

平成 20 年度 情報工学コース卒業研究報告要旨

石井（健）研究室	氏 名	太 田 知 宏
卒業研究題目	発話生成を目的とした Wikipedia からの文抽出	
<p>対話システムは大きく分けてタスク指向型対話システムと非タスク指向型対話システムの 2 つがある。タスク指向型対話システムは何らかのタスクを達成することを目的としており、話題を限定できるので、それに関する最小限の知識を備えていればよい。一方、非タスク指向型対話システムは相手を楽しませる対話をするのが目的となり、話題を限定できないので、必要な知識が膨大となる。</p> <p>相手を楽しませるような対話を実現するためには、対話システムが対話を盛り上げるような発話をする必要がある。しかしながら、現在の非タスク指向型対話システムの多くは対話が成り立つことのみを目的としており、対話を盛り上げるというところまで至っていない。</p> <p>対話を盛り上げるような対話システムを設計するため、本研究室では対話システム「KELDIC (Ken 's Laboratory Dialogue Computer)」の研究を進めている。その一つに人間のインタビューを模した「インタビュー KELDIC」がある。対話の名手であるインタビューは、対話の前に相手について十分調査をする。インタビューはこの周知な調査により、相手の好む話題や、それに関する知識を得ることができ、それを利用して対話を盛り上げることができる。</p> <p>そこで、本研究では対話システムが熟練したインタビューのように対話を盛り上げるような発話を行うことを最終的な目標とする。その前段階として、本論文では「Wikipedia」から対話を盛り上げるような文を発話候補文として抽出することを目的とする。抽出された発話候補文は発話の際に適切な形に修正されて利用される。</p> <p>本研究では以下の手順で処理を行い、発話候補文を抽出する。</p> <ol style="list-style-type: none">1. 前処理 Wikipedia から記事を検索し、タグを取り除いた後に形態素解析を行う。2. スコア付け tf-idf 法、語の共起、文の長さを利用して記事の各文に対してスコア付けを行う。3. 順位付け 発話として不適切な文を取り除いた後、スコア付けに従って記事の各文に対して順位付けを行う。4. 文の抽出 順位付けにより上位となった文を本論文で目的とする文として抽出する。 <p>本論文で提案する手法を用いて、Wikipedia の記事内の各文に対して順位付けを行い、上位の文に対して人間による評価を行った。その結果、評価を行った文のうち 67.3% の文が発話生成に利用可能であると判断され、さらにそのうち 72.0% の文が対話が盛り上がるような文と判断された。これにより Wikipedia は発話候補文を抽出するためのデータとして利用できることが確認された。</p>		