データベースシステム 115－6
情 報 学 基 礎 49－6
（1998. 5. 14）

単語の重要度に基づくテキストの要約

平尾 努　木谷 強
NTT データ通信 (株) 情報科学研究所
{hirao,tkitani}@lit.rd.nttdata.co.jp

あらまし
本稿では，単語の範囲内重要度に基づき文の重要度を決定し，要約文を抽出する方法について述べる．提案する手法では，単に重要度の高い文を要約文とするのではなく，隣接する文の重要度の変化に着目して要約文を抽出する．新聞記事を対象として実験を行ない，人手により抽出した正解データから適合率・再現率を求める方法と，提案手法により抽出した要約文に対するアンケート調査とによって提案手法を評価した．評価の結果，提案手法により，人間が最重要と考える文を平均 63%の精度で抽出することができた．また，提案手法を発展させることにより，精度の向上が期待できることもわかった．
キーワード 自然言語処理，自動要約，情報検索，情報抽出

Text Summarization based on Word Importance

Tsutomu HIRAO and Tsuyoshi KITANI
Laboratory for Information Technology
NTT DATA Corporation
{hirao,tkitani}@lit.rd.nttdata.co.jp

Abstract
In this paper, we describe a method of determining the importance of sentences based on word importance within a text region and extracting summary sentences. The summary sentences are not simply chosen by the importance, but are determined by the change of importance between adjacent sensences. This proposed technique was evaluated using newspaper articles. Two different evaluation methods: recall and precision rates by comparing the extracted summaries with manually selected ones, and human assessments to the extracted summaries. The proposed method extracted sentences that the human assessor judged as the most important on an average of 63%. We should be able to achieve a better accuracy when the proposal technique is futher improved.
Keywords Natural Language Processing, Text Summarization, Information Retrieval, Information Extraction

1 はじめに

近年，インターネットに代表されるネットワーク環境の発展や，CD-ROM，MO 等の大容量メディアの低価格化による普及により，大量の電子化テキストデータが氾濫している．このため，大量の情報から必要とする情報を効率良く選択する必要が生じている．従来から情報検索に関する研究が盛んに行

なわれているが，検索結果を適度な量に絞り込んだ場合でも，その内容はテキストを実際に読まなければ分からない．結局，情報の必要性は人間が目を通して判断する必要がある．このような負担を軽減するため，テキストの全文を読まなくとも内容を把握できる手段としてテキストの要約技術が活用されている．テキストを要約することで大意を汲みとることができれば，効率良く情報を取捨選択することができる．

本稿では，単語を範囲内重要度 [7] に基づき重み付けを行ない，文の重要度を決定し，文の重要度から要約文を抽出する方法について述べる．要約文は，単に重要度の高い文とするのではなく，文の重要度の変化からテキストをセグメントに分割し，それぞれのセグメントから抽出する．

以下，2 章では，自動要約技術の動向を述べ，3 章では要約文を決定するために必要となる単語の重要度について説明する．4 章では本稿で提案する要約手法を説明し，5 章では実装したシステムを用いた評価について述べる．また，6 章では，5 章の結果をもとに提案手法とその評価法を考察する．

2 従来の要約手法とその問題点

従来の計算機による自動要約の研究は，

- 生成派
- 抽出派

の 2 つに大きく分けることができる．

生成派の目的は，狭義の要約である．つまり，テキストの内容を理解した上で重要な情報を抽出し，それらをもとに可読な要約文を生成することを目指している．しかし，狭義の要約の実現のためには自然言語の意味理解や文生成などの技術が必要であるが，現状では完全に解決できないままに問題が残されている．

一方，抽出派の目的は広義の要約である．テキスト中の内容を把握する上で重要であると考えられる文章をそのまま抜き出す[1]ことで，要約と定義している．広義の要約の際の文の抽出手法には，

1. 言語の表層情報を用いる手法
2. 統計情報を用いる手法

が提案されている．言語の表層情報を用いる手法としては，接続詞等に着目して，文章構造を類推することにより，テキストの重要箇所を抽出する方法 [5] や，予め重要な情報が分かっている場合にパターンマッチングを用いる手法 [4] がある．統計情報を用いる手法としては，語の出現頻度情報を用いて重要度を決定し，重要語を含む文を要約として抽出する手法 [6] がある．また，言語の表層情報と，統計情報を組み合わせた手法も提案されている [9]．

言語の表層情報を用いる手法の問題点としては，特定の記述あるいは構造を持ったテキストにしか適用できないことが挙げられる．一方，言語の統計情報を用いた手法はあらゆる形式のテキストに対応できるが，単に重要語を含む文や重要度の高い文を抽出するため，テキストに複数の話題が存在する場合に対応できないという問題点がある．

本稿で提案する手法は，言語の統計情報を用いる手法に属するが，上述の問題点を解決するために，文の重要度の変化からテキストを複数のセグメントに分割し，セグメント毎に重要度の高い文を抽出する手法を採用する．本手法を用いることで，テキストに複数の話題が存在する場合にも対応できる．

3 単語の重要度

本稿で提案する要約手法では，文の重要度を決定するために語の統計情報を用いる．そこで本章では，[7] で提案されている範囲内重要度を用いた語の重要度の決定法について述べる．

3.1 範囲内重要度

従来の要約手法では，単語の出現頻度に基づく単語の重み付けが一般的に用いられる．これは，「多くのテキストに出現する語の重要度は低く，特定のテキストに多く出現する語は重要である」という仮定に基づいている．しかし，このような一般的な重み付けの方法では予め語の統計情報を持っていなければならなく，また，テキストの要約を考えた場合には，個々のテキストに閉じた単語の頻度情報等を使用する方が有効[2]である．そこで本稿では，範囲内重要度 [7] を適用して単語の重み付けを行なう．

[1] この際，要約の可読性，流暢さはあまり問題にされない．

[2] 我々がテキストの表層情報を元に要約するプロセスでは，テキスト内での繰り返しや出現位置等を手がかりとしている事に基づく．

3.1.1 範囲内重要度の定義

テキストの要約を作成する場合には，出現する語に関する関連性を考慮する必要がある．[7] では，

仮定 1 多くの特定範囲内で同時に出現する語どうしは関連性が高い

仮定 2 特定範囲に出現する語は，その範囲に出現する全語数が少ないほど重要性が高い

という 2 つの仮定を考え，ある単語 T の範囲内重要度 $C_r(T)$ を式 (1) で定義した．

$$C_r(T) = \frac{1}{M}\sum_{i}^{M}\frac{\varepsilon_i}{N_i} \tag{1}$$

ただし，

$$\varepsilon_i = \begin{cases} 1 & (A_i\text{にTが存在するとき}) \\ 0 & (A_i\text{にTが存在しないとき}) \end{cases}$$

式 (1) は，M 個の範囲 $A_1,..A_i,..,A_m$ 内に単語がそれぞれ，$N_1,..N_i,..,N_M$ 種類存在する時，範囲 A_i に存在する単語の範囲 A_i 内での重要度を $1/N_i$ と定め，範囲全体での重要度の平均を範囲内重要度と定義したものである．$C_r(T)$ は，$0 \leq C_r(T) \leq 1$ を満たし，値が大きいほど重要であることを表す．

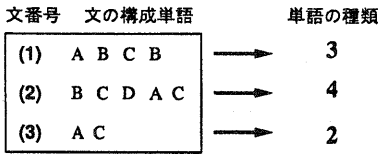

図 1: 範囲内重要度の決定例

本稿では，A_i の範囲を 1 文とする．図 1の例において，単語 A，B の重要度 $C_r(A)$，$C_r(B)$ はそれぞれ，

$$C_r(A) = \frac{1}{3}\left(\frac{1}{3}+\frac{1}{4}+\frac{1}{2}\right) = \frac{13}{36}$$

$$C_r(B) = \frac{1}{3}\left(\frac{2}{3}+\frac{1}{4}+\frac{0}{2}\right) = \frac{11}{36}$$

となる．単語 A，B とも同じ 3 回の出現回数であるが，範囲内重要度では，多くの範囲で出現する A の方が重要度が高い．

3.2 重要度の決定

前節では，単語の範囲内重要度について述べたが，単語の重要度を決定する要因として単語の出現頻度も考えられる．そこで，単語 T の出現頻度をテキスト全体の単語数で正規化した単語 T の出現確率 (式 (2)) を考える．

$$C_f(T) = \frac{Freq(T)}{\sum_k Freq(T_k)} \tag{2}$$

単語の重要度 $W(T)$ は，T の範囲内重要度 $C_r(T)$ と T の出現確率 $C_f(T)$ の和で定義する．

$$W(T) = \alpha C_r(T) + \beta C_f(T) \tag{3}$$

ただし，

$$\alpha, \beta \text{ は定数 } (\alpha+\beta=1;\ \alpha,\beta \geq 0)$$

以上で述べた範囲内重要度は，キーワード抽出法 [7] において，単語の重要度決定に有効な結果が得られており，本稿ではテキスト要約技術に適用する．

4 要約文の抽出法

本稿で行なうテキストの要約とは，テキスト中の重要文を抽出することである．

単語の重要度に基づき要約文を抽出する手法としては，

1. 重要度の高い単語を含む文を抽出する手法
2. 単語の重要度から文の重要度を計算し，重要度の高い文を抽出する手法

がある．

1 の手法では，重要単語を含む文を網羅的に抽出するため，重要語がテキストの多くの文に出現する場合には抽出文数が多くなる．また，重要語を含んでいれば，その内容にかかわらず文を抽出するという問題点もある．

2 の手法では，抽出する要約文は文の重要度順，あるいは，文の重要度が閾値を越えるなどの条件で抽出される．しかし，これらの抽出法ではテキストに複数の話題が含まれる場合にそれぞれの話題の重要文を抽出できない．

本稿では，このような問題点を解決するため，テキスト内における文の重要度の変化からテキストをセグメントに分割し，各セグメントから重要度の高い文を要約文として抽出する手法を提案する．

4.1 文の重要度

要約文として抽出する文の決定は，テキストにおける文の重要度に依存する．そこで，3章で定義した単語の重要度をもとに一文毎の重要度を定義する．テキスト中のある文 S_j の重要度 $W(S_j)$ は，以下の式 (4) で定義する．

$$W(S_j) = \sum_n W(T_n)/N \qquad (4)$$

式 (4) は，テキスト中のある文 S_j の重要度は，S_j に出現する単語の重要度の総和を S_j に出現する総単語数 N で正規化したものである．

4.2 要約文の選択

本稿では，単に重要度の高い文や重要度がある閾値を越えた文を要約文として抽出するのではなく，テキスト内の文の重要度の変化に着目し，抽出する文を決定する．これは，以下の仮定による．

仮定 テキストの文の重要度の変化が意味の変化を表す

テキストの先頭文の重要度から最終文までの重要度の変化から，図 2に示すグラフのように重要度の高低差により，「山」，「谷」ができる．上記の仮定に基づき，「谷」から「山」に移行する部分で意味が変わったと考え，「谷」部分でテキストを分割すると，複数のセグメントができる．このセグメントがいわゆる意味段落に相当するものと考えられる．つまり，テキストの話題が，それぞれのセグメントに相当するものと考えられる．本稿では，セグメント毎に重要度の高い文を要約文として抽出することで，従来では対応できなかった複数の話題を持ったテキストからの要約を可能とする．

5 評価実験

3章，4章に基づき要約システムを実装し，システムが抽出した要約文の精度を評価するための実験を行なった．要約の対象としては，検索ベンチマーク用テストセット $BMIR-J1$[3][1] に収録されている新聞記事から，テキストの文数が10行以上の任意の100件を用いた．表 1に使用した記事100件の詳細を示す．

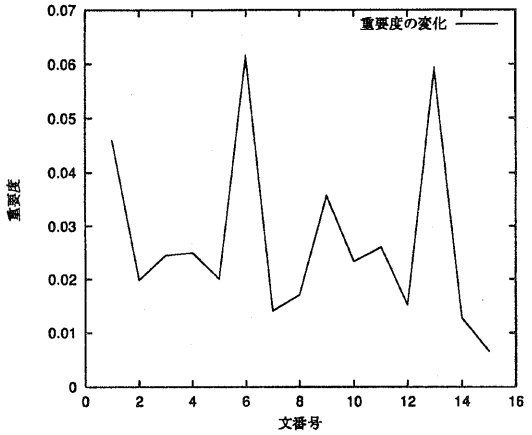

図 2: 文の重要度の変化

表 1: 記事の詳細

平均文数	平均サイズ
15文	1.4kbyte

5.1 要約文の抽出手順

本要約システムの手順を以下に説明する．

Step1 テキスト中の単語 (名詞) に式 (3) を用いて重要度を付与する．

Step2 記事のタイトルに含まれる単語がテキスト中に出現した場合には，Step1 で求めた重要度を n 倍にする[4]．

Step3 一文毎の重要度を式 (4) を用いて求める．

[3]株式会社 日本経済新聞社の協力によって，社団法人 情報処理学会・データベースシステム研究会・情報検索システム評価用データベース構築ワーキンググループが，1993 年 9 月 1 日から 12 月 31 日の日本経済新聞記事を基に構築した検索評価用データベース (テスト版) を利用．

[4]新聞記事は，そのタイトルである程度内容が理解できるようになっているため，タイトルに出現する単語は重要であると考えられることによる．なお，本実験では経験的に $n = 4$ とした．

Step4 テキスト中の文 S_j おける重要度 $W(S_j)$ と S_j の関数を関数 $F(S_j)$ と考える (図 2).

Step5 関数 $F(S_j)$ の極大値を与える S_j を要約文として抽出する.

Step6 Step5 で抽出した文 S_j の前後の文 S_{j-1}, S_{j+1} に対して,

$$\frac{W(S_{j-1})}{W(S_j)} \geq \gamma$$

$$\frac{W(S_{j+1})}{W(S_j)} \geq \gamma$$

を満たす S_{j-1}, S_{j+1} を要約文として抽出する. ただし, $0 \leq \gamma \leq 1$ とする.

以上, Step1 ～ Step6 で複数の文を要約文として抽出する. 抽出した要約文が要件より多い場合には, 抽出した要約文を対象として Step1 ～ Step6 を再度適用する. 要件を満たす文数にならない場合には, 要件に最も近い文数とする.

本実験では式 (3) の α, β の値をそれぞれ, 経験的に, 0.7, 0.3 とした. また, γ については, 文書全体の意味を表すために必要な文数が全体文数の約 3 割程度であると考え, 要約の文数が全体の 3 割になるよう γ の値を設定した.

5.2 正解データ

要約の評価のため, 100 件の記事を人手で要約し, 正解データを作成した. 作成法は, 100 件から任意の 10 件を取り出し, 10 数名の人に要約をしてもらい, その抽出傾向を参考に残りの 90 件について 2 人で要約を作成した. 抽出する文数は, 全体のシステムの出力と同等の 3 割程度とした. また, 人間の主観に基づいた要約が出来ているかを判断するため, 抽出文から最も重要と考える 1 文を選んだ.

5.3 評価

前述した 100 件の記事に対し, 5.1 節で述べた手法で要約を行なった. 要約結果の詳細を表 2に示す.

また, 5.1 節で述べた要約システムの評価のため, Precision (適合率) , Recall (再現率) という評価尺度を用いて評価した. 以下に, Precision, Recall の算出式を説明する.

表 2: 抽出した要約文の詳細

平均文数	平均サイズ
5 文	0.63kbyte

$$Precision = \frac{C}{A}$$

$$Recall = \frac{C}{B}$$

ただし, A はシステムの要約文数, B は人間の要約文数, C はシステムの要約文と人間の要約文において一致する文数である.

表 3に 100 件全体の Recall, Precision の平均を示す. また, 参考として, A 社のワープロソフトの要約機能を用いて要約した結果も併せて示す.

表 3: システムの精度 (100 件の平均)

	Precision の平均	Recall の平均
本システム	0.50	0.49
A 社	0.43	0.52

人間が最も重要と判断した文が抽出できている割合, 人間の抽出した文のうち少なくとも 1 文がシステムの要約文に含まれている割合を表 4に示す.

表 4: 重要文の抽出精度

	最重要文	すくなくとも 1 文
本システム	63%	99%
A 社	62%	100%

5.4 アンケートによる評価

人間が作成した要約の正解データ, システムが抽出した要約から, 5.3 節での評価結果で, Precision, Recall の高い文書と低い文書を各 10 文書づつ選びアンケートによる評価を行なった. それぞれの要約文書を以下の 5 段階で評価するようにした.

評価 5 大変良く分かる

評価 4 良く分かる

評価 3 普通

評価 2 少しわかりにくい

評価 1 わかりにくい

アンケートによる評価結果を表 5に示す (文書番号に続く数値は Precision, Recall を表す). アンケートの有効回答数は 11 件であった.

表 5: アンケートによる評価結果

文書番号	人間の要約	システムの要約
1 (1.0,0.80)	3.8	3.5
2 (0.67,0.80)	3.8	1.6
3 (0.60,0.60)	3.9	3.0
4 (0.50,0.40)	3.8	1.8
5 (0.50,0.33)	3.8	2.6
6 (0.43,0.50)	3.9	2.4
7 (0.38,0.50)	2.9	2.0
8 (0.40,0.40)	4.3	2.1
9 (0.25,0.25)	3.6	2.8
10 (0,0)	3.8	1.8

6 考察

6.1 提案手法について

4.2 節で述べたテキストの分割について, 5.4 節で用いた 10 文書[5]を対象に検証する. それぞれの新聞記事に設けられた段落と 4.2 節で述べた手法でテキストを分割した結果を比較すると, 1 文書の平均で人間が設けた段落数は 5 個, システムが分割したセグメント数が 5 個であった. また, 両者間で一致するセグメントは, 1 文書あたり, 平均で 1.9 個であった. 文書の先頭のセグメントでは 10 文書中 5 文書で一致し, 末尾のセグメントでは 10 文書中 4 文書で一致した. 文書の先頭, 末尾以外のセグメントで一致する割合は低いが, 分割の際に 1 文あるいは, 2 文ずれている. 以上より, 精度については今後更に検討が必要であるが, 4.2 節での仮定は, 段落情報が与えられていない場合には有効であると考える.

5.2 節の正解データを分析した結果, 人間が要約文として抽出した文は, テキストの先頭の 1 ～ 2 セグメントに集中していることが分かった. 本実験で使用したテキストが新聞記事であるため, 重要情報がテキストの先頭部分に集中し, 複数の話題が記述されることが少ないことが原因であると考えられる[6]. よって, テキストの後部の記述は重要度が低いと判断され, 抽出されにくい. しかし, 提案手法では, 文の重要度が極小となる全ての部分でテキストを分割するため, 人間が重要度が低いと判断したセグメントからも要約文を抽出しており, その結果が, 表 3, 表 5に反映したものと考える. よって, 各セグメントの重要度を計算し, 重要度の高いセグメントから要約文を抽出する等の対処法をが必要である. また, 同一テキスト内の異なるセグメントが, 意味的に等価である場合も考えられ, セグメントの類似度も考慮に入れる必要がある.

6.2 正解データについて

表 5より人間が作成した正解に対するアンケート結果から, 評価値は平均で 3.8 である. 人間の作成した要約であっても, 評価はあまり良くないないことから, 重要文に対する判断が人により異なることがわかる. これは, 必要とする情報が人により異なり, 重要と判断する文に差異が生じるためである. また, [8] では, 被験者 112 名が作成した要約を *Kappa* 統計 という尺度で評価した結果, 人間が作成した要約文の信頼性が低いことを報告している. これらのことから, 信頼性の高い正解を作成することが困難であることがいえる.

6.3 評価法と精度について

[2], [3] 等でも, 要約を Precision, Recall という指標を用いて評価している. しかし, この評価法が必ずしも要約の精度を表すとは限らない. 例えば, 表 5の文書 2 では, Precision, Rcall がそれぞれ, 0.67, 0.8 という値であるのに対し, アンケートでの評価値は, 1.6 と 10 文書中最低の評価となっている. この原因としては, 人間が最も重要であると判断した文を抽出できなかったことが挙げあら

[5] 1 文書あたりの平均文数は 14 文である.

[6] 複数の話題から重要文を抽出するためには新聞記事は向いていない.

れる．最重要文が抽出されていないことから，要約文の意味が理解できなくなるからである．逆に，人間が最も重要であると判断した文を抽出できれば，文書5のように，Precision，Recallが共に低くとも，2.6というある程度の評価値は得られると考えられる．

表3より，本システムのPrecsion，Recallはともに約50%である．PrecisionはA社のワープロソフトの要約機能を上回った．しかし，上述したように，この結果をそのままシステムの評価とするには問題がある．表4から人間が最も重要と考える1文を抽出する精度では，ほぼ同等の精度であり，2つのシステムの要約文を人間が評価した場合には，大きな精度差はないと予想される．つまり，要約の評価の際には，単にPrecision，Recallを用いるのではなく，人間の持つ重要文に対する優先順位を反映するような評価を考える必要がある．

検討段階であるが，4.2節に述べた手法でテキストをセグメントに分割した後，それぞれのセグメントに対し，4.1節の式(4)を用いて文の重要度を決定し，セグメント内で重要度が最も高い文を抽出するという手法を任意の20文書で試した．その結果，人間が最も重要と判断する文の抽出率は70%であった．表4では精度63%であったので，上述の手法を適用すると更に精度の向上が期待できる．

7 まとめ

単語の範囲内重要度に基づき，文の重要度の変化からテキストをセグメントに分割し，各セグメントの重要文を要約文として抽出する手法について述べた．Precision，Recallは100件の平均で，それぞれ，50%，49%であった．人間が最も重要と判断した文のうち平均63%を抽出できた．さらに，提案手法を用いてテキストをセグメントに分割した後，各セグメントを範囲として範囲内重要度に基づき文の重要度を計算して抽出する要約文を決定することにより，精度の向上が期待できることがわかった．抽出した要約文に対し，アンケートによる評価を行なった結果，Precision，Recallを用いた評価では，必ずしも人間の直観に合う訳でないことがわかった．人間が重要と考える文を含めることが必要であることが分かった．

今後の課題としては，より分かりやすい要約文の抽出手法を検討する必要がある．また，妥当性のある正解データの作成法，人間の直観を満たす要約文の評価法も検討課題である．

謝辞

システムの実装にあたり，数々の有効な意見を下さった，NTTデータ通信(株)技術開発本部ソフトウェア技術センタの喜多淳一郎氏に感謝いたします．また，評価に協力戴いた東京大学工学部計数工学科 桜井一俊氏に感謝いたします．

参考文献

[1] 芥子育雄ほか．情報検索システム評価用ベンチマーク Ver1.0(BMIR-J1)について．情報処理学会研究報告会，DBS-106-19,1996.

[2] 亀田雅之．段落間及び文間関連度を利用した段落シフト法に基づく重要文抽出．情報処理学会研究報告会，NLP-121-17,1997.

[3] 黒武者健一，芥子育雄．連想検索技術を利用した文書の要約．*Proceedings od Advanced Database Symposium '97*, 1997.

[4] 佐藤理史，佐藤円．ネットニュースグループfj.wantedのダイジェスト自動生成．自然言語処理，Vol. 3, No. 2, 1996.

[5] 山本和英，増山繁，内藤昭三．文章内構造を複合的に利用した論説文要約システムgreen．自然言語処理，Vol. 2, No. 1, 1995.

[6] 小部正人．文章抄録装置．特開昭61-117658,1988.

[7] 原正巳，中島浩之，木谷強．テキストのフォーマットと単語の範囲内重要度を利用したキーワード抽出．情報処理学会論文誌，Vol. 38, No. 2, 1997.

[8] 野本忠司，松本裕治．人間の重要文判定に基づいた自動要約の試み．情報処理学会研究報告会，NLP-120-11,1997.

[9] 渡辺日出雄．新聞記事の要約のための一手法．言語処理学会 第1回年次大会，1995.