Language Understanding

08 - Semantic Parsing

Hossein Zeinali

Introduction

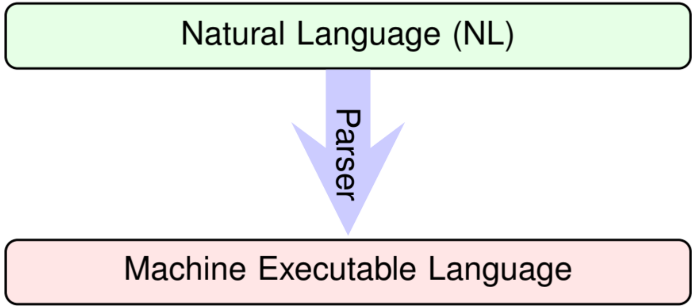
- **Semantic parsing** is the task of translating natural language into a formal meaning representation on which a machine can act.
  - Representations may be an executable language such as SQL or more abstract representations such as First Order Logic (FOL)
- Examples:

What states border Texas?

$$\lambda x.state(x)\ \wedge\ borders(x, Texas)$$

What is the largest state?

$$argmax(\lambda x.state(x)\ \wedge\ \lambda x.size(x))$$



Introduction

- Applications of semantic parsing
  - Machine translation
  - Question answering
  - Ontology induction
  - Automated reasoning
  - Code generation
- Types of semantic parsing:
  - Shallow
    - Identifying entities in an utterance and labelling them with the roles they play, sometimes known as slot-filling.
  - Deep
    - Producing precise meaning representations of utterances that can contain significant compositionality. Doing both intent detection and slot-filling.

Introduction

- Semantic parsing example: Querying a Database

NL What are the capitals of states bordering Texas?

↓ Parser

DB $\lambda x.\ \text{capital}(y, x) \wedge \text{state}(y) \wedge \text{next\_to}(y, \text{Texas})$

Introduction

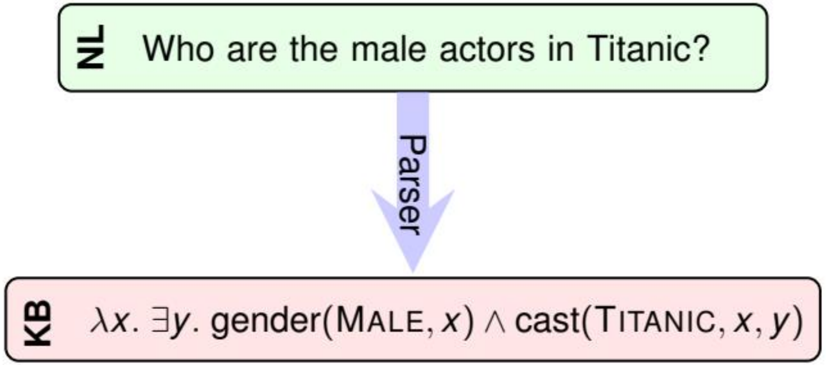
- Semantic parsing example: SQL Queries

which country had the highest carbon emissions last year

```
SELECT   country.name
FROM     country, co2_emissions
WHERE    country.id = co2_emissions.country_id
AND      co2_emissions.year = 2014
ORDER BY co2_emissions.volume DESC
LIMIT    1;
```

Introduction

- Semantic parsing example: Answering Questions with Freebase



Introduction

- Semantic parsing example: WikiTableQuestions

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

x = Greece held its last Summer Olympics in which year?

y = 2004

Logic Games From LSAT

- Six sculptures — C, D, E, F, G, H — are to be exhibited in rooms 1, 2, and 3 of an art gallery.
  - Sculptures C and E may not be exhibited in the same room.
  - Sculptures D and G must be exhibited in the same room.
  - If sculptures E and F are exhibited in the same room, no other sculpture may be exhibited in that room.
  - At least one sculpture must be exhibited in each room, and no more than three sculptures may be exhibited in any room.
- If sculpture D is exhibited in room 3 and sculptures E and F are exhibited in room 1, which of the following may be true?
  A. Sculpture C is exhibited in room 1.
  B. Sculpture H is exhibited in room 1.
  C. Sculpture G is exhibited in room 2.
  D. Sculptures C and H are exhibited in the same room.
  E. Sculptures G and F are exhibited in the same room.

Challenge: Anaphoric Adjectives

- "Sculptures D and G must be exhibited in the **same** room."

  (if (and (in D X) (in G Y)) (= X Y))

- "Sculptures D and G must be exhibited in the **red** room."

  (if (and (in D X) (in G Y)) (and (red X) (red Y)))

- "Sculptures D and G must be exhibited in the **large** room."

  (if (and (in D X) (in G Y)) (and (large X) (large Y)))

- "Sculptures D and G must be exhibited in the **smelly** room."

  (if (and (in D X) (in G Y)) (and (smelly X) (smelly Y)))

Challenge: Ambiguity

- **Ambiguity** can occur at various levels of NLP.
    - It could be Lexical, Syntactic, Semantic, and etc.
- Examples:
- “The car hit the pole while it was moving.”
    - The car, while moving, hit the pole.
    - The car hit the pole while the pole was moving.
- “The man saw the girl with the telescope.”
    - The man saw the girl carrying a telescope.
    - The man saw the girl through his telescope.
- “بخشش لازم نیست اعدامش کنید”

SHRDLU Demo (Winograd 1972)

- Find a block which is taller than the one you are holding and put it into the box.

  OK.

- How many blocks are not in the box?

  FOUR OF THEM.

- Is at least one of them narrower than the one which I told you to pick up?

  YES, THE RED CUBE.

CHAT-80

- Chat-80 was a natural language system which allowed the user to interrogate a Prolog knowledge base in the domain of world geography.
- It was developed in the early '80s by Warren and Pereira
- Proof-of-concept natural language interface to database
- Implemented in Prolog
- Hand-built lexicon & grammar
- Highly influential NLIDB system

Things You Could Ask CHAT-80

- Is there more than one country in each continent?
- What countries border Denmark?
- What are the countries from which a river flows into the Black sea?
- What is the total area of countries south of the Equator and not in Australasia?
- Which country bordering the Mediterranean borders a country that is bordered by a country whose population exceeds the population of India?
- How far is London from Paris?

Precision vs Robustness

Precise, complete understanding

SHRDLU CHAT-80

Brittle, narrow coverage

Robust, broad coverage

Fuzzy, partial understanding

Semantic Parsing vs. Machine Translation

- Both involve translating from one semantic representation into another.
- Both involve complex structures, often related in complex ways.
- Some techniques are transferable: coreference resolution, alignment, …
- But in machine translation, the target semantic representation is not machine-readable! Rather, it is human-readable.

```
(TravelQuery (Destination /m/0d6lp) (Mode TRANSIT))
```

directions	*to*	*SF*	*by*	*train*
Wegbeschreibung	*nach*	*SF*	*mit*	*dem Zug*

Standard Datasets and Evaluation Metrics

ATIS

- ATIS is a spontaneous spoken language data instead of the read speech.
- It contains 5,410 spoken queries for air travel information.
- Utterances were recorded, manually transcribed, and manually categorized into three different classes:
  - Class A: queries that can be interpreted without the context information;
  - Class D: queries that can be interpreted with the context information;
  - Class X: unanswerable queries (e.g. out-of-domain utterances).
- The corpus also contains a back-end database of US domestic flights:
  - Each utterance in class A or D is also manually associated with a target SQL query, together with the reference answer, which is obtained by executing the query on the database.

Training Dataset	Development Dataset	Test Dataset
4480 instances	480 instances	450 instances

GeoQuery

- GeoQuery contains a small database of information about United States geography. It has about 880 facts, represented as Prolog assertions. It also has SQL queries.
- The database mainly contains the following information:
  - States - their capitals, populations, areas, population densities, major cities, rivers and the bordering states
  - Cities - their populations and the states they are in
  - Rivers - their lengths and the states through which they flow
  - Mountains - their heights and the states they are in

Training Dataset	Test Dataset
600 queries	280 queries

GeoQuery

- GeoQuery contains a small database of information about United States geography. It has about 800 facts, represented as Prolog

Question	SQL query
what is the biggest city in arizona	SELECT CITYalias0.CITY_NAME FROM CITY AS CITYalias0 WHERE CITYalias0.POPULATION = (SELECT MAX(CITYalias1.POPULATION) FROM CITY AS CITYalias1 WHERE CITYalias1.STATE_NAME = "arizona") AND CITYalias0.STATE_NAME = "arizona"

Training Dataset	Test Dataset
600 queries	280 queries

Evaluation Metrics

- **Sentence/Utterance Level Semantic Accuracy (SLSA):** this metric measures the percentage of the correct semantic representations assigned to the sentences/utterances in a test set.

$$\frac{\#\ of\ sentences\ assigned\ the\ correct\ semantic\ representation}{\#\ of\ sentences}$$

- **Slot Error Rate** (SER): measures the slot level performance of a frame-based SLU system.

$$\frac{\#\ of\ inserted/deleted/substituted\ slots}{\#\ of\ slots\ in\ the\ reference\ semantic\ representations}$$

Evaluation Metrics

- **Slot Precision/Recall/F1 Score:**

$$Precision = \frac{\text{\# of reference slots correctly detected by SLU}}{\text{\# of total slots detected by SLU}}$$

$$Recall = \frac{\text{\# of reference slots correctly detected by SLU}}{\text{\# of total reference slots}}$$

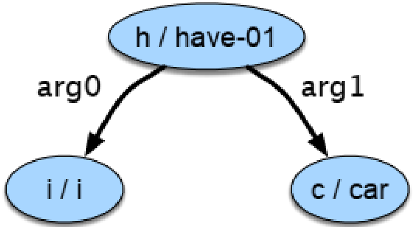
$$F_1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

Logical Representations of Sentence Meaning

Meaning Representations

- **Meaning Representations (MR)**: the meaning of linguistic expressions can be captured in formal structures called MR
- Example: MRs for the sentence ***I have a car*** using four MR languages

$$\exists e,y\ Having(e) \wedge Haver(e,Speaker) \wedge HadThing(e,y) \wedge Car(y)$$




```
(h / have-01
      arg0: (i / i)
      arg1: (c / car))
```

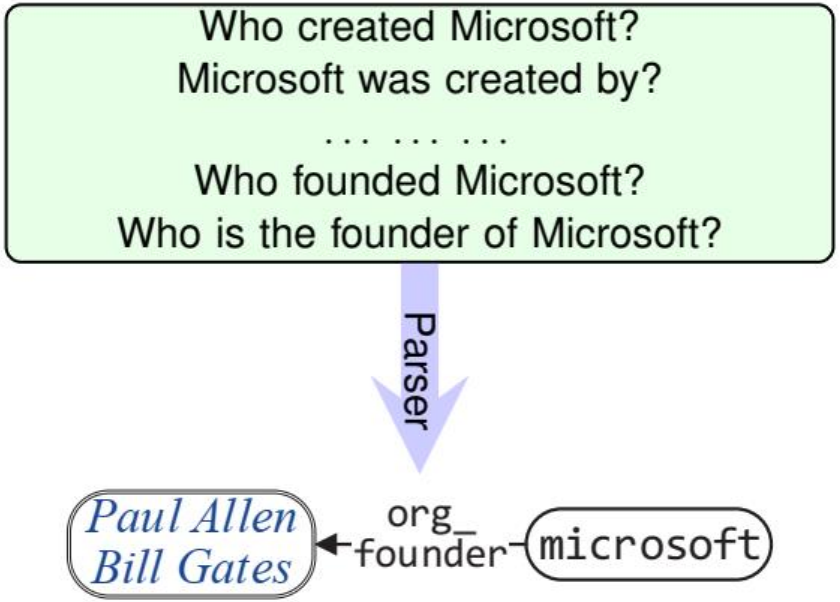
```
Having:
     Haver:   Speaker
     HadThing:   Car
```

Meaning Representations

- Problem with Natural Language (NL):
  - Much of natural language is unverifiable, ambiguous, non-canonical.
    - These are the main reasons that meaning representations are needed.
- But, meaning representations are:
  - **Verifiable**: a system's ability to compare the state of affairs described by a representation to the state of affairs in some world as modeled in a knowledge base.
  - **Unambiguous**: a sentence in NL can be ambiguous; a single linguistic expression can have more than one meanings. But MR cannot be ambiguous.
  - **Canonical**: distinct inputs that mean the same thing should have the same MR.

Meaning Representations

- Canonical example:



Direct Meaning vs Inference

- Compare following two sentences:
  - Does Maharani have vegetarian dishes?
  - Can vegetarians eat at Maharani?
- There is a common-sense connection between what vegetarians eat and what vegetarian restaurants serve.
  - Common sense reasoning task
- **Inference**: draw valid conclusions based on the meaning representation of inputs and background knowledge in the database.

First-Order Logic

- **First-Order Logic (FOL)** is a flexible, well-understood, and computationally tractable meaning representation language.
- **Term**: the FOL device for representing objects. Term defines an object in the world under consideration.
- **Constants**: refer to specific objects in the world being described.
- **Functions**: correspond to concepts that are often expressed in English as genitives such as *AUT's location.*
- **Variables**: are the final FOL mechanism for referring to objects.
- **Predicates**: are symbols that refer to, or name, the relations that hold among some fixed number of objects in a given domain.

First-Order Logic

$$
\begin{array}{rcl}
\textit{Formula} & \rightarrow & \textit{AtomicFormula} \\
& | & \textit{Formula Connective Formula} \\
& | & \textit{Quantifier Variable}, \ldots \textit{ Formula} \\
& | & \neg \textit{ Formula} \\
& | & (\textit{Formula}) \\
\textit{AtomicFormula} & \rightarrow & \textit{Predicate}(\textit{Term}, \ldots) \\
\textit{Term} & \rightarrow & \textit{Function}(\textit{Term}, \ldots) \\
& | & \textit{Constant} \\
& | & \textit{Variable} \\
\textit{Connective} & \rightarrow & \wedge \mid \vee \mid \implies \\
\textit{Quantifier} & \rightarrow & \forall \mid \exists \\
\textit{Constant} & \rightarrow & A \mid \textit{VegetarianFood} \mid \textit{Maharani} \cdots \\
\textit{Variable} & \rightarrow & x \mid y \mid \cdots \\
\textit{Predicate} & \rightarrow & \textit{Serves} \mid \textit{Near} \mid \cdots \\
\textit{Function} & \rightarrow & \textit{LocationOf} \mid \textit{CuisineOf} \mid \cdots
\end{array}
$$

First-Order Logic

- The existential quantifier:
  - There exists : ∃
  - For all: ∀
- Example:
  - 1: a restaurant that serves Mexican food near ICSI.

$$\exists x\ Restaurant(x) \wedge Serves(x, MexicanFood)$$
$$\wedge\ Near((LocationOf(x), LocationOf(ICSI))$$

  - 2: All vegetarian restaurants serve vegetarian food.

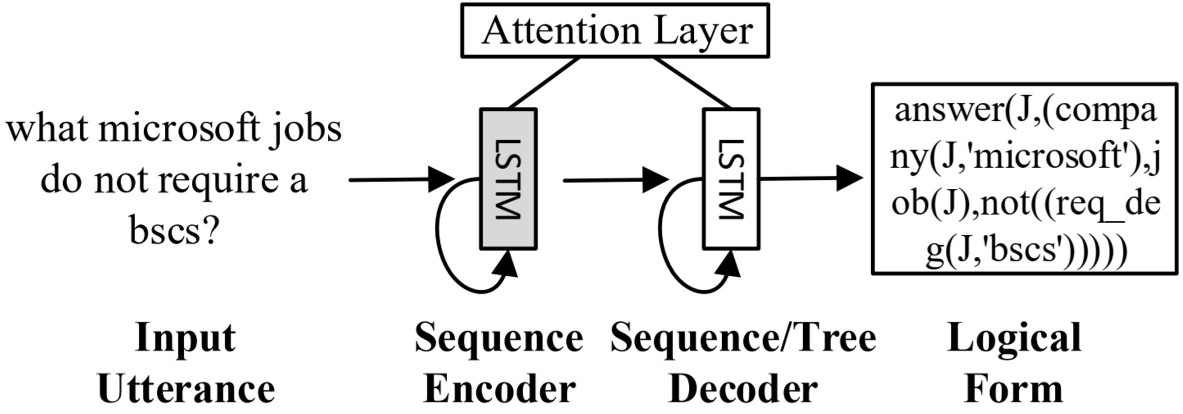
$$\forall x\ VegetarianRestaurant(x) \implies Serves(x, VegetarianFood)$$

Seq2seq for NL to MR

- Viewing MR as a word sequence, semantic parsing is just conditional language modeling (machine translation)

$$P\left(y_1, \dots, y_{|y|} \middle| x_1, \dots, x_{|x|}\right)$$

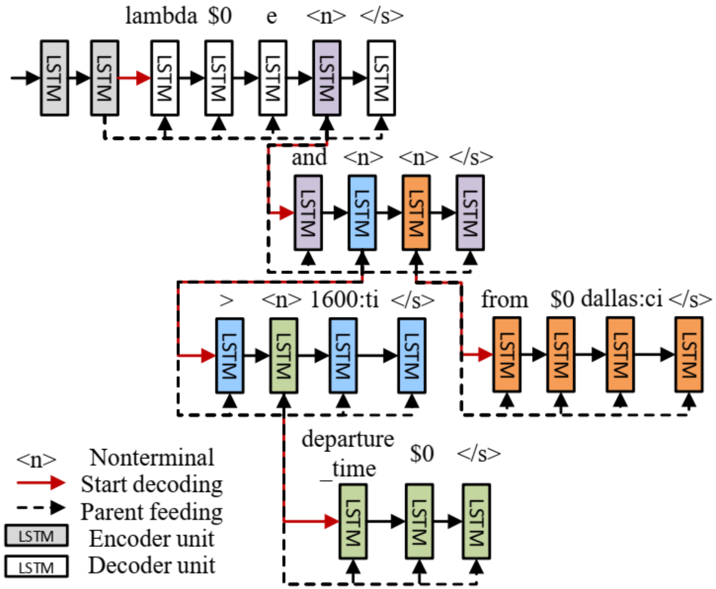
- So, it can be modeled using standard seq2seq models:



Source: Language to Logical Form with Neural Attention

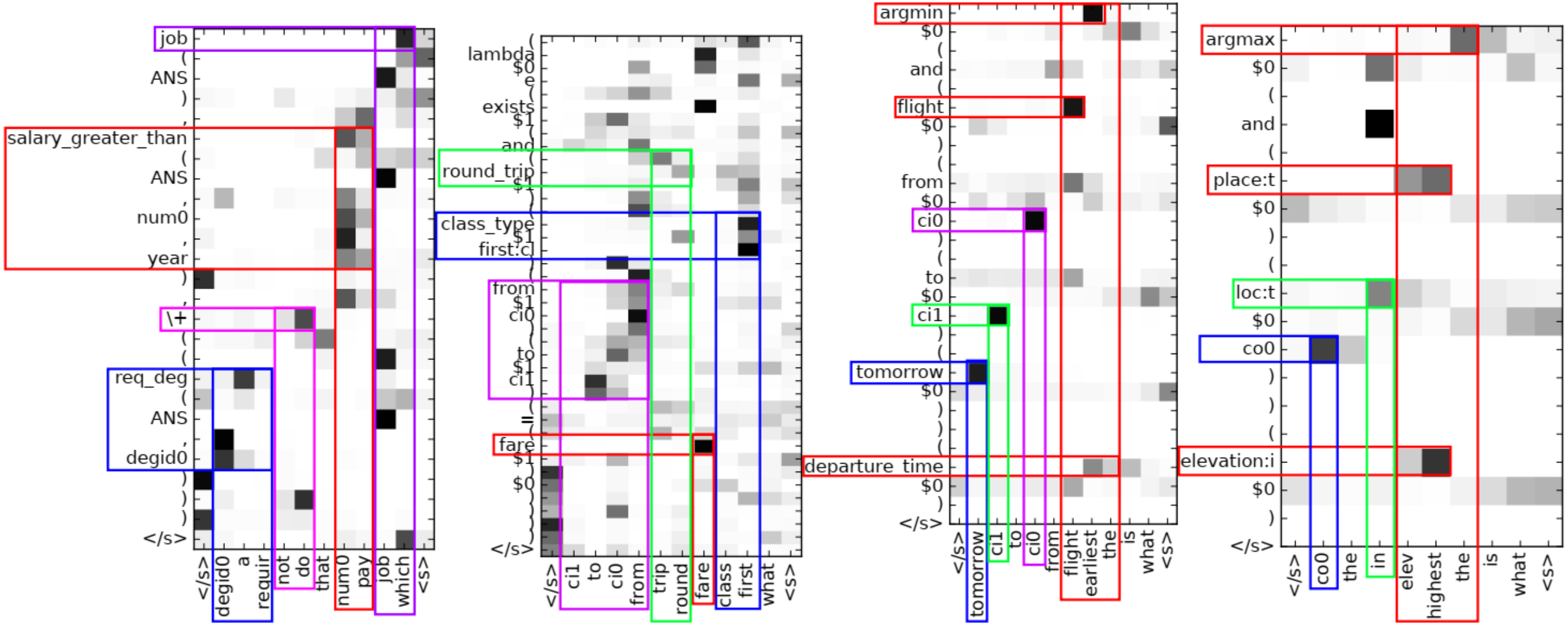
Seq2tree for NL to MR

- The seq2seq model has a potential drawback in that it ignores the hierarchical structure of logical forms.
- The decoder is fundamentally different from the seq2seq model and generates logical forms in a top-down manner.



Source: Language to Logical Form with Neural Attention

Seq2tree for NL to MR



(a) which jobs pay num0 that do not require a degid0

(b) what's first class fare round trip from ci0 to ci1

(c) what is the earliest flight from ci0 to ci1 tomorrow

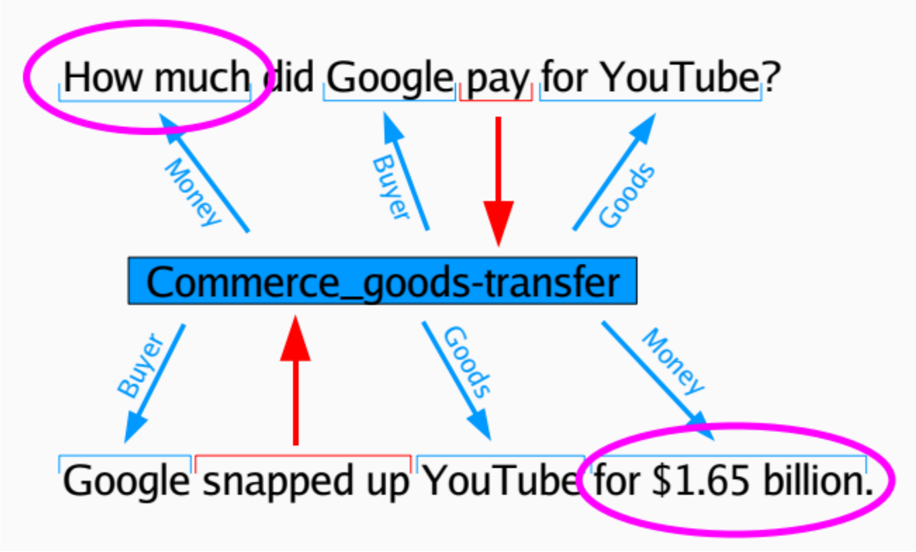
(d) what is the highest elevation in the co0

Source: Language to Logical Form with Neural Attention

Semantic Role Labeling

Semantic Role Labeling

- Semantic role labeling aims to model the predicate-argument structure of a sentence and is often described as answering "Who did what to whom".



Semantic Roles

Thematic Role	Definition
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The beneficiary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

Semantic Roles

Thematic Role	Example
AGENT	*The waiter* spilled the soup.
EXPERIENCER	*John* has a headache.
FORCE	*The wind* blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke *the ice*...
RESULT	The city built a *regulation-size baseball diamond*...
CONTENT	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	He poached catfish, stunning them *with a shocking device*...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	I flew in *from Boston*.
GOAL	I drove *to Portland*.

Semantic Roles

- John broke the window.
  AGENT THEME
- John broke the window with a rock.
  AGENT THEME INSTRUMENT
- The rock broke the window.
  INSTRUMENT THEME
- The window broke.
  THEME
- The window was broken by John.
  THEME AGENT
- The set of thematic role arguments taken by a verb is often called the **thematic grid**, θ-grid, or **case frame**.

Proposition Bank Corpus

- PropBank is a version of the Penn Treebank annotated with semantic roles.

Propbank	Frames
Arg0	proto-agent
Arg1	proto-patient
Arg2	benefactive, instrument, attribute, end state
Arg3	start point, benefactive, instrument, or attribute
Arg4	end point
ArgM	modifier (TMP, LOC, DIR, MNR, etc.)

- Arg2-Arg4 are often verb specific

PropBank Corpus Example

1. **increase.01** "go up incrementally"
   Arg0: causer of increase
   Arg1: thing increasing
   Arg2: amount increased by, EXT, or MNR
   Arg3: start point
   Arg4: end point
2. [$_{\text{Arg0}}$ Big Fruit Co.] increased [$_{\text{Arg1}}$ the price of bananas].
3. [$_{\text{Arg1}}$ The price of bananas] was increased again [$_{\text{Arg0}}$ by Big Fruit Co.]
4. [$_{\text{Arg1}}$ The price of bananas] increased [$_{\text{Arg2}}$ 5%].

BIO Labeling (Decoding)

- BIO id used for sequence labeling for a span-recognition problem.
- BIO labeling:
  - Label any token that begins a span of interest with the label **B**.
  - Tokens that occur inside a span are tagged with an **I**.
  - Any tokens outside of any span of interest are labeled **O**.

Housing	starts	are	expected	to	quicken	a	bit	from	August's	pace
B-ARG1	I-ARG1	O	O	O	V	B-ARG2	I-ARG2	B-ARG3	I-ARG3	I-ARG3

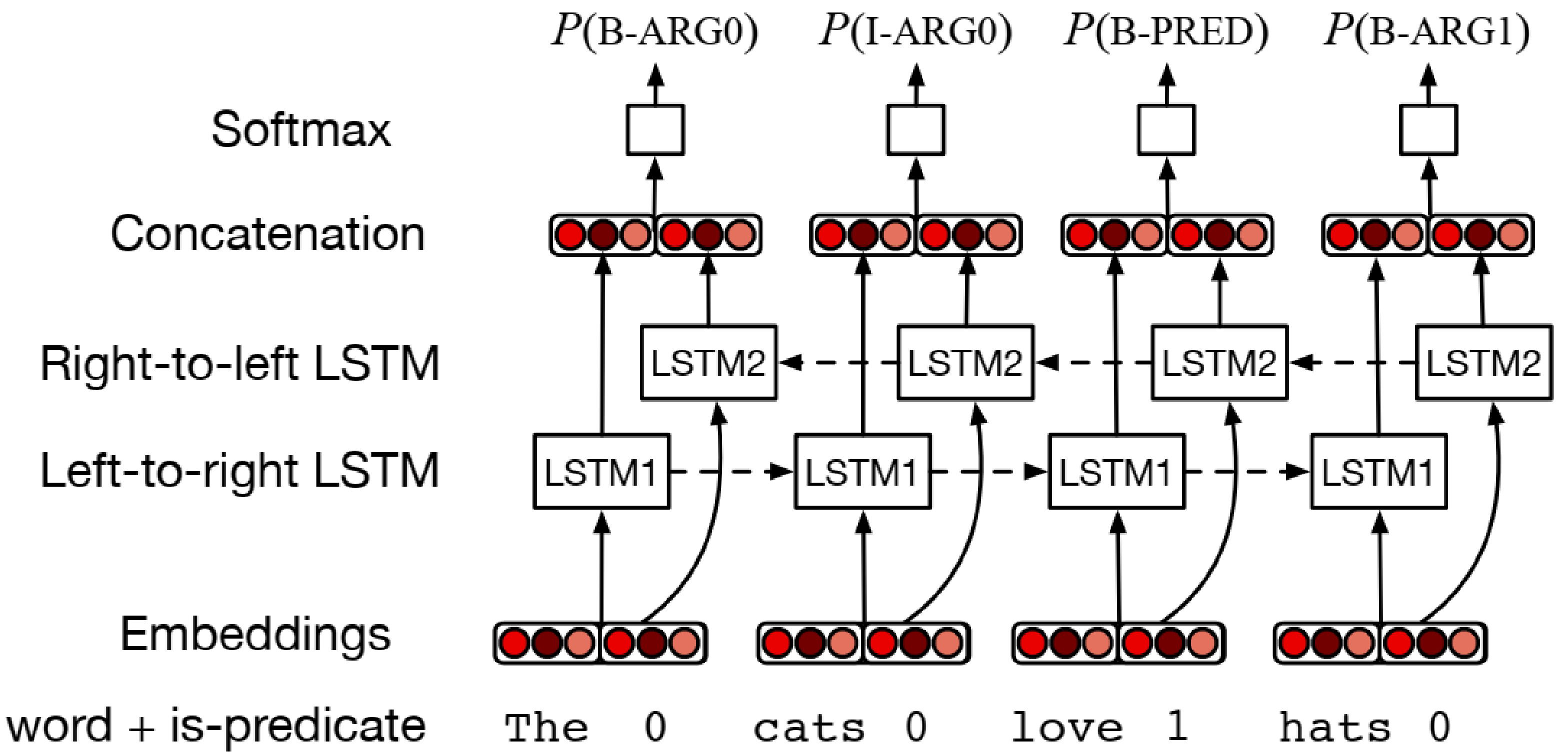
DNN Based SRL Method

- The standard neural algorithm for semantic role labeling is based on the bi-LSTM BIO tagger

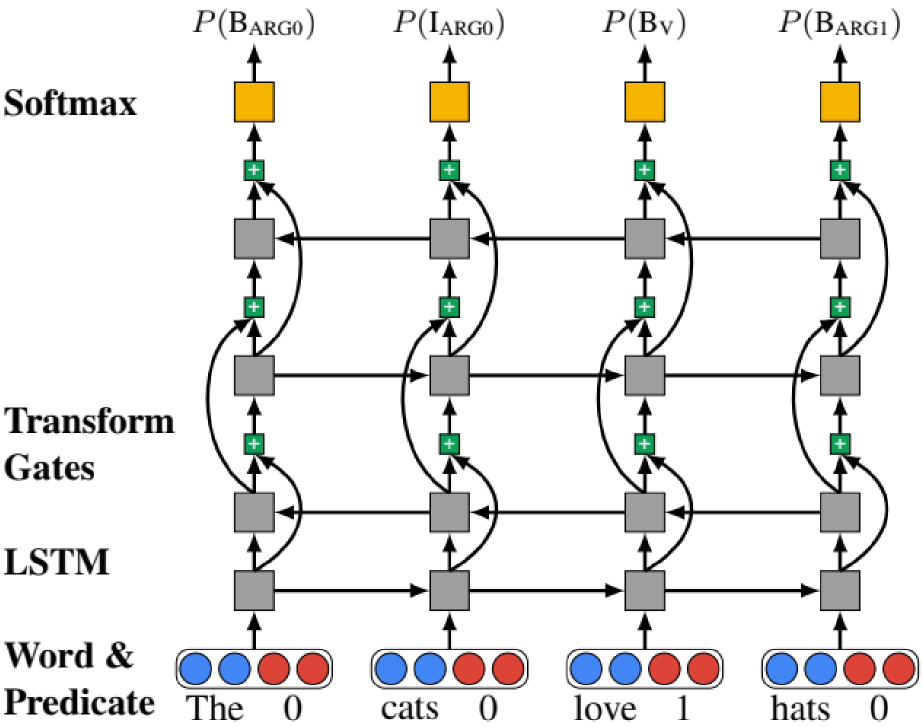
$$\hat{y} = \underset{y \in T}{\operatorname{argmax}}\, p(\boldsymbol{y}|\boldsymbol{w})$$

- As a network input:
  - Pre-trained embeddings of words, plus
  - An embedding for a flag (0/1) variable indicating whether that input word is the predicate.
- Standard bi-directional LSTM based network can be used
  - Highway layers can be used to connect these layers as well.
- The network output can be passed to simple softmax, or an additional CRF+Viterbi decoding, or Viterbi decoding on top of softmax output.

DNN Based SRL Method

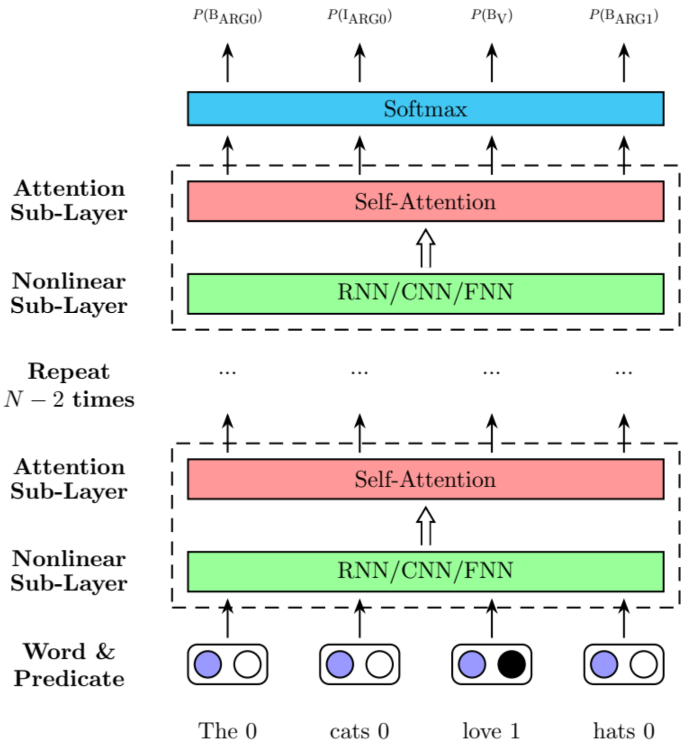


DNN Based SRL Method 2



Source: Deep Semantic Role Labeling: What Works and What's Next

Transformer for SRL



Source: Deep Semantic Role Labeling with Self-Attention

Intent Detection and Slot Filling

Semantic Frame

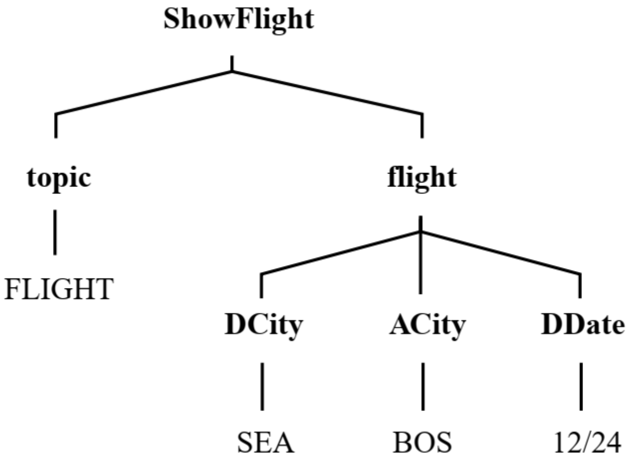
- The semantic structure of an application domain is defined in terms of the ***semantic frames***.
- **ATIS**: Air Travel Information System
    - Is an important milestone for the frame-based SLU
- Each frame contains several typed components called "***slots***".
    - The type of a slot specifies what kind of fillers it is expecting.

```
<frame name="ShowFlight" type="Void">
    <slot name="topic" type="Topic">
    <slot name="flight" type="Flight">
</frame>
<frame name="GroundTrans" type="Void">
    <slot name="city" type="City">
    <slot name="type" type="TransType">
</frame>
<frame name="Flight" type="Flight">
    <slot name="DCity" type="City">
    <slot name="ACity" type="City">
    <slot name="DDate" type="Date">
</frame>
```

Semantic Frame

- The meaning of an input sentence is an instantiation of the semantic frames.
- "Show me flights from Seattle to Boston on Christmas Eve."

```
<ShowFlight>
    <topic type="Freeform">FLIGHT</topic>
    <flight frame="Flight" type="Flight">
        <DCity type="City">SEA</DCity>
        <ACity type="City">BOS</ACity>
        <DDate Type="Date">12/24</DDate>
    </flight>
</ShowFlight>
```



- Attribute-value representation:

[**topic**: FLIGHT] [**DCity**: SEA] [**ACity**: BOS][**DDate**: 12/24]

Intents and Arguments

directions to SF by train

```
(TravelQuery
  (Destination /m/0d6lp)
  (Mode TRANSIT))
```

angelina jolie net worth

```
(FactoidQuery
  (Entity /m/0f4vbz)
  (Attribute /person/net_worth))
```

weather friday austin tx

```
(WeatherQuery
  (Location /m/0vzm)
  (Date 2019-05-10))
```

text my wife on my way

```
(SendMessage
  (Recipient 0x31cbf492)
  (MessageType SMS)
  (Subject "on my way"))
```

play sunny by boney m

```
(PlayMedia
  (MediaType MUSIC)
  (SongTitle "sunny")
  (MusicArtist /m/017mh))
```

is REI open on sunday

```
(LocalQuery
  (QueryType OPENING_HOURS)
  (Location /m/02nx4d)
  (Date 2019-05-12))
```

SLU Technical Challenges

- **Extra-grammaticality** – spoken languages are not as well-formed as written languages.
- **Disfluencies** – false starts, repairs, and hesitations are pervasive, especially in conversational speech.
- **Speech recognition errors** – Speech recognition technology is far from perfect. Environment noises, speaker's accent, domain specific terminologies, all make speech recognition errors inevitable.
- **Out-of-domain utterances** – a dialog system can never restrict a user from saying anything out of a specific domain, even in a system-initiated dialog, where users are prompted for answers to specific questions.

Traditional Methods for SLU

- Knowledge-based Solutions:
  - Has the advantage of no or less dependent on labeled data
  - Semantically Enhanced Syntactic Grammars
    - Replacing the low-level syntactic non-terminals with semantic non-terminals in grammars
  - Semantic Grammars
    - Directly models the domain-dependent semantics with a semantic grammar.
    - Usually, the used grammar here is very complicated with too many grammar rules.
- Problems with knowledge-based solutions:
  - Grammar development is an error-prone process.
  - It takes multiple rounds to fine tune a grammar.
  - Grammar authoring is difficult to scale up.

Traditional Methods for SLU

- Data-driven Approaches:
  - Can automatically learn from example sentences with their corresponding semantics annotated
  - Generative Models:

$$\hat{M} = \underset{M}{\operatorname{argmax}}\, P(M|W) = \underset{M}{\operatorname{argmax}}\, P(W|M)P(M)$$

  - Conditional Models:
    - Conditional Random Fields (CRFs) or Hidden State Conditional Random Fields (HCRFs) are commonly used

Intent Detection

- Intent Classification / Spoken Utterance Classification (SUC)
  - Usages:
    - Call routing
    - Call type classification
  - Classifying the user intent into pre-defined types of intents.
- Interactive Voice Response (IVR) systems

HMIHY:	How may I help you?
User:	Hi, I have a question about my bill (*Billing*)
HMIHY:	OK, what is your question?
User:	May I talk to a human please? (*CSR*)
HMIHY:	In order to route your call to the most appropriate department can you tell me the specific reason you are calling about?
User:	There is an international call I could not recognize (*Unrecognized_Number*)
HMIHY:	OK, I am forwarding you to the human agent. Please stay on the line.

Intent Classification vs. Frame Filling

- Intent classification does not care much about the arguments provided by the user
  - The arguments can be used only in the sense that they help make the right classification.
- While, detecting the arguments is the most important task in a semantic frame (slot) filling
- These two concepts are complementary:
  - The users' requests are first categorized into their goals and then the corresponding templates (frames) can be filled.
  - For semantic frame filling, usually a finer grain categorization is needed
    - I am calling to check whether you received my payment - Verify(Payment)
    - I am calling to make a payment - Make(Payment)

Intent Classification vs. Frame Filling

- Intent classification does not care much about the arguments provided by the user
  - The arguments can be used only in the sense that they help make the
- W
  se
- Th
  - led

Query: What flights are available from Pittsburgh to Baltimore on Thursday morning

Intent: flight info

Slots:

- from_city: pittsburgh
- to_city: baltimore
- depart_date: thursday
- depart_time: morning

  - I am calling to check whether you received my payment - Verify(Payment)
  - I am calling to make a payment - Make(Payment)

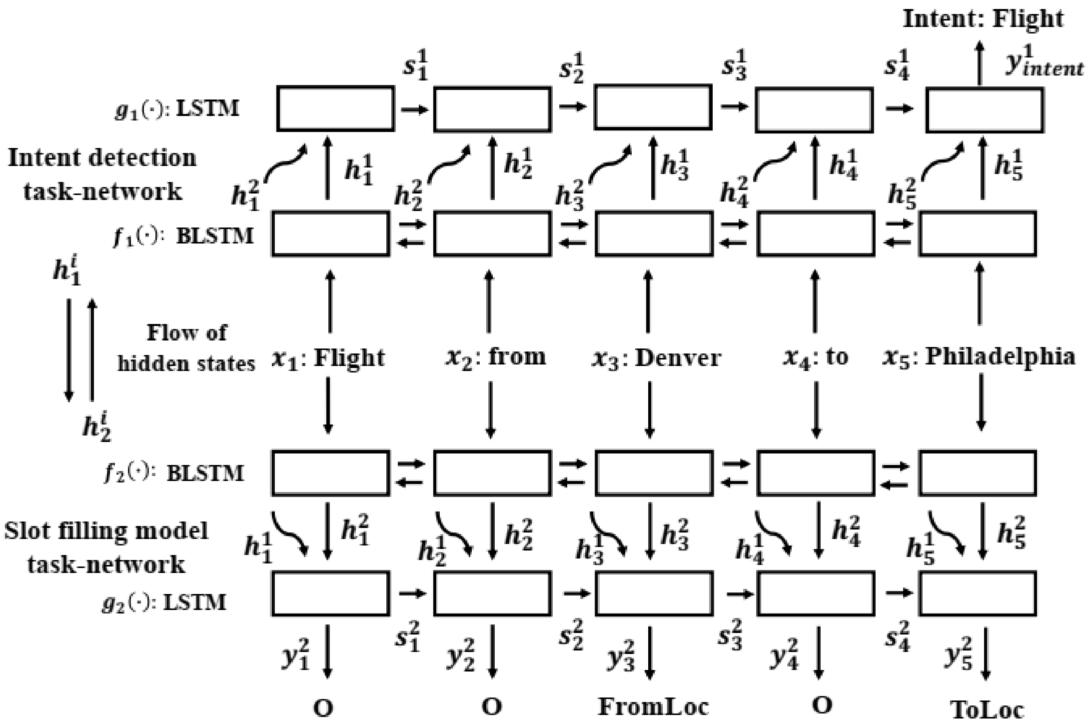
Intent Detection

- This task aims at classifying a given user commands/queries (speech utterance) X_r into one of M semantic classes (intents), $\hat{C}_r \in C = \{C_1, \dots, C_M\}$

$$\hat{C}_r = \underset{C_r}{\operatorname{argmax}}\, P(C_r | X_r)$$

- Semantic classifiers require operation with significant freedom in utterance variations.
- It should be able to generalize well from a small amount of training data.

Bi-model Based RNN Method



Source: A Bi-model based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling

Bi-model Based RNN Method

- Intent detection task:

$$s_t^1 = \phi(s_{t-1}^1, h_{n-1}^1, h_{n-1}^2)$$
$$y_{intent}^1 = \arg\max_{\hat{y}_n^1} P(\hat{y}_n^1 | s_{n-1}^1, h_{n-1}^1, h_{n-1}^2)$$

- Slot filling task:

$$s_t^2 = \psi(h_{t-1}^2, h_{t-1}^1, s_{t-1}^2, y_{t-1}^2)$$
$$y_t^2 = \arg\max_{\hat{y}_t^2} P(\hat{y}_t^2 | h_{t-1}^1, h_{t-1}^2, s_{t-1}^2, y_{t-1}^2)$$

Source: A Bi-model based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling

Bi-model Based RNN Method

- Asynchronous training:
    - Train two task-networks based on their own cost functions in an asynchronous manner.
- Both losses are defined using cross entropy as:

$$\mathcal{L}_1 \triangleq -\sum_{i=1}^{k} \hat{y}_{intent}^{1,i} \log(y_{intent}^{1,i})$$

$$\mathcal{L}_2 \triangleq -\sum_{j=1}^{n} \sum_{i=1}^{m} \hat{y}_j^{2,i} \log(y_j^{2,i})$$

Source: A Bi-model based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling

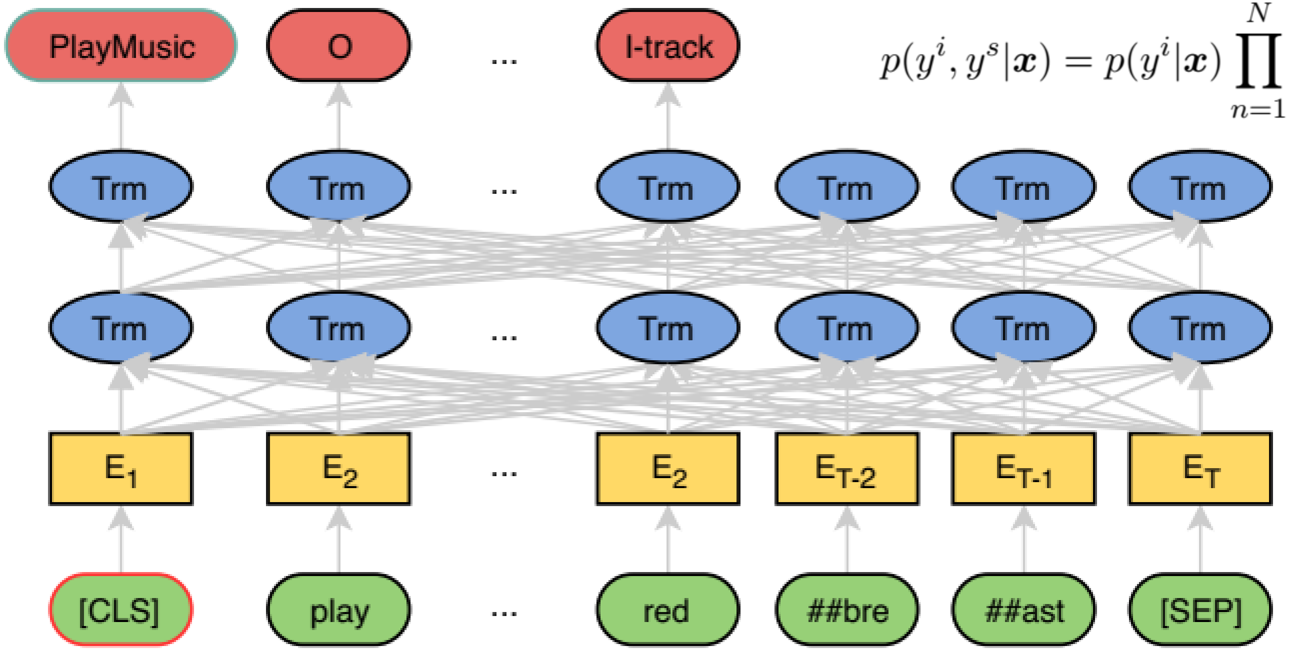
Bi-model Based RNN Method

- ATIS results:

Model	F1 Score	Intent Accuracy
Recursive NN (Guo et al., 2014)	93.96%	95.4%
Joint model with recurrent intent and slot label context (Liu and Lane, 2016b)	94.47%	98.43%
Joint model with recurrent slot label context (Liu and Lane, 2016b)	94.64%	98.21%
RNN with Label Sampling (Liu and Lane, 2015)	94.89%	NA
Hybrid RNN (Mesnil et al., 2015)	95.06%	NA
RNN-EM (Peng and Yao, 2015)	95.25%	NA
CNN CRF (Xu and Sarikaya, 2013)	95.35%	NA
Encoder-labeler Deep LSTM (Kurata et al., 2016)	95.66%	NA
Joint GRU Model (W) (Zhang and Wang, 2016)	95.49%	98.10%
Attention Encoder-Decoder NN (Liu and Lane, 2016a)	95.87%	98.43%
Attention BiRNN (Liu and Lane, 2016a)	95.98%	98.21%
Bi-model without a decoder	**96.65%**	**98.76%**
Bi-model with a decoder	**96.89%**	**98.99%**

Source: A Bi-model based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling

BERT Method



$$p(y^i, y^s|\boldsymbol{x}) = p(y^i|\boldsymbol{x}) \prod_{n=1}^{N} p(y_n^s|\boldsymbol{x})$$

Source: BERT for Joint Intent Classification and Slot Filling

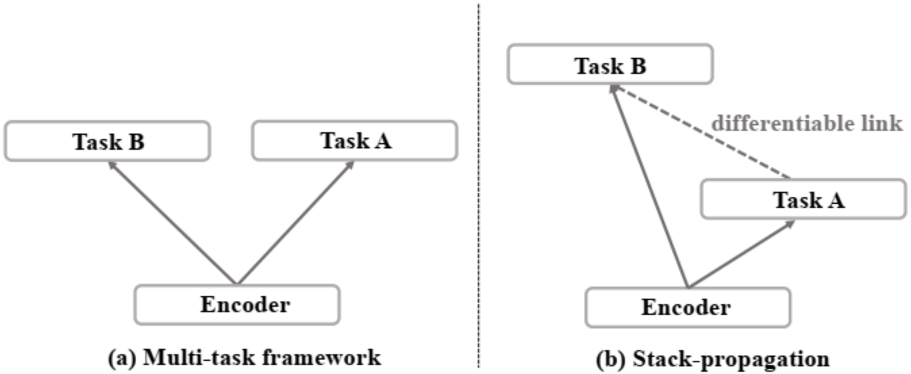
BERT Method

Models	Snips			ATIS		
	Intent	Slot	Sent	Intent	Slot	Sent
RNN-LSTM (Hakkani-Tür et al., 2016)	96.9	87.3	73.2	92.6	94.3	80.7
Atten.-BiRNN (Liu and Lane, 2016)	96.7	87.8	74.1	91.1	94.2	78.9
Slot-Gated (Goo et al., 2018)	97.0	88.8	75.5	94.1	95.2	82.6
Joint BERT	**98.6**	**97.0**	**92.8**	97.5	**96.1**	88.2
Joint BERT + CRF	98.4	96.7	92.6	**97.9**	96.0	**88.6**

Source: BERT for Joint Intent Classification and Slot Filling

Stack-Propagation Method

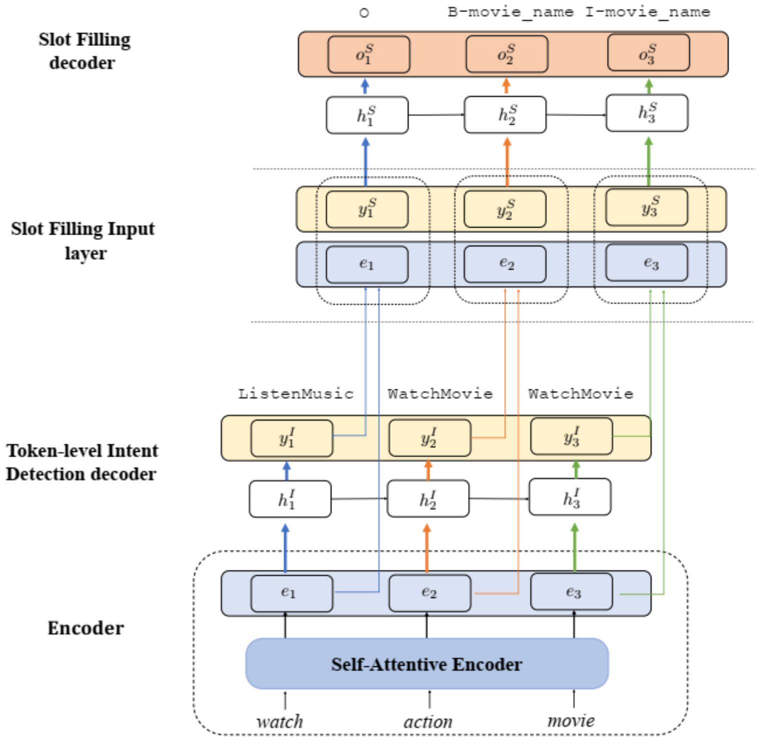
- Multi-task Framework vs Stack-Propagation



- The multitask framework cannot provide features from up-stream task to down-stream task explicitly.

Source: A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding

Stack-Propagation Method



$$\mathbf{y}_i^I = \text{softmax}\left(\mathbf{W}_h^I \mathbf{h}_i^I\right),$$
$$o_i^I = \text{argmax}(\mathbf{y}_i^I),$$

$$o^I = \text{argmax} \sum_{i=1}^{m} \sum_{j=1}^{n_I} \alpha_j \mathbb{1}[o_i^I = j],$$

Source: A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding

Stack-Propagation Method

- Joint Training
  - Convert the sentence level classification task into token-level prediction
- Intent detection objection:

$$\mathcal{L}_1 \triangleq -\sum_{j=1}^{m}\sum_{i=1}^{n_I} \hat{\mathbf{y}}_j^{i,I} \log\left(\mathbf{y}_j^{i,I}\right)$$

- Slot filling task objection:

$$\mathcal{L}_2 \triangleq -\sum_{j=1}^{m}\sum_{i=1}^{n_S} \hat{\mathbf{y}}_j^{i,S} \log\left(\mathbf{y}_j^{i,S}\right)$$

- The final joint objective:

$$\mathcal{L}_\theta = \mathcal{L}_1 + \mathcal{L}_2$$

Source: A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding

Stack-Propagation Method

Model	SNIPS			ATIS		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
Joint Seq (Hakkani-Tür et al., 2016)	87.3	96.9	73.2	94.3	92.6	80.7
Attention BiRNN (Liu and Lane, 2016)	87.8	96.7	74.1	94.2	91.1	78.9
Slot-Gated Full Atten (Goo et al., 2018)	88.8	97.0	75.5	94.8	93.6	82.2
Slot-Gated Intent Atten (Goo et al., 2018)	88.3	96.8	74.6	95.2	94.1	82.6
Self-Attentive Model (Li et al., 2018)	90.0	97.5	81.0	95.1	96.8	82.2
Bi-Model (Wang et al., 2018)	93.5	97.2	83.8	95.5	96.4	85.7
CAPSULE-NLU (Zhang et al., 2019)	91.8	97.3	80.9	95.2	95.0	83.4
SF-ID Network (E et al., 2019)	90.5	97.0	78.4	95.6	96.6	86.0
Our model	**94.2***	**98.0***	**86.9***	**95.9***	**96.9***	**86.5***
Oracle (Intent)	96.1	-	-	96.0	-	-

Source: A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding

Stack-Propagation Method

- Using BERT in the proposed framework:
  - replace the self-attentive encoder by BERT base model with the fine-tuning approach

Model	**SNIPS**			**ATIS**		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
Our model	94.2	98.0	86.9	95.9	96.9	86.5
Intent detection (BERT)	-	97.8	-	-	96.5	-
Slot filling (BERT)	95.8	-	-	95.6	-	-
BERT SLU (Chen et al., 2019)	97.0	98.6	92.8	96.1	97.5	88.2
Our model + BERT	97.0	99.0	92.9	96.1	97.5	88.6

Source: A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding

Thanks for your attention

References and IP Notice

- Some slides were selected from Bill MacCartney's slides and Mirella Lapata's slides
- Some graphics were selected **Slidesgo** **template**