Bayesian Spatio-Temporal approach to weather forecasting

Adam Pluck | Supervisor: Peter Flach

1 Updates

Week 7 - 12/11/18

Tasks completed:

The collection of data is complete with a total of 900 days (08/10/2018 - 22/04/2016) of weather data for each of the 8 locations. I've fully completed the programming for the framework to clean the data (e.g removing redundant data, ensuring missing data is dealt with etc...) and set it to manageable data. I've started to make frameworks to have better visual representations of the data too.

The data comes with a total of 33 different features of varying use. I've started to investigate the removal of correlated features. Since I have a lot of different features, I need to determine the effect multicollinearity (one feature can be produced relatively accurately by a combination of others) and negate it if present. This could be by dropping or combining features together

I've started to implement simple Gaussian process regression between two different features. I still need to read more around how to do a multiple input Gaussian regression (e.g using 5 different features' values to predict just a single one).

Problems and challenges:

Ensuring this project is computer science project rather than statistics - do I need to make sure I'm taking a Bayesian rather than Frequentist approach?

Task for the coming week

Two different avenues to explore: fuzzy Bayesian Network or a spatio-temporal kernel.

Potentially a convolution or combination of a spatial kernel and a termporal kernel

Read more papers about using a spatio-temporal kernel in the Gaussian process regression and look at how it is implemented.

Start to write the introduction

2 Project Outline

Classical weather forecasting methods generally consist of numerical weather prediction (NWP). This is the way in which computer simulations, in conjunction with sophisticated mathematical and physical models use the oceans and atmosphere to predict future weather conditions. These models do, however, come with an ever increasing complexity. Even with the most powerful supercomputers, "forecast skill" drops off after just the 6th day. These models are also fundamentally flawed by their intrinsic sensitivity and underlying chaotic nature further reducing efficacy as they produce further in time forecasts.

One way to overcome this is to take a data driven approach using machine learning techniques. This strategy is in its infancy when it comes to weather forecasting. Large meteorological entities such as the Met Office are, however, exploring methods such as pure Gaussian Process regression or a hybrid of both GP and NWP. These methods generally forecast for each city or region, assuming independence between each city. Although this is done with some success, particularly exceeding in the long term (6days +) forecasts, it could be argued that we are losing accuracy from that assumption. Weather features between cities are most likely correlated in some way. Hence I aim to explore the idea of taking a Bayesian Spatio-Temporal approach to weather forecasting, utilising both trends in time-series data of different weather features and also correlation in space between different locations of the United Kingdom. The ultimate goal is to produce a model that predicts precipitation in a given city using historical data from a variety of cities possibly as in Figure 1.

- Edinburgh(55.953251, -3.188267) DONE
- Glasgow(55.864239, -4.251806) DONE
- Newcastle(54.978252 , -1.617780) DONE
- Liverpool(53.408371,-2.991573) DONE
- Norwich(52.628101, 1.299350) DONE
- Bristol(51.454514, -2.587910) DONE
- London (51.509865, -0.118092) DONE
- Exeter(50.721802, -3.533620) DONE
  taking their lat/long readings from LatLong.net

The data-set will be taken from the company "Dark Sky" using their own API. I will be sampling daily weather data for the past few years. Many different weather features are provided, so I will need to be careful to overcome the curse of dimentionality and reduce the number of features down to a useful amount. We also want to remove or "merge" the highly correlated features to improve the model.

Although Spatio-Temporal forecasting research is also still in its infancy with regards to weather forecasting; there has been some serious progress in the field in general. From

My evaluation will be ... cross validation of some sort?

3 Project Plan - Predicting average temperature each day for the next 7 days

- Preface about the impact of weather on different departments (economy, agriculture, transport...) maybe 1 pageish
- Talk about classic weather forecasting methods 1 Page
- Talk about machine learning re forecasting
- Talk about how spatio-temporal forecasting may work 1 page
- Potential Methods e.g Bayesian Hierarchical Modelling
- Actual proposed model outline
- Programming:
  - Data collection and processing
  - Feature selection (e.g solving multicollinerity, removing corellated, PCA, etc...
  - Implementation of the model
  - Training the model
  - Testing the model
  - Evaluation of model
- Conclusion and evaluation of project

4 Very Rough Ideas

Need to use sine and cosine curves to removal the seasonality of the data.

Notation station:

Let x_i^j be the jth value from feature i.

Z(s,t)

If it were linear it would look something like this:

$$\mathbf{y_j} = \beta_0 + \mathbf{x}_1^j\beta_1 + \mathbf{x}_2^j\beta_2 + ... + \mathbf{x}_n^j\beta_n + \epsilon$$

where x_i is the ith feature

Or if we use Einstein summation notation:

$$\mathbf{Y} = \mathbf{x}_i\beta_i + \beta_0 + \epsilon$$

Potentially Gaussian Regression for historical data at all locations, then use kriging to generate a more accurate version considering spatial data.

Multicollinearity

5 Paper Reviews

5.1 Paper 1

Spatiotemporal Prediction of Ambulance Demand using Gaussian Process Regression

https://arxiv.org/pdf/1806.10873.pdf

5.2 Paper 2

"Estimation and prediction of weather variables from surveillance data using spatio-temporal Kriging"
https://upcommons.upc.edu/bitstream/handle/2117/112695/dalmau_17_kriging.pdf

Summary of report:

- AIR TRAFFIC CONTROL
- NWP = numerical weather predictions
- "Kriging is a geostatistical interpolation technique to create short- term weather predictions from scattered weather observations "
- Most interpolation methods estimate variable as weighted sum of observations from location [U+F0E0] further away = lower weight
- Geostatistical – data-driven statistical models that consider correlation between data e.g Kriging
- Kriging interpolation provides best estimate of variable Z at unmeasured location x from set of surrounding data points
- Two different methods :
  - do temporal regression for all cities then spatio regression to take into account the links between cities. i.e predict value for all cities, then use those values in kriging to predict a value for a single city.
- Two types of spatio-temporal variogram:
  - separable = combination of purely spatial and purely temporal variograms
  - non-separable. =. "more flexible to handle..." (above equation 17) spatio-temporal UK (UK-ST) variant

5.3 Paper 3

Improved space–time forecasting of next day ozone concentrations in the eastern US

http://www.soton.ac.uk/~sks/research/papers/sahuyipholland.pdf

https://www.sciencedirect.com/science/article/pii/S2211675313000195

Spatio-temporal modeling for real-time ozone forecasting

Sahu, S.K., Yip, S., Holland, D.M., 2009a. A fast Bayesian method for updating and forecasting hourly ozone levels. Environmental and Ecological Statistics 18, 185–207. Sahu, S.K., Yip, S., Holland, D.M., 2009b. Improved space–time forecasting of next day ozone concentrations in the eastern US. Atmospheric Environment 43, 494–501.

5.4 Paper 4

"A Bayesian spatio-temporal model for forecasting Anaplasma species seroprevalence in domestic dogs within the contiguous United States"

https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0182028&type=printable

- Looking at ticks in dogs across USA
- Show correlation between features to backup point of spatial correlation
- They use bayesain hierarchical spatio-temporal regression model, autocorrelated random effects are utilized to account for the spatio and temporal dependence.
- areal units = places —-> use this term
- $Y_s(t)$ is positive test for county s at year t
- they have a term for "spatio-temporal random effects used to account for the spatial and temporal dependence"
- conditional autoregressive model (CAR) captures the spatial dependence
- "markov chain monte carlo posterior sampling algorithm"

5.5 Paper 5

A Bayesian hierarchical spatio-temporal model for extreme rainfall in Extremadura (Spain)

5.6 Paper 6

Bowman DD, Liu Y, McMahan CS, Nordone SK, Yabsley MJ, Lund RB. Forecasting United States heartworm Dirofilaria immitis prevalence in dogs. Parasit Vectors. 2016; 9(1):540.

https://doi.org/10.1186/s13071-016-1804-y

5.7 Paper 7

"A Probabilistic Approach for Weather Forecast using Spatio-temporal Inter-relationships among Climate Variables"