E-045

マルチモーダル対話における発話意味役割推定

Semantic Role Analysis of the Utterances in a Multimodal Dialog

鈴木 優† 福井 美佳† 藤井 寛子† 宮澤 隆幸† 浦田 耕二† 住田 一男†
Masaru Suzuki Mika Fukui Hiroko Fujii Takayuki Miyazawa Kouji Urata Kazuo Sumita

1. はじめに

ブロードバンドネットワークや第 3 世代携帯電話の普及によりマルチメディアコミュニケーションが日常的に行なわれるようになってきた．現在はコミュニケーションそのものを楽しむためにマルチメディアを利用することが多いと考えられるが，今後は e-Learning や消費者相談など知識の伝達を目的とした場面でもマルチメディアコミュニケーションが主流になるだろう．マルチメディアコミュニケーションで伝達される知識は，知識共有の視点で言えばフロー情報にあたり，これを整理，洗練してストック情報とすることで知識コンテンツとして共有できる．しかしながらマルチメディアデータの編集は一般の利用者にとっては難しく，時間がかかるという問題がある．

筆者らは映像，音声とテキストを構造化した知識コンテンツ (マルチモーダルナレッジコンテンツ) にオンデマンドにアクセスできる MKIDS(Multimodal Knowledge and Information on Demand Service) の開発を行なってきた [1]．これまでの研究で知識コンテンツのオーサリング支援が重要な課題であるとの知見を得ている [2]．本研究ではマルチメディアコミュニケーションのデータからマルチモーダルナレッジコンテンツへの変換を支援するため，教師と学習者の間の質疑応答型の対話について，各発話の意味役割 (挨拶，質問，回答，相槌など) を意味役割間の遷移確率を考慮して推定する技術を開発した‡．将来的には推定した意味役割に基づいて対話を要約，整理し，マルチモーダルナレッジコンテンツの生成を支援すること目指している．本稿では開発した意味役割推定手法について説明し，対話から書き起こしたテキストデータを用いた実験により本手法の有効性を検討する．

2. 関連研究

対話の意味役割構造については 1980 年代から多くの研究がなされており [3]，近年では音声対話を対象とした発話意図分類に関する研究も報告されている．音声対話の場合，音声認識の結果から発話意図の分類を行なうが，認識結果には誤りが含まれることがあるため，韻律情報を手がかりとするなどのアプローチがとられている [4]．

コミュニケーションのデータからのコンテンツ生成を支援する研究としては，メールによるコミュニケーションの内容をスケジュールや Q&A といった目的に応じて集約する研究 [5] や，カメラとセンサを利用して話者の自然な動作を認識することで日常会話からの知識獲得を目指す研究 [6] などがある．

†(株) 東芝 研究開発センター 知識メディアラボラトリー
212-8582 川崎市幸区小向東芝町 1
Tel: 044(549)2240, Fax: 044(520)1308
Mail: masaru1.suzuki@toshiba.co.jp

‡本研究は情報処理振興事業協会平成 14 年度次世代ソフトウェア開発事業の委託により実施した．

実演	動作や操作手順など実演を交えながら説明する
アドバイス	方法についてのワンポイントアドバイスを実演を交えずに説明する
製品紹介	お薦めの物について紹介する
スポット紹介	お薦めの場所について紹介する

表 1: 収集した対話の課題

挨拶	挨拶やお礼の言葉
質問	回答を期待した問いかけと補足説明
回答	質問に対する答えと補足説明
相槌	相手の発言に対する合いの手
確認	相手の発言への復唱や要約
演示	実演を交えた説明
その他	上記に当てはまらない独り言など

表 2: 意味役割一覧

3. 対話データの収集と分析

分析のため 32 人の被験者による対話データの収集を行なった．対話は 2 人ずつ行ない，対話毎に予め教師と学習者の役目を与えておく．対話は質疑応答型とし，表 1 に示した課題のいずれかの話題について自由に対話してもらう．こうして得られた 121 件の対話データ (平均長 2 分 51 秒) について書き起こしテキストを作成した．同一話者の発話であっても 1 秒以上の間隔が開いた場合には次の発話とし，書き起こしテキストには発話時刻及び発話継続時間も記録した．書き起こしデータのうち 71 件 (2,911 発話) を訓練データとして分析に用いた．

訓練データの分析から質疑応答型の対話に特徴的な 7 種の意味役割を定義した (表 2)．訓練データに含まれる各発話をいずれかの意味役割に人手で分類し，これを正解として以下の分析を行なう．正解を付与した対話データの例を表 3 に示した．

4. 意味役割の推定

前述のように音声対話の意味役割推定に関する従来研究では音声認識誤りに対応するため韻律情報を手がかりとする方法が提案されている．本研究では対話における発話の意味役割の遷移に着目し，遷移確率に基づき最尤な意味役割遷移系列を選択することで，音声認識誤りの発生した発話についても意味役割を推定するというアプ

時刻	時間	発話者	意味役割	発話内容
0:03	0:02	学習者	挨拶	はい。こんにちはー。
0:05	0:00	教師	挨拶	こんにちはー。
0:06	0:04	学習者	質問	えーと、折り紙で、オルガンの折り方を、教えてください。
0:10	0:11	教師	演示	はい、わかりましたー。じゃあ、このように、しかくい、こ、なんていったらいいんでしょうねー、ちゃんと辺がこのように、平行にくるように持ちましてー
0:22	0:00	学習者	相槌	はい。
0:22	0:05	教師	演示	これを、この上下のを合わせて、細長く折ります。
0:31	0:02	教師	その他	伊藤さんのほうが早い。はは。
0:34	0:07	教師	演示	そうしましたらー、細くなったのを、またふたつに、折って、しかっけいにしてください。
0:42	0:00	学習者	相槌	はい。
0:45	0:04	教師	演示	え、折りましたのを、もう一度開いてください。えー、細長い状態まで。
0:48	0:00	学習者	相槌	はい。

表 3: 対話データの例

ローチをとる．本手法は以下に説明する 2 段階の処理からなる．

4.1 パタンマッチによる重み付き分類

正解を付与した訓練データを人手で分析し，各意味役割毎に特徴的と考えられる表現 (文字列パタン) を抽出した．得られたパタンは重み付きで利用するため，あまり絞り込まずに多くのパタンを集めることとし，前述の訓練データ (2,911 発話) から約 1,300 のパタンを抽出した．

発話者の役割 (教師又は学習者)r_iが与えられた時のパタン e_jと意味役割 s_kとの関係の強さ w_{ijk}を，r_iの発話における s_kの出現数 N_{ik}と，意味役割 s_kである発話におけるパタン e_jの出現数 n_{ijk}によって式 (1) のように定義する．訓練データから式 (1) によって求めた w_{ijk}の一部を表 4 に示した．

$$w_{ijk} = \frac{n_{ijk}}{N_{ik}} \tag{1}$$

各発話 u_lはすべてのパタン e_jと比較され，パタンが適合するとそれぞれの意味役割 s_kと対応するスコア S_{kl}に w_{ijk}が加算される．こうして各発話がそれぞれの意味役割に重み付きで分類される．

4.2 遷移確率による最尤意味役割系列の選択

役割 r_{i_1}の発話者による意味役割 s_{k_1}の発話から役割 r_{i_2}の発話者による意味役割 s_{k_2}の発話に遷移する確率 $p_{i_1k_1i_2k_2}$は，訓練データにおける役割 r_{i_1}の発話者による意味役割 s_{k_1}の発話数 $N'_{i_1k_1}$と，そのうち r_{i_2}による s_{k_2}の発話へ遷移した発話数 $n'_{i_1k_1i_2k_2}$によって

$$p_{i_1k_1i_2k_2} = \frac{n'_{i_1k_1i_2k_2}}{N'_{i_1k_1}} \tag{2}$$

と表せる．ここで同一話者 ($i_1 = i_2$) の発話が連続してもよい．また各意味役割が対話の先頭に起きる確率と最後に起きる確率についても同様に求められる．$p_{i_1k_1i_2k_2}$の例として訓練データにより求めた学習者の発話から教師の発話へ遷移する場合の遷移確率を表 5 に示した．例えば学習者 (先行発話) の「質問」に引き続いて教師 (後続発話) が「相槌」を発話する確率は 0.55 である．

前節のパタンマッチによって求めた発話毎の各意味役割に対するスコアと，各意味役割間の遷移確率とを用い，ビタビアルゴリズム [7] により最尤な意味役割系列を求める．ただしパタンマッチにおいて全ての意味役割でスコアが 0 となった発話があると全系列の確率が 0 となってしまうため，そのような発話については各スコアに等しい値を与えてからビタビアルゴリズムを適用する．

5. 実験と考察

実験のため，訓練データとは異なる 25 人の被験者により 20 件 (1075 発話) の対話データを収集した．収集した対話データから書き起こしテキストを作成し，人手によりいずれかの意味役割に分類したものを評価データとした．

音声認識誤りが発生すると発話にパタンが適合する割合が下がると考えられる．そこで音声認識誤りに対する遷移確率適用の有効性を調べるため，パタンマッチに用いるパタンの数を無作為に $\frac{3}{4}, \frac{1}{2}, \frac{1}{4}, 0$ と減じてそれぞれの場合の意味役割推定の正解率を求めた．図 1 に，訓練データを処理した場合 (closed) と評価データを処理した場合 (open) のそれぞれについて，パタンマッチによる意味役割分類で発話毎に最もスコア S_{kl}の値が大きい意味役割を選択した場合 (PM) と，遷移確率を考慮して得られた最尤系列によって各発話の意味役割を決定した場合 (TP) の，用いたパタンの数と正解率との関係を示す．

図 1 によるとパタン数が 0 であっても遷移確率による推定は open データで 50%以上の正解率となっており，質疑応答型の対話が定型的で意味役割の遷移に着目するのが有効であることが確認できた．

また，closed/open 共にパタン数が $\frac{1}{4}$程度まで少なくなるとパタンマッチによる分類の正解率が急に低くなっているが，この場合でも遷移確率による推定は比較的高い正解率を維持している．このことから本手法が音声認識誤りが発生した場合でも安定して意味役割を推定できると期待できる．

次に全てのパタンを利用した場合の closed と open それぞれにおける正解率の比較を表 6 に示した．遷移確率を利用した場合 (TP) に open は closed よりも正解率が 16 ポイント程低くなっている．しかし遷移確率を利用したことによる利得率は open の方が大きく，質疑応答型の対話における意味役割の遷移確率がデータの違いに対しても安定していることがわかる．一方で open データに対するパタンの網羅性の低さが課題であり，従来研究と同様に韻律情報を併用するなどの改善が必要となると考えられる．

6. まとめ

マルチメディアコミュニケーションのデータから知識コンテンツへの変換を支援するため，教師と学習者の間

	学習者							教師						
	相槌	挨拶	質問	回答	確認	演示	その他	相槌	挨拶	質問	回答	確認	演示	その他
あのー	0.01	0	0.14	0.20	0.6	0	0	0	0	0.14	0.07	0	0.05	0
ください	0	0	0.04	0.01	0	0.15	0.05	0	0	0	0.01	0	0	0.33
こうなって	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0
そうですね	0.01	0	0.01	0.07	0	0	0.05	0.02	0	0.14	0.02	0	0.05	0
でしょうか	0	0	0.12	0	0.02	0.08	0	0	0	0.07	0.01	0	0	0

表 4: パタンと意味役割の関係

		後続発話						
		相槌	挨拶	質問	回答	確認	演示	その他
先行発話	相槌	0.09	0.01	0.01	0.66	0	0.11	0.01
	挨拶	0.29	0.53	0.01	0	0	0	0
	質問	0.55	0	0.02	0.35	0.01	0.05	0.02
	回答	0.07	0	0	0.87	0	0.07	0
	確認	0.34	0	0	0.61	0	0	0
	演示	0.29	0	0	0.32	0	0.39	0
	その他	0.14	0	0	0.17	0	0.27	0.31

表 5: 学習者から教師への発話の遷移確率

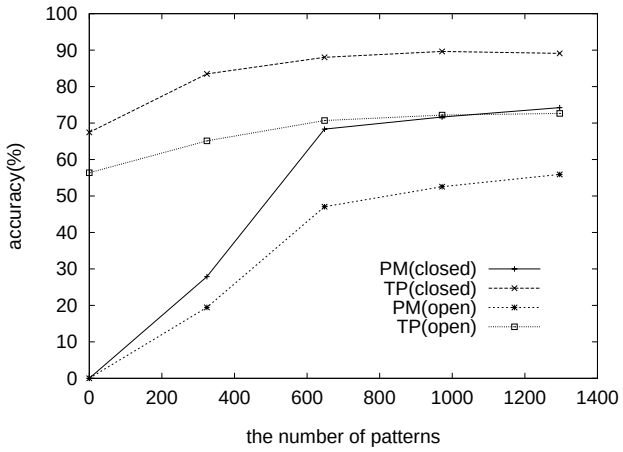

図 1: パタン数と正解率

の質疑応答型の対話について各発話の意味役割を推定する技術を開発した．実験の結果，パタンマッチが十分に機能しない場合であっても，各意味役割間の遷移確率を考慮して最尤意味役割系列を推定することで安定した精度を維持できることが示された．このことから音声対話で音声認識誤りが発生する場合にも本手法が有効であることが期待できる．

今後実際に音声認識と組み合わせた実験を行ない音声対話における本手法の有効性を確認する．さらにマルチメディアコミュニケーションを通じてマルチモーダルナレッジが生成/蓄積でき，検索/配信まで行なえる総合的なマルチモーダルナレッジ共有システムを開発し，熟練工の技能伝承や教育などの現場で実証実験を行なっていく．

	正解率 (PM)	正解率 (TP)	利得率
closed	74.23%	89.12%	20.06%
open	55.91%	72.65%	29.94%

表 6: 遷移確率利用の効果

参考文献

[1] 宮澤, 他, Bluetooth 機器に映像・ナレッジをオンデマンドで配信するシステム MKIDS, インタラクション 2002 予稿集, pp.51-52, 2002.

[2] 鈴木, 他, マルチモーダルナレッジ技術の展示案内システムへの適用, 人工知能学会誌, Vol.18, No.2, pp.177-182, 2003

[3] 石崎, 伝, 談話と対話, 東京大学出版会, 言語と計算 3, 2001

[4] 久保, 松居, 岡本, 協調学習環境における音声による議論の発話意図の分類に関する研究, 人工知能学会研究会資料, SIG-SLUD-A202-01, pp.1-6, 2002

[5] 原口, 梅木, 横田, メッセージ集約型コミュニティウェア GroupScribe, インタラクション 2003 予稿集, pp.229-230, 2003

[6] 山本, 坂根, 竹林, ユビキタスミーティングからのマルチモーダル知識獲得に関する研究, インタラクション 2003 予稿集, pp.73-74, 2003

[7] 北, 確率的言語モデル, 東京大学出版会, 言語と計算 4, pp.113-114, 1999