TOWARD COMPUTATIONAL MODEL OF UNDERSTANDING METAPHORS

Takenobu Tokunaga†, *Hozumi Tanaka*† *and Makoto Iwayama*‡

†Department of Computer Science, Tokyo Institute of Technology
‡Hitachi, Ltd., Central Research Laboratory

ABSTRACT

This paper presents a computational method of calculating the measure of salience in understanding metaphors. We mainly treat metaphors in the form of "A is (like) B," in which "A" is called *target concept*, and "B" is called *source concept*. In understanding a metaphor, some properties of the source concept are transferred to the target concept. In the transfer process, we first have to select the properties of the source concept that can be more preferably transferred to the target concept. The measure of salience represents how typical or prominent the property is and is used to measure the transferability of the property. By introducing the measure of salience, we have to consider only the high salient properties after the selection. The measure of salience was calculated from Smith & Medin's probabilistic concept according to Tversky's two factors; intensity and diagnostic factor.

1. INTRODUCTION

Natural language is a rich source of metaphors, and metaphors have strong relationship with the conceptual structure that has been acquired through our everyday life [2]. Thus, understanding metaphors is one of important research topic in natural language processing.

In this paper, we focus on metaphors in the form of "A is (like) B," in which "A" is called *target concept*, and "B" is called *source concept*. We consider the understanding metaphor as a transfer process of properties from the source concept to the target concept. For example, in the case of "A man is a wolf," some properties of "wolf" — "being vicious, dangerous, fierce, etc." are transferred to "man." As a consequence of the transfer, "man's" properties "being vicious, dangerous, fierce, etc." are highlighted.

This transfer process consists of the following three steps. First, we have to select properties of the source concept that can be transferred to the target concept. We call this step *selection step*. Secondly, we have to find the properties of the target concepts which correspond to the properties selected in the selection step. We call this step *mapping step*. Finally, we have to highlight or downplay the properties of the target concept according to the corresponding properties of the source concept. We call this step *variance step*. These steps are very similar to that of analogical reasoning [1]. In this paper, we focus on the selection step and show how this step is achieved.

In the selection step, we have to select the properties of the source concept that can be more preferably transferred to the target concept. We introduce *the measure of salience* that measures the transferability of the property. Generally, the measure of salience represents how typical or prominent the property is. By introducing the measure of salience, we have to consider only the high salient properties after the selection step. There are many properties that play little importance during the whole process of understanding metaphors. With respect to the above example, "wolf's" properties "being vicious," "being fierce," "being dangerous" are high salient properties and are more likely to be transferred to "man." On the other hand, "having two eyes," "having four legs," etc. are low salient properties and cut off at the selection step.

Many researchers have used salience in the process of understanding metaphors [4, 3, 8], but they have not described precisely how salience is calculated. It is necessary to show the foundation which salience was based on and the method of calculating salience based on the foundation. In this paper, we propose a method to calculate the measure of salience from Smith & Medin's probabilistic concept [5, 6] which has a grounding in probability theory. According to Tversky [7], we calculate the measure of salience in terms of two factors. One is intensity which refers to the signal-to-noise ratio; this is calculated from the entropy of properties. The other is diagnostic factor which refers to the classificatory significance of properties; this is calculated from the distribution of the property's intensity among similar concepts. Finally we briefly outline the whole process of understanding metaphors using the measure of salience.

2. PROBABILISTIC CONCEPT AND THE MEASURE OF SALIENCE

In this section, we describe our method of calculating the measure of salience. First, we briefly review *probabilistic concept* that Smith & Medin have proposed [5, 6]. Our measure of salience is calculated based on the probabilistic concept.

2.1. Probabilistic Concept

Probabilistic concept is composed of a set of properties, each of which has an attribute with a set of possible values. Probability is attached to each value. We use probabilistic concept as our concept representation.

Definition 1 Concept
A *Concept* denoted by *(C) is a set of *properties* S_i.

$$*(C) = \{S_1, S_2, \ldots, S_n\}$$

A property S_i is a pair of an *attribute* a_i and its *possible value set* V_i.

$$S_i = a_i : V_i$$

The possible value set V_i is a set of which element is a pair of a_i's *possible value* $v_{i,j}$ and its *probability* $w_{i,j}$ among the V_i.

$$V_i = \{v_{i,1}\#w_{i,1}, \ldots, v_{i,j}\#w_{i,j}, \ldots, v_{i,m}\#w_{i,m}\}$$

That is,

$$\sum_{j=1}^{m} w_{i,j} = 1$$

Most Likelihood Value (MLV) $v_{i,max}$ is the value with the highest probability among a possible value set V_i, and *Most Likelihood Property (MLP)* $S_{i,max}$ is the pair of attribute a_i and its MLV $v_{i,max}$ and is denoted by $a_i : v_{i,max}$.

Following is an example of *(Apple).

Example 1 Definition of *(Apple)

$$\text{*(Apple)} = \left\{ \begin{array}{ll} color : & \left\{ \begin{array}{l} red\#0.8 \\ green\#0.15 \\ brown\#0.05 \end{array} \right\} \\ shape : & \left\{ \begin{array}{l} round\#0.95 \\ cylindrical\#0.05 \end{array} \right\} \\ texture : & \left\{ \begin{array}{l} smooth\#0.9 \\ rough\#0.1 \end{array} \right\} \\ \vdots & \end{array} \right\}$$

"$color : \{red\#0.8, green\#0.15, brown\#0.05\}$" is a property, and "$\{red\#0.8, green\#0.15, brown\#0.05\}$" is its possible value set, where each real number is the probability of the value. "red" is the MLV and "$color : red$" is the MLP.

2.2. The Measure of Salience

Each property has a measure of salience which is a real number ranging from 0 to 1. 0 and 1 represents the lowest and the highest salience respectively. The measure of salience represents the typicality of a property and is used in understanding metaphors to decide which properties of a source concept might be more preferably transferred to a target concept. Many researchers have used the measure of salience in the same way as mentioned above [4, 3, 8], but they have not shown the precise method to calculate the measure of salience.

Tversky claims that salience is determined by two types of factors; intensitive and diagnostic [7]. The former refers to the signal-to-noise ratio and the later refers to the classificatory significance of properties. In the following sections, we show the method of calculating the measure of salience according to Tversky's two factors. The intensitive factor is calculated based on the entropy in information theory, and we call this measure *the Amount of Information of Property (AIP)*. The diagnostic factor is calculated based on the distribution of a property's AIP among similar concepts, and we call this measure *the Difference of Property (DP)*.

2.2.1. The Amount of Information of Property (AIP)

The first factor in calculating the measure of salience is the amount of information which a property has. This is calculated by the entropy of a possible value set V_i. Because the entropy is a measure of randomness, the lower the entropy is, the less random a possible value set V_i is, that means V_i has more redundant information. Intuitively, more redundant V_i means that its MLV $v_{i,max}$ occurs more frequent comparing with other values of V_i. It follows that the property S_i with more redundant V_i is the more typical and salient property.

For example, compare the following two properties:

$color{:}\{red\#0.6, green\#0.1, yellow\#0.1, blue\#0.1, brown\#0.1\}$
$color{:}\{red\#0.6, green\#0.4\}$

In these two possible value sets, probability of each MLV "red" are the same 0.6. But if we take account of the distribution of all elements in each possible value set, MLV of V_1 occurs more redundantly than that of V_2 because, the degree of concentration of V_1 is higher than that of V_2. In fact entropy (relative entropy) of V_1 (i.e. 0.7627 according to the definition below) is lower than that of V_2 (i.e. 0.9705), and it shows that V_1 is more redundant and has more information than V_2.

Definition 2 The amount of information of property(AIP)
Given a property $S_i = a_i : V_i$, in which

$$V_i = \{v_{i,1}\#w_{i,1}, v_{i,2}\#w_{i,2}, \ldots, v_{i,m}\#w_{i,m}\}$$

the AIP of S_i is denoted by $r(V_i)$ and calculated by the following expression:

$$r(V_i) = 1 - h(V_i) \ \text{(redundancy)}$$

where $h(V_i)$ is

$$h(V_i) = \begin{cases} 0 & \text{if } m = 1 \\ \frac{H(V_i)}{\log_2 m} & \text{otherwise.} \end{cases} \quad \text{(relative entropy)}$$

$H(V_i)$ is

$$H(V_i) = \sum_{j=1}^{m} w_{i,j} \log_2 \frac{1}{w_{i,j}} \ \text{(entropy)}$$

The AIP ranges from 0 to 1 depending on the diversity of a possible value set. If the diversity of a possible value set concentrates on only one value — for example $x : \{a\#1, b\#0, c\#0, d\#0\}$, the property has the highest AIP 1, because all instances has the value a. To the contrary, if the diversity of a possible value set is averaged — for example $x : \{a\#0.25, b\#0.25, c\#0.25, d\#0.25\}$, the property has the lowest AIP 0, because one can not successfully predict which value a instance has.

2.2.2. The Difference of Property (DP)

The second factor in calculating the measure of salience is the difference of a property among similar concepts. It is the distribution of a property's AIP among similar concepts. Intuitively, the more distinguished property from other similar concepts has the higher value of DP and this property becomes higher salient.

Definition 3 The Difference of Property (DP)
Given a concept $*(C)$'s property $S_i = a_i : V_i$ and a set of similar concepts $Sim(*(C))$ including $*(C)$, the DP of the S_i is denoted by $d(S_i)$ and calculated by the following expression:

$$d(S_i) = \frac{r(V_i)}{\sum^{*(C_j) \in Sim(*(C))} \sum^{S_k \in *(C_j)} r'(S_k, S_i)}$$

where $r'(S_k, S_i)$ is calculated by the following expression:

$$r'(S_k, S_i) = \begin{cases} r(V_k) & \text{if } S_{k,max} = S_{i,max} \\ 0 & \text{otherwise.} \end{cases}$$

In this paper, we define the similar concepts as the concepts that has the same parent node in the IS-A hierarchy.

The DP ranges from nearly equal to 0 to 1. If a MLP of a property is the unique MLP among similar concepts, this property is the most distinguished property and the DP of this property becomes 1. For example the MLP of penguin "$can\_fly : no$" is in this case, because all other birds have the different MLP "$can\_fly : yes$." To the contrary, if every similar concepts has the same MLP, the DP of the property which has the MLP become nearly equal to 0. For example, because all fruits have the MLP "$have\_seed : yes$," the DP of this property of apple becomes nearly equal to 0.

2.2.3. The measure of salience of properties

The measure of salience of a property is calculated by the following definition.

Definition 4 The measure of salience of a property
The measure of salience of a property S_i is calculated by the following expression:

$$salience(S_i) = r(V_i) \times d(S_i)$$

Because $d(S_i)$ is the rate of $r(V_i)$ occupying among similar concepts, $salience(S_i)$ represent the apparent AIP in similar concepts. For example, if $d(S_i)$ is 1, S_i is the most distinguished property and the apparent AIP is the same as the net AIP (i.e. $r(V_i)$) itself. If $d(S_i)$ is much lower, there are many similar concept that have the same MLP of S_i and the apparent AIP becomes lower than the net AIP.

3. WHERE DOES PROBABILITY COME FROM?

The probabilistic concept plays a key role in the proposed method. To implement the model as a working system, we need to obtain probabilities of each attribute values of each concept. We conducted preliminary experiments to obtain attribute value probabilities from language resources. In the experiments, we focus on adjective and noun collocations, since an adjective is expected to describe an attribute of the modified noun. The probability of attributes would be calculated based on the frequency of adjective-noun collocations.

We first decided a set of target nouns. Our goal is to implement the proposed model and evaluate it by comparing with the results of psychological experiments with human subjects. To decide the target nouns, we extracted instances of expression "*A no youna B* (B like A)" from on-line novels (*Aozora Bunko*). We collected 4,085 nouns in type at the position A (source concept), and 3,105 nouns at the position B (target concept) from 2,949 novels (about 100MB). Among these nouns, 924 nouns appeared both position A and B, which were decided as target nouns.

As for target adjectives, we referred to a medium size Japanese dictionary (*Iwanami Kokugo Ziten*) and a dictionary of a Japanese morphological analyzer (*Chasen*) to define 686 adjectives as the target. These adjectives were included in both dictionaries.

Then, we generated all combinations of an adjective and a noun by using these target sets, and submitted them to a

Web search engine (www.goo.ne.jp) as queries to verify if the collocations could be used. The number of pages which include the collocation is considered as their frequency.

Table 1. Example of collocations

Collocation with "*ringo* (apple)"	
Adjective	Frequency
oisii (delicious)	1,230
akai (red)	852
aoi (blue)	752
amai (sweet)	557
amazuppai (sweet and sour)	292
yoi (good)	273
takai (expensive)	186
ookii (big)	161
ooi (many)	135
tiisai (small)	134

Collocation with "*hoo* (cheek)"	
Adjective	Frequency
siroi (white)	859
tumetai (cold)	783
akai (red)	478
yawarakai (soft)	172
aoziroi (pale)	135
marui (round)	122
atui (hot)	111
utukusii (beautiful)	103
kawaii (cute)	73
usui (thin)	64

Table 1 shows a list of adjectives which frequently cooccur with "*ringo* (apple)" and "*hoo* (cheek)". To derive probabilities from such frequency data, we need to make groups of attribute values to identify a corresponding attribute. For example, "*akai* (red)", "*aoi* (blue)", "*siroi* (white)" and "*aoziroi* (pale)" should be grouped as values of attribute *color*. Based on this preliminary experiments, we are going to develop a grouping algorithm.

4. CONCLUSION

In this paper, we have proposed a method of calculating the measure of salience for understanding metaphors.

The measure of salience represents typicality of a property and can be used in various inferences as a measure of preference. This is an aspect of the utility of salience, and the understanding metaphors is one of them.

The measure of salience proposed in this paper is based on the probability attached to attribute's values, and we have shown the precise method how salience is measured from the probability. Our measure is based on the entropy in information theory and more formal than other system's score of salience.

We have conducted preliminary experiments to obtain probabilistic concepts by using language resources, and found the used method promising. Our future research plan includes implementing the model and conducting psychological experiments to evaluate our model with the collected data.

5. REFERENCES

[1] R. P. Hall. Computational approaches to analogical reasoning: A comparative analysis. *Artificial Intelligence*, 39:39–120, 1989.

[2] G. Lakoff and M. Johnson. *Metaphors we live by*. Chicago University Press, 1980.

[3] A. Ortony. Beyond literal similarity. *Psychological Review*, 86(3):161–180, 1979.

[4] A. Ortony. The role of similarity in similes and metaphors. In A. Ortony, editor, *Metaphor and Thought*, pages 186–201. Cambridge University Press, 1979.

[5] E. Smith and D. Medin. *Categories and Concepts*. Harvard University Press, Cambridge, MA, 1981.

[6] E. E. Smith, D. N. Osherson, L. Rips, and M. Keane. Combining prototypes : A selective modification model. *Cognitive Science*, 12(4):485–527, 1988.

[7] A. Tversky. Features of similarity. *Psycological Review*, 84(4):327–352, 1977.

[8] E. J. Weiner. A knowledge representation approach to understanding metaphors. *Computational Linguistics*, 10(1):1–14, 1984.