ADS 506 Module 2 Exercises: Chapter 4

Zachariah Freitas

2022-11-05

Chapter 4: Forecasting Methods: Overview (Pages 77-78) #1 and 2

1. A large medical clinic would like to forecast daily patient visits for purposes of staffing.

   a. If data are available only for the last month, how does this affect the choice of model-based vs. data-driven methods?

   Answer: Having limited data to use, we are better served trying model-based methods. This doesn't mean data-driven methods won't work, some will, but most tend to work better where there is more data available.

---

   b. The clinic has access to the admissions data of a nearby hospital. Under what conditions will including the hospital information be potentially useful for forecasting the clinic's daily visits?

   Answer: If the hospital admissions data is correlated to the clinic patient vists data, then it could be useful. Especially if the data is lag correlated. This way we have data that we can use in our forecast at the point of forecast.

---

   c. Thus far, the clinic administrator takes a heuristic approach, using the visit numbers from the same day of the previous week as a forecast. What is the advantage of this approach? What is the disadvantage?

   Answer: The main advantage is that It is simple, straight forward and allows for easy automation. They require less user input/bias and are considered advantageous when model assumptions are likely to be violated. The main disadvantage is that it may be too simplistic and not capture global seasonality, trends, or cyclical patterns.

---

   d. What level of automation appears to be required for this task? Explain.

   Answer: This probably should be automated because it is not a one time request since staffing decisions are probably made on a weekly to biweekly basis due to human resource limitations like union contracts and legal staffing requirements.

---

e. Describe two approaches for improving the current heuristic (naive) forecasting approach using ensembles.

Answer: The above heuristic approach uses t_{-7} to predict t. We could extend this by using a naive drift model for the last two weeks and add the two forecasts together and divide by two. This could be our first ensemble method.

For our second ensemble method, we could take the average visits for the last seven days and weight that lightly with something like 20% and add that to our first naive forecast and weight that more heavily like 80%. Thus taking into account the average of the last seven days with the previous weeks value. The two models have different time horizons.

---

2. The ability to scale up renewable energy, and in particular wind power and speed, is dependent on the ability to forecast its short-term availability. Soman et al. (2010) describe different methods for wind power forecasting (the quote is slightly edited for brevity):

   Persistence Method: This method is also known as "Naive Predictor". It is assumed that the wind speed at time $t + \delta t$ will be the same as it was at time t. Unbelievably, it is more accurate than most of the physical and statistical methods for very-short to short term forecasts.

   Physical Approach: Physical systems use parameterizations based on a detailed physical description of the atmosphere.

   Statistical Approach: The statistical approach is based on training with measurement data and uses difference between the predicted and the actual wind speeds in the immediate past to tune model parameters. It is easy to model, inexpensive, and provides timely predictions. It is not based on any predefined mathematical model and rather it is based on patterns.

   Hybrid Approach: In general, the combination of different approaches such as mixing physical and statistical approaches or combining short term and medium term models, etc., is referred to as the hybrid approach.

   a. For each of the four types of methods, describe whether it is model-based, data-driven, or a combination .

   Answer:

   Persistence Method: This is a data-driven approach.

   Physical Approach: This is a model-based approach. I say this cause I think the patterns are global in nature and because we are using parameterizations.

   Statistical Approach: This is a model-based approach.

   Hybrid Approach: This is a combination approach.

---

   b. For each of the four types of methods, describe whether it is based on extrapolation, causal modeling, correlation modeling, or a combination.

   Answer:

   Persistence Method: This approach is based on extrapolation.

   Physical Approach: This approach is based on causal modeling.

   Statistical Approach: This approach is based on correlation modeling.

   Hybrid Approach: This approach is based on a combination of all the methods above.

---

c. Describe the advantages and disadvantages of the hybrid approach.

Answer:

1. Advantages
    1. It reduces the variance of forecast errors. When forecast errors of the combined models are negatively or lowly correlated, we get an improvement of forecast errors.
    2. Forecast is less susceptible to bias. Different models will have different biases associated with them and combining them allows us to increase reliability of the forecast.
    3. It guarantees that the resulting forecast will not be the worst forecast.
    4. The absolute error will be smaller than the averag of the componet forecast errors.
2. Disadvantages
    1. Increased Costs
    2. Need analysts who are familiar with a number of methods.
    3. Need to ensure that a pre-determined rule for combining is agreed upon. Otherwise, people can find a forecast to suit their biases.
