自 然 言 語 処 理 149－11
音声言語情報処理 41－2
（２００２．５．２４）

単語の文書頻度と文の長さを利用した抄録縮約方式

小峰 恒[†] 絹川 博之[†] 中川 裕志[‡]

[†]東京電機大学大学院 工学研究科
komine@cll.im.dendai.ac.jp kinukawa@im.dendai.ac.jp
[‡]東京大学 情報基盤センター

近年，携帯電話や PHS を用いての Web ページの閲覧サービスのニーズが高まってきている．ところが，従来の Web ページはパソコンなど大画面での閲覧を前提としているため，携帯電話や PHS など表示画面の小さい端末では画面切り替えの回数も増え，使いにくいものとなっている．この点を解決するため，Web ページを縮約することが必要である．そこで Web 表示サービスされている図書概要を対象に，さらに短く縮約する方式を開発することとした．

まず図書概要の縮約方式として，単語の重み付けによる方式を採用することとし，tf 法，tf-idf 法による従来の単語重み付け方式と異なる，df 法による単語の重み付け方式を新たに提案する．実験評価の結果，新しく提案した df 法が一番高い精度，再現率となったが，文の長さも重要な情報であり縮約を行う際に考慮すべきであることもわかった．

そこで，単語の重みに文長を複合させた縮約方式を提案し，前述の 3 つの重み付け方式に文長を複合させて実験評価を行った．その結果，df と文長の組み合わせによる複合型抄録縮約方式の精度，再現率が最高値となることを確認できた．今後の課題として，他の文書集合への適応を検討している．

Sentence Extraction based on Document Frequency and Text Length

Hisashi Komine[†] Hiroshi Kinukawa[†] Hiroshi Nakagawa[‡]
[†]Graduate School of Engineering, Tokyo Denki University
[‡]Information Technology Center, The University of Tokyo

In recent years, the demand of Web pages browsing by using cellular phone or PHS is increasing. But it is difficult to browse Web pages by using mobile terminals. Because Web pages are made for large size display of personal computers in order to solve this problem, it is neccessary to summarize Web pages. Then we propose a new method, in order to summarize book abstracts.

We use the method of term weighting for summarizing book abstracts. The conventional sentence extraction methods are based on term frequency (TF-method) and term frequency by inverse document frequency (TF･IDF-method). We propose a new sentence extraction method that is based on document frequency (DF-method). The experimental evaluation shows that the DF-method is the highest precision and recall ratio of the three methods, but turns out the text length to be taken into consideration.

Then, we newly propose a sentence extraction based on document frequency and text length (DF･TL-method). Experimental evaluation shows that precision and recall ratio of the DF･TL-method are higher than that of DF method and the highest of all. As future subject, we will apply the method for other document sets.

1. はじめに

近年，技術の発達に伴い，PHS や携帯電話など Web ページを閲覧できる携帯端末が増えてきている．ところが，Web 上に存在するページのほとんどが，パソコンなどの大画面にて作成されているため，閲覧する側もパソコンであることを前提に入れて設計されており，冗長に書かれている場合も決して少なくない．そのため，表示領域が狭い携帯端末では，ほとんどの場合 Web ページが一画面では収まりきらない．また，携帯端末の多くでは，通信コストは通信データ量に比例する．このようにコストという面からでも，通常の Web ページは携帯端末に表示に向いていないことは明白である．

そこで，本研究では，図 1 のように携帯端末に直接 Web ページを送るのではなく，無駄な情報を省いて重要な個所のみを抽出し，小さい画面に表示するのに適切な長さに縮約する方式の開発を目的とした．

図 1　研究の方針

1.1. 本研究の目的

今回，研究の対象として，東京大学情報基盤センター[1]にて提供されている抄録データベースであるブックコンテンツ内の図書データを利用した．図書データの例を図 2 に記す．

その図書データから抄録である Description を縮約の対象とする．その内容は，以下に記した通りである．

(a) およそ 1000 文書程度の調査では文字数は 42～486 字であり，平均 181 字である．文数はおよそ 3 文となっている．近年の携帯電話では一画面に表示できる文字数が 50 字から 100 字程度となっており，携帯端末に表示するのには，文が長いと考えられる．

(b) 本の内容を簡潔に説明している．すなわち，本の要約とも言える．

これらの特徴をもつ抄録を，携帯端末表示画面の制約に合わせて，50 字以下にすることを目標とする．

ID: NB0430909
Title: ニューロ・ファジィ・遺伝的アルゴリズム
Series: エレクトロニクス実践シリーズ
Author: 萩原／将文【著】
Publisher: 産業図書
Year: 1994
Description: 本書では、新しい技術として特に大きな注目を集めているニューラルネットワークとファジィ、それから遺伝的アルゴリズムを総合的に扱い、個々の技術の背景から基本と応用、さらにそれらの融合方法までをポイントをおさえて解説しています。高校生程度の方々でも理解できるように、解説はできる限り平易になるよう心掛けました。
Contents: *1章 はじめに
*2章 ニューラルネットワークとは何だろう
*3章 ニューラルネットワークの実際
*4章 ファジィの原理を学ぼう
ISBN: 4782855087

図 2　東京大学のブックコンテンツ例

2. 重要文抽出の方式

2.1. 重要文抽出における従来の方式

重要文抽出の手法として従来([4][5])述べられているようないくつかの方式が知られており、以下のように分類できる．

これらの方式は，文の抽出を行う際に良く使用されるものである．

(1)　文の出現位置を基に文を抽出する方式

文書には，ある種のスタイルを持って書かれている場合が多い．その考えから，その規則性を用いて，重要だと思われる文を抽出するのがこの方式である．

例としては，lead 法が上げられる．lead 法は文の先頭を重要視し，文を抽出する方式である．新聞など，先頭に大まかな内容が書かれるような文書に対して，非常に有効な方式である．

(2) 単語の重み付けにより文を抽出する方式

文書中に使用される単語に，適当な技法で重み付けを行い，重みの高い語を重要とみなす．次に重要な単語が使われている文は重要であるという考えから，文を抽出する方式である．

従来の重み付け方としては，tf 法や tf-idf 法が知られている．

2.2. 図書抄録データの抄録中の重要個所

以下に, ブックコンテンツの図書抄録データの抄録の中で, 人手により重要と判断された文の例（下線部）を二つ示す.

例 1

本書は, 光・レーザをいろいろな工学分野に応用するという視点に立って書かれている. したがって, 応用する際に必要となる基礎的内容だけを, できるだけ直感的に理解しやすいように説明した.

例 2

初心者にとって良き入門書であると同時に, 熟達した研究者には実験手法の視野を広げ新しい発想のヒントになるように配慮した新しいタイプの実験書. 多くの実験室で共通の基礎的手法であり, 学部 4 年生あるいは修士課程の院生にとってその習得は必須. 測定系の制御やデータ収録・解析に不可欠なエレクトロニクス, センサー, トランスデューサー, コンピュータの具体的かつ実用的な解説を行う.

この例に前述の従来の方式を当てはめた場合を考えてみる.

(1) 抄録は, 本の内容を簡単に説明する必要があるために, 短く書かれている場合がほとんどである. そのために使われる単語が少なく, 例 1 のように, 全ての単語の単語頻度が 1 であることも少なくない. このような文書では, 先ほど述べた tf 法や tf-idf 法では, 単語の重み付けの際, 差が明確に現れないため, 抽出すべき箇所を特定するのに有効であるとは言えない.

(2) 例 2 のように, 図書の抄録は新聞のように重要部が先頭に書かれているとは限らないものが多いため, lead 法を用いても重要な効果が得られとは限らない.

このように, 図書の抄録から重要文の抽出を行うためには先ほど述べた従来の方式は適切ではない. したがって, 従来とは異なる新しい方式が必要であると考えられる.

2.3. 単語の重みを利用した抄録縮約方式

様々な図書抄録データを調査したところ, 抄録には, その書き方に特徴があるということがわかった.

(1) 『本書は』という出だしで書かれた文中には当該図書の要点が簡潔に書かれていることが多い. 2.2 の例 1 のように「本書は〜〜書かれている」とある場合, この内容を見れば, その部分が図書の内容を書いてあるということは予想できる.

(2) 『本書』の他に『解説』などの特徴語も, 同様の理由で要約を表す文には多く使われている. 2.2 の例における網かけ部分がそれにあたる. 本発表では, それらの本の内容を示す文によく使われる単語を**文章特徴語**と呼ぶことにする.

(3) 文章特徴語は多くの文書で使われており, 文書頻度（df）は高いと考えられる.

表 1 は, 上記ブックコンテンツから

① 『生命』&『バイオ』
② 『建物』&『建築』
③ 『物理』&『運動』
④ 『エレクトロニクス』
⑤ 『自然』&『環境』&『研究』
⑥ 『電気』&『通信』
⑦ 『現代』&『経済』&『金融』
⑧ 『ロボット』
⑨ 『政治』&『倫理』

以上の, 学術系のキーワード 9 つによって検索された文書, 計 3019 を対象に各単語の文書頻度を計算し, 降順に並べた結果である. 割合は, 文書全体に対し, どれほどの出現率であるかを示す.

表から『本書』や『解説』など, 図書の説明に重要だと思われる単語の文書頻度が高いことがわかる. また,『経済』や『現代』など, 学術系図書に無関係と思われる単語が含まれているのは, 検索キーワードとして用いたからである.

以上を踏まえて, 我々は df 法を提案する[3]. すなわち, 文章特徴語が多く使われている文は本の内容を示す文であるという考えから, df の高い語つまり文章特徴語を含む文を抽出する方式である.

従来, 単語の重み付けによる方式として用いられる tf 法, tf-idf 法と異なり, 文書頻度 df の高い語を重要な単語と見なして, 重要な単語を含む文を抽出する方式である.

これは, 直観的には重要文抽出に用いられる

情報の一つである手がかり表現を利用する方式と言うことができる.

表１　学術系図書抄録の文書頻度

順位	単語	文書頻度	割合 (%)
1	本書	955	31.6
2	解説	475	15.7
3	技術	446	14.8
4	研究	398	13.2
5	問題	303	10
6	基礎	268	8.9
7	経済	268	8.9
8	現代	266	8.8
9	書	253	8.4
10	分野	244	8.1
11	環境	232	7.7
12	日本	228	7.6
13	者	223	7.4
14	社会	220	7.3
15	世界	209	6.9
16	情報	205	6.8
17	理解	196	6.5
18	科学	185	6.1
19	理論	180	6
20	応用	177	5.9

2.4. 抽出すべき抄録単位

縮約の目標は 50 字としたが，表 2 で示されている通り，抄録の文の平均文字数は 60 字程度の文も多くみられた．そのため，今回の抽出する抄録単位は文ではなく, 節にすることにする．ここで言う節とは，句読点で区切られた単位を指す．ただし，以下の条件では区切らないこととする．

(1) 直前が接続詞，係助詞の場合

接続詞は単体で節になるために区切る必要がないためであり, 係助詞の場合はその直前が主格になるからである．

(2) 連続した名詞, 未知語の区切りとして使われている読点

読点は列挙の区切りとして使われている場合があり，その内容は続いているからである．

以上の条件で区切られた単位を節と定義し，今回の抽出する抄文の単位とする．

表 2　抄録文の平均文字数

	句点数	平均文字数
ｴﾚｸﾄﾛﾆｸｽ	2.90	51.06
物理	3.76	48.42
法律	2.73	58.60
全体	3.12	52.88

3. 単語の重みを利用した抄録縮約処理方式

本方式の実際の流れを説明する．

(1) ブックコンテンツ･データベースから抄録を抽出する．

(2) 抽出した抄録に形態素解析を行い, 品詞情報を取り出す．なお，本研究では形態素解析器として茶筌[2]を用いた．

(3) 抜き出した品詞情報を基に, 先ほどの定義にあわせて，抄録を節に分割する．

(4) 節ごとに算出対象語を抜き出す. 算出対象語とは, 文章特徴語の候補となりうる単語のことを指す．本方式では，品詞が名詞もしくは未知語の単語とする．ただし，非自立語の名詞は含めない．

(5) 抽出した算出対象語の重み計算を行う. 重み計算として，従来の tf 法，tf-idf 法，提案の df 法がある．

(6) 単語の重みの和を節の重みとして, 節の重み計算を行う．

(7) 上記(6)の結果より，抄文抽出を行う．

以上の単語の重み付けによる抄録縮約処理方式を図 3 に記す．

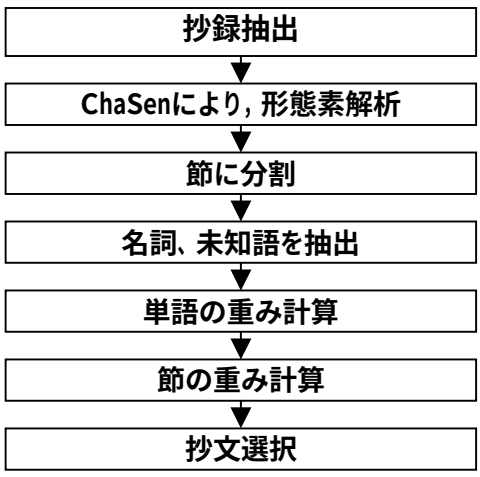

図 3　単語の重み付けによる抄録縮約方式

4. 単語重み抄録縮約方式の評価実験

上記の抄録縮約処理方式を用いて, 実験を行った. 単語の重み付けの際, 本研究で提案する df 法の他に, 比較対象として tf 法, tf-idf 法を選択した. この実験では, 抽出する節数を 2 とした.

また, df 法の場合は df の値が 1, tf 法, tf-idf 法の場合では tf の値が 1 の単語は単語の重み計算時に含めないこととする.

4.1. 実験対象データ

本実験では, 以下の 3 つの図書データを対象とした.

(1) 『エレクトロニクス』
(2) 『物理』&『出版年 1998 年～2001 年』
(3) 『法律』&『国』

ただし, 抽出する節数を 2 と定めたので 3 つ以上の節で構成される図書データのみを選んだ.

表 3 に図書データの詳細を記す. 節数にばらつきがあるのは, ジャンルによる影響だと考えられる. 一般的に, 理系より文系の図書の方が文字数, 節数は多い.

表 3 実験対象図書データ

	図書数	節数（平均）
ｴﾚｸﾄﾛﾆｸｽ	314	5.46
物理	310	6.02
法律	339	6.78

4.2. 正解作成

実験を行う際に, 正解を作成した. 以下に, その方法を記す.

三人に抄文単位に区切られた節から, 図書の特徴を示す節を 2 つ選択してもらう. 続いて, その結果から多数決を行い, 上位 2 つを正解の節とする. 本研究では, 作成した正解である節を正解節と呼ぶことにする.

4.3. 実験結果

正解の節数は 4.2 に記した通り, 1 文書につき 2 節となるので,『文書数×2』節になる. また, 抽出する節数も, 1 文書につき 2 つと定めたので, 同様に『文書数×2』節となる. つまり,

$$\text{精度}=\text{再現率}=\frac{\text{抽出した正解節数}}{\text{文書数}\times 2} \quad \text{(式 1)}$$

となり, 本実験では精度と再現率は同等である.

以上を踏まえて, 表 4 に実験結果を記す.

表 4 抄録縮約処理方式の精度, 再現率

	df 法	tf 法	tf-idf 法
ｴﾚｸﾄﾛﾆｸｽ	**0.629**	0.581	0.580
物理	**0.621**	0.548	0.539
法律	**0.578**	0.529	0.503

4.4. 実験結果の考察

(1) 実験に用いた 3 つの実験データすべてにおいて, df 法がもっとも性能が良く, 次いで tf 法, 一番低かったのが tf-idf 法という結果が得られた. tf 法より tf-idf 法の方が性能は低いということは, 図書の抄録に関しては, idf は正解節を抽出するのに適切な情報ではなく, df を用いる方が良い結果が得られるということである.

(2) 表 5 に節単位の文字数, 単語数の平均を記す. ここで言う単語とは, 算出対象語を指す. 表に示されているように, 正解節は, 全平均に比べて節長が長い. つまり, 正解節を抽出するのに, 節長が重要だと考えられる. 今回の実験結果では, できるだけ節長の影響を少なくするため, 値が 1 の単語を省いた. なぜなら, tf や tf-idf では値の差が小さいため, 値が 1 の単語を含めてしまうと結果が節長に大きく依存してしまうからである.

表 5 全体の節と正解節の対比

	全体		正解節	
	文字数	単語数	文字数	単語数
ｴﾚｸﾄﾛﾆｸｽ	27.2	5.65	36.1	8.52
法律	27.0	6.11	37.1	9.71
物理	26.8	6.16	37.3	9.21

図書の抄録に対する, 単語の重み付けによる方式では, df 法が tf 法, tf-idf 法より性能が高いことが示せた.

df 法の性能が高い理由は, 文書が特徴のある書かれ方をされているからであり, そのような文書集合に対して, 使われる手法であると考えられる. それはこの方式が, tf 法, tf-idf 法のように出現頻度などによる単語の重み付けによって節の抽出を行ってはいるが, 内容語で

はなく, 重みの大きい手がかり表現によって節を抽出する方式だからである.

5. 単語の重みと文の長さを利用した抄録縮約方式

4.4 節の表 5 より，節長が正解節を抽出するのに重要な情報であると考えられる．そこで, 現在の方式に文長を考慮に入れて抄録縮約を行うことにより，性能の改善を図る.

今回，文の長さを単語数で表すことにする.

5.1. 単語重み文長複合型抄録縮約方式

本方式の流れは以下のとおりである.

(1) 抄録を抽出, 形態素解析を行い，品詞情報を抜き出す.
(2) 品詞情報から，節に分割する．節の定義は前実験と同様である.
(3) 算出対象語を抜き出す.
(4) 単語の重み計算を行い, 同時に節中の算出対象語の個数（単語数）を数える.
(5) 単語の重みの和と単語数から, 節の重みを計算する.
(6) その結果から節の抽出を行う.

単語の重みによる抄録縮約方式（3.参照）と異なる点は，節の重みに，単語の重みの和だけではなく, 節中の算出対象語の個数も考慮に入れることである. 異なる情報を組み合わせるためには, 各々の情報をどのように組み合わせて一つの情報とするのか，調べる必要がある.

6. 単語重み文長複合型抄録縮約方式の評価実験

本方式を用いても，前方式と同様に df 法が tf 法，tf-idf 法に勝る結果が得られるのか検討する．さらに，単語の重み付けの値と単語数の比重をどのようにすればいいのかを調べる.

前実験では，比較対象として，tf 法，tf-idf 法を用いた際, 各値が 1 である単語を含めずに処理を行うことにより, 長さの影響をできるだけ受けないように設定した．だが，今回の方式では，3 つの単語の重み付け方式において，長さが考慮されていると考えることができる.

6.1. 実験内容

上記の抄録縮約方式を用いて，実験を行う．本実験の内容は以下のとおりである.

(1) 単語の重み付けの値は df，tf，tf-idf を用いて比較を行う.
(2) 実験データは前実験と同様である.
(3) 抽出する節数は一文書につき 2 個.
(4) 各値が 1 の単語も含める.
(5) 本実験結果も精度，再現率は同等である.

(4)以外は前実験と同様である.

6.2. 節の重み付け

節の重み計算を行う際には, 以下の二つを用いる.

$(1-\alpha)\text{weight} + \alpha\,\text{word}$ (式 2)

$(1-\alpha)\log(\text{weight}) + \alpha\log(\text{word})$ (式 3)

α 文長比重　$0<=\alpha<=1$

ここで weight とは df, tf, tf-idf などの単語の重みの和，word とは節中の算出対象語の個数である．(式 2)の手段は，weight と word の和であり, (式 3)の手段は weight と word の積である. この計算によって算出された値の上位 2 節を抄文として選択する. 節の重み付け計算時の流れを図 4 に記す.

本実験では和と積, 二つの方式で検討を行う. α の値を 0 から 1 の間を 0.0005 刻みで取り, 各数値の比重を用いて，精度, 再現率を算出する.

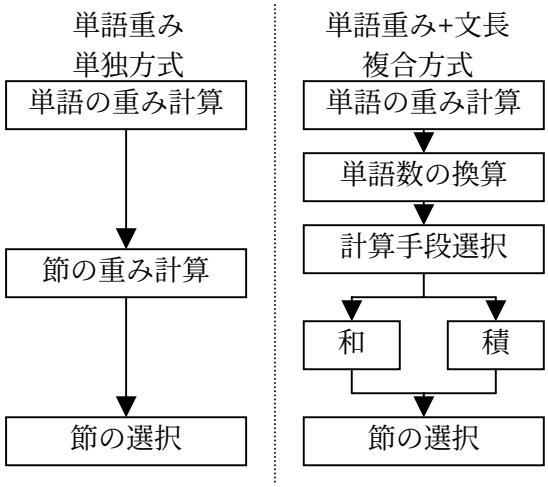

図 4 単語重み文長複合型抄録縮約処理方式

6.3. 実験結果

エレクトロニクスの(式 2), (式 3)による結果を図 5，6，物理の(式 2)，(式 3)による結果を図 7, 8, 法律の(式 2), (式 3)による結果を図 9, 10 に記す．横軸は α の値であり，縦軸は α =1 の時, すなわち算出対象語の個数のみで重み付けを行った時の精度, 再現率を基準とした場合の差を示す.

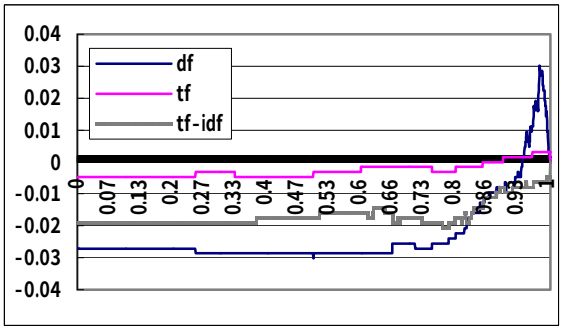

図 5　エレクトロニクス（(式 2)使用）

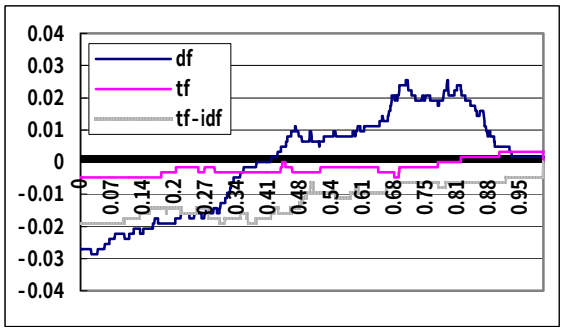

図 6　エレクトロニクス（(式 3)使用）

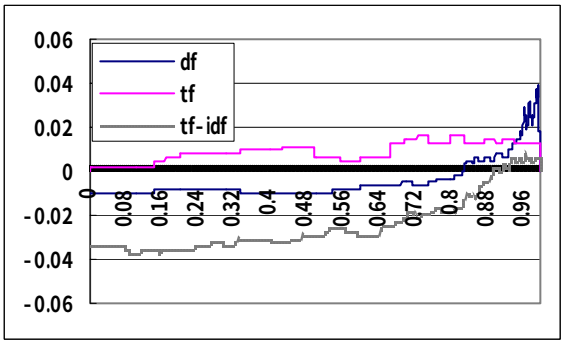

図 7　物理（(式 2）使用）

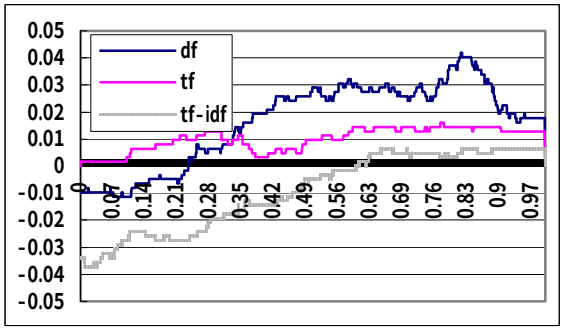

図 8　物理（(式 3）使用）

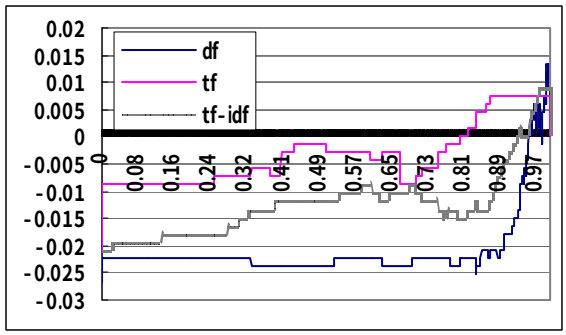

図 9　法律（(式 2）使用）

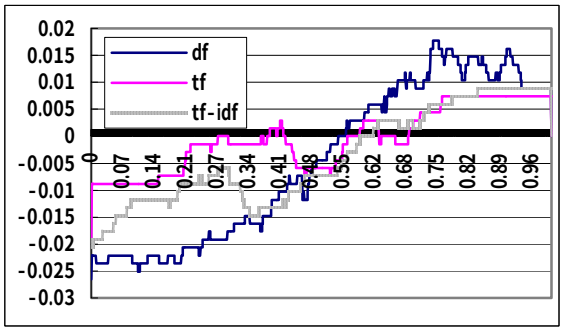

図 10　法律（(式 3）使用）

6.4.　考察

(1) 表 6 の通り，単語重み単独方式(式 1)，単語重み文長複合方式(式 3)の精度，再現率の最高値では 3 つの集合すべてにおいて df 法がもっとも性能が高い．すなわち，文長を考慮に入れた場合でも，df を用いるのが適切である．ただし，α の数値が低い状態では tf 法がもっとも性能が高い．

(2) 和と積，二つの最適化を行ったが，精度，再現率が最高値をとるのはエレクトロニクスでは図 5，図 6 から和，物理と法律では図 7～図 10 から積であることがわかる．このように，どちらが良いのか一概には言えない結果が得られた．また，表 7 のように，和の場合ではある一点で最大値をとるのに対し，積では山にばらつきが見られるため，積の場合，共通の最適な α の値を出すのは難しいと考えられる．

表 6　単語重み単独方式(式 1)，単語重み文長複合方式(式 3)の精度，再現率

	df 法			tf 法			tf-idf 法		
	1:単独	2:複合	3:(2-1)	4:単独	5:複合	6:(5-4)	7:単独	8:複合	9:(8-7)
ｴﾚｸﾄﾛﾆｸｽ	0.629	**0.678**	0.049	0.581	0.656	0.075	0.580	0.653	0.073
物理	0.621	**0.692**	0.072	0.548	0.666	0.118	0.539	0.656	0.116
法律	0.578	**0.631**	0.053	0.529	0.617	0.088	0.503	0.618	0.115

表 7 各単語の重み付け+文長における精度，再現率の最高値をとる文長比重αの範囲

		df 法+文長		tf 法+文長		tf-idf 法+文長	
		最高値	範囲	最高値	範囲	最高値	範囲
ｴﾚｸﾄﾛﾆｸｽ	(式 2)	0.683	0.9775	0.656	0.962~0.9995	0.653	1
	(式 3)	0.678	0.7035~0.7075 0.793~0.794	0.656	0.905~0.9995	0.653	1
物理	(式 2)	0.689	0.995	0.666	0.75 0.8005~0.833	0.658	0.9655~0.9665
	(式 3)	0.692	0.819~0.821	0.666	0.775~0.782	0.656	0.67~0.673 0.7018~0.7035 0.8215~0.9995
法律	(式 2)	0.622	0.9925~0.9945	0.617	0.867~0.9995	0.618	0.974~0.9995
	(式 3)	0.631	0.7435~0.756	0.617	0.7635~0.7855	0.618	0.84~0.9995

表 8　df 法と tf 法，tf-idf 法の差

	df 法と tf 法の差		df 法と tf-idf 法の差	
	単独	複合	単独	複合
ｴﾚｸﾄﾛﾆｸｽ	0.048	0.027	0.049	0.030
物理	0.073	0.026	0.082	0.034
法律	0.049	0.014	0.075	0.013

7. おわりに

7.1. 成果のまとめ

(1) 図書の抄録に関して df 法の精度，再現率は, 表 8 より単語の重み単独の場合 4~8%, 文長複合の場合 1~3%，tf 法，tf-idf 法よﾘ高い.

(2) (1)より図書の抄録においては，df 法は重要節抽出に有効である.

(3) 表 6 から，文長を複合することにより，df に関しては 5~7%程度，精度，再現率が上がることがわかった.

(4) ただし, 表 7 のように節の重みを加算, 乗算方式のいずれで計算するのか, また文長の比重をどのようにするのかは, さらに調査が必要である.

7.2. 今後の課題

抄録縮約方式には，問題点が残っている.

(1) 作成した節のスムージング

本方式では，節を抽出しただけに過ぎない. そのため，文として適切ではない場合もあり, さらに, 抽出された二節は文としてつながりのない可能性もある．また，制限である 50 字以下という条件も満たしていない場合もありえる．適切な文章に直すため, 節のスムージングについて検討が必要である. また, それに伴い, 節単位の抽出が本当に適切なのか判断する必要も出てくると考えられる.

(2) 他の文書集合に対する評価実験

今回は図書の抄録と限定したが, 他に適用できるかどうか, 評価実験を行う必要がある．本方式は手がかり語を情報とするため, 文書の書き方に特徴がなくてはならない. 候補としては, 新聞記事などが考えられる.

(3) 性能の向上

テキスト長を利用して, 性能を上げることは可能であることは示せた. だが, それでも 70%を超えることはなく, さらに性能の向上の検討が必要である.

参考文献

[1]東京大学情報基盤センター
http://www.lib.u-tokyo.ac.jp/

[2] 奈良先端科学技術大学院大学
http://chasen.aist-nara.ac.jp/

[3]小峰 恒，絹川 博之，中川 裕志
“文書頻度を利用した抄文選択方式”
第 64 回情報処理学会全国大会

[4]奥村 学, 難波 英嗣.
"テキスト自動要約に関する研究動向".自然言語処理, テキスト要約のための言語処理」特集号, Vol. 6, No. 6, pp.1-26 1999.7.

[5]Inderjeet Mani.
Automatic Summarization,
John Benjamins Publishing Company, 2001