Automatic Task Ontology Creation for Human-Computer Dialogue, using Word Net and Dialogue Corpora

Vladimir POPESCU, Jean CAELEN, Ioana HAGIESCU

Introduction

Research context

• Task-dependent human-computer dialogue for collaborative systems
• Knowledge-based approach ⇒ usage of task-specific ontologies
• Task ontologies ← usually handcrafted for each application domain ⇒ lack of portability & genericity

Goals

• Better generality regarding the application domain & task
• Automatic task ontology creation

Framework

• Generic task-oriented human-computer dialogue (Caelen, 1992-):
  – task-independent modules: Automatic speech recognition, Morpho-syntactic parsing, Dialogue control, Speech synthesis
  – task-dependent modules: Semantic analysis, Pragmatic interpretation, Answer generation
  – task-specific module: Task control - for explicit and implicit (goal-driven) task models

Solution for Ontology Creation

• Approach: rule-based, data-driven
• Resources:
  – universal ontologies: Word Net, SUMO, etc.
  – human-computer dialogue corpora, acquired through the Wizard of Oz method
• Solution: selection of task-specific sub-ontologies from universal ontologies, using relevant dialogue transcriptions

Ontology Creation Algorithm

1. Word clustering in semantic roles - agent ([agt]), object ([obj]), patient ([pat]), modifier ([mod]), predicate ([pred]);
2. Parsing of corpus utterances:
   (a) elimination of connectives and function words (viz. prepositions, conjunctions, ...);
   (b) lemmatization of semantically relevant words;
3. Mapping of lemmatized words to concepts in the universal ontology (e.g. Word Net);
4. Retrieval of the sub-graph containing the concepts found at step 3.:
   • graph kernel: the concepts found at step 3. as nodes;
   • graph extension: tournament graphs connecting these concepts to the universal concept (⊤) ⇒ hyponymic & hypernymic chains included;
   • graph edges: labels – ISA (instances → classes connections), AKO (“A Kind Of”, subclasses → classes connections).

Example

User: Hello, I would like a book on "X".

Machine : We have two books on "X", namely "Y" by author "A" and "Z" by author "B".

[agt] = {I, we, author}
[obj] = {book}
[pat] = {two}
<pred> = {like, have}

Conclusions

• General mechanism for automatic task ontology creation
• Parameterization of generic human-computer dialogue systems with regard to the task
• Rule-based data-driven approach, easily customizable for different languages and tasks

Prospects and Challenges

• Comparisons: automatically created vs. handcrafted ontologies for several tasks (e.g. room reservation, in the PVE project)
• Manual post-processing of the ontology automatically obtained ⇔ elimination of inconsistencies in universal ontologies (Word Net, SUMO, etc.)