Condensed Memory Networks for Clinical Diagnostic Inferencing

Aaditya Prakash[1] — *Brandeis University*
Siyuan Zhao — *Worcester Polytechnic Institute*

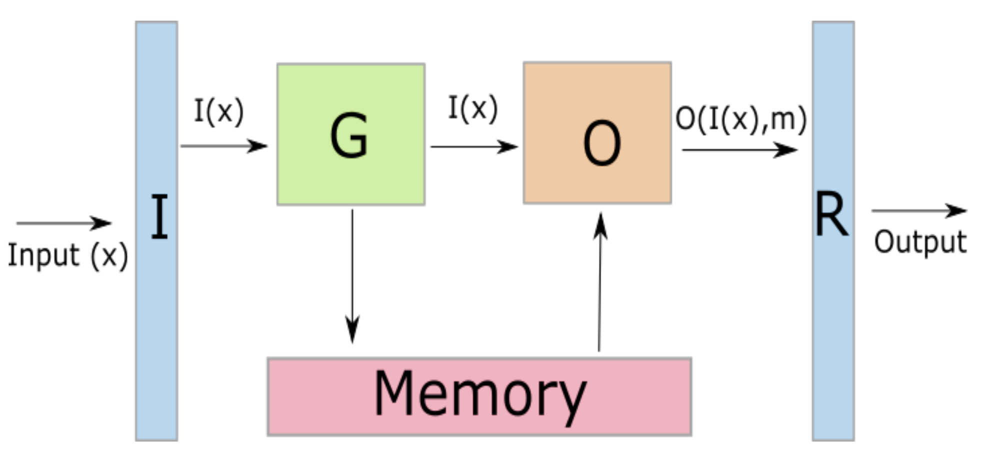
aprakash@brandeis.edu

Sadid A. Hasan, Vivek Data, Kathy Lee, Ashequl Qadir, Joey Liu, Oladimeji Farri

[1] *Artificial Intelligence Laboratory, Philips Research North America, Cambridge*

AAAI-17 San Francisco

Memory Networks



$$x_i = \{x_{i1}, x_{i2}, ..., x_{in}\}$$

$$m_i = \sum_j Ax_{ij}$$

$$p_i = softmax(u^T m_i)$$

$$= softmax(q^T B^T \sum_j Ax_{ij})$$

$$c_i = \sum_j Cx_{ij}$$

$$o = \sum_i p_i c_i$$

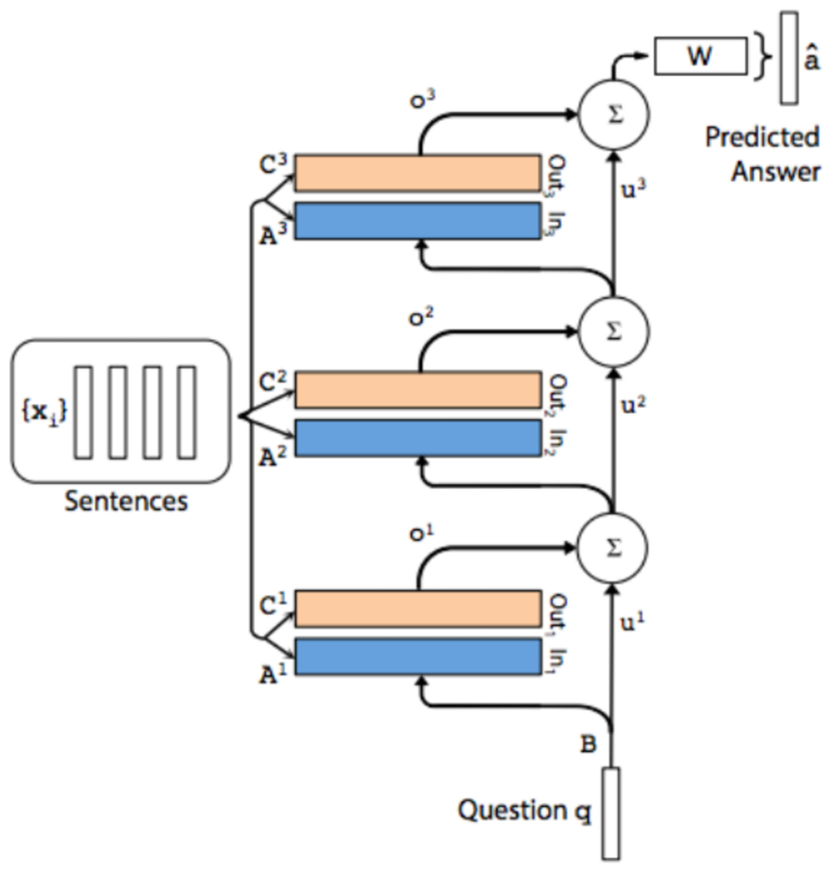
$$a = softmax(W(o+u))$$

I Convert the input (x) to internal representation (aka feature space)
- learn embeddings (bow/gru/lstm), use pre-trained vectors (glove/w2v)

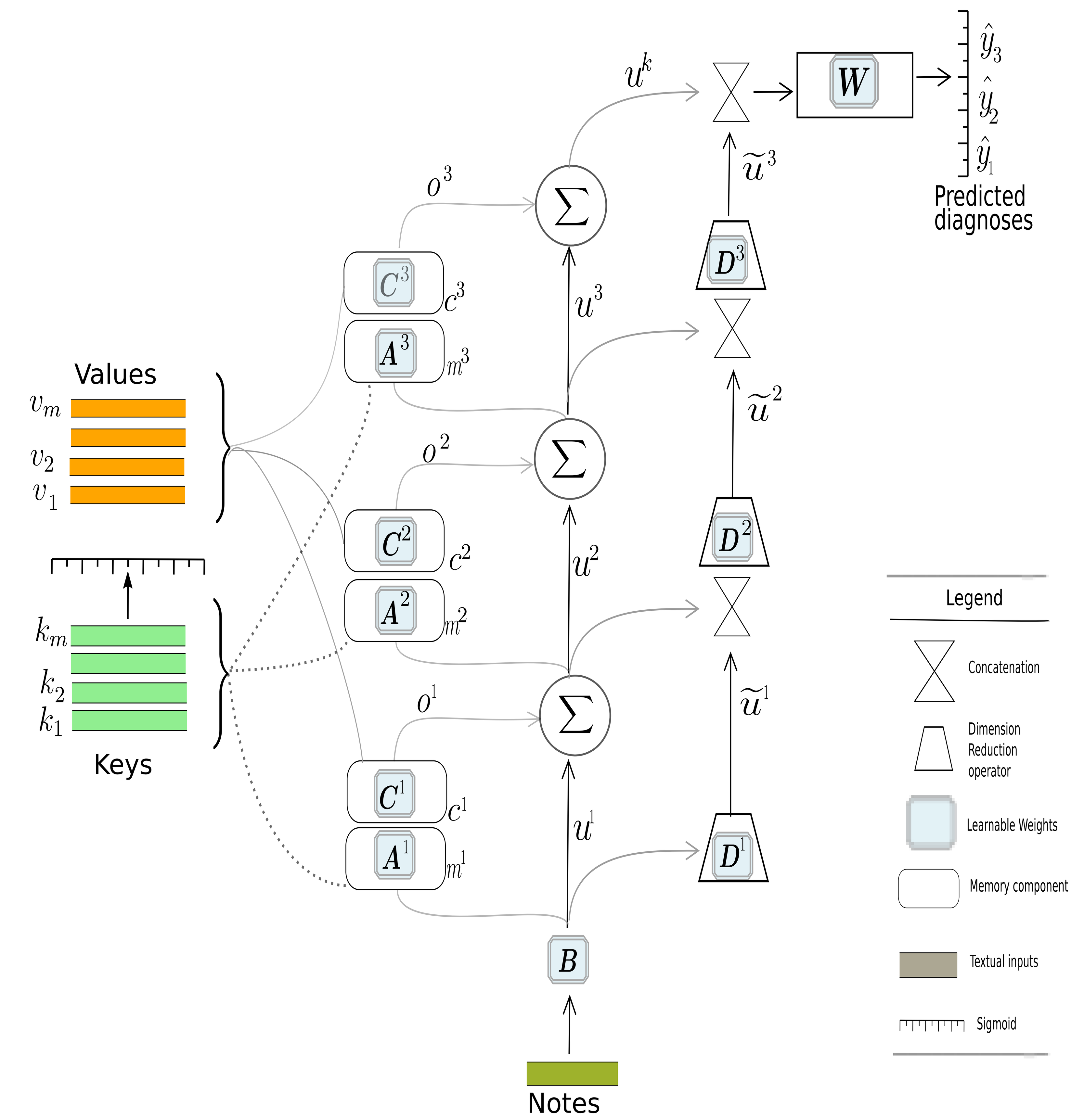
G Make a memory state (u), write initial data to the memory, update when necessary

O Generate the output state, which is combination of memory state and input

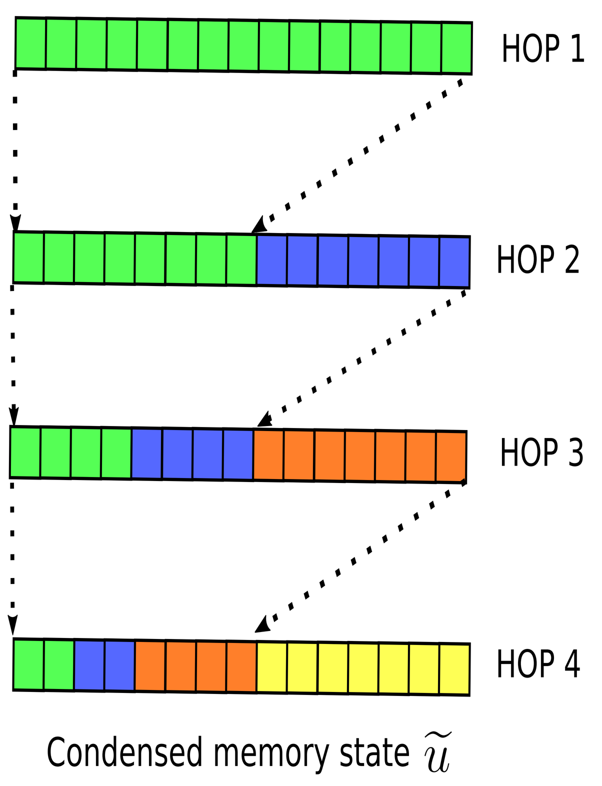
R Convert the output (O) into response as desired by the model



Condensed Memory Networks



Hierarchy in memory state



C-MemNN

$$o^k = \sum_i \text{Addressing}(u^k, m_i^k) \cdot c_i^k$$

$$u^{k+1} = u^k + o^k$$

$$\tilde{u}^{k+1} = u^{k+1} \oplus \boldsymbol{D_1} \cdot \tilde{u}^k$$

$$\hat{y}_r = \arg\max_{r \epsilon R} \frac{1}{1 + e^{-1 * (\tilde{u}^{k+1} \cdot \boldsymbol{W})}}$$

A-MemNN

$$\tilde{u}^{k+1} = \tilde{u}^k + \frac{\tilde{u}^{k-1}}{2} + \frac{\tilde{u}^{k-2}}{4} + \ldots$$

Dataset and Knowledge base

- MIMIC-III noteevents table - discharge summaries
- 58K ICU patients free-text clinical notes.
- Discard records with rare diagnoses (<50 records)

-Wikipedia pages of Diagnoses
-First paragraph and 'Signs and Symptoms' section.

Cardiac arrest

Cardiac arrest is a sudden stop in effective blood circulation due to the failure of the heart to contract effectively or at all[1]. A cardiac arrest is different from (but may be caused by) a myocardial infarction (also known as a heart attack), where blood flow to the muscle of the heart is impaired such that part or all of the heart tissue dies…
Signs and symptoms
Cardiac arrest is sometimes preceded by certain symptoms such as fainting, fatigue, blackouts, dizziness, chest pain, shortness of breath, weakness, and vomiting. The arrest may also occur with no warning …

Medical Note (partially shown)

Date of Birth: [**2606–2–28**] Sex: M
Service: Medicine
Chief Complaint:
Admitted from rehabilitation for hypotension (systolic blood pressure to the 70s) and decreased urine output. **History of present illness:**
The patient is a 76-year-old male who had been hospitalized at the [**Hospital1 3007**] from [**8–29**] through [**9–6**] of 2002 after undergoing a left femoral-AT bypass graft and was subsequently discharged to a rehabilitation facility.
On [**2682–9–7**], he presented again to the [**Hospital1 3087**] after being found to have a systolic blood pressure in the 70s and no urine output for 17 hours.

Diagnosis

Cardiorespiratory arrest.
Non-Q-wave myocardial infarction.
Acute renal failure.

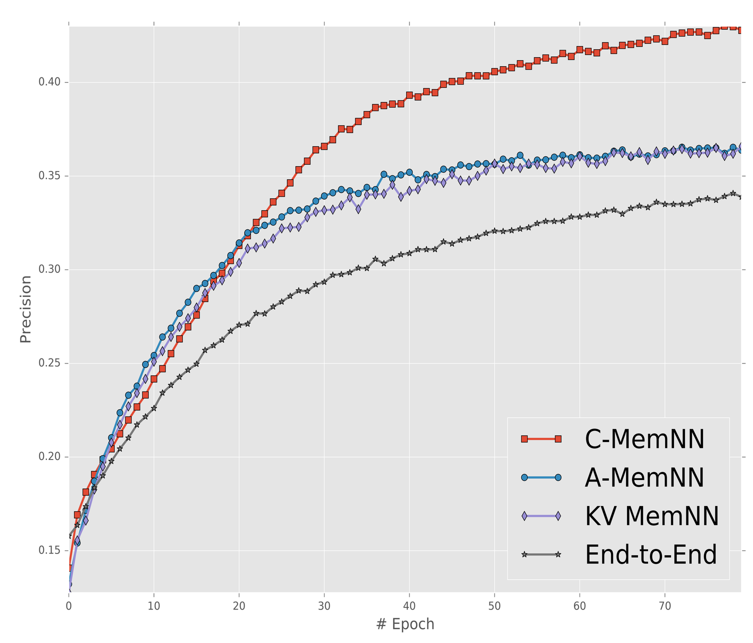
Training and Analysis

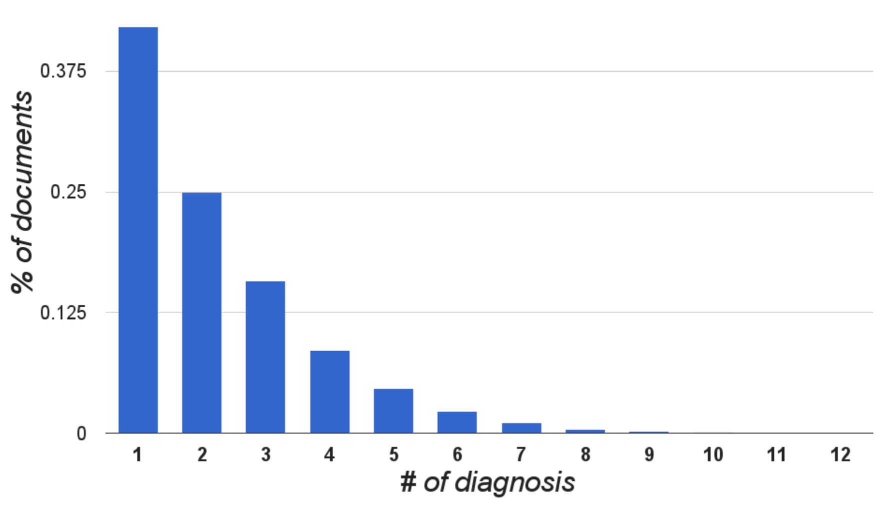
- Distribution of diagnoses has long tail.
- We found 4186 unique dignoses but most occur only once.
- 50 most common diagnoses cover 97% of the notes.
- 100 most common diagnoses covers 99.97% of the notes.
- We use maximum of 600 words in notes and wiki-pages.
- Vocabulary size was fixed to 20K. Embedding dimension was 500.
- Adam was used for optimization with learning rate of 0.001, batch size was 100.
- Loss is sum of cross entropy from prediction labels and prediction of memory slots.
- L2 loss and dropout of 0.5 used for regularization.

- AUC calculated by taking unweighted mean of AUC values.
- Average Precision is reported over top five predictions.
- Hamming loss is reported instead of accuracy because of multi-label classification.
- All results are from 10% of withheld test set after training for 100 Epoch.

- Our model outperforms other variants of memory networks.
- Difference in performance is more pronounced at higher hops.
- A-MemNN which does not add any extra parameters also performs better at higher hops.
- Across all models, higher hops is always better with diminishing return.
- C-MemNN has 30% more parameters compared to standard memory networks.
- Training loss plot (->) shows having more parameters in our model does not lead to overfitting.
- Because of parallel update of condensed memory state, inference time is same across models.

Results





Classes	# HOPS	Model	AUC (macro) ↑	Average Precision @5 ↑	Hamming Loss ↓
50 classes	3	End-to-End	0.759	0.32	0.06
		KV MemNN	0.761	0.36	**0.05**
		A-MemNN	0.762	0.36	0.06
		C-MemNN	**0.785**	**0.39**	**0.05**
	4	End-to-End	0.760	0.33	0.04
		KV MemNN	0.776	0.35	0.04
		A-MemNN	0.775	0.37	0.03
		C-MemNN	**0.795**	**0.42**	**0.02**
	5	End-to-End	0.761	0.34	0.04
		KV MemNN	0.775	0.36	0.03
		A-MemNN	0.804	0.40	0.02
		C-MemNN	**0.833**	**0.42**	**0.01**
100 classes	3	End-to-End	0.664	0.23	0.15
		KV MemNN	0.679	0.24	0.14
		A-MemNN	0.675	0.23	0.14
		C-MemNN	**0.697**	**0.27**	**0.12**
	4	End-to-End	0.672	0.24	0.15
		KV MemNN	0.683	0.24	0.13
		A-MemNN	0.689	0.23	0.11
		C-MemNN	**0.705**	**0.27**	**0.09**
	5	End-to-End	0.683	0.25	0.14
		KV MemNN	0.697	0.25	0.11
		A-MemNN	0.720	0.29	0.11
		C-MemNN	**0.767**	**0.32**	**0.05**