理工学専攻修士論文要旨

提 出 年 度 ： 2008 年度
提　出　日 ： 2009 年 1 月 29 日
専修コース ： 知能情報コース
学 生 番 号 ： 35607159
学 生 氏 名 ： 由原木 翔

（論文題目）
EM アルゴリズムを用いた自動要約システム ABISYS の要約の精度改善

（内容の要旨）
近年、インターネット上では、大量のテキストデータが溢れている。
これにより膨大な量の情報を取り扱うことが可能となった。だが、その反面、必要なデータのみを抽出することが、非常に困難になっている。また、携帯電話などの情報端末の普及から、情報から重要な箇所だけを抽出し、よりコンパクトにまとめる技術が必要不可欠となっている。そこで、原田研究室では自動要約システム ABISYS の研究が行われている。

ABISYS では、日本語文章における意味解析システム SAGE によって意味解析し、出力された意味解析結果から、名詞節の中から重要語を抽出し、これを中心に要約文を生成する。
今年度の ABISYS2008 ではこの重要語の識別関数の学習に EM アルゴリズムを導入した。従来の ABISYS で重要語判定に使用していた SVM（サポートベクターマシーン）は、語の重要度を、０と１のみで判別していた。だが EM アルゴリズムは、「どの程度重要か」といった重要度合いを確率として出すことができるため、より精度の高い学習が可能になる。

EM アルゴリズムを用いて、まず 5000 件の文章に対して人手で圧縮率 10%、30%、50%、で行い　文章中の各文節がそれぞれの圧縮率で残ったものとそれ以外 4 ランクに分け、それぞれの混合分布を構成する各クラスタの平均、標準偏差、重みを学習させる。この結果各分布のクラスタの数は、10%は 3 つ、30%は 8 つ、50%は 4 つ、それ以外は 5 つとなった。これによる文節の各ランクの確率を、逆順位平均によってまとめ、結果として出た統合確率を元に全名詞節をソートし、重要語の選別を行った。

評価にあたって、被験者 7 名に、人手の要約、SVM による要約、EM アルゴリズムによる要約の３つの要約および原文を提示し、原文の重要な内容をどの程度要約がカバーしているか、要約の読み易さの２つの評価基準で、要約を評価し、その順位の平均を算出した。結果、EM による学習が 1.97、SVM による学習が 2.47 となった。この結果、要約精度は改善されたといえ、本研究の EM アルゴリズム導入は妥当であったと言える。