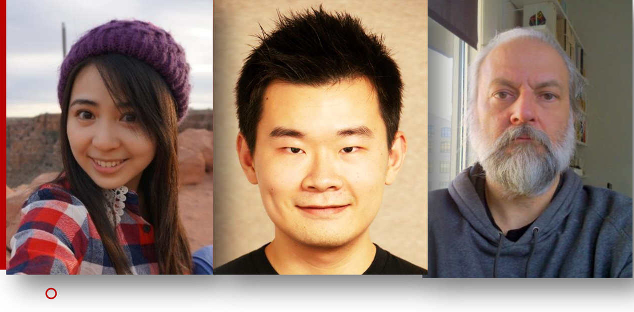
Learning Semantic Hierarchy with Distributional Representations for Unsupervised Spoken Language Understanding

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



Can a dialogue system automatically learn open domain knowledge?

Summary

- Motivation
  - Dialogue systems require a predefined semantic ontology; can it be learned from data?
  - A hierarchical ontology containing cross-slot information is crucial to SLU.
  - Word embeddings carry robust semantics.
- Approach
  1) HAC learns a hierarchical ontology based on FrameNet-parsed slot candidates and word embeddings.
  2) The slot importance estimated for different levels is integrated together to induce the ontology.
  3) The induced slots are used for training an SLU model.
- Result
  - With high-level information, the SLU model achieves 13% relative improvement on F1.
  - The learned hierarchy aligns well with the hand-craft mapping.

Low-Level Slot Importance Estimation

- Frame semantics parsing generates slot candidates (Chen et al., 2013; 2014)

can i have a cheap restaurant

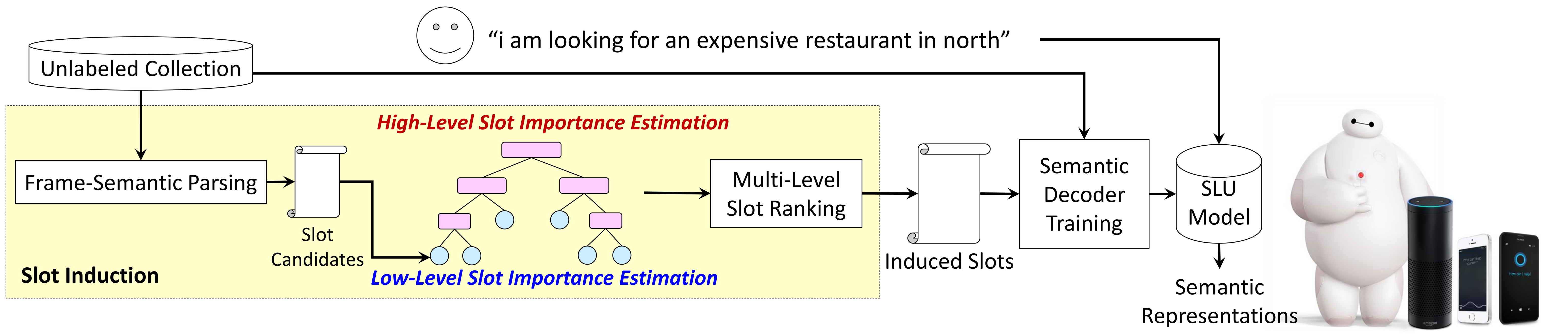
Frame: capability FT LU: can FE Filler: i	Frame: expensiveness FT LU: cheap	Frame: locale by use FT/FE LU: restaurant

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

- **Idea: rank domain-specific concepts higher than generic semantic concepts**

$$w^0(s) = (1-\alpha)\log f(s) + \alpha \log h(s)$$

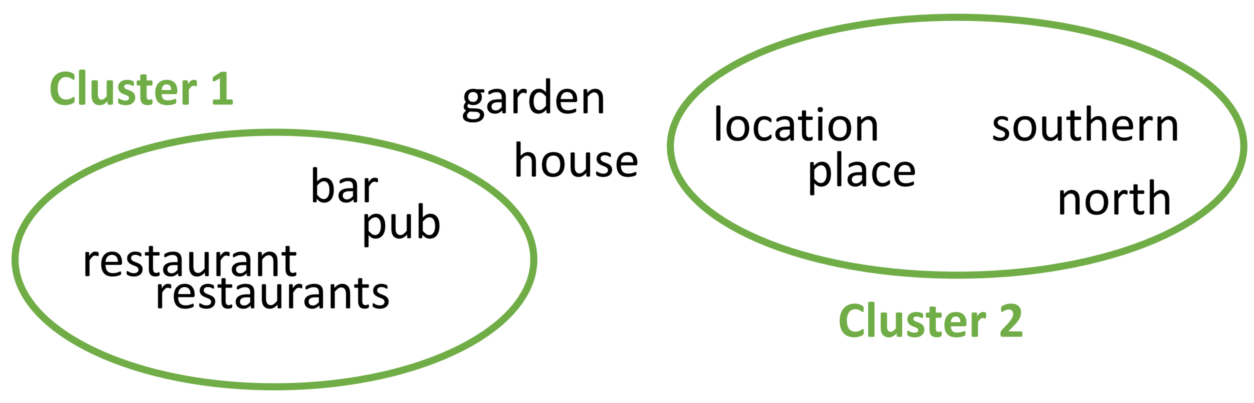
- $f(s)$: the slot frequency in the parsed corpus
  - ➢slots with higher frequency → more important
- $h(s)$: the coherence of slot-fillers
  - ➢domain-specific concepts focus on fewer topics → coherence can help measure slot prominence



High-Level Slot Importance Estimation

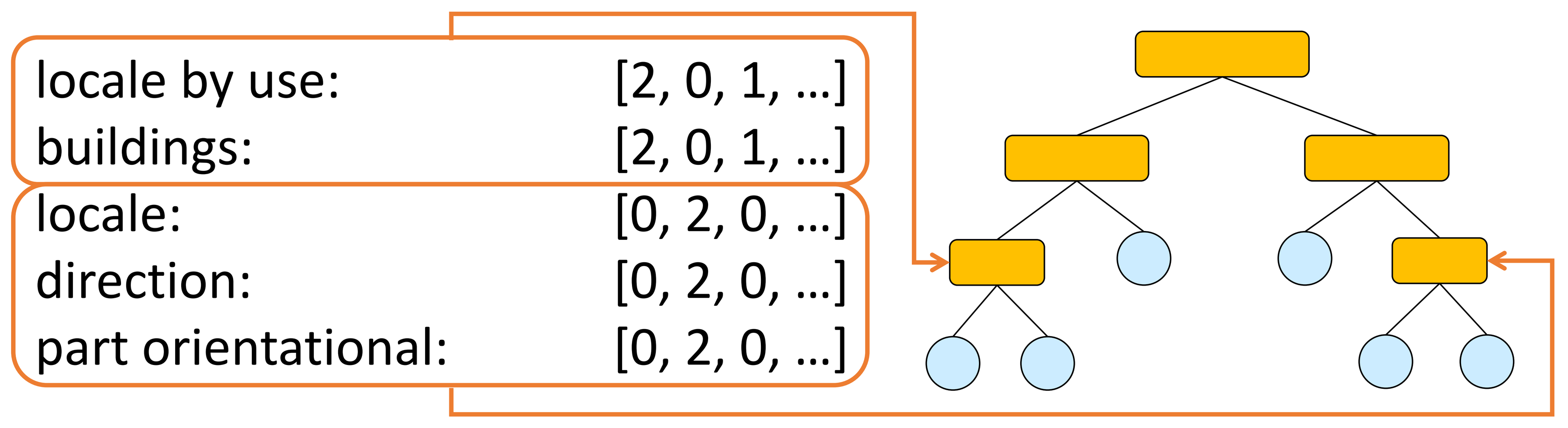
- **Hierarchical Agglomerative Clustering (HAC)** performs a bottom-up clustering approach by successively merging similar clusters together.
  - The distance between two clusters A and B is defined as $\frac{1}{|A||B|}\sum_{a\in A}\sum_{b\in B} d(a,b)$

- **Word-level clustering** groups the words with closer embeddings since they have similar contexts



➢Word embeddings help merge semantically similar words together.

- **Slot-level clustering** groups the slots with closer vectors built by the word-level clustering results



➢Different slot candidates generated by the frame semantic parser can be merged because they share similar clustering distribution.

- **Bottom-up slot importance estimation** estimates the high-level slot importance by aggragating the low-level importance:

$$w^{h+1}(s) = \frac{1}{|C^h(s)|}\sum_{c(s_k)=c(s)} w^h(s_k)$$

Multi-Level Slot Ranking

- **Idea: rank slots considering all different levels of the hierarchy**

$$w(s) = \sum_{h=0}^{H} \lambda_h w^h(s)$$

➢The final slot importance contains hierarchical information.

Experiments

- **Domain**: restaurant recommendation in an in-car setting in Cambridge (Word Error Rate = 37%)
  - Dialogue slots: addr, area, food, phone, postcode, price range, task, and type

Approach		Slot Induction AUC (%)	SLU F1 (%)
Baseline	Low-Level	79.50	60.27
Proposed	High-Level	81.28	67.94
	Multi-Level	**82.00**	**68.13**

Learned Semantic Hierarchy

- locale by use / building
- range / diversity / increment
- commerce scenario / expensiveness
- part orientational / direction / locale

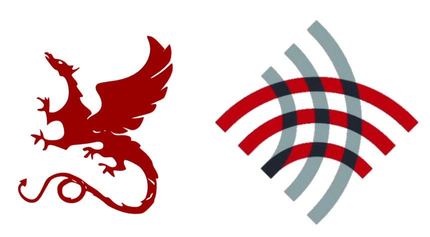
Mappings of Induced Slots (within the blocks) and Expert-Defined Semantic Slots (right sides of arrows)

- locale by use, building → type
- food, origin → food
- speak on topic → addr
- seeking, desiring, locating → task
- commerce scenario, expensiveness, range → pricerange
- part orientational, direction, locale, part inner outer → area
- contacting → phone
- sending → postcode

➢High-level information helps both slot induction and SLU performance, where the learned hierarchy aligns well with the manual mapping.

Conclusion

- We propose an **unsupervised approach** unifying semantics from **a hierarchical structure** to improve slot induction and SLU modeling.
- Our automatically induced semantic slots align well with reference slots.
- We show the feasibility of training an SLU model based on automatically induced slots and its promising performance for practical usage.



Carnegie Mellon University