

Review report of a final thesis

Czech Technical University in Prague

Faculty of Information Technology

Student: Bc. Ruhi Ravichandran
Reviewer: Ing. Jan Motl
Thesis title: Predictive models in Logistics: Comparison of traditional time series techniques with an artificial neural network model approach
Branch of the study: Knowledge Engineering (Master)

Date: 31. 5. 2015

<p><i>Evaluation criterion:</i></p> <p>1. Difficulty and other comments on the assignment</p> <p><i>Criteria description:</i> Characterize this final thesis in detail and its relationships to previous or current projects. Comment what is difficult about this thesis (in case of a more difficult thesis, you may overlook some shortcomings that you would not in case of an easy assignment, and on the contrary, with an easy assignment those shortcomings should be evaluated more strictly.)</p> <p><i>Comments:</i> The topic fits the scope of FIT as FIT moved away from developing general-purpose classifiers toward more focused topics like recommendation or image processing.</p> <p>Since education about time series at FIT ends with Poisson processes, I find the topic difficult.</p>	<p><i>The evaluation scale: 1 to 5.</i></p> <p>1 = extremely challenging assignment, 2 = rather difficult assignment, 3 = assignment of average difficulty, 4 = easier, but still sufficient assignment, 5 = insufficient assignment</p>
<p><i>Evaluation criterion:</i></p> <p>2. Fulfilment of the assignment</p> <p><i>Criteria description:</i> Assess whether the thesis meets the assignment statement. In Comments indicate parts of the assignment that have not been fulfilled, completely or partially, or extensions of the thesis beyond the original assignment. If the assignment was not completely fulfilled, try to assess the importance, impact, and possibly also the reason of the insufficiencies.</p> <p><i>Comments:</i> Experiment with different predictive models and measure their quality. - The predictive ability of the models was not assessed by testing error.</p> <p>Prepare informative features and evaluate their predictive power. Focus on historical sales data and various factors affecting the sales, e.g., promotions, seasonal purchasing trends. - The analysis was performed only with the historical sales data because data about promotions and discounts were not provided by the data supplier. Nevertheless, generally available predictors as daily temperature or holiday days could have been used to fit the sales, as some products, like sun creams, are truly seasonal commodities.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled</p>
<p><i>Evaluation criterion:</i></p> <p>3. Size of the main written part</p> <p><i>Criteria description:</i> Evaluate the adequacy of the extent of the final thesis, considering its content and the size of the written part, i.e. that all parts of the thesis are rich on information and the text does not contain unnecessary parts.</p> <p><i>Comments:</i> The thesis assignment form is not present in the electronic copy of the thesis.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = meets the criteria, 2 = meets the criteria with minor objections, 3 = meets the criteria with major objections, 4 = does not meet the criteria</p>
<p><i>Evaluation criterion:</i></p> <p>4. Factual and logical level of the thesis</p> <p><i>Criteria description:</i> Assess whether the thesis is correct as to the facts or if there are factual errors and inaccuracies. Evaluate further the logical structure of the thesis, links among the chapters, and the comprehensibility of the text for a reader.</p> <p><i>Comments:</i> There were some minor mistakes like: - The sum in equation 1.13 on page 24 should be over "i", not over "j". - The square operator in the definition of RMSE on page 26 should be inside of the brackets, not outside the brackets. - The null hypothesis for augmented Dickey-Fuller on page 35 should be for stationary series, not for non-stationary series.</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>90 (A)</p>
<p><i>Evaluation criterion:</i></p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p>

5. Formal level of the thesis

90 (A)

Criteria description:

Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 12/2014, Article 3.

Comments:

Omissions:

- The descriptions for equations 1.16 and 1.17 are identical (page 26).
- It is not defined in equation 1.16, what is variable "t".
- The value of r (correlation coefficient) always lies between -1 and 1 with negative values indicating... (the keyword "inclusive" is missing, page 28).

Inconsistency in spacing after commas:

- ...they have considered, Holt- Winters, ARIMA, and Regression models.

Inconsistency in spacing after dots:

- ...arrivals. The data shows...
- ...equally difficult to interpret. Hence my advice...

Inconsistency in spacing before dots:

- ...it was difficult to interpret them .
- ...affect on the accuracy of the forecasts.

Inconsistency in spacing before colons:

- Reduce the lead time : Lead...
- Inventory inaccuracy: Incorrect...

Inconsistency in formatting:

- The mean function on page 26 is once written with cursive and once with regular font.

Missing words:

- Goods that are frequently purchased by the consumer, example salt, sugar, bread, eggs.

Wrong order of words:

- These items are never stocked beyond the required quantity as they easily available and are inexpensive.

Missing dots:

- ...biscuits, etc the price...

Wrong casing:

- ...ariel detergent 30 wash blue...
- Some examples are delphi technique...

Typography:

- The "Hat" symbol (circumflex) is typed with caret symbol (page 19).

Grammar errors:

- An multi layer perceptron...
- ...we can also computing some statistics...
- Figure 1.12 shows different scatterplots depiction correlation between the carbon...
- ...we could see some seasonality in the data, and hence we decided to...

On the other end, I enjoyed a lot stylistic phrases like:

- ANN was pitted against Box-Jenkins method.
-it was a herculean task to collect the data in the first place.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

6. Bibliography

93 (A)

Criteria description:

Evaluate the student's activity in acquisition and use of studying materials in his thesis. Characterize the choice of the sources. Discuss whether the student used all relevant sources, or whether he tried to solve problems that were already solved. Verify that all elements taken from other sources are properly differentiated from his own results and contributions. Comment if there was a possible violation of the citation ethics and if the bibliographical references are complete and in compliance with citation standards.

Comments:

If you copy-paste definition of MAPE and scaled errors (on page 27) from <https://www.otexts.org/fpp/2/5> and move the inline reference to the end of the paragraph, you should also remove the reference to the reference ("They proposed...") as it is unclear who is "they".

On page 8 it is written: "Also in New Zealand it (FMCG) is regarded as the largest sector accounting for 5% of Gross Domestic Product (GDP)[29]" with reference to a web journal "Economy Watch". However, based on the official New Zealand statistic manufacturing in 2014 accounted for 11.1% of GDP. And the share of manufacturing on GDP was declining for the last 30 years (<http://www.treasury.govt.nz/economy/overview/2015/09.htm>).

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

7. Evaluation of results, publication outputs and awards

70 (C)

Criteria description:

Comment on the achieved level of major results of the thesis and indicate whether the main results of the thesis extend published state-of-the-art results and/or bring completely new findings. Assess the quality and functionality of hardware or software solutions. Alternatively, evaluate whether the software or source code that was not created by the student himself was used in accordance with the license terms and copyright. Comment on possible publication output or awards related to the thesis.

Comments:

Only training error and AICc (and it's alternatives like AIC and BIC) were measured. The testing error was not measured! Hence we actually do not know how accurate the forecasts are. All we know is how the methods can fit the time series.

Also, the error was measured only for the selected (representative) series. But no evidence is given that the representative series cover all the series in the data set. Hence someone could argue that the examined series were cherry picked. Furthermore, the proportional representation of the individual categories in the data set is not given. Therefore error over the whole data set cannot be even estimated.

Beside that, I don't have a reason to deviate from the standard 70-79 point assessment.

Evaluation criterion:

No evaluation scale.

8. Applicability of the results

Criteria description:

Indicate the potential of using the results of the thesis in practice.

Comments:

I would like to see a comparison of the evaluated methods with the predictions from Inventoro service (available from inventoro.com), which is provided by a Czech company focusing exclusively on demand forecasting and inventory management.

It would most likely take half a year to deploy the forecast methods benchmarked in the thesis as details like a workaround for too short time series would have to be developed (the thesis advises against time series shorter 3 years; however, over 50% of all the examined series are shorter 3 years).

Evaluation criterion:

No evaluation scale.

9. Questions for the defence

Criteria description:

Formulate any question(s) that the student should answer to the committee during the defence (use a bullet list).

Questions:

- 1)
 - a) Over the past decade a number of statistical techniques such as Holt-Winters, Exponential smoothing...
 - b) Winter's model performs better than Holt's model to forecast aggregate retail sales[26].
-> What is the difference between Winter's model, Holt's model and Holt-Winters technique?
- 2)
 - a) Over the past decade a number of statistical techniques such as Holt-Winters, Exponential smoothing, ARIMA, Box-Jenkins Models...
 - b) ARIMA (Autoregressive Integrated Moving Average) models also known as Box-Jenkins models...
-> What is the difference between ARIMA and Box-Jenkins models?
- 3) In the definition of ANN backpropagation with momentum is described. Describe learning algorithm for backpropagation without momentum.
- 4) The description of error term in ANN in equation 1.16 on page 26 is incomplete (variable "t" is not described). Explain the equation.
- 5) Another problem with percentage errors that is often overlooked is that they assume a meaningful zero. For example, a percentage error makes no sense when measuring the accuracy of temperature forecasts on the Fahrenheit or Celsius scales.
-> Explain.
- 6) They (MAPE) also have the disadvantage that they put a heavier penalty on negative errors than on positive errors.
-> Explain.
- 7) Is it more practical to use daily, weekly or monthly data for forecast one year in the future?
- 8) Would it be reasonable to assume that the only seasonality present in the data is with the lag of 1 year? And if yes, could we benefit from this constrain?
- 9) Could you compare accuracy of your predictions with predictions of inventoro.com service from Czech company focusing exclusively on demand forecasting and inventory management (preferably in the terms of training and testing errors)?
- 10) We also did not have information regarding factors that could affect the sales, like promotions, discounts, price hike or reduction. These sudden changes might have an affect on the accuracy of the forecasts.
-> Would it be reasonable to include temperature as one of the predictors?
- 11) The figure 2.10 shows the forecasts of all the four models discussed above. We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.4. Overall it is observed that the neural network model behaves the best followed by exponential smoothing, holt winters and ARIMA.
-> It looks like neural network in figure 2.10 has the biggest problem to model seasonality in the forecast. Can you discuss it?
- 12) The results are shown in table 2.8. If we look at the MAPE values we can see that neural networks again outperforms all the other models followed by holtwinters, ARIMA and exponential smoothing.
-> Do you think neuron networks outperforms other methods, if you look at Figure 2.14?
- 13) What are the training and testing errors on the whole data set? Do not only present the results but also describe the methodology of the validation on the testing set. Also provide interpretation of the results. Particularly comment on the difference between training and testing error.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

10. The overall evaluation

85 (B)

Criteria description:

Summarize the parts of the thesis that had major impact on your evaluation. The overall evaluation **does not** have to be the arithmetic mean or any other formula with the values from the previous evaluation criteria 1 to 9.

Comments:

If testing error was measured for the whole data set, the results would be more conclusive.

Signature of the reviewer: