ImageNet を用いた画像提示による会話補助の検討

Consideration of conversation assistance by image presentation using ImageNet

井深　翔大

Shota Ibuka

岡山大学 阿部研究室

Abe Laboratory, Okayama University

概要 本研究では，二者対話中に話題に沿った画像を提示することによって会話の補助を行うシステムを作成することを目指す．画像は分類学習用に収集された大規模データ ImageNet[1] を用いる．ImageNet の画像はカテゴリごとに分類されており，同じカテゴリに対して複数の画像を検索することができる．本報告ではこのシステム作成の前段階として，会話中の画像提示が有効であるかを，Wizrd of Oz を用いて検討する方法を考える．

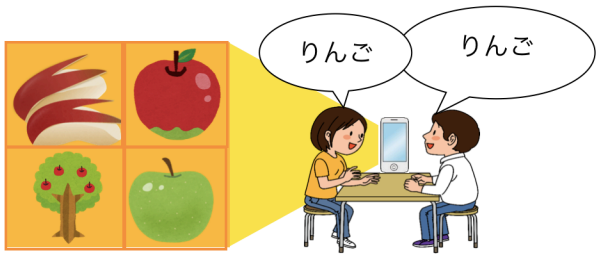

図 1: システムのイメージ

1　はじめに

音声認識の分野において，人とロボットが会話を行うための研究が活発に行われている．ユーザの質問に対してより自然な回答を行うために，Twitter から取得した大規模な対話コーパスを基に対話を生成する対話システム [2] や雑談対話システムにおける単語分散表現を用いた話題展開手法 [3] など，様々なアプローチが行われてきた．これらの研究では対話の自然性向上など一定の成果が見られたが，ロボットが人間と同等の対話を行うには至っていない．

そこで本研究では発想を変え，人間同士の会話を補助するためのシステムを作成する．人と人が会話を行う際には話題に対しての知識が必要不可欠である．話題に対しての共有する知識がないと会話が続かなくなったり，一方的な会話になる．よって共有する知識をシステムによって補うことが会話の補助につながるのではないかと考えられる．先行研究として「音声対話履歴を用いた tweet 検索に基づく話題提供方式の研究 [4]」を行なっている．この研究で作成したシステムでは，会話の補助方法として twitter 社の tweet と呼ばれる投稿を話題として提供している．結果として話題に関連する情報を提供することは，関連しない情報を提供することに比べ，話題が継続することがわかった．しかしシステム全体の評価としては，話題がすぐに思いつかなかったという意見が多かった．この原因の一つとして考えられるのは tweet の曖昧性である．話題と異なる tweet が検索されることも多く見られた．

そのため本研究では，ImageNet を用いて画像を提示することにより話題の提供を行う方法の検討を行う．画像を用いる理由は，瞬時に多くの情報を提供できると考えられるからである．

2　ImageNet

ImageNet は大規模な画像のデータベースである．約 3 万種類のカテゴリと，約 1400 万枚の画像が登録されている．画像には全て ID が振られており，画像の ID とそれに対応する画像の URL を記した words.txt，カテゴリと画像の ID の対応を記した gloss.txt が配布されている．システムはこの二つのファイルを組み合わせて走査することにより，ImageNet 内の画像検索を行う．例えば”りんご”の画像を検索する場合，まずカテゴリ名に”りんご”という単語が含まれるカテゴリ ID を検索する．カテゴリ ID が見つかった場合，カテゴリ ID に含まれる画像 ID を検索し，その画像の URL が得られる．

3　提案システム

システムの使用イメージは図 1 である．ユーザが共通する話題「りんご」について会話を行っている場合に，システムは「りんご」の画像を検索し表示を行う．また，システムの構成を図 2 に示す．システムは会話中の音声を認識する．そして認識した文から現在話題となっている単語を抽出する．表示部では抽出した単語を用いて ImageNet を検索し，話題の対象の画像を表示する．ImageNet には 1 つのカテゴリーに複数の画像が登録されている．画像の表示を行う際には複数枚表示する．ユーザは画像が表示されることによって話題についての具体的なイメージが沸き，会話の補助につながると考えられる．

4　Wizard of Oz 法による実験

現在システムは作成中であるが，その前に会話中における画像の提示が会話補助として有効かどうかを確かめるために，Wizard of Oz 法を用いて実験を行う．

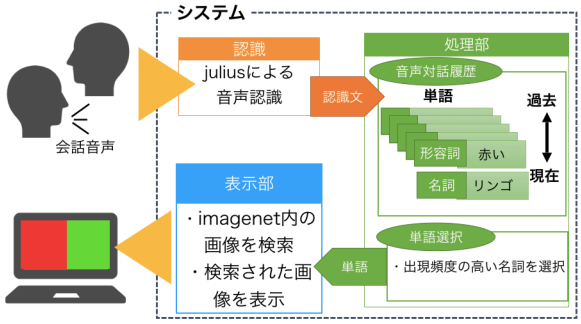

図 2: システムの構成

図 3: ImageNet を用いた画像の表示例

Wizard of Oz 法とは，システムのふりをした人間がユーザと対話する手法である．ユーザは実際にシステムを相手していると思いながら実験を行うため，得られるデータはより実際のシステムに近い状況であることが知られている．

被験者の二人組には，あらかじめ何の話題について会話を行ってもらうかを伝えておく．例えば「”りんご”についての会話を行ってください」のようにである．この条件のもと，被験者が以下のような会話を行ったとする．

話者 A　りんごといえば青森だよね
話者 B　いやいや，りんごは長野の方が有名でしょ
話者 A　青森と長野では品種が違うでしょ
話者 B　好みとしては僕はりんごより梨だな

実験者は”りんご”という単語が 2 回登場した時に，あらかじめ用意していた”りんご”の画像を表示する．表示方法は図 3 のようなものを想定している．図 3 の画像は実際に ImageNet で”apple”を検索し得られた画像の中の 4 枚である。偏った画像が表示されることを防ぐため，画像は常に 4 枚ほど表示する．またこの画像の選択は ImageNet で検索された画像の中から実験者がランダムで行いたいと考えている．

4.1　評価

被験者には複数の話題について同様の実験を行なってもらい，あえて画像の表示を行わなかった場合と比較して会話の盛り上がりを評価したいと考えている．盛り上がりの評価は以下の 4 つの項目で行う．

(a)　発話の長さ
(b)　発話の語彙結束性
(c)　声の高さ，速さ
(d)　発話のタイミング

4.1.1　(a) 発話の長さ

一度の発話が長ければ，会話が盛り上がっていると考えられる．よって 1 発話における平均の形態素数を評価する．

4.1.2　(b) 語彙結束性

語彙結束性は，発話間における意味の繋がりの強さを表す．対話において，発話と発話の意味的な繋がりが強いということは，ある話題についての詳細なやりとりがなされている考えられる．そして，話題について詳細な話に踏む込むということは，盛り上がっている可能性が高い．よってこの値を評価する．

4.1.3　(c) 声の高さ，速さ

人が盛り上がっている状態にある場合は，感情が高揚しているため，非盛り上がり時と比べて発声に変化が現れると考えられる．このため，音声の高さと速さを評価する．

4.1.4　(d) 発話のタイミング

発話間のタイミングは，盛り上がりと高い関連性を持つと考えられる．例えば，発話間の間隔が長く，場が沈黙で満たされている時間が長いほど，話者達は対話に消極的であり，対話は盛り上がっていない状態にあると考えられる．よってこの値を評価する．

5　まとめ

本報告では，二者対話中に話題に沿った画像を提示することによって会話の補助を行うシステムの作成を検討した．今後は Wizard of Oz の実験結果を見て修正を行いながらシステムを作成していく．

参考文献

[1] ImageNet(URL):www.image-net.org
[2] Bessho,F.,Harada,T.,Kuniyoshi,Y, “Dialog system using real-time crowdsourcing and Twitter large-scale corpus” Proc. of SIGDIAL, pp.227–231, Aug. 2012.
[3] 中野哲寛, 荒木雅弘, “雑談対話システムにおける単語分散表現を用いた話題展開手法” Proc. of 言語処理学会 , pp.269-272, 2015.
[4] 井深翔大, “音声対話履歴を用いた tweet 検索に基づく話題提供方式の研究” 平成 30 年度岡山大学工学部情報系学科特別研究報告書.