universität innsbruck

Natural Language to Logical Form

Kamal Zakieldin - Supervised by: Cezary Kaliszyk

Language to Logical Form with Neural Attention

Li Dong and Mirella Lapata
Association for Computational Linguistics, ACL, 2016

Agenda

- Overview
  - Motivation
  - Introduction
- Neural Network Terminology
- Problem Formulation
  - Models
  - Preprocessing
  - Datasets and Results

Agenda

- Overview
  - Motivation
  - Introduction
- Neural Network Terminology
- Problem Formulation
  - Models
  - Preprocessing
  - Datasets and Results

Overview

Content: Understanding natural language has not been treated consistently using logic.

Natural language understanding traditional approaches:

- Natural language processing - Semantic parsing
  - manually-built templates (SIRI)
  - grammar-based mapping (CCG parser)
  - database searching and matching (Watson by IBM)
  - entity mapping (Alexa by Amazon)
- Text entailment
  - Natural language inference without logic
- Logic entailment
  - Controlled natural language.

Overview

Content: Understanding natural language has not been treated consistently using logic.

Natural language understanding modern approaches:

- Text matching
  - Pattern matching using machine learning (Google Assistant)
- Logic entailment
  - ... ?

Overview

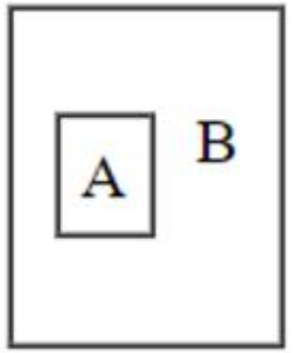
Content: Understanding natural language has not been treated consistently using logic.

Natural language understanding modern approaches:

- Text matching
  - Pattern matching using machine learning (Google Assistant)
- Logic entailment
  - Using deep learning to translate natural language sentences to logical formulas.

Motivation

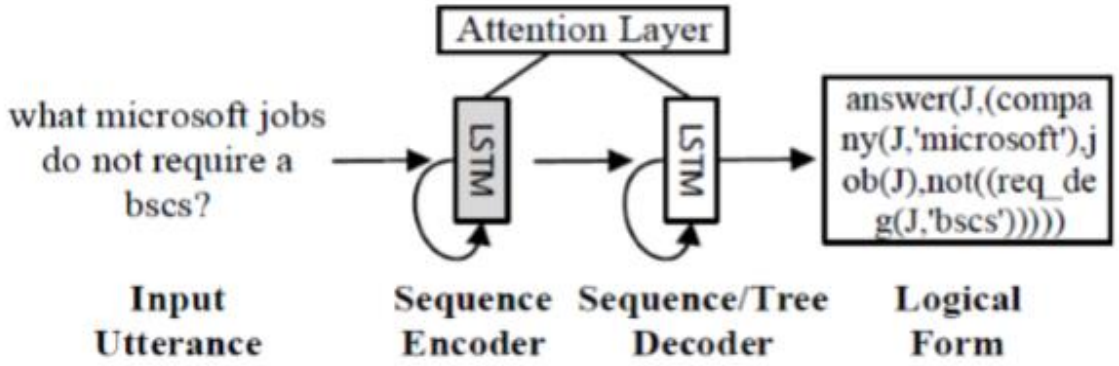
imagine we have a natural language sentence, and we can easily derive a Predicate logic formula of its meaning …



All A's are B's.
Only B's are A's.
$\forall X\ (a(X) \Rightarrow b(X))$

- We can infer if the sentence is valid or not.
- We can translate mathematical written sentences into formulas.
- We can summarize paragraphs.
- We can answer questions logically.
- And more …

Introduction



The main task is translating normal text to a formal representation such as:

- Logical form
- Structured queries

Agenda

- Overview
  - Motivation
  - Introduction
- Neural Network Terminology
- Problem Formulation
  - Models
  - Preprocessing
  - Datasets and Results

Neural Network Terminology

I. Machine learning.
II. Artificial Neural Network (ANN).
III. Recurrent Neural Network (RNN).
IV. Long Short Term Memory (LSTM).
V. Neural Attention.

Artificial Neural Network

Feed Forward Propagation:

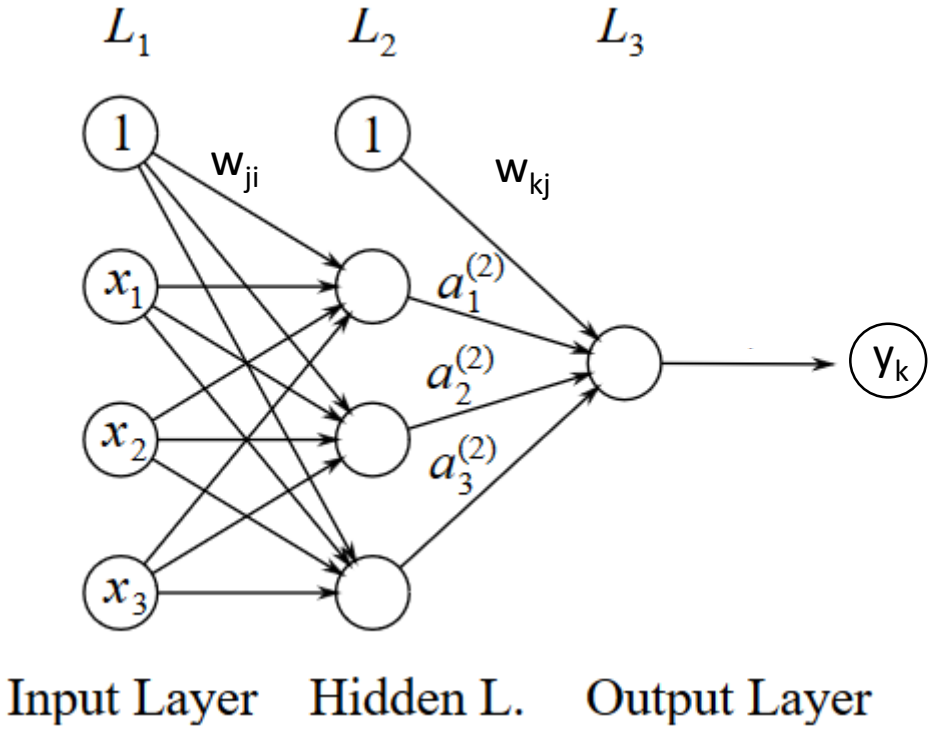
$$a_i^{(1)} = x_i$$

$$a_j^{(2)} = \sum_{i=0}^{n} w_{ji}^{(1)} a_i^{(1)}$$

$$z_j^{(2)} = h(a_j^{(2)})$$

$$a_k^{(3)} = \sum_{j=0}^{n} w_{kj}^{(2)} z_j^{(2)}$$

$$y_k = h(a_k^{(3)})$$



(J. Piater. Advanced machine learning course notes.)

Backpropagation

$$E = \frac{1}{2}(y_k - t)^2$$

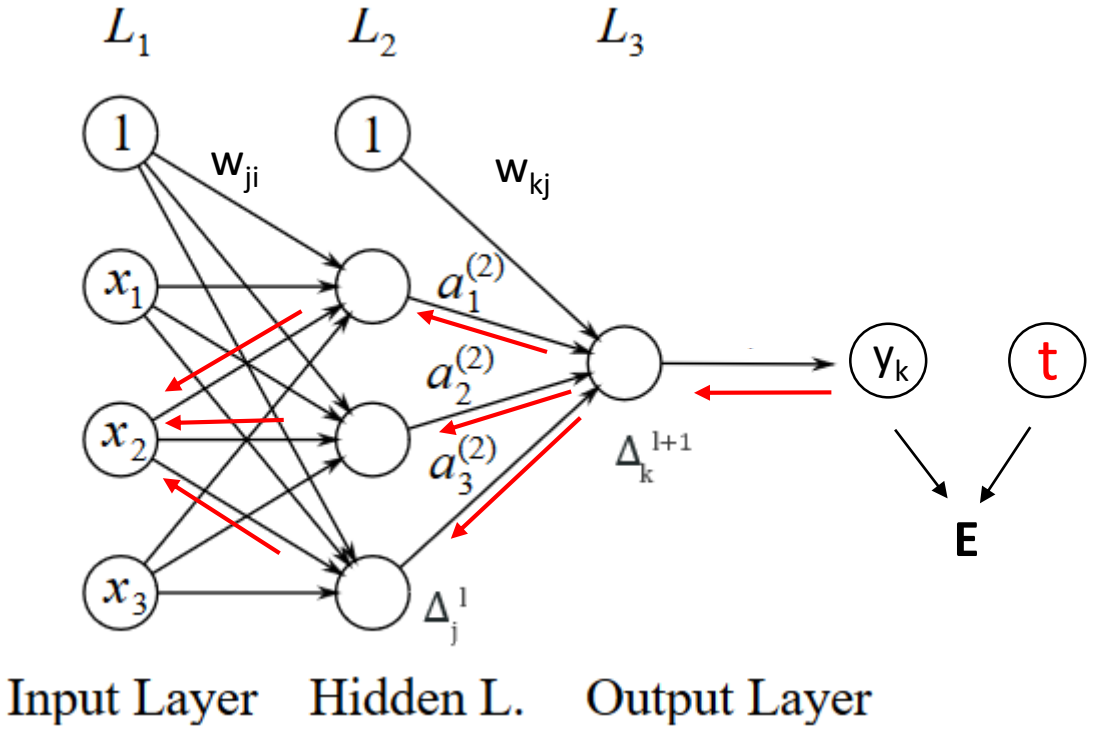
$$w_{kj}^{l} = w_{kj}^{l} - \alpha \frac{\partial E}{\partial w_{kj}^{l}}$$

$$\frac{\partial E}{\partial w_{kj}^{l}} = \frac{\partial E}{\partial a_k^{l+1}} \frac{\partial a_k^{l+1}}{\partial w_{kj}^{l}}$$

$$\frac{\partial a_k^{l+1}}{\partial w_{kj}^{l}} = z_j^{l}$$

$$\frac{\partial E}{\partial a_j^{l}} = \Delta_j^{l} = \sum_k \Delta_k^{l+1} \frac{\partial a_k^{l+1}}{\partial a_j^{l}}$$

$$= \dot{h}(a_j^{l}) \sum_k \Delta_k^{l+1} w_{kj}^{l}$$



(J. Piater. Advanced machine learning course notes.)

Recurrent Hidden Units

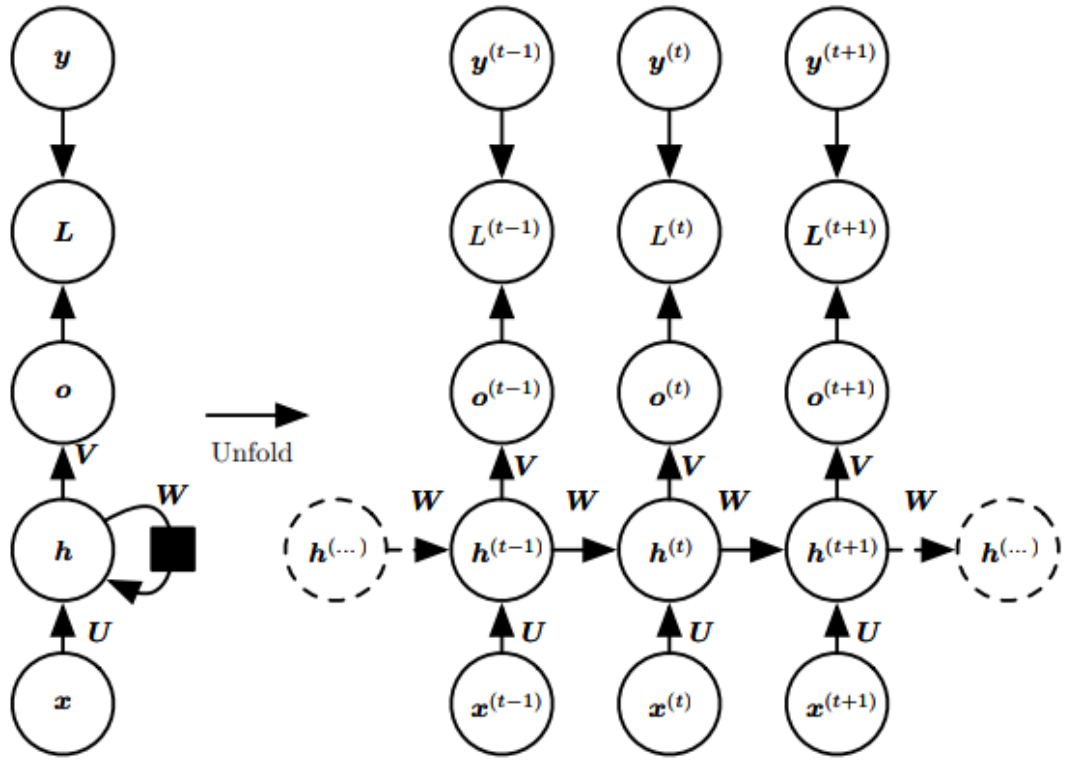


Figure 10.3

(Goodfellow 2016)

LSTM

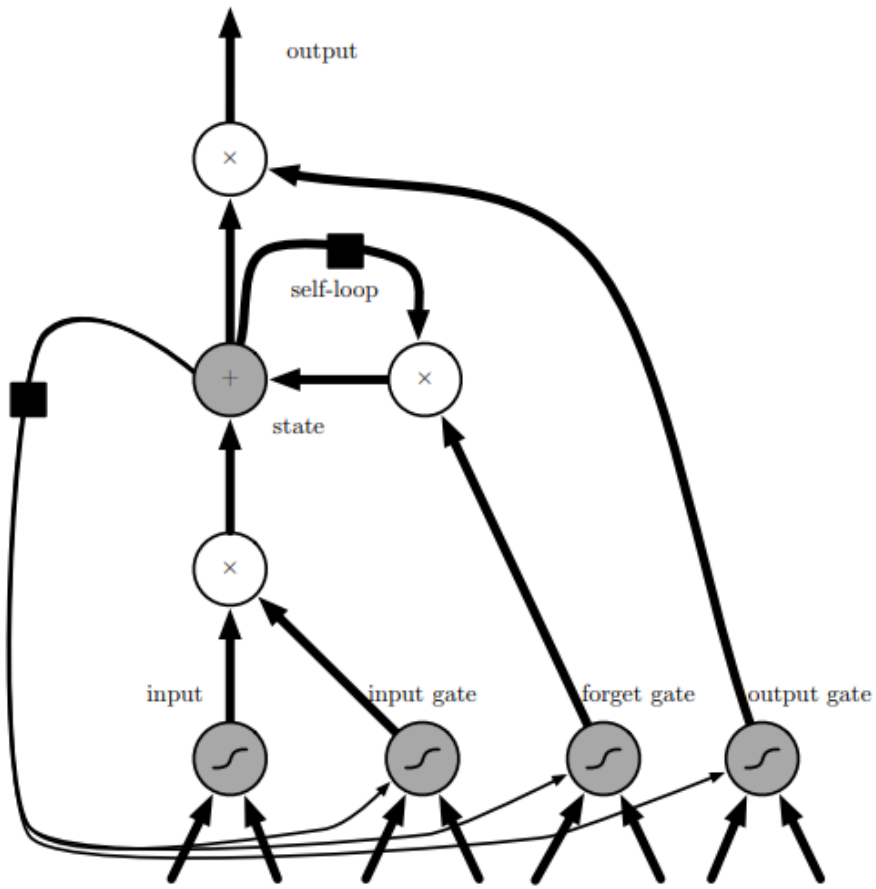


Figure 10.16

(Goodfellow 2016)

Sequence to Sequence Architecture

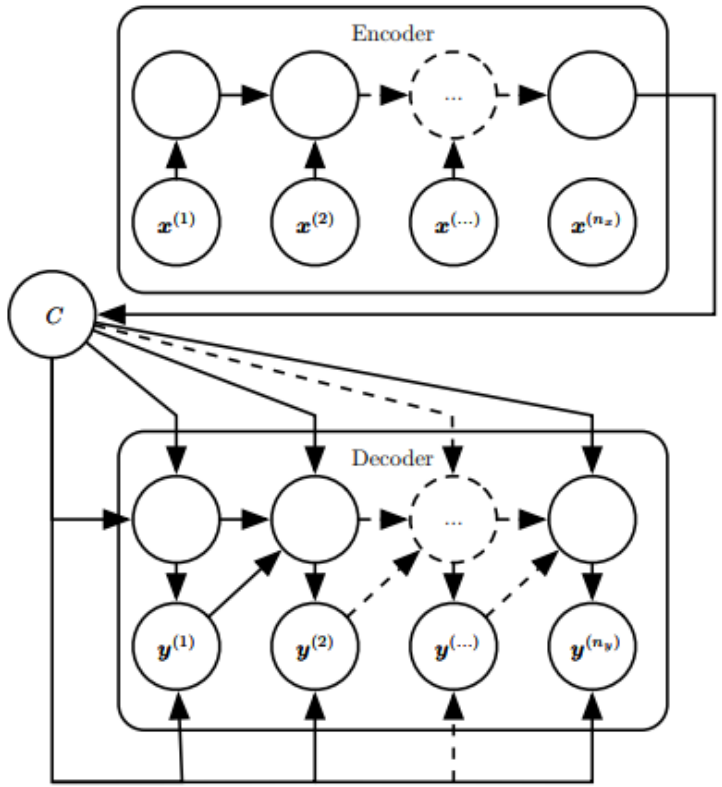


Figure 10.12

(Goodfellow 2016)

Neural Attention

- Idea of Neural Attention
  - Focus

Neural Attention

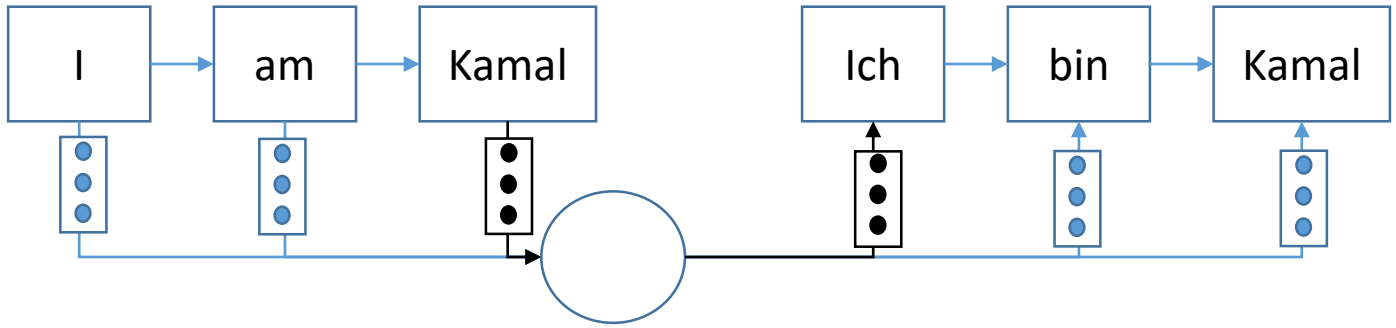
- Idea of Neural Attention
  - Focus

- Machine translation problems:
  - Length of the sentence.

Neural Attention

- Idea of Neural Attention
  - Focus

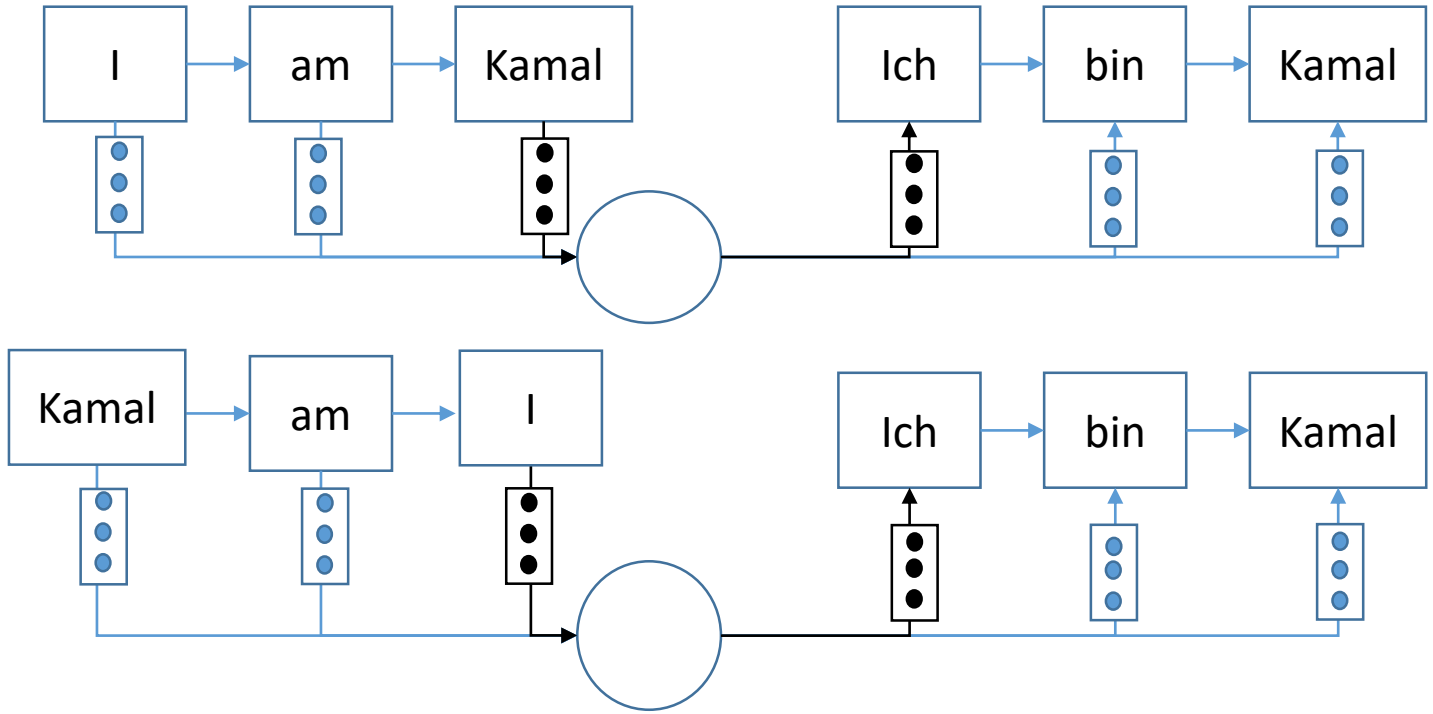
- Machine translation problems:
  - Length of the sentence.
  - Alignment relevant info.



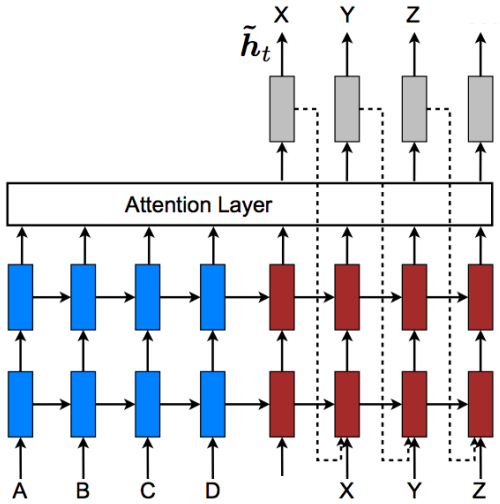
Neural Attention

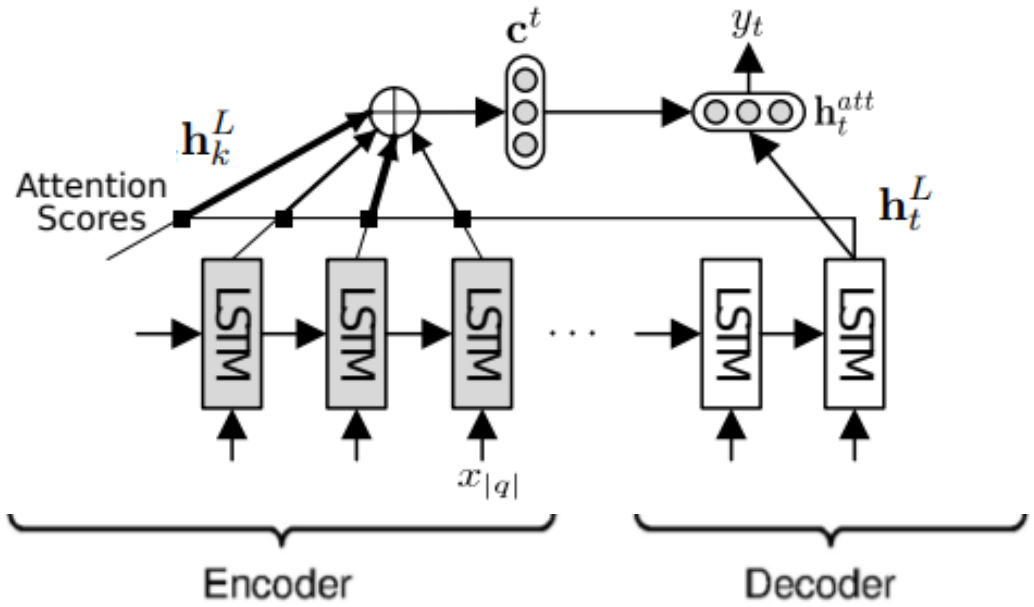
- Idea of Neural Attention
  - Focus

- Machine translation problems:
  - Length of the sentence.
  - Alignment relevant info.



Neural Attention





$$s_k^t = \frac{\exp\{\mathbf{h}_k^L \cdot \mathbf{h}_t^L\}}{\sum_{j=1}^{|q|} \exp\{\mathbf{h}_j^L \cdot \mathbf{h}_t^L\}}$$

$$\mathbf{c}^t = \sum_{k=1}^{|q|} s_k^t \mathbf{h}_k^L$$

$$\mathbf{h}_t^{att} = \tanh\left(\mathbf{W}_1 \mathbf{h}_t^L + \mathbf{W}_2 \mathbf{c}^t\right)$$

Agenda

- Overview
  - Motivation
  - Introduction
- Neural Network Terminology
- Problem Formulation
  - Models
  - Preprocessing
  - Datasets and Results

Problem formulation

Model maps **natural language input** $q = x_1 \ldots x_{|q|}$ to a **logical form** representation of its meaning $a = y_1 \ldots y_{|a|}$.

$$p(a|q) = \prod_{t=1}^{|a|} p(y_t|y_{<t}, q)$$

where $y_{<t} = y_1 \cdots y_{t-1}$

- Encoder encodes natural language input q into a vector representation.
- Decoder generates $y_1 \cdots y_{|a|}$ conditioned on the encoding vector.

Seq2seq Model

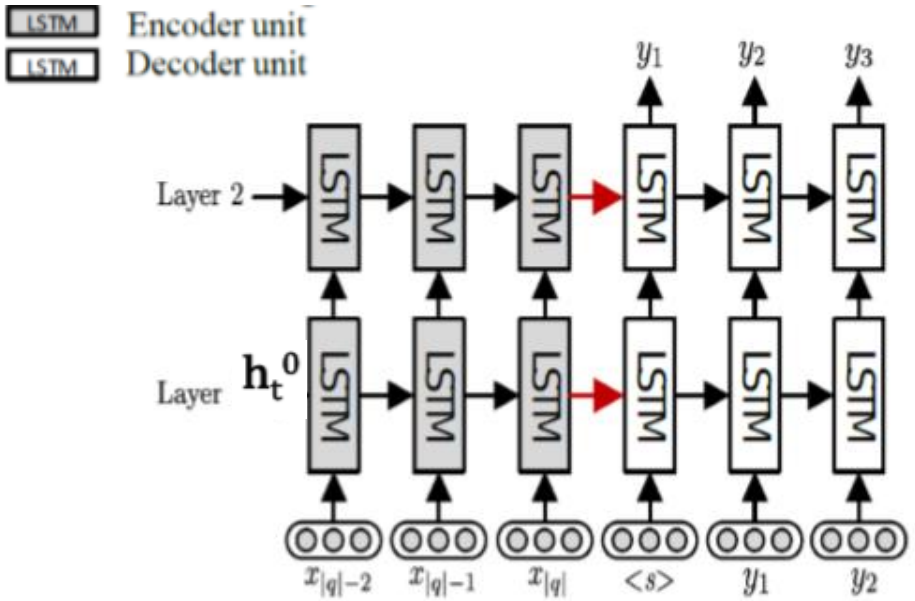


Figure 2: Sequence-to-sequence (SEQ2SEQ) model with two-layer recurrent neural networks.

Encoder:

$$\mathbf{h}_t^0 = \mathbf{W}_q \mathbf{e}(\mathbf{x}_t)$$

Decoder:

$$\mathbf{h}_t^0 = \mathbf{W}_a \mathbf{e}(\mathbf{y}_{t-1})$$

Predicted Output :

$$p(y_t|y_{<t}, q) = softmax(W_a \; h^L_t)^T \, e(y_t)$$

$$p(a|q) = \prod_{t=1}^{|a|} p(y_t|y_{<t}, q)$$

Seq2tree model

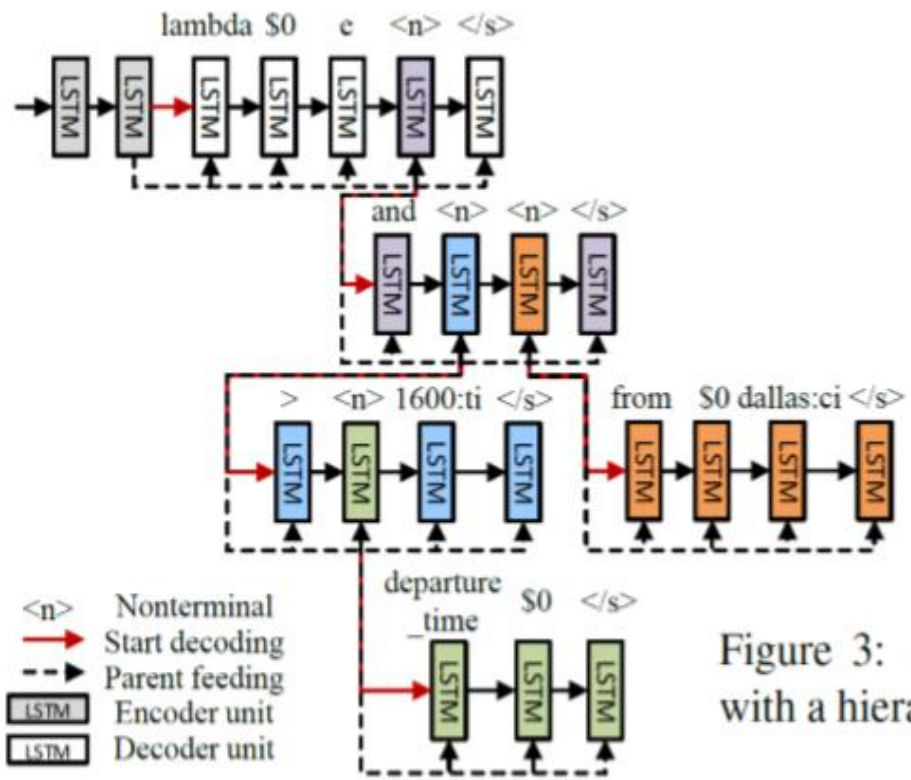
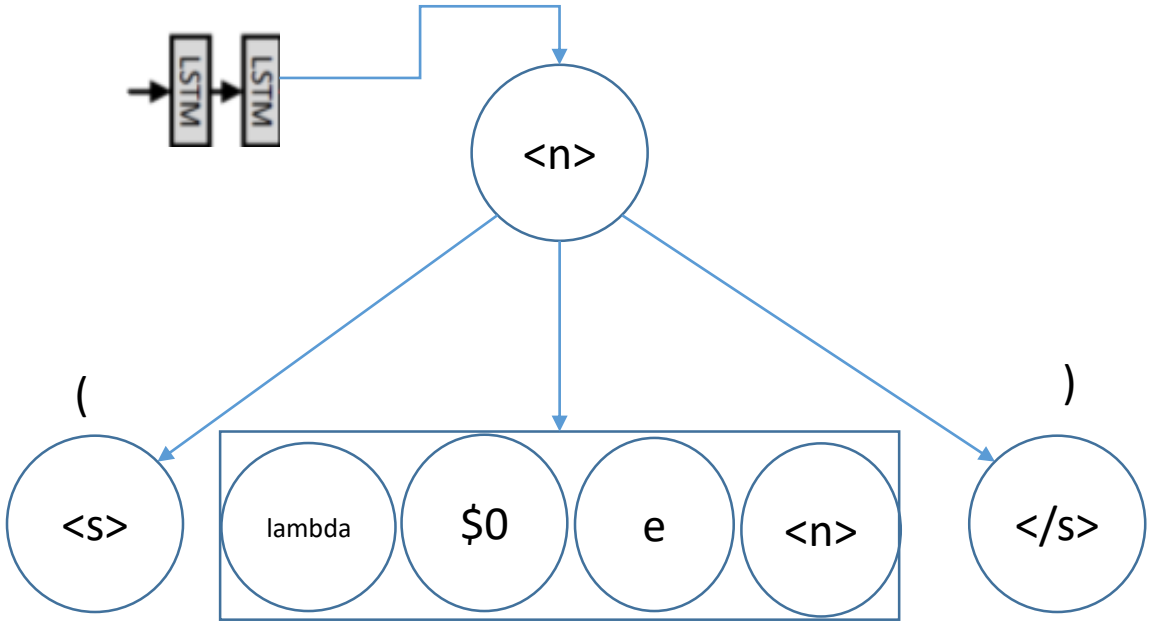


Figure 3: Sequence-to-tree (SEQ2TREE) with a hierarchical tree decoder.

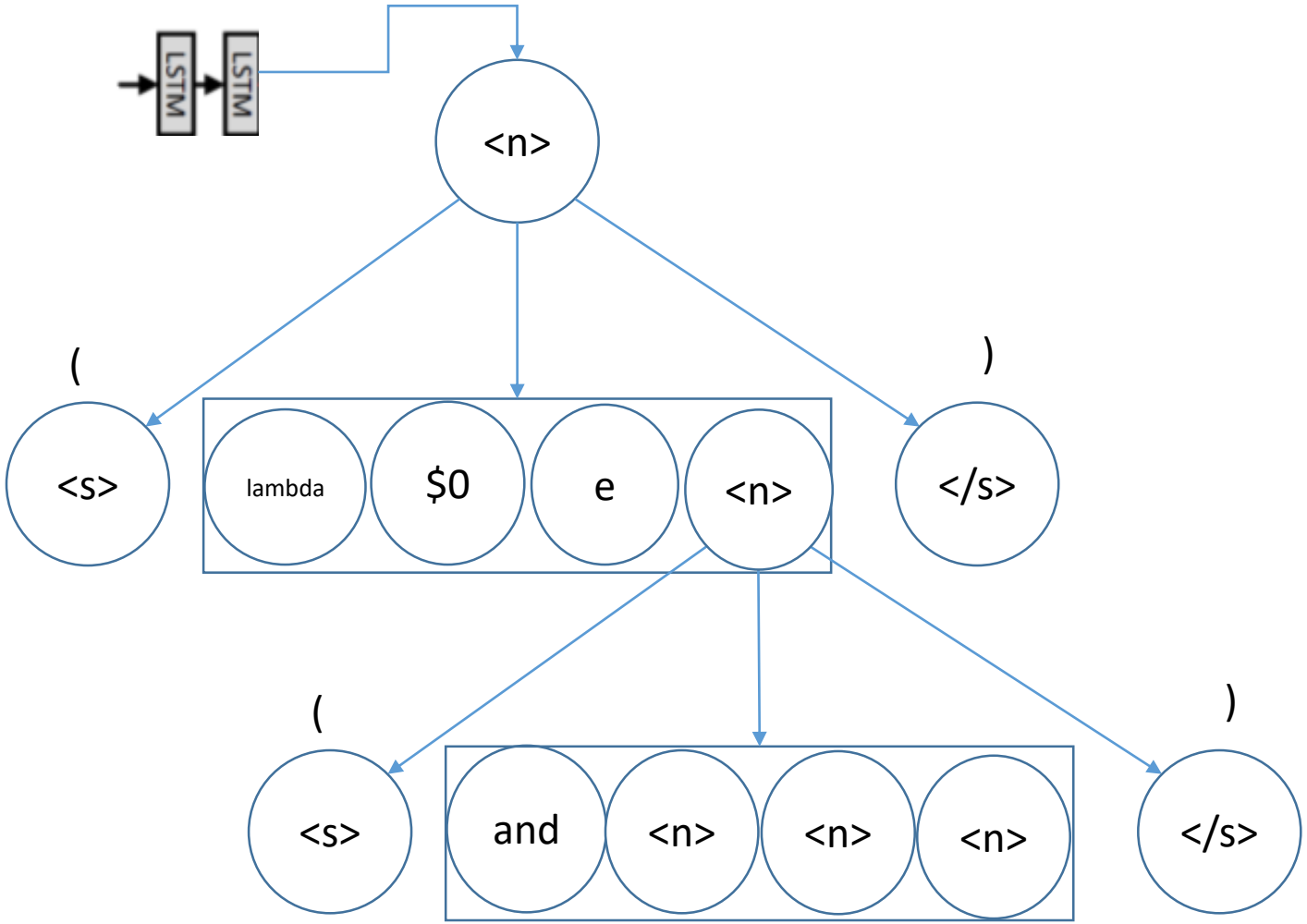
Seq2Tree

- Ex:
  - Input sentence:
    - Flight from Dallas to San_francisco leaving after 4 in the afternoon please.
  - Generated formula:
    - (lambda $0 e (and (> (departure_time $0) 1600 : t_i) (from $0 dallas : C_i) (to $0 San_francisco : C_i)))

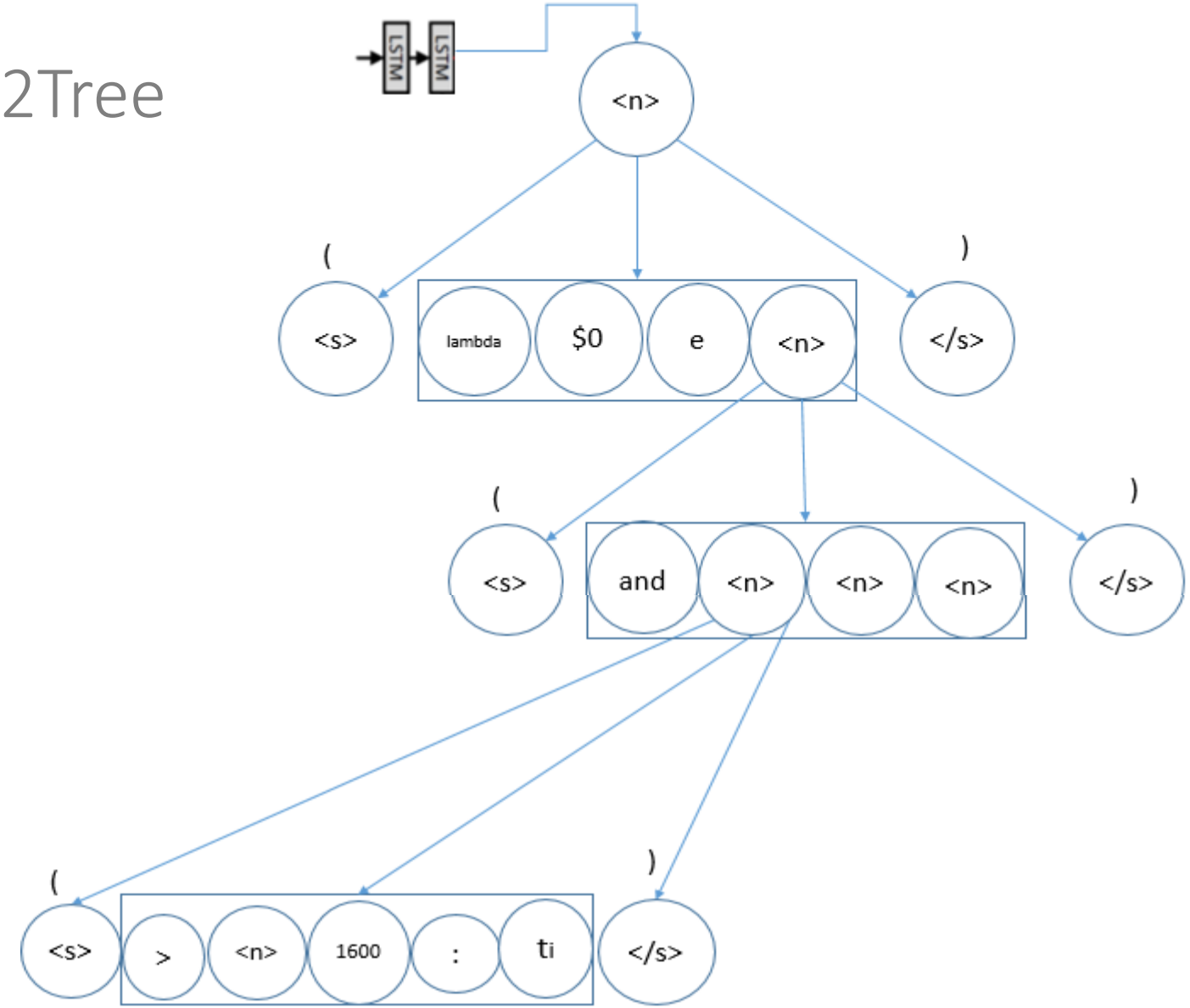
Seq2Tree



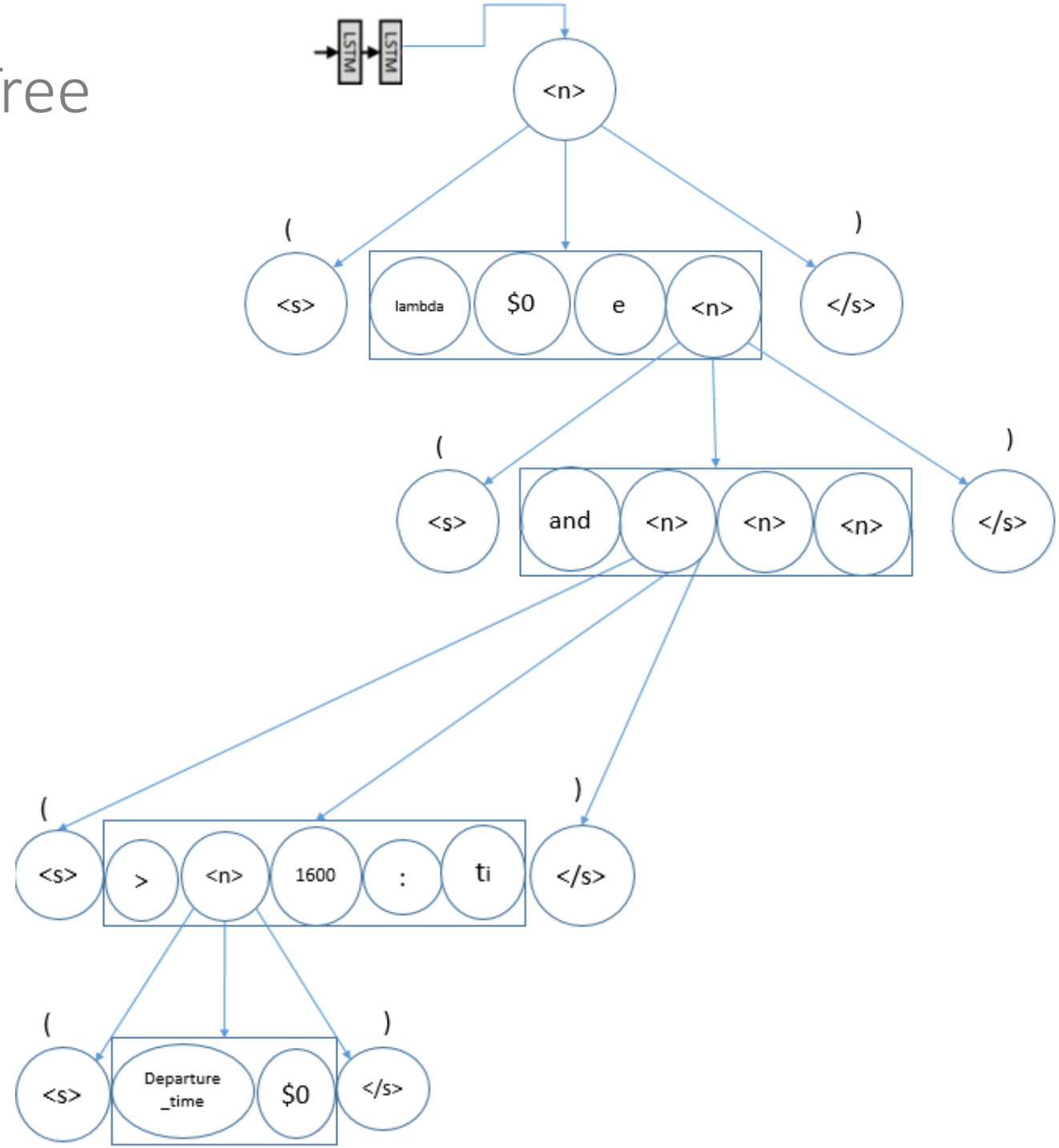
Seq2Tree



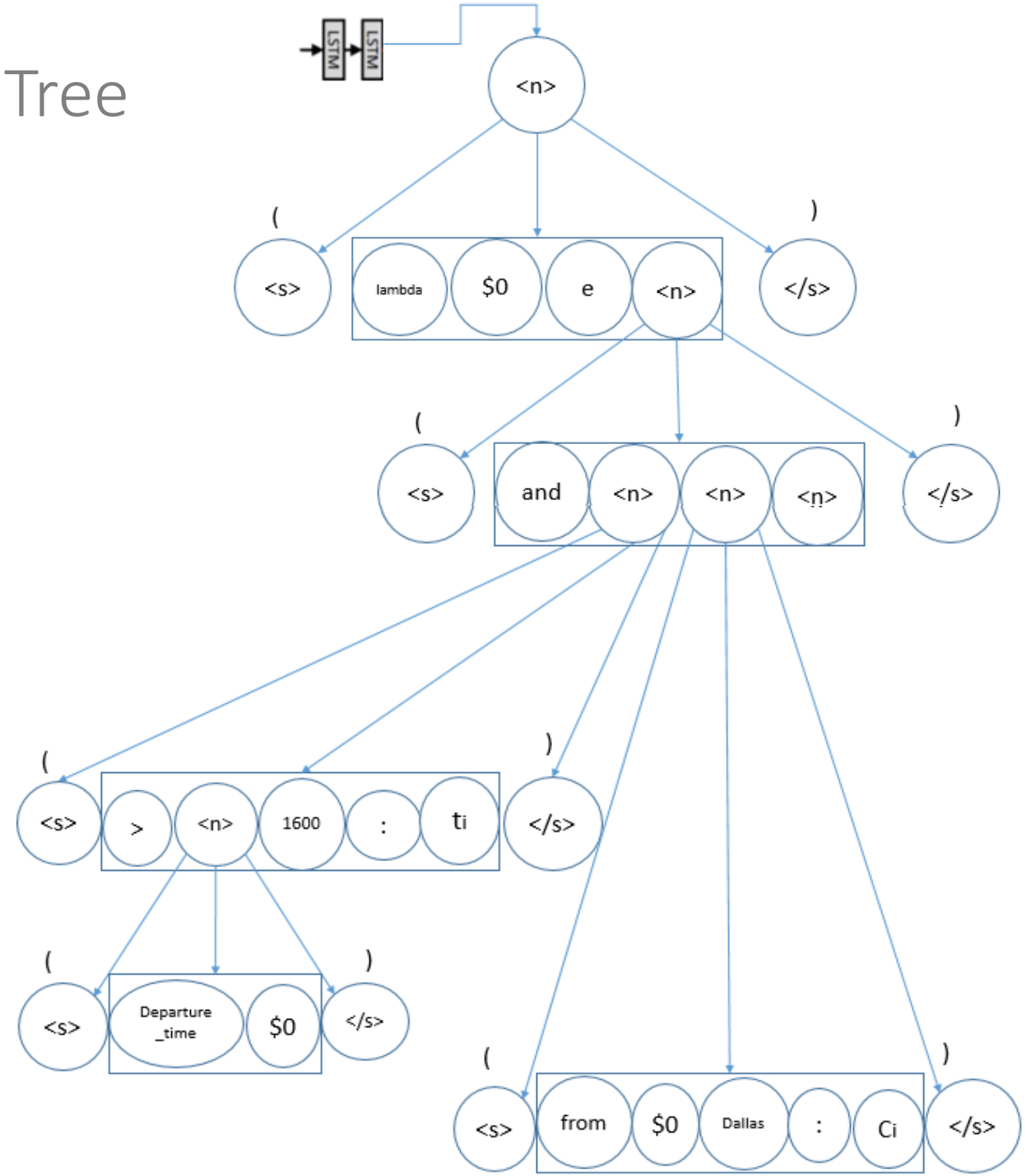
Seq2Tree



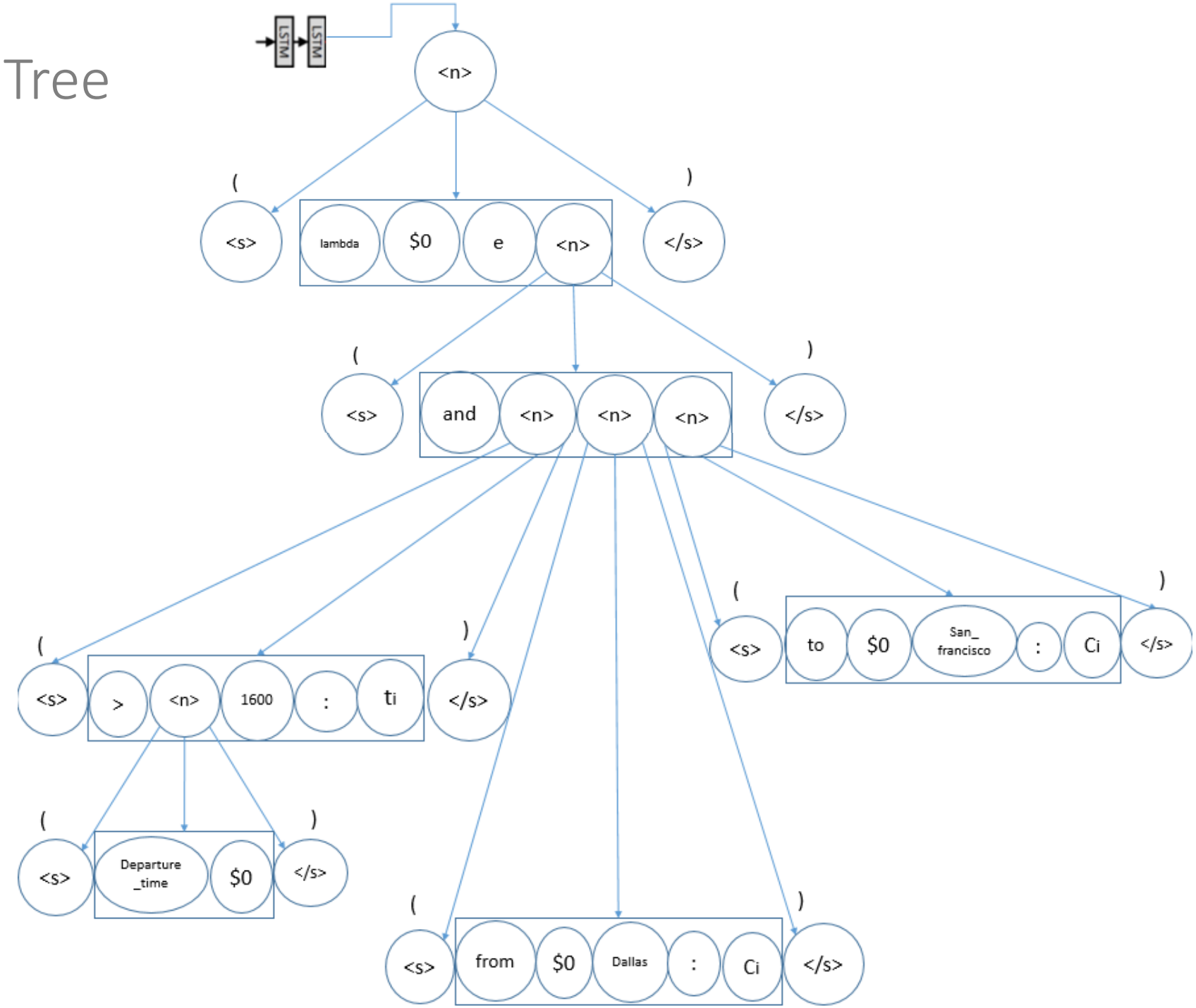
Seq2Tree


LSTM
LSTM
<n>
(
)
<s>
lambda
$0
e
<n>
</s>
(
)
<s>
and
<n>
<n>
<n>
</s>
(
)
<s>
></s>
<n>
1600
:
ti
(
)
<s>
Departure _time
$0
</s>

Seq2Tree


LSTM
LSTM
<n>
(
)
<s>
lambda
$0
e
<n>
</s>
(
)
<s>
and
<n>
<n>
<n>
</s>
(
)
<s>
>
<n>
1600
:
ti
</s>
(
)
<s>
Departure
_time
$0
</s>
(
)
<s>
from
$0
Dallas
:
Ci
</s>

Seq2Tree


LSTM
LSTM
<n>
(
<s>
lambda
$0
e
<n>
)
</s>
(
<s>
and
<n>
<n>
<n>
)
</s>
(
<s>
>
<n>
1600
:
ti
)
</s>
(
<s>
Departure _time
$0
)
</s>
(
<s>
from
$0
Dallas
:
Ci
)
</s>
(
<s>
to
$0
San_ francisco
:
Ci
)
</s>

Pre-processing

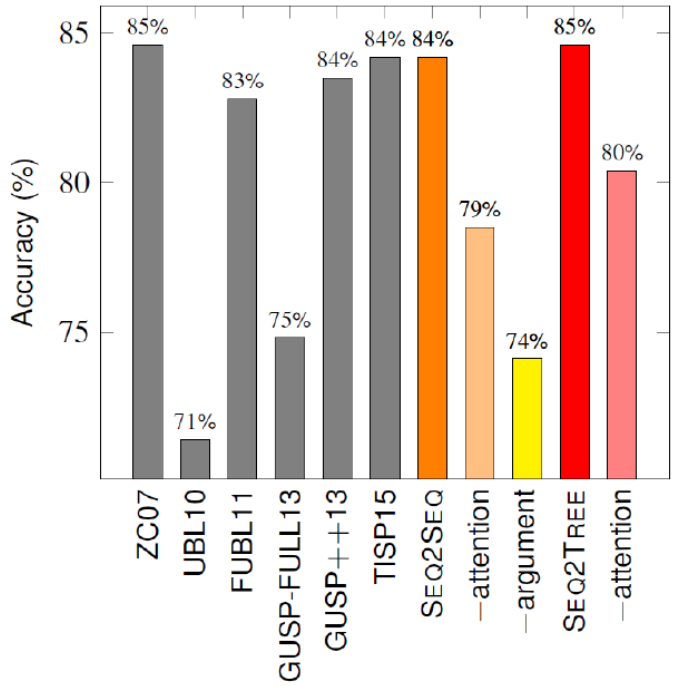
- Argument identification
  - Replace unknown rare entities with special symbol.
  - Replace numbers with special symbol.
- Misspelling correction.
- Stemming words.
- Pre-process subtrees for the seq2tree model.

- A post-processing step to recover them back.

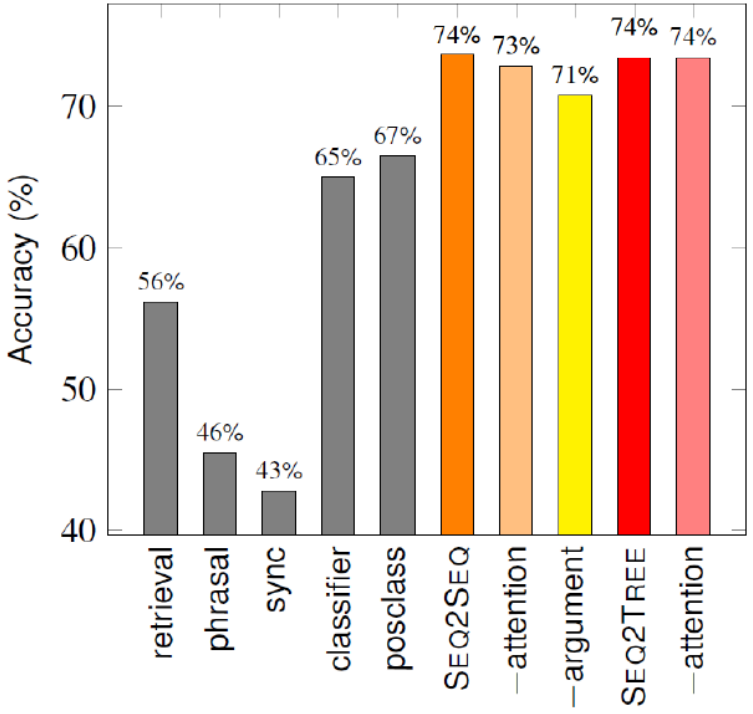
Datasets

Dataset	Num of examples	Description	Example
JOBS	640	Question about Jobs.	• what microsoft jobs do not require a bscs? ➢ answer(company(J,'microsoft '),job(J),not((req deg(J,'bscs'))))
GEO	880	U.S. Geography	• what is the population of the state with the largest area? ➢ (population:i (argmax $0 (state:t $0) (area:i $0)))
ATIS	5410	Flight booking system	• dallas to san_francisco leaving after 4 in the afternoon please ➢ (lambda $0 e (and (>(departure time $0) 1600:ti) (from $0 dallas:ci) (to $0 san_francisco:ci)))
IFTTT	85K	If this then that. are selected from various sources.	• Turn on heater when temperature drops below 58 degree ➢ **TRIGGER**: Weather - Current temperature drops below - ((Temperature (58)) (Degrees in (f))) ➢ **ACTION**: WeMo Insight Switch - Turn on - ((Which switch? ("")))

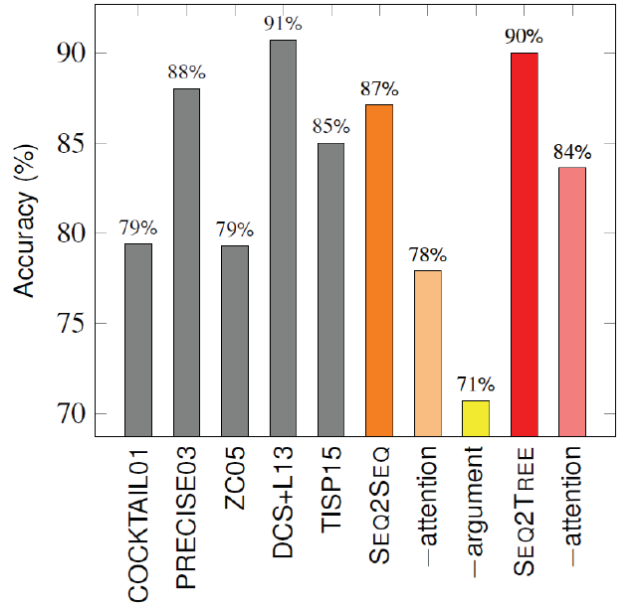
Results



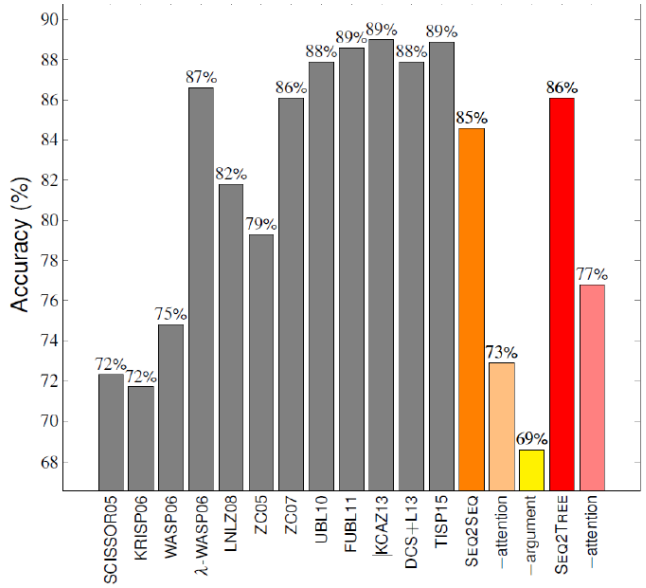
Evaluation results on ATIS.



Evaluation results on IFTTT.



Evaluation results on JOBS.



Evaluation results on GEO.

Conclusion

- Seq2tree model
- Neural attention is all you need.
- The potential of converting natural language sentence to predicate logic.
  - Then Reasoning.

References

- I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press 2016.
- Dong, Li, and Mirella Lapata. Language to Logical Form with Neural Attention, ACL 2016.
- Justus Piater. Advanced machine learning course: 703642.
- Tejas Khot, Language logical form presentation.
- Zhongen Li, Ohio Uni. Language to logcial form presentation.
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Thank you for your attention!

Kamal Zakieldin