Methods in Multivariate Time Series Analysis

Time Series with R/Python Capstone Proposal

Chang Shen
Advisor: Dhafer Malouche

1 Objective

A Multivariate Time Series is n time series within the same time frame,that is for any time t, $Y_t = (y_{1,t}, \cdots, y_{n,t})$. The analysis of Multivariate can be challenging since when modeling $y_{1,t}$, we need to include not only prior $\{y_{1,t-k}, \forall k\}$, but also interactions and latent information within the group of variables $\{y_{2,t-k}, \cdots, y_{n,t-k}, \forall k\}$. At the same time, we should keep the model as simple as possible to prevent the heavy calculation and over-fitting caused by too many parameters. We attempt to find a method to extract meaningful patterns and build general modeling frame for Multivariate Time Series data.

2 Datasets

The data set sourced from data mining GUI R package homepage **https://rattle.togaware.com/**. We use daily weather observations from numerous Australian weather stations(different cities) every day from 2008 to 2019, 142194 rows in all. Except the basic rainfall, the data sets include other 23 weather variables(time series), like the minimum and maximum temperature, class A pan evaporation, the direction and speed of wind gust, atmospheric pressure, etc.

3 Approach

3.1 EDA(Exploratory Data Analysis)

- In this part, I will apply some general data analysis methods like descriptive statistics summary, outliers, and missing values recognition and manipulation. After such process, our dataset would be trimmed into more a precise and reliable one. For our weather datasets, we can also conduct a comparison between different cities and time series classification among cities.
- As for the correlation analysis, we can implement different approaches like general ones, nonparametric ones and define Cross covariance matrix especially for Multivariate Time Series.

3.2 Model Variables Selection—Dimension Reduction

- In reality, our response variable can be related to multiple predictor variables. Some of the predictor variables can be uninformative or with information highly-overlap with other variables, which meaning we can use fewer variables and reach the same modeling accuracy. We can combine PCA, LASSO and empirical knowledge to decide final variables in use.
- I also intend to explore some time series related dimension reduction methods from former researchers.

3.3 Multivariate Time Series(MTS)

- Implement the vector autoregression (VAR) and VARMA model to predict the target variable with stationary hypothesis testing (if not, use the back shift operator to achieve stationary) and parameter estimation.
- Compare the VAR and VARMA model to other methods like LSTM(machine learning), Hidden Markov Models (HMM), State Space Methods and univariable ARIMA model.

4 Timeline

- (9/30) Prepare the outline for this project and literature reviews
- (10/17) Do baseline multivariate Analysis(realize the modelling R functions based on the related paper)
- (11/10) Find some insights, summary the result, improve the model and do method comparison
- (11/25) Implement the results visualization, attempt to build a R package and prepare for class presentation

5 Possible Issues and Expected Results

1.Possible problems I foresee with this project is my lack of coding technique to realize a complex algorithm in the paper or the predicted results can be meaningless even no better than univariate time series analysis .
2.Expect result: VAR method is superior to other methods in some conditions and improve the MTS analysis methods.

Note: I will provide an original R code and six pages paper with Abstract, Introduction, Methodology, Results and Discussion, Conclusion

References

[1] Rattle: A Graphical User Interface for Data Mining using R. https://rattle.togaware.com/

[2] tsdl: Time Series Data Library https://pkg.yangzhuoranyang.com/tsdl/index.html

[3] Multivariate Time Series Analysis in R http://past.rinfinance.com/agenda/2013/talk/RueyTsay.pdf

[4] Multivariate Time Series https://faculty.washington.edu/ezivot/econ584/notes/multivariatetimeseries.pdf

[5] The tutorial of fpp package https://otexts.com/fpp2/what-can-be-forecast.html

[6] A Multivariate Time Series Guide to Forecasting and Modeling https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide