

自然言語対話における発話話題の自動生成*

梶野 博仁、唐澤 博†

山梨大学 工学部‡

E-mail: {kajino, karasawa}@jewel.yamanashi.ac.jp

1 はじめに

本研究は、自然言語対話において図1の対話文の例に示すようなシステムとユーザとの対話で一つの話題について、ほぼ話し終わってしまっていると考えられる場合、次にシステムがどのような話題を発話すれば良いか、また、対話がとぎれないようにするためシステム側から話題の展開を行えるようにすることが目的である。

2 本研究モデル

まず、本システムの起動条件に当てはまる場合、図に示すように、ユーザの入力文より文脈情報の集合として定義する名詞を抽出する。ここで、意味的曖昧性解消に用いられるコネクショニスト・モデルの一つであるマーカパッシング [1] の手法を用いる。

マーカパッシング [1] は、辞書を意味ネットワークに変換し、対話文に登場する単語に対応するネットワーク上のノードに活性値を与え、活性伝搬によってその対話文中での各語の語義を選択する手法である。こうすることにより、対話文の文脈に関連度の大きな単語を抽出することができると考え、本研究モデルへの適用とした。

辞書より意味ネットワークを構築し、マーカパッシング [1] で次の発話文のキーを抽出し、知識ベースより発話文を決定する。

3 本システムの起動条件

本システムの起動条件は、以下のように定義する。

- (1) Bが現在の話題に対し、発話する内容が無くなった場合
 - 現在対話している内容について、知識ベースを検索しても何も無くすべてを話してしまった場合
- (2) Aが前の対話対で怒った場合

* Automatic topic generation for natural language dialog

† Hirohito Kajino, Hiroshi Karasawa

‡ Yamanashi University, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

対話文の例

A: もうすぐ夏休みですね。何か予定はあるんですか?
 B: バイクで北海道を横断しようかなと思ってます。
 A: ヘー、私は予定もないので、毎日バチンコですよ。
 麻雀はやりますか?
 B: やりますよ。北海道から帰ったら一緒にやりましょう。
 A: そうですね。連絡しますよ。

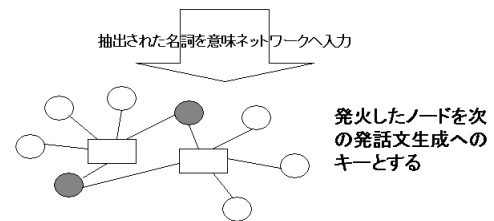


図 1: システム概要

- 対話相手が怒った場合、話題を展開することが対話を続ける一つの手段と考えられるが、システムが対話相手の感情を評価しなくてはならない。(例えば、対話相手が攻撃行動してきた場合など)

本研究では、(1)のみを扱うものとする。(2)は他研究 [2] に譲る。

4 名詞の抽出

まず、文脈より意味ネットワークへの入力候補となる名詞を抽出する。文脈とは、本研究では現在の対話文より前5文と仮定する。それぞれの対話文を形態素解析ツール「茶筌」 [3] を用いて、名詞のみの抽出を行う。

5 辞書

本研究では、意味ネットワークを構築するにあたり先行研究の話題連想辞書 [4] を用いる。話題連想辞書とは、インターネットの掲示板のカテゴリによる、95個の話題ラベルとインターネットの掲示板のトピックス名を使用したそれぞれの話題ラベルに関連度の強い単語(共起語)群の話題要素からなる共起辞書である。

6 次の発話文へのキーワードの自動生成

文脈より抽出された名詞と話題連想辞書を用いて、マーカパッシング [1] というコネクショニスト・モデルを行う。マーカパッシングとは、意味ネットワークに活性値を伝搬させてノードを発火させる手法である。

6.1 意味ネットワークの構築

意味ネットワークの構築は、図 3 に示すように話題連想辞書の話題ラベルより話題ラベルノードを、話題要素より話題要素ノードを作成しそれぞれ関連する話題ラベル、話題ノードへ活性リンクを張る。

本研究では、互いに相反すると考えられるノードが無い場合、抑制リンクによる結合は行わない。

6.2 活性値の伝搬

文脈から抽出した名詞が話題連想辞書の話題要素に含まれていれば、意味ネットワークの話題要素ノードに活性値を与える。活性値の与えられた話題要素ノードから活性リンクの張られたノードへ減衰率 n の活性値を持ったマーカを伝搬する。マーカが伝搬されたノードは、受け取ったマーカの数だけさらに n 減衰した活性値のマーカを活性リンクの張られたノードへ伝搬する。このような伝搬作業を繰り返し行い、伝搬される活性値がある閾値を下回ったら伝搬終了となる。

6.3 キーワードの決定

活性値の伝搬が終了したら、意味ネットワーク全体の話題要素ノードのうち活性値上位の話題要素を次の発話へのキーワードとする。ここで、本研究の目的は「話題の展開」にあるため、新たな話題（文脈に登場しない）の要素を抽出したい。よって、文脈にある対話で話し合われた話題と考えられる活性値第 1 位の話題ラベルの要素は削除する。これにより、「話題の展開」と定義する。

また、意味ネットワークへの入力に前の文の情報を扱うことにより、文脈を考慮した次の発話文へのキー

ワードとなる。

6.4 実験結果

活性伝播実験結果は、下の表 1 へ掲載する。実際に次の発話のキーワードとなるのは、出力話題要素の上位の単語となる。出力話題ラベル第 1 位の「読書・書籍・雑誌」の話題要素は候補から削除済である。今回の結果では、次の発話のキーワードは、良いものができていると考えているが、主観的な評価であるので、アンケート調査などで一般的な評価を行う。

7 まとめ

今回は、条件付きでシステム側から話題をふるための発話のキーワードを自動生成した。主観的な評価ではないので、一般的な評価を行っていく。その後、キーワードを用いて発話文を生成していく。

8 今後の課題

- アンケート調査結果でのシステムの客観的評価
- 発話文の生成
- 実データでの実験
- 活性伝播において、無駄な計算があるかの検討

参考文献

- [1] 長尾 真, 自然言語処理, 岩沼書店, 1996.
- [2] 上野, 唐澤: 感情モデルに基づく対話文の解析と生成, 山梨大学大学院工学研究科電子情報工学専攻修士論文, 2002.
- [3] 奈良先端科学技術大学院大学, 茶筌.
- [4] 笠原, 唐澤: 話題連想辞書の開発と談話理解への適用, 山梨大学工学部コンピュータメディア工学科卒業論文, 2002.

表 1: 実験結果

対話文 1	システムの出力話題			
	出力話題要素	活性値	出力話題ラベル	活性値
A: 最近、英語の勉強を頑張っています。 B: 英会話学校へ通っているんですか？ A: いいえ。 英語で書かれた村上春樹の小説を読んでいるんですよ。 B: それはいいですね。 洋画を観ることも、聞き取りにはいいらしいですよ。 A: そうですか。今度やってみます。	HP 心 教室 英語 小学生 勉強 子供 教育 通信 教師 絵本 クラブ 恋 女	55.901367 48.675781 45.747070 43.319336 43.006836 41.705078 38.660156 37.552734 37.552734 37.552734 37.104492 36.081055 35.313477 33.246094	読書・書籍・雑誌 英語・外国語・留学・資格とお勉強 日記 育児/パパ・ママ 創作 小説 学校・教育 まじめな話題いろいろ/ボランティア・災害対策 競馬・宝くじ・ギャンブル おもちゃ・キャラクター・コレクター モバイル・携帯電話・次世代 ファミリー・世代別 友達募集 クラシック・オーケストラ・合唱団	67.841797 49.496094 44.507812 36.517578 34.126953 27.119141 25.609375 15.035156 13.218750 10.306641 9.044922 8.935547 8.292969 8.179688