CH5115 :- PROJECT

Vignesh Kumar - S

CH18B118

BAYESIAN ONLINE CHANGEPOINT DETECTION

1)(a)

GOAL:- The goal of the paper is simple. Given a set of data points with a known family of distribution with (here Normal) with unknown parameters (/unknown parameters) and given that there are abrupt variations in the generative parameters, the goal is to estimate/predict the occurence of these changepoints and estimate the ~~vals~~ generative parameters in each partitions using a "message passing algorithm" through Bayesian principles

Message passing algorithm is nothing but use of recursion principles ~~to~~ in finding changepoints in a "ONLINE" scenario. The paper talks about finding the probability distribution of the "run length" or time since last changepoints.

~~CONTRIBUTION~~ USEFULNESS:

In process control, EEG analysis, DNA segmentation, econometrics and wherever there is a need of 'online' changepoint detection. It need not be stated that this works good with offline as the online behaves like a superset problem.

CONTRIBUTIONS

Most Bayesian approaches to changepoint detection is offline and hence looks at data after its been collected. This following paper focusses on online prediction and further using it to generate distribution of the next "unseen data".

This is quite important requirement where abrupt changes occur like leaving a refridgerator open, a robot entering a completely new room etc, pressure change due to valve failure etc.

ALGORITHM EXPLANATION

OBJECTIVE

Let $\vec{D} \equiv \{x_1, x_2, \ldots x_t\}$ data points where we have observed till t data points.

OBJ:- To Find $p(x_{t+1} | \vec{D})$ and $p(r_t | \vec{D})$

NOTATIONS

- $\vec{D} \underset{\text{equivalent to}}{\equiv} \vec{x}_{1:t} \underset{\text{eq. to}}{\equiv} x_{1:t}$
- $r_t \equiv$ run length value at time t

→ Posterior predictive :- $p(x_{t+1} | x_{1:t})$

→ Run length posterior :- $p(r_t | x_{1:t})$

Note:- the notations may be used interchangeably in the solution written.

Let $x_t^{(r)}$ denote set of observations with run length r_t.

⊛ POSTERIOR PREDICTIVE

$$\boxed{p(x_{t+1} \mid \vec{x}_{1:t})} = \sum_{l=0}^{t} p(\underbrace{x_{t+1} \mid r_t = l, \vec{x}_{t-l:t}}_{\text{UPM predictive}}) \cdot \underbrace{p(r_t = l \mid \vec{x}_{1:t})}_{\text{RL posterior}}$$

UPM predictive

(Underlying Probabilistic model)

⊛ Runlength (RL) posterior :

$$\boxed{p(r_t \mid \vec{x}_{1:t})} = \frac{\sum_{r_{t-1}} \overbrace{p(x_t \mid r_{t-1}, \vec{x}^{(l)})}^{\text{UPM predictive}} \cdot \overset{\text{Change pt prior}}{p(r_t \mid r_{t-1})} \cdot \underbrace{p(r_{t-1}, \vec{x}_{1:t-1})}_{\text{MESSAGE}}}{\sum_{r'_t} p(r'_t, \vec{x}_{1:t})}$$

→ Notice that UPM predictive links both the formulation and needs to be evaluated first.

Let's define change point prior now :-

Let T be a discrete, nonnegative RV for current runlength. Let $f(T)$ denote that current run length probability for T. Let $S(T)$ be survival function; which is the probability T takes a value greater than T.

$$S(T) = P(T \geq T) = \sum_{r'=T}^{\infty} f(T')$$

$$H(T) \text{ (HAZARD FUNCTION)} = \frac{f(T)}{S(T)}$$

PHYSICAL MEANING :-

provided a change point has not occured at runlength T, what is the probability that it will occur at T.

$$\text{so } p(r_t | r_{t-1}) = \begin{cases} H(r_{t-1}+1) & r_t = 0 \\ 1 - H(r_{t-1}+1) & r_t = r_{t-1}+1 \\ 0 & \text{otherwise} \end{cases}$$

$\Rightarrow$ If T is geometric RV, $\underline{H(T) = p}$. [our case $\frac{1}{\lambda}$]

$$\Rightarrow p(r_t | r_{t-1}) = \begin{cases} \frac{1}{\lambda} & \text{if } r_t = 0 \\ 1 - 1/\lambda & r_t = r_{t-1}+1 \\ 0 & \text{o.t} \end{cases}$$

UPM predictive

$$\text{UPM} \equiv p(x_t | r_t, \vec{x}^{(\ell)})$$

On using a ~~stated~~ for conjugate prior in deriving it, UPM becomes a closed form.

Let the hyperparameters by ν_{prior} & χ_{prior} of the conjugate.

UPM predictive can be modelled by a closed form distribution with new parameters ν_t, χ_t. (which can be found as a update rule)
eg :- for exponential family models,

Posterior predictive has same exponential family form with parameters ;

$$\nu' = \nu_{prior} + N,$$
$$\chi' = \chi_{prior} + \sum_{n=1}^{N} u(\vec{x}_N)$$

In otherwords, we can compute ν', χ' to find posterior predictive without integration

① Algorithm

1. Set priors and initial conditions.

- We assume change point occurs at the first point itself. $\Rightarrow p(r_0 = 0, \vec{x} = NULL) = p(r_0) = 1$

$$\nu_1^{(0)} = \nu_{prior}$$
$$\chi_1^{(0)} = \chi_{prior}$$
$$t = 1$$

2. Observe new data x_t.

3. Compute UPM predictive probabilities. (for each possible run length value)
l 0 to $t-1$

$$\pi_{t-1}^{(l)} = p(x_t \mid \nu_{t-1}^{(l)}, \chi_{t-1}^{(l)})$$ → This closed form we know (because we used conjugate prior)

4. Compute growth probabilities.
Earlier, we have written expressions for $p(r_t, \vec{x}_{1:t})$

But r_t can either be zero (change point occurs at current time or $r_t = r_{t-1} + 1$.)

→ We call the first term as growth probability. Since $r_t = r_{t-1} + 1$, there is no summation involved.

$$p(r_t = l, \vec{x}_{1:t}) = P(r_{t-1} = l-1, \vec{x}_{1:t-1})\, \pi_{t-1}^{(l)} (1 - H(r_{t-1}))$$

5. Change point probability :- Probability that run length drops to zero.

$$P(r_t = 0, \vec{x}_{1:t}) = \sum_{r_{t-1}} p(r_{t-1}, \vec{x}_{1:t-1})\, \pi_{t-1}^{(l)} (H(r_{t-1}))$$

6.a Compute Evidence (Normalizing factor since we need

$$P(r_t \mid \vec{x}_{1:t}) = \frac{P(r_t, \vec{x}_{1:t})}{\boxed{\sum P(r_t, x_{1:t})}}$$

↓

Evidence

6.b and find $P(r_t \mid \vec{x}_{1:t})$

7. Update sufficient statistics. Earlier we have said that update rule would be found. Using MAPHY, or derivation, we can find the rule.

$$\nu_{t+1}^{(0)} = \nu_{prior}$$
$$\chi_{t+1}^{(0)} = \chi_{prior}$$
$$\nu_{t+1}^{(l+1)} = \nu_t^{(l)} + 1$$
$$\chi_{t+1}^{(l+1)} = \chi_t^{(l)} + u(x_t)$$

8. Perform prediction

$$P(x_{t+1} \mid D) = \sum_{r_t} P(x_{t+1} \mid x_t^{(r)}, r_t)\, P(r_t \mid x_{1:t})$$

9. Return to step 2.

SHORTCOMINGS

- Offline estimations are more accurate.
- A common weakness is that the online approaches tend to be less accurate in high-dimensional problems.
- Bayesian online CP detection method has "high sensitivity" due to the prior distribution on the number and location of the changepoints". There is enough reasons that we may choose to work with different priors and those completely influence the outcome.