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.

- **Idea**
  - Frame semantics parsing generates slot candidates (Chen et al., 2013; 2014)
  - We propose a dialogue system that can automatically learn semantic knowledge.

- **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).
- We show in a cheap restaurant example:
  - Frame: capability
  - Frame: expensive
  - Frame: locale by use
  - Frame: restaurant

- **Idea**
  - Rank-domain specific concepts higher than generic concepts
  - \( w(s) = (1 - \alpha) \log f(s) + \alpha \log h(s) \)
  - \( f(s) \): the slot frequency in the parsed corpus
  - \( h(s) \): the coherence of slot-filler

**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.

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

- **Word embeddings help merge semantically similar words together.**

- **Bottom-up slot importance estimation** estimates the high-level slot importance by aggregating the low-level importance:
  - \( w^{h}(s) = \frac{1}{|C(h)|} \sum_{s_j \in C(h)} w^{l}(s_j) \)

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

- **Multi-Level Slot Ranking**
  - Idea: rank slots considering all different levels of the hierarchy
  - \( w(s) = \sum_{i} w_i^{h}(s) \)
  - The final slot importance contains hierarchical information.

**Experiments**

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

- **Approach**
  - Baseline: Low-Level AUC (%): 79.50
  - High-Level: 81.28
  - Multi-Level: 82.00

- **Slot Induction**
  - Proposed
  - SLU F1 (%): 60.27
  - 67.94
  - 68.13

- **Conclusion**
  - Learned Semantic Hierarchy helps both slot induction and SLU performance, where the learned hierarchy aligns well with the manual mapping.