

2) 基于对比的表征学习

大多数现有的表征学习工作集中在自然语言处理和计算机视觉领域

第一次将对比学习用于变化点检测

多变量时间序列的表征学习

提出一种通用的方法：深度扩展卷积网、基于负采样的无监督三重损失函数

对比预测编码（CPC）：使用自动回归模型来学习潜在嵌入空间中的表示。目的是在一个抽象的、全局的信号表示中学习。可以学习不同数据模式的有效表示。

一个深度网络编码器将信号映射到一个较低纬度的潜在空间，应用自动回归模型预测未来的帧。对比性损失函数使当前的帧和未来帧的密度比之间的相互信息最大化。

三.方法

1.问题的定义：

给定一个多变量时间序列 $\{x_1, x_2, \dots, x_T\}$ 的 T 观测值， x_i 属于 R^d （也就是 d 维实变量），估计与时间序列属性变化相关的时间（ t ）。将变化点（或者称之为片段边界）定义为未来无法从该点之前的数据中预测的时间点（其实就是这个点发生了变化，不能用以前的数据预测该点之后的数据，就是分界点）。

未来表征和预期表征之间的异同可以作为检测过渡到下一个片段的措施。（就是按照之前数据的预测结果如果和实际结果差很多，可能就是出现了分界点，就是到了下一个片段）

2.TS-CP2概述

一些用于视频和时间序列变化点和时间异常检测方法：用自动回归模型进行预测，变化点就是与预测误差显著增加的相关样本中检测出来的。但是预测误差很大程度取决于数据分布，建议使用表征学习提取一个紧凑的潜在表征，它对数据的原始分布不变。

（下面说的时间窗口应该就是一个时间节点、时间序列）

采用与CPC方法类似的方法，学习一种使连续时间窗口之间的相互信息最大化的表征。（也比之前的单一属性做表征好，可以综合多个属性的信息）

一个自动回归的深度卷积神经网络对每个时间序列窗口进行编码（特征提取、信息综合）；在这个编码基础上采用一个3层的全连接网络，产生一个更紧凑的嵌入式表示；

计算连续时间窗口的嵌入之间的余弦相似度，用以估计变化点。较小的相似度值相关的时间间隔有较高的可能性成为变化点。

对比学习的方法用来训练编码器。

接下来开始具体的方法。

3.表征学习:

TS-CP2方法的核心是编码器，将连续的时间窗口转换成一个紧凑的嵌入表征，这个表征用来学习有关概念。

采用自动回归的深度卷积神经网络WaveNet学习我们的编码表示，不要LSTM，因为时间卷积网络（TCN）通常可以在来纳许数据中产生卓越的预测性能，而且通常更容易训练。

编码器的结构：两块具有64个大小为4的kernel滤波器的TCN和三层扩张，其速率分别为1、4、16。然后TCN由一个简单的三层投影头与ReLU激活函数和批量归一化组成。

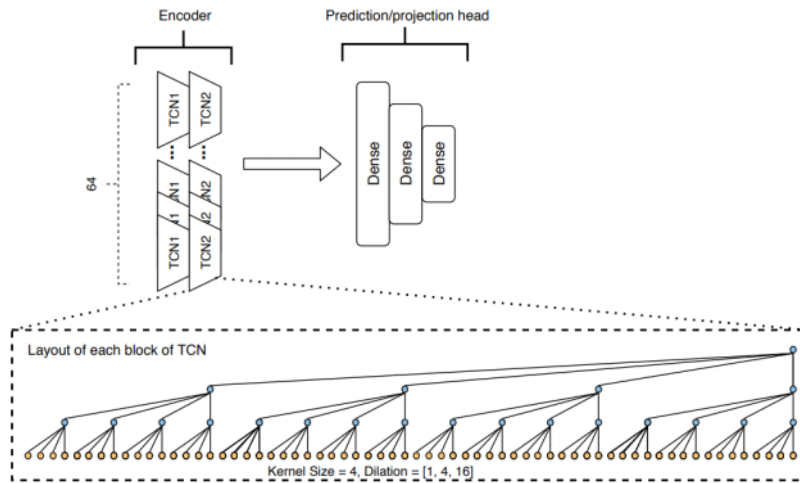


Figure 3: The encoder architecture for TS-CP2. We use two stacks of TCN with kernel size 4 and dilation sizes of 1, 4, and 16 followed by three Dense layers as the prediction head.

成对的历史和未来时间窗口送入编码器，将每个窗口编码映射到一个较低纬度的空间（使用一个具有三个隐藏层的MLP神经网络）

对比学习使用历史和未来的窗口对来训练一个嵌入式表征。

每个训练实例都由一对连续时间间隔的正样本和一组跨时间间隔的负样本组成。

(1) 成本函数，采用InfoNCE损失函数，基于噪声对比估计法，最初为自然语言表征提出。

该函数被定义为最大化连续时间窗口之间的相互信息。

一对正的时间相邻区间 (h_i, f_i) ，历史窗口 (h_i) 和未来窗口 (f_i) ；一组 $K-1$ 的负样本对， h_i 和 f_j 的间隔在整个序列当中很短。

$$\rho_i = \frac{\exp(\text{Sim}(h_i, f_i)/\tau)}{\sum_{j=1}^K \exp(\text{Sim}(h_i, f_j)/\tau)} \quad (1)$$

$$\mathcal{L} = - \sum_{i,j} y_{ij} \log(\rho_i) + (1 - y_{ij}) \log(1 - \rho_i) \quad (2)$$

$$y_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (3)$$

$$\mathcal{L} = \sum_i -\log(\rho_i) \quad (4)$$

(2) 负向抽样（提出一个更简单的抽样策略）

提示：正样本对表示相似，负样本对表示不同。

因此正样本对通常是连续的时间窗口

随机抽取正样本对，用此构建每个批次的负样本对。

随机选 k 个连续窗口的随即对 (h_i, f_i) 作为正样本对，正样本对之间要满足最小时间距离的约束条件。

每个批次采用其他 $k-1$ 个正样本对的未来窗口作为负样本对。

时间序列通常是非平稳的，因此，在时间上相互分离的窗口很可能表现出远比相邻窗口弱的统计依赖性。

4.变化点检测模块

实验的基本过程：将成对的历史和未来窗口转化为紧凑的嵌入，计算被测试的时间序列中嵌入对之间的余弦相似度和余弦相似度的移动平均值之间的差值，并应用寻峰算法来查找差值函数中的局部最大值。与这些局部最大值相关的时间间隔并认为是变化点。

四.实验

1.数据集

展示方法在各种应用当中的有效性：网络服务流量分析、人类活动识别和移动应用程序使用分析（均是时间序列）。

(1) 雅虎基准数据集，最广泛应用的异常检测基准之一。

包含了不同趋势、季节性和噪音的时间序列，包括随机的异常变化点。

(2) HASC挑战赛的数据集，多个传感器收集的人类活动数据。用这个数据集检测变化点目的是为了找到身体活动之间的过渡，比如：“停留”、“行走”、“慢跑”、“跳过”

(3) USC-HAD：类似HASC，是人类活动的数据

2.基线方法 (Baseline Method)

提出的TS-CP2方法的性能与五种最先进的无监督变化点检测技术进行比较，包括ESPRESSO、FLOSS、aHSIC、RuLSIF、KL-CPD。

简单介绍一下这五种方法。

3.评价指标

模型的性能以F1分数来评价。检测到变化点的误差范围是评价每种CPD方法性能的一个重要因素。

每个变化点的估计值位与实际变化点的误差范围内认为是positive。多个变化点估计值在范围内，只有最近的估计值是positive，其余的都是negative。没有落在范围内是negative。

4.微调和敏感度分析

TS-CP2方法对于不同方面的敏感性。（也就是某个方面改变了对结果影响大不大）

1) 窗口大小 (历史和未来间隔的长度)

2) 批量大小 (batchsize)

当编码长度和窗口大小进行平均时, 批量大小和检测性能之间有一个单调的递增关系。

3) 编码长度: 被选中的编码大小范围在4到20维之间

最佳的编码长度取决于窗口大小和批量大小

维度是4的编码对于每个窗口都显示出较弱的性能, 没有足够的能力来表示时间序列的关键特征

检测性能和编码长度之间的关系不是单调递增的, 最长的编码往往被证明不如8~12维的编码

5. 基线比较

提议的TS-CP2方法的性能与三个数据集的五个基线方法进行了比较。

1) 雅虎基准评估:

TS-CP2远胜于五种基线方法

2) USC-HA数据集评估:

TS-CP2在每个误差范围内都优于其他的基线表现。TS-CP2是唯一一个在最小误差下提供高F1分数的方法, 这意味着它在变化点发生的一秒钟内可靠地检测到它。

3) HASC数据集 (规模最小, 变化点最多, 密度最大) 评估:

TS-CP2在每个不同的窗口大小方面取得了第二高的性能水平, 但它仍然取得了最高的平均F1分数。

4) 讨论:

TS-CP2完胜非深度学习的方法

KL-CDP是最近的、有竞争力的基于深度学习的变化点检测, 也比不上TS-CP2

基线CPD方法利用了时间序列的一个特定属性的突然变化, 但是这不能有效适应不同类型的数据集。比如雅虎, 它的变化点由微妙缓慢演变的属性组成。

TS-CP2是利用一个共享的描述来编码历史和未来窗口, 一旦表征模型被训练出来, CPD的实现就很简单, 因为只涉及到表征的对比。

它有可能在低资源设备上实现在线操作, 基准方法不能在线应用, 因为它们需要考虑相当大的一批数据来捕捉重复模式或优化基于熵的损失函数。

五. 结论

提出一种新型的自监督CPD方法TS-CP2用于检测时间序列的变化点。

TS-CP2从历史样本中学习嵌入表征, 预测时间序列的未来区间。

第一个采用对比学习的CPD方法, 为每一帧提取一个紧凑而有信息量的表征向量, 根据后续帧的学习表征之间的一致性来估计变化点。

评估了三个数据集上与其他六个著名方法检测变化点的能力，优于其他方法，并且可以检测在线应用的变化。

个人的总结，就是有多维度的时间序列数据，需要检测其中的异常点。

过去经常用的方法（非深度学习的）：一种是统计方向，就是类似分段最小二乘，寻找边界让某种函数最优化；还有一种是针对时间序列形状检测变异点。以及二者的混合。

可能的思路就是，根据历史数据进行预测，如果现实数据与历史数据差别很大，说明变异点出现了。但是这通常是针对一个属性，有可能与统计分布有关。

因此提出了自监督的对比编码预测的方法。自监督就是自己生成一种表征，而生成表征的方式就是根据对比编码，采取历史和未来窗口以及正负样本对的方法，设定损失函数，训练出一种编码方式。类似卷积神经网络，把时间窗口投影到一个编码（表征），利用这个编码的余弦相似度判断异常点。

经过实验，发现性能非常好（相比之前已知的优秀方法）。