Adaptive Knowledge Base for Japanese-to-Braille translation

Satoshi Ono Takashi Yamasaki Shigeru Nakayama
Department of Information and Computer Science,
Faculty of Engineering, Kagoshima University
1-21-40 Korimoto, Kagoshima, 890-0065 Japan
{ono, sc101056, shignaka}@ics.kagoshima-u.ac.jp

Abstract

This paper proposes a method for generating adaptive knowledge base involving two knowledge representations – rule and case. Combining rules and cases makes it possible to solve problems accurately and quickly, and to acquire new cases from problem-solving results. The proposed method does not require manually adjustment of the thresholds and provides highly qualified solutions. This paper also proposes a Japanese-to-Braille translation system which uses the adaptive knowledge base as mentioned above. Experimental results have showed that the case acquisition can reduce errors, and that the threshold adjustment significantly reduces segmentation errors.
keywords: knowledge base, rule-based reasoning, case-based reasoning, threshold adjustment, Japanese-to-Braille translation

1 Introduction

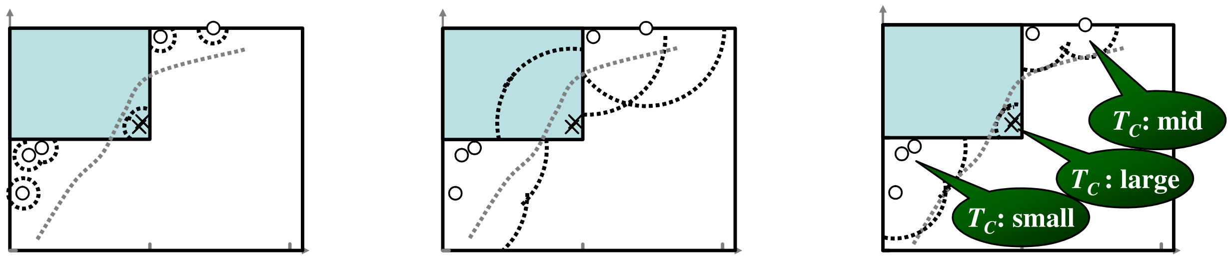
Rules and cases are valuable knowledge representations that mutually supplement drawbacks of each. While rules embody understanding that has been codified over the years by experts, cases contain the knowledge of a domain in a relatively unprocessed form. Rules are appropriate for representing general domain knowledge and cases are appropriate for representing the exceptions of that knowledge[1, 2, 3]. By combining rules and cases, it is possible to solve problems accurately and quickly, and to acquire new cases from problem-solving results[4, 5].

In previous methods using both rules and cases[4, 5, 6, 7], rule-based reasoning (RBR)[8] is performed first, and then case-based reasoning (CBR)[1, 2, 3] is performed. RBR solves a problem tentatively, but also chooses an exceptional case set. Retrieving the chosen cases, instead of all cases, reduces the processing time of case application, consequently a drawback of CBR. In the previous methods, only a common threshold of similarity is used to judge which exceptional cases should be applied. Although the threshold is manually adjusted, such adjusting is difficult because the appropriate threshold differs with each exceptional case.

In this paper, we propose an adaptive knowledge base (AKB), which is composed of a rule base, an indexed case base and a method for adjusting cases' thresholds. The proposed method eliminates the manual adjustment of the threshold and provides higher qualified solutions than the existing methods with the same knowledge base. In AKB, each case has a threshold in order to reuse exceptional cases that the existing methods cannot reuse because of the unified threshold. The thresholds are automatically adjusted after case acquisition. Providing a threshold for each case makes it easier to automatically adjust thresholds. Moreover, adjusting the threshold for each case increases the opportunity of reuse and decreases the risk of incorrect use.

This paper also proposes a Japanese-to-Braille translation system aJ2B which uses AKB as mentioned above. Japanese-to-Braille translation is a task involving two procedures – sentence segmentation and kanji-to-kana conversion. Accurate automatic translation is difficult due to the ambiguous, complicated rules peculiar to Japanese-to-Braille translation and Japanese language itself. A Japanese sentence is a string of characters concatenated without blanks, so spaces must be inserted between words to get a proper interpretation. Kanji must be converted to kana, because Braille characters expressing Japanese correspond only to kana.

Experimental results shows that the case acquisition can reduce errors, that the threshold adjustment significantly reduces segmentation errors, and that the proposed system reaches almost the same translation quality as the most popular software on the market called "Extra Ver.4.0."



(a) Boundary with large T_c. (b) Boundary with small T_c. (c) Boundary with adjusted T_c.

Figure 1: Boundaries depending on T_C.

2 Proposed adaptive knowledge base

2.1 Overview

In an attribute space, rules are represented as hyperrectangles and cases are represented as points whose coverage areas become Voronoi diagram. Although RBR quickly solves problems by applying a few rules, irregular or exceptional problems are difficult for RBR to solve. CBR can accurately solve these problems by applying many cases. Unfortunately, processing time increases proportionally to the number of cases used.

AKB utilizes rules to represent general knowledge, and cases to represent exceptional knowledge. This causes the knowledge base to maintain only negative cases, which have different operators from rules that they belong to. Positive cases, which have the same operator as rules that they belong to, are used only for learning process, not for solving problems. Consequently, the proposed knowledge base solves problems more accurately than RBR and more quickly than CBR because of less amount of knowledge than CBR.

2.2 Problem solving process

The fundamental idea of combining rules and cases is to apply the rules to a target problem to produce a draft solution; but if the target problem is judged to be compellingly similar to a known exceptional case of the rules, then the exceptional case is applied rather than the rules[4, 5, 6, 7]. The idea above is therefore realized through the following procedure:

Step 1: Use the rules to select an operator to apply.

Step 2: Search for exceptional cases that would contradict this choice of operator, stopping if and when a compelling case is found.

Step 3: If a compelling case was found, apply the operator it suggests, else proceed to apply the operator suggested by the rules.

2.3 Threshold adjustment

In the existing method using both rules and cases, the case coverage areas become hypersphere whose radiuses are decided by the unified threshold T_C in the attribute space. Large T_C prevents incorrect case application, but it also decreases the chances to reuse cases correctly are also decreased. In contrast, small T_C makes case coverage areas large and increases the changes to apply cases, but it increases the possibility of inadequate case application.

The AKB involves a threshold adjustment method for the hybridization of rules and cases. First, in the proposed knowledge base, each exceptional case has a threshold for deciding whether the case should be applied. This enables control of influence sphere at every exceptional case. Second, each threshold are adjusted automatically at a learning stage, while the existing method requires a developer to adjust the unified threshold manually. The AKB, therefore, reuses cases which are not reused efficiently in the existing method and prevents reusing cases which are reused incorrectly in the existing method.

The thresholds are adjusted one by one by the following procedure:

Step 1: Initialize a threshold T_{C_x} of an exceptional case C_x by $1.0 - \delta_T$.

Step 2: If $T_{C_x} \leq T_{C_x}^{(L)}$, terminate.

Step 3: For all learning data P_L in the root rule of C_x:

Step 3a: Apply C_x if $Sim(P, C_x) > T_{C_x}$.

Step 3b: If no error occurs, $T_{C_x} \leftarrow T_{C_x} - \Delta_T$ and go to Step 2.

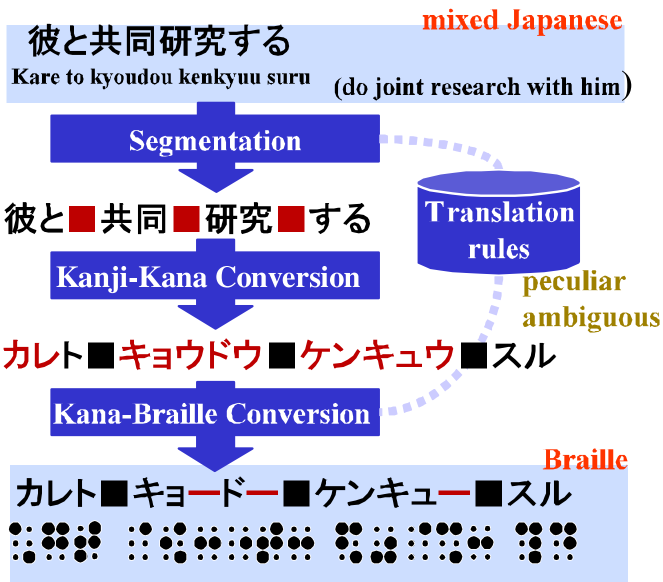


Figure 2: Japanese-to-Braille translation process.

Step 3c: If errors occur, $T_{C_x} \leftarrow T_{C_x} + \Delta_T$ and terminate.

3 Application to Japanese-to-Braille translation

The task for translating Japanese into Braille is defined as translation of mixed Japanese into Braille, and is done in two steps – sentence segmentation and kanji-to-kana conversion. A Japanese sentence is a string of characters concatenated without blanks, so spaces must be inserted between words to get a proper interpretation. Kanji must be converted to kana, because Braille characters expressing Japanese correspond only to kana.

The many translation rules that must be obeyed, are themselves ambiguous and full of exceptions[7, 9]. In sentence segmentation, for example, rules require that semantic and phonetic information be considered. In kanji-to-kana conversion, rules for distinguishing ordinary vowels are unclear, i.e., 'ア', 'イ', 'ウ', 'エ', and 'オ', and the symbol 'ー', denoting a long vowel, are used in writing, whereas a long vowel in regular Japanese is only used to express words of foreign origin or imitation sounds. It is thus very difficult to represent rules so that they are followed automatically by a computer.

Cases and rules segment sentences and revise strings in a draft. A rule consists of an operator (example: insert a space), conditions for applying the operator, and a priority score to resolve rule conflicts. A condition is stated by checking the value of an attribute. Attributes are obtained by morphological analysis, e.g., parts of speech, character type, a mora, and reading which is a pronunciation written in kana. All of those attributes are defined as symbolic attributes.

A case consists of an operator, a set of attributes of morphemes to which the operator of this case makes revision, and an identification number of a root rule, i.e., the rule for which the case is an exception. We define an object to which an operator of a rule or a case makes revision, i.e., a string or an interval between characters, as a spot.

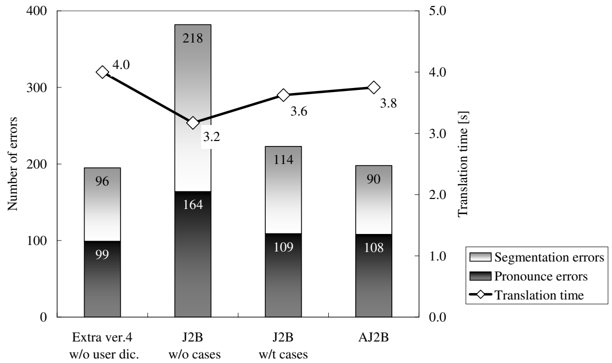
4 Experimental results

We compared aJ2B and the most popular Japanese-to-Braille translation software on the market — Extra Ver.4.0. A document "The Copyright Act" is used for evaluation. In the document, all sentences are divided into two groups — even sentences for training and odd ones for evaluation. The opposite experiments in which odd sentences are used for training and even sentences for evaluation are also performed. J2B[6, 7] is also tested for evaluating the effectiveness of threshold adjustment and case acquisition.

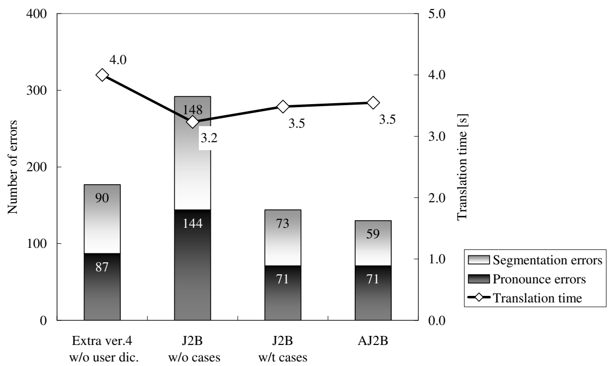
J2B and aJ2B utilizes a Japanese morphological analyzer called ChaSen with modified dictionary which does not involve extra foreign compound words. aJ2B parameters are decided as follows: $T^L_{C_R} = 0.9$, $T^L_{C_S} = 0.9$, $\Delta_T = 0.01$.

Figure 3 shows the number of errors and translation time of the proposed method and Extra in each experiment. Errors are categorized into pronounce and segmentation errors.

As shown Figure 3), case acquisition in J2B is significantly effective in decreasing errors. This is because "The Copyright Act" is a document that defines laws and that involves a lot of the same wording and quite similar wording. The important point to note is that the case acquisition in J2B can eliminate almost half of segmentation errors . This indicates that the flexibilty of cases is effective in representing exceptional segmentation knowledge of Japanese-to-Braille translation. In addition, the threshold adjustment can also reduce segmentation errors to almost 80% of J2B with cases. This indicates that the threshold adjustment can expand the flexibility of each case. In contrast, the threshold adjustment has no effectiveness on pronounce errors because cases for revising the pronounce



(a) The Copyright Act: Even senteneces are used for learning and odd sentences are used for test.



(b) The Copyright Act: Odd senteneces are used for learning and even sentences are used for test.

Figure 3: Experimental results

errors should be applied only to the same or quite similar wordings.

A glance at the graphs in Figure 3 also reveals that the threshold adjustment does not affect the translation time, although translation time increased by using cases.

5 Conclusions

In this paper, we proposed an adaptive knowledge base composed of rules and cases. The proposed knowledge base acquires exceptional cases from existing Braille documents and adjusts their threshold for judging whether the case should be applied or not. Experiments in Japanese-to-Braille translation problem showed that acquiring cases and adjusting their thresholds reduce errors.

Acknowledgements

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientist (B), No.16700412, 2004–2006.

References

[1] C. K. Riesbeck and R. C. Schank (1989), Inside Case-Based Reasoning. Hillsdale, New Jersey: Lawrence Erlbaum

[2] R. C. Schank, and A. Kassand, and C. K. Riesbeck (1994), Inside Case-Based Explanation. Hillsdale, New Jersey: Lawrence Erlbaum

[3] B. Bartsch-Spörl, M. Lenz, and A. Hubner (1999), Case-based reasoning - survey and future directions, Proceedings XPS-99, Springer Verlag, LNAI

[4] A. R. Golding and P. S. Rosenbloom (1991), Improving rule-based systems through case-based reasoning, in In Proceedings of the Ninth National Conference on Artificial Intelligence, pp. 22–27

[5] —— (1996), Improving accuracy by combining rule-based and case-based reasoning, Artificial Intelligence, vol. 87, pp. 215–254

[6] S. Ono, Y. Hamada, Y. Takagi, et al.(2000), Interactive japanese-to-braille translation using case-based knowledge on the web, in Proceedings of the Sixth Pacific Rim International Conference on Artificial Intelligence, pp. 638–646

[7] S. Ono, Y. Takagi, Y. Hamada, et al.(2003), Japanese-to-braille trnslation and error correction support using case-based knowledge, (in Japanese) Human Interface, vol. 5, no. 4, pp. 491–498

[8] F. Hayes-Roth, D. Waterman, and D. Lenat (1983), Building Expert Systems. Addison-Wesley Publishing Co., Reading

[9] E. Suzuki, S. Ono, T. Hiraoka, et al. (1997) Interactive japanese sentence segmentation system for translating japanese into braille, in Proceedings of the Natural Language Processing Pacific Rim Symposium, vol. 1, 1997, pp. 621–624