

自然言語と論理をつなぐ知識表現 KRNL の UNL による拡張

Extension of KRNL, which joins Natural Language and Predicate Logic, by UNL

谷口 智哉*¹
Tomoya Taniguchi

友部 博教*²
Hironori Tomobe

松尾 豊*³
Yutaka Matsuo

石塚 満*¹
Mitsuru Ishizuka

*¹ 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
*² 名古屋大学情報科学研究科
Graduate School of Information Science, Nagoya Univ.
*³ 産業技術総合研究所 サイバーアシスト研究センター
Cyber Assist Research Center, National Institute of Advanced Industrial Science and Technology

We think extension of KRNL (Knowledge Representation for Natural Language) that is knowledge representation adapted for Natural Language. In concrete terms, we show adoption of Relations of UNL (Universal Networking Language) for Conceptual Relations of KRNL. And we introduce Certainty factor into Conceptual Relations.

1. はじめに

「人間のように思考し賢く振舞うコンピュータ」を作ることは人類の夢である。

この夢の実現に向けて、コンピュータによって人間の知能機能の実現を目指す人工知能の分野では、コンピュータに高度な知識を実装するべく人間の思考プロセスを分析し、それをコンピュータ上でシミュレートする研究が行われてきた。具体的にいえば、知識をコンピュータ上で表現し、それを利用して新しい知識を生成する処理の研究がなされてきた。この表現が知識表現であり、この処理が推論と呼ばれる。

知識表現と推論の研究を、実用に結びつけるのに障害となるものの一つが、知識表現にそって知識を生成する処理がほとんどの場合、人手による処理であることである。

一方、近年において電子情報資源の生成・蓄積・管理が発展し、多様な情報資源が大量に集積されつつある。さらにインターネットなどのインフラストラクチャーの整備が進んだことにより、これらの情報資源を多くの人が利用できるようになった。これら大量の情報資源の大部分は人間が読むことを前提とする自然言語で書かれており、機械が直接処理することに適していない。この問題を克服しようとしているのが、自然言語処理の研究である。

これら両者を結びつけることによって、Web 上の情報資源を、コンピュータが推論するときの知識として利用できれば有用である。そのために自然言語文から容易に変換できるような知識表現を提案すべきであろう。

本稿では、石塚研究室で提案している自然言語文に適合した知識表現 KRNL(Knowledge Representation for Natural Language)に対して、自然言語を意味ネットワークで表現する UNL の技術を用いて拡張について述べる。その際には推論を考慮した知識表現にするように考慮した。

また自然言語文における表現の任意性を吸収するために類義語辞書を用いることを考えている。このために KRNL への確信度の概念の導入について述べる。

2. KRNL (Knowledge Representation for Natural Language)

友部、石塚らが提案・研究している自然言語表現に適合した

連絡先: 谷口 智哉, 東京大学大学院情報理工学系研究科電子情報学専攻, 東京都文京区本郷, 03-5841-6755, tani@miv.t.u-tokyo.ac.jp

知識表現「KRNL(Knowledge Representation for Natural Language, 以下 KRNL)」について述べる。

2.1 KRNL の目的

現在までに、知識を組み合わせて推論するための表現として、述語論理、命題論理などの論理表現が研究されてきた。一階述語論理は、数学的基礎に立脚し、完全性と健全性を持つ演繹推論系が存在し、表現力も豊かであるため、自然言語と論理に着目した研究はいくつか行われている[石塚 96]。

しかし、Web 上の自然言語は、複数ユーザによって記述されているので、表現の任意性の問題が大きく、同じ知識の表現が異なると知識の組み合わせによる推論が困難となることが問題となる。

このため KRNL では、自然言語文を述語論理に変換する際に、表現の任意性に一定の緩い制約を課している。これによって組み合わせによる推論に必要な述語記号の一致を確保し、また自然言語に内在する規則を予め与えておくことで、自然言語で表現された知識の利用を簡単かつ効率的に行うものである。

KRNL では、自然言語を主語と述語を関係付ける概念関係式という表現に置き換える。つまり、概念要素と概念要素をその関係によって結びつけた概念関係式を知識の単位とする。この概念間の関係を概念関係と呼ぶ。この概念関係は、自然言語で書かれた文との対応から、数種類に限定されている。また、概念関係は方向性がある。概念関係同士を結びつけ、その概念関係から意味を読み取り、新しい概念関係式をつくるという概念縮合によって推論を行う。

以下個々について述べる。

2.2 概念要素

知識を構成する最小単位となる単語を原子概念と呼ぶ。分子概念は、原子概念を中心とし、その概念を修飾する0個以上の、修飾ラベル付き分子概念によって構成される。修飾ラベルを表1、表2にあげる。以下、自然言語から原子概念に変換する例を挙げる。

「花子にバラをあげる」
→ (あげる (OBJ バラ) (OBJ 花子))

表1 動詞句を修飾するラベル

| ラベル | 意味 | 例文 |
|-----|--------------------|--------------------------------------|
| AGT | AGenT 動作の主体 | 父が食べるりんご りんご (MOD 食べる (AGT 父)) |
| OBJ | OBject 動作を受ける対象 | りんごを食べる 食べる(OBJりんご) |
| ETC | 上記以外 | |

表2 名詞句を修飾するラベル

| ラベル | 意味 | 例文 |
|--------|--------------|----------------------|
| ISAMOD | ISA-MODifier | 犬のポチ ポチ(ISAMOD 犬) |
| HASMOD | HAS-MODifier | 花子の本 本(HASMOD 花子) |
| ETCMOD | ETC-MODifier | 赤い車 車(ETCMOD 赤い) |
| ATOF | Attribute OF | 花の色 色(ATOF 花) |

2.3 概念関係

KRNL では分子概念間の関係を概念関係と呼ばれる関係で記述する。さまざまな概念間の関係から同地関係や上位-下位関係、全体-部分関係など特に推論において有用で一般的なものを選んでいく。また自然言語で記された知識は「状態」を表す知識と「動作」を表す知識に分類されることも概念関係の選出において考慮している[中川 96]。

概念関係[ISA][INC]

概念関係[ISA][INC]は包含関係を表す。A が B に包含される場合は「A[ISA]B」となり自然言語の「A は B である(A が B を含む)」に対応し、その逆が[INC]である。

概念関係[HAS][POF]

概念関係[HAS][POF]は所有の関係を表す。A が B を所有している場合は「A[HAS]B」となり、「A は B の一部である。」という意味に相当する。また「A は B にある(場所の部分関係)」も表す。

概念関係[EQ]

概念関係[EQ]は等価関係を表す。自然言語での「A と B は同じである」に対応する。

概念関係[VAL]

概念関係[VAL]は属性値の関係を表す。A の属性 B が C である場合「B(ATOF A) [VAL] C」となり、自然言語「A の B は C である」という表現になる。

概念関係[DO][PV]

概念関係[DO]は主語と述語の関係を表す。A の述語が B であるなら「A[DO]B」と表す。これは自然言語の「A は B する」に対応する。[PV]は自然言語の「A は B される」に対応する。

以上の概念関係は方向性をもっているため、分子概念の位置が逆になると、その2つの分子概念を結びつける概念関係も変わる。次に例を挙げる。

$$a [ISA] b \Rightarrow b [INC] a$$

2.4 概念関係式

概念関係式は前述の分子概念、原子概念、概念関係を用いて次のように表される。

概念関係式:

$$\text{分子概念} [\text{概念関係} / \text{真値値}] \text{分子概念}$$

但し、真値値は省略可能でデフォルトは1である。真値値が0のときは否定を表す。次に自然言語文を概念関係式に変換する例を挙げる。

「太郎は花子にバラをあげる」

$$\rightarrow \text{太郎}[\text{DO}][\text{あげる} (\text{OBJ} \text{バラ}) (\text{OBJ} \text{花子})]$$

2.5 変数・論理演算子と IF-THEN ルール

KRNL では変数や論理演算子も用いることができる。<AND>は論理積、<OR>は論理和の意味を持つ。

さらに<IF>と<THEN>を用いることによってルールを記述することができる。<IF>の後に前提となる概念関係式を<THEN>の後に結果となる概念関係式を記述することによって表現する(すなわち IF 部を本体、THEN 部を頭部とするホーン節を表現する)。また\$で始まる原子概念は変数であり、この変数は一階述語ホーン節論理におけるように全称作用素で修飾されているものとする。

2.6 概念縮合

KRNL による推論は概念縮合と前述のルールによって行われる。概念縮合は、2つの概念関係式から新しい知識となる概念関係式を導き出す推論である。2つの概念関係式が持つ概念関係動詞の意味的な結合によって、あたらしい概念関係式を導く。この意味的な結合を概念関係積と呼ぶ。次にその結合を用いた推論の例を示す。

概念関係式1: 人間[DO]死ぬ

概念関係式2: ソクラテス[ISA]人間

概念縮合分子: ソクラテス[ISA]人間[DO]死ぬ

概念関係積: [ISA] × [DO] = [DO]



結論: ソクラテス[DO]死ぬ

2.7 KRNL と自然言語・述語論理への変換

自然言語から KRNL へ変換する機構は図のようになっている。自然言語から KRNL へ変換する場合は、変換システムによりいくつかの候補が示されるので、それを人が手直すことにより、KRNL への変換を行うので半自動であると言える。KRNL から一階述語論理への変換は自動変換されるので Prolog による推論が可能となる(図1)。

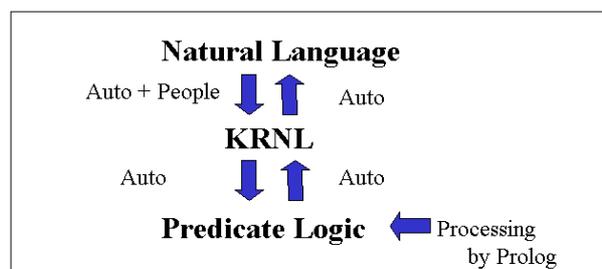


図1 KRNL の変換システム

3. UNL (Universal Networking Language)

Universal Networking Language (以下, UNL) はコンピュータのための言語である[UNDL]. UNL は, 人間のための自然言語の機能を模写したものである. その結果, 自然言語で表現される全ての情報や知識を, UNL で, 曖昧性なく表現することができる. UNL は Universal Word (以下 UW), Relation, Attribute, UNL Knowledge Base から構成される. また UNL はハイパーノードを含む意味ネットワークの形式で情報や知識を表現している.

UNL は翻訳のための中間言語として開発されており, 全ての言語は UNL を通して翻訳される. 例えば, 日本語から英語に変換するときは, 日本語から UNL に変換し, その UNL を英語に変換する.

本章ではこの UNL について述べる.

3.1 UNL System

UNL System は, UNL, 言語サーバ, UNL の基礎的なツールで主に構成されている[UNL System]. UNL System では, 自然言語で書かれた情報資源は言語サーバにより UNL と相互変換できるので, 言語障壁を超えた人間のコミュニケーションが可能となる. 自然言語から UNL に変換するエンコーダと, UNL から自然言語に変換するデコンバータは, 15 の言語において研究開発が進められており, 多言語基盤のアーキテクチャも完成している. UNL System が将来定着する可能性は大きいと考えられる.

3.2 UNL の所有権

UNL は国際連合 (以下, 国連) の中で誕生し, 国連大学高等研究所が中心となって研究開発が行われてきた. UNL の所有権は国連にあり, それゆえに人類の財産でもある. さらに, UNL は全ての人々が自由に使えることを保証するために, 国連の名前で特許申請が行われている.

3.3 Universal Word と Universal Knowledge Base

UW は UNL の語彙を構成する. UW の概念を次に示す.

UW: <Head Word> + <Constraint List>

<Head Word> は英単語の概念を表し, <Constraint List> で, その曖昧性をなくす.

例えば, <Head Word> が state である場合には次の 4 つの UW が考えられる.

```
state(icl > express)
state(icl > region)
state(icl > abstract)
state(icl > government)
```

但し, icl(InCLude)は上位概念を表す. このような UW の意味や役割を曖昧性なく定義するために UNL Knowledge Base (以下 UNL KB) が用意されている. UNL KB は, UW の包含関係に基づく階層関係を含む, UW 間の可能な関係を全て定義したものである.

3.4 Relation

UNL における Relation は, 2 つの UW 間の関係を明確にするために Relation ラベルによって表現されている. 次に Relation ラベルの一部について説明する.

agt(agent)

agt は動作の主体を表し, agt(do, thing) の形である. do は動作, thing は主体となる. 次に例を挙げる.

car runs... →

agt(run(icl > act(agt>volitional thing)), car(icl>vehicle)

obj(affected thing)

obj は, 作用を直接受けるものを表す. obj は obj(be, thing) や obj(do, thing) などの形をとる. 第1引数となる UW は出来事や状態を表し, 第2引数となる UW は作用を受けるものとなる. 次に例を挙げる.

cure the patient... →

obj(cure(agt>thing,obj>thing),patient(icl>person))

aoj(thing with attribute)

aoj は状態や属性を表す. aoj は aoj(be, thing) や aoj(thing, thing) などの形をとり, 第1引数の UW は状態や属性を表し, 第2引数の UW はその受け手を表す. 次に例を挙げる.

...leaf is red. → aoj(red(aoj>thing), leaf(pof>plant))

3.5 Attribute

UNL における Attribute は, 話者が意図した主観的な意味を表す. 例えば, 話者がどこに注目しているか, 可能かどうか (英語で言う can), また英語でいう完了形を表現するためなどに用いられる. Attribute は 「@+Attribute」 で表現される. 次に Attribute の一部をあげる.

@def(already referred) ex) the book you lost

@past(happened in the past) ex) It was snowing yesterday.

3.6 自然言語文の UNL への変換例

以上を用いて自然言語文を UNL で表現した例を挙げる. また, それを意味ネットワークで表した例を図 2 に挙げる.

The people build a huge tower.

→ agt(build(agt>thing,obj>thing), people(icl>person).@def)
obj(build(agt>thing,obj>thing)), tower(icl>building))
aoj(huge(icl>big), tower(icl>building))

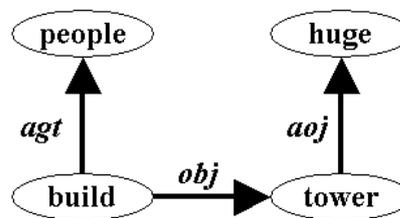


図 2 UNL のグラフ表現の例

4. KRNL の拡張

KRNL には問題点がある. そのいくつかをあげると, 自然言語から KRNL への変換機構が脆弱であること, 自然言語の表現力より KRNL の表現力が弱いので KRNL から自然言語への変換がうまくいかないこと, KRNL での単語の任意性に対応していないことである.

本章では, これらを改善するために KRNL の拡張を考える.

4.1 KRNL と UNL の対応

自然言語と KRNL 間の変換は 2 章で述べたとおり, 変換支援ツールを使い人が行うものであり, またその機構も簡単な構文解析とパターンマッチングを使用するものである. 変換機構は脆弱であると考えられる. また対応している自然言語は日本語のみである.

そこで自然言語から変換する際に UNL のシステムを利用することを考える. そのために UNL の二項関係と KRNL の概念関

係・ラベルの変換について考えると、UNL Relation の AGT(agent)は KRNL の AGT に、OBJ(object)は OBJ に、POS(possession)・POF(part-of)は[POS], HASMOD に対応し、それ以外の UNL Relation は KRNL において ETCMOD, ETC に対応する。しかし ETCMOD, ETC に対応する UNL の Relation の中で「動作が行われる時間と場所に関するラベル」は、頻繁に現れるので、KRNL にも対応するラベルを導入すべきだろう。

また上記以外でも UNL のラベルを用いて推論した場合に、その推論の妥当性をみることで、KRNL への UNL ラベルの導入を考え、KRNL の表現力の向上を考えている。

4.2 KRNL への確信度の導入

また 2 章でも述べたが、KRNL で使用される単語については、知識を記述する人による任意性を考慮していない。つまり、KRNL では単語が完全に一致することでしか、知識が同じであると判断できない。例えば「次郎は学校に行く」と「次郎は学校に登校する」はほぼ同様の意味を表しているのに、KRNL では全く別の知識として扱うか、全く同じ知識としてしか扱えない。

そこで UNL で使われる単語 UW を定義している UNL Knowledge Base(以下 UNL KB)の辞書を用いて、任意性を吸収することを考えている。現在は KRNL で主に使われている単語が日本語であるので、日本語の類義語辞書で代用することも考えている。

類義語の概念を KRNL に適用するために KRNL に確信度(確率と考えても良い)を導入する。確信度は、知識が正しいとされる確率である。つまり、今までは事実でしか推論をしなかった KRNL に仮説推論を入れることになる。表現は次のように概念関係と確信度をセットで表記する。

[概念関係: 確信度]

例えば、チンパンジーとゴリラが類義語辞書で近い関係であり、同じ意味とする確信度が 0.7 だと考えられる場合は次のように表す。

チンパンジー[EQ: 0.7]ゴリラ

2 章で述べたとおり否定を表すのは確信度が 0 のときである。例えば、チンパンジーと人工知能は同じではないことを表すには

チンパンジー[EQ:0]人工知能

となる。

この確信度を用いることによる利点を説明する。例えば「太郎は家が建てる」という事実は、KRNL で次のように示される。

KRNL1: 太郎[DO: 1]建てる(OBJ 家)

このときに「太郎は家を建設するか」という質問を考えると、質問は KRNL で

KRNL2: 太郎[DO: 1]建設する(OBJ 家)

と表せ、この表現が KRNL1 に当てはまらないので NO と答える。

ここで確信度を導入し、「建設する」は「建てる」と同じ意味であることが確信度 0.9 である、ということを知り、辞書からわかったとすると、

KRNL3: 建設する[EQ:0.9]建てる

と表せ、KRNL1,3 を概念縮小させることにより

KRNL4: 太郎[DO: 0.9]建設する(OBJ 家)

という知識が導かれ、質問に YES と答えることができる。このときには、質問者に 0.9 という確信度も示す必要がある。

また、「太郎が家を建てる」という知識の確信度が 0.8 だった場合を考えると KRNL は

KRNL1': 太郎[DO: 0.8]建てる(OBJ 家)

となり、「太郎は家を建設するか」という質問に対しては KRNL1' と KRNL3 を概念縮小させて

KRNL4': 太郎[DO: 0.72]建設する(OBJ 家)

となり、答えは Yes:0.72 となる。答えの確信度である 0.72 は 2 つの知識の確信度を掛け合わせたものである。確信度の計算法については、考案中であるが、確信度の低いものを組み合わせれば確信度がさらに下がることや、後述の仮説推論の考えからも妥当だと考えている。

また、現在 KRNL の推論は Prolog で実装されているが、この確信度を導入するためには仮説推論ができる効率の良いシステムが必要である。

仮説推論とは、真か偽か不明な事柄をとりあえず真と考えて(仮説を立てて)推論を進め、矛盾なくゴール(観測された事実)が説明できれば、建てた仮説は正しかったと考える推論である[松尾 01]。基盤性と実用性の両面で有用な枠組みであるが、NP 完全または困難な問題であるため、推論速度が大きな問題となり、高速に仮説推論問題を解く手法がいくつか開発されている[大澤 95][松尾 01]。

仮説推論をする際に最も妥当な推論は、最も確信度の高い組み合わせで推論をすることである。その場合には確信度からコストを計算し、最小のコストの推論(最尤仮説)を見つけることである。ここでコスト計算として、

$$C = -\log P$$

(コスト: C 確信度: P)

を定義する。このコストは、確信度 0 の知識のコストを無限大にして推論に使われないようにすることや、確信度 1 の知識のコストは 0 にして、今までの事実と同様に扱えることを考慮している。

5. おわりに

自然言語と適合する知識表現として KRNL を紹介し、その拡張のために自然言語処理技術のひとつである UNL を紹介し、KRNL の拡張について述べた。

これからの研究方針として、UNL を利用した自然言語から KRNL への変換システムの構築や確信度を考慮した推論機構を利用できるようにすることを考えている。

また、確信度を使う際に人間関係を考慮して確信度を利用者によって変化させることを考えたい。このときには、人間同士の信頼度のベクトルを用いることを考えている。

[石塚 96] 石塚 満: 知識の表現と高速推論, 丸善, 1996.

[中川 96] 中川 裕志: 順接複文における主語の共参照関係の分析, 自然言語処理, Vol.3, No.2, pp.59-74, 1996.

[Iwanska 00] Iwanska, L. and Shapiro, S. C.: Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language, AAAI Press/MIT Press, 2000.

[大澤 95] 大澤 幸生, 石塚 満: 多項式時間仮説推論を達成するネットワーク化バブル伝搬法の述語論理への拡張, 人工知能学会誌, Vol.10, No.5, pp.731-740, 1995.

[松尾 01] 松尾 豊, 石塚 満: コストに基づく仮説推論の 2 種の連続値最適化問題 への置換法とその協調による推論法, 人工知能学会誌, Vol.16, No.5 B, 2001.

[UNDL] UNDL: Universal Networking Language, <http://www.un dl.org/>.

[UNL System] The UNL System, <http://www.unl.ias.unu.edu/unl sys/>.