

接続詞「にもかかわらず」を含む文からの因果知識の獲得

富川昌則[†] 石川勉[†]

拓殖大学工学部情報工学科^{††}

1. はじめに

コンピュータに人間のような柔軟な処理をさせるには常識知識の獲得とその活用が不可欠となる。常識知識のうち事実知識に関しては、従来から百科辞典や国語辞書からの獲得、インターネットを利用した不特定多数からの獲得等、広く研究されてきている[1-3]。これに対し因果知識の獲得については、いくつかの研究例[4,5]はあるがまだ広く研究されているとはいえない。常識的な因果知識は、冒頭の目的、特に柔軟な判断のための推論処理では極めて重要となる。本報告では、接続詞「にもかかわらず」を含む文からの因果知識(否定的)の自動獲得法について提案する。

2. 基本的な考え方

接続詞「にもかかわらず」は、広辞苑には“前文の内容を受け、それに反対の内容を下に続ける”とあるが、一般には次の二つの形態で使われることが多い。

- ① 前文の条件の下で、後文の内容は一般には起こらないことを表す(例: 雨が降っているにもかかわらず運動会は実施された)。
- ② 前文に対して、それと反対の内容を単純に対比的に表す(例: 太郎は真面目に掃除をしているにもかかわらず次郎は外で遊んでいる)。

①の形態で使われるときは、一般に“前文なら後文でない”ということの意味し、前文をP、後文をQで表せば、“ $P \rightarrow Q$ ”の形の知識に変換できる。すなわち、この形態の文からは否定的な因果知識を獲得可能である。従って、ここではこの形態を知識獲得の対象とする。

ただし、後文が否定文の場合には“ $P \rightarrow Q$ ”となるが、この場合、PはQの必要条件でしかないため除外することとする。例えば、“晴れているにもかかわらず運動会が行われない”の場合、“晴れている”はあくまで“運動会が行われる”の必要条件にすぎない。

なお、これらの形態の識別は、前文と後文で意味上の主語が異なるか否かにより行い、異なる場合は②とすることとした(“輸出が増加するにもかかわらず景気が良くなるらない”のように①である場合もあるが判定が困難なため)。ただし、主語が異なっても“AはBを～するにもかかわらずBは～する”のように判定が容易な文は獲得対象とした。

3. 獲得対象とする文型と変換形式

(1) 知識の表現形式

知識は基本的には述語論理により“述語(ラベル:引数, …)”形式で表す[6]。例えば、“医者(は)患者(を)治す(は)”

“治す(agt:医者,obj:患者)”のように表す。なお、二者が等値の関係を示す場合は述語“EQ”を用いて“EQ(atr:太郎の年齢,val:20歳)”のように表す(以下、EQ 述語と呼ぶ)。

(2) 対象とする文型と基本的な変換形

対象とする文型は以下の2パターンとする。

i) 前文(P)が単語(名詞句を含む)の場合(文型1)

Pが単語(名詞と形容詞)、名詞を連結した複合語、名詞を助詞で連結した名詞句(形容詞と名詞の連結も含む)の場合である。この場合、基本的には以下のような形式で知識獲得を行う。

- $P(\text{inst:Qの主語}) \rightarrow \sim Q \dots$
- $\text{EQ}(\text{atr:Pの上位概念, val:P}) \rightarrow \sim Q \dots$

ii) 前文(P)が節の場合(文型2)

この場合、PとQをそれぞれ独立した文ととらえ、以下の形式で知識獲得を行う。

- Pの主動詞(agt:Pの主語, …) $\rightarrow \sim Q \dots$

4. 変換法

4.1 変換処理の流れ

変換の流れは以下の通りである。

i) 文型の分類

形態素解析、構文解析(それぞれ茶釜[7]、南瓜[8]を利用)を行い、Pの構造から文型1または文型2に分類する。この際、表面的には文型2のような文でもPを単語として扱うべき場合もあり、後述のように意味的な主語の判定を行う。

ii) 述語知識への変換

文型1の場合は単純にはPを述語式には変換できない。このため普通名詞、普通名詞+サ変名詞、形容詞の3パターンに分類し、それぞれ補足語を補い述語式に変換する。また、文型2の場合は後述のように文の整形を行い、P、Qをそれぞれ述語式に変換する。

iii) 知識の一般化

述語式の引数に固有名詞が含まれている場合は、知識の一般化を行う。具体的には、固有名詞を普通名詞や変数で置換する。

4.2 文型1の変換法

(1) Pが普通名詞の場合

その単語の属するカテゴリーを調べ、補足語を補い述語式に変換する。具体的には、以下の例のように単語が具体物に属する場合は、補足語をQの主語とし述語式(引数のラベルはinst)に変換する。また、単語が抽象物に属する場合は、Pの上位概念を補足語としEQ述語に変換する。この例では、“状況”が“雨”の上位概念である。

例) “P=馬”の場合は“馬(inst:Qの主語)”

“P=雨”の場合は“EQ(atr:状況,val:雨)”

(2) Pが“普通名詞+サ変名詞”の複合語の場合

普通名詞とサ変名詞に分け、サ変名詞に“する”を加え動詞化する。次に、EDRの動詞共起パターン副辞書[9]を用いて助詞を補い、述語式に変換する(図1)。

A Method for Acquiring Causality Knowledge from Sentences Including Conjunction “NIMO-KAKAWARAZU”

[†]Masanori Tomikawa, Tutomu Ishikawa

^{††}Department of Computer Science, Takushoku University

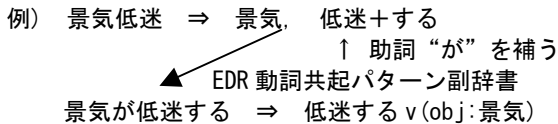


図 1: 普通名詞+サ変名詞の変換法

(3) P が形容詞の場合

形容詞を名詞化し, その名詞のカテゴリーから補足語を補い EQ 型述語に変換する. 例えば“P=暑い”の場合“暑さ”と名詞化し, それより補足語“状態”を補い“EQ(atr:状態, val:暑い)”と変換する.

4.3 文型 2 の変換法

4.3.1 文の整形

構文解析結果を利用して P と Q の意味上の主語を特定し文を整形する. 例えば以下の例文の場合,

- 例1) 太郎は雨にもかかわらずビーチに飛び出した
- 例2) 太郎は風邪を引いているにもかかわらずプールに入った

例1)では“太郎”は Q の主語と判定されるため, “P=雨”, “Q=太郎はビーチに飛び出した”とする. すなわち, 文型 1として処理する. また, 例2)では“太郎”は P の主語と判定されるが, この場合はこの主語は Q の主語でもあるとし, “P=太郎は風邪を引いている”, “Q=太郎はプールに入った”とする.

4.3.2 述語知識化

(1) 述語が動詞の場合

“A は B を C する”などの文型であり, “C(agt:A,obj:B)”のように変換する. ここで, ラベル付けはフィルモアの格文法を用いる. 具体的には EDR 電子化辞書の日本語共起副辞書を利用して変換する.

(2) 述語が形容詞の場合

全 15 パターン[6]で分類を行う. 例えば, “A は B”の場合は“B(sbj:A)”, “A は B が C”の場合は“C(sbj:A,s-obj:B)”のように表す. ラベルは格助詞より決定する.

(3) 述語が名詞の場合

基本的には, “A は B である”の文型であり, “inst”ラベルを用いて“B(inst:A)”と変換する. ただし, “A”や“B”の意味カテゴリーをシソーラスで調べ, 一定の上下関係がない場合, あるいは“B”が数値や単位に関する語の場合は EQ 述語を用いて変換する.

4.4 知識の一般化

(1) 述語式の述語が動詞の場合

この場合は, 固有名詞のカテゴリーを調べ, “人間”, “場所”, “組織”等の 8 種類の普通名詞の中からこの上位概念を選定する. 選定した上位概念によりこの固有名詞を以下の例のように置き換える.

- 例) EQ(atr:状況, val:雨)
 ⇒ 飛び出す(agt:太郎, plc:ビーチ)
- ⇒ EQ(atr:状況, val:雨)
 ⇒ 飛び出す(agt:人間, plc:ビーチ)

ここで, “人間”に置き換えられた場合には, さらに“男性”, “女性”への詳細化を試みる. 具体的には, 述語および他の引数と“人間”, “男性”, “女性”との共起を調べ, より多く共起するものと置換する. 以下に例を示す.

- 例) 怪我をする(agt:太郎) ⇒ 力仕事をする(agt:太郎)
 ⇒ 怪我をする(agt:男性) ⇒ 力仕事をする(agt:男性)

(2) 述語式の述語が名詞の場合

文型 1 で “inst”ラベルがつく場合であり, この場合は以下の例のように固有名詞を変数(X)に置き換える.

- 例) 社長(inst:太郎) ⇒ 一貧乏(sbj:太郎)
 ⇒ 社長(inst:X) ⇒ 一貧乏(sbj:X)

5. 実験と評価

インターネット上より「にもかかわらず」を含む文 437 個を採取した. このうち知識獲得の対象となる文は 195 文, 対象外の文は 242 文であった.

対象外の 242 文の内訳は, 対比関係を示す文と後文が否定文の文が 86 文, “にもかかわらず”の使い方があまり適切ではない文が 166 文である. 後者には例えば, “私の子供は中学生にもかかわらず使い捨てコンタクトを使用している”のような文も含まれる. この文を 2 章の①とみなしそのまま知識獲得を行うことも考えられるが, 得られる知識は必ずしも常識的といえるか疑問が残る(その人の感覚に大きく依存する). よって, 今回は獲得の対象からはずしたが, この種の文をいかに自動的に識別するかは今後の大きな課題である.

対象となる 195 文について前章までの手法を適用し知識獲得を行った. その結果, 正しく獲得出来たものは 117 文であった. 正しく変換できなかった 78 文の内訳は, 述語知識への変換が共起辞書データの不足により正確にされなかった文 56 文, 前文の内容が係わっているため一文からだけでは人間でも正確に変換できない文が 22 文であった.

6. まとめ

接続詞“にもかかわらず”を含む文から否定的な因果知識を獲得する方法について提案した. 本手法は“にもかかわらず”の前文(P)と後文(Q)から“P→Q”の形式の知識を述語論理形式で獲得するものである. 現段階の正解率は約 60% であるが, 手法の改良により約 90% 程度まで向上可能である.

参考文献

- [1] D.B.Lenat,R.V.Guha,K.Pittman,D.Prattand M.Shepherd:Cyc;Toward Programs with Common Sence,Comunications of the ACM,Vol.33.No.8(1990)
- [2] マインドピクセル, <http://www.mindpixel.com/>
- [3] 佐々木智彦, 増子公德, 野中昌行, 石川勉“国語辞書からの常識知識の獲得”, 情報処理学会第 65 年全国大会講演文集(2), pp137-138(2003)
- [4] 乾 孝司, 乾 健太郎, 松本 祐治, “接続標識「ため」に基づく文章集合からの因果関係知識の自動獲得法”, 情報処理学会論文誌, val45, p919, (2004)
- [5] 佐藤 浩史, 笠原 要, 松沢 和光, “表層的因果知識ベースによる事象推移予測方式”, 情報処理学会第 56 回全国大会, Vol. 2, pp251-252(1998)
- [6] 佐々木 智彦, 石川 勉, “連結定数で統合された素式群による複文の述語知識表現とそれへの変換法”, FIT 2004 第 3 回科学技術フォーラム講演論文集(pp147-148(2004)
- [7] 茶筌:<http://chasen.aist-nara.ac.jp/hiki/ChaSen/>
- [8] 工藤拓, 松本裕治:チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol 43, No 6, pp1834-1842(2002)
- [9] 日本電子化辞書研究所:EDR 辞書 電子化辞書, 日本語共起パターン副辞書, (1996)