Bayesian Networks for Time Series Forecasting: A Case Study

Gordon Rios, Antonino Marvuglia, Rick Wallace

Cork Constraint Computation Centre (4C)
University College Cork, Ireland

S4 ENVISA Colloquium 2010 – June 21, 2010
Forecasting Problem

Perform long range or n-step ahead forecasts on one, two or more datasets of a selection of 6 datasets (each containing 11 time series) on transportation data as accurately as possible, using methods from computational intelligence and applying a consistent methodology.

The data consists of 6 datasets with 11 time series with different time frequencies, including yearly, quarterly, monthly, weekly, daily and hourly transportation data.
Our Objectives

• Automate N-Step Forecasts (i.e. 12 months ahead)
• Maintain high “interpretability”
• Adapt to wide range of series types (e.g. nonstationary, nonlinear, etc.)
• Explore use of Bayesian Networks in performing forecasts (substitute “automated” for “consistent”)
Dataset C: Monthly Time Series

#1 (n = 48)

#2 (n = 48)

#3 (n = 198)

#4 (n = 172)

#5 (n = 118)

#6 (n = 118)

#7 (n = 118)

#8 (n = 57)

#9 (n = 227)

#10 (n = 132)

#11 (n = 228)
Model Identification, Stationarity, Linearity

<table>
<thead>
<tr>
<th>Name of the time series</th>
<th>ETS</th>
<th>Value of the KPSS test</th>
<th>Stationary</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>(M,N,M)</td>
<td>0.174</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>C2</td>
<td>(M,N,M)</td>
<td>0.130</td>
<td>YES</td>
<td>YES++</td>
</tr>
<tr>
<td>C3</td>
<td>(M,N,M)</td>
<td>0.191</td>
<td>YES</td>
<td>YES++</td>
</tr>
<tr>
<td>C4</td>
<td>(A,A,N)</td>
<td>0.407</td>
<td>NO**</td>
<td>NO</td>
</tr>
<tr>
<td>C5</td>
<td>(A,N,A)</td>
<td>0.032</td>
<td>YES</td>
<td>YES+</td>
</tr>
<tr>
<td>C6</td>
<td>(M,N,M)</td>
<td>0.028</td>
<td>YES</td>
<td>YES++</td>
</tr>
<tr>
<td>C7</td>
<td>(A,N,A)</td>
<td>0.115</td>
<td>YES</td>
<td>NO++</td>
</tr>
<tr>
<td>C8</td>
<td>(M,N,M)</td>
<td>0.079</td>
<td>YES</td>
<td>YES++</td>
</tr>
<tr>
<td>C9</td>
<td>(A,Ad,N)</td>
<td>0.273</td>
<td>NO**</td>
<td>NO</td>
</tr>
<tr>
<td>C10</td>
<td>(M,Ad,M)</td>
<td>0.127</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>C11</td>
<td>(A,N,A)</td>
<td>0.160</td>
<td>YES</td>
<td>NO</td>
</tr>
</tbody>
</table>

ETS := (Error, Trend, Seasonality)

Linearity Tests:
+ DVV & (REV | C3)
++ DVV & REV & C3

KPSS Reject stationarity
* p < 0.05
** p < 0.01
Partial Auto Correlation Function
Monthly Plots

Series No. 1

Series No. 2

Series No. 3

Series No. 4

Series No. 5

Series No. 6

Series No. 7

Series No. 8

Series No. 9

Series No. 10

Series No. 11
Monthly Plots: \texttt{diff (t)}
Comparative Forecasting Methods

- ARIMA & Exponential Smoothing State Space
  - Model Identification via (Hyndman)
  - Parameter Estimation
  - forecast package in R (Hyndman)
  - N-Step Forecasts: cascade

- Neural Network
  - Single Hidden Layer
  - Model Training / Testing
  - nnet package in R (Ripley)
  - N-Step Forecasts: cascade

- Bayesian Network for N-Step Forecasts
  - Structure learning (PEBL, bnlearn in R, Banjo)
  - Parameter learning (bnlearn)
  - Inference (bnlearn for fitted values, forecasts???)
Forecast Models

State Space & Neural Network: $n$ step cascade of one step ahead forecasts ...

\[ \tilde{y}_t = E\left[ y_t \mid y_{t-1} \cdots y_{t-12} \right] \]
\[ \tilde{y}_{t+n} = E\left[ y_{t+n} \mid \tilde{y}_{t+n-1} \cdots \tilde{y}_t, y_{t-1} \cdots y_{t-12+n} \right] \]

Bayesian Network: joint estimation of $n$ periods ...

\[ \{ \tilde{y}_t \cdots \tilde{y}_{t+n} \} = E\left[ y_t \cdots y_{t+n} \mid y_{t-1} \cdots y_{t-12} \right] \]

Choice of 12 periods of conditioning is based on this monthly task but otherwise is arbitrary ... also “month of year” is excluded though it’s generally an important factor ...
Fitted Results: Last 12 Months
Fitted Accuracy: Mean Absolute Scaled Error (MASE)

One step ahead forecasts ....

\[ q_t = \frac{y_t - \hat{y}_t}{\frac{1}{n-1} \sum_{i=2}^{n} |y_i - y_{i-1}|} \]

\[ MASE = mean(|q_t|) \]

For legacy purposes forecast competition required symmetric Mean Absolute Percentage Error (sMAPE) but good case is made to prefer MASE (Hyndman & Koehler)

<table>
<thead>
<tr>
<th>Name of the time series</th>
<th>BAYES</th>
<th>ETS</th>
<th>NNET</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.92</td>
<td>0.54</td>
<td>0.03</td>
<td>25</td>
</tr>
<tr>
<td>C2</td>
<td>0.49</td>
<td>0.36</td>
<td>0.03</td>
<td>25</td>
</tr>
<tr>
<td>C3</td>
<td>0.86</td>
<td>0.58</td>
<td>0.81</td>
<td>175</td>
</tr>
<tr>
<td>C4</td>
<td>2.017</td>
<td>0.85</td>
<td>0.92</td>
<td>149</td>
</tr>
<tr>
<td>C5</td>
<td>0.67</td>
<td>0.44</td>
<td>0.28</td>
<td>95</td>
</tr>
<tr>
<td>C6</td>
<td>1.20</td>
<td>0.34</td>
<td>0.53</td>
<td>95</td>
</tr>
<tr>
<td>C7</td>
<td>0.87</td>
<td>0.78</td>
<td>0.40</td>
<td>95</td>
</tr>
<tr>
<td>C8</td>
<td>0.61</td>
<td>0.49</td>
<td>0.11</td>
<td>34</td>
</tr>
<tr>
<td>C9</td>
<td>8.39</td>
<td>1.03</td>
<td>0.86</td>
<td>204</td>
</tr>
<tr>
<td>C10</td>
<td>0.99</td>
<td>0.35</td>
<td>0.48</td>
<td>109</td>
</tr>
<tr>
<td>C11</td>
<td>1.03</td>
<td>0.60</td>
<td>0.64</td>
<td>205</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>1.64</strong></td>
<td><strong>0.58</strong></td>
<td><strong>0.46</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td><strong>0.92</strong></td>
<td><strong>0.54</strong></td>
<td><strong>0.48</strong></td>
<td>-</td>
</tr>
</tbody>
</table>
N-Step Forecasts

holdout last 12 months: ets
N-Step Forecasts

*holdout last 12 months: ets & nnet*
Structure Learning: bnlearn and Banjo

bnlearn
http://www.bnlearn.com/

Banjo Version 2.2.0
High scoring network, score: -5260.3243
Project: nge1 series
User: gordon
Dataset: C.4.dat

http://www.cs.duke.edu/~amink/software/banjo/
Bayesian Network: Parameter Learning

- Conditional model selection impacts both structure learning and parameter learning
- Discrete or Gaussian
- Needed hybrid in order to capture discrete “month of year” effects but not implemented in package \texttt{bnlearn}
Bayesian Network: Inference

- Approximation by Markov Chain Monte Carlo which is “more art than science” while exact inference is intractable (Koller & Friedman 2009)

Only tractable for tree or forest networks (see Inference in Bayesian Networks by Scott Davies and Andrew Moore) but do we have that for our models?
Well, no ... for example (series #4)

Month Y11 drastically fails (even the in-sample fitted values) since backwards inference would be required ... possible but near total absence of good tools ...
Acknowledgements

This work was supported under Framework Six within CREEDS (Constraint Reasoning Extended to Enhance Decision Support) supported by the European Commission, Marie Curie Transfer Of Knowledge (TOK) scheme.

http://4c.ucc.ie/creeds
Questions?

Automated N-Step Univariate Time Series Forecasts with Bayesian Networks

Gordon Rios, Antonino Marvuglia, Richard Wallace

Cork Constraint Computation Centre (4C), University College Cork (UCC), IRELAND

g.rios@4c.ucc.ie, a.marvuglia@4c.ucc.ie, r.wallace@4c.ucc.ie