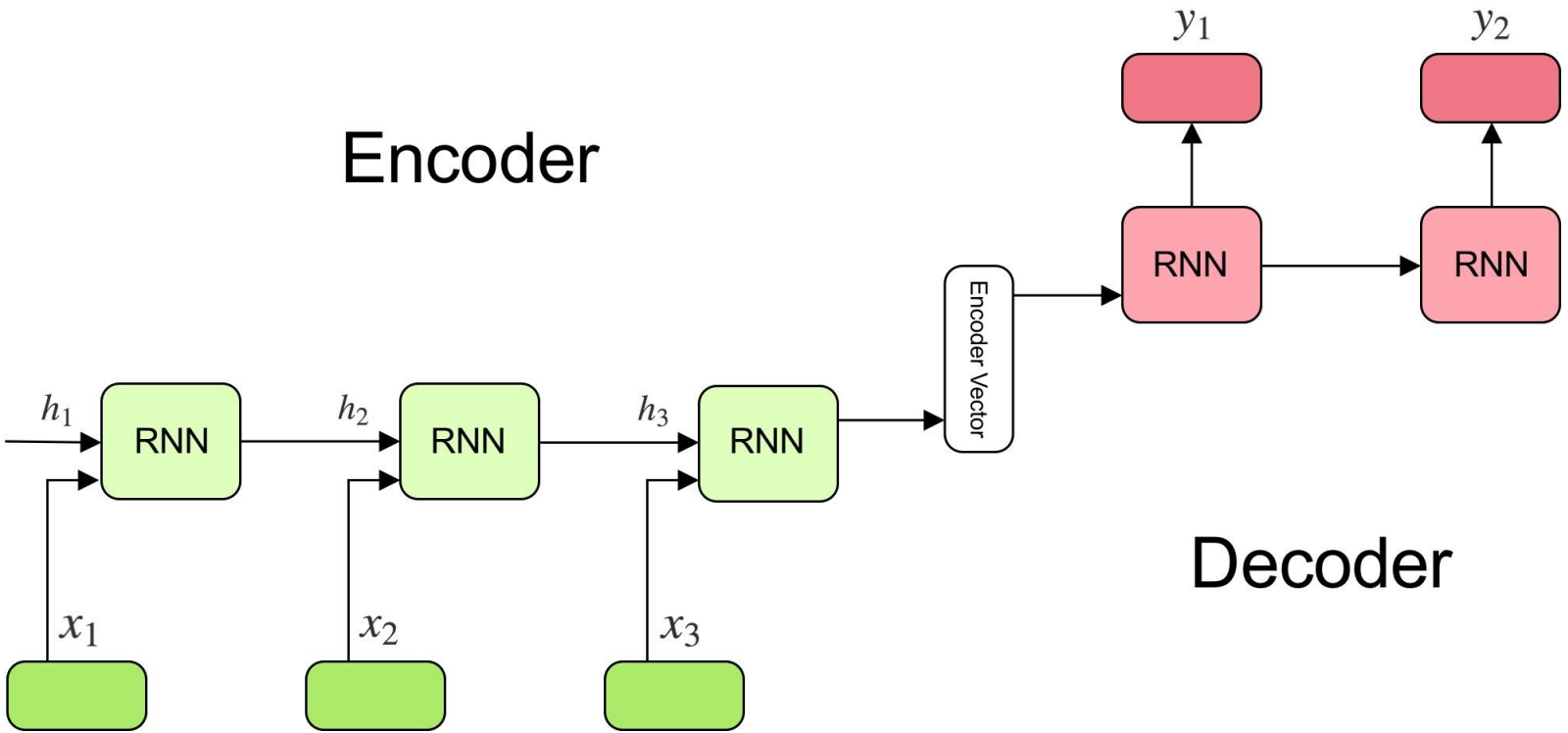
Seq2seq. Attention. Chatbots.

Маша Шеянова, masha.shejanova@gmail.com

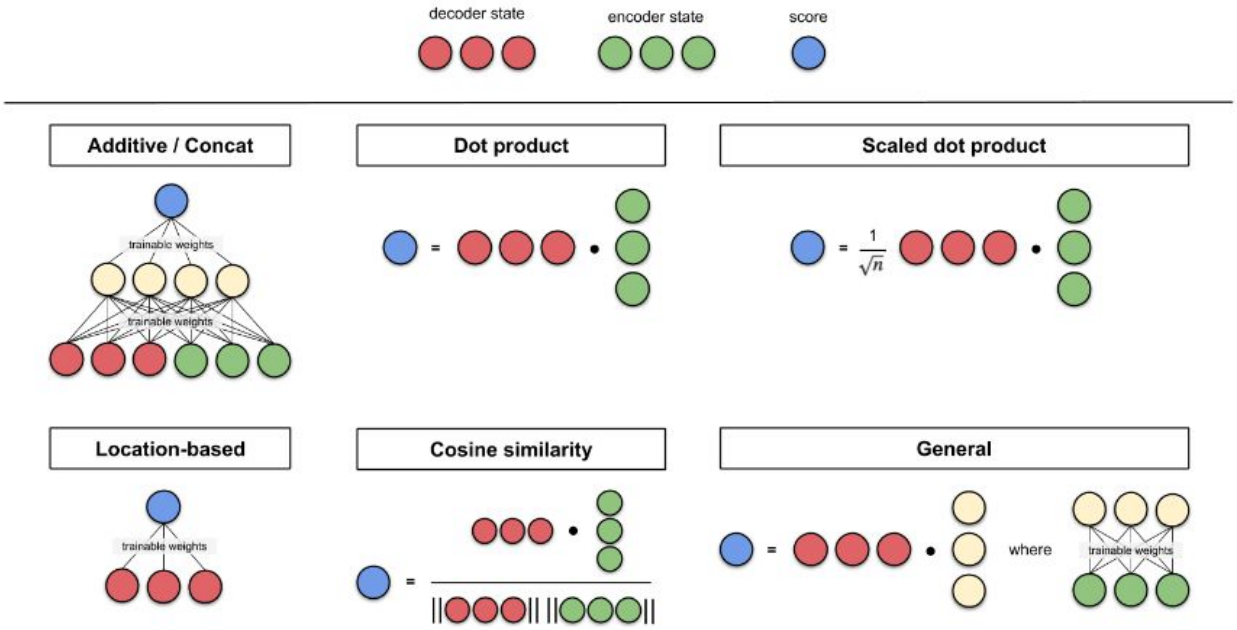
Seq2seq + attention

seq2seq (стандартная)



seq2seq + attention

Виды attention (источник)



Виды attention (источник)

Name	Alignment score function	Citation
Content-base attention	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \text{cosine}[\boldsymbol{s}_t, \boldsymbol{h}_i]$	Graves2014
Additive(*)	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \mathbf{v}_a^\top \tanh(\mathbf{W}_a[\boldsymbol{s}_t; \boldsymbol{h}_i])$	Bahdanau2015
Location-Base	$\alpha_{t,i} = \text{softmax}(\mathbf{W}_a \boldsymbol{s}_t)$ Note: This simplifies the softmax alignment to only depend on the target position.	Luong2015
General	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \boldsymbol{s}_t^\top \mathbf{W}_a \boldsymbol{h}_i$ where $\mathbf{W}_a$ is a trainable weight matrix in the attention layer.	Luong2015
Dot-Product	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \boldsymbol{s}_t^\top \boldsymbol{h}_i$	Luong2015
Scaled Dot-Product(^)	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \frac{\boldsymbol{s}_t^\top \boldsymbol{h}_i}{\sqrt{n}}$ Note: very similar to the dot-product attention except for a scaling factor; where n is the dimension of the source hidden state.	Vaswani2017

Развитие идеи: Transformers

Чатботы

Какие чатботы бывают

- goal-oriented
  - попросить построить маршрут
  - попросить включить песню
  - спросить о погоде
- “болталка”
  - “мне скучно; кто себя создал; ...”
  - бот для психологической помощи
- всё и сразу :)
  - состоит из того и другого

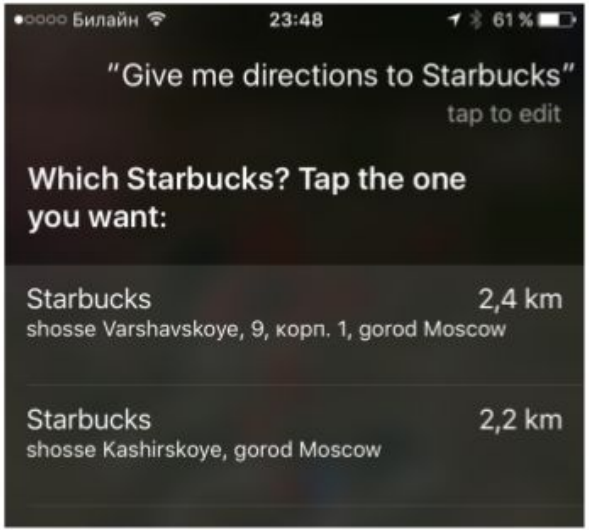
How to train your dialog system

Andrey Zimovnov

Yandex

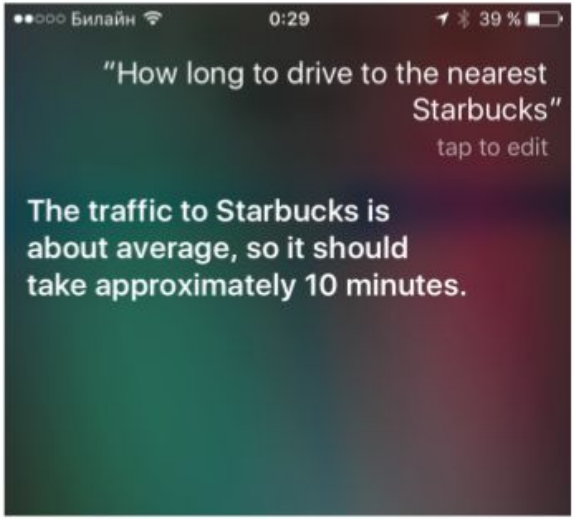
Siri

- Voice assistant from Apple
- How to make one?



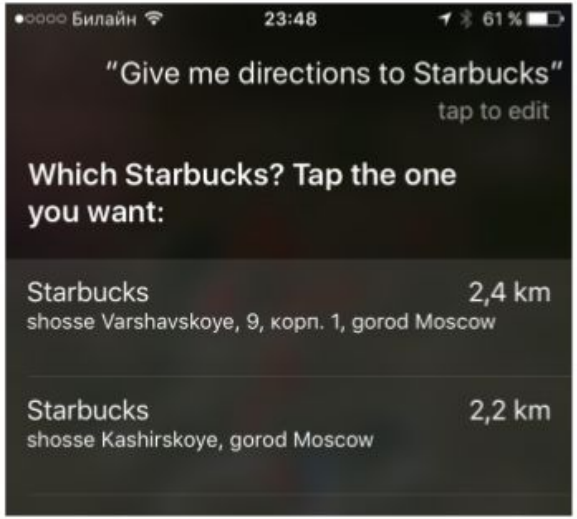
Intent classification

- Intent is **a class of an action** that assistant will execute
- Limited set of intents (navigation, music, notes, calendar, … actions)



Intent: **nav.time.closest**

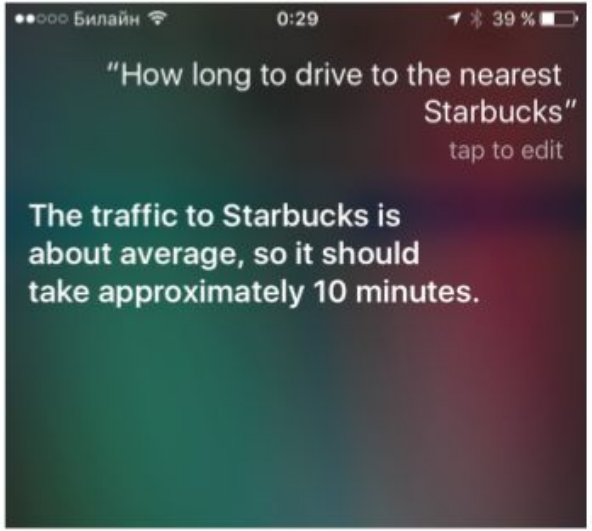
Intent: **nav.directions.closest**



Intent: **nav.directions**

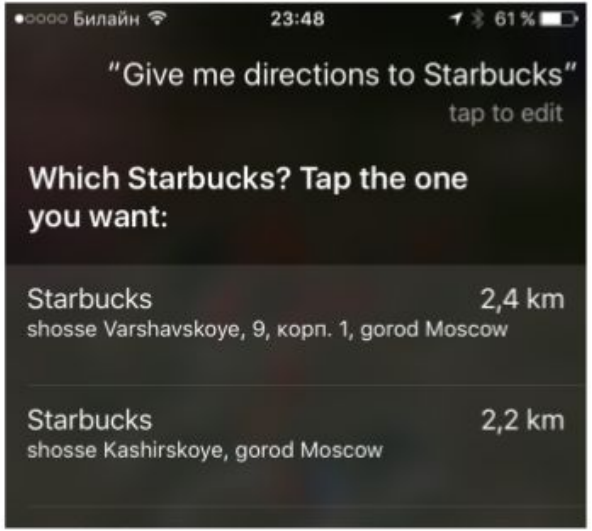
Slot filling

- Slot is **a parameter of an action**
- Limited set of slots (POIs, cities, streets, music artists, … names)



Slots: **@POI.NAME**{Starbucks}

Slots: **@POI.NAME**{Starbucks}



Slots: **@POI.NAME**{Starbucks}

Two classification tasks

- Assume that we have an ASR result (text from speech)

- Marked up utterances (12k):
  - take me to **@NUMBER**{ 1 0 8 2 } **@STREET.NAME**{ east el camino real } **@CITY**{ sunnyvale } → **nav.route.start**

- Intent classification (80 classes):
  - **media.play, nav.find, phone.call, …**

- Slot filling (25 slots):
  - **@MEDIA.ARTIST, @CITY, @CONTACT.NAME, …**

Slot filling as classification

- Assign tag to each word:
  - **B-@CITY** – beginning of the slot **@CITY**
  - **I-@CITY** – continuation of the slot **@CITY**
  - **O** – this word is not used in slots

Sentence	first	class	fares	from	boston	to	denver
Slots	B-class_type	I-class_type	O	O	B-fromloc	O	B-toloc
Intent	airfare						

- Tag classification for each word

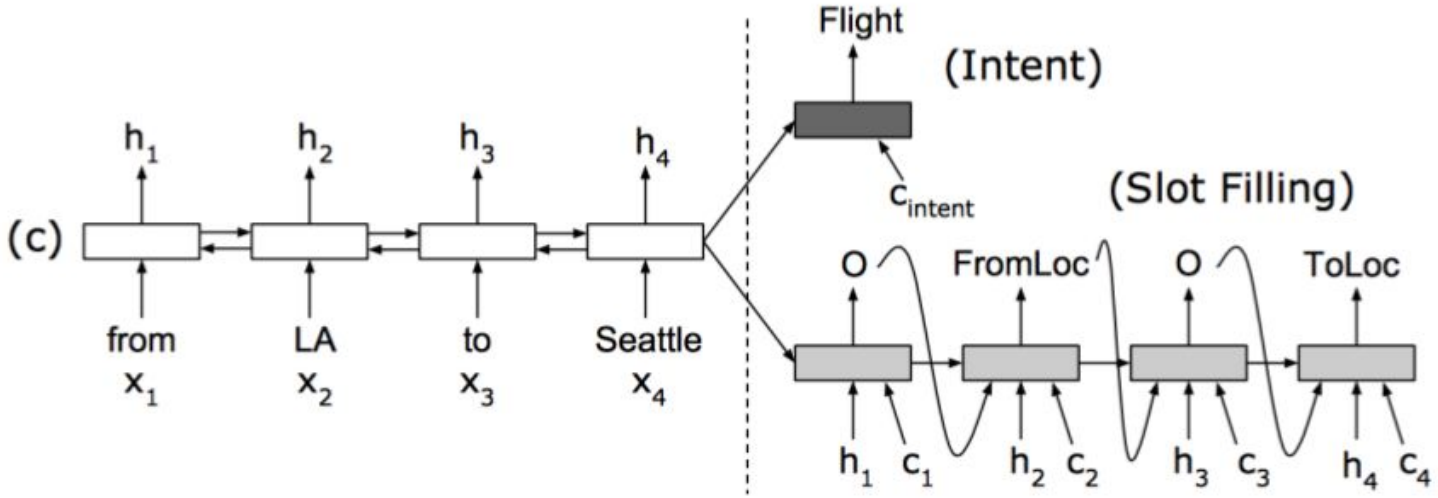
BIOES coding example

- Example for 4 lexicon dictionaries
- B, I, O, E, S are later encoded as one-hot vectors

Text	Hayao	Tada	,	commander	of	the	Japanese	North	China	Area	Army
LOC	–	–	–	–	–	B	I	–	S	–	–
MISC	–	–	–	S	B	B	I	S	S	S	S
ORG	–	–	–	–	–	B	I	B	I	I	E
PERS	B	E	–	–	–	–	–	–	S	–	–

Deep learning

- Encoder-decoder model for **joint** intent detection and slot filling
- Encoder is a bidirectional RNN (LSTM)
- With aligned inputs ($\mathbf{h}_i$ on the right) and attention ($\mathbf{c}_i$ on the right)



Joint training helps, so they say

- Paper results on ATIS (Airline Travel Information Systems) dataset:

Training	Slots F1	Intent % error
Independent training for slot filling	95.78	-
Independent training for intent detection	-	2.02
Joint training for slot filling and intent detection	**95.87**	**1.57**