

Глубинное обучение

Лекция 6: Нейросети в задачах обработки текстов

Лектор: Антон Осокин

ФКН ВШЭ, 2018



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Виды задач для нейросетей

- Классификация
 - Примеры: определение темы, sentiment analysis
 - Учитывать последовательность или нет?
 - Bag-of-words, RNN
- Разметка последовательности
 - Примеры: определение частей речи, chunking
 - Локальный классификатор, RNN, CRF, RNN-CRF (BiLSTM-CRF)
- Последовательности в последовательность (разной длины!)
 - Примеры: машинный перевод, аннотация (summarization), диалоги
 - Авторегрессионные модели: seq2seq (+attention), ByteNet, и т.д.
- Синтез текста
 - Примеры: описание изображения (captioning), диалоги, искусство?
 - Авторегрессионные модели на основе обученных представлений

План лекции

- Непрерывные представления слов (embedding)
 - word2vec, FastText
- Обработка последовательностей
 - Seq2seq
- Механизм внимания
 - Seq2seq + attention
 - Логические диалоги: memory networks

Как вставить текст в нейросеть?

Непрерывные представления слов (word embeddings)

- Позволяют строить непрерывные представления дискретных объектов
- Непрерывные представления – это способ поместить текст в нейросеть
- Представление – вектор по индексу (слова, символы, n-граммы)
- Представления могут обучаться совместно с моделью
- Предобученные представления
 - Обучены на больших корпусах текстов
 - Обучены без ручной разметки (self-supervision)
 - Freeze, fine-tune, train from scratch?

Представления word2vec (skipgram)

[Mikolov et al., 2013]

- Обучение на предсказании контекста по слову
 - Обучение на корпусе текстов без разметки (self supervision)
- Вспомогательная задача:
 - Предсказываем каждое слово из контекста отдельно

Source Text	Training Samples					
<table><tr><td>The</td><td>quick</td><td>brown</td></tr></table> fox jumps over the lazy dog. ➡	The	quick	brown	{the, quick} {the, brown}		
The	quick	brown				
<table><tr><td>The</td><td>quick</td><td>brown</td><td>fox</td></tr></table> jumps over the lazy dog. ➡	The	quick	brown	fox	{quick, the} {quick, brown} {quick, fox}	
The	quick	brown	fox			
<table><tr><td>The</td><td>quick</td><td>brown</td><td>fox</td><td>jumps</td></tr></table> over the lazy dog. ➡	The	quick	brown	fox	jumps	{brown, the} {brown, quick} {brown, fox} {brown, jumps}
The	quick	brown	fox	jumps		
The <table><tr><td>quick</td><td>brown</td><td>fox</td><td>jumps</td><td>over</td></tr></table> the lazy dog. ➡	quick	brown	fox	jumps	over	{fox, quick} {fox, brown} {fox, jumps} {fox, over}
quick	brown	fox	jumps	over		

image credit:
[Chris McCormick](#)

Представления word2vec (skipgram)

[Mikolov et al., 2013]

- Обучение на предсказании контекста по слову
 - Обучение на корпусе текстов без разметки
- Предсказываем каждое слово из контекста отдельно
 - Текущее слово w ; слово из контекста v
 - Для каждого слова – 2 представления (in , out)
 - Совместимость – скалярное произведение $in_w^T out_v$
 - Полезные – представления in
 - Модель с softmax $P(v | w, \theta) = \frac{\exp(in_w^T out_v)}{\sum_{v'} \exp(in_w^T out_{v'})}$
 - Медленная нормировка
 - Обычное решение – Noise Contrastive Estimation (NCE)

$$\text{loss}(w, v) = \log(1 + \exp(-in_w^T out_v)) + \sum_{\text{random } v'} \log(1 + \exp(in_w^T out_{v'}))$$

Представления word2vec (skipgram)

[Mikolov et al., 2013]

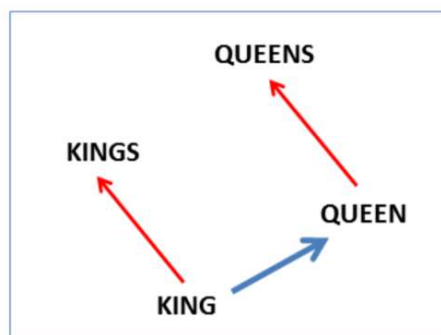
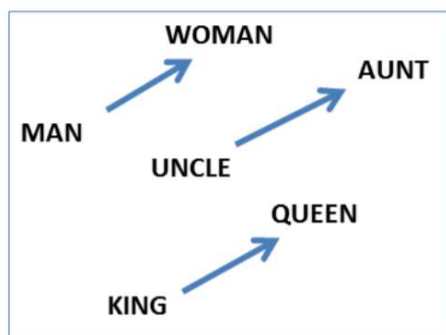
- Обучение на предсказании контекста по слову
 - Обучение на корпусе текстов без разметки
- Предсказываем каждое слово из контекста отдельно
 - Используются представления *in*
- Достоинства
 - Ближайшие соседи (cosine distance = норм. скал. произв., корпус GoogleNews)
 - university: student, teacher, teaching, students, schools
 - Putin: Medvedev, Vladimir_Putin, President_Vladimir_Putin, Prime_Minister_Vladimir_Putin, Kremlin
 - putin: lol, mr, don't, obama, Hahaha
 - obama: dems, americans, washington, america, libs

Source: http://bionlp-www.utu.fi/wv_demo/

Представления word2vec (skipgram)

[Mikolov et al., 2013]

- Обучение на предсказании контекста по слову
 - Обучение на корпусе текстов без разметки
- Предсказываем каждое слово из контекста отдельно
 - Используются представления *in*
- Достоинства
 - Ближайшие соседи (cosine distance = норм. скал. произв., корпус GoogleNews)
 - Арифметика над представлениями
 - $\text{king} - \text{man} + \text{woman} = \text{queen}$



Source: [Mikolov et al., 2013]

Представления fastText (skipgram)

[Bojanowski et al., 2017]

Как в word2vec:

- Обучение на предсказании контекста по слову
 - Обучение на корпусе текстов без разметки
- Предсказываем каждое слово из контекста отдельно

Новая идея:

- Добавить информацию о символах слова (через n-граммы)

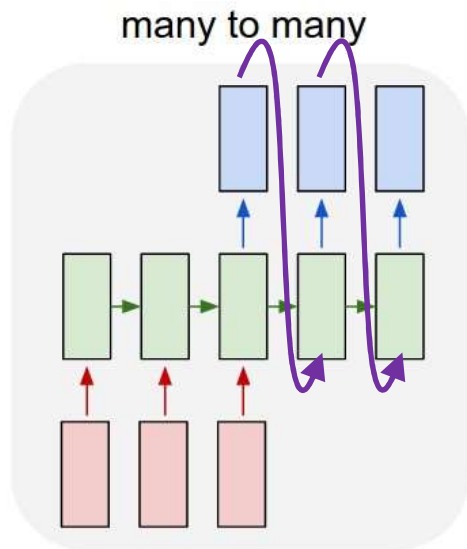
$$\text{in}_w = \text{word}_w + \sum_{p \in \text{n-grams}(w)} \text{part}_p \quad \text{in}_w^T \text{out}_v \Rightarrow \text{word}_w^T \text{out}_v + \sum_{p \in \text{n-grams}(w)} \text{part}_p^T \text{out}_v$$

- “where” = “<where>”, “<wh”, “whe”, “her”, “ere”, “re>”
- Важно использовать длинные n-граммы ($n \leq 6$)
- Достоинства:
 - Близость по написанию
 - Слова вне словаря, опечатки и т.д.

Код и данные на fasttext.cc

Модель seq2seq [Sutskever et al. , 2014]

- Модели для предсказания последовательностей разной длины



Входы, память, выходы

Входы – представления входов

Память – слои RNN или CNN

Выходы – шансы слов из словаря
(logits, идут в logsoftmax)

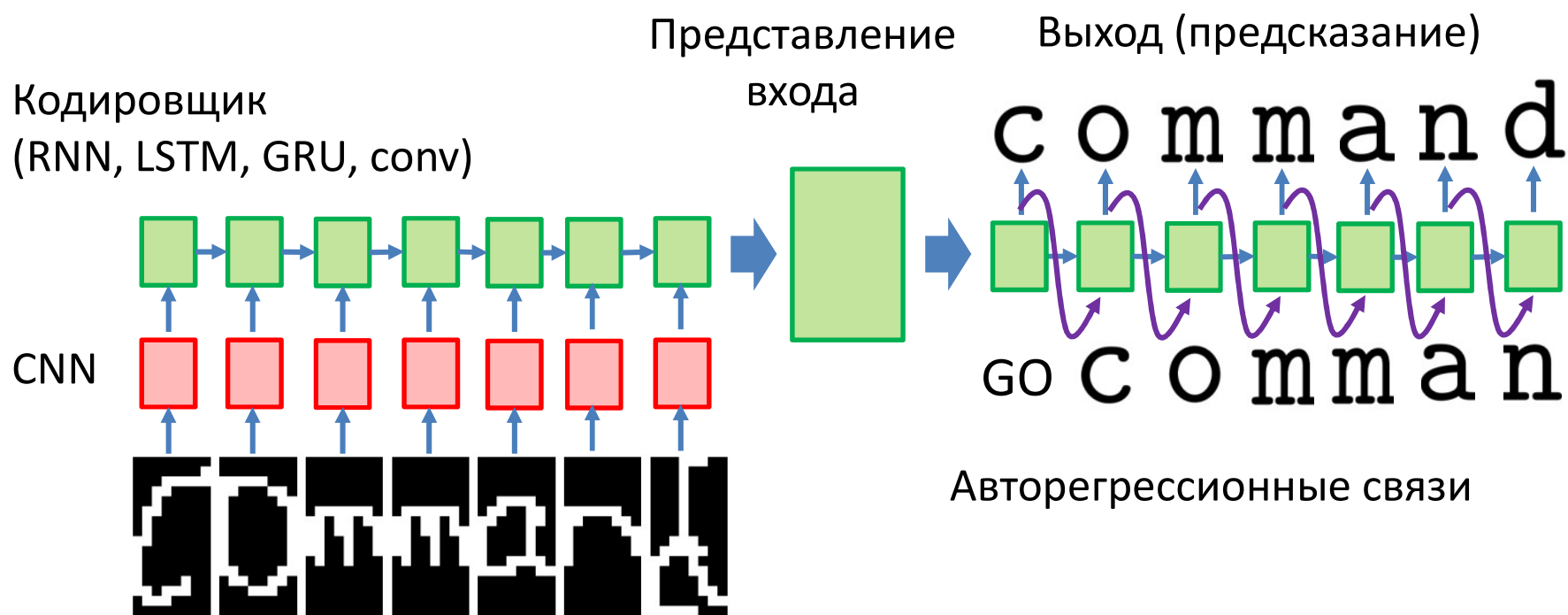
Аutoreгрессионные связи (→)
передают решение о текущем слове

На входы → подаются представления
выходного алфавита

image credit: [Andrej Karpathy](#)

Последовательное предсказание

- Пример:  → command



Если не фиксирована длина выхода, то используют символ EOS (с барьером)

Обучение авторегрессионных моделей

- Обычный способ – метод максимального правдоподобия на каждом шаге декодировщика

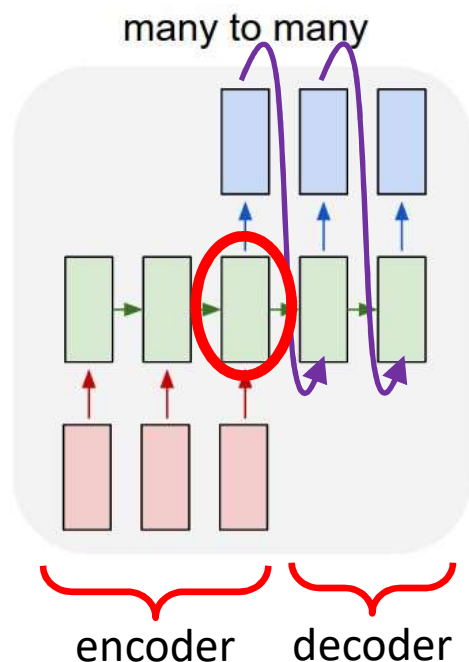
$$P(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = P(y_1 \mid \mathbf{x}, \boldsymbol{\theta})P(y_2 \mid y_1, \mathbf{x}, \boldsymbol{\theta})P(y_3 \mid y_2, y_1, \mathbf{x}, \boldsymbol{\theta}) \dots$$

- Teacher forcing – на вход декодировщику подаются правильные ответы
- Проблема:
 - Модель видит только правильные траектории
 - Не знает, что делать при ошибке
 - Как исследовать траектории (exploration)?
 - Связь с «Обучением с подкреплением»
- Можно использовать любые функции потерь
 - Информацию о качестве каждого предсказания!

[LeBlond et al., 2018]

Модель seq2seq [Sutskever et al. , 2014]

- Модели для предсказания последовательностей разной длины



Модель encoder-decoder

○ – представление всего входа

Модель плохо работает для длинных последовательностей

Причина: представление входа – вектор фиксированной размерности
(не может представить весь язык)

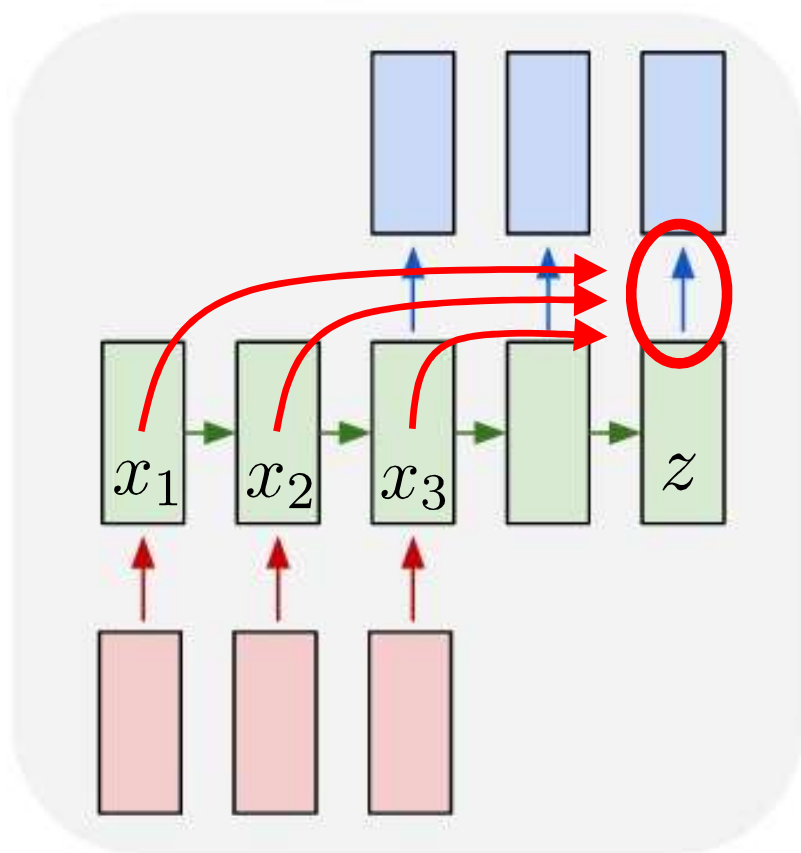
Решение: механизм внимания
(attention)

image credit: [Andrej Karpathy](#)

Модель seq2seq с вниманием

[Bahdanau et al. , 2015]

- Модели для предсказания последовательностей разной длины



Внимание «выбирает релевантные элементы памяти»

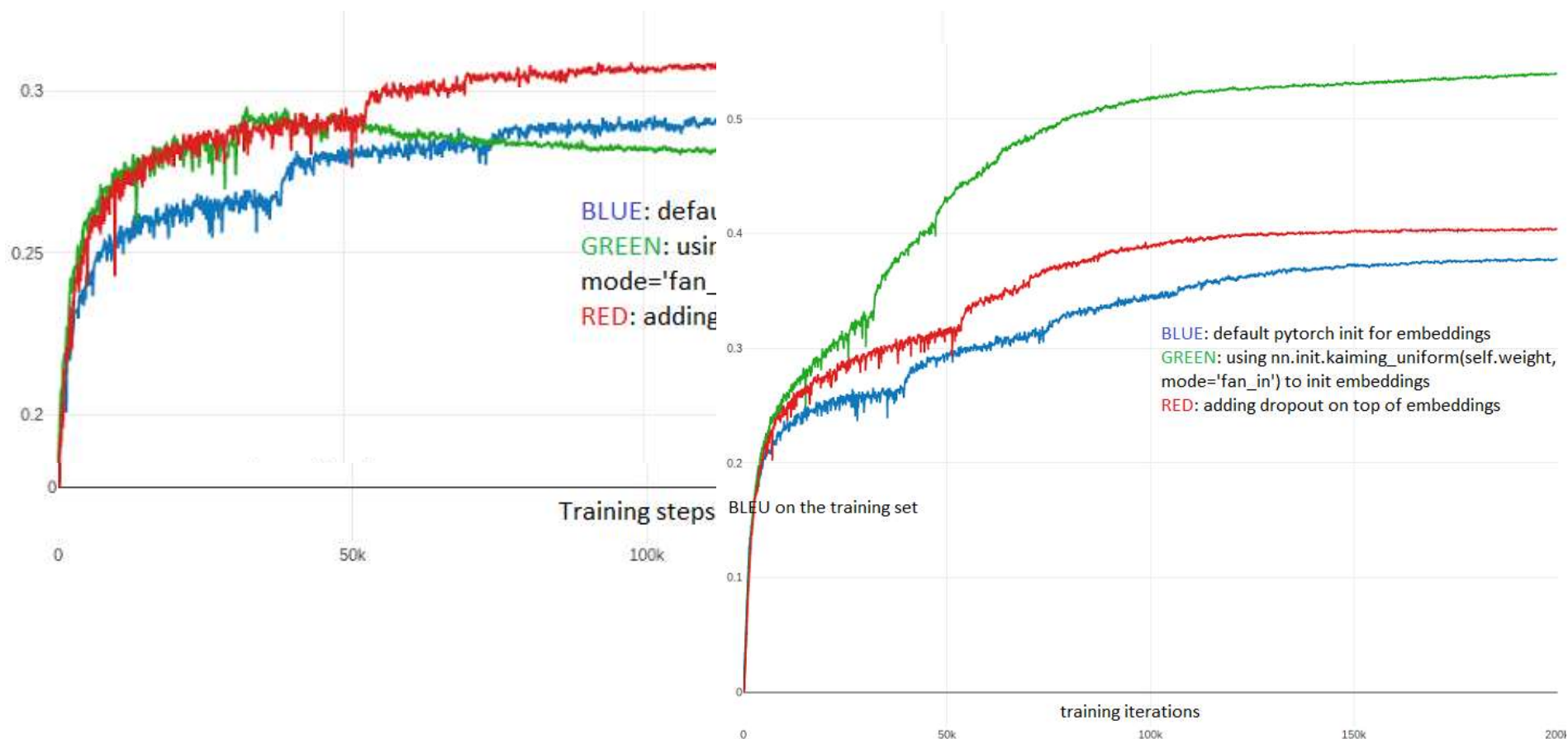
Модель внимания:

- релевантность
$$s_i := \text{score}(x_i, z) = \begin{cases} x_i^T z \\ W[x_i; z] \end{cases}$$
- вероятности
$$a_1, a_2, \dots := \text{softmax}(s_1, s_2, \dots)$$
- контекст
$$c := \sum_i a_i x_i$$
- новые признаки $\tilde{z} := [c; z]$
- soft-argmax

image credit: [Andrej Karpathy](#)

История из жизни о дебаге нейросетей

- Пример из личной практики
seq2seq с вниманием для машинного перевода
БАГ: медленное обучение и потерянные 3 BLEU



Логические ответы на вопросы

Memory networks

[Sukhbaatar et al., 2015]

- Диалоговым агентам важно запоминать большой объем информации, искать релевантные факты, совмещать факты
- Задача: ответ на вопрос по последовательности фактов (bAbI)

John was in the bedroom.

Bob was in the office.

John went to kitchen.

Bob travelled back home.

Q: Where is John? A: kitchen

SUPPORTING FACT

John is in the playground.

Bob is in the office.

John picked up the football.

Bob went to the kitchen.

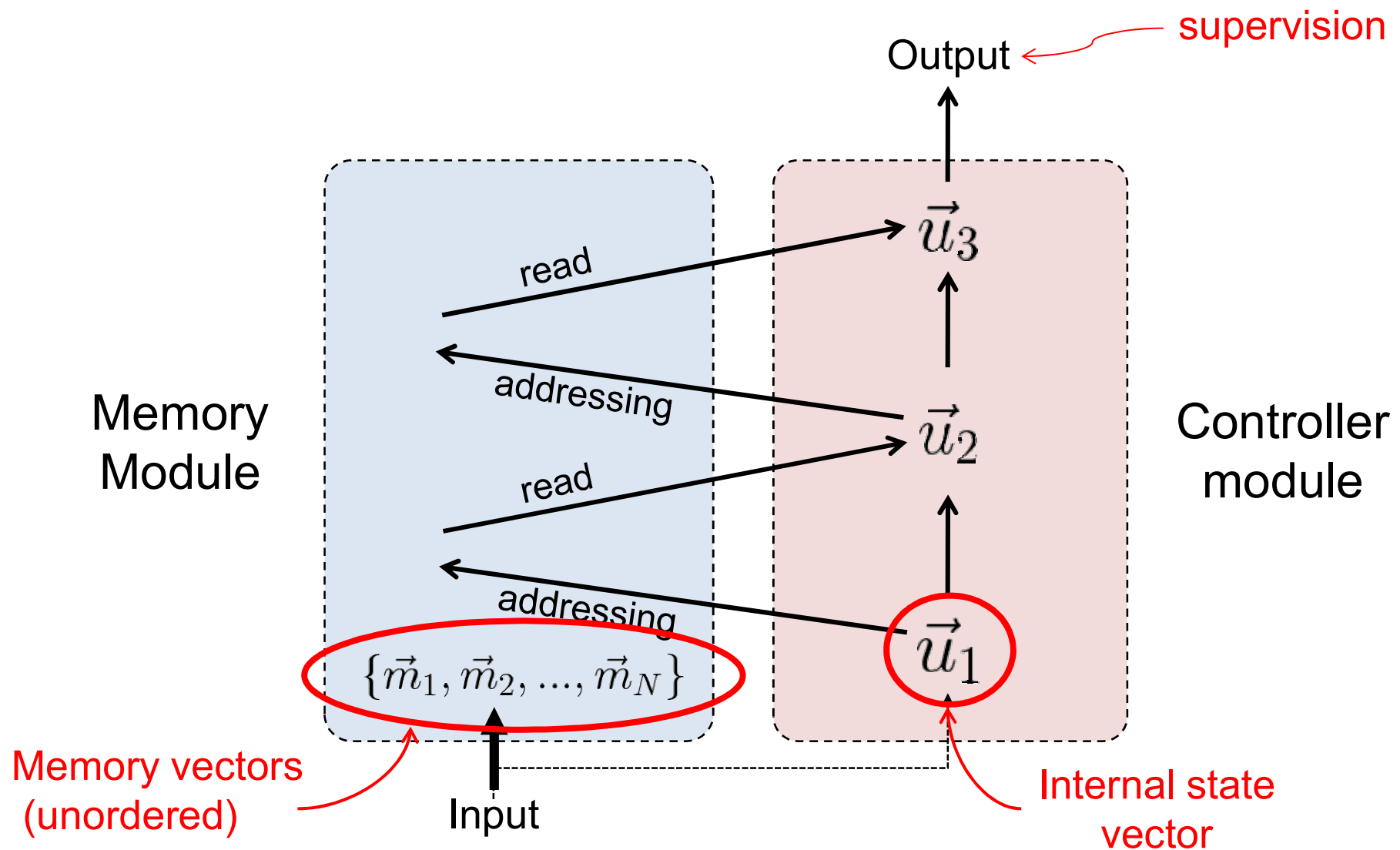
Q: Where is the football? A: playground

SUPPORTING FACT

SUPPORTING FACT

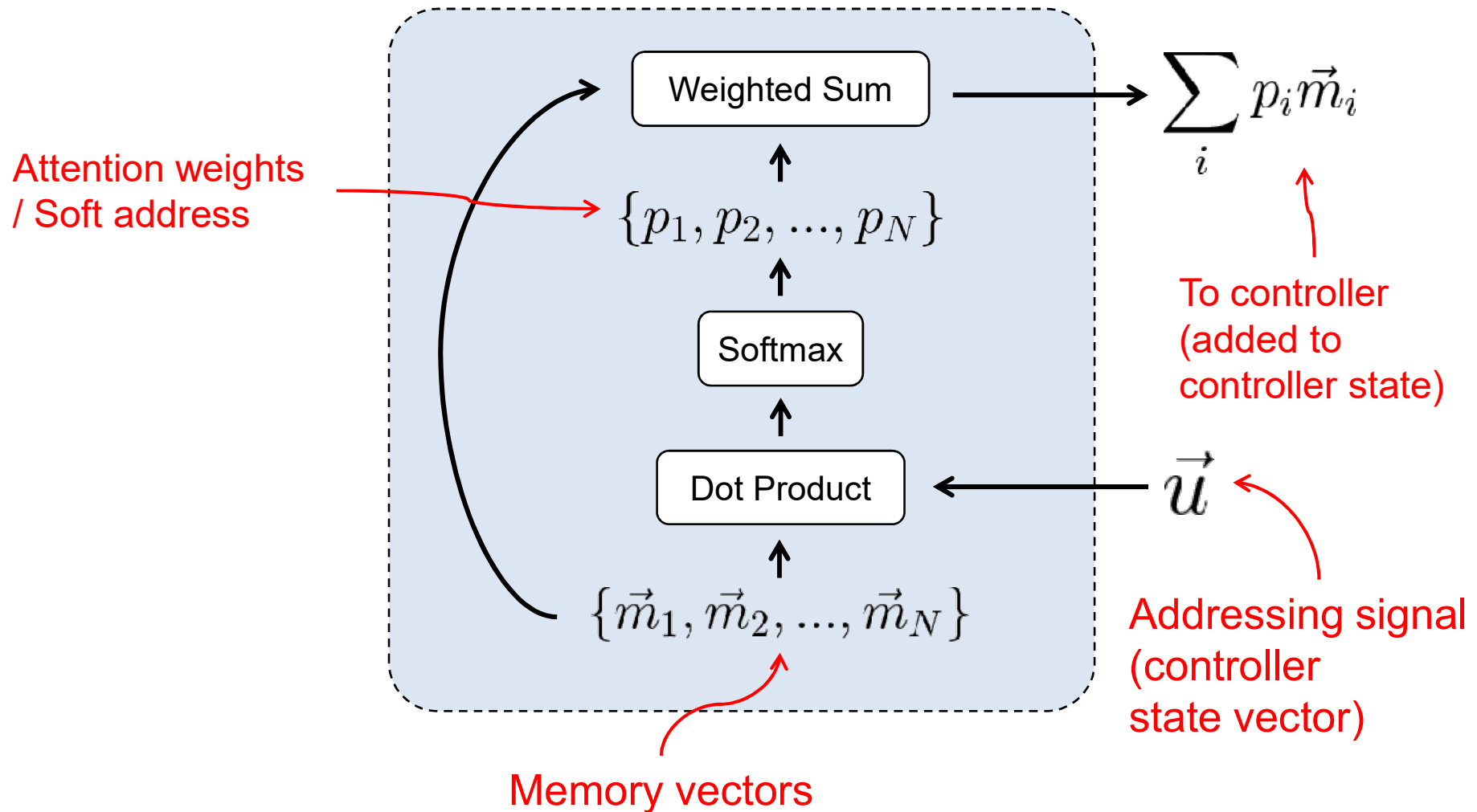
slide credit: Jason Weston

Архитектура MemN2N



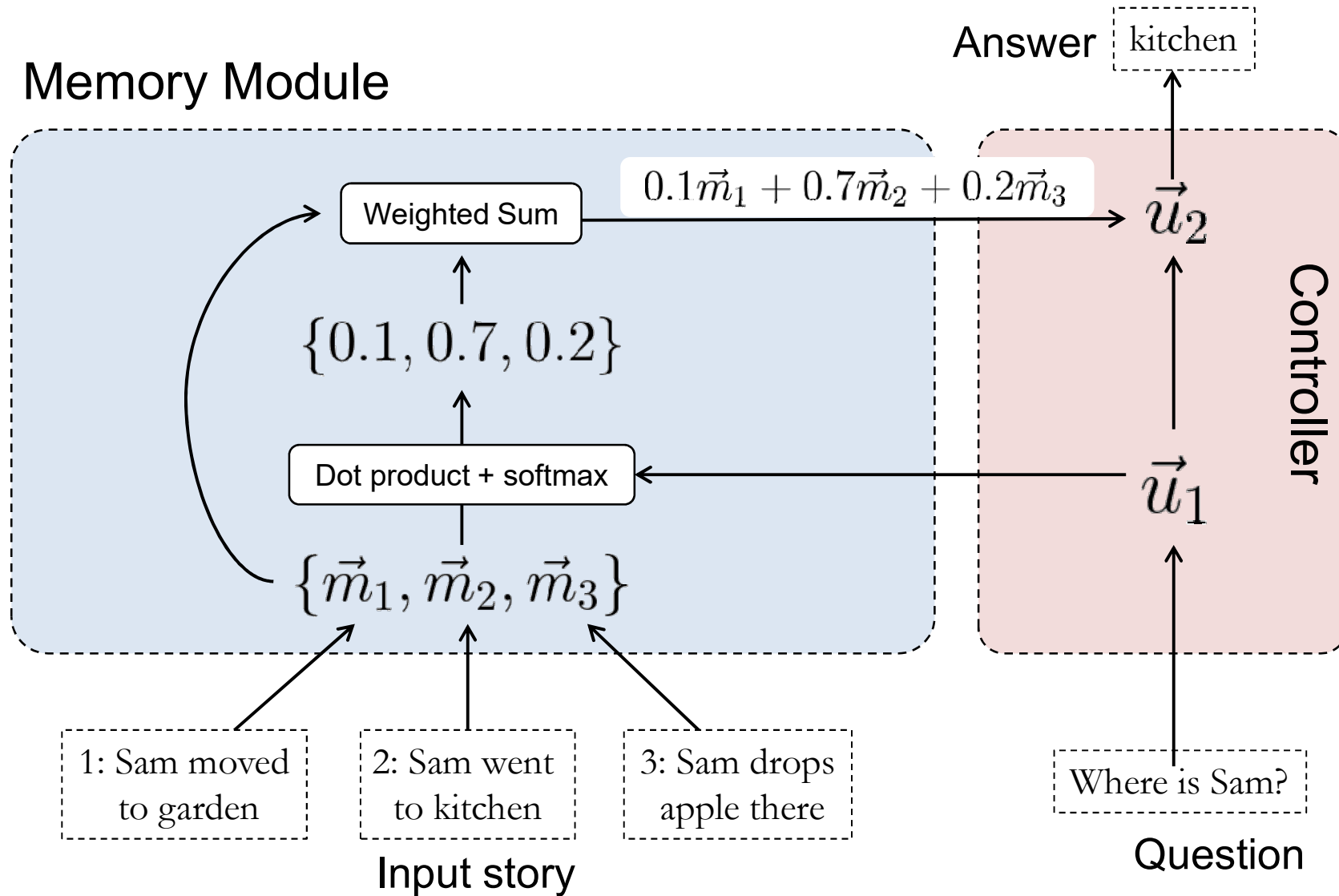
slide credit: Jason Weston

Модуль памяти



slide credit: Jason Weston

Вопрос и ответ



Представление факта: BoW, BoW + Pos, RNN

slide credit: Jason Weston

Attention during memory lookups

Samples from toy QA tasks

Story (1: 1 supporting fact)	Support	Hop 1	Hop 2	Hop 3
Daniel went to the bathroom.		0.00	0.00	0.03
Mary travelled to the hallway.		0.00	0.00	0.00
John went to the bedroom.		0.37	0.02	0.00
John travelled to the bathroom.	yes	0.60	0.98	0.96
Mary went to the office.		0.01	0.00	0.00
Where is John? Answer: bathroom Prediction: bathroom				

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow Prediction: yellow				

Story (2: 2 supporting facts)	Support	Hop 1	Hop 2	Hop 3
John dropped the milk.		0.06	0.00	0.00
John took the milk there.	yes	0.88	1.00	0.00
Sandra went back to the bathroom.		0.00	0.00	0.00
John moved to the hallway.	yes	0.00	0.00	1.00
Mary went back to the bedroom.		0.00	0.00	0.00
Where is the milk? Answer: hallway Prediction: hallway				

Story (18: size reasoning)	Support	Hop 1	Hop 2	Hop 3
The suitcase is bigger than the chest.	yes	0.00	0.88	0.00
The box is bigger than the chocolate.		0.04	0.05	0.10
The chest is bigger than the chocolate.	yes	0.17	0.07	0.90
The chest fits inside the container.		0.00	0.00	0.00
The chest fits inside the box.		0.00	0.00	0.00
Does the suitcase fit in the chocolate? Answer: no Prediction: no				

20 bAbI Tasks

	Test Acc	Failed tasks
MemNN	93.3%	4
LSTM	49%	20
MemN2N 1 hop	74.82%	17
2 hops	84.4%	11
3 hops	87.6%	11

slide credit: Jason Weston

Заключение

- Обработка языка активно использует нейросети
- Очень большая область – много задач
 - Есть успехи!
- Представления, Seq2seq, внимание, etc.
- Понимание смысла – очень сложная задача!