1T-06

マルチモーダル対話システム HermesⅢの開発

スホルクワイク　亜蘭　　谷津　元樹　　原田　実
青山学院大学理工学部情報テクノロジー学科

1　はじめに

近年、GoogleAssistant や Siri などの対話型インタフェースや Alexa や Clova などのチャットボットといった雑談対話型システムが普及している。しかし、これらのシステムには質問応答ができるものが少なく、また雑談がテンプレート的なものが多く飽きがきてしまう。本研究では、ユーザの発話を質問か雑談かどうかを判断し、適切な応答を Agent の身振りを伴って発話し、話し相手の様子を理解して話しかけもできる自然な対話を実現したマルチモーダル対話システム HermesⅢを開発する。

2　システムの構成

システムの起動からシステム発話までの流れは図 1 にも示すように以下のような手順になる。

1. エージェント側からユーザへ発話を促すシステム発話をする。
2. ユーザの発話待ち状態。
3. ユーザ発話が行われた場合、Hermes[1]による雑談応答または Metis[2]による質問応答によってシステム発話を生成する。
4. ユーザ発話を感情分析システム STM[3]により感情分析し、ユーザ感情を獲得する。
5. 獲得されたユーザ感情と生成されたシステム発話を、本システムの GUI である MMDAgent[4]が読み取れるコマンドに変換する。これによって 3 章で述べるようにユーザ感情に従って表情や身振りを変えさせる。
6. ユーザ発話待ち状態になり 2.へ戻る
7. ユーザ発話が一定時間行われなかった場合、カメラ撮影で画像を取得し、DeepWatcher2[5]を用いてキャプション生成を行い、話しかけをして 5.へ移動する。

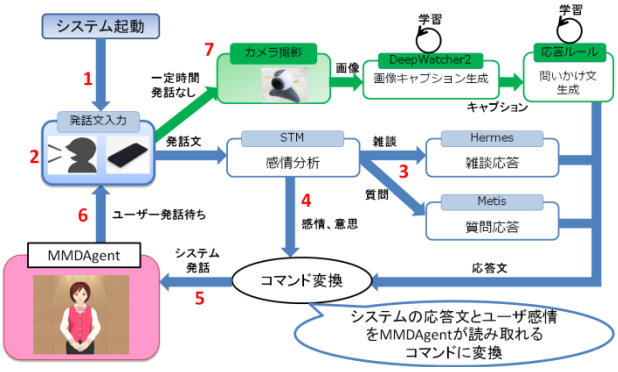

図 1 システム構成図

Chat and question dialog system HermesIII with talking to function based on perception of user`s apperance
Schalkwijk Alan, Yatsu Motoki and Harada Minoru
Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University

2.1　雑談応答

雑談応答システム Hermes を用いてユーザ発話文のキーワードをもとに Twitter から関連文を抽出し、その中から最も関連性が高いものを応答文として選び発話する。(図 2①)

またユーザの発話に指示詞が含まれていた場合、照応解析システム AnasysD を用いて指示先を特定し、指示詞を指示先の語に置き換える。例えば、図 2④のように"それで Youtube を見ています"という発話文の場合、"それ"という指示詞が含まれているため直前の応答文(図 2③)と発話文を連結し意味解析、照応解析を行い指示先である"スマホ"を特定する。そして入力文の"それ"を"スマホ"に置換し、キーワード抽出、Twitter 検索を行うことによって図 2④のような応答文を獲得する。

2.2　質問応答

質問応答システム Metis を用いて web から知識文を獲得し、質問文と知識文の類似性の計算を行い最も類似度が高い文から回答箇所を特定して図 2②のように第 3 候補まで出力する。類似度が非常に高い、または”～とは?”など答えが 1 つに定まるような質問文が来た場合、図 2⑤のように類似度が最も高い文だけを回答する。また質問文の中に指示詞が来た場合、2.1 と同様に指示詞の補完を行う。

2.3　システムからの話しかけ

システムとの対話中に間が空いた場合、カメラ撮影によって得られた画像からキャプション生成を行い、生成されたキャプションから応答ルールに基づいて話しかけ文を生成する。(図 2③)

2.3.1　キャプション生成

キャプション生成には DeepWatcher2 を用いる。まずキャプション生成させたい画像を物体検出に入力し元の画像と人物のみを切りとった画像に分ける。元の画像をあらかじめ学習させた「位置に関する言語モデル」に入力し位置と人物を表すキ

ャプションを生成する。人物画像は「人物に関する言語モデル」に入力し人物のみの特徴を表したキャプションを生成する。本システムでは、人物に関するキャプションを利用し話しかけ文を生成している。また、もし生成されたキャプションの中に人を表す語が含まれていない場合はカメラの前にユーザがいないと判断し、システムは応答ルールを使わずにカメラの前にユーザを来させるように促す話しかけを行う。

2.3.2 応答ルール

応答ルールには、実際に生成されたキャプションとそれに対応する応答文が用意されており、生成されたキャプションと最も類似度が高いキャプションに対応した応答文を出力する。また生成されたキャプションと応答ルール中のキャプション部に色や動物などの共通の属性を表す語が含まれている場合、応答文中の属性語を生成されたキャプション文中の属性語に置き換える。

話し手	発話文	
システム	何か話してみてください	
ユーザ	最近寒くなってきましたね	
システム	おはようございます♪、今日もみんなにいいことが起こりますように♪、最近寒くなっているので風邪には気をつけましょうね！	1
ユーザ	冬なので近いうちに温泉に入ろうと思っています	
システム	夏に入る温泉もいいが、やはり秋から冬の間の温泉はまた格別だと私は思う･･･。冬が楽しみだな。	
ユーザ	温泉で有名な観光地はどこですか？	
システム	回答候補が3つあります。 候補1　箱根 候補2　飛騨の山に囲まれた岐阜県屈指の下呂温泉 候補3　鹿児島県指宿温泉	2
システム	スマホで何をしているのですか？	3
ユーザ	それでYoutubeを見ています	
システム	スマホから見るとYoutube調子悪い、低スペって訳でもないのに、まともに生放送が見れん･･･	4
ユーザ	富士山の高さは？	
システム	富士山の高さは一般に3,776M	5
ユーザ	さようなら	
システム	さようならー。またお話ししましょうね	

図 2 対話事例

3 ユーザ発話に対する動作の選択

HermesⅢでの GUI として MMDAgent を用いる。ユーザ発話をテキストマイニングシステム STM で感情抽出を行い、HermesⅢによって生成された回答と感情データをコマンド変換し、ソケット通信を用いて MMDAgent に送信する。MMDAgent 側ではコマンドを実行し応答を発話し、指定された動作をする。STM で抽出される感情は、10 種類に分けられる。本システムでのユーザ発話の感情とキャラクターの動作の対応を表 1 に示す。またユーザが発話していない状態が何秒か続くと、問いかけるような動作を行うことによってユーザからの発話を促す。

表 1 ユーザ発話の感情とキャラクターの動作の対応表

ユーザ発話の感情	動作
喜	笑う
怒	お辞儀
哀	励ます
驚	驚く
安	軽く笑う
怖	悲しい顔
厭	悲しい顔
恥	軽く笑う
好	笑う
昂	ガッツポーズ

4 評価実験及び結果

事前に用意したユーザ発話文を被験者に入力させ、システムの応答に対する印象を各項目について 5 段階(1～5)で評価した。比較対象として Siri と GoogleAssistant と HermesⅢを用いた各評価項目に対する平均得点を表 2 に示す。

表 2 対話の印象評価得点（平均）

評価項目	HermesⅢ	Siri	Google Assistant
全体的に自然な会話か	3.7	2.7	2.3
質問応答の正確性	3.7	3.7	4.7
雑談応答は適当か	3.0	2.0	1.3
会話の連続性	3.3	1.3	1.3
エージェントの動作は適切か	3.3		
問いかけは適切か	3.0	1.0	1.0

参考文献

[1]山崎亮嘉，大森悠平，柳田菜々，谷津元樹，原田実：”指示詞を含む発話にも対応する雑談・質問対話システム Hermes”，情報処理学会第 80 回全国大会論文集，7Q-7，(2018.3).
[2] 高附勇介，谷津元樹，原田実：”高精度の Factoid/NonFactoid 型質問に対応する質問応答システムMetis”，情報処理学会第80回全国大会論文集，7Q-8，(2018.3).
[3]西脇　剛,保立哲志,原田実:“意味解析に基づくテキストマイニングシステム STM”,情報処理学会第 69 回全国大会論文集,2C-03,第 2 分冊 pp. 89-90.（2007.3).
[4]李 晃伸，大浦 圭一郎，徳田 恵一：“魅力ある音声インタラクションシステムを構築するためのオープンソースツールキット MMDAgent
掲載誌 電子情報通信学会技術研究報告 ：信学技報 111(364):2011.12.19・20 p.159-164
[5] 小林豊．鈴木諒，谷津元樹，原田実，:“深層学習による日本語キャプション生成システムの開発”，第 17 回インタラクティブ情報アクセスと可視化マイニング研究会発表予稿集(2017.11).