CH5115 PROJECT

BY: S VISHAL
CH18B020

Q1)a) Goal of the Paper :

The main goal of the paper is to detect 'change-points' online in the generative parameters of a data sequence adopting a Bayesian approach. They make the assumption that the parameters before and after a change point are independent of each other. They approach this problem by estimating the length of the current "run"

Contribution of the paper and Why is it useful

~~The authors state that~~

The authors state that most Bayesian approaches prior (pun not intended) to their approach use retrospective and offline approach ~~to~~ change pt detection. If that is so, then this work ~~work constitutes~~ is of immense value.

Many applications essentially require online detection, let's say a plant where certain vital states of system are being estimated or consider a medical application where ~~the~~ some signal indicating the health of the vital system of a man is presented, immediate detection of abnormalities might prove to be necessary.

The authors also cite an example of brightness change detection in vision systems which ~~extent~~ ~~order~~ mandate the requirement of online detection.

Although frequentist approaches have been developed, Bayesian gives the additional advantages it always carries with it: i) use of prior information and a ii) probabilistic distribution of the required parameter. The latter advantage might prove to be useful in many scenarios.

The approach is versatile in the sense that ~~it~~ there is a clear cut separation between the implementation of the change-point algorithm and the implementation of the model. This means we can use this algorithms across ~~to~~ process with different models. This course project itself is a nice example ~~for~~ this:

→ In q1) we apply the algorithm to predict parameters (mean) of a distribution.

→ ~~In~~ q2)

→ In q1) we use it to find change points in which the parameter of the distribution (mean) is changing

→ In q2) we use it to ~~d~~ find change points in an AR(1) model where the variance of the endogeneous white-noise driving force chngs (variance)

So the algorithm works for 2 very different scenarios!

How the Algorithm Works.

To detect changepoints the authors try to compute the run length distribution (conditioned on the data)

Run length is defined as the number of datapoints which has been following a specific generative process. Alternatively, it is the time since the last change point occured

By its very definition $r_{t+1} = \begin{cases} r_t + 1 & \text{probability } p \\ 0 & 1-p \\ \text{any other value} & 0 \end{cases}$

where p is a valid probability

[Example of run length values]

So by predicting r_t we will know whether changepoint has occured at the datapoint x_t, because at change point r_t will be 0

Reqd. value : $P(r_t \mid x_{1:t})$ $\forall\, r_t$

(Note : $r_t \leq n$ where n is the index of the current data pt)

$\Rightarrow$ we need $\dfrac{P(r_t, x_{1:t})}{P(x_{1:t})}$

~~To get~~ Note that $P(x_{1:t}) = \sum\limits_{r_t} P(r_t, x_{1:t})$ ($\forall r_t$)

marginal.

So we need the joint distribution.

For that, after some neat simplifications (and assumption of Markov Property - for predicting r_t we need only r_{t-1}) the authors derive a recursive eqn of the form:

$$P(r_t, x_{1:t}) = \sum P(r_t \mid r_{t-1})\, P(x_t \mid r_{t-1}, x_t^{(r)})\, P(r_{t-1}, x_{1:t-1})$$

$P(r_t \mid r_{t-1})$ can be modelled using a hazard function (like geometric distribution)

Now we need the ~~term~~ term: $P(x_t | r_{t-1}, x_t^{(r)})$

We can write it ~~as~~ as some

$$P(x_{t+1} | x_{1:t}) = \int_{\Theta} P(x_{t+1} | \theta)\, P(\theta | x_{1:t})\, d\theta.$$

where $P(x_{t+1} | \theta)$ is the so called predictive -

and θ is the parameter(s) involved in

the DGP of x. And $P(\theta | x_{1:t})$ is the posterior

Here the author's assume exponential family

likelihoods and conjugate exponential prior $P(\theta | \chi, \nu)$

which ensures an exponential posterior.

~~Bec~~ This allows modularity wherein we can update

the ~~param~~ hyperparameters χ and ν indepdently

of run length predictions (thus, ensuring the delineation

of change point algorithm and model).

~~Thus,~~ • For initial conditions we ~~learn~~ the

paper considers 2 cases:

① $P(r_0 = 0) = 1$ (change point occured apriori before the first datum.) of data

② Some recent observations of a subset is used to get a prior over the initial run length as the normalized survival function.

$$P(r_0 = \tau) = \frac{1}{Z} S(\tau) \;;\; S(\tau) = \sum_{t=\tau+1}^{\infty} P_{gap}(g = t)$$

$Z \rightarrow$ normalising const

Thus, putting together all these pieces, the algorithm is complete! (we get a value for $P(r_t \mid x_{1:t})$)

PERFORMANCE & SCOPE FOR IMPROVEMENT

→ Three datasets are considered in the paper and the algorithm performs reasonably well in all 3 cases. It predicts the changepoints within a little error

→ It detects changes in both mean (NMR well log data) and variance (DOW JONES RETURNS data)

→ The algorithm is computationally very friendly (as long as $P(x_{t+1} \mid r_t)$ is evaluated to be a simple function / exact solution for the integral exists)

So, it can easily keep up pace with the arrival of new data.

→ However despite its very good performance in the project data as well as the original datasets the authors have used, there are some concerns.

(i) Underflow error: (<1)

As we keep on multiplying probabilities the values of the joint probabilities get closer and closer to zero ie eventually become so insignificant the computer can't evaluate resulting in the failure of the algorithm.

One alternative is to take logarithm of probabilities and work with those instead. But ~~other~~ ~~the~~ ~~and~~ more robust solutions are needed.

A closely related issue is that the $P(r_t \mid x_{1:t})$ vector (along r_t) is likely to be sparse, so some approximation or modification in the ~~space~~ implementation needs to be made to ensure optimal ~~ex~~ space usage (ie time as well since this is $O(N)$ complexity)

(ii) Initial conditions:

This is an issue which we face in Bayesian estimation i.e. choice of prior & hyperparameters.

One needs to test for the robustness of this algorithm in case there are errors in the prior or the model or the run length assumption ($P(r_0 = 0) = 1$)

→ Further Improvements could also be made for approximations of the predictive in case the exact solution is unavailable

As a closing remark, I would like to state that the algorithm is immensely useful and as long as the applications are chosen carefully and is definitely a big contribution for the field of online Bayesian changepoint detection