Proposal Summary

Outline of the project

The goal of the proposed work is to build an ontology –a formal representation of organized knowledge or information of concepts and their relationships in hierarchy— that is semantically rich. Specifically, this project aims to introduce syntactic/semantic restrictions and axioms to each object of the ontology; i.e., the ontology objects have information on their syntactic selectional restrictions and semantic role preferences as well as their definitions in logical forms.

This project mainly uses two publicly available lexical databases: WordNet and VerbNet. WordNet provides definitions of find-grained word senses and VerbNet provides verb groupings and syntactic and semantic selectivity of the groups. Both WordNet and VerbNet offer a very shallow taxonomy – i.e., only few different hierarchical levels exist between groups of words. The proposed research focuses on improving on this weakness of the two lexical databases by using their available resources; (i) from WordNet, by obtaining around 150,000 words and their definitions, the ontology expands its vocabulary size which will add more layers to the type hierarchy and (ii) from VerbNet, verb groupings give syntactic restrictions and an approximate measure of which verbs should be close to each other in the ontology.

Building the ontology involves automatic computation of parsing and subsuming definitions. First, the definitions of fine-grained word senses of WordNet are parsed by TRIPS parser. The sentences that the parser outputs are in TRIPS logical forms and current TRIPS ontology types. Next, by using the existing hierarchy of TIRPS ontology, definitions are organized in new hierarchy according to their subsumption relations to each other. Subsumption relations are mainly based on the definitions' words' TRIPS hierarchy, in the order of priority from the main predicate (most preferred) to adverbial modifiers (least preferred).

Resources Needed

It is important to evaluate the tentative result of the ontology model based on the common knowledge shared among the people who understand the language in which the ontology is written. For this reason, the proposed research requires hiring native English speakers who are not involved in this project and collecting judgments from them. The human labor resources are required for two parts: testing whether the hierarchical relationships are well-established and whether the semantic roles are correctly tagged in the ontology objects.

First, participants are asked to answer how much they agree to the hierarchy between given two words in different groups. Considering that the participants are not familiar with the ontology, instead of presenting them the names of ontology objects, the experiment adopts another way to ask this question by using entailment, i.e., one sentence is true if all the others are also true, implied from the subsumption relationships between two words. The participants are present with two sentences – two exactly same sentences except that one includes a word and the other includes the ancestor of the word, and asked to answer to the question whether the second sentence holds true when the first sentence is true. For example, to test a verb hierarchy of the ontology, a participant is given the following two sentences: “we ate chicken at 7am” and “we consumed chicken at 7am”. Since it is true that “we consumed chicken at 7am” when “we ate chicken at 7am” is true, the ontology correctly guessed that ‘eat’ is a descendant of ‘consume’.

Second, to evaluate semantic role restrictions for verbs in the ontology, hiring participants who have linguistics knowledge, specifically in semantic roles, is required. The participants are asked to give scores to the sentences that are generated by TRIPS parser based on their tagged semantic roles. Once the data are collected, the semantic restrictions of the sentences that are given low scores need to be revised. In these strategies, collecting human annotations and judgments helps iteratively re-modeling decisions to develop the ontology.

Potential Benefits

Improving ontology has merits in its applications in the realm of artificial intelligence, especially in natural language processing. First, the proposed project helps managing the knowledge data on natural language, from specific vocabularies and to vague concepts as well as their relationships and attributes in compact representations. Formally specified and well-organized ontologies make it easy for computer to use the data for automated process.

Second, the taxonomic relationships between words in the ontology are useful for automatic extraction of event knowledge from given text or speech. Specifically, commonsense knowledge regarding entailment can be extracted from the subsumption relations. For example, given the sentence “Susan ate chicken”, the machine can infer that Susan consumed meat, following the hierarchy of ‘chicken’-‘poultry’-‘meat’ and ‘eat’-‘consume’.

Third, the broader coverage of an ontology with expanded vocabularies becomes useful in measuring the semantic distance between words. It helps disambiguating word senses by looking at the context of dialogues. For instance, when the machine is given a word ‘bass’, which has two word-senses (fish and guitar), it needs to measure the semantic distance between a word from the conversation context and ‘bass’ in order to choose the correct sense. If the dialogue involves a word ‘instrument’, then the machine measures the semantic distance between ‘instrument’ and the two senses of ‘bass’. Since the semantic distance between ‘instrument’ and ‘bass’ (guitar) is shorter than that between ‘instrument’ and ‘bass’ (fish) – i.e., ‘instrument’ and ‘bass’ (guitar) are closer to each other in the ontology’s hierarchy than ‘instrument’ and ‘bass’ (fish) – the machine is expected to interpret the word ‘bass’ as a bass guitar.

The aforementioned points make big contributions to alleviating current issues in human-machine communications. Improving the ontology leads to the improvement of a parser, which brings on the improvement of dialogue systems. Applications of dialogue systems are in the following areas. In education, as computer system can converse with a human, tutoring by robots gives students with diverse needs the opportunities to learn at a low cost. In welfare, social robots are beneficial to those who seek friends but might have social anxiety. In health care, dialogue systems could be one’s personal assistant and remind to take medicine every day at the right time. In communication, under an emergent situation, dialogue systems understand questions correctly and provide informative responses to resolve a problem when there are no experts around to help. Therefore, improving ontologies directly impacts and enhance many areas of an individual’s wellbeing.