Online Sparse Temporal Disaggregation

Matthew Speers

Supervisor: Luke Mosley

August 2020

Sparse Temporal Disaggregation

Temporal dissagregation

The process of deriving high frequency data from low frequency data through the use of related high frequency indicator series.

A **sparse** method attempts to simplify the model by discounting the effects from some indicator series.

$$\hat{\beta}_\lambda \in \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{2}||y - X\beta||_2^2 + \lambda||\beta||_1.$$

Converting to an online framework

What about when we have real time observations?

Ideally:

1. new data is received sequentially,
2. our model can be reliably updated to consider this new data,
3. this update is efficient to compute.

We can use online methods to achieve this.

Applications of Online Sparse Temporal Dissagregation

Weather forecasting: Forecasting models often need higher precision measurements than have been recorded. See [Barton et al., 2020].

Economic indicators: For example, using supermarket checkout data to analyse inflation.

Examples of online algorithms

Onine gradient descent (OGD): Uses gradient calculated from entire data set in update rule. See [Hazan et al., 2008].

Stochastic gradient descent (SGD): Estimates the gradient to use in update rule. See [Bottou, 2010].

Regularised dual averaging (RDA): Uses the whole regularisation term at each step to ensure sparsity. See [Xiao, 2009].

Forward-backsplitting (FOBOS): Interleaves analytical minimsation steps with subgradient steps. See [Singer and Duchi, 2009].

Alternating Direction Method of Multipliers (ADMM)

ADMM: Solves convex optimization problems by breaking them into easier to handle smaller pieces.

Take the minimisation problem:

$$\min ||(y - X\beta)||_2^2 + \lambda||z||_1$$

such that $\beta - z = 0$.

$$\mathcal{L}(\beta, z, \mu) = ||(y - X\beta)||_2^2 + \lambda||z||_1 + \mu^T(\beta - z) + \frac{\rho}{2}||\beta - z||_2^2$$

See [Suzuki, 2013] for more information.

Alternating Direction Method of Multipliers (ADMM)

The iterative update rules can be found to be:

$$
\begin{aligned}
\beta^{k+1} &= \arg\min_{\beta} \mathcal{L}(\beta, z^k, \mu^k) \\
z^{k+1} &= \arg\min_{z} \mathcal{L}(\beta^{k+1}, z, \mu^k), \\
\mu^{k+1} &= \mu^k + \beta^{k+1} - z^{k+1}.
\end{aligned}
$$

Problem: ADMM requires that all the data be stored in memory.

Possible Solution: OADM, see [Wang and Banerjee, 2013].

Online Alternating Minimisation Algorithm (OAM)

Proposed by [Li and Li, 2020].

OAM benefits compared to OADM:

- OADM depends heavily on new data, OAM does not,
- OAM is based on recursive least squares,
- OAM can give closed-form solutions for LASSO.

Pseudocode for OAM

Algorithm 1 OAM algorithm

Initialize with $P_0 = I$, $\alpha_0 = \alpha_{-1} = 0$, $\theta_0 = 0$.
for $n = 0, 1, 2, \cdots$ **do**
  $c_{n+1} = 1/(1 + \varphi_{n+1}^T P_n \varphi_{n+1})$;
  $d_n = \alpha_n - \alpha_{n-1}$;
  $g_{n+1} = c_{n+1} P_n \varphi_{n+1}$;
  $e_{n+1} = y_{n+1} - \varphi_{n+1}^T \theta_n$;
  $P_{n+1} = P_n - g_{n+1} g_{n+1}^T / c_{n+1}$;
  $\theta_{n+1} = \theta_n + g_{n+1} e_{n+1} + \mu P_{n+1} d_n$;
  $\alpha_{n+1} = \text{Soft}(\theta_{n+1}, \lambda/\mu)$.
end for

Combining ADMM & OAM

ADMM: Constrains the difference between β and z, but is memory intensive.

OAM: Faster, but doesn't consider difference between β and z.

Solution: Combine both algorithms into a 'New OAM' algorithm.

Data Analysis

Target Variable: Quarterly GDP

Indicator Series Used: VAT diffusion indices for both agriculture and construction, road traffic data at ports, monthly business surveys for production and services, retail sales

Variables Selected: VAT construction, MBS services, MBS production and XL vehicles

Tuning

For an ℓ_1-penalised least-squares optimisation problem:

$$\hat{\beta}_\lambda \in \arg\min_{b \in \mathbb{R}^p} \frac{1}{2}||y - Xb||_2^2 + \lambda||b||_1,$$

the selection of λ is very important.

Hedging parameter selection [Chretien et al., 2018] could be looked into as a way of doing this.

Other further work

Some more ideas:

- Test the combined algorithm using more data
- Find ways of choosing prioritising old/new data

References I

Barton, Y., Sideris, I. V., Germann, U., and Martius, O. (2020).
A method for real-time temporal disaggregation of blended radar–rain gauge precipitation fields.
Meteorological Applications, 27(1):e1843.

Bottou, L. (2010).
Large-scale machine learning with stochastic gradient descent.
In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer.

Chretien, S., Gibberd, A., and Roy, S. (2018).
Hedging parameter selection for basis pursuit.
arXiv preprint arXiv:1805.01870.

References II

Hazan, E., Rakhlin, A., and Bartlett, P. L. (2008).
Adaptive online gradient descent.
In *Advances in Neural Information Processing Systems*, pages 65–72.

Li, J. and Li, X. (2020).
Online sparse identification for regression models.
Systems & Control Letters, 141:104710.

Singer, Y. and Duchi, J. C. (2009).
Efficient learning using forward-backward splitting.
In *Advances in Neural Information Processing Systems*, pages 495–503.

References III

Suzuki, T. (2013).
Dual averaging and proximal gradient descent for online alternating direction multiplier method.
In *International Conference on Machine Learning*, pages 392–400.

Wang, H. and Banerjee, A. (2013).
Online alternating direction method (longer version).
arXiv preprint arXiv:1306.3721.

References IV

Xiao, L. (2009).
Dual averaging method for regularized stochastic learning and online optimization.
In *Advances in Neural Information Processing Systems*, pages 2116–2124.