How to Make Chatbots More Intelligent: Computational Semantics Beyond Events and Roles

Lê Hồng Phương

<*phuonglh@hus.edu.vn*>
Data Science Laboratory, Hanoi University of Science
Vietnam National University, Hanoi

May 9, 2018

Content

Data Science Laboratory – DSL

- The Data Science Laboratory is at VNU University of Science, Hanoi.
- Officially created in late December 2017 as a part of the Department of Mathematics, Mechanics and Informatics.
- Head of the DSL: Dr. Le Hong Phuong
- Seven teams:
  1. Natural language processing and text mining
  2. Image processing and computer vision
  3. Mathematical finance
  4. Optimization and operations research
  5. Applied statistics and probability
  6. Cryptography and data security
  7. Dynamical systems and biological mathematics
- http://mim.hus.vnu.edu.vn/dsl/

Two Main Components of a Chatbot System

AI Engine

language understanding & intent detection

Dialog Engine

state machine that executes context-driven workflows with scope variables

- **Machine Learning Component**: Build, test ML models
- **Knowledge Services**: Entity modeling, representing domain knowledge as semantic graphs

An Example Chatbot

Lunch Ordering

Utterance	Intent
M: Hello, can I help you?	greeting
H: Yes, I'd like to have some lunch.	askMenu
M: Would you like a starter?	askStarter
H: Yes, I'd like [a bowl of chicken soup], please.	chooseStarter
M: And what would you like for your main course?	askMainCourse
H: I'd like [a grilled cheese sandwich].	chooseMainCourse
M: Would you like anything to drink?	askDrink
H: Yes, I'd like [a glass of coke], please.	chooseDrink
M: Do you need anything else?	askOther
H: No, thank you.	confirmation

Every chatbot needs **intent detection** and **entity extraction**.

Is Your Bot Intelligent?

- Intent detection and entity extraction are not sufficient to make your chatbots intelligent.
- They cannot answer questions such as:
  1. `What is three plus four?`
  2. `What is the largest prime less than 2018?`
  3. `What states border Texas?`
  4. `What is the tallest mountain in Vietnam?`
  5. `Which restaurant has the most rating?`
  6. `Who is the spouse of Donald Trump?`
  7. `Where was Barack Obama born?`
  8. `What is the cover price of X-men?`
  9. `Where did Walt Disney live before he died?`

Semantic Parsing

A semantic parsing maps natural language utterances into an intermediate logical form, which is executed to produce a denotation.

A simple arithmetic task

- Utterance: `What is three plus four?`
- Logical form: `(+ 3 4)`
- Denotation: `7`

A question answering task

- Utterance: `Where was Donald Trump born?`
- Logical form: `(placeOfBirth ‘‘Donald Trump’’)`
- Denotation: ‘‘`New York City`’’

Semantic Parsing

A travel agent bot

- Utterance: `Show me flights to Hanoi leaving tomorrow`
- Logical form: `(and (type flight) (destination Hanoi) (departureDate 2018.05.10))`
- Denotation: `(list ...)`

Semantic Representations

- Semantic representations are generally **logical forms**, which are expressions in a fully specified, unambiguous artificial language.
- There are a variety of different formalisms:
  - lambda calculi
  - natural logics
  - diagrammatic languages
  - programming languages
  - robot controller languages
  - Grammar formalism based schemes:
    - Dependency Formalism
    - Combinatory Categorial Grammar (CCG)
    - Head-Phrase Structure Grammar (HPSG)
    - Lexicalized Tree Adjoining Grammar (LTAG)
  - database query languages
  - **knowledge-based query languages** (SPARQL, etc.)
  - **lambda dependency-based compositional semantics**

Logical Forms

- A **logical form** is a hierarchical expression. Primitive logical forms represent concrete values:
  - `(boolean true)`
  - `(number 11)`
  - `(string "`*xin chào*`")`
  - `(date 2018 05 09)`
  - `(fb:en.donald_trump)`
  - `(list (number 2) (number 3))`

Logical Forms

- Logical forms can be constructed recursively by a function name followed by arguments, which are themselves logical forms:
  - `(call + (number 3) (number 4))`
  - `(call java.lang.Math.cos (number 0))`
  - `(call .indexOf (string “ai for life”) (string “if”))`
  - `(call .substring (string “ai for life”) (number 3) (number 6))`
  - `(call if (call < (number 3) (number 4)) (string yes) (string no))`
- We can execute a logical form and get the denotation (answer):

```
(execute (call + (number 3) (number 4)))
```

$\Downarrow$

```
(number 7)
```

Parsing Utterances to Logical Forms

- How to map a natural language utterance into a logical form?
- The key framework is **compositionality**:
  - *The meaning of a full sentence is created by combining the meanings of its parts.*
  - Meanings are represented by logical forms.
- Classical and powerful approach: use a **formal grammar**.

Formal Grammar

- A **grammar** is a set of **rules** which specify how to combine logical forms to build more complex ones in a manner *that is guided by the natural language*.
- An example grammar:

```
(rule $Expr ($PHRASE) (NumberFn))

(rule $Operator (plus) (ConstantFn (lambda y (lambda x (call + (var x)
   (var y))))))

(rule $Operator (times) (ConstantFn (lambda y (lambda x (call * (var x)
   (var y))))))

(rule $Partial ($Operator $Expr) (JoinFn forward))

(rule $Expr ($Expr $Partial) (JoinFn backward))

(rule $ROOT ((what optional) (is optional) $Expr (? optional)) (IdentityFn))
```

Parsing Utterances to Logical Forms

- Now, the utterance "`What is three plus four?`" should give the output `(number 7)`
- A longer sentence such as "`What is three plus four times two?` should give two derivations:
  - `(number 14)`
  - `(number 11)`
- A **parser** is an actual algorithm that takes the grammar and generates those derivations.
  - INP: `What is three plus four?`
  - OUT: `(derivation (formula (((lambda y (lambda x (call + (var x) (var y)))) (number 4)) (number 3))) (value (number 7)))`

Parsing Utterances to Logical Forms

- A given utterance might be consistent with multiple logical forms, creating ambiguity.
- INP: `What is three plus four times two?`
- OUT:
  1. `(derivation (formula (((lambda y (lambda x (call + (var x) (var y)))) (((lambda y (lambda x (call * (var x) (var y)))) (number 2)) (number 4))) (number 3))) (value (number 11)))`
  2. `(derivation (formula (((lambda y (lambda x (call * (var x) (var y)))) (number 2)) (((lambda y (lambda x (call + (var x) (var y)))) (number 4)) (number 3)))) (value (number 14)))`

Parsing Utterances to Logical Forms

- **Computational challenge**: the number of candidate logical forms is in general exponential in the length of the sentence.
- For example, in a simple grammar, even with just 9 atomic numeral expressions to consider, the number of logical forms for the short sentence "`two times two plus three`" is $9^3 \times 3^2 \times 2 = 13,122$.
- In the question "`What kind of system of government does the United States have?`" (Berant et al., 2013):
  - the phrase "`United States`" maps to 231 entities in the lexicon,
  - the verb "`have`" maps to 203 binaries,
  - the phrases "`kind`", "`system`", and "`government`" all map to many different unary and binary predicates.

Learning

- Machine learning concerns the *ability to generalize* from a set of past observations or experiences in a way that leads to improved performance on a future task (T. Mitchel, 1997).
- In **supervised learning**, the experiences take the form of input/output pairs $(\mathbf{x}, \mathbf{y})$.
- A ML system has 3 integral pieces:
  1. a **feature representation** of the data
  2. an **objective function**
  3. an algorithm for **optimizing** the objective function

Learning – Feature Representation

- The most important part of ML is determining a suitable representation of the data.
  - The key idea in *linear* classification is to map each $(\mathbf{x}, \mathbf{y})$ pair to a d-dimensional feature vector $\phi(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d$.
- The objective function and associated scoring function define the learning goal.

$$\begin{aligned}\mathrm{Score}(\mathbf{x}, \mathbf{y}; \mathbf{w}) &= \mathbf{w} \cdot \phi(\mathbf{x}, \mathbf{y}) \\ &= \sum_{j=1}^{d} w_j \phi(\mathbf{x}, \mathbf{y})_j\end{aligned}$$

- We associate each feature j with a real-valued weight w_j. For a new input $\mathbf{x}$, we simply predict the output with the highest score:

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} \mathrm{Score}(\mathbf{x}, \mathbf{y}; \mathbf{w})$$

Learning – Objective Function

- The goal of training is to optimize the weight vector $\mathbf{w}$ so that the score of the correct output $\mathbf{y}$ is generally larger than the score of the incorrect output $\bar{\mathbf{y}}$ on a training set.
- There are many objective functions, a common one is *the multiclass hinge loss objective*:

$$\min_{\mathbf{w} \in \mathbb{R}^d} \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{D}} \max_{\bar{\mathbf{y}}} [\mathrm{Score}(\mathbf{x}, \bar{\mathbf{y}}; \mathbf{w}) + c(\mathbf{y}, \bar{\mathbf{y}})] - \mathrm{Score}(\mathbf{x}, \mathbf{y}; \mathbf{w}),$$

where $c(\mathbf{y}, \bar{\mathbf{y}})$ is the prediction cost of an example, usually

$$c(\mathbf{y}, \bar{\mathbf{y}}) = \begin{cases} 0, & \text{if } \mathbf{y} = \bar{\mathbf{y}} \\ 1, & \text{if } \mathbf{y} \neq \bar{\mathbf{y}} \end{cases}$$

Learning – Optimization

- There are many ways of solving the optimization problem.
- One simple method is **stochastic gradient descent** (SGD) which iteratively computes the gradient of the objective function.

The SGD Optimization Algorithm

Data: $\mathcal{D}$ a set of training sample
Result: Parameter vector $\mathbf{w}$
initialize $\mathbf{w} \leftarrow \vec{0}$;
while *(not converge)* **do**
  for $(\mathbf{x}, \mathbf{y}) \in \mathcal{D}$ **do**
    $\tilde{\mathbf{y}} \leftarrow \arg\max_{y'} \text{Score}(\mathbf{x}, \mathbf{y}'; \mathbf{w}) + c(\mathbf{y}, \mathbf{y}')$;
    $\mathbf{w} \leftarrow \mathbf{w} - \alpha \left[\phi(\mathbf{x}, \tilde{\mathbf{y}}) - \phi(\mathbf{x}, \mathbf{y})\right]$;
  end
end
return $\mathbf{w}$;

Datasets

1. The FREE917 dataset: 917 questions involving 635 Freebase relations, annotated with lambda calculus forms.
2. The OVERNIGHT dataset: 13,682 examples of languages utterances paired with logical forms across eight domains (*basketball*, *blocks*, *calendar*, *housing*, *publications*, *recipes*, *restaurants*, *social network*).
3. WEBQUESTIONS dataset: 5,810 questions with answers available in Freebase (possible entities, values or list of entities).
4. VITK dataset: 879 question-answer pairs, annotated with SPARQL queries against DBPedia.

Datasets

```
<qa type="NUM">
  <question>Số nhân viên của FPT là bao nhiêu?</question>
  <cypher>
    START x=node:DBPediaIndex(key="FPT") RETURN DISTINCT x.sốNhânViên, x.chiTiết LIMIT 20
  </cypher>
  <correctCypher>true</correctCypher>
  <answer>
    FPT, tên viết tắt bằng tiếng Anh của Công ty cổ phần FPT (tên cũ của Công ty là
    Công ty cổ phần phát triển đầu tư công nghệ FPT), là một tập đoàn kinh tế tại Việt Nam
    với lĩnh vực kinh doanh chính là cung cấp các dịch vụ liên quan công nghệ thông tin.
  </answer>
  <correctAnswer>true</correctAnswer>
  <speed>134</speed>
</qa>
<qa type="HUM">
  <question>Ai là vợ của Tôn Đức Thắng?</question>
  <cypher>
    START x=node:DBPediaIndex(key="Tôn Đức Thắng") MATCH x-[r:vợChồng]-y RETURN DISTINCT
      y.chiTiết LIMIT 20 START x=node:DBPediaIndex(key="Tôn Đức Thắng") RETURN DISTINCT
      x.vợChồng, x.chiTiết LIMIT 20
  </cypher>
  <correctCypher>true</correctCypher>
  <answer>Đoàn Thị Giàu</answer>
  <correctAnswer>true</correctAnswer>
  <speed>153</speed>
</qa>
```

Accuracy

Dataset	Accuracy
Free917	71.30%
Overnight	79.60%
WebQuestions	32.90%
Vitk	76.90%

- The best accuracy scores of published semantic parsing systems are promising.
- Semantic parsing is a very difficult task. We are still at day 1.

Summary

- Many existing chatbot systems have mostly used a swallow semantic representation of text, disregarding significant meaning encoded in human language.
- A deep understanding and reasoning of natural language is required in order for a chatbot to be more intelligent (better understand an input utterance and produce an appropriate action).
- Recent logical and statistical approaches have identified methods for mapping utterances to meanings efficiently.

Take Home Messages

Message #1

Chatbots can be made more intelligent only by computer scientists with a deep knowledge of computational semantics.

Message #2

Chatbot startups will be likely to fail if they set goals that are too ambitious for current advances of computational linguistics.

References

1. Jonathan Herzig and Jonathan Berant, "Neural Semantic Parsing over Multiple Knowledge-bases", *Proceedings of ACL*, 2017.
2. Jonathan Berant et al., "Semantic Parsing on Freebase from Question-Answer Pairs", *Proceedings of EMNLP*, 2013.
3. Percy Liang and Christopher Potts, "Bringing Machine Learning and Compositional Semantics Together", *Annual Reviews of Linguistics*, 2015.
4. Phuong Le-Hong and Duc-Thien Bui, "A Factoid Question Answering System for Vietnamese", *Proceedings of WWW Companion*, 2018.
5. SEMPRE 2.0 semantic parsing system[1]

[1] *https://github.com/percyliang/sempre*