Project 2 - Report

Phil Bailey

October 27, 2018

Overview

The goal of this project was to predict Walmart sales for individual departments within various stores. Essentially we are doing times-series forecasting and need to determine a model that will provide both accuracy and enough performance to handle the large volume of depts/stores we are working with. I tested various models from the **forecast** package in **R**, including:

- naive
- snaive
- tbats
- snaive+stlf
- tslm+stlf

NOTE: though all methods above were tested, only naive, snaive, and tslm+stlf are reported in final output with WMAE scores.

In the end, the best model proved to be **a combination of tslm+stlf**. This model produced a WMAE of 1596.

Technical Details

Initially naive and snaive (i.e. seasonal naive) were tested. They both run very fast, with naive giving poor accuracy and snaive giving a respectable WMAE of 1880. I attempted the tbats model next, but the performance was to slow to complete on my PC, so gave up on this one. Finally I wanted to try stlf as it is known to provide good results in general. The problem is that stlf cannot be used until our train data has 2 full seasons. So I needed to create a kind of 'mixed' model where we use some other method until we get to 2 full seasons and then switch over to stlf. I tried both snaive and tslm (a linear forecast model) as the supplement to stlf. They both worked well, but tslm+stlf produced the best results with a WMAE of 1596.

System Specs & Runtime

System Specs: **Windows 10, 2.7GHz, 8GB Ram**

Processing time to complete the 3 models over 10 folds: **50 minutes**

Final Results

Table 1: WMAE

naïve	snaive	tslm_w_stlf
2079	2262	2042
2589	1787	1440
2254	1779	1435
2823	1716	1597
5156	2400	2328
4218	1697	1674
2270	2087	1613
2144	1750	1336
2221	1720	1274
2372	1681	1226

Table 2: Mean WMAE over 10 folds

naive	snaive	tslm_w_stlf
2813	1888	1596