

2012 年度 修士論文要旨
自然言語処理による日本語文章の自動生成

関西学院大学大学院 理工学研究科
情報科学専攻岡留研究室 杉本 亘

現在、ソーシャルメディアを代表する多様な手段で多くの人が日々情報を発信し、twitter上のbotのような文章の自動生成が盛んに行われている。しかしbotのようにスロット形式やテンプレートを用いた文章生成では、あらかじめ用意した文や決まった形の文しか生成できないため表現できる幅が狭い。本研究では、任意の話題に関する幅広い文章生成を目的にする。そのために、Web上に存在する情報を利用し、文章中の文はコピーせず、単語だけ抜き出しそれと、独自の構造文法を用いることで、テンプレートを用いない幅広い表現の文章を自動生成する。本研究ではWeb上の文章を素材として利用し、任意の「名詞」と「動詞」に関する文章を生成する。まず生成したい話題に関連した「名詞」と「動詞」に対してweb検索を行う。検索結果について、クエリと関連性の高いwebページをランキング上位に、関連性の低いページをランキング下位にリランキングを行う。取得したwebページの上位30ページを素材としてその中に含まれる単語を品詞ごとに収集し、後続文の主語選定のためにページごとに文も収集する。文章構築には句構造文法を用いる。ランダムに句構造文法からSVC, SV00といった文型を選択し、対応する品詞にそれぞれ単語を当てはめ文を生成する。その文と関連のある単語を次の文の主語として選定し、句構造文法により次の文を生成する。この作業を繰り返すことで文章を生成していく。

句構造文法を用いることで、テンプレートやスロット法を使うことなく、単語を素材とした幅広い表現が可能な文の自動生成ができた。文章の繋がりに関しては新たなアプローチが必要と考えられる。