

2ZA-06 自然言語に近い表現を可能にする CCR による知識共有*

坂本 栄里奈[†] 友部 博教[†] 木本 協司[‡] 伊庭 斉志[†] 石塚 満[†]

[†]東京大学工学部電子情報工学科

[‡]ミレーヌコーポレーション

1. はじめに

情報の中で自然言語で表された情報は非常に多く、これを知識として共有できるようにすることが重要になってきている。知識を述語論理で表せばコンピュータでの処理、共有は容易になるが、述語論理表現は例えば何を述語記号にして表現するかなどで任意性があり、一つの自然言語文が一意的な論理表現にはならないという問題がある。これにより、多数の人が記述する知識を共有するには必ずしも適していない。人間にわかりやすい自然言語表現に近く、対応がつけやすく、かつ推論機能も持つ表現形式が必要になると思われる。

我々はその手法のひとつとして、CCR (Concept Chemical Representation) を用いることを考えている。これは、(有) ミレーヌコーポレーションの木本協司が考案した「概念化学プログラミング」[2]の考えをもとにした知識表現法である。「概念化学プログラミング」は、発明・発見の概念過程は化学分野の架橋反応に類似しているとの着想から考案されたものである。CCR を用いて自然言語で表された知識を表現し、推論加工することによって、より高度の利用を可能にすることを目標とする。

将来的には WEB 上の文章なども扱うことを考えているのだが、そういったものは書き手によって書き方が異なり、現状では統一的な知識として書きなおすのが難しいと思われる。そこで今回は、一般の自然言語情報よりも比較的書式が統一されており、正しい知識が記述されている電子情報通信ハンドブック[3]上の知識を CCR に書きなおして、各種の利用ができるかについて検証を行った。

2. CCR の知識表現法

CCR では、知識は次のような概念関係式で表される。

「ソクラテスは人間である」

⇒ ソクラテス [bg] 人間

以下、CCR による知識表現の概要について述べる。

概念関係式

CCR では、以下のように2つの概念要素を概念関係で結んだ概念関係式で知識を表現する。

概念要素 [概念関係 / 真理値] 概念要素

真理値は0 (偽) から1 (真) までの連続値を取り、省略時は1とする。

概念関係

2つの概念要素を結びつける概念関係には、次のようなものがある。

[eq] 恒等関係 [bg] [ct] 帰属関係

[be] 連結関係 [cp] [hv] 構成関係

[bg]と[ct]などは逆行の関係にあり、これを用いることにより概念要素の並べ替えができる。

a [bg] b ⇒ b [ct] a

概念要素のリスト

概念要素は次のようにリストを取ることができる。

a((b₁,b₂)(c₁,c₂)…)

ここでリスト中のb₁・c₁はラベルを表し、b₂・c₂はそれを伴う語を表す。

例えば、「太郎は早く家を出る」という文はリストを用いて次のように表される。

太郎 [do] 出る((を 家) (adv 早く))

複雑な文の表現

複雑な文は変数(大文字で表す)を用いて分解して表現する。また、名詞の修飾は「@」を用いて表す。

例えば「妹は英語を勉強する学生です」という文は、

*Knowledge Sharing based on CCR allowing Representation close to Natural Language

Erina Sakamoto[†], Hironori Tomobe[†], Kyouji Kimoto[‡], Hitoshi Iba[†], Mitsuru Ishizuka[†]

Department of Information and Communication Engineering, School of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

妹 [bg] 学生@A

A [eq] 勉強する((を 英語))

と表す。

論理記号

論理記号として<and> (積)、<or> (和)、<if>~<then> ~ (条件文) <some> (∃)、<all> (∀) も用意されている。なお、否定は真理値を0とすることにより表現する。

<if>~<then>~とXなどの変数を用いることにより、以下の例のようにルールを記述することができる。

<if> (X [bg] 鳥) <and> (X [hv] 足@長い)

<then> X [bg] ダチョウ

3. 推論法

CCR では、概念縮合という推論法を用いる。

これは、複数の概念関係を意味的に縮合させて一つの概念関係を作り、これによって新たな知識を得る方法である。[x]と[y]を縮合させて[z]になる場合、

$a[x]b, b[y]c \Rightarrow a[z]c$

となる。この他にも、入力されたルールも推論に用いられる。

4. CCR によるハンドブック知識の

記述と利用

ハンドブック[3]のデータベース分野の記述を CCR によって表した(この自然言語からの CCR への変換は、将来は理解に基づく自動化も考えているが、当面は人手で行う)。CCR への変換での注意すべき点、課題には以下のような事項がある。

- ・各文を書き換える時に用語を補う。
- ・各文について用語との関係を考え、直接には書かれていない知識を補う。主語についても、主語のない文では、適切な主語を補う。
- ・文同士のつながりを CCR でどう表現するか。例えば、「～である。よって…である。」の「よって」はどう表現するか(論理表現を用いても共通の課題である)。

CCR では入力した質問(クエリー)に対して後ろ向きに演繹推論を行う機能と共に、キーワードを入力すると前向き推論の連鎖により関連する知識を出力する機能を備えている。これにより、ユーザの思考を補助することがで

きる。

データベース分野の記述を表した CCR 知識ベースに対して、「銀行システム」というキーワードを入力したときに、前向き推論により出力された関連知識の例を以下に示す。

(トランザクションシステム [ct] 銀行システム)

(トランザクションシステム [do] 反映する

((を 変更@データ)(に システム)(adv 早く)))

(銀行システム [do] 反映する

((を 変更@データ)(に システム)(adv 早く)))

(トランザクション [eq/0] プログラム@一般)

(トランザクション@各 [do] 満足する

((を 規則@特定)(において プロトコル形)))

(プロトコル形 [ct] 二相施錠方式)

(プロトコル形 [ct] 時刻印方式)

(二相施錠方式 [do] 使われる((adv 最も)

(の中では 実現方式@直列可能性)))

5. CCR の意義

論理では述語記号は任意に設定しうるが、CCR では概念関係(論理での述語記号にあたる)を有限個に制限し、概念要素(論理での引数にあたる)の数を2個に限定している。よって CCR は論理の完全性を継承し、また、人間が自然言語文章を CCR の形式に変換する際、ある程度機械的に行えるといえる。これにより、複数のユーザによる知識共有の表現形式として適していると考えている。今後、この効用を規模を拡大して検証すると共に、発想支援等の機能も組み込んでいきたい。

参考文献

- [1] 友部博教、木本協司、伊庭斉志、石塚満：自然言語に近い知識表現を可能にする概念化学プログラミングの実現、情報処理学会全国大会、No.1N-10(1999.3)
- [2] 木本 協司：「研究開発における発想とその支援ツール 概念化学プログラミング—構造化中間言語—」、研究開発マネジメント、1998年9月
- [3] 電子情報通信学会(編)：エンサイクロペディア 電子情報通信ハンドブック、オーム社、1998年