



卒業研究報告書

平成 27 年度

レポート評価支援システム構築に向けた文章の特徴量の抽出

指導教員 内田 眞司 准教授

氏 名 蘆田 誠人

平成 28 年 2 月 29 日 提出

奈良工業高等専門学校 情報工学科

レポート評価支援システム構築に向けた文章の特微量の抽出 要約

近年、あらゆるところで様々な分野の研究や開発が行われている。研究や開発の結果を書き記したものを論文と呼ぶ。その論文は年々減少傾向にあるが、研究者の数は増加傾向にある。新たに何らかの新しい研究や開発を試みる人達は増加しているということである。その研究者たちが結果を残すと論文を作成することになる。論文の作成工程で、一応書いては見たものの次の工程である推敲が難しい。そこで本研究の目的は、初心者の論文推敲を支援するためのシステム構築に向けた学術論文の特微量の抽出を行うことである。特微量を抽出し、その特微量と作成した論文の特微量を比較し、近い数値が出力されれば、既出の論文とよく似た構造になっていることが言える。しかし、それは表層的な部分に限定される。表層的とは外面的に見たもので、内容など内面的なものは考慮しない。よって、表層的な部分の特微量を抽出することを目的とした。

本研究では、論文の特微量を抽出するためにテキストマイニングというツールを用いた。そのツールを用いて、定義した特微量を抽出した。その結果、1文あたりの文字数と文節数、段落間関係の評価、主語・述語間の距離の特微量は、推敲支援システム構築の際に用いることができる特微量であったが、1段落あたりの文節数の特微量は、構築の際に用いるのには難しい結果となった。

特微量を抽出した結果、構築の際に用いることができる特微量も抽出できたが、物足りない特微量や、場合によってはシステムに適用しないという可能性も視野に入れつつ、再検討すべき課題であると考ええる。また本研究では、論文とレポートを比較したためお互いの特徴が行き違う場面が見られた。論文は1段落が長いために段落数が少ない傾向がある。一方レポートは、1段落が短い代わりに、段落数が多い傾向がある。これにより、1段落あたりの文節数を比較しても、所望の結果を得ることができなかった。この問題を修正することも今後の課題の一つと言える。

目次

第1章	はじめに	1
第2章	理論	3
2.1	論文の読みにくさ・わかりにくさ	3
2.1.1	文章の長さ	3
2.2	段落間関係の評価	4
2.2.1	段落間の関係の定量化	4
2.2.2	段落の構造	6
2.3	主語・述語間の距離	8
2.4	テキストマイニング	9
2.5	提案手法	10
第3章	評価実験	11
3.1	1文あたりの文字数の比較	11
3.2	1文あたりの文節数の比較	12
3.3	1段落あたりの文節数の比較	13
3.4	段落間関係の評価	14
3.5	主語・述語間の距離	16
第4章	関連研究	17
4.1	論文等における文の長さに関する検討	17
4.2	トップダウン・ボトムアップな文章構造作成のための推敲支援システム	17
4.3	法律条文の自動図解化に向けての研究	18
第5章	考察	19
5.1	1文あたりの文字数・文節数	19
5.2	1段落あたりの文節数	19
5.3	段落間関係の評価	19
5.4	主語・述語間の距離	20
5.5	今後の課題	20
第6章	まとめ	21

第 1 章 はじめに

近年,あらゆるところで様々な分野の研究や開発が行われている. その研究の成果として具体的にまとめられた文章のことを論文という. その論文は世間に発信され,誰でも閲覧することができるようになる.

1990 年代の日本の学術論文発表数はおよそ 90000 本であったのに対し,2000 年代半ばの発表数は 120000 本前後と発表数が飛躍的に増えたのだが,そこから徐々に減少傾向にあり,2010 年の発表数は 110000 本と減っている[1]. 減少した理由の一つに国立大学の法人化が挙げられる. 国立大学が法人化されたことにより,教員の定員削減や政府支出研究費の減額などの措置が取られるようになった. 結果的に論文発表する機会が減少し,論文数も減少した. 一方で研究者数に着目してみると,年々増加している.[2] 言い換えれば,新たに何らかの研究を試みる人達が増えてきているということである. その研究成果として論文を作成するが,その作成手順の一部である推敲に着目した.

推敲とは,自分が作成した文章をじっくり読み直し,おかしい所や悪いところを見つけてそれを修正していく作業のことをいう. 論文を推敲するにあたり,着目するポイントは数多く存在する. まず全体に目を通してみて,内容が首尾一貫しているかということや,論文全体の長さが適当なのか,章立ては適当かといったことなどが挙げられる. 段落や文章に絞って目を通すと,段落の長さはもちろん,1 つの段落に 1 つの内容が取り上げられているかということや,主語述語の対応がついているか,修飾語と被修飾語はできるだけ近いところに配置されているかなどが挙げられるが,これだけの量でもほんの一部にしか過ぎないのである. 論文を書き慣れている研究者にとっても,推敲がいかに大変な作業であるのかが理解してもらえるはずだ. しかし,その作業を初心者だからと言って,素通りする訳にもいかず,必ずしなければならない作業の一つである. 論文を書いたことがなく,推敲の仕方がわからない人がもし,おかしい所や悪い所を見つけたとしても,その場所をどのように変えればいいのかまではわからない人が多数だと考える.

そこで,そのような問題を解決するのに,研究を始めてからまだ間もない研究者を対象とする表層的な推敲支援を自動的に行うシステムを作成する. そのシステム構築にあたり,あらかじめシステム内に特徴量と呼ばれる文章の特徴を定量化した数値を格納しておかなければならない. 本研究では,その格納しておく特徴量を世間に公開されている学術論文から抽出することを目的とする. 本研究で扱う特徴量は,文節数や文字数など,1 文に着目するものや,段落間の関

係性など，段落に着目して定量化したものとする．ただし，本研究では表層的な推敲支援システムを構築すると仮定しているので，誤字脱字や品詞の用法など内面的な推敲に関しては一切考えないものとする．

第 1 章では，本研究における背景などを記した．第 2 章では，本研究で必要になる概念やツールの基本的なことを記した．第 3 章では，評価実験という題目で，実験方法とその結果を記した．第 4 章は，本研究と関わりがあり，参考にするものが多かった論文の概要を簡単に記した．第 5 章では結果についての考察，第 6 章ではまとめを記した．

第2章 理論

本章では,研究を行うにあたり,用いられる方法や概念を紹介し,本研究の手法について説明する.

2.1 論文の読みにくさ・わかりにくさ

2.1.1 文章の長さ

学会誌に発表されている論文を読んでいると,わかりにくい表現があったり,疲れが出ることがよくある. そのわかりにくさの原因は,日本語の文法が正しくないことや今読んでいる文と次の文の繋がりが悪く,内容を理解するのが難しいといったことが挙げられる. ただ,挙げた原因のどちらも内面的な問題であり,考慮しないこととする.

表層的な部分だけを見てわかりにくさの原因を探ると,文章の長さが長くなっていたり,主語と述語があまりに離れていて読みづらいといった数々の問題点が挙げられる.

論文内の文章の長さについて清水は,論文の一文の文章の長さが文章のわかりやすさを左右すると提唱している.[3][4] また清水は,論文の文字数と比較する為に社説やショートレター,天声人語の文字数や文節数の解析を試みた. その結果を下記の表 2.1 に示す.

表 2.1 各文の調査数と平均数[4]

種類	調査数	文字数	最大値	文節数
論文	23	57.5	190.8	18.5
ショートレター	23	56.9	163.2	19.1
社説	117	44.5	96.4	17.6
天声人語	58	27.9	67.3	13.6

表 2.1 は清水が解析した結果をまとめたものである. 表 2.1 内の文字数は 1 文あたりの文字数,最大値はその種類のものの最大である 1 文あたりの文字数を表す. 文節数は,1 文を文節ごとに区分けしたものの総数を表す.

文節とは,日本語の言語単位の一つで,文の構成要素で文の実際の言葉として不自然にならない程度に区切ったとき得られる最小のひとまとまりのものである.[5]

まず,文字数に着目すると論文の 1 文あたりの文字数はショートレターの約 1.01 倍,社説の約 1.29 倍,天声人語の約 2.06 倍になっている. 続いて論文の 1

文あたりの文字最大数はショートレターの約 1.17 倍,社説の約 1.98 倍,天声人語の約 2.84 倍になっている。最後に論文の文節数は,ショートレターの 0.97 倍とやや少なめとなったが,社説の約 1.05 倍,天声人語の約 1.36 倍とやや多い結果となった。以上の結果より,論文は他で出版されている文章などと比較して長くなっている傾向がある。文章の長さが長いとその文章はわかりにくいので,論文は比較的わかりにくいものである.[3]

2.2 段落間関係の評価

2.2.1 段落間の関係の定量化

段落とは,長い文章の中の一つの主題をもってまとめた部分のことを指す。板倉らは,段落間の内容に一貫性がない場合,内容を十分に伝えることができなくなり,段落一貫性が低い場合,主題となる内容が読者に伝わりにくくなると指摘する.[7] よって論文をわかりやすく読んでもらうためには,段落間の一貫性を高くする必要があると考える。段落間の一貫性があるというのは,段落ごとの内容をできるだけ似るように作成するということである。山手らは,段落間関係の評価に基づく文章構造支援システム作成にあたり,そのシステム内に段落間関係の評価値を求めるアルゴリズムを設計した。その評価値が高いほど,関係があり,ある基準を超えると推敲を終了するというものである.[8] 本研究では,着目する段落とその前の段落で文節を比較して,すべての文節数と比べてその前の段落と重複している文節数がどれくらいの割合を占めているのかを解析し,その割合が高いほど,重複している文節数が多く,一貫性があるというシステムを想定している。また,石塚は段落の情報量を求める式を定義しており,その式をこのシステムの目的に合うように改良した式を用いて求めた情報量で判断するようにする.[9] 改良した式を下記の式 2.1 に示す。

$$I_i = -\log \frac{A_i}{T_i} \quad (2.1)$$

式 2.1 は 2 段落間の関係を情報量として結果を出す式である。この式の結果として出力される情報量 I_i で段落間の関係进行评估する。右辺の $\log$ の中身は確率で表している。 A_i は $(i-1)$ 段落と i 段落で重複している文節数を表す。 T_i は段落 i の総文節数を表す。重複している文節が全ての文節のどれくらいを占めているかにより確率が求められる。

例えば,下記に示すような段落間関係があったとする。

1. i 段落と $(i-1)$ 段落の間に重複している文節が 5 つあり,段落 i の総文節数が 100
2. i 段落と $(i-1)$ 段落の間に重複している文節が 7 つあり,段落 i の総文節数が 100
3. i 段落と $(i-1)$ 段落の間に重複している文節が 10 つあり,段落 i の総文節数が 150

段落間関係 1 を式(2.1)で定義したものを用いて情報量を求める.

$$\begin{aligned}
 I_i &= -\log \frac{A_i}{T_i} \\
 &= -\log\left(\frac{5}{100}\right) \\
 &= -\log(0.05) \\
 &\approx 1.30
 \end{aligned}$$

情報量は上記で示したように求める. 求めた結果を表 2.2 に示す. 段落間関係 2,3 の計算式は割愛する.

表 2.2 情報量の計算結果

段落間関係	情報量
1	1.30
2	1.15
3	1.18

表 2.2 は段落間関係 1,2,3 の情報量をまとめた表である. 段落間関係 2 の情報量が一番少なく,段落間関係 1 の情報量が一番大きくなっている. 要するに,段落間関係 2 が内容の一貫性が高いと言え,段落間関係 1 が内容の一貫性が低いと言える. このように情報量が大きくなっている段落間には,一貫性が欠けているので推敲する必要がある. 理解しやすい論文を作成するには段落間の情報量をできる限り小さくし,主題から大きくずれないことが大事である.

2.2.2 段落の構造

論文を作成する際,大まかに分けるとトップダウンとボトムアップと混合構造の3つの手法がある. トップダウンとは,論文の大構造から先に決め,そこから詳細な下部構造を作っていくアプローチの仕方である. ボトムアップは論文のテーマに沿ったアイデアをブレインストーミングのような形で次々に生み出し,それらを1アイデア1パラグラフとしてまとめていく手法である. [10]

本研究における混合構造は, トップダウンとボトムアップが同じ文章中に共存する構造である.

トップダウンとボトムアップのイメージ図を図 2.3,図 2.4 に示す.

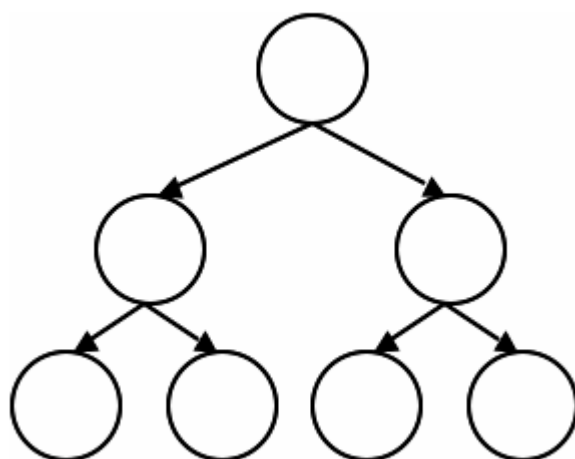


図 2.3 トップダウンのイメージ図

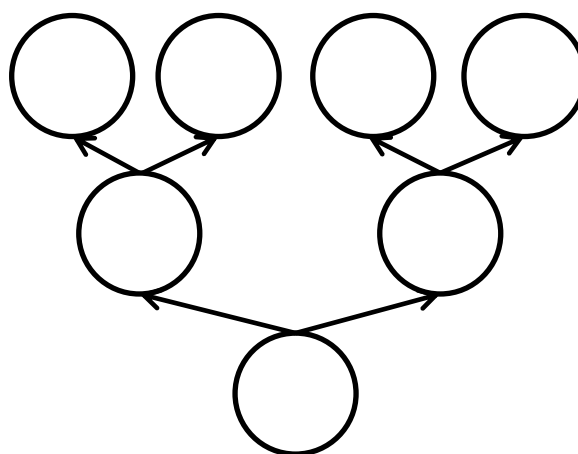


図 2.4 ボトムアップのイメージ図

トップダウンは先に結論を冒頭に書き,そこから話を広げ議論する形になっているのに対し,ボトムアップは細かいことを冒頭で述べ,そこから話をまとめ

て結論に持っていく形である．しかし、論文をボトムアップで作成すると、様々なデメリットが生じる．まず1つはブレインストーミング等で思い浮かべたアイデアをもったいないからと言って、論文のどこかに入れてしまうと、論文の主張を弱め、読者が議論の流れを追うのを邪魔してしまうことである．もう1つは先に細かいことを並べて書いてしまうと、読者側は読むことがついていけなくなる可能性があるからである．主にこの2点が論文を作成する際に障害となってくる恐れがあるので、できる限り避けたほうがよい．その為、トップダウンを用いて作成するのがよいと考える．比較的短い文を扱ったため、混合構造はあまり出現しなかった．本研究で特徴量を抽出する際に用いたのは、論文の1章で比較的短い文であった．長い文を扱うと混合構造が多く出現すると推測される．

トップダウンは結論を始めに書き、そこから話を広げていくので、徐々に情報量が少なくなっていくのが理想である．理想的な推移をしている論文の情報量の推移を表したグラフを図2.5に示す．

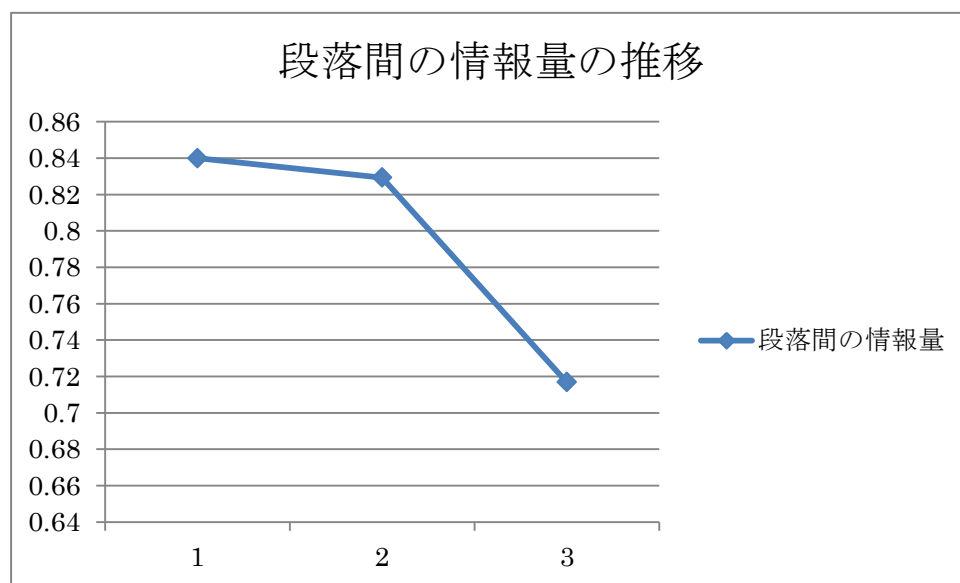


図 2.5 段落間の情報量の推移

図2.5は段落間の情報量の推移の理想の形であった論文の情報量の推移を表したグラフである．第1段落はその前の段落がないため、新語が続出する．よって情報量はかなり大きくなるので、割愛する．第2段落、第3段落ではその前の段落に使われている旧語の数が徐々に増加するため、総語数に比べて新語の割合が減少してくる．よって情報量は減少しだす．論文を作成する際の段落の組み立ては図2.5のように、段落が後になるにつれ情報量が少なくなるようにするのが理想的な形である．

主語と述語の距離が離れている文がどれほど読みにくく, わかりにくいものなのか説明する. 説明に用いる文を下記の図 2.2 に示す.



論文以外の文章を推敲する際のポイントの 1 つとして，できるだけ主語と述語の距離を縮める，といったポイントがある．よって，距離を縮めることにより，主語と述語がはっきりしたわかりやすい文章になる．本研究で想定しているシステムにおいて，主語・述語の距離の特徴量は小さい方がよいと捉えることとする．

2.4 テキストマイニング

テキストマイニングとは、定型化されていない文章の集まりを自然言語解析の手法を用いて単語やフレーズに分割し、それらの出現頻度や相関関係を分析して有用な情報を抽出する手法やシステムのことを指す.[6] 本研究ではテキストマイニングを統計解析向けのプログラミング言語である「R」を用いて行う。Rでテキストマイニングを用いて行う際に、パッケージをインストールしなければならない。そこで本研究で用いたパッケージである RMeCab を紹介する。

RMeCab とは指定したファイル形態素解析エンジン MeCab に渡して形態素解析を実施し、MeCab から結果を受け取る関数が用意されているパッケージである。このパッケージを用いて、文字数や文節数の特徴量を計測する。

本研究では、テキストマイニングの 2 つの長所を活用した。まず 1 つは、テキストデータから定量的に解析をするものである。この長所を用いると、文の長さなど推敲する基準が明確となる。2 つ目の長所は、アルゴリズムさえ与えれば大規模なデータも自由自在に扱える。本研究では、50 本の論文を解析し、特徴量を抽出するので、テキストマイニングを用いることは作業の効率化に繋がる。

2.5 提案手法

本研究では,論文の特徴量の抽出する手法を提案する.

特徴量を抽出する論文の対象は,情報・通信分野の学術論文をランダムで選んだ 100 本である. より現在の論文のスタイルの特徴量を抽出したいため,発表年が 2004 年から 2015 年までのものに限定する.

本研究で抽出する特徴量の一覧を下記に示す.

- 1 文あたりの文節数
- 1 文あたりの文字数
- 1 段落あたりの文節数
- 主語・述語間の距離
- 段落間の情報量

1 文あたりの文節数と 1 文あたりの文字数,更に 1 段落あたりの文節数と 1 段落あたりの文字数は,R でテキストマイニングを行い抽出する. 1 文あたりの文字数は R に解析したいファイルを入力し,適当なプログラムで処理することにより,文字数が出力される. 1 段落あたりの文字数も同様に解析したいファイルを入力し,プログラムで処理すると文字数が出力される. 1 文あたりの文字数を出力させることと違うところは,段落の変わり目で文字数カウントを変更させなければならないところである. その部分を変更させればプログラムを流用させることができる. 1 文あたりの文節数を抽出する場合,R に既存で存在する関数では抽出することができない. そこで R のパッケージの 1 つである RMeCab をインストールし,このパッケージに実装済みの関数 RMeCabText 関数を用いて文節数を抽出する. 1 段落あたりの文節数の抽出も既出の手法で行うが,1 文あたりであるところを 1 段落あたりの文節数に変換させなければならないことに注意しなければならない.

主語・述語間の距離の抽出は,動的プログラミング言語である「Python」のソースコードに形態素解析エンジン「CaboCha」をインポートさせて行う. プログラムの処理中に主語と述語を自動的に決定する機能を実装し,主語と述語の距離を計測する.

段落間の情報量の抽出には,R のパッケージ「RMeCab」でテキストマイニングを行う. 段落間の情報量を抽出する際,あらかじめ抽出した 1 段落あたりの文節数を基に計算する. また 2 つの段落で重複している文節を抽出し,その文節数をカウントし,総文節数に占める割合を求める. その割合を式(2.1)の $\log$ の中身に代入して情報量を求める.

第 3 章 評価実験

2 章に示した提案手法に基づき，実験を行った．

特徴量を抽出する論文の対象は，情報・通信分野の学術論文をランダムで選んだ 100 本である．より現在の論文のスタイルの特徴量を抽出したいため，発表年が 2004 年から 2015 年までのものに限定する．

レポート評価支援システム構築のための特徴量抽出であるので，その抽出したものと奈良高専の 2 年生が情報工学実験 I の際に作成し，高評価を受けたレポートと低評価を受けたレポートを比較した．

3.1 1 文あたりの文字数の比較

上記に示した条件に当てはまる学術論文 100 本をテキストマイニングで解析した結果と，学生が作成したレポートの高評価なものと低評価のものをまとめた表 3.1 に示す．

表 3.1 1 文あたりの文字数の比較

	論文	学生レポート(高評価)	学生レポート(低評価)
平均	63.72	84	82.375
最大文字数	208	113	146
最小文字数	8	48	43
標準偏差	30.90	22.88	33.54

100 本の論文を解析したところ，世間に公開されている学術論文の 1 文あたりの文字数は約 64 文字であることがわかった．続いて高評価の学生レポートを解析すると約 84 文字，低評価の学生レポートに関しては，約 82 文字という結果になった．論文と学生レポートを比較すると，最大文字数においては論文が多い結果となり，また最小文字数は論文が少ない結果になった．学生レポートの標準偏差に関しては，高評価のレポートのほうが低い結果となった．低評価のレポートと論文を比較すると，論文の値が小さくなっている．

以上より，文字数が多いから読みにくいとは一概には言えないが，1 文あたりの文字数は減らしたほうがよいと言える．

3.2 1文あたりの文節数の比較

上記に示した条件に当てはまる学術論文 50 本をテキストマイニングで解析した結果と、学生が作成したレポートの高評価なものと低評価のものをまとめた表 3.2 に示す。

表 3.2 1 文あたりの文節数の比較

	論文	学生レポート(高評価)	学生レポート(低評価)
平均	35.68	47.63	46.13
最大文節数	102	65	74
最小文節数	7	29	22
標準偏差	16.69	13.46	16.89

50 本の論文を解析したところ、世間に公開されている学術論文の 1 文あたりの文節数は約 36 文節であることがわかった。続いて高評価の学生レポートを解析すると約 48 文節、低評価の学生レポートに関しては、約 46 文節という結果になった。論文と学生レポートを比較すると、最大文節数においては論文が多い結果となり、また最小文節数は論文が少ない結果になった。学生レポートの標準偏差に関しては、高評価のレポートのほうが低い結果となった。低評価のレポートと論文を比較すると、ほぼ等しい結果が出た。

以上より、文字数同様に文節数が多いから読みにくいということは一概には言えないが、1 文あたりの文節数の偏りは減らしたほうがよいと言える。

3.3 1 段落あたりの文節数の比較

上記に示した条件に当てはまる論文 50 本をテキストマイニングで解析した結果と、学生が作成したレポートの高評価なものと低評価のものをまとめた表 3.3 に示す。

表 3.3 1 段落ごとの文節数の比較

	論文	学生レポート(高評価)	学生レポート(低評価)
平均	131.49	85.5	49.14
最大文節数	564	129	71
最小文節数	21	40	28
標準偏差	52.39	17.19	17.2

段落ごとの平均文節数を計測したところ、学生レポートは論文よりも少なくなっている。1 文あたりの文節数では論文のほうが上回っていた。この結果より、学生レポートは論文と比較して、1 文あたりの文節数は上回るが、1 段落あたりの文節数が少なくなるのは段落を分けることが多いためだと考えられる。

次に標準偏差に着目すると、学生レポートが小さくなっていることがわかる。段落間の文節数の誤差が比較的少ないということが言える。

3.4 段落間関係の評価

段落間関係の評価を行った結果を、構造ごとに説明する。

評価結果を基に、学生レポートの高評価であるレポートと、低評価のレポートを比較する。高評価を受けた学生が作成したレポート2つの評価結果を表3.8と表3.9に示す。

表 3.4 高評価の学生レポート(1)

段落番号	総文節数	重複語数	情報量
4_1	40	0	－
4_2	129	29	0.648192
4_3	124	38	0.513638
4_4	49	16	0.486076

表 3.5 高評価の学生レポート(2)

段落番号	総文節数	重複語数	情報量
5_1	38	0	－
5_2	125	4	1.49485
5_3	92	23	0.60206
5_4	54	7	0.887296
5_5	37	9	0.613959

表3.4の学生レポートは、後の段落に進むにつれ、情報量が減っていることがわかる。よってトップダウン構造がなされている。表3.5の学生レポートは、3段落目の情報量が少なくなっているものの、それを除くとトップダウン構造になっていることがわかる。

続いて、学生レポートの低評価を受けたレポート2つを表3.6と表3.7に示す。

表 3.6 低評価の学生レポート(1)

段落番号	総文節数	重複語数	情報量
6_1	57	0	－
6_2	28	4	0.845098
6_3	71	5	1.152288
6_4	42	3	1.146128
6_5	36	3	1.079181
6_6	51	7	0.862472
6_7	59	11	0.729459

表 3.7 低評価の学生レポート(2)

段落番号	総文節数	重複語数	情報量
7_1	64	0	－
7_2	73	18	0.60805
7_3	94	15	0.797037
7_4	34	4	0.929419
7_5	66	2	1.518514

表 3.6 の学生レポートは、情報量にばらつきがあり、トップダウンとボトムアップが混合したレポートになっている。表 3.7 のレポートは、ボトムアップ構造になっている。この 2 つのレポートは構造が混合してしまったり、違ったものにしていたが為に、読み手である先生の誤解を招いた可能性がある。

3.5 主語・述語間の距離

100本の論文を1文ずつに分けて、主語と述語の距離を測定した。その結果を表3.8に示す。

表 3.8 1文ごとの主語・述語間の距離の比較

	論文	学生レポート(高評価)	学生レポート(低評価)
平均	4.85	5.28	7.71
最長距離	34	18	21
最短距離	0	0	0
標準偏差	6.14	6.28	7.86
中央値	3	1	7

(単位は文節)

表3.8は1文ごとの主語・述語間の距離を比較した表である。まず、平均を比較すると論文の特徴量が小さい値を示しており、論文を目指すべき指標にすることが可能である。最長距離は、論文が大きい値になったのは、1189文と34文を比較した文章の数が大きく異なることが原因として挙げられる。最短距離は全てにおいて0という結果になった。これは距離が0という意味を示すものなので、主語と述語が繋がっていると解釈できる。標準偏差は、学生レポートの低評価のものが比較的大きな値を示し、ばらつきが大きくなっていると言える。

第 4 章 関連研究

本研究の内容と関連のある先行研究が 3 つあるので紹介する。

4.1 論文等における文の長さに関する検討[4]

本論文は,東京工業大学大学院社会理工学部研究科の名誉教授である清水康敬が 1997 年の日本教育工学会で発表したものである。学会誌に掲載されている論文はわかりにくいとよく言われる。また,自分の専門分野から離れている内容を読む場合はわかりにくいことがあるかもしれない。筆者が論文を読み,論文はわかりにくいと感じた理由の一つに文の長さと指摘した。では,実際論文というものは他の文章を扱っているものと比較し,論文における 1 文あたりの長さが長いことを研究で示した。また,論文が長文になる理由を考察として示した。論文と新聞記事などと比較すると,論文が圧倒的に長いことがわかり,「～である」や「～であり,」などの表現の仕方によって,文を長くしてしまっている傾向がある。また,論文を読みやすくするには,文を極力短くし,句読点を適宜使って文を切るなどの対策をとるとよいと考察した。本研究では,論文のわかりにくさの理由の 1 つである文の長さについて参考にした。論文は極力短く作成するのがよいと述べられているので,推敲支援システムを実装する際に,推敲ポイントとして取り上げることが検討したい。

4.2 トップダウン・ボトムアップな文章構造作成のための推敲支援システム[8]

本論文は,広島市立大学大学院情報科学研究科の山手と砂山が 2013 年に人口知能学会に発表されたものである。文章を書く機会は様々なところである。本人がわかりやすく書いたつもりでも,他人が読んでわかりにくい場合,筆者と違う解釈をしてしまう場合がある。自分の意図を読者に正しく伝えるためには自分で書いた文章構造を推敲する必要がある。しかし,文章構造の推敲を自分でするには限界がある。自分で簡単に文章構造の推敲が可能な環境があることが望まれる。そこで文章を段落間のつながりに着目した段落間関係の評価に基づく文章構造推敲支援システムの構築を目的として研究を行った。段落間関係の評価を条件付き確率で表現し,一定の確率を超えると推敲を終了するというシステムを作成した。評価実験で 8 名の人に 2000 字程度の作文をしてもらい,そのシステムを使用して推敲してもらったところ,段落間のつながりの強さを変える修正をアシストする機能が役立ったと感じた被験者が多かった。このアシスト機能があることによって文章構造の推敲を促すことができる。

4.3 法律条文の自動図解化に向けての研究[9]

本論文は,亜細亜大学経営学部の石塚によって 2006 年に電子情報通信学会
日常的に法律を気にする人は少ないが,紛争や災難に巻き込まれた場合,現行法
規に従い,問題解決を図ることが要求される. しかし細かいことでいちいち法
律専門家に頼っていてはきりがなく,私達は必要な時に内容を理解できなけれ
ばならないことが要求されている. 法律は論理的に書かれていて一般人が理解
するのには少し時間がかかる. また,1 文が相当な長さの条文も見られ,他の条
項がしばしば参照されている. そこで重要な事項や探したい事項の位置づけを
明確にするためのマクロ的な図解化の問題について,各条文の総単語数と新語
数をカウントすることにより情報量を定義し,情報量の分布を用い木構造のチ
ャートが得られることを提案する. 本研究では,その情報量の定義している式
を参考にして段落間の情報量を定義した.

第 5 章 考察

本実験の結果に基づき，レポート評価支援システムにどのように活用できるかを考察した．

5.1 1 文あたりの文字数・文節数

3.1 節，3.2 節で世間に公開されている学術論文を解析し，1 文あたりの文字数，文節数を計測した．結果は，文字数と文節数共に，学生レポートのほうが多くなった．また標準偏差に関しては，低評価なものが高い数値を出した．よって，文字数および文節数は少なければ少ないほど読みやすいレポートである．また，各文における文字数と文節数の偏りが少ないほうが読みやすい論文と推定される．

上記の結論により，レポート評価支援システム構築の際に用いることが出来る特徴量であると推定される．

5.2 1 段落あたりの文節数

3.3 節で 1 段落あたりの文節数を計測した．結果は論文のほうが多かった．高評価を受けた学生は比較的論文寄りではあるものの，近いとは言にくい．論文とレポートを比較したときの大きな欠点である，段落の長さや段落数が明らかに違うという点が顕著に現れた．1 段落が長い論文では文も長いので文節数も多くなってしまう．一方，段落数が多いレポートは，1 段落の文節数が少ないのでこのような結果が出てしまった．この特徴量を活用するには，もう少し改善させなければならないと考える．

5.3 段落間関係の評価

3.4 節で段落間関係の評価を定量的に計測した．その結果より 3 種類の文章構造があることがわかった．トップダウン構造，ボトムアップ構造，トップダウンとボトムアップの混合の構造の 3 種類だが，高評価の学生レポートは全てトップダウン構造で作成されていた．よって，読み手に誤解を招くことなく読んでもらえる論文の構造はトップダウンであると考えられる．この特徴量は，レポート評価支援システム構築の際に用いることが出来る特徴量の 1 つと推定される．

5.4 主語・述語間の距離

3.5 節で、主語・述語間の距離を計測した。結果は論文が一番小さい値となった。学生レポートは、高評価、低評価共に論文より大きな値をとり、主語と述語が比較的離れている文章が多い結果となった。主語・述語間の距離が短い方が、読者に誤解されることが少ない文章になるので、レポート評価支援システムを構築する際に用いることが出来る特徴量であると考ええる。

5.5 今後の課題

まず 1 つ目は、5.2 節で取り上げた 1 段落あたりの文節数の特徴量抽出の改善である。結果を見てわかる通り、低評価のレポートが論文と同じ特徴量を示したりしていたことがあった。予想の段階だが、1 段落あたりの文節数で論文とレポートの因果関係がない可能性もあるので、この特徴量を適用しない可能性も視野に入れつつ、改善しなければならない。

続いて 2 つ目は、段落間関係の評価の拡張である。表 3.7 から見ても分かる通り、世間に公開されている論文はトップダウンとボトムアップの混合構造で作成されている論文が大半である。しかし、本研究ではトップダウンでの評価しか対象にしていないので、他の 2 種類も対応できるように拡張することである。

第 6 章 まとめ

本研究では、レポート評価支援システム構築における特徴量の抽出を試みた。特徴量とは、1 文あたりの文字数や、文節数、段落間の評価、主語・述語間の距離を定量的に表した数値のことを指す。その特徴量を抽出し、レポート評価支援システムにどのように活用できるかを検討した。

まず、世間に公開されている学術論文と学生が作成したレポートの 1 文あたりの文字数と文節数を抽出し比較した。その結果、高評価を受けた学生のレポートが論文に近い数値を出力し、低評価の学生レポートは比較的離れた結果が出た。この結果より、レポート評価支援システム構築の際に用いることが出来る特徴量の 1 つと推定される。

次に、収集した論文と学生のレポートから 1 段落あたりの文節数を抽出した。抽出した値を比較すると、論文とレポートにおける因果関係があまり無いように思わせる結果が出た。この結果のままでは、レポート評価支援システム構築の際に用いることが出来ないなので、再検討が必要である。

3 つ目に収集した論文と学生のレポートから段落間関係の評価を行った。一般的に論文を作成するときはトップダウンが良いと提唱されている中で、高評価の学生レポートは全てトップダウン構造で作成されていることがわかった。よって、この特徴量はレポート評価支援システム構築の際に用いることができる 1 つの特徴量であることが推定される。

4 つ目は主語・述語間の距離を抽出した。論文と学生レポートを比較し多結果、論文が 1 文の主語・述語間の距離が短いことがわかった。主語・述語間の距離が短い文章は読者の誤解を招きにくいことから、良い論文を目指す良い指標になり、システム構築の際に用いることが出来る特徴量であると推定される。

今後の課題として、上記にも挙げた通り、1 段落あたりの文節数の再検討と段落間関係の評価の構造の種類を増やすことが挙げられる。

参考文献

[1]ある医療系大学長のつばやき-あまりにも異常な日本の論文数のカーブ
[online]

<http://blog.goo.ne.jp/toyodang/e/26f372a069cbd77537e4086b0e56d347>

[2]文部科学省：科学技術要覧 平成 23 年度版 海外及び日本の科学技術活動の概要 ～ ii 研究人材～,参照(2015-11-29) [online]

http://www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2011/07/12/1307510_2.pdf

[3]清水康敬「わかりにくい論文の特徴と改善の視点」,日本音響学会,1993

http://ci.nii.ac.jp/els/110003110637.pdf?id=ART0003516225&type=pdf&lang=jp&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1449729306&cp=

[4]清水康敬「論文等における文の長さに関する検討」,日本教育工学会,1997

http://ci.nii.ac.jp/els/110003026311.pdf?id=ART0003486810&type=pdf&lang=jp&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1449732617&cp=

[5]文節(ぶんせつ)とは - コトバンク [online]

<http://kotobank.jp/word/%E6%96%87%E7%AF%80-128566>

[6]IT 用語辞典 e-Words テキストマイニングとは [online]

<http://e-words.jp/w/テキストマイニング.html>

[7]板倉由知,白井治彦,黒岩丈介,小高知宏,小倉久和「様々な文書を対象とした段落一貫性の解析」,情報処理学会,2009

https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=62647&item_no=1&page_id=13&block_id=8

[8]山手砂都美,砂山渡「トップダウン・ボトムアップな文章構造作成のための推敲支援システム」,人工知能学会,2013

<https://kaigi.org/jsai/webprogram/2013/paper-174.html>

[9]石塚隆男「法律条文の自動図解化に向けての研究」,電子情報通信学会,2006

<http://ci.nii.ac.jp/naid/110004662887>

[10] 小論文・レポートの書き方 - トップダウン・ボトムアップ - [online]

<http://shouronbun.com/topdown.html>

付録

- 1：砂山渡他：テキストデータマイニングのための統合環境(TETDM)，電子情報通信学会，言語理解とコミュニケーション 111(119), 15-20, 2011
- 2：竹岡志朗他：テキストマイニングを用いたイノベーションの普及分析(<特集>情報経営への言語的アプローチ)，日本情報経営学会，日本情報経営学会誌 35(1), 72-86, 2014
- 3：喜田昌樹：データマイニングの視点から見たテキストマイニングの三つの利用法(<特集>情報経営への言語的アプローチ)，日本情報経営学会，日本情報経営学会誌 35(1), 4-18, 2014
- 4：豊嶋章宏他：語の連鎖構造とテキストマイニングに基づく概念ベースの構築，情報処理学会，第76回全国大会講演論文集 2014(1), 165-166, 2014
- 5：山中裕樹他：テキストマイニング技術を応用したメソッドクローン検出手法の提案，情報処理学会，研究報告ソフトウェア工学(SE) 2013-SE-182(28) 1-8, 2013
- 6：泉朋子他：複数の言語的特徴を用いた日本語述部の同義判定，言語処理学会，自然言語処理 = Journal of natural language processing 20(4) 539-561, 2013
- 7：松原裕之：テキストマイニングを用いたエンジニアリングデザイン教育の振り返りの分析，情報処理学会，情報教育シンポジウム 2014 論文集，2014
- 8：美馬秀樹他：近代文献のデジタルアーカイブ化とテキストマイニング，情報処理学会，研究報告人文科学とコンピュータ(CH) 2012-CH-95(4) 1-8, 2012
- 9：大崎隆比古他：古典史料における人名・地名を用いたテキストマイニング，情報基礎とアクセス技術(IFAT) 2010-IFAT-100(2) 1-8, 2010
- 10：井坪将他：古典史料における人名・地名情報を用いた共起関係の可視化，情報処理学会，じんもんこん 2010 論文集 2010(15) 203-210, 2010
- 11：上拾石博紀他：カスタマーレビューを用いた商品利用者の意見分析，人工知能学会，人工知能学会全国大会論文集 27, 4p, 2013
- 12：田中良典他：テキストマイニングによる若者の観光需要の推定～大学生のライフスタイルに着目して～，研究報告人文科学とコンピュータ(CH) 2011-CH-90(10), 1-8, 2011
- 13：石田淳他：テキストマイニングシステム STM の感情・意思分析機能の組み込みを中心とした機能拡充，情報処理学会，研究報告自然言語処理(NL) 2010-NL-196(5) 1-8, 2010
- 14：中島正人他：テキストマイニングによる成功経験の抽出，情報処理学会，全国大会講演論文集 第72回(人工知能と認知科学), 41-42, 2010

- 15：大崎隆比古他：古記録の地名情報によるテキストマイニング手法の提案，情報処理学会，じんもんこん 2009 論文集 2009(16) 217-224, 2009
- 16：末永高志他：テキストマイニングのためのドメイン別単語辞書の構築方法，情報処理学会，研究報告バイオ情報学（BIO） 2009-BIO-19(23) 1-6, 2009
- 17：和泉潔他：テキストマイニングによる金融市場の月次動向分析，人工知能学会，知識ベースシステム研究会 85 69-74, 2009
- 18：佐野真：感性品質評価語辞書を利用したテキストマイニング，情報処理学会論文誌データベース（TOD） 45(SIG07(TOD22)) 43-53, 2004
- 19：寺田昭他：文脈情報による同義語辞書作成支援ツール，情報処理学会，情報処理学会研究報告自然言語処理（NL） 2006(124(2006-NL-176)) 87-94, 2006
- 20：溝口達也他：Web コミュニティの信頼度判定に関する研究，情報システム学会，第 5 回全国大会・研究発表大会，2009
- 21：根本啓一他：テキストデータの構造化を支援する対話的マイニングシステム，人工知能学会，人工知能学会全国大会論文集 28 1-4, 2014
- 22：杉浦隆博他：数値による新聞記事テキストマイニングシステムの提案(E 分野:自然言語・音声・音楽), FIT(電子情報通信学会・情報処理学会)運営委員会，情報科学技術フォーラム一般講演論文集 6(2), 161-164, 2007
- 23：村田真樹他：大規模記事群からの数値固有表現情報のテキストマイニング可視化システム，情報処理学会，情報処理学会研究報告自然言語処理（NL） 2008(33(2008-NL-184)) 25-32, 2008
- 24：柿村幸貴他：対話的テキストマイニングのためのタグ付け用検索基盤，情報処理学会，情報処理学会研究報告グループウェアとネットワークサービス（GN） 2008(7(2008-GN-066)) 39-46, 2008
- 25：丸山広他：学生実験支援システムの構築と運用，情報処理学会，情報処理学会研究報告コンピュータと教育（CE） 2004(100(2004-CE-076)) 25-32, 2004
- 26：緒方進他：Web 上のテキスト情報を用いた人物評価手法，情報処理学会，情報処理学会研究報告自然言語処理（NL） 2005(1(2004-NL-165)), 9-14. 2005
- 27：村上明子他：複数の筆者の表記の違いを利用した同義表現抽出，情報処理学会，情報処理学会研究報告自然言語処理（NL） 2004(73(2004-NL-162)), 117-124, 2004
- 28：宅間大介他：テキストデータを用いた問題の早期発見手法，情報処理学会，情報処理学会研究報告自然言語処理（NL） 2004(73(2004-NL-162)), 19-26, 2004
- 29：上野友司他：係り受けの 2 部グラフと共起関係を利用した同義表現抽出，情報処理学会，情報処理学会研究報告自然言語処理（NL） 2004(1(2003-NL-159)), 169-176, 2004

- 30：須藤一弘他：報道内容の印象分析のためのテキストマイニング手法，情報処理学会，全国大会講演論文集 第 72 回(人工知能と認知科学), 357-358, 2010
- 31：竹内広宣他：ライフサイエンス分野向けテキストマイニング用辞書における不適切エントリーの抽出，情報処理学会，情報処理学会研究報告情報学基礎 (FI) 2005(108(2005-FI-081)), 37-44, 2005
- 32：柏崎礼生他：広域分散仮想化基盤のストレージ評価の最新動向，情報処理学会，インターネットと運用技術シンポジウム 2013 論文集 39-42, 2013
- 33：服部峻他：語の同位関連と性質の継承関連を用いた概念階層の Web からの抽出，電子情報通信学会，電子情報通信学会技術研究報告．DE，データ工学 107(131), 127-132, 2007
- 34：松原靖子他：情報拡散過程のダイナミクス：非線形モデルの提案と情報予測，情報通信学会，情報処理学会論文誌データベース (TOD) 6-5 11-22, 2013
- 35：小林雄太他：共起ネットワークを用いたクラスタ性によるテキスト分類，情報通信学会，第 74 回全国大会講演論文集 2012(1), 537-538, 2012
- 36：鈴木晋：文章構造を付加した日本語テキストに対する情報検索方法の検討，情報処理学会，第 75 回全国大会講演論文集 2013(1), 531-532, 2013
- 37：岡本隆明：古文書・典籍を対象とした文字管理システムとその可能性，情報処理学会，情報処理学会研究報告人文科学とコンピュータ (CH) 2008(47(2008-CH-078)), 77-84, 2008
- 38：松本圭祐他：テキストの内容理解促進を補助するための動的なマルチメディアコンテンツ生成に関する検討，情報処理学会，情報処理学会研究報告エンタテインメントコンピューティング (EC) 2006(24(2006-EC-003)), 121-128, 2006
- 39：江上修平他：文章構造と意味ネットワークを用いた句抽出による自動要約，電気関係学会，電気関係学会九州支部連合大会講演論文集 2004(0), 626-626, 2004
- 40：遠藤雅樹他：地域サイト及びブログの観光情報融合のためのキーワード自動抽出手法の検討，情報処理学会，研究報告データベースシステム (DBS) 2013-DBS-158(2), 1-8, 2013
- 41：松林優一郎他：日本語述語項構造アノテーションに関わる諸問題の分析，情報処理学会，研究報告自然言語処理 (NL) 2013-NL-214(12), 1-18, 2013
- 42：大野一樹他：係り受け関係の階層化とその共起に基づいた構文木モデルを利用した構文解析手法の提案，情報処理学会，研究報告自然言語処理 (NL) 2013-NL-214(6), 1-6, 2013

- 43: 斉藤いつみ他: 正規・崩れ表記のアライメントに基づく表記崩れパタンの抽出, 情報処理学会, 研究報告自然言語処理 (NL) 2013-NL-214(5), 1-9, 2013
- 44: 但馬康宏: 多値ラベルによるニュース記事のコメントに対する感情推定, 情報処理学会, 研究報告数理モデル化と問題解決 (MPS) 2013-MPS-95(18), 1-6, 2013
- 45: 内海康他: ベイズ階層言語モデルと Semi-Markov SHDCRF の協調学習による教師なし形態素解析, 情報処理学会, 研究報告自然言語処理 (NL) 2013-NL-213(10), 1-9, 2013
- 46: 寺内美奈他: NVSG 形態表記のための日本手話語彙分類法, 情報処理学会, 研究報告自然言語処理 (NL) 2013-NL-213(9), 1-4, 2013
- 47: 金久保正明: 形態素解析手法と通俗的単語群に基づく類音文変換システム, 情報処理学会, 情報処理学会論文誌 54(7), 1937-1950, 2013
- 48: 鍛冶伸裕他: 未知語を考慮した形態素解析のための単語ラティスの効率的な生成方法, 情報処理学会, 研究報告自然言語処理 (NL) 2013-NL-211(12), 1-8, 2013
- 49: 岡輝晃他: 未整備の歴史的文献への濁点の自動付与アプリケーション, 情報処理学会, じんもんこん 2012 論文集 7 191-198, 2012
- 50: 松本裕治他: 形態素・係り受け解析済みコーパス管理・検索ツール「茶器」, 情報処理学会, 研究報告自然言語処理 (NL) 2010-NL-199(18), 1-6, 2010
- 51: 溝渕正剛他: 着目キーワードとつながりのある埋もれた語の抽出手法の検討, 情報処理学会, 研究報告知能システム (ICS) 2010-ICS-161(3), 1-6, 2010
- 52: 白川真澄他: 格フレームを考慮した Web 検索スニペット解析による動作関係抽出, 情報処理学会, 研究報告データベースシステム (DBS) 2010-DBS-151(38), 1-7, 2010
- 53: 斎藤祐一郎他: 自然言語処理を用いた企業相関関係の取得, 情報処理学会, 研究報告インターネットと運用技術 (IOT) 2010-IOT-11(4), 1-5, 2010
- 54: 鈴木敬史他: 代表・派生関係を利用した日本語機能表現の解析, 情報処理学会, 研究報告自然言語処理 (NL) 2010-NL-199(6), 1-9, 2010
- 55: 張嵐他: 受け入れテスト駆動開発による実行可能な仕様に関する検証結果, FIT(電子情報通信学会・情報処理学会)運営委員会, 情報科学技術フォーラム講演論文集 14(1), 229-230, 2015
- 56: 稲船哲也他: Java プログラムの逆方向解析, FIT(電子情報通信学会・情報処理学会)運営委員会, 情報科学技術フォーラム講演論文集 14(1), 227-228, 2015

- 57：大橋輝他：組み合わせテスト集合生成におけるテストケース候補数の最適化について，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 225-226, 2015
- 58：長田知之他：故障木図からのモデル検査式導出手法の提案，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 223-224, 2015
- 59：鴨生悠冬他：ファイル読み込み処理における AnT オペレーティングシステムの OS サーバの優先度逆転抑制法の評価，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 221-222, 2015
- 60：澤田淳他：通信制御サーバを用いた AnT の OS サーバ入替え機能の評価，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 219-220, 2015
- 61：森山英明他：CPU 資源の割当て制御による KVM の動作制御評価，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 217-218, 2015
- 62：大山恵弘：ハイパバイザによる災害警報通知，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 213-216, 2015
- 63：杉本洋輝他：仮想化環境におけるゲスト OS ページキャッシュの監視によるホスト OS ページキャッシュのヒット率の向上，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 211-212, 2015
- 64：寺岡明彦他：システムコール発行頻度に基づくバッファキャッシュ制御法における重要度更新の遅延方式，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 209-210, 2015
- 65：井川尚幸他：無誤差演算のための最小演算精度の提案と評価，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 205-208, 2015
- 66：大石真生他：GPU による回遊中心性及び利便中心性の並列計算，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 201-202, 2015
- 67：荒川隆他：結合ソフトウェア ppOpen-MATH/MP の性能評価と最適化，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 199-200, 2015
- 68：福永隆文：スイッチを用いないトラス接続 PC クラスタの実装と評価，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 197-198, 2015
- 69：平松和剛他：ストレージイベント駆動型ジョブ制御システム：TaNAS，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 195-196, 2015
- 70：齊藤隆之他：HPC システムにおける障害監視および動的再構成機能，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 193-194, 2015
- 71：中村正治：MDA：その数学的定式化，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1), 189-192, 2015

- 72：小倉直徒他：Java メソッドの純粋性の変化と不具合の混入および除去に関する調査，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1)，187-188, 2015
- 73：野口孝文：プログラミング教育におけるオブジェクトの構成比率による課題プログラムの分類と特徴，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1)，185-186, 2015
- 74：菊池拓磨他：バレーボール動画におけるサブデータ抽出に関する研究，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1)，183-184, 2015
- 75：井田正明：コレスポンデンス分析におけるデータ変動の影響，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1)，181-182, 2015
- 76：米澤剛他：エンドユーザ向けゲーム開発ツールの試作：トランプゲームを対象としたオブジェクトベースの解釈実行方式，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1)，177-180, 2015
- 77：多田圭佑他：VDM++ 分散モデル実行環境を用いたプロトタイピングと非同期 UI への適用，FIT 運営委員会，VDM++ 分散モデル実行環境を用いたプロトタイピングと非同期 UI への適用
- 78：夏目貴史他：オープンソースソフトウェア開発におけるコミュニティ活動の課題分析とその解決方法の提案，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(1)，167-172, 2015
- 79：齊藤尊他：過去の PBL の情報を用いた工数見積り支援ツールの開発，日本ソフトウェア科学会第 31 回大会講演論文集
- 80：井上哲哉：アプリケーションの挙動に影響を及ぼしている要因の測定方法に関する考察，FIT 運営委員会，アプリケーションの挙動に影響を及ぼしている要因の測定方法に関する考察
- 81：樽松理樹：ブロック単位の語句の出現頻度に基づく特許分類支援システムの提案，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(2)，233-238, 2015
- 82：飯尾淳：オープンソースソフトウェア技術の傾向分析，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(3)，85-88, 2015
- 83：宮本優他：連立一次方程式解法理解のためのタブレット PC を用いた教育用ソフトウェア開発，FIT 運営委員会，情報科学技術フォーラム講演論文集 14(4)，393-394, 2015
- 84：蝦名亮平他：リアルタイムバースト解析手法の提案，FIT 運営委員会情報処理学会，情報処理学会論文誌データベース (TOD) 5(3)，86-96, 2012
- 85：田中優斗他：未来の在室情報を予報する大室管理システム，情報処理学会，情報処理学会論文誌 54(9)，2265-2275, 2013

- 86：折原幸治他：二語の共通周辺文字列の長さに着目した語文脈類似判定，情報処理学会，二語の共通周辺文字列の長さに着目した語文脈類似判定
- 87：角田史記他：交通系 IC カードを利用した鉄道輸送障害時の影響を定量化する方法の研究，情報処理学会，情報処理学会論文誌データベース (TOD) 6(3), 187-196, 2013
- 88：吉川友也他：ユーザの潜在特徴を考慮したソーシャルネットワーク上の情報拡散モデル，情報処理学会，情報処理学会論文誌データベース (TOD) 6(5), 85-94, 2013
- 89：鍋谷洋介他：遺伝的アルゴリズムによる荷重メジアンフィルタの多目的最適化に基づく設計，情報処理学会，情報処理学会論文誌数理モデル化と応用 (TOM) 6(3), 106-116, 2013
- 90：塚田将太他：遺伝的アルゴリズムを用いた基底関数構築と非線形時系列予測，情報処理学会，情報処理学会論文誌 54(12), 2472-2480, 2013
- 91：秋岡明香：ストリーム解析処理のベンチマーク構築に向けたタスクグラフ自動生成ツール，情報処理学会，研究報告計算機アーキテクチャ (ARC) 2013-ARC-207(33), 1-6, 2013
- 92：小林泰三他：過渡現象数値計算のボトルネックとポスト処理連携，情報処理学会，研究報告計算機アーキテクチャ (ARC) 2013-ARC-207(23), 1-7, 2013
- 93：安井雄一郎他：ULIBC ライブラリを用いた共有メモリ型並列アルゴリズムの高速化，情報処理学会，ハイパフォーマンスコンピューティングと計算科学シンポジウム論文集 2014, 106-115, 2013
- 94：黒田明義他：高次元メッシュ/トラスネットワークにおける実アプリケーションの通信最適化手法—「京」上の Tofu ネットワークを例に—，情報処理学会，ハイパフォーマンスコンピューティングと計算科学シンポジウム論文集 2014, 97-105, 2013
- 95：佃洗摂他：属性値の同一性・相補性に着目したオブジェクト集合検索手法の提案とその観光地データへの適用，情報処理学会，情報処理学会論文誌データベース (TOD) 6(5), 49-61, 2013
- 96：藤原靖宏他：時々刻々と成長するグラフのための直径モニタリング，情報処理学会，情報処理学会論文誌データベース (TOD) 6(5), 1-10, 2013
- 97：高田雅美他：特異値分解アルゴリズムの性能評価のための大きな条件数を持つ行列作成，情報処理学会，情報処理学会論文誌数理モデル化と応用 (TOM) 6(3), 75-86, 2013
- 98：田村一樹他：OCR による文字認識誤りを考慮した重み付きトピックモデルに関する検討，情報処理学会，情報処理学会論文誌数理モデル化と応用 (TOM) 6(3), 66-74, 2013

99：東山翔平他：カテゴリ階層を考慮した構造化パーセプトロンによる固有表現抽出，情報処理学会，情報処理学会論文誌数理モデル化と応用（TOM） 6(3), 43-52, 2013

100：浜田龍平他：水中ワイヤレスセンサネットワークにおける伝搬遅延を考慮したタイムスロットスケジューリング，情報処理学会，情報処理学会論文誌数理モデル化と応用（TOM） 6(3), 1-10, 2013

謝 辞

本研究に際して，様々なご指導を頂きました内田眞司准教授，並びに副査を担当して頂きました井上一成教授に深く御礼申し上げます．また，中間発表にてアドバイスを頂いた方々にも感謝の意を表します．