

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 4 月 3 日現在

機関番号：32638

研究種目：基盤研究 (C)

研究期間：2010～2012

課題番号：22560406

研究課題名（和文）日本語文からの論理式生成と推論法の研究

研究課題名（英文）Generating the Well-Formed Formulas from Japanese Sentences and the Inference Scheme

研究代表者

石川 勉 (ISHIKAWA TSUTOMU)

拓殖大学工学部・教授

研究者番号：10281370

研究成果の概要（和文）：自然言語インターフェースの質問応答システムや対話型ロボットへの応用を目指し、日本語文から機械的に論理式を生成する手法およびその論理式を用いて効率的に推論する手法について研究した。また、それら技術を広く利用可能とすべく具体的なプログラムモジュールとして具現化しツール化を図った。さらに、以上の技術の有効性を確認するため、これらを利用した自由対話システムを開発し学会等でデモ公開を行った。

研究成果の概要（英文）：We have studied a mechanism that can generate well-formed formulas from Japanese sentences automatically and an inference scheme to process the formulas, aiming to apply to several intelligent systems such as semantic information retrievers, dialogue robots and so on. We have also developed the tools to make use of these mechanisms in concrete application systems. Usefulness of the tools has been shown through the demonstrations of the open-ended dialogue system using them.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	700,000	210,000	910,000
2011年度	1,400,000	420,000	1,820,000
2012年度	800,000	240,000	1,040,000
総計	2,900,000	870,000	3,770,000

研究分野：工学

科研費の分科・細目：電気電子工学・システム工学

キーワード：論理式、推論、日本語文、意味処理、対話システム、概念ベース

1. 研究開始当初の背景

インターネット情報の爆発的増加に伴い Web 検索技術には目覚ましい進歩が見られるが、まだユーザが欲しい情報が直ちに得られ

るレベルには至っていない。また、MUC(Message Understanding Conference) 等における質問応答技術についても、パターンマッチの域を脱したとは言い難い。

また、最近の人型ロボットはその動作に関しては著しい進歩が見られるが、人間との自然な対話という点では未熟である。対話に関してはチケット販売等限定的な応用では実用システムになりつつあるが、話題を限定しない場合は、未だに Eliza の域を脱したとは言いがたい(所謂、人工無脳 (chatterbot) のレベル)。

これらは主に自然言語文に対する意味的な扱いが不十分なためである。すなわち、前述のような電子化情報の有効利用さらには対話システムの高度化には、自然言語文を一定の知識表現に変換しそれを用いてある程度深いレベルでの推論処理を施す必要がある。さらに実用的応用のためには、これらは一貫した自動化が可能でなければならない。従来の研究では、完璧性や論理の厳密性を追求するあまり、この視点が極めて希薄と言える。

2. 研究の目的

本研究は、日本語文から機械的に論理式を生成し、それを用いて効率的に推論する手法を確立することを目的とする。自然言語の持つ曖昧性と真偽を問題とする論理とは本質的に相容れない面があり、文の論理式表現は極めて難しい課題である。また、その論理式を用いた推論でも、多義性等から式自体の正当性が保証されないため、実用性のある解を得ることは容易ではない。これに対し、本研究では完璧性を追及するのではなく、“概略的あるいは近似的でも実用的な結果を得ること”を狙う。応用としては、近未来的には自然言語インターフェースの質問応答システム、究極的には対話型ロボットの脳を想定している。

3. 研究の方法

本研究の主要課題は、日本語文からの論理式生成、その論理式を処理する推論法、およびそれらの有効性を確認するためのデモンストラシステムの構築である。

本研究では、前述したように論理式生成から推論まで一貫した自動化を想定しているため、各課題の研究成果を互いにフィードバックしつつ段階的に研究を進める。この際、既存の利用可能な自然言語処理ツール(茶笥、南瓜といった形態素・構文解析ソフト等)・辞書類(EDR 電子化辞書等)を積極的に活用し、研究の効率化を図っていく。

各課題の具体的な進め方は、論理式生成ではまず一階述語論理式への変換法を確立し、変換ツールとして具現化する。次にこれをこれまで開発してきた順序ソート論理表現に変換する方法を開発する。

推論法については、まず知識の不完全性と複雑な知識構造に対処可能な方式を確立し、基本的な一階述語論理ベースの推論エンジンとして具体化する。次に、順序ソート論理への拡張、本格的な推論エンジンとしての構築、推論処理の高速化法の研究へと展開していく。

デモンストラシステムの構築は、上記課題の成果が具現化できるレベルに達した段階で着手し、研究の進捗に合わせ適宜その改良を行っていく。

4. 研究成果

本研究の成果は以下の通りである。

(1) 日本語文からの論理式生成

① 想定する知識表現法

本研究では、論理的に真偽が定まる(事象や現象等)全ての日本語文を対象とした。これら文を単文、複文共に以下の形式の述語式 L あるいはその結合である論理式で表現する。

$$L = sP(r_1:t_1, \dots, r_n:t_n)$$

ここで、P は述語部であり、文の主節の述部を構成する単語(動詞、形容詞、名詞のいずれか)とする。P が動詞、形容詞の場合は終止形を、名詞の場合(述部が「～である」の場合)はそのまま用いる。 $r_i:t_i$ はラベル付き引数(項)であり、 t_i が引数本体、 r_i がそれと述語との関係を表すラベルである。引数本体は基本的には、その文の述部と関連する名詞または名詞句である。ラベル r_i は述語部により異なり、動詞の場合は agt(主格), obj(対象格), plc(場所格)等の深層格(EDR 電子化辞書で用いられるラベルを拡張)を、形容詞や名詞の場合は sbj, inst のような新たに設定したラベルを用いる。また、s は文の様相や態を表す識別子であり、過去(*), 進行(g), 受身(u), 丁寧(t), 否定(!)等とする。推論処理においては識別子が一致することを前提にラベル毎に照合をとり、単一化等の処理を行っていく。

複文は上式 L で主節を表し、従属節は L 中の P や t_i に埋め込んで表現する。具体的には従属節を L' (以下、これを節述語と呼ぶ)、それにより修飾される語を a としたとき $[a]\{L'\}$ という形で埋め込む。ここで、節述語 L' は L と構造上は基本的に同一である。ただし、L が真偽を表すのに対し、L' は事象やアクションを表す。なお、L' ではその述語部 P と a の間に格関係が存在する場合には該当するラベルの引数部に記号 # を付与する。

なお、条件を表す複文については、一つの述語式では表現できないため、主節と従属節をそれぞれ独立の文として扱い、節毎に述語式を用いて表し、それぞれ連言や含意の論理記号を用いて連結する。

以下に、本形式での表現例を示す。

- ・「太郎は昨日、大阪に行った」
⇒ *行く (tme:昨日, agt:太郎, gol:大阪)
- ・「太郎は大学生である」
⇒ 大学生 (inst:太郎)
- ・「刑事は犯人が隠れている家を探している」
⇒ 探す (agt:刑事,
obj:[家]{隠れる (agt:犯人, gol:#)})
- ・「花子は太郎が経営している会社の社員である」
⇒ [[会社]{経営する (agt:太郎,
obj:#)}]の社員 (inst:花子)
- ・「春が来れば花が咲く」
⇒ 来る (obj:春)→咲く (obj:花)

②論理式生成法

変換対象の文を、形態素・構文解析(それぞれ、茶釜、南瓜を利用)し、単語の品詞情報、文節の係り受け、格助詞、構文番号などを獲得する。その後、これら情報をもとに述部の品詞毎に、意味解析を行い論理式に変換する。

文の述部が動詞の場合は前述した情報とEDR辞書(動詞共起パターン副辞書および単語辞書)を基に必須格を定める。同辞書に登録されていない動詞については、IPAL動詞辞書から同類語を求め、その動詞のEDR辞書中の共起パターンを利用して必須格を定める。

その後、任意格の判定を行うが、これについては格助詞の前方名詞のカテゴリ情報を基に決定する。即ち、各格助詞についてラベル毎に該当し得る意味範囲をシソーラス上のカテゴリ番号を基に設定し(この範囲を以下、照合範囲と呼ぶ)、前方名詞がこの照合範囲に含まれるか否かをチェックして決定する。

述部が名詞の場合、具体的には「AはPである」の文型では、基本的にはPとAはクラス・インスタンスあるいは上位下位の関係を表すとし、ラベル“inst”を用いて、P(inst:A)と変換する。

これらの関係がない場合(文中の語間が等値の関係にある場合等)、述部にその関係を表す述語(以下、EQ述語)を用いて変換する。例えば、「太郎は20歳である」は“EQ(atr:太郎の年齢, val:20歳)”と変換する。この際、この例のようにAとPの関係を示すため、Aに対して補足語H(この例では“年齢”)を補いEQ(atr:AのH, val:P)と変換する。このHの獲得は、予め作成しておいた補足語テンプレートを用い、これとPが属す意味カテゴリとを照合することで行う。

述部が形容詞の場合は、ラベルを文型と格助詞ごとに決定し、述語式に変換する。変換パターンは14種設定し、例えば「私は試験勉強で忙しい」の場合は、“忙しい (sbj:私, s-obj:試験勉強)”と変換する。

③本生成法のツール化

上記の論理式生成法を具体的に利用可能とするため、プログラムとして実装しツール化を図った。本ツールは、文法的に正しく省略の無い日本語文を入力すると前述した形式の論理式を出力するもので、我が国ではほとんど目にしないものと言える。これについては、ホームページを通してインターネット上で公開した。

(2)推論法

①一階述語論理ベースの推論法

これまで開発してきた概略推論法 ARSK を、概念間の意味的な包含関係も考慮できるように拡張した。さらに、単語だけでなく名詞句等の構造を持った概念も扱えるように改良した。

ARSK は、基本的には欠落知識を類似する知識で補い、概略的な解を得る導出処理に基づく推論法である。導出処理では、2つの親節 $A \setminus B$ と $A' \setminus C$ から、 A と A' が単一化可能または A が A' を意味的に包含しているならば信頼度“1”の導出節 $B \setminus C$ を導く。また、 A と A' の意味的な類似度が閾値よりも大きい場合にもその導出節を生成する。この場合は、その類似度を導出節の信頼度とみなし導出処理毎に累積し解の信頼度とする。また、その値が閾値以下になった場合には解が無いとして推論を停止させる。

この意味的な包含・類似関係の判定は述語間および同一のラベルの引数間で行い、述語式全体の類似度はそれらの最小値としている。類似度の算出にはこれまで開発してきた概念ベース(概念間の意味的な類似性を判定するための知識ベース)を用いている。

なお、この推論法で類似度の閾値を“1”とすれば、類似関係を扱わない一般の一階述語論理での推論法(SLD法)となる。

②本推論法のツール化

上記の推論法を具体的に利用可能とするため、プログラムとして実装しツール化を図った。本ツール(推論エンジン)は(1)②の出力をそのまま推論処理できるため、両者の併用により日本語入力の質問応答システムや各種コンサルティングシステム等を容易に実現できる。なお、本ツールは関連研究者に配布している。

③順序ソート論理への拡張

①の推論法を順序ソート論理に拡張した。従来の順序ソート論理では、ソート概念はあらかじめソート階層(サブソート関係)が定義された単純な概念に限定されていた。これに対し、本研究では名詞句のような構造を持った概念を許容すると共にソート階層を推論時に動的に判定可能なように、新たな推論規則を考案し順序ソート論理自体の枠組みも拡張した。

対話システム等の動的かつ自然言語文を

対象としたシステムでは、あらかじめ知識を定義しておくことは不可能であり、このような枠組みが必須となる。さらに、本推論法ではソート自体にも変数の使用を可能とした。

(3) デモシステムの構築

本研究の有効性を確認するため、デモシステムとして、(1)②と(2)②のツールを利用した自由対話システムを構築した。

本対話システムは、質問応答、常識知識獲得・利用・精練、吉凶相槌機能等を有する。現時点で実現されている対話例を以下に示す。

[質問応答]

USER : 私は車が好きだよ
COM : ほおー
USER : 私はよくスピード違反をします
COM : ふうん
USER : あと、私は度胸があるよ
COM : へえー
USER : 私は何に向いていますか？
COM : レーサーに向いていると思うよ！

[常識知識獲得・利用・精練]

USER : 花子はタレントです
COM : すごいですね
USER : だから、花子は綺麗です
COM : ナルホド
USER : 良子もタレントです
COM : じゃあ、良子さんも綺麗なんだね
USER : いや、綺麗じゃないです
COM : どうして？
USER : お笑い系だからね
COM : ふーん
USER : 明子はタレントです
COM : じゃあ、明子さんは綺麗なんだね
USER : いいえ
COM : じゃあ、明子さんもお笑い系なんだね

[吉凶相槌]

USER : 先週、宝籤に当たった
COM : ラッキーですね
USER : やっと風邪が治りました
COM : よかったですね
USER : 今日、財布を落としちゃった
COM : 気を落とさないで

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計3件)

- ① 高柳俊祐、上条敦史、石川 勉 : 日本語文から拡張型述語論理式への自動変換ツール: CONV、人工知能学会論文誌 Vol. 27 NO. 5B pp. 271-280 (2012. 9) 査読有
- ② 高柳俊祐、上条敦史、石川 勉 : 日本語

文を想定した述語論理ベースの知識表現法とそれへの自動変換法、拓殖大学理工学研究報告 vol. 12 No. 1 pp. 13-21 (2011, 3) 査読有

- ③ 石川 勉 : 日常語をベースとした順序ソート論理による知識表現法とその推論処理法、拓殖大学理工学研究報告 vol. 11 No. 2 pp. 61-63 (2010, 11) 査読有

[学会発表] (計6件)

- ① 石川 勉 : 日本語文を想定した知識処理ツール、人工知能学会第 26 回全国大会 AI レクチャ: ツールボックス 2012. 6. 14 山口県教育会館
- ② 石川 勉 : 言葉で考えるコンピュータ---日本語文の論理処理---、人工知能学会第 25 回 ことば工学研究会 SIG-LSE-C002 pp. 21 2011. 11. 25 拓殖大学
- ③ 高柳俊祐、中山敬太、石川 勉 : ペット型ロボットを想定した知識処理利用の雑談型自由対話システム、人工知能学会第 25 回全国大会 3C2-0S19-8 2011. 6. 1 いわて県民情報交流センター
- ④ 中山敬太、石川 勉 : 備忘録知能を有する雑談型の自由対話システム、情報処理学会第 73 回全国大会 1S-4 2011. 3. 4 東京工業大学
- ⑤ 佐藤雄也、石川 勉 : 知識の意味的な包含・類似関係を利用した日本語向き概略推論法、情報処理学会第 73 回全国大会 6Q-5 2011. 3. 4 東京工業大学
- ⑥ 高柳俊祐、久野由隆、石川 勉 : 演繹・帰納・発想のメカニズムを利用した雑談型の自由対話システム、第 9 回情報科学技術フォーラム (FIT2010) 3N-5 2010. 9. 7 九州大学

6. 研究組織

(1) 研究代表者

石川 勉 (ISHIKAWA TSUTOMU)
拓殖大学・工学部・教授
研究者番号: 10281370