Predicting Patient State-of-Health using Sliding Window and Recurrent Classifiers

Adam McCarthy[1]* and Chris Williams[2,3]

[1]Department of Computer Science, University of Oxford [2]School of Informatics, University of Edinburgh [3]Alan Turing Institute, London

Objective

Bedside monitors in Intensive Care Units (ICUs) frequently sound incorrectly, slowing response times and desensitising nurses to alarms, causing true alarms to be missed [1]. We compare sliding window predictors with recurrent predictors to classify patient state-of-health s from ICU multivariate time series x at every timestep t, i.e. $P(s_t|x_{t-l:t+r})$, where l is the amount of past context and r is the amount of future context.

1. Data

Channels Heart Rate, Systolic and Diastolic Arterial Blood Pressure, Systolic Intracranial Pressure

Annotations Stable, Blood Samples, Endotracheal Suction, Damped Traces, X Factor (abnormal)

Dataset 27 patients admitted to the Neuro ICU at the Southern General Hospital in Glasgow [4]

Max Sequence Lengths Ranging from 1,149 timesteps (19mins) to 153,678 timesteps (42 hours) depending on the event

2. Methodology

Sparse Input Sequences Only a small proportion of timesteps contained annotations (a class imbalance), so we used the annotated events as input sequences. We included context information of length equal to the event on either side, under the assumption that this would contain stable periods.

Model Inputs We supplied only the blood pressure channels to the blood sample and damped trace predictors, and all channels in other cases.

3. Hyperparameter Selection

Sliding Window Hyperparameters

- Number of hidden layers in $\{1, 2, 3\}$
- Number of hidden units h $(4 \leq h \leq 2048)$
- Length of the segments $4 \leq l \leq 49$ and $0 \leq r \leq 10$
- Learning rate μ $(0.001 \leq \mu \leq 0.1)$

Recurrent Hyperparameters

- Number of hidden cells c $(8 \leq c \leq 128)$
- Number of hidden units h $(4 \leq h \leq 2048)$
- Learning rate μ $(0.001 \leq \mu \leq 0.1)$

We performed Bayesian optimisation by fitting a Gaussian Process prior over our observations of performance and generating new hyperparameters using an acquisition function which computes the expected improvement [3]. This improved the selection time considerably over grid search.

* work completed whilst at University of Edinburgh.

4. Sliding Window (MLP)

We extracted the following features

- Least squared fits of a line of segments of the traces
- Exponentially weighted moving average
- Pulse pressure
- The first order differences of the sequences

using a feature extraction layer with appropriate initial weights, which were fine-tuned through backpropagation. To operate on segments of the input sequence, connections were removed as necessary. This fed into a number of ReLU layers and finally, a sigmoid output layer trained to output $P(s_t|x_{t-l:t+r})$.

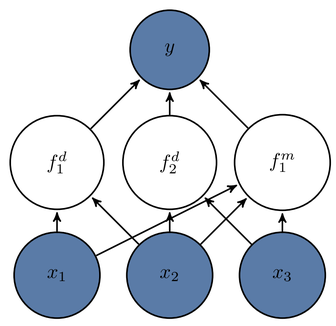


Figure 1: An example MLP with feature extraction layers. $\mathbf{x}$ is the input layer and $\mathbf{f}$ is a linear layer which computes the first order differences and mean of the input. The weights for f_1^d and f_2^d are $(+1, -1)$ and the weights for f_1^m are $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

5. Recurrent (RNN)

To investigate whether a recurrent predictor could learn long term correlations in the input, we directly supplied the input channels instead of extracting features, except for the Damped Trace event, where we replaced the systolic blood pressure with the pulse pressure.

- GRU hidden cells with a single sigmoid output unit at each timestep
- Truncated backpropagation [6] over 256 timesteps to make computation tractable

We found with the MLP experiments that including futre context r of up to 10 seconds improved classification. We therefore delayed the targets by 10 seconds during training.

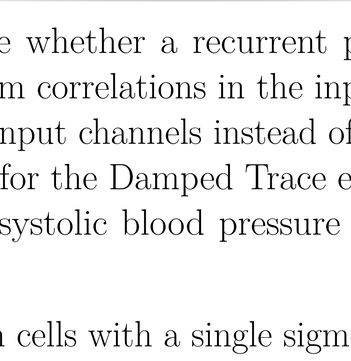
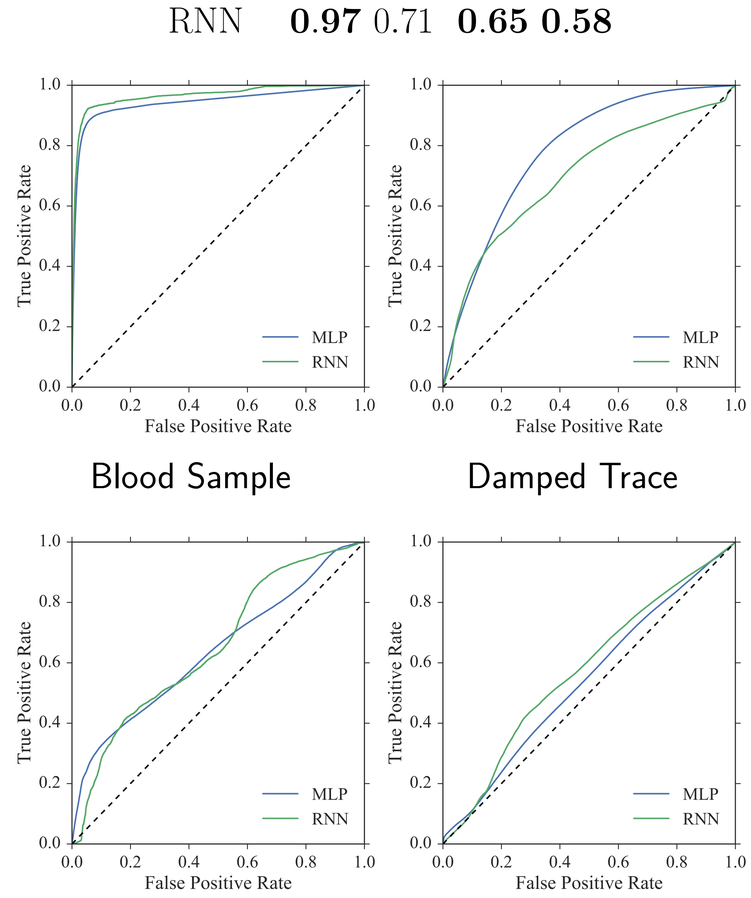


Figure 2: The RNN used in our experiments. x_t is the vital signs data, y_t is $P(s_t|\mathbf{x})$ and hidden cells are GRUs.

6. Results

We performed nested cross-validation [2] (3 folds in an outer loop and leave-one-patient-out in an inner loop) and concatenated the results.

AUC	BS	DT	SC	X
DSLDS	0.94	**0.78**	0.64	0.56
MLP	0.94	**0.78**	0.63	0.54
RNN	**0.97**	0.71	**0.65**	**0.58**



Blood Sample | Damped Trace

Endotracheal Suction | X Factor

7. Empirical Observations

Empirically, we made the following observations, which could warrant further study (see [5] sections 4 and 5 for more thorough discussion).

Adapting to baseline physiology Sliding window predictors would incorrectly classify low pulse pressures as damped traces, whereas recurrent predictors would wait for a reduction in pulse pressure before making a classification.

Classifying long-term events Events are often obscured by other pathology. RNNs can maintain state through these disturbances, whereas sliding window predictors will only classify the beginning and end. However, RNN hidden state can decay too slowly if event ends are not well delineated.

Noisy input sequences The RNN was better at handling very volatile input because the hidden state causes the predictions to effectively be smoothed, in comparison to the MLP which produces very volatile predictions in response to this volatility.

Computational complexity RNNs are much more computationally expensive to train because the sequences must be zero padded to equal length.

References

[1] M-C. Chambrin. Alarms in the intensive care unit: how can the number of false alarms be reduced?. Critical Care 5.4(1), 1. 2001

[2] S. Varma and R. Simon. Bias in Error Estimation when using Cross-Validation for Model Selection. BMC Bioinformatics, 7(1), 1. 2006.

[3] J. Snoek, H. Larochelle, and R. P. Adams. Practical Bayesian Optimization of Machine Learning Algorithms. Advances in Neural Information Processing Systems. 2012.

[4] P. Lal, C. K. I. Williams, K. Georgatzis, P. Hawthorne, C. McMonagle, I. Piper, and M. Shaw. Detecting Artifactual Events in Vital Signs Monitoring Data. Technical Report, University of Edinburgh and Glasgow. 2015.

[5] A. McCarthy. An Evaluation of Sliding Window and Recurrent Predictors for the Classification of ICU Time Series. MSc dissertation. University of Edinburgh. 2016.

[6] A. Graves. Generating sequences with recurrent neural networks. arXiv:1308.0850. 2013.