Automatic Acquisition and Application of Translation Rules from Sentences to Semantic Representations Using Inductive Learning

Eigo Mori  Yoshikazu Miyanaga  Koji Tochinai  Kenji Araki
Hokkaido University  Hokkai-Gakuen University

1 Introduction

For the deep analysis of sentences such as Natural Language Understanding, sentences are needed to be transferred to semantic representations to enable a computer to handle their meanings (Greene, 1986) (Ishizak, 1995). The traditional methods of translation from sentences to semantic representations consist of several processes such as morphological analysis, syntactic analysis and semantic analysis. In the method, the rules for each analysis are designed manually. For constructing a system which can handle actual sentences, an expanded set of rules are needed. However, it is a difficult task to expand the rule-sets with avoiding conflicts among the rules. To solve this problem, we propose an automatic acquisition method using inductive learning. In the proposed method, the system acquires translation rules from sentences to semantic representations automatically through generalizations by applying known rules and inferring partial rules (Araki and Tochinai, 1992) (Mori et al., 1996). The translation rules are connected to associations-keys by weighted links, and suitable application rules are selected by pulling links from several association-keys. The system sets up the priority of each rule depending on the context by setting the weight of the links during the rule acquisition stage. The rule application is carried out as a verification of selected rules and it enables the system to operate processes simultaneously.

For analyzing the meanings of sentences, the ways of handling its ambiguity are very important. Generally the ambiguity is solved using various types of information such as context and commonsense. In the proposed method, thesaurus information is given as background-information. Disambiguation is carried out by this thesaurus information and link-weight settings during the rule acquisition stage. In this paper, the algorithm of the rule acquisition and the rule application are described. Experiments confirming the efficiency of our method are also shown.

2 Semantic Representations

2.1 Case Frame

In the proposed method, we adopted a deepcase frame as a semantic representation and we selected English as the input language (EDR, 1995) (Charniak and Wilks, 1976). The case frame is defined as a set of pairs and the pairs are feature and its value which are shown as the following:

\[ Ss \equiv \{[F1, V1], [F2, V2], \ldots, [Fn, Vn] \} \] (1)

The feature Fi is classified into 3 types, i.e. Concept Feature, Case Feature and Peripheral Feature.

2.2 Values of Features

2.2.1 Concept Feature

Here we describe the values of each feature (Vi). The value of Concept Feature (HEAD) corresponds to a principal meaning of a semantic representation. The meaning of a verb in a sentence and the meaning of a noun in a noun phrase represents the value of the concept feature commonly (Sells, 1985). In general, a word has plural meanings and to classify those meanings is not simple. In the proposed method, we classified the meanings using 2 processes shown below (Mori et al., 1995).


We adopted ”Word List by Semantic Principles (Bunrui-Goihyo)” as a thesaurus. This thesaurus classifies Japanese words into the 6 layered classes. Therefore this thesaurus put 6 digits number to each concept. Due to the Process 2 each word concept from the same Japanese translation has the same thesaurus information.

Word concept data, thesaurus data and morphological information are given to the system in advance. The data is given by the format: (Word/Word Concept/Morphological Info./Thesaurus Info.). The morphological information is given as a serial number in the words having the same concept. A part of the word concept dictionary is shown in Fig.1. The concepts of pronouns are defined as PRO and the concepts of idioms are defined like "LOOK-FORA" because whole "look for" corresponds to a concept.
2.2.2 Case Feature and Peripheral Feature

The case feature has a semantic representation as its value. It includes the case information subcategorized by verbs. The case features are defined for deep semantic roles. For instance, subject of "hit" has EXP, subject of like has AGT as its case feature. Examples of the case features are listed in Table 1 (B. Blake, 1994).

The peripheral features give additional information to the case frame such as the modality, the aspect and the person. Values are selected from the prepared ones for each feature. The peripheral features and their values are listed in Table 2.

2.2.3 An example of semantic representation

An example of a semantic representation in our method is shown in Fig.2. The concept of "like" is selected as a value of concept feature for a whole sentence, EXP, THEME and DGR are added as the case features. Each case takes the concept of a noun and a verb as a value of a concept feature in the corresponding phrase. TENSE and S-ATTR are added as the peripheral features of a whole sentence and NUMBER and DEF are added for the peripheral features in the embedded case frame.

3 Algorithms of Rule Acquisition and Application

3.1 Translation rules and link expression

We define a "translation rule" in this method as a simple pair of generalized surface expressions and a semantic representations. For making it easy to handle this representation on a computer, the surface expression is expressed by a list expression. The translation rules are described as the following:

\[ [W1, W2, ..., Wn] \quad (SEF) \]
\[ \Rightarrow \quad [[F1, V1], [F2, V2], ..., [Fn, Vn]] (SRF) \]

(2)

under the condition that

\[ Wi \in \{\text{Words}, \text{On}, m > \} \quad (3) \]

We call the generalized surface expression in a rule: Surface Expression Fragment (SEF) and a semantic expression in a rule: Semantic Representation Fragment (SRF). The symbol \( \text{On} \) (RM) represents a generalized part where must be replaced with a surface expression (in the SEF) and a semantic representation (in the SRF). A pair of the RMs having same numbers in the SEF and the corresponding SRF must be replaced with ones between which a translation rule exists. The symbol \( < m > \) represents a morphological information and works for the preceding RM. \( m \) corresponds to a morphological information in the word concept dictionary. An example of the translation rules is shown in Fig.3. In Fig.3 \( \text{O2} \) in the SEF is expected to be replaced with a verb and \( < 2 > \) implies that this verb is inflected due to the morphological information. (See Fig.1)

When a system applies rules to a input sentences, it needs to select a appropriate rules for the rule application. However, it is difficult
task for the system to select the suitable rule. In the proposed method, the system associates the association-keys with rules. The system does not have a rule-scanning process because rules are selected by pulling links with the maximum weight from the association-key. We call this method of the rule selection "rule association". There are three types of the rule association as the followings:

**Type 1.** Association of a SEF using a word as a key

**Type 2.** Association consists of 2 steps as the following:

Step 1. It associates a word with a word concept by pulling a dictionary link

Step 2. It associates the word concept with a SRF by pulling a association link
directly or through thesaurus links

**Type 3.** Association of a upper SRF using a SRF

When the system obtains a SEF or a SRF by pulling the links from an association key, opposite fragment (the SRF or the SEF) is obtained by pulling a rule link. The links are classified into four types.

**Association link:** Links from the association-keys to rules

**Rule link:** Links linking between a SEF and a SRF

**Thesaurus link:** Links representing thesaurus information

**Dictionary link:** Links from words to word concept

The thesaurus link and the dictionary link are constructed from the information in the word concept dictionary. The thesaurus links represent the 6 layered classes and connect nodes from the lowest level to which the word concepts is connected (THP:x1.x2.x3.x4.x5.x6) to the highest level (THP:x1). All these links are described in a link system and are distinguished by the attribute which each link has. A part of the link system is shown in Fig.4.

In Fig.4 the association from [a] is classified into Type 1, the association from THP.2.1.5.6.3.1 is classified into Type 2 and the association from [HEAD, @1], [NUM, SG], [DEF, -] is classified into Type 3.

Each node also holds not only links from the direct sources but the records of the lower association source (RLAS) to improve the accuracy of the rule selection. When the system selects a link to pull, both its weight and the RLAS are considered.

Our experimental system adopts a value calculated by a formula $lw \times zw$ for selecting rules. Here $lw$ represents the weight of its link and $zw$ represents the weights of the RLAS. However, in the case that RLAS does not exist, $zw \div 0 > 0$ is given to the formula in stead of the $zw$.

### 3.2 Application of translation rules

This proposed method has two stages, a rule acquisition stage and a rule application stage. These stages are carried out independently. Algorithm used in the rule application stage is also used in the rule acquisition stage. Therefore we mention the rule application algorithm first. We assume that rule acquisition and the link weight setting have been completed by the rule learning stage.

The system accepts an input sentence from the first word in order and selects rules by the association. The later words are checked the suitability to the activated (associated) rule. The suitability is examined as the association possibility for the activated rule by pulling the links using the input word as the association-key. During the examining process, the thesaurus data is used as the background information for the disambiguation. When it finds the suitability is lower than a parameter, the application of the activated rule is interrupted. Conditions which cause the interruption are shown below.

**Condition 1.** In case of the input word does not suit to activated rule. This includes a case in which the input word does not suit to the morphological information in the SEF.

**Condition 2-a.** Semantic representation produced for replacing the RM does not have a association link to the activated SRF.

**Condition 2-b.** The relation degree between the word concepts produced for replacing the RM and the activated SRF is lower than the parameter.

Here the relation degree between the word concept and the SRF is estimated at maximum when there is a direct link connecting the word concept with the SRF and at minimum when there is no path even through the top layer thesaurus link. When the system finds a path to the SRF through the thesaurus link and the layer of the thesaurus link is higher than the parameter, the rule application is interrupted because of the condition 2-b.

When the system accepts the input sentence, it copies the link state of the first word on working memory. The links on the working memory have variables showing one of three states (WAITING/ACTIVATED/FINISHED). When the links are copied on the working memory, their values are set to WAITING as a first value. When a link is used for the association, it becomes ACTIVATED and when the rule application is interrupted, it becomes FINISHED.

The system obtains a first rule by pulling a WAITING link with the maximum weight from a first word using the method 1 or the method 2 and continues the analysis of the later words using this rule.

For the RM, the system applies rules recursively and the links of those referred rule are copied on the working memory at each time. When an interruption is occurred, the system carries out the backtrack and obtains a rule by pulling a WAITING link with maximum weight and the present ACTIVATED link becomes FINISHED. When the system completes the rule application and still receives the inputs, it associates the SRF with upper rule by the method 3 and continues analysis using
this new rule.

By copying links on the working memory, the misapplication of the rules which might be occurred when a rule is used more than twice for a sentence is prevented because the states of links are copied for each case. Processes corresponding to the traditional analysis such as morphological analysis, syntactic analysis and semantic analysis are carried out simultaneously by the rule application.

An example of the rule application which contains the backtrack is shown in Fig.5. In the example, a word concept (PRO), a SRF and a SEF were associated by the method 2. Then upper rule was associated using the method 3. However the relation degree from LIKE.A (a word concept corresponding to "like") to the SRF was lower than a parameter and then the application was interrupted. The system carried out the backtrack and obtained a new rule. The new rule had a path to the SRF through the thesaurus links. For the part "a car," recursive rule application was carried out using "a" as an association-key under method 1. The new rule selected after the backtrack succeeded to be applied until the end of the sentence.

3.3 Acquisition of translation rules

The translation rules are learned from database consisting of the sentences and the corresponding semantic representations. In this stage, the system checks whether it is possible to translate a sentence to a semantic representation using the known rules. When it fails to translate the sentence, the system is forced to acquire new rules. The system constructs links to the new rule and increase the weight of the links by which the system succeeds the translation. During the rule acquisition stage, the weights of all links are weakened certain extent each time when system operates a pair of data. This algorithm enables it to reduce the weight of the non-efficient link gradually. The amount of the link thickening and weakening depends on the weight of the link at that time. Here we defined that all weights of the links are restricted in $0 \sim 1$. In the current system, the amount of the link thickening($aw$) is fixed (See Fig.6a) as the following:
\[ aw = t_1 \] (4)
and the amount of the link weakening($dw$) is determined by the next formula. (See Fig.6b)
\[ dw = ax^2 + bx + c \] (5)

When the maximum amount of the link thickening $t_2$, the minimum amount of the link weakening $t_3$ and the quadratic coefficient $a \leq t_2 - t_3$ is determined, $b$ and $c$ are calculated as the followings:
\[ b = t_3 - t_2 - a \] (6)
\[ c = t_2 \] (7)

When the system analyze the sample data, it works by means of the same algorithm in the rule application stage basically. The produced semantic representations are compared with the given semantic representation and when they differ, the system is forced to construct another semantic representation. When a rule application succeeds partially, the situation is saved. Finally, a case in which the system succeeds to construct a correct semantic representation, the links used to produce it are thickened. Cases in which the system can not succeed to construct the correct semantic representation, the system judges that it does not have enough rule for analyzing and infer partial rules for analyzing the values of case features which have not been replaced with the RM. For inferring partial rules, the system tries to extract parts of sentence corresponding to each value of case feature by the procedure below.

Procedure 1 It refers to the word concept dictionary and mark the word having the corresponding value to the concept feature in the target semantic representation.

Procedure 2 It looks for a rule which can be applied to the target data by replacing the part containing marks put in procedure 1 with the RM. When it has such rule, the parts are selected as the parts for partial rules.

Procedure 3 A case in which procedure 2 fails, it infers the parts containing the marks set in procedure 1 as the part for partial rules.

The marks set in procedure 1 have a structure like #n1-n2-n3-f, n1 and n3 shows a level of the analysis, f shows a case feature for which the system is analyzing. The rules used in the procedure 2 is selected by pulling the rule link from a SRF. The parts in the procedure 3 indicates a part divided by the marks containing different top level location number or the RM. A pair of the target semantic representation and an inferred part of the sentence is memorized as a new rule after being replaced the word concept with a RM. When the target semantic representation has case features, these procedures are carried out recursively and new rules are acquired. When the system finishes to replace the values of every case features, the value and the inferred part is replaced with the RM. A case in which a pair of the SRF and the SEF produced by this process is unknown as a rule, it is acquired as a new rule. The association links are set from the SRF of each partial rule to the upper SRF and the association links directly and through the thesaurus links are set from the word concept to the upper SRF. A case in which the first word of the SEF is not the RM, an association link is set from the word to the SEF.

This method of partial rule inference does not
guarantee to obtain appropriate rules. For instance, when a word exists more than twice in a sentence, the system might fail to analyze the relation between a word concept and a word. However, the rules acquired by the misapplication do not work efficiently in the later rule application process and are dismissed gradually.

An example of the rule application process is shown in Fig. 7. In Fig. 7, the system infers a SEF and acquires a partial rule because it fails to replace the value of the case feature (THEME) by the known rules. "computer games" is selected by the inference using the known rules and "computer" is selected as a part corresponding to the value of the SEF by recursive analysis. The parts corresponding to the value of the concept feature and the case feature are replaced with the RM and then the system acquires it as a new rule.

4 Experiments

4.1 The system for experiment

We constructed an experimental system which follows the proposed method. The first weight of the dictionary links and the thesaurus links which are set by referring the word concept dictionary were set to 1.0. When the system acquires new rules, the weight of the rule links and the association links were set to 0.8. From the result of the preparatory experiment, the amount of the link thickening (t1), the amount of the maximum link weakening (t2), and the amount of the minimum link weakening (t3) and the quadratic coefficient (a) were set as the followings:

\[ t_1 = 0.1, t_2 = 0.005, t_3 = 0.00001, a = 0.00499 \]  (8)

The interruption of the rule application occurred when the system could not find the path even though it used the thesaurus link located in deeper than level 4. The parameter zw used when the RLAS does not exist was set to 0.1.

4.2 Experiment

As the sample data, we used 751 English sentences from two elementary level English textbooks (all sentences) and put semantic representations on each sentence (A. Ota et al., 1991). We also prepared the word concept dictionary using the method we described. The system is given these data and carried out the rule application and the rule learning alternately for each pair of sentences and semantic representations. First the system tried to produce a semantic representation for an input sentence and the case in which the produced semantic representation accords to the given semantic representation, we estimated that it succeeds to produce a correct semantic representation.

Then the system tried to learn rules from the data. This procedure were carried out for all data continuously. Fig. 8 shows the transition of the correct translation rate (CTR). Here the CTR is defined as the following:

\[ CTR = \frac{\text{Number of correct translation}}{\text{Number of inputs}} \times 100 \]  (9)

4.3 Result and consideration

A symbol "A" in Fig. 8 shows the point at which the sample data switched from a text to another one. The sample data includes relative complicated sentences in the later parts of each book, because we used the textbooks as sample data. Therefore the CTR was expected to become worse in the later part even though the system had more rules than it had had in the former part. However as showing in Fig 8, after the point "A" the descending rate was decreased and the correct translation rate (CTR) started to increase after it received 600 pairs of data. This result shows that the learning of the rules worked efficiently in this experiment. The average CRT was 32.2% (before "A") and 48.2% (after "A"). The erroneous-translation rate (ERT)
gree among the semantic representations like we did among the word concepts and to allow the system to associate with the rules through the semantic representation which has high relation degree. However we need to examine this method carefully because this might prevent the appropriate interruption.

5 Conclusion

In this paper, we proposed an automatic acquisition method of the translation rules from sentences to semantic representations using inductive learning. The acquired rules are described using the weighted links, and an input sentence is transferred to a semantic representation directly using the acquired rules. Experiments confirmed the efficiency of our method are also shown. We are planning to carry out the experiments for more data and apply this method to designing the flexible language understanding system.

References

B.Sells. 1985. Lectures on Contemporary Syntactic Theories. CSLI(Stanford Univ.).