

# 事例・ルール間変換による知識編成方式と 日本語点字翻訳の分かち書き問題への適用

小野 智司<sup>†</sup> 鈴木 恵美子<sup>††</sup>  
宮下 和雄<sup>†††</sup> 西原 清一<sup>†</sup>

ルールベース推論を用いた問題解決システムにおける知識獲得の問題を解消する方式として、帰納的学習や事例ベース推論が広く研究されている。事例ベース推論は、事例の量を増やすことで様々な例外状況に対応することが可能となり、精度の高い問題解決を行うことができる反面、事例数に比例して実行速度が低下する。また、帰納的学習により知識数を削減することができるが、例外的な状況への対応が難しくなり精度が低下する。本論文では、事例とルール間の相互変換によって知識ベースの編成を行う方式を提案する。本方式は、問題に適した知識ベースを編成することで、精度と速度のバランスがとれた推論を行うことができる。また、逐次的な事例獲得が可能のため、対話型システムに適している。本方式を日本語点字翻訳における分かち書き問題に適用し、実験的にその有効性を検討した。

## Bidirectional Transformation of Knowledge between Cases and Rules with Applications to Sentence Segmentation in Japanese-Braille Translation

SATOSHI ONO,<sup>†</sup> EMIKO SUZUKI,<sup>††</sup> KAZUO MIYASHITA<sup>†††</sup>  
and SEIICHI NISHIHARA<sup>†</sup>

In case-based reasoning, a large case base copes with various exceptional situations; however, its computational cost necessary to apply knowledge increases in proportion to the number of cases. Inductive learning is known to reduce the amount of knowledge losing partly the case's faculty for exceptional situations. In this paper, we propose a new method which makes use of both cases and rules, where the knowledge base is optimized by bidirectional transformation between cases and rules to fit in with problem's property. Especially, our method is effective for interactive systems because it enables incremental learning preserving the balance between accuracy and efficiency of reasoning. The method is applied to the sentence segmentation problem in the Japanese-to-Braille translation, whose experimental simulations show the effectiveness of our method.

### 1. はじめに

ルールベース推論は、専門家の領域知識をルールで表現することにより、効率的に問題解決を行うことができるため、多くのエキスパートシステムで用いられてきた。しかし、人手でルールを生成する作業は膨大な労力が必要とされるだけでなく、すべての領域知識をルールで表現することが困難な問題も多く存在する。このため、専門家の経験など過去の事例からルールを

導出する帰納的学習や<sup>1)</sup>、事例を問題解決に直接利用する事例ベース推論 (Case-Based Reasoning: CBR) などが注目されている<sup>2),3)</sup>。

ルールと事例は相補的な関係にあることが指摘されている<sup>2)</sup>。一般に、ルールは因果関係が明確な知識表現である。ルールは高速に適用することが可能である反面、あらゆる知識をルールで表現することが困難であるという問題点を持つ。これに対して、事例は因果関係を不問のまま容易に知識獲得を行える反面、事例探索に時間がかかるという問題点を持つ。

ルールと事例のそれぞれの特徴を生かすため、この2つの知識表現を併用する方式も提案されている<sup>4)~8)</sup>。Anらは、CBRを行う前にルールを導出し、それらによって属性の重み付けや、ノイズの除去を行う方式を提案している<sup>4)</sup>。Goldingらは、ルールの例外的な知

<sup>†</sup> 筑波大学電子・情報工学系

Institute of Information Sciences and Electronics, University of Tsukuba

<sup>††</sup> 東京家政学院筑波女子大学短期大学部情報処理科

Tokyo Kasei Gakuin Tsukuba Women's University

<sup>†††</sup> 電子技術総合研究所

Electrotechnical Laboratory

識を事例で表現し、ルールベース推論を行った後に事例ベース推論を行う方式を提案している<sup>5)</sup>。Domingos は、事例ベース学習と帰納的学習を統合し、事例集合から最適な事例とルールの組合せを導出する RISE アルゴリズムを提案している<sup>6)</sup>。このような、CBR や帰納的学習を用いた問題解決方式により、効率的な知識ベースシステムの開発が可能である。

しかしながら、これらの方式を適用することが難しい問題もある。点字翻訳における分かち書き問題は、計算機上での自動化が非常に難しいとされている問題であり、領域知識が曖昧かつ不安定であるとされている。この問題は、対話的に処理を進めて逐次的に事例を獲得していく方式が有効である<sup>8)</sup>。このような問題に CBR を適用する場合、より多くの事例を保持することで精度の高い問題解決を行うことが可能になる反面、事例数に比例して実行速度が低下し、対話応答性を維持することが難しくなる。RISE を適用し、複数の事例から 1 つのルールを導出することで実行速度を改善することができるが、事例からルールを導出する一方向の変換のみを行うために、ルールが例外的な状況に対応することが難しくなり、精度が低下する場合がある。領域知識が不安定な問題において、対話的に高品質の問題解決を実現するには、精度と速度のバランスのとれた推論を行うように、知識を調整することが重要である。

本論文では、事例とルール間の双方向変換によって知識ベースの動的な編成を行う新たな方式を提案する。提案する方式は、コンパクトでしかも問題に適した知識ベースを編成することで、処理速度と精度のバランスのとれた推論を行うことができる。また、提案する方式を日本語点字翻訳における分かち書き問題へと適用し、実験的にその有効性を確認する。

以下、2 章では、知識の粒度について定義し、推論の性能と粒度の関係について述べる。3 章では、まず本方式の特徴および基本方針について述べ、次に提案する方式の知識表現、構成と処理手順、およびアルゴリズムについて説明する。4 章では、日本語点字翻訳における分かち書き問題について述べる。5 章では、分かち書き問題に応用した実験により、RISE、CBR ならびに本方式の性能を比較検討する。

## 2. 知識の粒度と推論

以下では、事例およびルールの両者を含めて知識と呼び、知識の有限集合を知識ベースとする。また、事例を構成する各属性を軸とする空間を事例空間と呼ぶ。属性の値は有限の離散値とする。

知識の抽象性を示す指標の 1 つとして事例空間  $U$  における知識  $K$  の粒度  $G(K, U)$  を  $\log_{|U|}|K|$  により定義する。ここで、 $|U|$  は  $U$  のサイズ、すなわち各属性における値の総数の積、 $|K|$  は  $K$  中の値組の総数である。知識  $K$  が事例の場合は  $|K| = 1$  なので事例の粒度はつねに 0、ルール  $R$  の粒度  $G(R, U)$  は  $0 < G(R, U) \leq 1$  となる。粒度は 0 に近いほど細かく、1 に近いほど粗い。

一般に、複数の事例を 1 つのルールに変換することは複数個の粒度 0 の知識を粗い粒度の知識 1 つにまとめることであり、知識のデータ量を削減することができる。しかし、粗粒度のルールでは例外事項の扱いが難しくなり、精度の改善が困難になる。これに対し、CBR のように事例を問題解決に直接利用すると例外的な状況にも対応することができる。しかし、様々な例外に対応するためには、大量の事例を保持する必要があるが、実行速度が低下してしまう。したがって、精度、速度ともに高い推論を行うには、一般的な領域知識は粒度の粗いルールとして、例外的な領域知識は事例または粒度の細かいルールとして保持することが重要である。

RISE における知識ベースの編成方式は、事例からルール、すなわち知識の粒度が粗くなる方向の変換のみを含んでいる。これにより速度を改善することができるが、領域知識が不安定な問題や訓練データを十分に用意することができない問題など、逐次的な事例獲得が有効な問題では、粗粒度のルールが例外的な状況に対応できなくなった場合に、細粒度の知識へ戻すことができない。また、RISE の知識表現では、知識の粒度調節が困難であり、問題に適した粒度の知識を導出することが難しい場合がある。

## 3. 提案する方式

### 3.1 提案する方式の基本方針および特徴

本論文では、事例とルール間の双方向の変換を許して知識ベースの編成を行う方式を提案する。本方式は、コンパクトでしかも問題に適した知識ベースを動的に編成することで、速度と精度のバランスのとれた推論を行うことができる。

提案する方式の基本方針および特徴を以下に示す。方針 1. 事例・ルール間で双方向変換を行うことにより知識ベースの再編成を行う。事例からルールを導出する変換（事例→ルール変換）は高速化を、ルールを事例へと還元する変換（ルール→事例変換）は精度の改善を目的とするものである。事例・ルール間で双方向変換を行うことで、適切な粒度を持つ知識を保持す

ることができ、精度と速度のバランスがとれた推論を行うことができる。

方針 2. ルールの条件部を調整するパラメータを導入することで、導出するルールの粒度を制御する。ルールの条件部において、第  $i$  属性で許される値の個数を制限するパラメータ  $V_i$  を導入する (3.2 節参照)。  $V_i$  を調整することにより、一般的な知識と例外的な知識の類別のレベルを調整し、問題に適した知識ベースを導出することができる。

方針 3. 問題解決用の知識ベースとそれを再編成するための保管事例ベースとを区別する。問題解決には小規模の知識ベースを用い、知識ベースの再編成時には大量の事例を用いる。これにより、コンパクトで高品質な知識ベースを編成することができる。

### 3.2 知識表現

本方式では、事例とルールの表現方式を統一する。これにより、事例とルール間の双方向の変換を容易に行うことができる。また、ルールの適用方法を一般的な exact-match 方式のかわりに best-match 方式とし、事例とルールの適用方法を統一する<sup>6)</sup>。これにより問題解決アルゴリズムを単純化することができる。

知識  $K$  は、既述したように事例・ルールの両方を表すが、次のように統一的に定義する。

$$K = (p, C)$$

$p, C$  はそれぞれ条件部、結論部である。  $K$  が事例の場合は  $p = (e_1, e_2, \dots, e_m)$  であり、属性  $i$  に関する属性値  $e_i$  の論理積である。  $m$  は属性の個数である。  $K$  がルールの場合は  $p = (A_1, A_2, \dots, A_m)$  となり、属性  $i$  に関する条件  $A_i$  の論理積である。ただし、  $A_i$  は‘属性  $i$  で許される値’の有限集合  $\{r_{i,1}, \dots, r_{i,j}\}$  または ‘\*’ ( don't care ) である。この集合のサイズ  $j$  は  $V_i$  以下である。ただし、  $V_i$  はルールの属性  $i$  に関する条件で許される値の個数を制限するためにあらかじめ与えておくパラメータであり、一般的な知識と例外的な知識との類別のレベル、すなわち両者の粒度の差を調整する働きを持つ。  $V_i$  が 1、すなわち RISE と同じ知識表現のときは、属性は 1 つの具体値または ‘\*’ のいずれかであり、 ‘\*’ の多い一般的な知識と、具体値のみからなる最細粒度の例外的知識とに分断される傾向がある。  $V_i$  を大きくすることで適正規模の粒度の知識を形成できるようになる。事例およびルールの結論部  $C$  はクラスであり、知識が適用された結果を示す。

### 3.3 構成と処理手順

本方式は知識の適用、獲得、変換の 3 つの機能、および知識ベース、事例アーカイブの 2 つの知識データ

ベースより構成される ( 図 1 )。知識ベースは推論に用いる知識、すなわち事例とルールからなるデータベースである。事例アーカイブはこれまでに獲得した全事例を保管するデータベースであり、知識ベースの再編成時に用いる。

本方式の処理手順を図 2 に示す。図 2 の訓練段階で事例-ルール変換により知識ベースの編成を行うが、事例のみからなる知識ベースを用いて問題解決を開始することもできる。また、図 2 の問題解決段階で追加的な事例獲得を行うことを逐次的な事例獲得と呼ぶ。

本方式は、次々と与えられる問題を順次解決し、その合間を利用して知識ベースを再編成することができる。与えられる個々の問題  $P$  は各属性の値の組であり、事実上事例の条件部  $p$  と同じ構造を持つ ( $P = (p_1, \dots, p_m)$ )。また、問題解決とは、知識ベース内の知識を適用することにより、問題を分類すべき

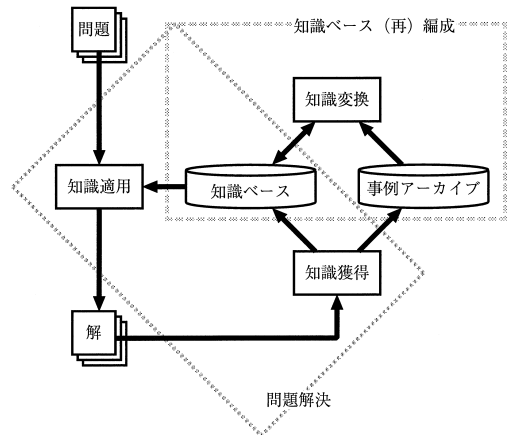


図 1 本方式の機能構成

Fig. 1 The functional structure of the proposed method.

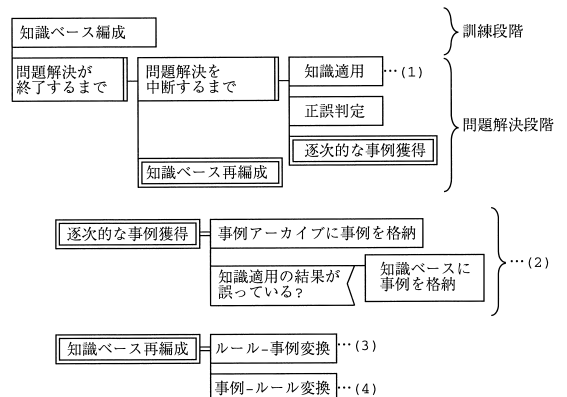


図 2 本方式による問題解決の流れ

Fig. 2 The problem solving process of the proposed method.

クラス  $C$  を決定する処理である。知識を適用した結果が誤っている場合は、 $P$  および正しいクラスを組にし、事例として知識ベースに追加する。知識ベースの再編成は、まず精度の低いルールを元の複数の事例へ還元し、次により多くの事例からより精度の高いルールを導出することで行う。以下、図 2 の (1)~(4) について説明する。

(1) 知識適用

対象とする問題と知識ベース内のすべての知識との距離を計算し、最も距離の近い知識を適用する。そのような知識が複数個存在する場合は、最も精度の高い知識を優先する。それでも一意に決まらない場合のタイブレークはランダムに行う。

問題  $P$  と知識  $K$  の距離  $\Delta(P, K)$  は、 $P$  と  $K$  の  $p$  とを比較することで次式のように計算する<sup>6)</sup>。

$$\Delta(P, K) = \sum_i \delta(i) \tag{1}$$

$\delta(i)$  は第  $i$  属性の距離で、 $K$  がルールの場合には次式で計算する。

$$\delta(i) = \begin{cases} VDM(x_i, A_i) & \text{if } A_i \neq * \\ 0 & \text{if } A_i = * \end{cases} \tag{2}$$

式 (2) の  $VDM$  (Value Distance Metric) は、属性値とクラスの共起頻度に基づいて、記号属性における属性値間の距離を計算する方法である<sup>3),6)</sup>。 $K$  が事例の場合は、 $\delta(i) = VDM(x_i, e_i)$  として計算する。なお、事例と知識間の距離も、 $P$  を事例の  $p$  に置き換えることで計算することができる。

次に、知識  $K$  の精度は次式を用いて計算する。

$$\frac{n_+ + 1}{n_+ + n_- + c} \tag{3}$$

$n_{+(-)}$  は事例アーカイブ内において知識  $K$  に含まれる正(負)事例数、 $c$  はクラスの個数である。正事例とは知識  $K$  と同じクラスを持つ事例であり、負事例とは知識  $K$  と異なるクラスを持つ事例のことである。クラス数は問題を定義することによって決まる。知識が事例を‘含む’とは、事例が持つ属性値が知識の条件部を満たすことである。式 (3) では、ごく少数の事例のみを含む知識の精度が高くなってしまふのを防ぐために、分母に  $c$  を、分子に 1 を加算してある<sup>6)</sup>。

(2) 逐次的な事例獲得

問題に知識を適用した後、問題と正しいクラスとを組にしたものを事例として事例アーカイブに追加する。知識の適用結果が誤っていた場合には、さらに知識ベースにも事例として追加する。知識ベースに格納された事例は、知識ベースの再編成が行われるまでは

事例として利用され、再編成が行われるときに、より有用な粒度の知識となる。

(3) ルール-事例変換

ルールの適用結果が誤りである割合が高くなり、ルールの精度が閾値を下回るようになった場合には、そのルールを元の複数個の事例へと戻す。

図 3 (a) にルール-事例変換の模式図を示す。‘+’ および ‘x’ は事例アーカイブ内の事例を示し、同じ記号で表される事例は同じクラスを持つものとする。閉ループは知識ベース内の知識を示している。状態 1 では、ルール  $R$  が 1 つ存在し、その精度は式 (3) より、 $(5 + 1)/(5 + 0 + 2) = 0.86$  である。 $R$  を問題に適用した結果が誤っており、状態 2 のように  $R$  とは異なるクラスを持つ事例が 2 つ獲得されたたすると、 $R$  の精度は  $(5 + 1)/(5 + 2 + 2) = 0.67$  となる。ここで、ルール-事例変換の閾値を 0.8 とすると、 $R$  の精度が閾値を下回るので、 $R$  を生成する際に用いた事例へと変換する(状態 3)。

ルール-事例変換は、次項 (4) に先だって、精度の低いルールを事例(粒度 0)へと戻す処理である。

(4) 事例-ルール変換

知識  $K$  の抽象化は、事例アーカイブ  $A$  内の最も距離が近い正事例を含むように  $K$  の条件部を  $A_i = A_i \cup \{e_i\}$  と変更することで行う<sup>6)</sup>。このとき、 $A_i$  に含まれる値の個数が  $V_i$  を超える場合は  $A_i = *$  とする。上記の処理を 1 ステップとする。

事例-ルール変換は、知識ベース内の全知識に対し

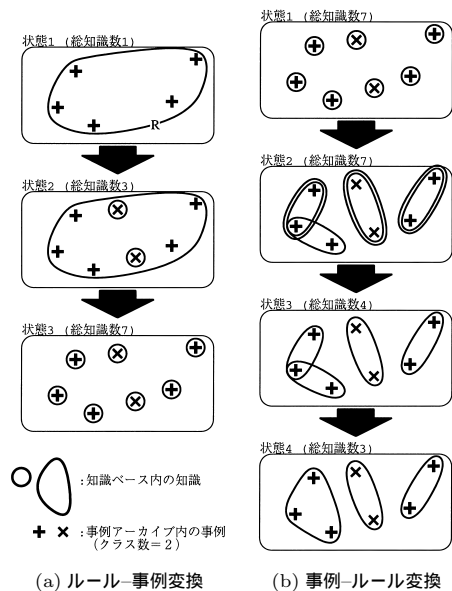


図 3 知識変換の模式図

Fig. 3 A diagram of knowledge conversion.

て1ステップ抽象化を施す処理を1ループとし、知識ベースの精度が向上しない状態が $l$ ループ続くまで行う。ここで $l$ は事例-ルール変換の終了を制御するパラメータである。 $l$ を大きくすると、より丁寧な事例-ルール変換が行われ、知識ベースの規模がコンパクトになる。知識ベース $K$ 内の知識 $K$ を1ステップ抽象化する際は、 $K$ を抽象化した知識 $K'$ に置き換えた知識ベース $K'$ の精度 $Acc(K', A)$ を計算し、 $Acc(K, A) \leq Acc(K', A)$ であれば、 $K$ を $K'$ に置き換える。このとき、 $K'$ と重複する知識、すなわち同一の条件部、結論部を持つ知識が $K$ 内に存在する場合は $K'$ を削除する。なお、知識ベースの精度 $Acc$ はleave-one-outクロスバリデーション法により計算する<sup>6)</sup>。

図3(b)に事例-ルール変換の模式図を示す。状態1は知識ベース内に事例のみ存在している状態であり、総知識数は7である。この状態から、事例-ルール変換を1ループ実行すると状態2のようになる。ここで、同じ事例を含む重複知識については1つだけ残し他を削除すると状態3のようになり、総知識数は4に減少する。状態3からさらに事例-ルール変換を1ループ行うと状態4のようになり、総知識数は3となる。

事例-ルール変換は粒度の細かい多数の知識から粒度の粗い少数の知識を導出するもので、速度の改善を図るものである。

### 3.4 本方式の位置づけ

本方式は、獲得した知識数に比例して実行効率が低下するというCBRの問題点を克服するために、事例と、より抽象的な知識表現であるルールを併用し、知識の抽象化・調整を行う能力を付加したものである。

また、本方式はRISEと比較すると、ルール-事例変換、パラメータ $V_i$ 、事例アーカイブの導入によって、より柔軟な知識の粒度の最適化を図っている。したがって、逐次的に事例獲得を行う問題で、RISEにおいて精度の改善が難しいような場合も、本方式は精度を改善することができるものと考えられる。

本方式はGoldingらの方式<sup>5)</sup>と組み合わせることにより、人手で生成されたルールベースと併用することができる。Goldingらの方式におけるCBRを本方式に置き換えることで、より高速かつ高精度な問題解決が可能である。

## 4. 点字翻訳における分かち書き問題への応用

ここでは、前章で提案した方式を、日本語点字翻訳の処理において最も難しいとされる分かち書き問題に適用する。日本語を点字に翻訳するには漢字をかなに

変換するとともに、点字の規則に従って分かち書きを行う必要がある。点字における分かち書きとは、基本的には文節ごとに区切りを挿入することである。しかし、一般の文節区切りで1つの文節と認定される単位についても、長い複合語などは、後に続く語の品詞、長さや発音の自然さに応じてより細かい分かち書きを行う必要がある<sup>9)</sup>。このように分かち書きは一般の文節区切りとは異なり、意味や発音までも考慮した複雑で曖昧な多くの点字に特有の規則に基づいているため、計算機上で自動的に分かち書きを行うことは非常に難しい<sup>9)</sup>。

分かち書き問題において、CBRに基づく方式が提案されている<sup>8)</sup>。これは、高精度の分かち書きを行うことが可能な方式であるが、事例数の増加による速度の低下が問題となる。提案する方式を分かち書き問題に適用することにより、高精度を維持しつつ実用的な時間内で分かち書きを行うことができると考える。

本論文では、分かち書き問題を、「対象文書における各形態素間の区切り方を、前後の形態素情報をもとに決定する問題」とする。また、1形態素間を1問題とし、前後各2形態素の情報を1問題が持つ属性とする。区切り方、すなわちクラスは3種類{区切らない、区切る(1個または2個の空白を挿入する。)}とする<sup>9)</sup>。対象文書の解析には、音声合成用の形態素解析を用いる。

## 5. 評価実験

### 5.1 実験方法

本方式を分かち書き問題に適用し、日本語文書(情報処理の入門テキスト、全80,490問題、596段落)を用いて実験を行い、本方式の性能評価を行った。

1回の実験では、まず訓練段階において、10,000個(1~71段落)の問題から2,500個の訓練用問題をランダムに選択して、以下の処理を行った。すなわち、訓練用問題から距離の計算に必要な属性とクラスの共起頻度情報を獲得したのち、事例-ルール変換により知識ベースを編成した。次に問題解決は、残りの70,490個の問題(71~596段落)に対して行った。実際にユーザが使用する状況を想定し、段落単位で知識適用、正誤判定および事例獲得を繰り返すものとした。正誤判定は実際に点字翻訳ボランティアが翻訳したものを正解として行った。事例-ルール変換の実行時間制御パラメータ $l$ は2に設定した。

距離は訓練段階で利用した2,500問題に含まれる属性情報のみを用いて計算し、問題解決段階で逐次的に獲得した事例の属性情報は利用しないものとした。

表 1 本実験で用いた分かち書き問題における属性とその値の個数  
Table 1 The attributes and the number of values per attribute.

属性 id	属性名	値の個数	属性 id	属性名	値の個数
1	字種	796	4	拍数	11
2	品詞	206	5	アクセント	8
3	字面の長さ	13	6	文節情報	4

表 2 パラメータ  $V_i$  に関する実験結果  
Table 2 Comparison as for parameter  $V_i$ .

$V_{1-2}, V_{3-6}$	正解率 (%)	知識適用時間 (msec)	$V_{1-2}, V_{3-6}$	正解率 (%)	知識適用時間 (msec)
1,1	93.2 (0.4)	4.0 (1.0)	3,3	93.8 (0.7)	10.5 (0.8)
2,1	94.1 (0.3)	7.3 (0.5)	4,1	93.6 (0.4)	9.3 (0.6)
2,2	94.5 (0.3)	8.9 (0.6)	4,2	93.7 (0.6)	10.3 (0.2)
3,1	93.7 (0.4)	9.0 (0.5)	4,3	93.2 (1.1)	10.8 (0.4)
3,2	94.2 (0.5)	9.8 (0.4)	4,4	93.1 (1.1)	10.9 (0.4)

括弧内の数値は標準偏差を表す。

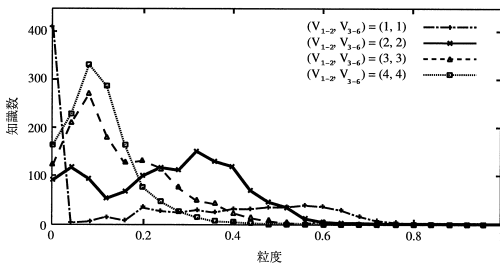


図 4  $V_i$  と粒度ごとの知識数  
Fig. 4 Experimental results on granularity.

性能評価は、正しく推論を行った割合 (正解率: %) および 1 問題あたりの平均推論時間 (知識適用時間: msec) の 2 項目について行った。

### 5.2 本方式の特性に関する実験と評価

#### 5.2.1 パラメータ $V_i$ の調整による性能変化

パラメータ  $V_i$  は導出するルールの粒度を調整する働きを持つ。この値による性能変化を評価するための実験を行った。問題解決段階における事例獲得および知識ベースの再編成は行わないものとした。また、表 1 に示すように、第 1 属性 (字種) と第 2 属性 (品詞) の属性値の個数が他の属性に比較して多いため、 $V_1$  と  $V_2$  を同一の値  $V_{1-2}$  に設定し、同様に  $V_3 \sim V_6$  も同一の値  $V_{3-6}$  に設定した。各設定について 5 回実験を行ったときの正解率および知識適用時間の平均を表 2 に示す。表 2 より、 $(V_{1-2}, V_{3-6}) = (1, 1)$  のとき、すなわち RISE と同様の知識変換を行う場合に最も高速であるが若干正解率が低く、 $(V_{1-2}, V_{3-6}) = (2, 1), (2, 2)$  および  $(3, 2)$  のときに正解率が高いことが分かる。

さらに、 $(V_{1-2}, V_{3-6}) = (1, 1), (2, 2), (3, 3)$  および  $(4, 4)$  の場合における、知識ベース内の知識の粒度

表 3 比較する方式の一覧  
Table 3 The methods compared with our method.

	方式 1	方式 2	方式 3	本方式	方式 4	方式 5
	逐次的な事例獲得	—	—	—	—	—
事例-ルール変換	—	—	—	—	—	—
ルール-事例変換	—	—	—	—	—	—

—: この機能を含む。 —: 事例を知識ベースに格納せずに、事例アーカイブにのみ格納する。

表 4 知識変換による性能変化  
Table 4 Comparison on knowledge transference methods.

方式	正解率 (%)	知識適用時間 (msec)
方式 1	94.5 (0.3)	8.9 (0.6)
方式 2	96.4 (0.1)	16.0 (0.4)
方式 3	96.8 (0.1)	12.2 (0.6)
本方式	97.0 (0.1)	14.8 (0.3)
方式 4	94.3 (0.1)	7.2 (0.3)
方式 5	95.5 (0.2)	10.8 (0.4)

括弧内の数値は標準偏差を表す。

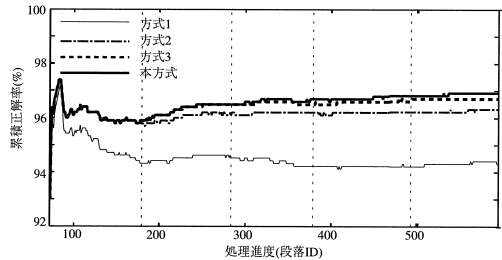


図 5 知識変換による正解率の変化  
Fig. 5 Experimental results on accuracy.

を調べた。各粒度における知識数の分布を図 4 に示す。図 4 より、 $(V_{1-2}, V_{3-6}) = (1, 1)$  の場合、一般的な知識と例外的な知識との類別が過剰に行われ、粒度が 0.04 ~ 0.16 のルールがほとんど存在しないことが分かる。これに対し、 $V_i$  が 2 以上の場合は、粒度が 0.04 ~ 0.16 のルールを導出できていることが分かる。このように、本方式はパラメータ  $V_i$  を調整することで、導出する知識の粒度を調整することができる。

なお、以後の実験では  $(V_{1-2}, V_{3-6}) = (2, 2)$  とした。

#### 5.2.2 知識ベース再編成の効果

本方式における知識変換の効果の評価するために実験を行い、表 3 の前半における、方式 1, 2, 3 および本方式の 4 つの方式を比較した。表 3 に各方式が 3.3 節の機能のうちのどれを採用しているかを示した。各方式は、訓練段階で事例-ルール変換により知識ベースを編成するものとした。方式 3 および本方式は、あらかじめ設定した段落 (176, 281, 386, 491 段落) が終了した時点で、ユーザが作業を中断するものと想定し、知識ベースの再編成を行った。ルール-事例変換

時のルール削除の閾値は 0.9 に設定した．訓練用問題を変えて，各方式について 5 回実験を行ったときの正解率および知識適用時間の平均を表 4 の前半に示す．また，そのうちの 1 回の実験における正解率の変化を図 5 に示す．図 5 において，処理進度とは，処理が終了した段落の番号を表し，累積正解率とは，その時点までの正解率を表す．垂直方向の破線は，この時点で知識ベースの再編成が行われたことを示す．

表 4 および図 5 より，方式 1 が他の 3 つの方式よりも正解率が低いことから，逐次的な事例獲得により正解率が向上することが分かる．また，方式 2 に比べて方式 3 の正解率が高いことから，問題解決段階における事例-ルール変換を行うことで，新たに獲得した知識の粒度が高くなり，推論に利用されやすくなったと考えられる．一方，図 5 の 400 段落以降において，本方式の正解率は方式 3 を上回りかつその差が徐々に大きくなっていくことが分かり，ルール-事例変換によって正解率を改善できていることが分かる．

### 5.2.3 事例アーカイブの必要性

事例アーカイブは，知識および知識ベースの精度計算を行う処理と，事例-ルール変換において知識を 1 ステップ抽象化する処理において参照される．事例アーカイブの働きを確認するために，表 3 における，方式 1, 4, 5 の 3 つの方式を比較した．各方式を 5 つの異なる訓練用問題集合について実行したときの正解率および知識適用時間の平均を表 4 の後半に示す．

方式 1 に比べて方式 4 は若干正解率が低く，知識ベースおよび事例アーカイブに事例を格納せずに知識変換を行っても正解率が向上しないことが分かる．一方，問題解決段階で事例アーカイブに事例を格納する方式 5 の正解率は方式 1 および 4 よりも高い．これは，事例アーカイブ内の事例が増加することで，知識および知識ベースの精度計算，ならびに知識の抽象化をより適切に行うことができるためと考えられる．

### 5.3 CBR および RISE との比較実験

CBR, RISE および本方式の性能を比較する実験を行った．CBR は，各段落で正誤判定を行った後，知識の適用結果を事例ベースに格納するものとした．各方式を 5 つの異なる訓練用問題集合について実行したときの正解率および知識適用時間の平均を表 5 に示す．また，そのうちの 1 回の実験における正解率の変化，知識適用時間と知識ベースの再編成に必要な時間の変化，および知識数の変化をそれぞれ図 6，図 7，図 8

表 4 において，方式 3 と本方式の正解率が正規分布をなすと仮定すると，両者の平均の差は対応あり  $t$  検定によって有意水準 0.05 で有意差があるといえる．

表 5 CBR および RISE との比較

Table 5 Comparison on proposed method, CBR and RISE.

方式	正解率 (%)	知識適用時間 (msec)
本方式	97.0 (0.1)	14.8 (0.3)
CBR	97.5 (0.1)	202.4 (0.6)
RISE	94.0 (0.1)	6.8 (0.5)

括弧内の数値は標準偏差を表す．

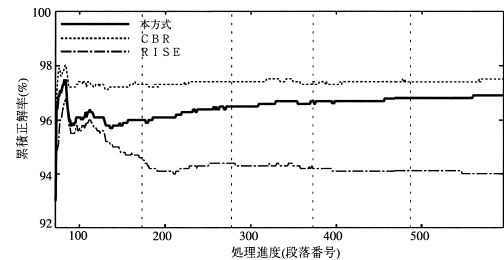


図 6 CBR および RISE との比較 (正解率)

Fig. 6 Experimental results on accuracy.

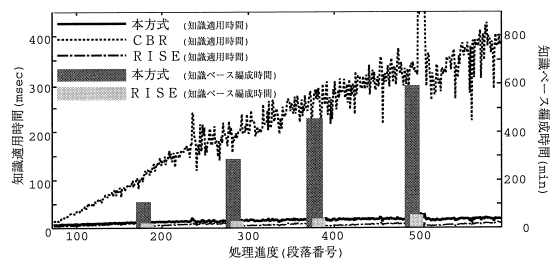


図 7 CBR および RISE との比較 (知識適用時間)

Fig. 7 Experimental results on knowledge applying time.

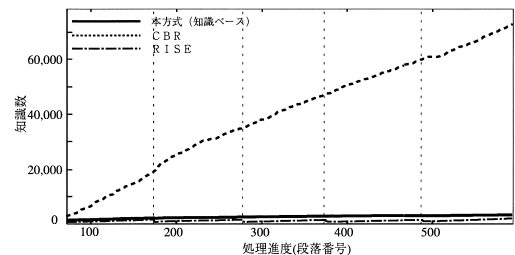


図 8 CBR および RISE との比較 (知識数)

Fig. 8 Experimental results on the numbers of knowledges.

に示す．図 6，図 8 における垂直方向の破線の時点で，RISE および本方式は知識ベースの再編成を行った．

表 5 より，本方式を CBR および RISE と比較すると，最終的に本方式が RISE の正解率を 3.0% 上回り，CBR の正解率を 0.5% 下回っている．また，図 6 より，逐次的に事例を獲得しているにもかかわらず，CBR は 150 段落以降でほぼ一定の正解率であり，RISE は

徐々に正解率が下降するのに対し、本方式は 150 段落以降も徐々に正解率を改善できている。

図 7 および図 8 より、CBR では処理が進んで事例数が増加するにつれ、知識適用時間が増大しているのに対して、RISE および本方式はほぼ一定の時間で知識を適用することが分かる。最終段落近くでは、本方式が CBR の 20 分の 1 以下の時間で処理できている。

図 7 より RISE および本方式は、処理が進むにつれて、知識ベースの再編成を行う時間が増加することが分かる。また、RISE と比較すると、当然ではあるが、本方式は知識ベース編成のためにより多くの時間が必要となる。

#### 5.4 考 察

##### (1) RISE との比較

本方式は RISE よりも知識適用に若干時間がかかるが、精度の面では RISE を上回る。表 5 より、1 段落あたりの平均誤り箇所数は CBR が 3.0 箇所、本方式が 3.5 箇所であるのに対し、RISE は 7.1 箇所であり、CBR や本方式の 2 倍以上の箇所知識適用の結果が誤っていた。よって、ユーザは本方式を利用することにより、RISE を利用した場合に比べて半分の箇所を修正すればよいことになる。本方式が RISE よりも精度が高かったのは、本方式が採用している 3 つの方針 (3.1 節) が分かち書き問題に対して有効であったためと考えられる。

本方式はパラメータ  $V_i$  の導入により、問題の性質に応じた柔軟な知識表現が可能である。図 4 より、RISE の知識表現 (本方式で  $V_i = 1$  とした場合と同様) では、一般的な知識と例外的な知識の類別が過剰に行われ、両者の間の粒度を持つ知識が導出されていないことが分かる。これに対し本方式は、 $V_i = 2$  と設定することで知識の分布が双耳峰形となっており、一般的な知識と例外的な知識との類別がバランス良く行われていることが分かる。

また、本方式は RISE と同様の事例-ルール変換に加えてルール-事例変換も行うため、逐次的な事例獲得が行われる問題において、より精度を改善することができる。図 5 ではルール-事例変換の効果がわずかであったが、分野の異なる複数の本を点字翻訳する場合のように領域知識が変化する場合や、ばらつきや偏りのある事例集合から知識ベースを導出する場合はルール-事例変換の効果がより大きく現れると考える。

さらに、本方式は事例アーカイブの導入により、逐次的な事例獲得が重要な問題において、RISE よりも多くの事例を参照して知識ベースの再編成を行える。

##### (2) CBR との比較

本方式と CBR を比較すると、本方式は CBR よりも若干精度が低いが、実行速度の点では本方式が CBR を大きく上回る。図 7 より、最終段落近くにおける 1 段落あたりの平均処理時間を比較すると、CBR が 47.0 秒、本方式は 1.8 秒となる。対話型点字翻訳システムでは、ユーザが操作を行ってから 3 秒以内にシステムから反応があることが望ましいとされており、CBR は対話処理を実現することが難しく、本方式は対話応答性を十分に確保しつつ高精度で処理を行うことができるといえる。

また、本方式を、問題解決を行うクライアントと知識ベースの再編成を行うサーバとに分けて実装することにより、クライアントは、図 8 に示すように CBR と比較して非常にコンパクトな知識ベースのみを保持すればよく、記憶コストの点でも優れている。

##### (3) 知識ベースの再編成に必要な時間

本方式は事例アーカイブ内の事例数に比例して知識ベースの再編成、特に事例-ルール変換に要する時間が増大する。しかし、知識ベースの再編成はユーザが問題解決を行っていない時間を利用して行うため、対話応答性に影響がないと考えられる。一般的に、1 冊の本を点字翻訳するのに必要な時間は短くても 2、3 週間といわれており、本方式はユーザが作業を行っていない膨大な時間を利用して知識ベースを再編成することができる。また、本方式における事例-ルール変換では、各ルールの抽象化を段階的に行うので、任意の時点で事例-ルール変換を終了または中断/再開することができる。1 つのルールを 1 ステップ抽象化するのに必要な時間はたかだか 0.8 秒程度であり、ユーザとの対話時における入力待ちなどの空き時間を利用して再編成を遂行することができる。さらに、既述のようにクライアントサーバシステムとして実装することにより、計算能力の高いサーバに知識ベースの編成を行わせることもできる。

## 6. おわりに

事例とルールから構成される知識ベースを用いて推論を行い、事例-ルール間の双方向変換により知識ベースの編成を行う方式を提案した。提案する方式は、対象とする問題の性質に知識ベースを適応させ、精度と速度のバランスがとれた問題解決を行うことができる。また、本方式を点字翻訳における分かち書き問題に適用し、この問題における本方式の有効性を確認した。分かち書き問題は属性値の総数が多い、例外的な知識が多数存在する、事前に適切な訓練データを準備する



ことが困難である，問題解決の合間に空き時間が存在するといった特徴を持つ．

本方式は，逐次的な学習が有効である問題において，高精度を維持しつつ実行効率を改善できる．よって，対話型システムのように，知識を逐次的に獲得するシステム，かつ処理時間に制約があるシステムにおいて有効であると考えられる．今後，パラメータの自動調整，知識変換の高速化，および本方式の有効範囲の明確化を行う予定である．

謝辞 形態素解析プログラムをご提供いただいた日本電信電話(株)サイバースペース研究所殿，正解データ作成にご協力いただいたボランティアの皆様，本研究を進めるにあたって有益なご助言をいただいた本学工学研究科水野一徳氏に深く感謝いたします．

### 参 考 文 献

- 1) 仁木和久，石崎 俊：概念の帰納的学習，人工知能学会誌，Vol.3, No.6, pp.695-703 (1988).
- 2) 小林重信：事例ベース推論の現状と展望，人工知能学会誌，Vol.7, No.4, pp.559-566 (1992).
- 3) Cost, S. and Salzberg, S.: A weighted nearest neighbor algorithm for learning with symbolic features, *Machine Learning*, Vol.10, pp.57-78 (1993).
- 4) An, A., Cercone, N. and Chan, C.: Integrating Rule Induction and Case-Based Reasoning to Enhance Problem Solving, *ICCB*, pp.499-508 (1997).
- 5) Golding, A.R. and Rosenbloom, P.S.: Improving accuracy by combining rule-based and case-based reasoning, *Artificial Intelligence*, Vol.87, pp.215-254 (1996).
- 6) Domingos, P.: Unifying Instance-Based and Rule-Based Induction, *Machine Learning*, Vol.24, pp.141-168 (1996).
- 7) 福井哲也：日本語自動点訳ソフト 4 種の精度の比較，第 2 回視覚障害リハビリテーション研究発表大会論文集，pp.114-117 (1993).
- 8) 高木喜次，小野智司，宮下和雄，西原清一：表層解析に基づく点字用日本語分かち書きへの事例ベースの適用，情報処理学会研究報告，99-NL-129 (1999).
- 9) 鈴木恵美子，小野智司，狩野 均：点字翻訳ボランティアのための対話型分かち書き支援システム，自然言語処理，Vol.5, No.4, pp.95-110 (1998).
- 10) 小野智司，鈴木恵美子，宮下和雄，西原清一：点字翻訳問題を対象とした事例とルールの併用による適応型問題解決について，人工知能学会研究会資料，SIG-KBS-9903-3, pp.13-18 (2000).
- 11) Ono, S., Hamada, Y., Takagi, Y., Nishihara, S. and Mizuno, K.: Interactive Japanese-to-

Braille Translation Using Case-Based Knowledge on the Web, *PRICAI*, pp.638-646 (2000).

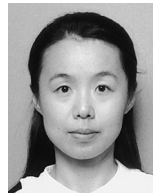
(平成 11 年 9 月 3 日受付)

(平成 12 年 9 月 7 日採録)



小野 智司(学生会員)

1997 年筑波大学第三学群情報学類卒業．1999 年同大学院理工学研究科修士課程修了．現在，同大学院工学研究科博士課程在学中．知識処理の研究に従事．



鈴木恵美子(正会員)

1981 年筑波大学第三学群情報学類卒業．1983 年同大学院理工学研究科修士課程修了．同年，日本アイ・ビー・エム(株)入社．同社東京基礎研究所において日本語文書校正支援システム・日英機械翻訳システムに関する研究に従事．1990 年より東京家政学院筑波短期大学勤務．現在，同筑波女子大学短期大学部情報処理科助教授．工学博士．電子情報通信学会，科学教育学会，計量国語学会各会員．



宮下 和雄(正会員)

1983 年東京大学工学部精密機械工学科卒業．1985 年同大学院工学系研究科修了．同年松下電器産業(株)入社．1990~1992 年カーネギーメロン大学ロボティクス研究所客員研究員．1995 年通産省工業技術院電子技術総合研究所に入所．現在，主任研究官．工学博士(大阪大学)．1999 年 1 月より筑波大学連携大学院助教授併任．分散協調問題解決，事例ベース学習，知的生産システム等に関する研究に従事．AAAI, IEEE CS 各会員．



西原 清一(正会員)

1968 年京都大学工学部数理工学科卒業．同年，同大学大型計算機センター助手．1975 年より筑波大学電子・情報工学系．現在，同教授．工学博士．1982~1983 年ヴァージニア工科大学，1998 年 IASA．グラフィックスと CAD，組合せ探索アルゴリズム，知識処理，制約充足問題，複雑系の研究に従事．著書に「データ構造」(オーム社)等．1975 年情報処理学会論文賞．電子情報通信学会，人工知能学会，ACM，IEEE 各会員．