

雑談型自由対話システムにおける常識知識の獲得・精練・利用法

高柳 俊祐[†]石川 勉[†]拓殖大学工学部情報工学科[†]

1 はじめに

現状の対話システムはチケット予約等の分野を限定したものがほとんどであり、限定しないシステムの研究例は少ない[1]。人工無脳[2]と呼ばれる対話システムもあるが、それらは基本的に Eliza 型であり、知識処理を行っていない。

我々は、知識を扱いかつ話題を限定しない友達感覚の対話システムの実現を目指し研究している。これまで、推論処理を用いた簡単な質問応答機能、吉凶情報を利用した相槌機能等を実現してきたが、基本的にユーザー主導のシステムであった。これに、特定の話題に対するスクリプト対話機能、常識知識獲得・利用機能を付加し、コンピュータに主導権を持たせられるシステム[3]を新たに開発している。

本報告では、このうち後者の常識知識獲得・精練・利用法について提案する。

2 基本的な考え方

自然な対話の中には常識として利用できるような知識が間接的（無意識）に発話されていることがある。たとえば、「太郎は金持ちだ」「太郎は社長だからね」等の発話で、一般的には「 α は B だよ」「 α は A だからね」のような表現である。これらは“対象の特性や性質を表す文”や“一般的な因果を表す文”でよく見かけられる。ここでは、これらを“A(x)→B(x)”という形式で常識知識として獲得し、さらに対話や質問応答の中で利用する。具体的には、対話では意味的に前件部に一致する発話があるとき後件部を連想し、それを自然言語文に変換し確認文として発話応答する。それに対するユーザの返答は、肯定か否定であり、否定の場合には“どうして？”等の返答をして、それに対する応答文を例外知識として獲得し、常識知識の精練を行う。

以上は間接的な知識獲得であるが、ユーザが直接システムに教え込むような場合もある（以下、教え込みと呼ぶ）。例えば、“社長は金持ちだよ”等の発話である。これについても同様に知識獲得・精練・利用を行う。

3 知識の獲得

ここでは、以下のように知識を述語形式で表現する。

例) “太郎は金持ちである”

⇒金持ち(inst: 太郎) …(1)

引数部には述語との意味的な関係を表すラベルを付加する。このラベルとしては、基本的に述語部が動詞の場合には深層格、名詞の場合には inst、形容詞の場合は sbj のような新たに設定したラベルを用いる。なお、(1)式

Acquisition, Refinement and Usage of Common Knowledge in Open-ended Dialog System

†Shunsuke Takayanagi, Tsutomu Ishikawa

†Department of Computer Science, Takushoku University

への変換は、当研究室で開発している変換プログラム[4]を利用する。

以下、獲得対象とする文および具体的な獲得法を示す。

(i) 対象の特性や性質を表す文

以下のパターンを獲得対象とする。

パターン 1) 発話 1: α は B (だ)

発話 2: α は A だからな

パターン 2) 発話 1: α は A だ

発話 2: R, α は B (だ)

ここで、 α は固有名詞、A は一般名詞、B は特性や性質を表す名詞・形容詞・形容動詞である。R は、『だから』、『それゆえ』、『従って』、『なので』等の理由を表わす接続詞である。

常識知識として獲得するのは、基本的には、A が具体的物を表す名詞（ただし、人称を表す名詞は除く）、B は名詞の場合には抽象を表す名詞および人の能力や社会集団を表す名詞である。また、B が形容詞・形容動詞の場合は感情を表す語以外の場合とした。これらのチェックは日本語語彙体系[5]、IPAL 辞書[6]を用いて行う。

知識獲得は、それぞれの発話を (1) 式の述語形式に変換し、固有名詞 α を変数 x に置き換え、両述語を含意記号で結合することにより行う。具体的には、“A(x)→B(x)”という形式で基本となる常識知識として獲得する。

なお、教え込みの場合は、上記のパターンの発話 1 と発話 2 を結合した “A は B (だ)” という発話パターンを対象とする。知識獲得は、発話文を前記の変換プログラムにより “B(A)” の形に変換し、その後 A、B を個別の述語として前記と同様に含意記号で結合することにより行う。

(ii) 一般的な因果を表す文

一般的な因果を表す文には以下の例があげられる。

例) “株価が下がるね” 「景気が悪いからだよ」

よって、(i) と同様に以下を獲得対象とする。

パターン 1) 発話 1: Y

発話 2: X だからね

パターン 2) 発話 1: X

発話 2: R, Y だよ。

ここで、X は“原因を表わす文”、Y は“結果を表わす文”である。R の接続詞は (i) と同様である。

知識獲得も (i) と同様であり、それぞれの発話を述語形式に変換し、以下の形式で獲得する。

$X' \rightarrow Y'$

ここで、 X' と Y' は X, Y を (1) 式の述語形式に変換したものである。例えば、先の例では “悪い(sbj: 景気) → 下がる(obj: 株価)” となる。

また、教え込みの場合は、X と Y を、含意を表す接続助詞で結合した “X ならば Y” のような発話パターンを対象とする。先の例では “景気が悪ければ、株価が下がる” となる。ここでは、接続助詞としては『ば』、『たら』、『と』、『なら』とした。知識獲得は、接続助詞

表 1 対話例

対話文 (U: ユーザ, C: コンピュータ)	述語知識	知識ベース
① U 太郎は社長だよ	社長 (inst: 太郎)	
② C へー		
③ U だから、太郎は金持ちだよ	金持ち (inst: 太郎)	→ 社長 (inst: x) 一金持ち (inst: x)
④ C そうなんだ		
⑤ U 次郎も社長だよ	社長 (inst: 次郎)	
⑥ C じゃあ、次郎さんも金持ちなんだね	金持ち (inst: 次郎)	← 社長 (inst: x) 一金持ち (inst: x)
⑦ U いいえ	～金持ち (inst: 次郎)	
⑧ C どうして？		
⑨ U 負債が多いんだよ	多い (inst: 次郎, s-obj: 負債)	→ 社長 (inst: x) ∧ ~多い (inst: x, s-obj: 負債) 一金持ち (inst: x)
⑩ C ふーん		
⑪ U 三郎も社長だよ	社長 (inst: 三郎)	
⑫ C じゃあ、三郎さんも金持ちなんだね	金持ち (inst: 三郎)	← 社長 (inst: x) ∧ ~多い (inst: x, s-obj: 負債) 一金持ち (inst: x)
⑬ U いいえ	～金持ち (inst: 三郎)	
⑭ C じゃあ、三郎さんも負債が多いんだね	多い (inst: 三郎, s-obj: 負債)	← 社長 (inst: x) ∧ ~多い (inst: x, s-obj: 負債) 一金持ち (inst: x)
⑮ U はい		
⑯ C へー		

の前後で文を 2 つに分け、それぞれを標準形に直して、前記と同様に行う。

なお、この知識は次章の利用を考慮し (i) とは別の知識ベースに保管する。

4 知識の利用

獲得した常識知識は、対話では (i) を、質問応答では (i) (ii) を利用することとした。対話での利用例を図 1 に示す。

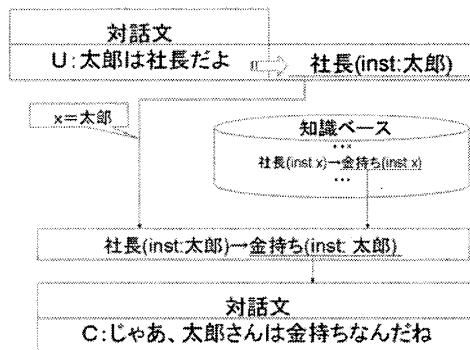


図 1 常識知識の利用例

ユーザが意味的に知識の前件部と一致する文 “A(α)” を発話した場合、獲得した常識知識 “A(x) → B(x)” を用いて、“B(α)” を生成し、それを自然言語文に変換して応答する。この応答は図 1 のような確認発話とする。この応答に対するユーザの返答は肯定か否定のどちらかであり、肯定の場合は相槌を、否定の場合は理由の問い合わせを行う。それに対する応答は例外知識として獲得し、次章のように知識の精練に用いる。なお、肯定否定の判定は、それを表す単語をキーワードとして登録しておき、それを用いて行う。

また、質問応答では、それ用の問題解決知識の一部としてそれに付加し、推論処理の中で利用する。

5 知識の精練

例えば、図 1 の確認発話が否定され、理由として “負債が多い” が発話されたとき、これを用いて “社長 (inst: x) ∧ ~多い (inst: x, s-obj: 負債) → 金持ち (inst: x)” のように知識の精練を行う。

すなわち、理由を表す発話を述語形式に変換し、固有名詞を変数に置き換え、常識知識の前件部に否定形で加える。一般的には以下の知識とする。

$$A(x) \wedge \sim C(x) \rightarrow B(x)$$

“C(x)” が理由を表す部分である。

この精練した知識は対話では以下のように利用する。ユーザが意味的に “A(x)” に一致する文 “A(α)” を発話した場合、まず “B(α)” を用いた確認発話をを行う。この発話をに対し、肯定の場合は相槌を、否定の場合は “C(α)” を用いた確認発話をを行う。例えば、上記の例では、“太郎さんは負債が多いんですか” と応答する。さらに、それがまた否定された場合は、理由を聞き以上と同様に前件部に加え、再度利用する。以上を繰り返して常識知識の精練を図る。

6 対話例

実現している対話例を表 1 に示す。

7 まとめ

対話における発話から常識知識を獲得し、それを利用・精練する方法について提案した。この方法を用いた対話システムは、使っているうちに知識が増加し、いわゆる成長型のシステムとなる。なお、本大会のデモセッションで開発中の対話システム（本手法も内蔵）を公開している。

参考文献

- [1] 柴田雅博、富浦洋一、西口友美：雑談自由対話を実現するための WWW 上の文書からの妥当な候補文選択手法、人工知能学会論文誌 24 卷 6 号 G p507-519
- [2] 人工無脳は考える、
<http://www.ycf.nanet.co.jp/~skato/muno/>
- [3] 久野由隆、石川勉：ちょっと賢いお友達感覚の自由対話システム、情報処理学会設立 50 周年記念（第 72 回）全国大会デモ
- [4] 上條淳史、石川勉：名詞のカテゴリ情報と格助詞の特性を用いた任意格の推定法、第 8 回情報科学技術フォーラム E-005 p257-258
- [5] 池原、宮崎、白井、横尾、中岩、小倉、大山、林：日本語語彙体系、岩波書店
- [6] IPAL 辞書、情報処理振興事業協会