Supplement for ChangeDAR: Online Localized Change Detection for Sensor Data on a Graph

CCS CONCEPTS

• **Information systems** → *Data mining*;

KEYWORDS

Online, Localized, Change Detection, Sensor Data, Graph

1 TIME COMPLEXITY FOR TIME SERIES MODEL FITTING STEP

One of the key steps of our ChangeDAR-S algorithm computes the bitsave scores. Recall that the bitsave score from adding a change at time t at node v is:

$$\begin{aligned} \Delta_v(t) &= \text{Cost}(X_v(t-w), \cdots, X_v(t+w)) \\ &- \text{Cost}(X_v(t-w), \cdots, X_v(t-1)) \\ &- \text{Cost}(X_v(t), \cdots, X_v(t+w)) \end{aligned} \tag{1}$$

Also recall that the Cost terms in Eq. (1) are defined as the number of bits needed to encode the given sequence under various time series models. In this section, we consider the constant (mean), autoregression, and seasonal autoregression models and show that the bitsave scores at each time tick (e.g. time tick t in the above expression) can be computed in amortized constant time.

Note that computing Eq. (1) naively would require at least $O(w)$ time, since we would compute the modelling error at time ticks $t - w$ to $t + w$ individually. To speed this up, we use an online approach: note that when we move from time t to $t + 1$, the Cost expressions in Eq. (1) change only by shifting the window by one step. The fact that it mostly remains the same makes it possible to update this Cost expression to its new value in constant time.

Lemma 1. *Under the constant, autoregression, and seasonal autoregression models, the bitsave score at time t and node v in Eq. (1) can be computed in amortized $O(p^3)$ time, where p is the AR order (which we treat as constant).*

Proof. First consider an AR(p) model [1], with the form:

$$X_v(t) \sim \phi_0 + \sum_{i=1}^{p} \phi_i X_v(t-i)$$

Rewriting this in the using linear model form:

$$\underbrace{\begin{bmatrix} X_v(t-w) \\ X_v(t-w+1) \\ \vdots \\ X_v(t+w+1) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & X_v(t-w-1) & \dots & X_v(t-w-p) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_v(t+w-1) & \dots & X_v(t+w-p) \end{bmatrix}}_{\mathcal{X}} \underbrace{\begin{bmatrix} \phi_0 \\ \vdots \\ \phi_p \end{bmatrix}}_{\beta}$$

By standard formulas for linear regression [2], the residual sum of squares is

$$\text{RSS} = y^T y - y^T \mathcal{X}(\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T y$$

Note that when moving from time t to $t + 1$, $y^T y$ can be easily updated in constant time, by adding the square of the newly added entry of y and subtracting the square of the entry that is being removed. Similarly, $\mathcal{X}^T y$ can be updated in $O(p)$ time, while $\mathcal{X}^T \mathcal{X}$ can be updated in $O(p^2)$ time, after which computing $(\mathcal{X}^T \mathcal{X})^{-1}$ is $O(p^3)$.

Combining, this shows that RSS, and hence the $\text{Cost}(X_v(t-w), \cdots, X_v(t+w))$ term of Eq. (1), can be computed in $O(p^3)$ time. The other two terms of Eq. (1) can also be updated in $O(p^3)$ time in the same way: note that in all 3 cases, when moving from t to $t + 1$, we essentially have a linear regression which is changing by one sample, which we can update in the same way. In conclusion, for AR(p) models, we can compute Eq. (1) in amortized constant time. In addition, the constant (mean) model is just an AR(0) model, so also requires amortized constant time.

Finally, the seasonal AR SAR(p) model [1] with period m can also be expressed in linear model form with $p + 1$ coefficients in almost the same way:

$$\underbrace{\begin{bmatrix} X_v(t-w) \\ X_v(t-w+1) \\ \vdots \\ X_v(t+w+1) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & X_v(t-w-m) & \dots & X_v(t-w-m\cdot p) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_v(t+w-m) & \dots & X_v(t+w-m\cdot p) \end{bmatrix}}_{\mathcal{X}} \underbrace{\begin{bmatrix} \phi_0 \\ \vdots \\ \phi_p \end{bmatrix}}_{\beta}$$

Hence, using the exact same argument as for AR, the SAR(p) model also allows us to compute Eq. (1) in amortized constant time. □

REFERENCES

[1] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control.* John Wiley & Sons.

[2] Norman R Draper and Harry Smith. 2014. *Applied regression analysis.* Vol. 326. John Wiley & Sons.

CIKM 2018, October 2018, Turin, Italy

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4