

# リード文の文短縮， および情報補足によるニュース記事の要約

情報科学科 小林 幸 指導教員：山村 毅

## 1 はじめに

現在，インターネットの普及により，ニュースでは世界各地の出来事も報道されるようになった．国内でのニュースも合わせると，相当量のニュースが日々我々の元へと入ってくる．迅速な状況理解のためには，これらから内容把握を行って，必要なものを取捨選択する必要がある．ニュースの内容把握の一般的な指標として，ニュースのタイトルがあるが，タイトルを見るだけで内容把握をすることは難しい．タイトルより多くの情報量を得るためにはニュースの内容を要約することが考えられる．そこで本研究では，ニュースの選別や時間短縮に役立てるため，入力されたニュース記事から要約文を作成する．

## 2 要約

要約とは，文章などの論点・要点を短くまとめて表すことを指す．例えばあらすじは物語全体の流れを簡潔にまとめたもので，これも要約といえる．一般的な要約が指すものは，このように一つの文章に対する要約であることが多い．しかし要約には，文内要約という一文のみに対する要約もある．

文内要約の手法は様々だが [1]，最も広く使用されているのは元の文から文節単位，または単語単位で不要な語句を削っていく文短縮手法である．他にもあらかじめ定型文を用意して元の文から抽出した単語を当てはめる手法 [2] や，重要度の高い文節を抽出しそり外の文節を削除する手法などがある．

## 3 要約文作成手法

本研究では，入力されたニュース記事のリード文を係り受け解析を利用して，固有表現などの不要語の削除，文短縮を行うことで要約文を作成する．リード文のみでは情報が不足していると判定した場合は，その情報を捕捉できる一文を本文より選出し追加文とする．

### 3.1 文短縮手法

文短縮手法は，文を係り受け解析し，根と根にかかる文節を短縮文の基本とし，それらの文節にかかる文節が必要かどうかを判断し，文節を短縮文に採用していく．要・不要の判定には，以下に示すような採用条件を用いる．

- 動詞，形容詞，形容動詞に係る文節
- 名詞に情報を捕捉する文節
- 重要注目語（逮捕，逃走など）を含む文節
- 短縮文に重要注目語以外の動詞がない時，動詞を含む文節
- 短縮文の文節数が 2 以下である時，短縮文に含まれる文節に係る文節

全ての条件で採用する文節がなくなれば文短縮完了とする．

### 3.2 追加文

追加文が必要かどうかは，重要注目語（加害者や被害者に関する単語）がリード文と本文にどのように含まれているかで判定する．追加文が必要な場合，重要注目語が含まれる文に対して重み付けをし，最も重みが大きい文を追加文として，リード文

と同様の処理を行う．

## 4 評価実験

評価は 2 種類の手法によって行う．

1 つは，YAHOO!JAPAN ニュースから収集した 500 のインターネットニュース記事にあらかじめ人手で理想要約文を作成し，これらとシステムが出力した要約文を比較し共通に含む名詞数を用いて，精度，再現率，F 値の評価基準で評価する．もう 1 つはシステム要約文を実際に人に見てもらい要約できているかを 3 段階（ ， ， × ）で評価してもらう．

それぞれの評価結果を表 1，2 に示す．全体的な評価はいずれも 6 割前後であった．

表 1 理想要約文による評価結果

	リード文のみ	リード文 + 追加文
精度	0.532	0.519
再現率	0.713	0.757
F 値	0.578	0.590

表 2 人手による評価結果

	人手
	256
	106
×	62
評価	0.618

## 5 今後の課題

結果として 6 割前後の評価を得ることができたが，精度と再現率に差ができてしまった．これは係り受け解析を用いた文短縮手法にまだまだ欠点が多く，不要な文節を完全に削ることができなかったからである．このことから，文短縮手法には係り受け関係だけでなく，別の指標を作り複数の観点から要・不要を判定する必要があると考えられる．

また不要語の削除により，係り受け木から切り離される文節が出てしまう場合があった．それらの文節に対して係り先を書き換えるなどの処理を行い係り受け木を修復することが可能であるのか考えたい．

## 参考文献

- [1] 奥村学，難波英嗣 共著，” 知の科学 テキスト自動要約”，オーム社 (2005)
- [2] 山本和英，牧野恵，池田諭史，” 類似用例文の部分的置換による文短縮”，情報処理学会研究報告自然言語処理 (NL)，Vol.2006,No.53,pp.21-28(2006)