

表層格支配パターンの類似性を用いて応答を選択する対話システム*

1M-03

小磯拓也 池田剛 乾伸雄 小谷善行

(東京農工大学工学部情報コミュニケーション工学科)

1.はじめに

大量の対話事例を使い入力文とその対話事例とを比較して最もマッチしていると考えられる事例の応答にあたる文を出力する自然言語自由対話を行うシステムにおいても重要な処理は、事例のマッチングである。従来のキーワードマッチング[1]では構文情報を使用しないでマッチングを行ってきたが、さらに妥当なマッチング処理を行うために構文情報を加味することを考えた。しかし自由対話で用いられる文の特性により文法などの厳密な構文情報を用いるのには限界がある。そこで我々は、自由対話における文においても見ることができる構文情報の 1 つである表層格支配を用いたマッチングを提案する。これによりキーワードマッチングでは対応しきれなかった係り受け情報も加味され、より柔軟なマッチングが実現可能になった。

2.事例とのマッチング

2.1 表層格支配パターンへの変換

図 1 のように n 個の形態素列からなる形態素列と、格と係り受け元のフレームへのリンクを対とした要素 m 個のリストからなる表層格フレームを定義する。表層格支配パターンはこの表層格フレームが階層的に連結されたものと定義する。

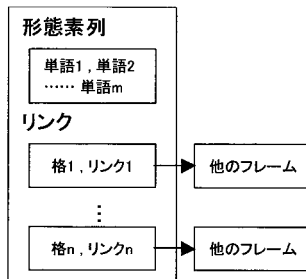


図 1 表層格フレーム

文を係り受け解析しその結果から文節を表層格フレームに、係り受け関係をフレーム間のリンクに変換することで表層格支配パターンに変換する。同じ格に係る場合、例えば「私は林檎が好きだ」のような場合は後の単語を優先する (この場合は「林檎は」が格で繋がる単語となる)。図 2 は文「私は林檎を食べた」を表層格支配パターンに変換した例である。この表層格支配パターンは 3 つの表層格フレームで構成さ

れており、最上位の文節にあたるフレーム A に対するリンクを表層格支配パターンを参照するためのリンクとして利用する。フレーム B と C はフレーム A を係り受け先とする文節であり、フレーム A からそれぞれのフレームの格を関係係名にしたリンクが張られている。

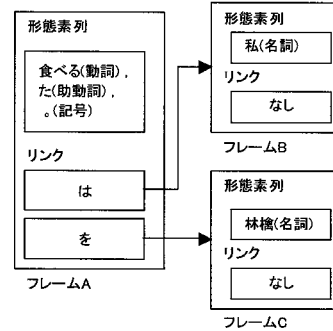


図 2 表層格支配パターンの例

2.2 マッチングの方法

2.2.1 コストについて

表層格支配パターンのマッチングのためにコストを定義する。コストは 2 つの表層格支配パターンがどのくらい似ているかを表す数値で、小さければ小さいほど 2 つの表層格支配パターンが似ていると定義する。同一の表層格支配パターンの比較した場合に最小値 0 を取り、2 つの表層格支配パターンで異なる部分があるほど増加する。

2.2.2 マッチング方法

表層格支配パターンのマッチングは次式で定義される。関数 J と F は 2 つの形態素列を引数とし、関数 C は 2 つの表層格フレームと構文における重要度を表すパラメータ z の 3 つを引数とする。比較したい 2 つの表層格支配パターンへのリンクと z に初期値 s を代入したものを引数として関数 C を呼び出すと 2 つの表層格支配パターン間のコストを返す。関数 J は自立語の比較のために用いられ、単語が等しい場合、品詞が等しい場合、全く違う場合の 3 通りに定義したコストを返す。また関数 F は引数の形態素列を比較し付属語が片方にのみ存在する場合にその種類によって定義しコストを返す。付属語は文の意味にあまり関係ないと考えたためコストは自立語に定義するもよりも小さくしているが、疑問や命令、否定を表すような文の意味に大きく影響すると考えられる付属語に対しては自立語よりも大きくなるようにコストを定義する。関数 L は引数の表層格フレームの持つリンクを調べ、片方にのみ存在する格の数に応じてコストを返

*A dialog system which chooses a response using similarity between a surface case rule patterns
Takuya KOISO, Takeshi IKEDA, Nobuo INUI, Yoshiyuki KOTANI
Tokyo University of Agriculture and Technology Department of Computer, Information and Communication Sciences

す。関数Cは関数J、F、Lを実行して得られたコストをパラメータzで割ったのちに、2つの表層格フレームで共通の格が持つリンクとその引数zを定数p倍したものを引数として再帰呼び出しを行って呼び出して得られたコストを加えて戻り値とする。

$$J(x, y) = \begin{cases} j_1 (\text{単語が等しい}) \\ j_2 (\text{自立語の品詞のみ等しい}) \\ j_3 (\text{品詞も異なる}) \end{cases}$$

$$F(x, y) = \begin{cases} f_1 (\text{付属語が双方に存在する}) \\ f_2 (\text{推論・時制}) \\ f_3 (\text{否定・命令・疑問}) \end{cases}$$

$$L(x, y) = a \times \sum (\text{片方にのみ存在する格})$$

$$C(x, y, z) = \frac{1}{z} \left(\begin{matrix} J(x \text{の形態素列}, y \text{の形態素列}) + \\ F(x \text{の形態素列}, y \text{の形態素列}) + \\ L(x, y) \end{matrix} \right) + \sum_{i, j=x, y \text{で共通した格が持つリンクの対}} C(i, j, z \times p)$$

2.2.2 主語になるフレームへの重み付け

2つの表層格支配パターンで主語にあたる表層格フレームが異なる場合、目的語やそれ以外の単語にあたる表層格フレームが異なる場合に比べて文の意味に与える影響が大きいと考えた。そこで関数Cの再帰呼び出し部分において主語になると考えられる「格が『は』である」である場合には、得られた値をb倍することで主語と思われる表層格フレームに対して重み付けを行う。

2.2.3 コストが閾値を超えた場合

マッチングの結果得られたコストがあらかじめ定めた閾値cを超えた場合は、表層格支配パターンの最上位の表層格フレームが異なりマッチングが失敗したと考えられる。マッチングが失敗した場合返される応答は不自然なものになる事が予想されるため、コストが閾値を超えた場合には登録されている返事からランダムに選択し相槌として使用することで対話を続けられるようにした。

2.2.4 選んだ事例に対するコスト修正

単調な対話になるのを防ぐために、一度答えた応答事例のコストに対して定数dを加える事で(コスト+dとする)連続して同じ応答が返らないようにする。コスト修正は一回応答をする度に定数eずつ減らし最終的には0に戻る。

3 実験結果

システムを実現しテスト文による実験と対話による実験を行った。パラメータは $j_1=0.0$ 、 $j_2=0.8$ 、 $j_3=1.0$ 、 $f_1=0.0$ 、 $f_2=0.05$ 、 $f_3=3.0$ 、 $a=0.05$ 、 $p=10$ 、 $b=3.0$ 、 $c=1.0$ 、 $d=0.5$ 、 $e=0.05$ 、と設定した。

3.1 テスト文による実験

マッチングを行うシステムを実現し、テスト文を用いて実験を行った。その結果構文情報を加味したマッチングができていた事が示された。例えば A:「私は林檎を食べる。」 B:「私はミカンを食べる。」 C:「彼は林檎を食べる。」の3つをそれぞれ比較した場合(AB間のコスト=0.08) < (AC間のコスト=0.24) < (BC間のコスト=0.32)となり、目的語よりも主語に重点を置いてマッチングができていることが分かる。また A:「私は林檎を食べる。」 B:「私はミカンを食べる。」 C:「私は林檎を食べない。」を比較した場合には(AB間のコスト=0.08) < (AC間のコスト=0.3) < (BC間のコスト=3.08)となり、否定文と肯定文との比較ではコストが大きくなっていることが分かる。さらに、A:「私は林檎を食べる。」 B:「私はミカンを食べる。」 C:「今日はいい天気だ。」ならば、(AB間のコスト=0.08) < (AC間のコスト=1.34) = (BC間のコスト 1.34)となり、全く異なる文の比較にはコストが大きくなっているが分かる。

3.2 対話システム

登録事例数 24643 の対話システムを実現し、ユーザとの対話実験を行い 105 事例のデータを得た。そのうち 56 事例では最上位の単語がマッチした場合、24 事例では最上位の品詞がマッチした場合、残りの 23 事例ではまったくマッチしない場合であった。表1の通りコストが0.8より小さい場合は最上位の単語がマッチしていたと考えられ、自然な応答になっているように思われる。格も含めて完全にマッチした場合が1例もなかったが、事例数が少ないため事例数が増えればマッチする事例が出ると考えられる。

表 1 対話結果

応答	コスト	事例	入力
たからなこが	0	そうです。	そうだね。
はあ。	0.04	ないんです。	もう、さすがに話すことがない。
はい。	0.84	そりゃあんまりじゃないの。	どうもこうもない。
それどこかしたの？	15	コンピュータサイエンスです。	あのねえ。

4. おわりに

応答事例をデータとして持ち、それを用いて応答文を返すシステムを作成した。

参考文献

- [1] 江部剛明: 会話データとのキーワードマッチングを行い、応答文を決定する対話システム, 第 58 回情報処理学会全国大会論文集 2, pp.281-282, 1998.
- [2] 吉村賢治: 自然言語処理の基礎, サイエンス社(2000)