構文片を用いた要約文生成

村松祐希，山本 和英
長岡技術科学大学 電気系
E-mail:{muramatsu,ykaz}@nlp.nagaokaut.ac.jp

1 はじめに

自動要約の研究はさまざまに存在する。近年、文章の重要な部分を確率モデルにより求める方法が主流となっている。しかし、確率モデルによって計算される単位は形態素や文節が主であり、出力された文章が不自然な場合がある。そこで、我々は係り受け関係を持つ 2 文節を単位とする「構文片」を用いることにした。従来の単位で要約するよりも、人間が見て自然な繋がりを持つ文節が増えることによって、要約文を生成することが容易であると考える。

自動要約の研究として、入力された文章の単語や文節に重要度を設け、重要と判断された部分を抽出して要約文を出力する研究がある。

重要度を用いる研究として、1 文を圧縮して要約文を作成する文圧縮の研究では堀ら[1]の研究がある。単語の重要度 TF・IDF と連接確率を用いている。重要な単語を含み、日本語として繋がりのよい要約文を作成することを目的した手法である。動的計画法を用いることにより、効率的な要約がされている。

また小黒ら[2]は、「文節重要度」と「係り受け整合度」を重要度としている。これにより、要約文を出力するための文節を導出している。文節単位の要約により単語単位より自然な要約がされている。

重要度の変わりに、入力文章と類似用例文の類似度を用いる研究として牧野ら[3]の「用例利用型要約」の研究がある。あらかじめ人手で要約された文を要約するための用例文として利用する手法である。用例文を選択するために、「助詞の一致」、「固有表現タグの一致」、「単語間類似度」を設けることで類似用例文の選択を行っている。入力文章に対して 1 文の要約文を生成している。類似用例文を探す際、単語の情報を用いている。頻度をもとに単語間類似度を考慮しているため、適切でない用例文が対応付けられる場合がある。結果として、最終的に人間が見て不自然な要約結果になることがある。このため、我々は構文片を用いた。青木ら[4]は評判情報の評価極性を構文片に付与することで、従来手法よりも高い精度で評判情報を抽出している。構文片とは青木らが提案している係り受けの最小単位である。

構文片を用いることにより 3 つの利点が挙げられる。1 つ目は、「靴をはきます」と「白い息をはきます」では「はきます」意味が異なる。構文片は文節単位より正確に文の意味が得られると考える。2 つ目は構文片が構文構造の 1 片であることにより、文の復元が可能であると考える。これにより、従来手法よりも自然な要約文が生成できると考える。また、係り受け情報を用いることにより、構文片の 2 文節の場所が隣接していなくても問題がないと考える。3 つ目は係り受け情報のみを用いるため抽出が容易であると考える。シソーラスを用いることで単語間類似度より意味的に近い類似用例文が選択できると考える。

2 事前準備

2.1 構文片

文の類似性を算出するために、文の木構造を考える方法がある。青木らが提案した構文片の定義は次のようになっている。構文片とは係り受けの最小単位で、係り受け関係の修飾要素と被修飾要素の対である。

しかし、この条件では入力記事に対して十分に構文片を網羅することが出来ない。よって本稿は係り受け解析を行って抽出できる修飾要素と被修飾要素全ての対を構文片として考える。以降、係り元の文節を「前項」、係り先の文節を「後項」と呼ぶことにする。入力記事から取り出した構文片を $Pi[i_{me} \rightarrow i_{mc}]$、（$i_{me}$ は前項の文節、i_{mc} は後項の文節、矢印は係り受け関係）用例文から取り出した構文片を $Ps[s_{me} \rightarrow s_{mc}]$とする。

2.2 用例文

用例文はあらかじめ人手で要約された要約文とする。用例文に含まれている構文片は用例文の復元が可能である。つまり、用例文から抽出した構文片は人間が要約するときにどのような内容で要約したらよいのかという経験や文法等を含んでいると考えることができる。用例文の構文構造を利用して、入力記事を 1 文へ要約する。

人手で要約された要約文はさまざまな形で用意をすることが出来るが、本稿では Nikkei-goo[1]からメール配信された要約文を用例文として使用した。この要約の記事は牧野ら[2]の観察の結果、1～3 文で構成されており、重要な内容を最初に表記する傾向が強いとされている。また 2 文目以降が無くても要約文として成立する場合が多いため、用例文として使用するのは要約記事 1 文目に限定することにした。

3 提案手法

3.1 システム構成

図1にシステムの流れを示す。

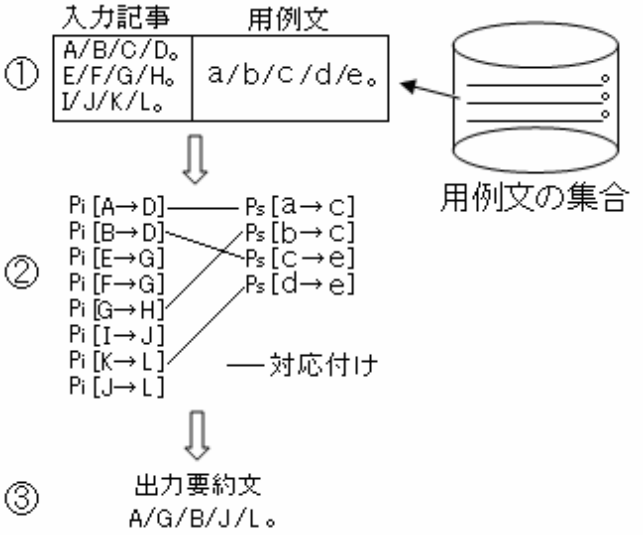

図 1 用例利用型要約の手法概要

システム構成は以下に示す3つの処理で構成される。

①.用例文の選択：入力記事と意味的に近い用例文を選択する。

②.構文片の対応付け：入力記事と用例文の構文片を比較して類似しているものを対応させる。

③.要約文の生成：対応付けられた構文片を組合わせて1文を生成する。

用例文の選択については 3.3 節で、構文片の対応付けは 3.4 節で、対応する構文片の組合わせは 3.5 節で示す。

3.2 構文片の類似度

2 つの構文片について、どの程度意味が近いかを求めるために EDR 概念辞書[2]を用いて類似度を計算する。本研究では、EDR 概念辞書に存在するノード間の距離を、長尾ら[5]が示した類似度と置き換えて考える。まず入力記事から抽出した構文片と類似用例文から抽出した構文片を用意する。

各構文片から内容語を全て抽出し、(内容語は名詞、動詞、形容詞とし、非自立語は除く) EDR 概念辞書にあるシソーラスを用いてノード間の距離を計算する。複合名詞や複数の形態素で成り立っている語を考慮するため、EDR 概念辞書と最長一致した内容語列を用いる。例として、ある文節から連続している内容語を 3 つ取り出したとする。(内容語 a,b,c とする) 取り出した内容語から EDR 概念辞書に登録されている最長の連続した形態素を探す。ただし、連続した内容語の最後尾の形態素が含まれているものを優先的に探すことにする。

abc というノードが EDR 概念辞書に登録されていない場合、次に探す語は bc、そして c となる。c が EDR 概念辞書に登録されていない場合、内容語 abc は除外して考える。

構文片から EDR 概念辞書に登録されている全ての内容語を抽出し、内容語の類似度を式 (1) から算出する。ただし、内容語によっては、同じ表記でも複数のノードが存在する場合がある。その場合には類似度が最大になるノードを選択する。これを前項の文節同士、後項の文節同士で行う。前項から算出した類似度と後項から算出した類似度の加算平均を式(2)に示した構文片の類似度とする。

例を示す。入力記事の構文片を Pi「私は→歩く」、用例文の構文片を Ps「あなたは→走る」として、Pi、Ps の類似度を計算することにする。まず、各構文片から全ての内容語を抽出する。Pi の場合は「私」、「歩く」。Ps の場合は「あなた」、「走る」になる。次に前項同士の内容語(「私」、「あなた」)の類似度、後項同士の内容語(「歩く」、「走る」)の類似度を計算する。計算して求めた 2 つの類似度を加算平均したものを式(2)に示す構文片の類似度として考える。

$$sim_c = \frac{2(dk+1)}{di+dj} \quad (1)$$

(di,dj,dk は、内容語 s_{me}(又は s_{mc})のノードの深さ,内容語 i_{me}(又は i_{mc})のノードの深さ, i_{me} と s_{me} の共通上位ノードの深さ)

$$sim(Ps,Pi) = \frac{sim_c(s_{me},i_{me}) + sim_c(s_{mc},i_{mc})}{2} (2)$$

3.3 用例文の選択

用例文の選択では入力記事と用例文に含まれている構文片を比較し、入力記事と内容の似ている用例文を取得する。内容の類似性をどの部分に注目して計るかという問題が生じるが、牧野ら[2]は入力記事と用例文の対応付けにおいて、文節間での「助詞の一致」、「固有表現タグの一致」、相互情報量を用いた「単語間類似度」を用いている。しかし、本稿は構文片の類似度を用いてこの問題を解決する。

入力記事と用例文の類似度を計る。図 2 を用いて説明する。入力記事から全ての構文片を抽出する。Pi は入力記事の構文片の集合である。用例文 E_1 についても同様に行う。(Ps も同様である) 次に、用例文 E_1 の構文片 Ps_1 から見て、入力記事の構文片に最も類似度が高い構文片を選択する。これを用例文の構文片 (Ps_1～Ps_i) に対して全て行う。ただし、最大類似度が 0 の構文片について考慮しない。用例文の構文片から得られた全ての最大類似度に対して、加算平均を求める。式(5)で示すように、この値を入力記事と用例文 E_1 との類似度として考える。

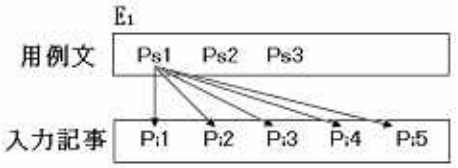

図 2 入力記事と用例文の類似度

$$Pi \in \{P_{i1}, P_{i2}, P_{i3}, \cdots P_{im}\} (3)$$

$$Ps \in \{P_{s1}, P_{s2}, P_{s3}, \cdots P_{sn}\} (4)$$

$$sim_E = \frac{\sum_{s=1}^{n} sim(Ps, \arg\max_i(Pi))}{n} (5)$$

これを全ての用例文 (E_1～E_N) に対して行い、類似度が最も高く評価された用例文を入力記事に対して最も内容が似ているとして考える。

3.4 構文片の対応付け

3.3節の用例文の選択では、入力記事の1つの構文片に対して、用例文の複数の構文片が対応付けられる場合がある。本節では、構文片を組み合わせることによって要約文を生成するまでの方法を述べる。図3に入力記事に対して用例文が選択された場合の構文片の対応付け例を示す。

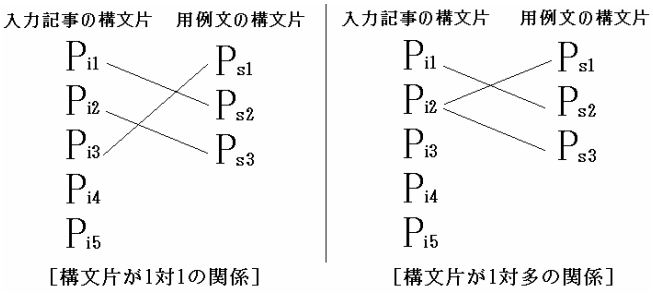

図 3 構文片の対応付け例

用例文の構文片が入力記事の構文片に対して 1 対 1 の関係が取れている場合に注目する。用例文の意味に近い構文片が入力記事の構文片と対応が取れていると仮定する。対応付けが取られなかった入力記事の構文片については用例文の構文片とは意味的に近くないと考えたため、対応付けが取れなかった構文片に対しては要約文に使用しないことにする。使わなかった構文片を削除することによって情報を落とすことが出来る。用例文から見て最大類似度が複数存在する場合は、文頭に近い構文片を選択する。

次に用例文の複数の構文片が重複して入力の構文片と対応した場合について考える。1 対 1 の関係が取れている場合と同様に対応が取られなかった構文片は要約に使用しない。1 対 1 の場合と比べて最終的に残る構文片は減ることになるが、意味的に同じものが選択されているので、効率的な要約が行われていると考えることができる。

本手法の対応の取り方は用例文から見て入力記事の構文片に最も類似度の高い構文片を取る方法である。よって、入力記事の複数の構文片が重複して 1 つの用例文の構文片に対応付けられることはない。

3.5 要約文の生成

3.4 節で入力文から得られた構文片を用いて要約文を生成する方法について述べる。以下に例を示す。

＜入力記事＞
北海道連合海区漁業調整委員会がまとめた九八年の秋サケの最終漁獲状況で、漁獲量は三千九百七十六万二千六百四十一尾で前年同期比一四･八％減となった。海水温の上昇が不漁の原因とされる。ただ単価は上昇、金額ベースでは約三百二十六億九千五百万円と同七･二％の増。

＜用例文＞
１０月末の外貨準備高は８４１７億９２００万ドルで前月末比１７億７１００万ドル減。

＜出力要約文＞
不良の原因とされる漁獲量は前同期比一四･八％減となった。

入力記事に対して用例文が選択された後、用例文の構文構造に従って、用例文の構文片と類似度が高い入力記事の構文片が当てはめられて要約文が生成される。この場合、入力記事と用例文の構文片の類似度に関する対応付けは表 1 のようになる。用例文の「10 月末の」という文節は入力記事の「不良の」という文節に対応付けられているので、出力する文節となる。用例文の「外貨準備高」が含まれる文節は入力記事の「原因とされる」、「漁獲量は」の複数に対応付けられている。この場合は、構文片の類似度の計算に使用した文節同士の類似度を用いて、類似度が高い入力記事の文節を選択する。例では「原因とされる」が選択されているのが分かる。

表 1 構文片の対応付け例

用例文の構文片	入力記事の構文片
Ps1[10 月末の→外貨準備高]	Pi1[不良の→原因とされる]
Ps2[外貨準備高は→17 億 7100 万ドル減]	Pi2[漁獲量は→一四･八減となった]
Ps3[8417 億 9200 万ドル→17 億 7100 万ドル減]	Pi3[漁獲量は→一四･八減となった]
Ps4[前年末比→17 億 7100 万ドル減]	Pi4[前同期比→一四･八減となった]

4 評価実験

提案手法の有効性を確認するため従来手法である牧野らの手法と比較実験を行った。用例文の集合は従来手法と全く同じ用例文を用いた。よって 2007 年 12 月までの用例文を使用した。

･用例文の集合
用例文の集合には、日経ニュースメール Nikkei-goo から配信されているニュースの要約文を用いた。このニュース要約文は人手で作成されているものであり、1999 年 12 月から 2007 年 12 月までの 27036 件を用いた。

･評価実験データ
日本経済新聞 1998 年[3]の記事 100 件と各記事に対する要約文を用いた。要約文には入力記事の 1 文目と、第 3 者による要約文を 2 文用意した。

4.1 自動評価尺度による評価

現在までに多くの要約システムの自動評価尺度が提案されている。本稿では DUC や TSC などで多く用いられている BLEU 値と ROUGE-4 値を用いて自動評価[4]を行う。これらは入力文に対してあらかじめ作成した人手による要約文とシステムが出力した要約文を比較する。結果を表 2 に示す。

表 2 自動評価尺度 BLEU,ROUGE-4 による評価結果

評価尺度	従来手法	本手法
BLEU 値の平均 (100 件)	0.41	0.081
ROUGE-4 値の平均 (100 件)	0.85	0.64

表 2 より本手法は従来手法より下回った結果が得られたことが分かる。従来手法と本手法を比較したときについて評価値の低下率を計算した。BLEU 値は 80%,ROUGE-4 値は 24%であった。BLUE 値が人手による要約文とシステムが出力した要約文の適合率と考える。ROUGE-4 値は同様に再現率と考える。本手法は適合率に対する低下率が特に高いことが分かる。実際に人手による要約文とシステムが出力した要約文を観察しても、全く合致していない文が出力されてしまう場合があった。

4.2 人手評価

評価者 2 人に本手法によって出力した要約文と従来手法の要約文の比較評価を行った。評価者にはシステムに入力した文章を読んでもらい、2 つのシステムが出力した要約文について内容がより適切である要約文を選んで評価した。結果を表 3 に示す。

表 3 内容適切性評価結果

	従来手法	本手法
評価者 A	82/100 (文)	18/100 (文)
評価者 B	82/100 (文)	18/100 (文)

表 3 は評価者 2 人が 4.1 節で使用した文章と同じ文章(100 件)について評価した結果である。従来手法に比べて本手法は内容適切性が大きく下回る結果となった。評価者 2 人どちらにも従来手法より本手法が適切であると判断された 7 文に対する自動評価の結果を観察した。全ての文に対して、従来手法が本手法よりも高い評価値であることが分かった。次に日本語としての可読性(自然さ)を評価した。表 4 は内容適切性評価と同じ評価実験データを使用した結果である。結果を表 4 に示す。

表 4 可読性評価結果

	従来手法	本手法
評価者 A	70/100 (文)	30/100 (文)
評価者 B	44/100 (文)	56/100 (文)

評価者 A は本手法が従来手法に比べて、下回ると評価した。評価者 2 人の結果を合算して考えると、本手法は従来手法より下回る結果となった。しかし、評価者 B の評価では本手法が従来手法よりもわずかに上回る結果となった。構文片の係り受け関係を使用することにより、日本語として自然な要約文を生成できた可能性がある。

5 考察

5.1 EDR 概念辞書の網羅性

用例文の集合と評価実験データに対して、取り出された内容語がどの程度 EDR 概念辞書 (日本語単語登録数 27 万語) で網羅されているのかを調査した。結果を表 5 に示す。

表 5 EDR 概念辞書の合致率 (用例文 27036 件、評価実験データ 100 件)

	用例文の集合	評価実験データ
形態素数	1,003,459	16,188
内容語数	677,855	9,998
EDR 概念辞書との合致数	294,445	4,394

表 5 は形態素数が入力記事から取り出された形態素数を示し、そこから内容語と判断された数が内容語数である。EDR 概念辞書との

合致数は内容語と EDR 概念辞書に登録されている単語と合致した数を示す。形態素数に対する内容語の割合を計算したところ、用例文の集合は68%であり、評価実験データでは62%であった。しかし、内容語に対して EDR 概念辞書が網羅している形態素は半分以下であることが分かる。ここから、形態素解析辞書との合致率が低い中で、内容語の取り出し方に問題があると考える。複数の内容語が 1 文節に存在する場合、後方から優先させて一番連続して合致する形態素を選択している。複合語は後方の形態素に意味が強く出るという仮定で行っているので、この仮定に対しては再検討が必要だと考える。

本手法の有効性を確認するためには、EDR 概念辞書との合致率が高い条件で同様に評価する必要があると考える。

5.2 入力記事と用例文の類似度

入力記事と用例文の類似性を測るため、評価者 1 人が次の評価を行った。入力記事に対して従来手法と本手法の選択した用例文を読んでもらい評価を行った。調査には評価実験データを使用した。評価基準は次の 2 値である。

・評価 1.経験的に似ている文である。
・評価 2.経験的に似ている文でない。

結果を表 7 に示す。

表 7 入力記事と用例文に関する類似度調査

評価 1	評価 2
47/100 (44%)	53/100 (8%)

評価 2 が評価 1 より上回る結果となった。入力記事に対する用例文の選択を行う処理の部分で、評価 1 になった件数が半分以下であることが分かる。原因として入力記事と用例文の類似度に問題があると考える。入力記事と用例文の類似度は、用例文の構文片から見て入力記事と類似度が最大となる値を加算平均で計算している。構文片の中に存在する内容語はシソーラス上で複数のノードが存在する場合がある。本手法は複数のノードが存在する場合、類似度が最大となるノードを選択している。しかし、最適となるノードは文脈によって変化するので、類似度が最大となるノードでは不適切な場合がある。文脈に対して、内容語の最適なノード選択を行うことが課題である。上位ノードの情報を用いることで、誤ったノード選択を減らすことが出来るので、より意味的に近い類似度が計算できると考える。しかし、誤ったノードが上位に存在した場合、類似度を別途考慮する必要がある。また誤ったノードの基準を設ける必要がある。

5.3 内容語、構文片の類似度

入力記事と用例文の類似度の計算に用いた内容語、構文片の類似度について計算方法に問題がなかったか調査した。内容語と構文片それぞれの類似度がどの程度人間の主観に近い指標か調査した。調査方法は類似度が最大値 1 ではなかった内容語、構文片の中で上位 100 件について人手評価を行った。調査には評価実験データを使用した。調査基準は次の 2 値である。

・評価 1.経験的に似ていると判断できる。
・評価 2.経験的に似ていると判断できない。

評価値 1 の割合について、結果を表 8 に示す。

表 8 内容語、構文片の類似度調査

	内容語	構文片
評価者 A	62/100 (62%)	34/100 (34%)
評価値 B	38/100 (38%)	22/100 (22%)
評価値 C	64/100 (64%)	42/100 (42%)

内容語で評価1は平均で55%であったが、構文片に拡張することによって、33%となった。

内容語の類似度について次のように考える。内容語の類似度の観察結果、類似度が高いほど評価1が多い傾向であることが分かった。そこで、内容語の類似度が 0.9 以上の中で評価 1 の一致率を調べた。0.9 以上の内容語は 14 件中、評価者 2 人以上が評価 1 とした件数は 13 件あった。具体例として「経常黒字」と「利益」、「決定」と「決める」、「森首相」と「セディジョ大統領」などが挙げられる。内容語は類似度が 0.9 以上であると過半数で評価 1 が得られることが分かった。この結果より、0.9 以上の内容語の類似性は人間の主観に近いと考える。

構文片の類似度について同様に考える。類似度が 0.9 以上の構文片は 2 件であった。評価者 2 人以上が評価 1 となった件数は 0 件であった。具体例として表層形で「政府雇用対策案が明らか」と「予定であることを明らかにした」、「信託業法改正案が明らか」と「意向を明らかにした」があった。

構文片の類似度を内容語の類似度による加算平均のみでは類似度を計算出来ないと考える。構文片の類似度を改善することにより、入力記事に対して意味的に近い用例文が出来ると考える。格フレームを考慮した機能語の類似性などが改善点として挙げられる。

6 結論

入力記事に対して意味の近い要約文を生成するため、構文片とシソーラスを用いて要約を行うシステムを構成した。評価では自動評価、人手評価共に従来手法より下回る結果となった。入力記事と用例文の類似度の加算平均が原因であると考える。内容語のノードが複数存在した場合の処理について再検討が必要である。内容語の類似度と主観の一致性は類似度が0.9を越える必要があることが分かった。内容語から構文片の類似度に拡張して考える際、機能語の考慮が必要であることが分かった。今後は、格フレームを用いた機能語の形態素について、類似度を検討する予定である。また、構文片の類似度から構文木の類似度に拡張させたいと考えている。

参考文献

[1] 堀智織,古井貞熙,単語抽出による音声要約生成法とその評価,電子情報通信学会論文誌,Vol.J85-D-Ⅱ,No.2 pp.200-209 , 2002.

[2] 小黒玲,尾関和彦,張玉潔.文節重要度と係り受け整合度に基づく日本語簡約アルゴリズム.自然言語処理,Vol.8,No.3,pp.3-18, 2001.

[3] 山本和英.牧野恵,要約事例を用例として模倣利用したニュース記事要約.自然言語処理学,vol.15,No.3,pp.115-158,2008.

[4] 青木優,山本和英.構文片を用いた分野の同定を必要としない意見・評判情報抽出.電子情報通信学会技術研究報告 NLC2007-88,pp.7-12,2008.

[5] 長尾真(編),岩波講座ソフトウェア科学 15 自然言語処理,岩波書店,pp.236-240,1996.

使用した言語資源及びツール

(1) 日経ニュースメール,NIKKEI-goo, https://letter.goo.ne.jp/nkgmail/member.cg
(2) EDR,概念辞書,情報通信研究機構(NICT) http://www2.nict.go.jp/r/r312/EDR
(3) 日本経済新聞全記事データベース 1998 年度版, 日本経済新聞
(4) 自動要約評価ツール BLEU,ROUGE-4(日本語版),広島市立大学 竹澤研究室. http://www.nlp.its.hiroshima-cu.ac.jp/