

平成28年度 情報工学コース卒業研究報告要旨

長尾 研究室	氏 名	Jung Dawoon
卒業研究題目	発表スライド内の用語と関連した質問の自動生成に関する研究	

大学のセミナーなどの発表を伴う会議で活発な議論が展開されるためには、参加者が発表の内容を正しく理解する必要がある。しかし、発表スライド内には説明が不足している用語が含まれることがある。このような場合でも会議の雰囲気や立場によっては、参加者自身が用語の意味について質問しにくいことがある。

本論文では、発表スライド内に出現する用語のうち、説明が不足している用語を検出する手法について述べる。検出された用語の意味を問い合わせる質問文を自動生成することにより、コミュニケーションロボットなどが代行して問い合わせるシステムを構築した。このシステムの実行の流れを図1に示す。

スライド内のある用語に対して、その意味が分からない議論参加者が少なくとも1名は存在する用語を説明が不足している用語とする。これらの用語をコミュニケーションロボットが自律的に質問するためには、発表スライドから説明が不足している用語を検出し、その用語に適した質問文を生成する必要がある。調査により、用語を質問する発言においては用語が分からない理由が述べられる傾向があること、及び、その理由に応じて発表者の回答の仕方が異なる場合があることを観察している。説明が不足している用語に対して該当する理由を抽出できれば、用語とその理由に基づき、あらかじめ用意されたテンプレートを選択することによって質問文を生成できる。

発表スライド内の説明が不足している用語について質問を自動生成するために、発表スライド内のあらゆる用語に対して、該当する用語を検出し、不足している理由を推定する手法を提案する。検出、及び、推定にはSVMを用いる。SVMで使用する素性として、用語の言語的な特徴に加えて、用語が出現するスライドのテキスト構造や発表者の過去の発表スライドなどの文脈の特徴を用いる。

発表スライド内の説明が不足している用語の検出実験とその理由の推定実験を行った。実験データとして著者の所属している研究室のゼミで使われた発表スライドに含まれる用語を用いた。学部生及び大学院生が実験データ中の分からない用語と回答した用語と、その理由を正解データとした。非言語的な特徴による素性を除去した場合を比較手法として、説明が不足している用語の検出実験とその理由の推定実験を行った結果、比較手法ではF値が29.5%であり、提案手法ではF値が45.1%であった。これらの結果から、提案手法は比較手法よりF値が高く、文脈の特徴の利用により検出性能が上がることを確認した。

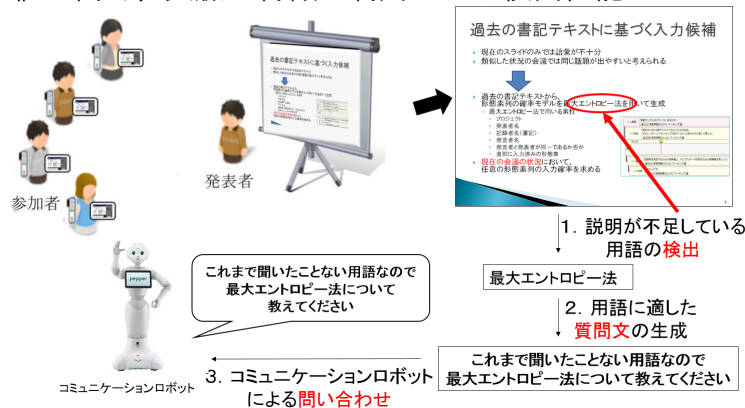


図1: システムの実行の流れ