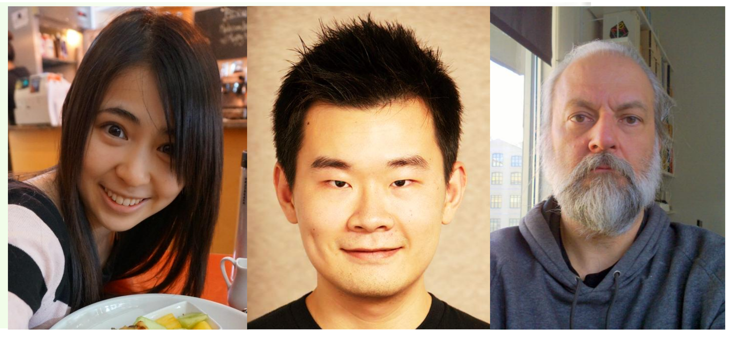
Carnegie Mellon
lti

UNSUPERVISED INDUCTION AND FILLING OF SEMANTIC SLOTS FOR SPOKEN DIALOGUE SYSTEMS USING FRAME-SEMANTIC PARSING

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



1. The Task

- Motivations
  - SDSs require predefined semantic slots to parse users' language input into unified semantic representations
  - Defining semantic slots requires the involvement of a domain expert
  - Frame semantics theory provides generic semantic information

> Given a collection of unlabeled raw audios, can we use the frame semantics theory to automatically induce and fill the semantic slots in an unsupervised fashion?

- Approaches
  - Obtain slot candidates using a state-of-the-art frame-semantic parser
  - Propose a clustering-based ranking model to distinguish generic semantic concepts and domain-specific concepts
- Results
  - Automatically induced semantic slots have MAP of 69.36% for ASR-transcribed data

2. Probabilistic Frame-Semantic Parsing

- Frame Semantics theory states that the meaning of most words can be expressed on the basis of semantic frames:
  - frame (F): the frame "food" contains words referring to items of food
  - frame elements (FE): a descriptor frame element within the "food" frame indicates the characteristic of the food.
  - lexical units (LU): the values of the corresponding frame element, such as "milk"
- SEMAFOR is a state-of-the-art semantic parser for frame semantic parsing trained on a linguistically-principled semantic resource FrameNet.

> How to map and adapt the FrameNet-style frame-semantic parses to the semantic slots in the target semantic space for practical use in the spoken dialogue systems?

- We parse all ASR-decoded utterances using SEMAFOR and extract all frames from semantic parsing results as slot candidates, where the LUs that correspond to the frames are extracted for slot-filling.
  - Slot candidate
  - Slot filler

can i have a cheap restaurant

Frame: capability	Frame: expensiveness	Frame: locale by use
FT LU: can FE LU: i	FT LU: cheap	FT/FE LU: restaurant

generic semantic concept (useless for SDS) — domain-specific concept

3. Slot Ranking Models

- **Main idea: rank domain-specific concepts higher than generic semantic concepts**
- Rank the slot candidates by integrating two scores

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

  - $f(s_i)$: the frequency of each candidate slot in the SEMAFOR-parsed corpus

> slots with higher frequency may be more important

  - $h(s_i)$: the coherence of values the slot corresponds to

> We assume that domain-specific concepts should focus on fewer topics and be similar to each other, so the coherence can help measure the prominence of the slots.

slot: quantity — a, one, three, all — lower coherence in topic space

slot: expensiveness — cheap, inexpensive, expensive — higher coherence in topic space

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

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

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

  - We have corresponding cluster vectors

$$C(s_i) = \{\mathbf{c_1}, ..., \mathbf{c_j}, ..., \mathbf{c_J}\}$$

$$\mathbf{c_j} = [c_{j1}, ..., \boxed{c_{jk}}, ...c_{jK}]$$

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, which are more specific, which is preferable for slots of SDS.

- Spectral clustering

> 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 other approaches

  - For each word $w = [r_1, ..., r_i, ...]$

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

> We assume that two words are topically related when they occur in the same utterance.

  - 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

4. Experiments

- **Domain**: restaurant recommendation in an in-car setting in Cambridge* (WER = 37%)
  - Dialogue slot (total #slot = 10):
    addr, area, food, name, phone, postcode, price range, signature, task, and 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 that are actually used in a real world dialogue system can be induced automatically in an unsupervised fashion using our approach.

- **Slot Filling Evaluation**
  - For each slot, we compute F-measure by comparing the lists of extracted slot fillers with the induced slots and the slot fillers in the reference list

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

* M. Henderson, M. Gašić, B. Thomson, P. Tsiakoulis, K. Yu, and S. Young, "Discriminative spoken language understanding using word confusion networks," in SLT, 2012.

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 Induction and Slot Filling Evaluation**
  - MAP-F1-Hard/Soft: weight the MAP score with F1-Hard or 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 mismatch the reference slot, all the slot fillers will be judged as incorrect fillers.

5. Conclusions

- 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 that corresponds to a domain-specific SDS setting.
- Our experiments show that automatically induced semantic slots align well with the reference slots created by domain experts.
- We focus on the slot-filling tasks that extract the slot-filler information from those automatically induced slots.

☺ Schedule a meeting with the author? ycchen@cs.cmu.edu vivian.ynchen