
離散 GA
降幡建太郎

1 今回の報告

今回は、大学院入試の勉強のため、研究を休ませていただいた。そのため、研究報告ができないので、かわりに、院試で提出した小論文の内容を簡単に記す。

2 自然言語処理での現在の研究

本節では、自然言語処理において文章理解支援にかかわる研究の現状を述べる。

● 文書の自動要約

現在の自動要約の研究は、主に文章中から重要部分を抽出することを中心に研究されてきている。重要文の抽出には、テキスト中の語句の出現頻度の利用、先頭に来る文は重要である可能性が高いといった位置情報の利用、関連性のある語彙が複数の文で用いられることによる意味的なつながりを利用する方法などがある。また、それらの手がかりを複数統合して用いた方法も研究されている。最近では、そういった情報の最適な統合方法を機械学習で見つける研究などが行われている。今後は、指示代名詞の照応解析を利用した手法、抽出文の再構成の研究も重要になってくる。

● 言い換え

言い換えるの研究は始まったばかりで、まだ例が少ない。その中の一つに、本年度から NAIST の研究室に着任された乾助教授らの研究がある。実際のプロトタイプシステムとして、FUNNE, KURA という 2 つのアプリケーションを開発している。同研究では問題を、文の言い換え、否定や受動態などの言い換え、動詞や名詞などの言い換え、などに分けて取り組んでいる。前 2 つに比べて、動詞や名詞などは、言い換え規則を定めにくいいため、大規模な知識が必要になる。さらに、そのような知識をどのように集めるのかという課題もある。

● DL-MT

実際の日本語読解システムとして、Web 上で公開されている DL-MT がある。このプログラムは、JAIST の自然言語処理学講座の日本語読解支援システム構築チームによって開発された。これは、ユーザーが日本語テキストを入力すると形態素解析し、単語上にマウスポインタを示すと、英訳および単語の意味が表示されるというものである。

この他にも、指示代名詞などの照応解析の研究なども始まっている。私は現状を調べるために文献調査を行ったが、このような文書を扱う研究はまだ始まったばかりで、現実のシステムへの実装はまだ時間がかかりそうである。しかし、一方では、形態素解析や、係り受け解析に関しては、かなり実用的なレベルに達している。

3 言い換えについて

文章の分析を行う際に、読者がよくわからない部分に遭遇したとする。このとき、その部分に関して説明している箇所や、同じような箇所を探し出し、読者に提示しなければならない。その際、同じ概念を述べている文であるかを判断しなければならない。そのときに、言い換え技術が有効である。

言い換え技術は、自動要約にも適用できると考える。すなわち、本文中の重要箇所を抜き出した後、それをまとめた文章に再構成する必要がある。そのときに言い換え技術を利用できる。

言い換え技術の適用範囲はそれだけではない。例えば、現在の機械翻訳は、訳文がぎこちなく、とても自然な日本語とはいえない。そのような訳文に対して言い換え技術することによって、自然な日本語に置き換えるという利用法が考えられる。

また、言い換えには、単純に語と語が対応関係を取れるようなものばかりではなく、もっと概念的に広義の言い換えるの研究も行う必要がある。自動要約も、広義の言い換えである。

以上のように、言い換えは未知の部分が多いが、非常に適用範囲が広いように思う。今後、自然言語処理において、大変重要な位置を占めるのではないかと思われる。

4 翌月への課題

- 並列分散 GA のプログラムを作成し、正しく動作させる。
- パラメータで選択法、交叉法等を変更できるようにする。
- 階層型の島モデルを実現する。