Memory (Neuron) Networks

Petr Baudiš ⟨pasky@ucw.cz⟩

FEE CTU Prague; Ailao

Jan 2016

Motivation

"Normal" NNs: Map input FV to output label
MemNN task: Map input FV to *a past memory*
Perfect setting: NLP and Question Answering
bAbI dataset (simple stories)

Mary went to the bathroom. John is in the playground.
John moved to the hallway. John picked up the football.
Mary travelled to the office. Bob went to the kitchen.
Where is Mary? A:office
Where is the football? A:hallway

Motivation

"Normal" NNs: Map input FV to output label
MemNN task: Map input FV to *a past memory*
Perfect setting: NLP and Question Answering
bAbI dataset (simple stories)

Mary went to the bathroom. John is in the playground.
John moved to the hallway. John picked up the football.
Mary travelled to the office. Bob went to the kitchen.
Where is Mary? A:office
Where is the football? A:hallway

Motivation

“Normal” NNs: Map input FV to output label
MemNN task: Map input FV to *a past memory*
Perfect setting: NLP and Question Answering
bAbI dataset (simple stories)

Mary went to the bathroom. John is in the playground.
John moved to the hallway. John picked up the football.
Mary travelled to the office. Bob went to the kitchen.
Where is Mary? A:office
Where is the football? A:hallway

MemNN Architecture

- IGOR — Introduced in *(Weston, 1410.3916)*
- I: Map input to features (numeric vectors)
- G: Generalization — store input in memory
- O: Output — combine input and memory to output features
- R: Representation — convert numeric vectors to data

Basic MemNN (Weston, 1410.3916)

- I: Bag-of-words embeddings
- G: Memory is just a list we append to
- O: Pick memory by learning model $s(x, y) = x^T \cdot U^T \cdot U \cdot y^T$
- R: Select memory from the word by another s-like function

Many extensions — multi-hop inference, memory hashing, write time model, etc.

Mary went to the bathroom. John is in the playground.
John moved to the hallway. John picked up the football.
Mary travelled to the office. Bob went to the kitchen.
Where is Mary? A:office

Deep Sentence Selection (Yu, 1412.1632)

Key problem: **Evaluating pairs of sentences**

- U matrix is a common trick
- In *(Yu, 1412.1632)*: $s(x, y) = \sigma(x^T M \cdot y^T)$
- In *(Weston, 1410.3916)*: $s(x, y) = x^T U^T \cdot U y$
- Generative model — (i) the matrix projects between vector spaces of embeddings; (ii) dot-product is a measure of similarity
- Possible to train by simple regressions
- More complex projection models possible (e.g. GRU, LSTM)

Better MemNN (Weston, 1502.05698)

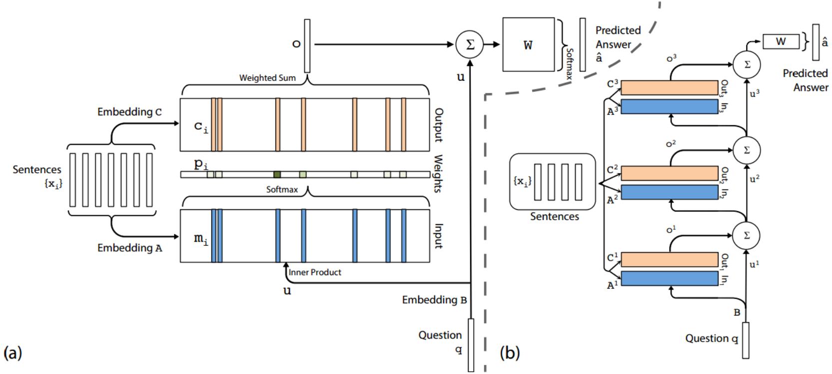
- MemNNs continue to develop besides the original paper
- (Originated at Facebook)
- Adaptive memory: Flexible number of hops
- Better relevancy scoring models (nonlinearity, MLP)
- Bag-of-ngrams

Better MemNN performance (Weston, 1502.05698)

	Weakly Supervised		Uses External Resources	Strong Supervision (using supporting facts)						
TASK	N-gram Classifier	LSTM	Structured SVM COREF+SRL features	MemNN Weston et al. (2014)	MemNN ADAPTIVE MEMORY	MemNN AM + N-GRAMS	MemNN AM + NONLINEAR	MemNN AM + NG + NL	No. of ex. req. ≥ 95	MultiTask Training
1 - Single Supporting Fact	36	50	99	100	100	100	100	100	250 ex.	100
2 - Two Supporting Facts	2	20	74	100	100	100	100	100	500 ex.	100
3 - Three Supporting Facts	7	20	17	20	**100**	**99**	**100**	**100**	500 ex.	**98**
4 - Two Arg. Relations	50	61	98	71	69	**100**	73	**100**	500 ex.	80
5 - Three Arg. Relations	20	70	83	83	83	86	86	**98**	1000 ex.	**99**
6 - Yes/No Questions	49	48	99	47	52	53	**100**	**100**	500 ex.	**100**
7 - Counting	52	49	69	68	78	86	83	85	FAIL	86
8 - Lists/Sets	40	45	70	77	90	88	94	91	FAIL	93
9 - Simple Negation	62	64	100	65	71	63	**100**	**100**	500 ex.	**100**
10 - Indefinite Knowledge	45	44	99	59	57	54	**97**	**98**	1000 ex.	**98**
11 - Basic Coreference	29	72	100	100	100	100	100	100	250 ex.	100
12 - Conjunction	9	74	96	100	100	100	100	100	250 ex.	100
13 - Compound Coref.	26	94	99	100	100	100	100	100	250 ex.	100
14 - Time Reasoning	19	27	99	99	100	99	100	99	500 ex.	99
15 - Basic Deduction	20	21	96	74	73	**100**	77	**100**	100 ex.	**100**
16 - Basic Induction	43	23	24	27	**100**	**100**	**100**	**100**	100 ex.	94
17 - Positional Reasoning	46	51	61	54	46	49	57	65	FAIL	72
18 - Size Reasoning	52	52	62	57	50	74	54	**95**	1000 ex.	93
19 - Path Finding	0	8	49	0	9	3	15	36	FAIL	19
20 - Agent's Motivations	76	91	95	100	100	100	100	100	250 ex.	100
Mean Performance	34	49	79	75	79	83	87	93		92

MemN2N (Sukhbaatar, 1503.08895)

End-to-end MemNN, weaker supervision



Conclusion

- Sorry for the brief presentation
- By itself still only toy tasks — simple questions, small number of words, small memory

Also check out:

- Keras MemNN examples
- Dynamic Memory Networks *(Kumar, 1506.07285)*
- Reading Children's Books *(Hill, 1511.02301)*

`pasky@ucw.cz`
`petr.baudis@gmail.com`

Thank you for your attention!

If you have good ideas, good data and fast computers, you can do almost anything. – Geoffrey Hinton