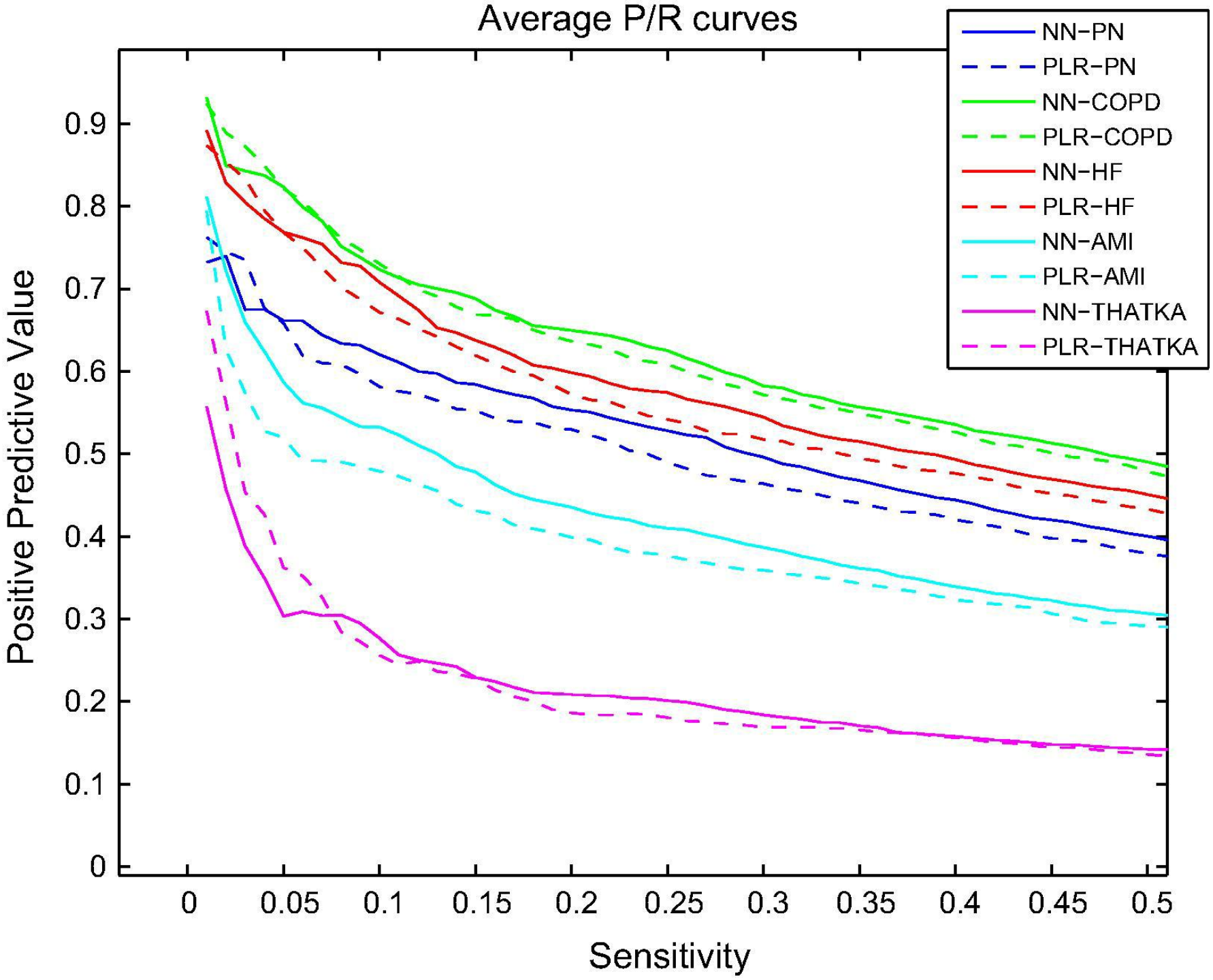
EMR Terminology

- ICD-9, CPT
- ICD-10
- SNOMED (DNS for disease)
- HL7
- HIPAA 835 Transaction set
- DICOM
- LOINC

Hospital Readmission Prediction (2015 Futoma et al)

- A study based on EMR database over 3.9m admissions, of 1.3m patients, @New Zealand
- Comparing Deep Learning model with 5 conventional statistical prediction methods
- DL model based on a RBM, feed forward and back propagation
- This simple model does not take into consideration of repeated diagnostic observation. So it impacts prediction accuracy.


Average P/R curves
NN-PN
PLR-PN
NN-COPD
PLR-COPD
NN-HF
PLR-HF
NN-AMI
PLR-AMI
NN-THATKA
PLR-THATKA
Positive Predictive Value
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Sensitivity

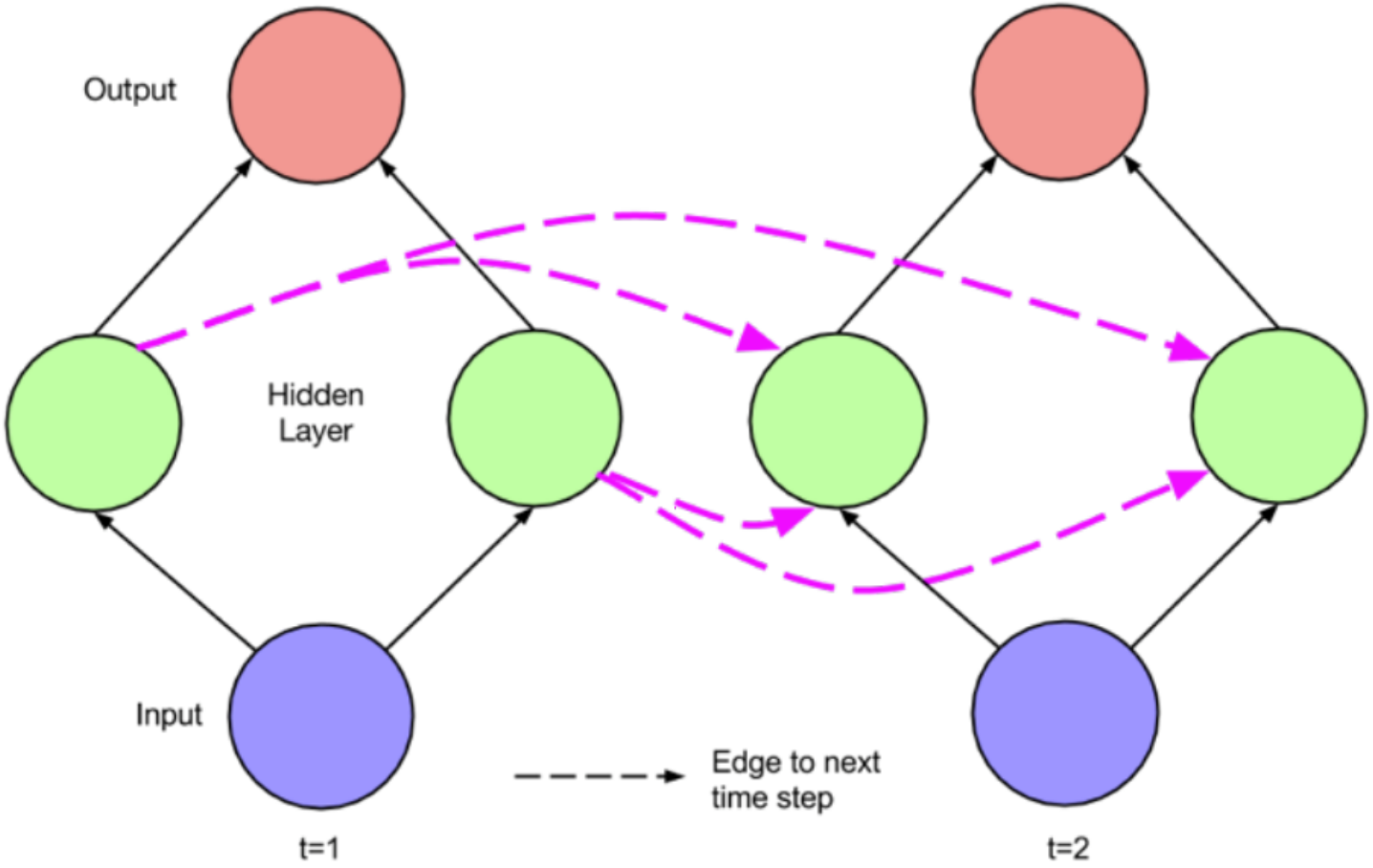
Results

- The deep neural networks consistently had better AUC (3/5 times they were significantly better at a 0:05 level), but involved a substantial amount of tuning of parameters, requiring a large amount of CPU time.
- The benefit of RBN is there is no need for data annotation
- However, this model does not take account of time-variance of data
- More advanced modeling will likely increase the prediction performance.

Review of Recurrent Network

- https://arxiv.org/abs/1506.00019: Sequence Learning to model data with time label by Z. Lipton
- https://arxiv.org/abs/1511.05942: Predicting Clinical Event by RNN

Recurrent Net (Unfolded)



$$h^{(t)} = \sigma(W_{hx}\boldsymbol{x}^{(t)} + W_{hh}\boldsymbol{h}^{(t-1)} + \boldsymbol{b}_h)$$

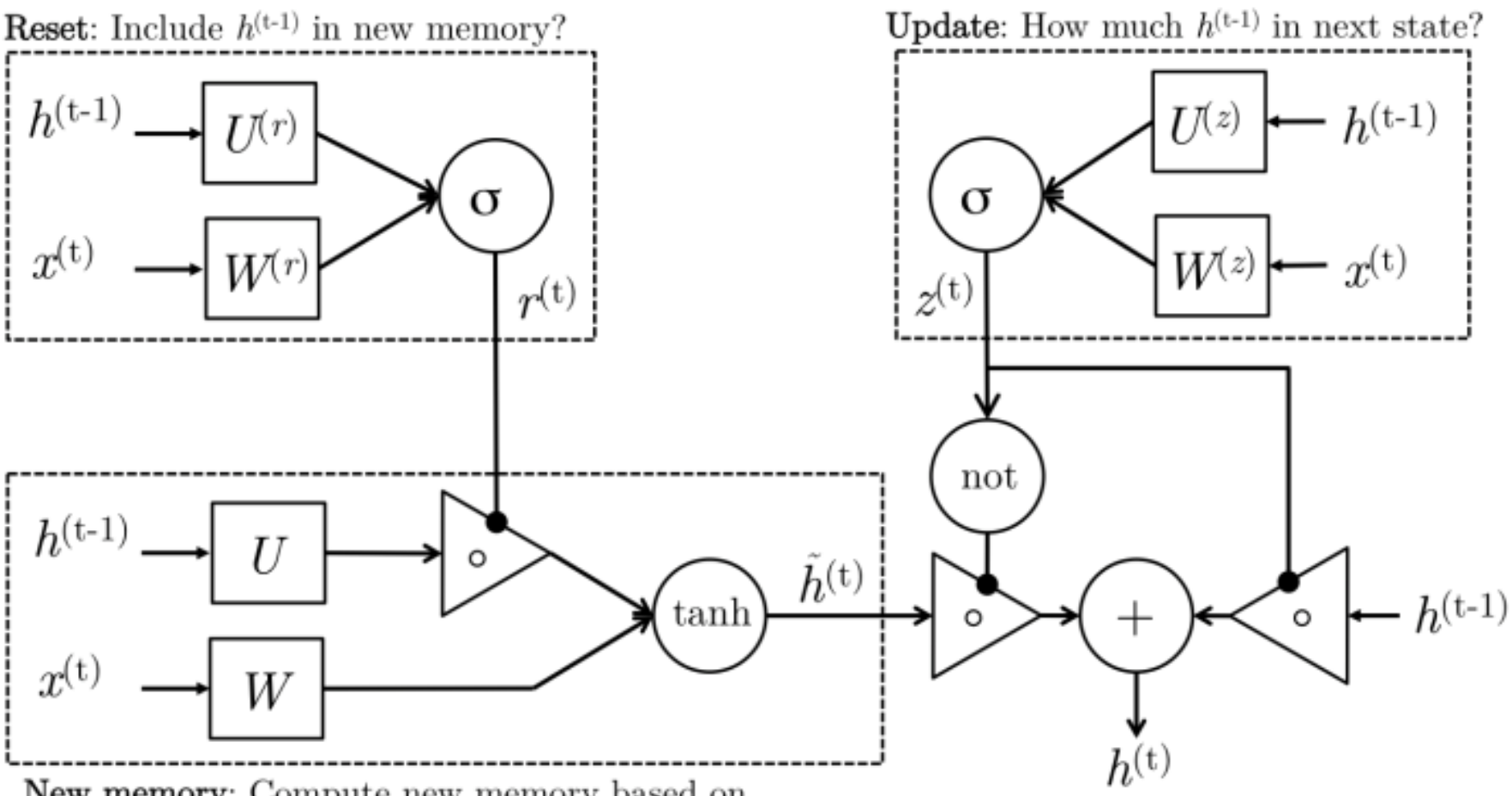
$$\hat{\boldsymbol{y}}^{(t)} = \text{softmax}(W_{yh}\boldsymbol{h}^{(t)} + \boldsymbol{b}_y)$$

Difficulty of Training RNN

- RNN is essentially a very deep NN
- Learning long term time dependencies is difficult (Bengio et al., 1994), (Pascanu et al. 2012)
- Vanishing Gradients problem: The error signal decrease too quickly as layer get deep and the front layers train very slowly.” It forgets too quickly”!
- There seems to be no successful system using plain RNN

Doctor AI

- Used an RNN with Gate Recurrent Units (GRU)
- GRU is a simpler form to introduce “memory” to a network
- This study uses 5 years of patient visit data (ICD-9), to predict the next diagnostic code in future visit


Reset: Include $h^{(t-1)}$ in new memory?
Update: How much $h^{(t-1)}$ in next state?
$h^{(t-1)}$
$U^{(r)}$
$x^{(t)}$
$W^{(r)}$
σ
$r^{(t)}$
$U^{(z)}$
$h^{(t-1)}$
σ
$W^{(z)}$
$x^{(t)}$
$z^{(t)}$
not
$h^{(t-1)}$
U
$x^{(t)}$
W
tanh
$\tilde{h}^{(t)}$
+
$h^{(t-1)}$
$h^{(t)}$
New memory: Compute new memory based on current word input $x^{(t)}$ and potentially $h^{(t-1)}$

RNN with GRU Results

- Theano source codes on Github
- No need for more than two hidden layer with GRU and Dropout
- Capable to predict “when” the next diagnostic will happen. However, result can be improved with financial status, life style, means for transportation. This has significant implication to activity data to be captured by Sabre!!

Diagnose with LSTM RNN, Lipton et al

- Using hospital emergency room to predict patient outcome.
- LSTM, like GRU, contains info outside of the normal flow of RNN in a gated cell, to allow “remember” past events without suffering vanishing gradients.
- Unlike GRU, LSTMs use a “forget gate” to block/filter the input/output, and is said to improve performance.
- LSTM attempts to allow selective forget. Some of the sequence data might not have causal relationship (my back pains 2 years ago do not usually causes to my high blood pressure today). This allows network to construct useful causal relationship.
- LSTM is particular important to medical data in my opinion, because many EMR data tend to be erratic or missing.

RNN with LSTM Results

- Learning is accelerated
- Over-fitting is reduced
- Require less data
- Although performance improvement is moderate, this achievement is still remarkable considering it is hard to train LSTM successfully