

平成17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	小泉 康 治
卒業研究題目	自由記述プロフィールからの話題抽出	
<p>近年、日本語を用いた人とコンピュータとの対話システムは多くの分野で研究がなされている。そのほとんどは、接客などの特定の目的を持つタスク指向型システムであり、既に自然な会話を実現している。一方で、雑談によって、人を楽しませることを目的とした非タスク指向型システムは少ない。その原因の一つとして、システムがユーザの好む話題を理解せずに対話をするため、ユーザが話題に乗れず話が続かないことが挙げられる。そこで、本研究では、システムがユーザの好きな話題に関する情報を事前に獲得し、それを用いることで楽しい対話を実現することを目指している。</p> <p>本研究は対話システム KELDIC(Ken's Laboratory Dialogue Computer) を対象とする。KELDIC は対話コーパスからスクリプト(対話における発話集)を自動生成し、その中から発話を選択することで対話を行う。現在は発話の選択を人間が行っており、全自動ではない。</p> <p>本論文では、発話選択に役立てるため、あらかじめユーザに書いてもらったプロフィール文書から、ユーザが好む話題の情報をシステムが自動的に抽出する手法を提案する。プロフィール文書には「趣味は園芸です」など、記述したユーザの趣向などが記入されている。これによって、生成されたスクリプトから適切なものを自動的に選択し、発話までを全自動で行うことが期待される。</p> <p>本手法では、(趣味, 園芸) というような、(属性, 値) の対を「話題対」と呼んで話題の情報として扱う。まず、話題抽出対象のプロフィール文書とは別の文書から、人間が正しいと思う話題対を事前に用意し、「理想話題対」と呼ぶ。次に話題抽出対象のプロフィール文書から、係り受け構造を用いて話題対を取り出し、その中から条件を満たすものを抽出する。条件は2つ有り、条件1は抽出候補の話題対が理想話題対と同一であるかどうかであり、条件2は上位概念を利用し、抽出候補の話題対と理想話題対との類似度を求め、類似した理想話題対が多数存在するかどうかである。</p> <p>手法の有効性を求めるために、50人分のプロフィール文書を用いて抽出実験を行った。抽出結果のうち、人間が見て正しいものの割合を適合率とした。また、元のプロフィール文書を見て、人間が抽出を期待している話題対をどれだけ抽出できたかの割合を再現率として評価した。その結果、条件1のみを満たす話題対を抽出する方法では適合率は88.5%、その時再現率は37.9%となった。また、条件2を加えたところ、適合率が80.9%のとき、再現率が最大54.3%となった(ただし、適合率80%以上を条件とした)。よって、適合率が十分高いことと、二つの条件を用いたことによって再現率が上昇したことから、本手法が話題抽出に有効であることがわかった。</p>		