

対話文脈解析における場面同定手法に関する研究*

藤田 明日香, 唐澤 博†

山梨大学大学院‡

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

1 はじめに

自然言語対話システムにおいて、計算機と人間のコミュニケーションが円滑に行われることは大きな課題である。本研究では、話されている話題の中の場面という点を対話文脈解析に適用することを目的として研究を行っている。これまでに動的な入力文における場면을推論、同定するシステムを開発し、そのシステムの場面同定能力が人間に類似するという結果を得た [1]。場面という点からの文脈生成が可能になった。現段階としてはシステムの結果をより人間に近づけるための改良作業を行い、またそれを対話文脈解析に適用した場面情報適用型形態素解析へ応用する試みを行っている。

2 研究の位置づけ

英単語の曖昧性解消を目的とした、ある 1 場面の構成要素群のみからの静的な場面同定を実施した研究がある [2]。彼らは人間の推論の基本的部分として、神経回路網と論理記号処理による制御および検証を目的とした PDAI & CD (Parallel Distributed Associative Inference & Contradiction Detection) アーキテクチャ [3] を基に実現されたシステム WAVE (Winner Associative Voting Engine) [4] を実装している。本研究はこの研究に基づき、さらに拡張したものである。

3 場面連想辞書の再構築

これまで使用してきた場面連想辞書 [1] は専門的な場面を多く含み、実際の会話にそれらの語が出現することはあまりなく、1 テキスト中の全名詞数に対する場面連想辞書の語の割合は 12 % 程度であった。また異表記の問題なども考慮し、場面ラベルも含めたこれまでの場面連想辞書の語を EDR 日本語単語辞書 [5] の既存の概念番号に 1 対 1 対応で置き換える作業を行った。ただし場面ラベルは概念重複不可のため 1 対多も可とし

*Research on the scene identification technique in dialog context analysis

†Asuka Fujita, Hiroshi Karasawa

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

た。再構築した場面連想辞書は (a) 概念レベル (b) 語彙レベルでは表 1 に示す通りである。

表 1: 場面連想辞書

	(a)	(b)
全場面数	134 場面	134 場面
重複ありの全要素数	5244 概念	9211 語
重複なしの全要素数	3283 概念	8621 語
1 場面あたりの平均要素数	39.13 概念	68.74 語
1 要素あたりの平均所属場面数	1.60 場面	1.07 場面
1 概念あたりの平均語彙見出し数	-	2.63 語

4 場面同定システムの改良

本研究において開発した場面同定システム [1] は図 1 に示す通り、1 文入力毎に、場面連想辞書に含まれる名詞を抽出、活性伝搬して場면을同定し、設定値より下の活性値は足切り、残った活性値には 1 未満の係数をかけ、コピーバックするという処理プロセスであった。

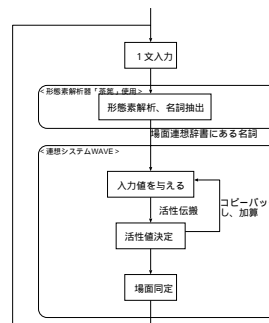


図 1: 場面同定システムの処理概要

これまでに連想システム WAVE の主な計算部分を一部変更したので以下にその変更した点を示す。

4.1 活性伝搬

これまで場面ラベルの活性値計算は場面要素の活性伝搬終了後、その結果を元に行われていた。そのため入

力文に場面ラベルが語として陽に出現しても全く考慮されていなかった．そこで先行研究の計算方法 [2] に変更し，場面ラベルも場面要素群とともに発火させることにした．ただし場面ラベルが文中に出現したときはその場面が発火しやすくなるように，その入力活性値を 2 倍する．また入力文に入力単語がないときは前の文脈をそのまま継続することが適当だと考え，リンク数で割らない値を自分にだけ伝搬することにした．これにより，活性低下のみ行うことになる．さらに場面ラベルは概念が複数対応することがあるが，その場合は平均値をその場面ラベルの活性値とする．

4.2 コピーバック係数

これまでコピーバック係数を固定ではなく動的に設定する方法を提案し，実験を行ってきた [1]．しかし場面が切り替わるときの係数がほぼ 0 になり文脈をほとんど消してしまうという問題点が解決できなかったため，固定値に設定し直した．

4.3 足切り処理

これまで計算時間短縮のため足切り処理を行っていた [1]．しかし足切り処理による実験結果への悪影響が出たため，さほど計算時間の短縮にならないこの方法を撤回し，足切り処理を行わないことにした．

5 実験

これまでに変更してきた点を踏まえ，実データ 3 つ [6] を用いてコピーバック係数 0.9 で場面同定実験を行った．使用したテキストと実行結果について表 2 にまとめる．ただし，場面同定妥当率とは 1 文毎に，同定された場面が妥当であるかないかを主観的に判断し，全入力文数からみて算出した割合である．

表 2: 使用したテキストと実行結果

	実データ 1	実データ 2	実データ 3
入力文数	36 文	46 文	127 文
全名詞数	92 語	71 語	207 語
連想辞書所属名詞数	11 語 (12%)	18 語 (25%)	82 語 (40%)
場面ラベル出現数	2 語 (2%)	11 語 (15%)	36 語 (17%)
場面同定妥当率	69%	33%	66%

5.1 考察

実データ 2 の場面妥当率が低い理由としては，形態素解析の時点での誤りと，システムに存在しない場面について多く語られていることが挙げられる．前者の問

題は，今後場面情報適用型形態素解析システムによって解決されるのではないかと期待している．辞書の再構築により語ではなく概念レベルで発火させることになり，表記の揺れ，同義語の問題などが解消された．しかし入力テキスト中で発火する語が増えたことで，予想していなかった語が発火し，期待通りの結果にならない場合もあった．また場面ラベルが文中に出現するとその場面が選択されることがほとんどであったが必ずしもそれが期待通りであるとは言えない．場面ラベルへの概念対応付けに問題があると考えられる．コピーバック係数については，以前よりも期待するような結果が得られ，妥当な値として 0.9 が求まった．

6 場面情報適用型形態素解析への応用

本研究で開発，改良した場面同定システムの出力を形態素解析に適用して，場面文脈に依存した妥当な形態素解析結果を得る試みを行っている．そのために，筆者等が開発した最小コスト法に基づく既存の形態素解析サーバを改造し，このコスト値を場面同定システムの出力で修正する仕組みを実装した．今後，多量の実テキストデータを与えて期待する効果が得られるか評価する．

7 まとめ

これまでは自作データを実験に使用し，それなりの成果を出したが，実データを使用してみると，場面同定能力が人間とあまり一致していないことがわかった．テキストによらず場面同定妥当率を上げることと形態素解析への応用について評価することが今後の課題である．

参考文献

- [1] 藤田，唐澤．場面連想辞書の開発と談話理解，情報処理学会第 65 回全国大会，Vol.2, pp159-160, 2003.
- [2] 角田，田中．辞書ベース連想による場面同定に必要な文脈情報量の推定，情報処理学会第 48 回全国大会，Vol.3, pp171-172, 1994.
- [3] 角田，田中．*PDAI* & *CD* に基づく意味の学習および文脈依存の多義性解消，電子情報通信学会技術研究報告，Vol.DE93-1, pp1-8, 1993.
- [4] 角田，田中．汎用並列システム *WAVE*，電子情報通信学会技術研究報告，AI92-39, Vol.92, No.184, pp17-26, 1992.
- [5] EDR 日本語電子化辞書研究所．*EDR* 日本語単語辞書，EDR 日本語電子化辞書研究所.
- [6] 音声情報学会．対話データコーパス，音声情報学会.