Studies on understanding metaphors using knowledge of noun properties

Tatsuo Mishima (510096)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 8, 2007

Keywords: understanding metaphors,property name,property,salience,corpus.

Among various language phenomena, understanding metaphors by computers is difficult. Our language almost contain metaphoric phrases and it's pointed out that metaphors are concerned with our conceptual hierarchy. So, understanding metaphors is very important for NLP, but traditional ways of NLP may tail to understand metaphors. For example "ringo no youna hoo (cheeks like an apple)" means "cheeks are red". But the concept "red" can't be derived from syntactic or semantic analysis. We need knowledge that an apple has typical property "red". In NLP domain, psychological experiments are major to get knowledge for understanding metaphors. In recent years, however, WWW and corpus are available for building large scale knowledge automatically.

So, this paper presents a computational model for understanding metaphors. Our system builds knowledge base of nouns by getting co-occurrence frequency of noun and noun's property from corpora, and selects semantic correct property for understanding metaphor expression.

We define understanding metaphor "A no youna B (B like A)" is to determine a correct a pair of property and property name that is transferred from A to B. A property name is classification of property such as "color" or "figure" etc. and a property is a noun's feature such as "red" or "round". For example, our system outputs a property pair $(color, red)$ as transferred property from A to B for the phrase "ringo no youna hoo". We used thesaurus to define properties and property names, adjectives for properties and semantic classes for property names. We also remove properties and property names that are not suitable for understanding metaphors. As a result, we prepare 100 property names and 3006 properties.

An important factor to understanding metaphors is salience of property, that is how typical the property is. To measure salience , we need knowledge about properties of nouns. To handle various metaphors, we get co-occurrence frequency of nouns and properties from corpora by pattern matching for constraction of large scale noun knowledge base. But, a pair of noun and salient property may not appear often in corpora. Therefore, using simple patterns only can't get high frequency of salient property in accordance with human's intuition. So we find patterns that can extract salient property more precisely than simple patterns, and use those patterns to improve system's precision.

We built a computational model for understanding metaphors based on knowledge of nouns that was built by above way. As a result of understanding metaphor, the system outputs a pair of property name and property that has the highest score. A score is designed considering the following conditions: giving high score when difference between candidate property's frequency and other's is large, putting weights for frequency obtained by co-occurrence pattern that can extract salience property better than simple pattern, giving high score to candidate property name that tends to modify B. We ignore candidate pairs of property name and property that has few co-occurrence because they are unreliable..

In the experiment, the proposed method is applied for phrases of "A no youna B", that are extracted from corpora. As a result , as for selecting both correct property name and correct property, precision is 62% and as for selecting only correct property , precision is 71%. While, baseline that simply outputs property which has the highest co-occurrence, achieves 32% precision. This results shows the effectiveness of the proposed method. According to the results of some additional experiments, we found that major three ideas of this research, calculation salience by property name, using pattern extracting salient property precisely, considering with relation between property name and noun B, are useful for understanding metaphors.

For the future works, we need to find other patterns extracting salient property to construct better knowledge of properties of noun to understand metaphor. We found problems that we sometimes treat similar property, like "kuroi (black)" and "dosu-kuroi (dark black)", as different properties and co-occurrence frequency of them are also different. Therefore, the system sometimes can't understand that a certain noun has typical property "black". So we need to redifine the set of properties, for example "black" and "dark black" should be defined the as same property.