

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*

3W-8

江部利明 小島丈幸 乾伸雄 小谷善行
(東京農工大学工学部電子情報工学科)

1 はじめに

実際に行われた会話を大量にデータとして持つ対話システムを設計し、実現した。本システムは、ユーザから文が入力されたら、データ文の中から入力文に一番最も類似した文を探し、入力文に対する応答文を決定する。システムは ELIZA 同様、マッチング以外に知識処理機構を持たない。マッチングは、それぞれの文からキーワードを取り出し、それを比較することで行う。データ文とのヒット率をあげることで自然な対話を目指す。

2 キーワードマッチング

入力文とデータ文とのマッチングには文中のキーワードを用いる。この方法をキーワードマッチングと呼ぶことにする。この章ではキーワードの抽出方法とマッチング方法について述べる。

2.1 キーワードの取出し

文からは名詞、動詞、形容詞の三つのキーワードを取り出す（どれかがない場合やキーワードが一つもない場合もある）。なるべく簡単な作業でキーワードを取り出すようにするため、形態素解析結果を用いる。文中に出てきたすべての名詞、動詞、形容詞に得点を与えて、最も高いものをキーワードとする。

2.1.1 動詞、形容詞の得点

文中に出てきたすべての動詞、形容詞には文字の数と種類によって点数をつける。

文字 $c_1 \dots c_n$ からなる単語を w とする。文字 a の得点 $s(a)$ を次のように決める。

$$s(c_i) = \begin{cases} 1 & (\text{ひらがなのとき}) \\ 1.5 & (\text{カタカナのとき}) \\ 2 & (\text{漢字のとき}) \end{cases}$$

形容詞の得点 $E_a(w)$ 、動詞の得点 $E_v(w)$ は次のように計算する。

$$E_v(w) = E_a(w) = \sum_{i=1}^n s(c_i)$$

例えば、「大きい」は「大」が漢字で2点、「き」と「い」はひらがなでそれぞれ1点で合計4点となる。

2.1.2 名詞の得点

文中に出てきた名詞は文字の数と種類に加え、後ろに接続する助詞、品詞情報によって得点付けされる。

接続での得点を $K(w)$ 、品詞情報での得点を $H(w)$ とし、名詞の得点 $E_n(w)$ を次のように計算する。

$$K(w) = \begin{cases} 3 & (\text{「は」と接続するとき}) \\ 2 & (\text{「が」と接続するとき}) \\ 1 & (\text{「も」と接続するとき}) \end{cases}$$

$$H(w) = \begin{cases} 3 & (\text{固有名詞、複合名詞のとき}) \\ 1 & (\text{一般名詞のとき}) \end{cases}$$

$$E_n(w) = \sum_{i=1}^n s(c_i) + K(w) + H(w)$$

2.2 マッチングの方法

2.2.1 キーワードが一つもない場合

キーワードが一つもない文とは、たとえば、「こんにちは。」や「はい。」などの文である。このように、入力文にキーワードとなる語がない場合、同じくキーワードを持たないデータ文と字面でのマッチングをとる。つまり、文字列として一致するかどうかを判断する。一致するものがない場合、キーワードを持たないデータ文をランダムに一つ選び、それを最も類似した文とみなす。

2.2.2 キーワードがある場合

この場合のマッチングを、「大きい書店ならもっと良い本が売っているかもしれない。」という入力があったときを例にして説明する。まず、2.1 で述べた得点付けに基づいてキーワードを取り出す。なるべく事例データとマッチさせるため、入力文からはキー

*A Natural Dialogue System using Case Based Data, Toshiaki EBE, Takeyuki KOJIMA, Nobuo INUI, Yoshiyuki KOTANI, Tokyo University of Agri. and Tech., Dept. of Computer Science.

ワードを二つ取り出す。この文のキーワードは次のようになる。

- ・動詞キーワード ①売っ
- ・名詞キーワード ①本 ②書店
- ・形容詞キーワード ①大きい ②良い

そして、次の①～⑦の順番でマッチするデータ文を検索する。

- ① 取り出したキーワードのすべての組み合わせ
(売っ, 本, 大きい)、(売っ, 本, 良い)、
(売っ, 書店, 大きい)、(売っ, 書店, 良い)
- ② 形容詞キーワードは何でもいい
(売っ, 本, 〃)、(売っ, 書店, 〃)
- ③ 名詞キーワードは何でもいい
(売っ, 〃, 大きい)、(売っ, 〃, 良い)
- ④ 動詞キーワードは何でもいい
(〃, 本, 大きい)、(〃, 本, 良い)、
(〃, 書店, 大きい)、(〃, 書店, 良い)
- ⑤ 動詞キーワードだけあっている
(売っ, 〃, 〃)
- ⑥ 名詞キーワードだけあっている
(〃, 本, 〃)、(〃, 書店, 〃)
- ⑦ 形容詞キーワードだけあっている
(〃, 〃, 良い)、(〃, 〃, 大きい)

3 応答文の生成

3.1 マッチするデータがある場合

3.1.1 キーワードを入れ替える

入力文とデータ文のキーワードが入れ替えられるときは入れ替えたものを応答文とする。例えば、「ご飯が食べたい。」(食べる、ご飯、(なし))という入力文があり、事例中の「うどんが食べたい。」(食べる、うどん、(なし))とマッチしたとする。このデータ文に対する応答文が

「うどんはいやだ。」((なし)、うどん、(なし))
の場合、ご飯とうどんを入れ替え、
「ご飯はいやだ。」((なし)、ご飯、(なし))
という文を作る。

3.1.2 データをそのまま出力

キーワードが入れ替えられないときはデータにある文をそのまま応答文とする。先ほどの例で、データ文とその応答文が

「うどんが食べたい。」
「いいね。」

であった場合、

「いいね。」
を応答文とする。

3.2 マッチするデータがなかった場合

2.2 で説明したように、入力文はなるべくデータ文とマッチさせるようにしているが、マッチするデータがない場合がある。このとき、事例データの中からランダムに一つの文を選び、「ところで、」の後ろにつけた文を応答文とする。これにより、話題を変える役目をする。

4 対話システムの実現と使用実験

システムを実現し、東京農工大学繊維博物館主催の子供科学教室に出展し、1998年12月、小中学生約60名を対象に2回使用実験を行なった。システムにはあらかじめ435個のデータを持たせておいた。システムの全発話1291のうち、マッチするデータがある場合の発話が1130あり、ヒット率は約87.5%であった。データの数を増やせば、さらに高い数字が期待できると思う。

5 終わりに

会話データを持ち、文生成に利用する対話システムを作成した。

今後の課題としては、

- ・ヒット率をあげる。
→データの数を増やす。
→マッチング方法を改良する。(シソーラスを用いた方法など)。
- ・話題管理を行い、対話の内容がそれないようにする。
などがあげられる。

参考文献

- [1] 長尾真編：自然言語処理, 岩波書店, (1996).
- [2] 佐川ほか：柔軟な対話制御機構を持ったコンサルテーションシステム, 情報処理学会論文誌, Vol.29, No.4, pp. 350-358, (1988).
- [3] 佐川ほか：自然言語対話システムにおけるユーザモデルの更新に伴う対話の際プランニングに関する考察, 情報処理学会論文誌, Vol.35, No.6, pp. 1042-1049, (1994).
- [4] 美馬ほか：類似検索を用いた情報検索システム, 言語処理学会第2回年次大会発表論文集, pp. 113-116, (1996)