Author’s response to reviews

Title: Recursive neural networks in hospital bed occupancy forecasting

Authors:
Ekaterina Kutafina (ekutafina@mi.rwth-aachen.de)
Istvan Bechtold (istvan.bechtold@st-marienkrankenhaus.de)
Klaus Kabino (kkabino@mi.rwth-aachen.de)
Stephan Jonas (sjonas@mi.rwth-aachen.de)

Version: 1 Date: 07 Sep 2018

Author’s response to reviews:

Dear Editors,

Please find in the submission files our revised manuscript. The reviewers’ comments were addressed and the required changes made.

We corrected two citations (now [1-2]), which were not properly indexed earlier.

We modified several sections in introduction, discussion and outlook to better reflect limitations and comparison to state of the art as proposed by the reviewers.

We have also changed the format of Table 4 into Figure 5 as suggested by the second reviewer.

Please find a detailed reply to the reviewers’ questions on the following pages.

Best regards,

Ekaterina Kutafina and Stephan Jonas

Rodney Jones, Ph.D. (Reviewer 1):

Nice work.

We would like to thank the reviewer.
Several points. Firstly, I note that the model seems to consistently perform poorly on the weekends. Some comments are required.

Fig. 4 might indeed give this impression, however the performance on the weekend days is not significantly different from the week days. We have attached separately a figure showing the average and standard deviation of MAE per day of week for the evaluation period 2014/2015.

We have also added an according sentence in the results section.

Two older papers by G Davis and W Lowell in 1998 and 1999 in Am J Medical Qual on neural networks are probably worth a mention. My own work on bed occupancy has shown that the number of deaths are a primary driver of bed occupancy, see http://www.hcaf.biz/2010/Publications_Full.pdf and this is also worth a mention.

The paper by G. Davis et al. [28] was added to the Background chapter.

The Reviewer’s paper was mentioned in the Discussion:

In case of longer times of prediction, other factors might play role. For instance, Jones [40] suggested number of hospital deaths as a possible bed occupation predictor.

I have also found that bed occupancy and emergency admissions show on/off switching behaviour and this on/off behaviour may well show up in the data on model performance listed by year in the paper. Some comments on how well the model may cope with unexpected on/off switching would be valuable.

We have added on/off switching along with other similar events into the new limitations section of the manuscript:

On the other side, some limitations arise from the short history of the proposed model. Local one-time events such as disasters, diseases outbreaks or events with slower periodicity (Olympic Games, soccer championships) might not be predictable and could have a higher influence on the following year. Similarly, Jones [38] reported on/off switching (rapid increase and decrease) of hospital bed occupancy with a cycle length of two years, which can also not be modeled with a history of less than one cycle length. The usage of the multiple years can have a smoothening effect and reduce the possible error in these cases. However, neither on/off switching nor other events with such an effect could be observed in this work.

Temperature is also a well-known factor in emergency admissions (and deaths) and since this is not measured some comments regarding this would be welcome. Hence considerably expand the scope of introduction and discussion.

Due to the medium-scale prediction of 60 days up to one year, no reliable weather or temperature prediction is available for use in the model. Thus, we have not included temperature as a
predictor into the model. However, we have addressed this issue in the outlook-section of our manuscript:

Another possible factor to take into account is weather, specifically excessive heat or cold, or other external factors such as flu outbreaks. Because of the length of our goal forecast (60 days), which is much longer than reliably available detailed weather forecasts, we decided to not incorporate it in the model. However, with the help of Bayesian modelling [39], this kind of uncertainty can be taken into account.

Cristiano André Da Costa (Reviewer 2):

The article is well written and organized. The topic is interesting and has a good contribution to the scientific community. Following, I list some comments that could improve the final version of the article:

The authors would like to thank the reviewer.

- In introduction, you could also consider the following related work:

http://doi.org/10.1097/CIN.0000000000000421

The paper suggested by the Reviewer was mentioned in the Conclusions, as a possible way to integrate the NARX prediction with an automatic decision system.

In particular, it can be integrated into an automatic decision model, e.g. similar to developed by Grübler et al. [41].

- Method is well described. In data curation you could give some examples of removed / missing data. It is not clear how you chose the training and evaluation data.

We have updated both the curation paragraph to include examples of text:

Data curation. In step 1 (removal of missing values), entries with missing data are deleted from the records (i.e., missing date or patient number). These entries correspond to about 2% of the overall data, so the removal will not essentially affect the model. Next, removal of non-necessary data (step 2), the data from clinics not taking part in the bed pool is removed (i.e., ICUs or specialty clinics not in the shared bed pool). These entries correspond to about 50% of the overall data.

In addition, the training/testing data selection strategy was enhanced by an example for better understanding:

As an example, if a history of one year and a delay of 2 is used to predict 60 days starting May 1st 2008, the following data would be used. Training data would be the data of the 365 days prior
to May 1st 2008, which is May 2nd 2007 (due to leap year) to April 30th 2008. For the prediction of the first day (May 1st 2008), the bed occupancy data of April 29 and April 30 2008 are used as input for the NARX. For the second day (May 2nd 2008), the first prediction (May 1st 2008) and the occupancy at the last day of April is used as input. In the next step, only the newly predicted occupancies are used (closed loop).

- A ROC curve wouldn't also be a good evaluation for the model?

The ROC curve can only be performed on binary classifiable data. In this specific case we either have error rate or total number of beds (or wards) as a measure of distance from ground truth. Thus, an ROC curve cannot be plotted for this type of data. We could dichotomize our results, e.g., within an “acceptable window of error”, but that is usually not more informative than error rates and was therefore omitted.

- In Figure 2, it is not clear the use of the colors, however explained in the Figure.

The colors have been chosen to correspond to Figure 1 but we forgot to put the reference in. We have added this reference so the color choice should be more clear.

- I think there is too much tables in the results section. Maybe, as a suggestion, reduce to 1 or 2 and use 1 or 2 additional graphs (for instance, for yearly predictions). The only graph used is for a 60 days prediction example.

We have replaced Table 4 with Figure 5. However, due to the scaling across almost two magnitudes between the error measures MAX and GE, we were not able to replace Tables 1-3 with according graphs.

- Discussion is good and compares with some related works. I miss a paragraph or two commenting on limitations.

We have changed the discussion and added a dedicated limitations paragraph:

On the other side, some limitations arise from the short history of the proposed model. Local one-time events such as disasters, diseases outbreaks or events with slower periodicity (Olympic Games, soccer championships) might not be predictable and could have a higher influence on the following year. Similarly, Jones [38] reported on/off switching (rapid increase and decrease) of hospital bed occupancy with a cycle length of two years, which can also not be modeled with a history of less than one cycle length. The usage of the multiple years can have a smoothening effect and reduce the possible error in these cases. However, neither on/off switching nor other events with such an effect could be observed in this work.
You could better highlight the scientific contribution in conclusion.

We have highlighted the contribution more clearly in the conclusion:

Recurrent NARX networks were successfully used for time series data modelling in other areas, but to our best knowledge the presented work is the first application of NARX to hospital bed planning. With 6.24 MAPE on 60 days forecast, our model is competitive to the current state of art, while not using any sort of personal patient’s data

- There is no 2018 reference. Furthermore, no reference from articles of BMC Medical Informatics and Decision Making.

References from 2018 (Gruebler [41], Jones [38]) and from BMC Medical Informatics and Decision Making (Schmidt et al. 2013 [5]) were added.

- Maybe as a future work, you could consider other supporting data as variables for the method. For instance, excessive temperatures, storms, traffic jams, etc. Why NARX was chosen?

We agree that several other factors are contributing on a short-scale to admissions. However, due to the medium-scale prediction of 60 days up to one year, no reliable weather or temperature or traffic prediction is available for use in the model. Thus, we have not included temperature, which is otherwise a known indicator of emergency hospitalization, as a predictor into the model. However, we have addressed this issue in the outlook-section of our manuscript:

Another possible factor to take into account is weather, specifically excessive heat or cold, or other external factors such as flu outbreaks. Because of the length of our goal forecast (60 days), which is much longer than reliably available detailed weather forecasts, we decided to not incorporate it in the model. However, with the help of Bayesian modelling [39], this kind of uncertainty can be taken into account.

Justification of the choice of NARX is based on prior work, which has shown NARX to outperform other methods on comparable data. A more detailed justification can be found in the methods section:

Moreover, multiple publications provide comparisons between stochastic models (autoregressive-moving-average (ARMA), autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), generalized autoregressive conditional heteroskedasticity (GARCH) and a special type of a recurrent Artificial Neural Network: nonlinear autoregressive model with exogenous terms (NARX) [33].

While it is difficult to make generalizations, comparisons based on specific data, such as chaotic laser time series [34], wind speed [35] or refrigeration compressors production [36] tend to agree that NARX are superior to stochastic methods, particularly for multi-step forecasting [34].
The error was corrected.