2014/02/13 Sphinx Lunch

Best Student Paper Award

@ 2013 IEEE Workshop on Automatic Speech Recognition and Understanding – Dec. 9-12, 2013

Unsupervised Induction and Filling of Semantic Slot for Spoken Dialogue Systems Using Frame-Semantic Parsing

Yun-Nung (Vivian) Chen, William Yang Wang, and Alexander I. Rudnicky

Language Technologies Institute
School of Computer Science
Carnegie Mellon University

Overview

Introduction

Approach

Experiments

Conclusion

Overview

Introduction

Approach

Experiments

Conclusion

In 2011…

- “… *I have long been looking for good knowledgeable student to take on the task of getting conversational agents to* ***move beyond prerecorded interaction*** *(or clunky spoken dialog systems)* …”

- “… *Wow! Sounds like they know how to achieve the ultimate machine intelligence at CMU*…”

Spoken Language Understanding (SLU)

- SLU in dialogue systems
  - **SLU maps NL inputs to semantic forms**

    "I would like to go to $\underbrace{\text{Shadyside}}_{\text{location: Shadyside}}$ $\underbrace{\text{Tuesday}}_{\text{date: Tuesday}}$."

  - **Semantic frames, slots, & values**
    - often manually defined by domain experts or developers.

What are the problems?

Problems with Predefined Slots

- **Generalization:** may not generalize to real-world users.
- **Bias propagation:** can bias subsequent data collection and annotation.
- **Maintenance:** when new data comes in, developers need to start a new round of annotation to analyze the data and update the grammar.
- **Efficiency:** time consuming, and high costs.

Can we automatically induce semantic slots only using raw audios?

Overview

Introduction | Approach | Experiments | Conclusion

Probabilistic Frame-Semantic Parsing

(Das et al., 2010; 2013) on ASR-transcribed utterances.

First Step

FrameNet

- FrameNet (Baker et al., 1998)
  - a linguistically-principled semantic resource, based on the frame-semantics theory.
- Example
  - **Frame** (*food*): contains words referring to items of food.
  - **Frame Element**: a descriptor indicates the characteristic of food.
    - "low fat milk" → "milk" evokes the "food" frame; "low fat" fills the descriptor FE.

Frame-Semantic Parsing

- SEMAFOR (Das et al., 2010; 2013)
  - a state-of-the-art frame-semantics parser, trained on manually annotated FrameNet sentences.

The Panacea?

- Unfortunately…

Good!

Good!

Bad!

can i have a cheap restaurant

Frame: expensiveness
FT LU: cheap

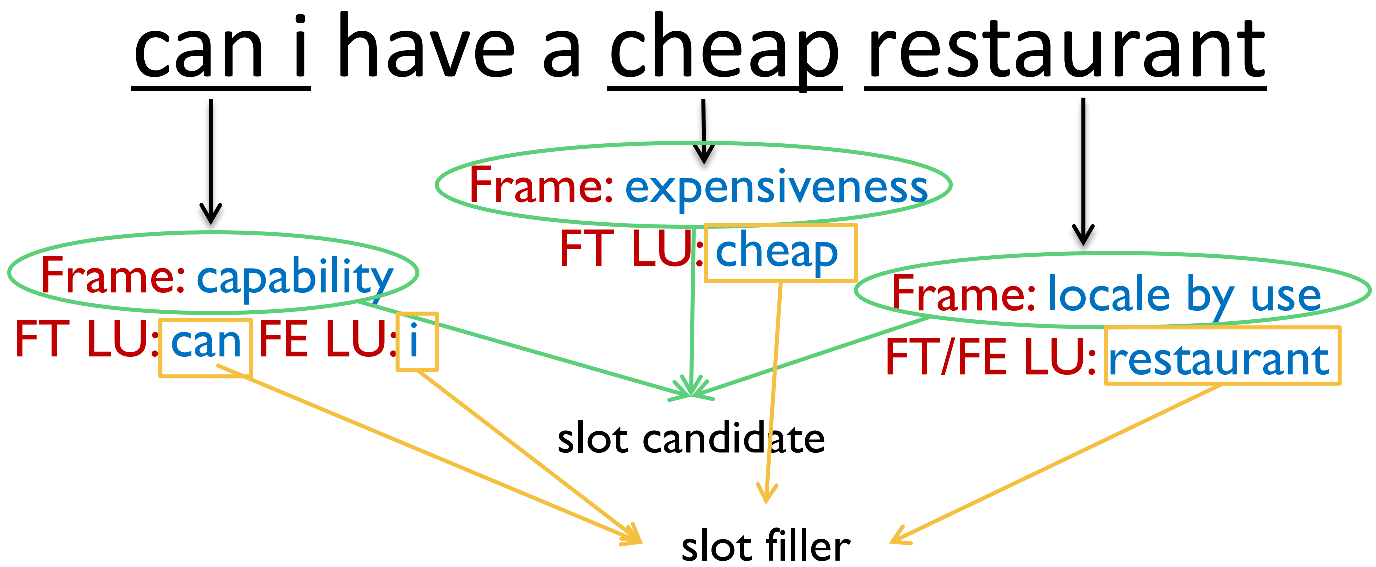
Frame: capability
FT LU: can FE LU: i

Frame: locale by use
FT/FE LU: restaurant

Task: adapting ***generic*** frames to ***task-specific*** settings in dialogue systems.

As A Ranking Problem

- Main idea
  - Ranking domain-specific concepts higher than generic semantic concepts



Slot Ranking Model (1/3)

- Rank the slot candidates by integrating two scores

$$w(s_i) = \log f(s_i) + \alpha \cdot \log h(s_i)$$

the frequency of the slot candidate in the SEMAFOR-parsed corpus

slots with higher frequency may be more important

the coherence of slot fillers

domain-specific concepts should focus on fewer topics and be similar to each other

slot: quantity

one
a
three all

lower coherence in topic space

slot: expensiveness

cheap inexpensive
expensive

higher coherence in topic space

Slot Ranking Model (2/3)

- Measure coherence by word-level context clustering
  - For each slot s_i,

$$V(\boxed{s_i}) = \{\mathbf{v_1}, \ldots, \boxed{\mathbf{v_j}}, \ldots, \mathbf{v_J}\}$$

slot candidate: expensiveness

corresponding value vectors: "cheap", "not expensive" (from the utterances with s_i in the parsing results)

  - We have corresponding cluster vector

$$C(s_i) = \{\mathbf{c_1}, \ldots, \mathbf{c_j}, \ldots, \mathbf{c_J}\} \quad \mathbf{c_j} = [c_{j1}, \ldots, \boxed{c_{jk}}, \ldots c_{jK}]$$

the frequency of words in v_j clustered into cluster k

  - Measure coherence measure by pair-wised cosine similarity

$$h(s_i) = \frac{\sum_{\mathbf{c_a}, \mathbf{c_b} \in C(s_i), \mathbf{c_a} \neq \mathbf{c_b}} \mathrm{CosSim}(\mathbf{c_a}, \mathbf{c_b})}{|C(s_i)|^2}$$

The slot with higher $h(s_i)$ usually focuses on fewer topics, more specific, and preferable for slots of SDS.

Slot Ranking Model (3/3)

- Spectral clustering
  - For each word $w = [r_1, \ldots, r_i, \ldots]$

    r_i = 1 when w occurs in the i-th utterance
    r_i = 0 otherwise

    Assume that the words in the same utterance are related to each other

    - The approach can be summarized in five steps:
      1. Calculate the distance matrix
      2. Derive the affinity matrix
      3. Generate the graph Laplacian
      4. Eigen decomposition of L
      5. Perform K-means clustering of eigenvectors

Reasons why spectral clustering:
1) can be solved efficiently by standard linear algebra
2) invariant to the shapes and densities of each cluster
3) projects the manifolds within data into solvable space and often outperform others

Overview

Introduction | Approach | Experiments | Conclusion

Experiments

- Dataset
  - Cambridge University SLU corpus (Henderson, 2012)
    - restaurant recommendation in an in-car setting in Cambridge
      - WER = 37%
      - vocabulary size = 1868
      - 2,166 dialogues
      - 15,453 utterances
      - dialogue slot (total # = 10): addr, area, food, name, phone, postcode, price range, signature, task, type

Slot Induction Evaluation

- MAP of the slot ranking model
  - measure the quality of induced slots based on induced and reference slots via the mapping table

Approach	MAP	
	ASR	Manual
Frequency	67.31	59.41
K-Means	68.45	59.76
Spectral Clustering	**69.36**	**61.86**

The majority of the reference slots used in a real world SDS can be induced automatically in an unsupervised way.

The mapping table between induced and reference slots

Induced Slot	Reference Slot
Speak on topic	Addr
Part orientational	Area
Direction	
Locale	
Part inner outer	
Food	Food
origin	
(NULL)	Name
Contacting	Phone
Sending	Postcode
Commerce scenario	Price range
Expensiveness	
Range	
(NULL)	Signature
seeking	Task
Desiring	
Locating	
Locale by use	Type
building	

Slot Filling Evaluation

- MAP of the slot ranking model
  - For each slot, we compute F1by comparing the lists of extracted and reference slot fillers

The top-5 F1-measure slot-filling corresponding to matched slot mapping for ASR

SEMAFOR Slot	Locale by use	Speak on topic	Expensiveness	Origin	Direction
Reference Slot	Type	Addr	Price range	Food	Area
F1-Hard	89.75	88.86	62.05	36.00	29.81
F1-Soft	89.96	88.86	62.35	43.48	29.81

F1-Hard: the values of two slot fillers are exactly the same
F1-Soft: the values of two slot fillers both contain at least one overlapping words

Slot Induction and Filling Evaluation

- MAP-F1-Hard / MAP-F1-Soft
  - weight the MAP score with F1-Hard / F1-Soft scores

Approach	MAP-F1-Hard		MAP-F1-Soft	
	ASR	Manual	ASR	Manual
Frequency	26.96	27.84	27.29	28.68
K-Means	27.38	27.99	27.67	28.83
Spectral Clustering	**30.52**	**28.40**	**30.85**	**29.22**

When the induced slot mismatches the reference slot, all the slot fillers will be judged as incorrect fillers.

Overview

Introduction

Approach

Experiments

Conclusion

Conclusion

- We propose an unsupervised approach for automatic induction and filling of semantic slots.
- Our work makes use of a state-of-the-art semantic parser, and adapts the linguistically principled generic FrameNet-style outputs to the target semantic space corresponding to a domain-specific SDS setting.
- Our experiments show that automatically induced semantic slots align well with the reference slots created by domain experts.

Thanks for your attention! ☺

Q & A