INFORMATICS PROFESSIONALS. LEADING THE WAY.

A Multi-Task Framework for Monitoring Health Conditions via Attention-based Recurrent Neural Networks

Qiuling Suo[1], Fenglong Ma[1], Giovanni Canino[2], Jing Gao[1], Aidong Zhang[1], Pierangelo Veltri[2], Agostino Gnasso[2]

[1] University at Buffalo, NY, USA;
[2] Magna Graecia University, Catanzaro, Italy

Background

- Modeling longitudinal healthcare data is challenging
  - Heterogeneous: static & dynamic variables
  - Irregular sampled: patient’s irregular visits, incomplete recording, etc.
  - Results need to be reasonably interpreted
- Monitor health record sequences using machine learning method
  - Able to handle the temporality of multivariate sequences
  - Able to memorize the long-term dependencies among variables
  - Regression or Markov models: memoryless, may miss severe symptoms in the past

Prediction Framework

Health data

Demographics

Image

Lifestyle

Survey

• • •

A patient has multiple visits;
Each visit contains multiple variables.

Learning

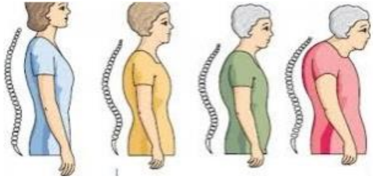
Feature representation

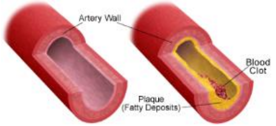
Learning Algorithm

Prediction

Classification

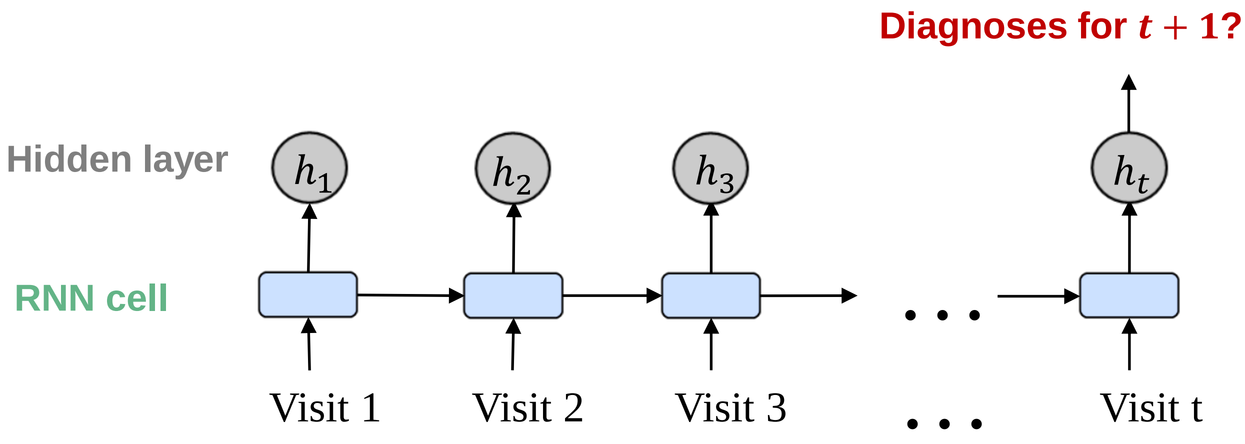
Clustering



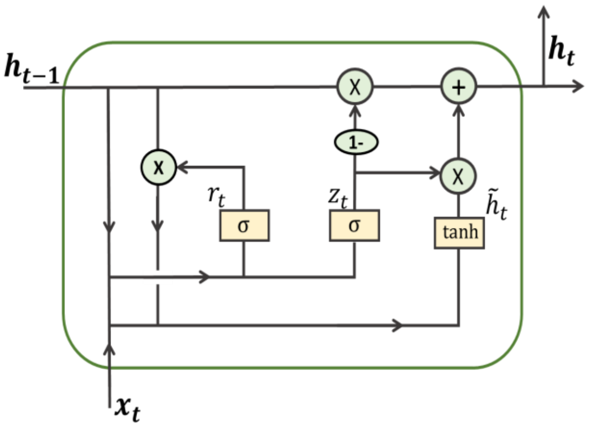


- How to learn an appropriate representation for the longitudinal health data?

RNN for Sequences



Gated recurrent unit (GRU)



update gate $z_t = \sigma(W_z x_t + U_z h_{t-1})$

reset gate $r_t = \sigma(W_r x_t + U_r h_{t-1})$

intermediate state $\tilde{h} = \tanh(W x_t + U r_t \circ h_{t-1})$

hidden state $h_t = z_t \circ h_{t-1} + (1 - z_t) \circ \tilde{h}_t$

RNN for Sequences

- Can memorize the characteristics of sequential variables
- GRU variant uses *reset* gate and *update* gate to control how much information from previous state should keep around
- Limitations:
  - Get more impact from last few timestamps, and may not discover major influences from earlier time
  - May not be able to memorize all the past information of long sequences
- Adding attention mechanism to RNN

Attention Mechanism

- Automatically calculate the importance of each past visit to current prediction

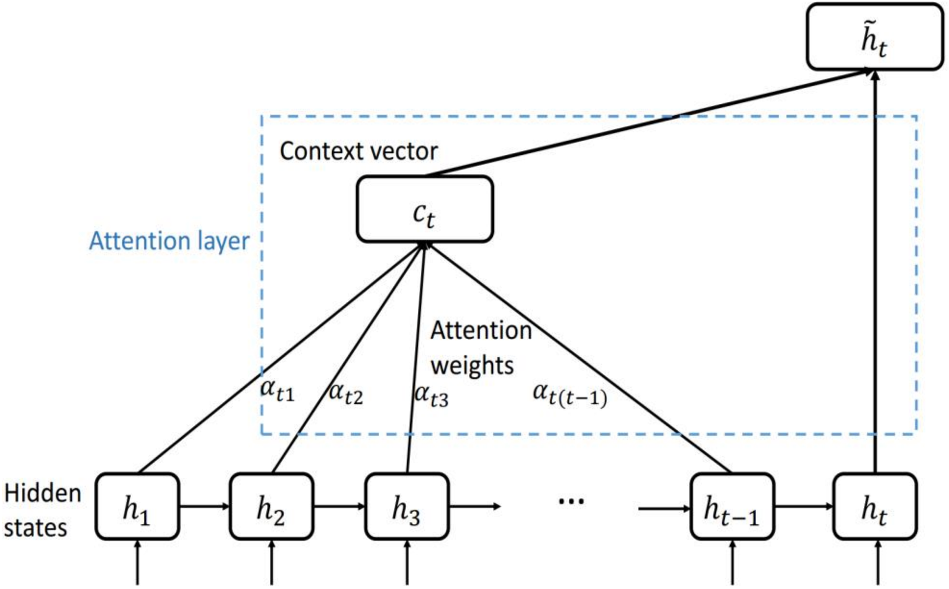


Fig. Location based attention.

- Get weight solely from current state h_i

$$\alpha_{ti} = f(Wh_i + b)$$

- Aggregate past information in context vector c_t

$$c_t = \sum_{i=1}^{t-1} \alpha_{ti} h_i$$

- Combine context vector c_t and current hidden state h_t

$$\tilde{h}_t = f(W[C_t; h_t])$$

- $\tilde{h}_t$ is the latent representation for all the past visits

Attention Mechanism

- To capture the relationships between current hidden state and all the previous states

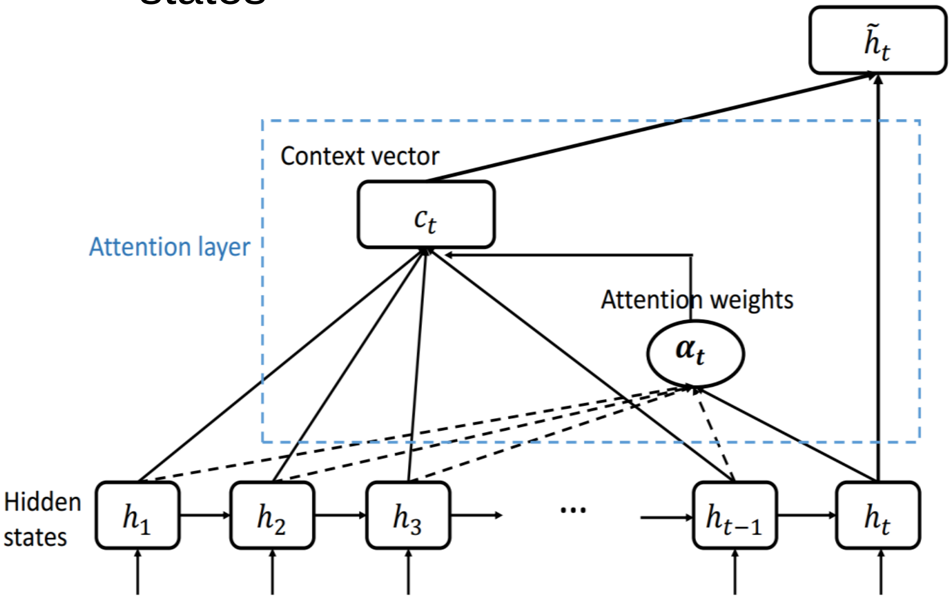


Fig. General attention and concatenation-based attention.

- Measure relationship between h_t and h_i

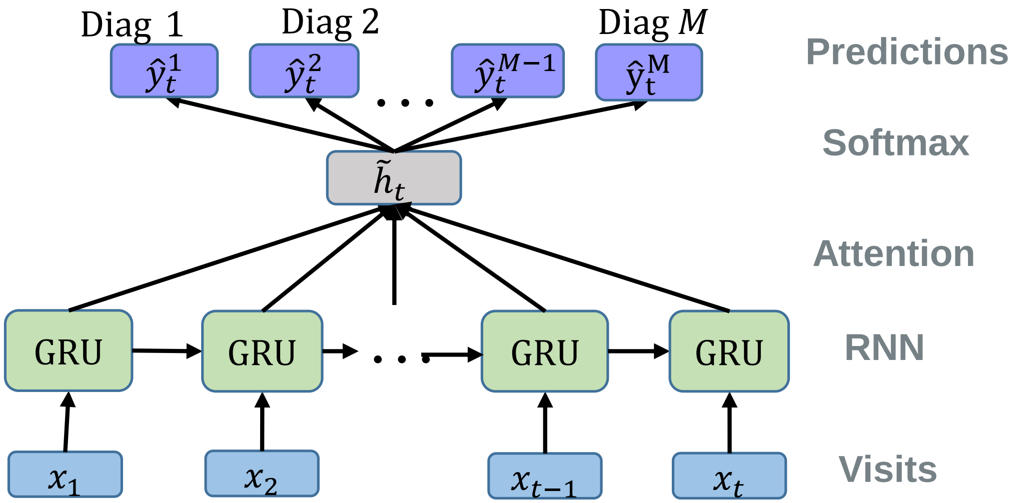
$\alpha_{ti} = f(h_t^{\mathrm{T}} W h_i)$ **General attention**

$\alpha_{ti} = vf(W[h_t; h_i])$ **concatenation-based attention**

- Aggregate past information in context vector c_t
- Combine context vector c_t and current hidden state h_t
- $\tilde{h}_t$ is the latent representation for all the past visits

Multi-task Learning Framework

- **Multi-diagnoses prediction**
  - Predict the status of multiple diagnoses
  - Each task is to predict one diagnosis
  - Each diagnosis has multiple severity levels
- **Multi-task learning:**
  - Learn a shared representation from all the features
  - Perform task-specific prediction on top of the shared information



M is the number of tasks(diagnosis variables we monitor)

Two Real World Examples

- The Study of Osteoporotic Fractures (SOF)
  - Predict bone health status (normal, osteopenia and osteoporosis) of different body regions (hip, femoral neck, etc)
- Cardiovascular disease (CD) data from *University Hospital of Catanzaro, Italy*
  - Predict level (normal, low/high abnormal) of different blood tests (MYO, CKM, TRHS, etc)

Table: Statistics of datasets

Dataset	SOF	CD
Number of patients	5,318	2,055
Number of visits	22,313	18,758
Average number of visits per patient	4.19	9.13
Number of normal claims	25,145	221,642
Number of low abnormal claims	55,399	17,407
Number of high abnormal claims	31,021	79,837
Total number of features	42	17
Number of monitored diagnoses	5	17

Prediction results

Method	Osteoporosis fracture		Cardiovascular blood test	
	Avg.# Correct	Accuracy	Avg.# Correct	Accuracy
Median	10,509	0.8209±0.0057	32,253	0.7616±0.0013
LR	10,125	0.7909±0.0069	34,836	0.8225±0.0077
RNN	10,769	0.8412±0.0042	36,167	0.8540±0.0051
RNN_l	**10,822**	**0.8454±0.0031**	36,443	0.8605±0.0056
RNN_g	10,805	0.8440±0.0027	36,423	0.8600±0.0059
RNN_c	10,816	0.8449±0.0023	**36,560**	**0.8632±0.0051**

- Baselines
  - Median considers each diagnosis separately
  - Logistic regression (LR) feeds input directly into classifiers, not memorizing historical information
- RNN: Memorize long-term dependencies
- RNN+attention
  - RNN_l location-based attention, RNN_g general attention, RNN_c concatenation-based attention
  - Fully take all the previous visit information for prediction

Visit Interpretation (Case Study)

- Attention score is used to measure the importance of historical visits to current prediction

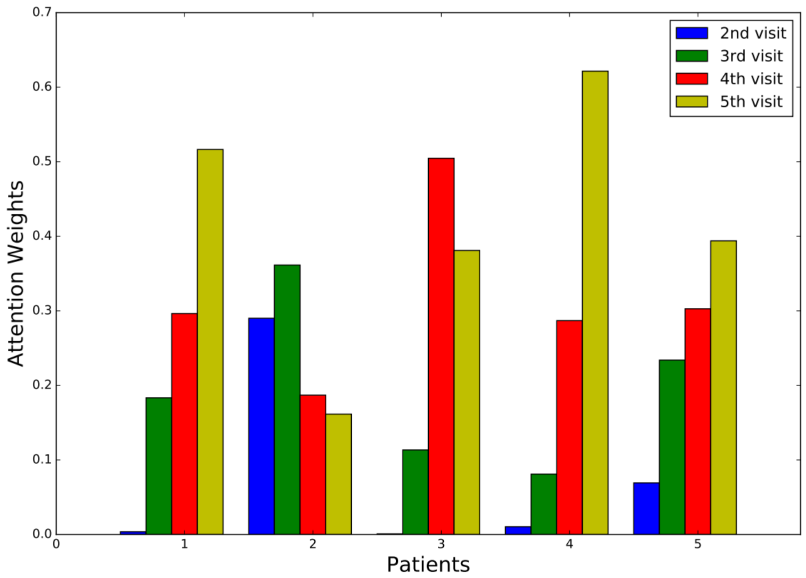


Fig. Attention weights of five patients, each with four visits on SOF data.

- The last visit is usually the most important. (patient 1, 4&5)
- Attention mechanism can identify earlier important visits (patient 2&3)
- Interpreting visit importance can help doctors to better monitor disease progression

Visit Interpretation (Case Study)

Table: BMD diagnoses and attention scores of 2nd patient. 0 is normal, 1 is osteopenia. Note that 2 (osteoporosis) does not occur for this patient.

Diagnoses\Visits	V_2	V_3	V_4	V_5	V_6
Total hip	0	0	0	0	0
Femoral neck	1	1	0	0	1
Intertrochanteric	0	0	0	0	0
Trochanteric	0	0	0	0	0
Wards	1	1	1	1	1
Attention weights	0.290	0.361	0.187	0.162	–

- Using V_1 to V_5, predict V_6
- V_4 and V_5 are closer to V_6 in terms of time, V_2 and V_3 have the same diagnoses as V_6.
- Attention mechanism assigns larger weights to V_2 and V_3, indicating the variable conditions are similar to V_6.

Conclusion

- RNN architecture
  - memorize health characteristics of longitudinal records
  - promising results on predicting risk of diseases
- Attention mechanism
  - further improve accuracy, and interpret visit importance
  - provide suggestions for doctors to pay attention on information from specific time points
- Monitor multiple diagnoses simultaneously
  - learn shared hidden causes and confounding risks
  - can help doctors to make more precise decisions on controlling risks

Questions

Question: RNN is designed to memorize sequential information by recursively updating its hidden states. In our longitudinal health data prediction, we add attention mechanism on RNN in the visit level. The advantage of using attention mechanism is:

A. Reduce the dimension of feature representation
B. Measure the contribution from historical states
C. Extract high level abstract features
D. Investigate feature correlations

Answer

A. Reduce the dimension of feature representation

B. Measure the contribution from historical states

C. Extract high level abstract features

D. Investigate feature correlations

Explanation: Attention mechanism makes RNN more powerful. RNN is impacted more by the latest information, and may not be able to discover major influences from earlier timestamps. We add attention in the visit level, to measure the importance of each visit timestamps. The attention based RNN can take all the previous visit information into consideration, assign different importance scores for the past visits.

Thank you!

Questions?