

7P-4

文章理解を支援する知識の学習に関する研究

藤田 悟, 相田 仁, 猪瀬 博, 齊藤忠夫
(東京大学工学部)

1. はじめに

自然言語理解システムに利用するための文脈に関する知識は、扱う文章の内容・環境・あるいは時間の経過によって変化するため、その作成自体に困難さを持つ。

本稿で取り上げる FOCUS (Frame Oriented Context Understanding System) は、同一のテーマについて記述した複数の文章を経験として蓄積し、その中から文章に共通して存在する文脈に関する知識(文脈情報)を再構成することを目的としている。学習対象としては、イラン・イラク戦争に関する新聞記事を用いている。以下、文脈情報学習についての基本概念、及び FOCUS の概要について論じる。

2. 動的再構成可能な記憶構造と文脈2.1 文章の作成・理解過程と文脈

人間が文章を作成する際は、主張を表現するために、頭の中にくいつかの state を用意し、それらに関連づけながら一連の並びを生成する。FOCUS ではこれら state 間の関連づけられた並びを文脈としてとらえている。実世界中の文は、この思考世界の構造を mapping 関数によって、展開することにより得られる。

次に、実世界中の文章を理解する方法であるが、まず前述の mapping 関数に対する逆関数が必要である。これは、自然言語という実世界中の表現から、人間の内部表現である原言語を導く関数にあたる。しかし、これだけでは各 state は推定できるが、それらのつながりを抽出することはできない。そこで、もう一つ、実世界中における文と文の間の関係を発見する必要がある。この関係を relation と呼ぶ。

これら state と、relation の解析を行ないながら文章理解を進める際に、人間は文脈に関する過去の膨大な経験を利用していると考えられる。この文脈理解を補助する知識は、「一般化された文脈」「文脈の枠組」といった性格を持つものである。本システムは、この文脈情報の学習を目的としている。

2.2 文脈情報の動的再構成に関する提案

1文ごとに文章の内容の理解が深まり、それに適した文脈情報を要求する場合を仮定する。このような場合、最初に必要な文脈情報とは、非常に一般化されたものであり、次第にスコープを狭くして、特定の経験に近いレベルでの情報が重要となる。この考え方に基づく、帰納的学習により一つの文脈情報に一般化していくことは無意味なことであり、新しく入力された文章の理解の補

助をする文脈情報は、その都度過去の経験の中から、その文章に合ったスコープで再構成するのがもっとも適切である。

2.3 動的再構成可能な記憶構造

このような文脈情報を再構成可能な記憶構造として、記憶の階層構成法を提案する。以下に各レベルを紹介する。

EM (Event Memory)

各文章の詳細な内容にわたる記述が state の下に記述される。

CM (Causal relation Memory)

EMと同じレベルに位置し、ここに、EM中に記述された state 間の relation が記述される。

GEM (Generalized Event Memory)

EM中の state のうち、同じまたは近い意味の述語をもつ state を一つにまとめあげ、以下それぞれ格に対しても一般化を行なう。

GCM (Generalized Causal relation Memory)

GEM中の state 間の relation を記述する。CMの一般化されたものという立場もあり、同じ種類の relation がまとめあげられる。

SM (Situation Memory: 状況記憶)

GEM、GCM中で特に関係が強いと思われる組をSMの中に表現し、GEM、GCMへのポインタを設ける。

3. 文脈情報学習システム FOCUS3.1 FOCUS の構造3.1.1 FOCUS の知識構造a. state の表現

state の内部表現は、基本的に格フレーム表現である。中心となる述語は"isa"というフレーム内に記述している。これは、述語自身はその文の上位語であって、その述語に行為主や対象などといった制約がついて文が成り立っているという観点に立っているからである。名詞句の表現にも同様に"isa"というフレームが利用される。

b. relation の表現

人間は、様々な relation を理解できる。例えば、理由-行動関係であり、原因-因果関係である。しかし、ここでは、文と文のつながりを発見できれば充分であり、なるべく文の表層構造だけで決定できる relation を用いる態度をとった。以下に、現在のとこ

Acquisition of Knowledge for Natural Language Understanding

Satoru FUJITA, Hitoshi AIDA, Hiroshi INOSE, Tadao SAITO

University of Tokyo

ろ利用している relation およびその説明を示す。

- IR 情報源の動作を示す文と情報を示す文との間の relation。
- SEQ 時間的に連続な動作を示す文の間にある relation。
- PARA 時間的に並列している文の間の relation。
- COM 文中のある名詞を関係代名詞的に修飾している relation。
- BUT 二つの文が否定的つながりを持つような relation。

例えば、「～₁し、～₂した。」という文があった場合、～₁と～₂という文の間に SEQ という relation を設ける。

3.1.2 FOCUSの学習方式

FOCUSの全体像を図1に示す。知識は、EM・CM・GEM・GCM・SM中に蓄積され、必要に応じて context maker によって文脈情報が再構成される。parser、frame maker、relation finder の3つは現在のところ人手で行なっている。その他の部分について、以下に説明を加える。

- a. EM constructor フレーム表現された文章の内容全てをEM中に展開する。
- b. CM constructor relation finder により発見された relation を用いて、EM中の state 間に relation を設定する。
- c. GEM reconstructor FOCUSの中心をなす部分である。EM中に新しい文章が展開されることに起動され、EM中の各 state を単位として、state と state の比較を行なう。
- d. GCM reconstructor CMに書かれている情報を利用して、GEM中に展開された一般化された state 間に relation の link を設定する。
- e. SM reconstructor GEMの state とGCMの relation の情報を用いて、共起する state-relation の並びの primitiveを発見し、SM中にそれを展開する。
- f. context maker 文脈情報要求時にのみ起動される。GEM・GCMの情報を用い、SMを参照しながら文脈を構成する。現在2種類の文脈生成モードがある。
 - i) 複数の文章を指定して、それらに共通する文脈情報としてもっともらしいものを抽出する。
 - ii) 文あるいは文章を格フレーム表現の形で入力して、それに関連する文脈情報を推定する。

3.2 文脈情報学習の実験結果

現在、イラン・イラク戦争に関する記事を20個入力している。(毎日新聞1985年1月8日～4月2日) 複数段落にわたる記事は、その第1段落のみを学習の対象とした。SMを見ることによって、どのような種類の文脈が存在したかを概観することができる。3文章以上によって支持された situation は、次の3つである。

- ・ イラク軍スポークスマンは、イラク軍の飛行機が攻撃し、帰還したと発表した。

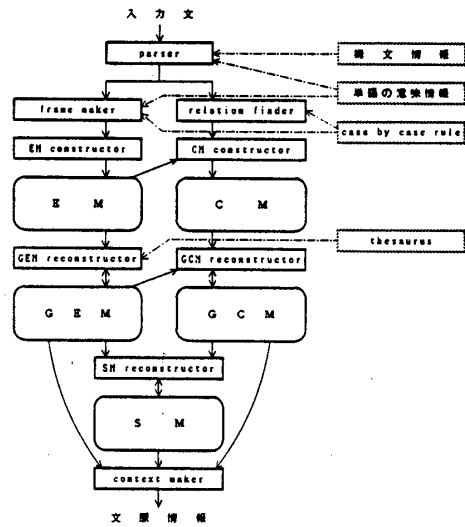


図1 FOCUSの全体像

- ・ イラク軍スポークスマンは、イラク軍の飛行機が石油積み出し基地を攻撃したと発表した。
- ・ イラク軍スポークスマンは、軍が、南部戦線においてイランに対する新たな攻勢を開始すると発表した。

3.3 文献検索システムとしての応用

FOCUSが文脈情報を抽出する際に、文章を格フレームの形式でシステムに蓄積するため、それを利用して文章の検索システムを作成することができる。比較は、入力文とGEM及びEM中に蓄積された情報との間で行なう。こうすることにより、自然言語による検索文で、新聞記事自身である一次情報を用いた検索システムが可能となる。

4. 評価・検討

他システムとの比較を含めて、FOCUSの特徴を次に述べる。

- (a) multiple concept, multiple step の学習
- (b) 最初にシステムに学習の核となる大枠を与える必要がない。
- (c) 様々なスコープで文脈情報を参照することが可能

5. おわりに

文脈情報の学習という点について、本稿で論じた再構成による学習は、一応の成果をおさめた。従来の文脈に関する学習システムに比べ、一般化の行き過ぎという最大の問題点を回避し、かつ、様々なスコープにより文脈情報を再構成できるという点で、文脈情報学習のための興味ある一方式である。増大する記憶・入力段の意味処理・得られた知識の応用等が、今後の課題である。

<参考文献>

[1] 藤田、猪瀬、斉藤：「新聞記事からの意味的構造の抽出」、情報処理学会第31回全国大会、1985。
 [2] 堀：「自然言語理解システムのための知識の表現と帰納的学習」、東京大学学位論文、1983。