

レポートのテーマ関連度と意見文抽出による情報量評価

川口 俊明[†] 砂山 渡[†]

近年、情報の多様化により、大学のような教育機関で電子レポートを提出する機会が増えてきた。手書きのレポートであれば、人間が目で見ると採点する必要があるが、電子的なデータとして存在するレポートであれば、そのレポート内の単語情報を計算機が自動的に取り出すことができるため、計算機による自動採点や、採点支援環境の構築により、人間の負荷の軽減が期待される。

本研究では、各レポートの情報量を、「テーマに関連する意見文、事実文の含まれる量」として定義し、この情報量に基づいて、レポート集合を2次元インタフェース上に視覚化するシステムを提案する。インタフェースを用いることで、ユーザが適切な情報量に応じたレポートを発見するのを支援する。

Report Evaluation based on Theme Relevancy and Opinion Extraction

TOSHIAKI KAWAGUCHI[†] and WATARU SUNAYAMA[†]

In recent years, as the Internet has grown, electronic reports have come to be used in educational organizations such as universities. Though reports written by hand must be evaluated by hand except for stereotype descriptions or numerical answers, electronic reports can be rated by computer.

This paper presents a report evaluation support system that displays each report with their rates based on information quantity in a two dimensional interface. Information quantity of a report is defined as the amount of sentences that describes facts and opinions. This system supports that users can evaluate reports by the informative contents.

1. はじめに

近年、情報の多様化によりメール、チャット、レポートなど多くの情報が電子的にやりとりされるようになってきており、大学など教育機関の講義課題としても、電子レポートを提出する機会が増えてきている。手書きのレポートであれば、人間が目で見ると採点する必要があるが、電子的なデータとして存在するレポートであれば、そのレポート内の単語情報を計算機が自動的に取り出すことができるため、計算機による自動採点や、採点支援環境の構築により、人間の負荷の軽減が期待される。

人間がレポートを評価する場面として、複数の評価者が評価に関わる場面、また学習者同士が評価を行う場面なども考えられるが、本稿では一人の評価者が全てのレポートを評価する場면을対象とする。

大学など教育機関の講義レポートは、文章を書くプロではないごく普通の個人が書いたものであるため、必ずしも、レポート課題と内容の一貫性が保たれてい

るわけではない。さらに、考察部分においては、自分の考えを述べる部分にも関わらず、意見文がないレポートも少なくない。そのため、意図がうまくつかめないレポート、当たり前のことを淡々と書いてあるレポート、無駄に長いレポートなどが存在しており、このようなレポートにはあまり高い評価は与えられない。

そこで本研究は、電子的に集められたレポートの「レポートのテーマ関連度」、「意見文」を評価し、レポート評価者に客観的なレポートの評価値を2次元インタフェース上に視覚化するシステムを提案する。レポートのテーマと関係ある部分、意見文、およびレポート集合全体の評価値の様子を提示することで、レポートの採点を支援できると考えられる。

本研究では、各レポートの情報量を「テーマに関連する意見文、事実文の含まれる量」と定義する。

2. 関連研究

2.1 レポートの自動評価

レポートを自動評価するシステムが多数提案されてきており、その一つとして、プログラミング課題のための自動採点システム¹⁾や、小論文の自動採点システム^{2),3)}などがある。これらのシステムは、課題に対す

[†] 広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

る解答の正当性を評価し、その提出された内容をキーワード情報から採点することを目的としている。テキストの構成をキーワード情報から採点するという点では、本研究と共通しているが、テーマ関連語から評価するという点で異なる。すなわち、文章のテーマ関連語を評価することにより、文章の一貫性を評価する。

また、類似するレポートの発見、評価支援を行う研究として、 \cos 類似度を用いた類似度計算法⁴⁾、 $tf \cdot idf$ を用いたベクトル空間法による類似度計算法⁵⁾、 n -gram 解析を用いた類似度計算法^{6),7)}、Smith-Waterman アルゴリズムを用いた類似度計算法⁸⁾ を評価基準として採点する研究がある。これらのシステムは課題に対する正当性ではなく、単語情報からレポート間の独自性を相対的に評価することを目的としている。本研究でも、独自性の高い単語を評価の対象としているが、評価指標が独自性一つではなく、テーマとの関連や、意見文の量から総合評価する点で異なる。

2.2 意見文抽出

従来から意見文抽出に関する研究は多数ある。文章から意見文を抽出する手法の多くは新聞やコラムなど説明的文章から機械学習をし、作成したデータベースを参照するものが多い。このデータベースからナイーブ・ベイジ法⁹⁾、Support Vector Machine¹⁰⁾、繰り返し学習¹¹⁾、手がかり表現を用いた単語の意味カテゴリ辞書の作成¹²⁾ を分類基準とする研究がある。これらの研究は、文末表現などの単語情報から意見文抽出する点で本研究と共通している。しかし、これらの研究は機械学習し、意見文の特徴が辞書を作成しているので、システムが大掛かりなものになりやすく、膨大なデータベースが要求されるという問題点があるのに対して、本研究では、機械学習をせず、文末表現のみから意見文を抽出すると言う点で異なる。

また、文章要約手法の多くは文章中の各文に対して重要度を与え、重要な順に文を抽出して出力する。重要度の評価方法としては、高頻度の単語が多数含まれるほど重要とするものや、テーマに関連する単語を多数含むほど重要とするものなどがある^{13)~16)}。本研究では、意見文と重要文はほぼ同じ意味であり、単語情報から重要文を抜き出す点では同じであるが、このような要約システムの場合、要約として選ばれなかった文は出力に含まれることはない。提案するシステムでは、意見文には強調して出力し、それ以外の文でも、単語情報を含めて出力している。そのため、要約システムでは見落とされがちな文を確認できたり、重要な文のテキスト全体の中での位置を直感的に把握することが可能になると考えられる。

2.3 情報の視覚化

情報を羅列するだけでは理解が非常に難しいため、直感的に理解するために情報を視覚化する研究がある。テキスト集合を分類してクラスタリングを行う研究¹⁷⁾ や、テキスト集合から話題構造を取り出して視覚化する

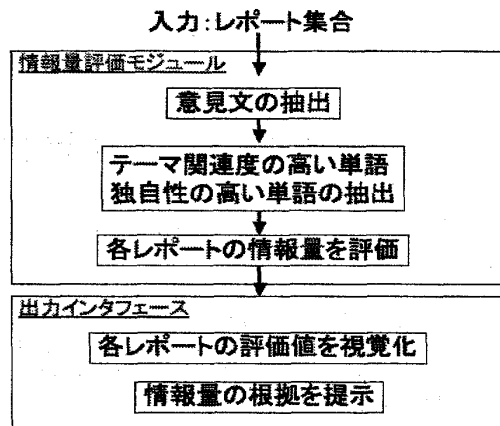


図1 レポートの情報量評価支援システムの構成

る研究¹⁸⁾もある。これらの研究では、各テキストがどの話題と関連が強いかを視覚的に表すことを目的としており、個々のテキストの関係ではなくテキスト集合全体の評価値を把握する本研究の提案システムとは異なる。

また、レポート間の \cos 類似度と多次元尺度構成法により、二次元平面上に散布図を表示するシステム¹⁹⁾では、評価者の客観的な評価を支援し、評価時に前に見たレポートの内容によって影響を受ける系列効果を減じている。

また、レポートのキーワードと文章量を軸として、レポートを色によって表すシステム²⁰⁾は、レポート評価者の客観的な評価支援を行うとともに、評価者の動機付けを与えている。

2.4 本研究の位置付け

本研究では電子的に集められたレポート全体のレポート課題に関連する結果、結果に対する意見の量を計ることでレポート文書を文章の長さによらない評価を与え、間、レポートの情報量をインタフェース上にグラフ構造や、情報量の根拠を含めたレポート本文表示で視覚的に表現することで、レポートの情報量における評価の指標をユーザに与えることを目的とする。

3. レポートの情報量評価支援システムの構成

本章では、レポートの情報量を評価するための支援システム(図1)について述べる。まず、ある課題に対するレポート集合をシステムへの入力とする。情報量を評価するモジュールでは、各レポートから抽出した品詞情報をもとに意見文を抽出し、レポート間の情報量を計算し、この情報量をもとに各レポートに評価値を与える。出力インタフェースでは、情報量の評価値の大きさに基づく座標上に、各レポートを配置する。また、レポートに与えた情報量の根拠を含めた本文を表示する。以下で、本システムの詳細について述べる。

表 1 意見文の文末表現

意見の表現	文末表現
考え	思う, 考える, 感じる, 言える
推量	う, かも
要望	たい, 望ましい
自発	れる, られる
主張	べき, ちがいない

3.1 レポート集合の入力

本システムへの入力は、一つのトピックについて電子テキストの形式で書かれたレポート文章の集合とする。ある1つの課題のレポート集合を入力とする。特にレポート内の考察部分のみを抜き出して入力とする。またレポート課題には、単に全員共通の正解を導くだけのものではなく、人によって解答の導き方が異なる課題や、考察など各個人の意見を述べる部分など、人によって内容が異なる部分が含まれると仮定する。この仮定は、レポートの独自性を評価することに意味をもたせるための条件となる。

対象となる入力テキストは、句点によって一文ごとに切り分けが可能なテキストで、かつ形態素解析によって品詞の抽出が可能なテキストとする。なお形態素解析には、「茶釜」²¹⁾を用いた。「茶釜」は品詞分解された単語をそれぞれ、出現系、読み、基本形、品詞、段、活用系に分けて出力する。

3.2 意見文の抽出

本節では、レポート本文から意見文を抽出する方法について述べる。入力されたレポートの各文の文末表現に注目し、表1の文末で終わっており、かつ、名詞数が3個以上の文を意見文とする。本研究では、文末とは、句点より前4単語以内としている。表1に該当しなかった文を事実文とし、さらにこのうち、名詞数が3個未満の文をどちらにも該当しない文として分類する。

3.3 各レポートの情報量の評価

本節では、各レポートの評価値を計算する方法について述べる。各レポートの情報量を、意見文とその他の事実文ごとに、各文に含まれる「テーマと関連する語」および「独自性の高い単語」をもとに100点満点で評価する。

テーマと関連する語の集合 R は、多くのレポートに出現する名詞(文書頻度上位20%)とし、独自性の高い単語の集合 D は、少数のレポートにしか出現しない名詞、動詞、形容詞(文書頻度下位5%)とする。名詞の中でも、形態素解析により、一般、固有名詞、サ変接続、形容動詞語幹、副詞可能、接尾-助数詞に分類される単語を対象とした。

一つのレポートに含まれる意見文の集合 A と事実文の集合 B を別々に評価し、各文に含まれる、テーマ関連語の数 $n(R)$ と独自性単語の数 $n(D)$ により、文 k の評価値 S_k を式(1)で与えた後、レポート i に、式(2)の情報量の評価 $Score(i)$ を与える。

$$S_k = \min\{\alpha(n_k(R) + n_k(D)), M\} \quad (1)$$

$$Score(i) = \left(\sum_{a \in A} S_a + \sum_{b \in B} S_b \right) \left(1 - \left| 0.5 - \frac{N(A)}{N(ALL)} \right| \right) \quad (2)$$

ただし、 α , M , はレポート集合 X に含まれる平均文数 $Average(X)$ によって式(3)、式(4)のように決められる。

$$\alpha = \frac{100}{5 \cdot Average(X)} \quad (3)$$

$$M = \frac{100}{Average(X)} \quad (4)$$

また、 $\sum S_a$ と $\sum S_b$ はそれぞれ50点を越えないものとする。また、 $N(A)$, $N(ALL)$ はそれぞれ、レポート中の意見文の数、全文数を表す。

また、意見文と事実文が同程度に含まれている場合には評価が高くなる。

ちなみに、一度評価に使われた単語は、二度と評価には用いられない。

3.4 出力

各レポートの情報量による評価値を視覚化する出力インタフェースを、図2の左半分、各レポートの情報量の根拠を右半分に示す。各レポートは、出力インタフェース中の円形ノード、およびそのノード名(仮名)で表され、その評価値がついた情報量の根拠を表す詳細をフォントの違いにより提示する。このインタフェースの出力方法について、以下で詳細を述べる。また、ノードの色の違いや、想定する本システムの使用方法については次節で後述する。

まずランダムに座標を与えることでインタフェース上の初期配置を作成する。

次に、初期配置をもとに、同じ評価値のレポートが同心円上来るように配置を移動する。あらかじめ、評価値 $Score(i)$ は、式(5)により $svalue(i)$ に変換しておく。同心円は、最高点、すなわち最も情報量が高いと考えられるレポートがインタフェースの中心付近に、レポートの情報量が低くなるにつれて中心から遠くなるように描く。すなわち、レポート i の座標(点 P_i) を式(6)によって点 P'_i に移動させることで図2の左半分のような出力インタフェースを得られる。ただし、点 O はインタフェースの中心、 $margin$ はインタフェースの中心部分の表示に余裕を持たせるためのマージン、 OP_i や OP'_i は点 O と点 P_i または点 P'_i をつなぐベクトル、 $|OP_i|$ はベクトル OP_i の長さを表す。

$$svalue(i) = \frac{Score(i) - \min_k \{Score(k)\}}{\max_k \{Score(k)\} - \min_k \{Score(k)\}} \quad (5)$$

$$OP'_i = |svalue(i) + margin| \cdot \frac{OP_i}{|OP_i|} \quad (6)$$

この同心円による配置と評価値の高低を直感的に理解するために、表2に示す基準でインタフェースに5つ

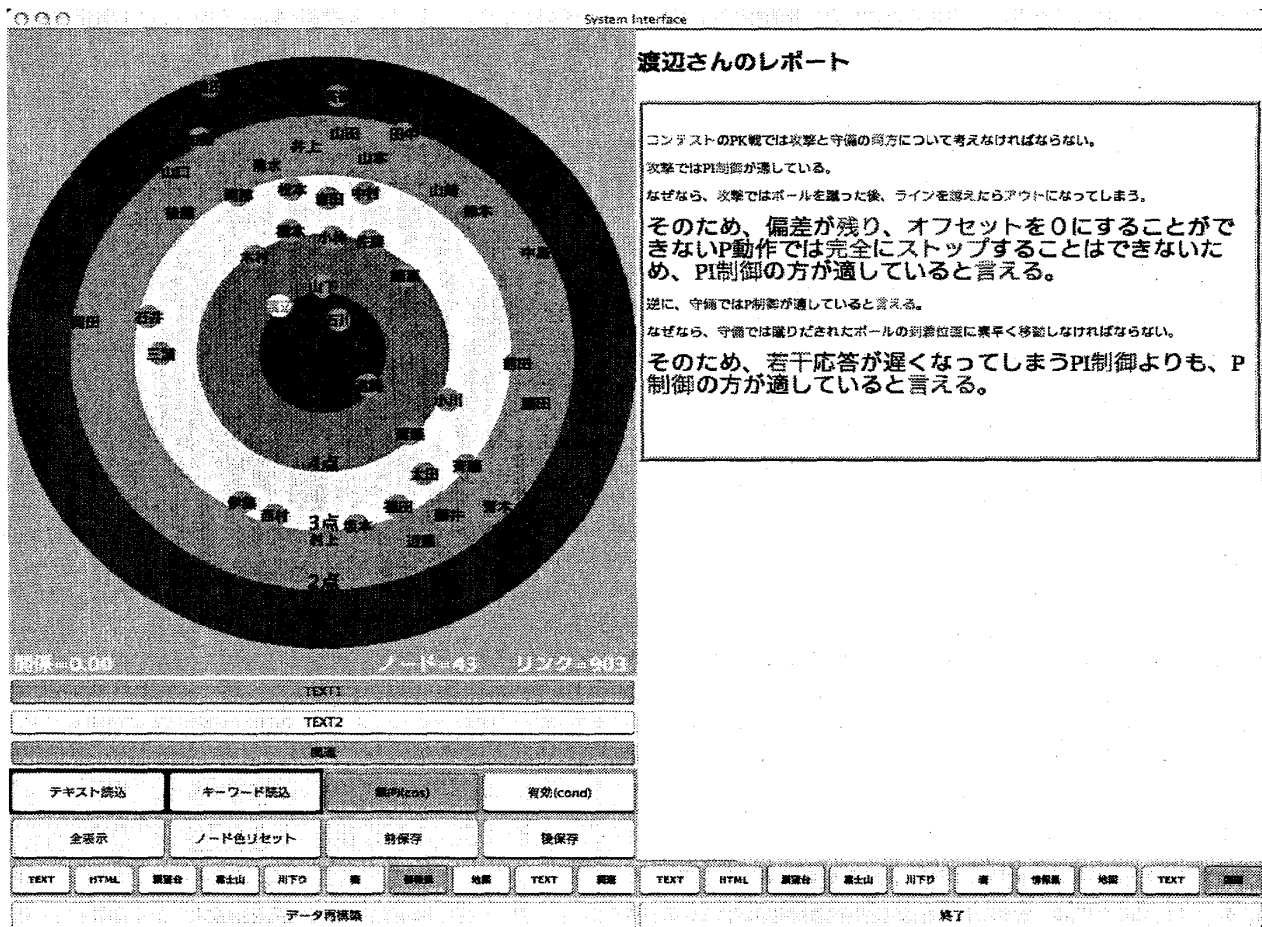


図2 レポートの情報量評価支援システムの構成

表2 色の塗り分け基準

色	評価値の分類基準	図中の点数
赤	$svalue(i) \geq 0.8$	5点
オレンジ	$0.6 \leq svalue(i) < 0.8$	4点
黄	$0.4 \leq svalue(i) < 0.6$	3点
緑	$0.2 \leq svalue(i) < 0.4$	2点
青	$svalue(i) < 0.2$	1点

の円を表示し、それぞれを色で塗り分ける。 $svalue(i)$ を用いることで、相対的に情報量の多いレポートと情報量の少ないレポートを視覚的に確認することが可能になる。

ユーザが特定のレポート（図2中では渡辺）をインタフェース上で選択（マウスでクリック）した際には、そのレポートの情報量の根拠を表すレポート本文が右側に表示される。この部分にはシステムが抽出した意見文を大きなフォントで表示し、テーマ関連度の高い単語を茶色、独自性の高い単語を緑色で表示している。本研究で定義した情報量の根拠をフォントの大きさや、色で表示することで、レポートに与えられた情報量の根拠を視覚的に確認することが可能になる。

3.5 レポートのテーマ関連度と意見文抽出による情報量評価システムの使用法

本節では、レポート評価者が、本システムを用いて実際に評価を行う手順について述べる。レポートのテーマ関連度と意見文抽出による情報量評価システムの全体画面は図2の通りである。画面の左半分は、前節で述べた出力インタフェースを表示し、右半分に情報量の根拠を示したレポート本文を表示している。なお、本システムのインタフェース全体は 1280×1000 ピクセルの解像度で作成した。

以下に、本システムを用いたレポートの評価の手順を示す。

1. 初期画面で各レポートの座標を保存
 2. 画面左側の同心円配置、右側に表示されるレポートの情報量の根拠をもとに評価
 3. システムが出力する点数と異なる点を与える際には、左側のレポートをドラッグ、再配置
 4. 各ノードの座標を再保存
操作前後のノード座標を保存することでレポートの修正の様子を調べることができる。
- また、レポートをクリックするとノードの色が最高

表3 クリック回数とそれに応じたノード色

クリック回数	ノード色
0回(初期)	茶色
1回	桃色
2回	水色
3回以上	黄緑色

で3段階変化し、ユーザにレポートの確認済み、確認忘れがわかるようにしてある。クリック回数とそれに対応するノード色を表3に示す。なお、図2中の白色のノードは現在選択中のノードを指す。

ユーザはレポートが出力されたインタフェース上において、各レポートの位置や、情報量の根拠をみることで、レポート集合の評価の全体像を把握できると考えられる。また、テーマ関連度の高い単語、独自性の高い単語、意見文を見ることで、内容がわかりにくいレポート、主張のないレポート、無駄に長いレポートを発見できることが期待される。

4. 予備実験

本章では本システムの意見文の抽出方法の妥当性を検証した実験、人間がつけるレポートの評価値の傾向を検証した実験について述べる。

4.1 意見文抽出実験

本節では本システムの意見文の抽出方法の妥当性を検証した実験について述べる。

4.1.1 実験内容

実験は Web 上でを行い、大学の C 言語による「プログラミング演習」、学生実験の「ロボット作成実験」の授業で実際に提出された学生レポート、人工知能学会論文誌²²⁾に掲載された論文の合計3種類のテキスト集合を用いて、各テキストから意見文を抜き出してもらった。各テキスト集合について、プログラミング演習レポートと学生実験レポートは考察部分を、論文は「結果と考察」と書いてある節から抜き出している。被験者には、情報科学を専攻する大学生、大学院生20名であり、テキストの内容を理解するための素養を有していた。

各テキスト集合に対して、被験者に与えた実験手順は以下の通りとした。

- 1 テキスト集合を一文ずつ読み、筆者の意見を述べている文(意見文)だと思ったらチェック
- 2 全文の判断が終了したらチェックしたデータを送信
各テキスト集合の詳細を以下に示す。一人の被験者に対して3種類全てのテキスト集合の意見文抽出を行ってもらい、人工知能学会論文誌の論文については、被験者の半数には1から5を、残り半数には6から10の論文を読むように割り振った。

4.1.2 実験結果

本システムが抽出する意見文と6割以上の被験者が意見文であると回答した文の適合率と再現率を表4に

表4 本システムと被験者が回答した意見文の適合率と再現率

テキスト集合	適合率	再現率
プログラミング演習	0.69	0.76
学生実験	0.97	0.93
論文	0.87	0.60

示す。

4.1.3 考察

まず、本研究で扱うようなレポート、すなわちレポートプログラミング演習レポートと学生実験レポートについて考える。学生実験レポートは適合率、再現率、ともに非常に高いのに対してプログラミング演習レポートは適合率、再現率ともに比較的低い。この原因として挙げられるのは、プログラミング演習レポートは全文中のうち意見文の占める割合が他のレポート集合に比べて低いことが考えられる。加えて、学生実験のレポートは大学3年生後期のレポートであるのに対して、プログラミング演習レポートは大学2年生前期のレポートであるため、プログラミング演習レポートのレポートとしての程度の低さも原因の一つと言える。その理由として、プログラミング演習レポートには課題に対する考察というよりも、「関数を多く使うと本当に複雑になると痛感した。」「関数を呼び出すアルゴリズムを書く際にどう書けばいいのかイマイチわからない。」「しかし引き値を理解するのはなかなか難しかった。」「などの課題に対する簡単な感想文が多い傾向にあった。被験者には「筆者の意見を述べている文を抜き出してください」という指示しか与えなかったため、このような文も意見文として抽出され、再現率が低下したと考えられる。また、「きちんと動作したため問題点や改良点はないと思われる。」「どうやって10人分の偏差値を出力しようか考えた。」のような特に意味のない文も多数あり、適合率も低下したと考えられる。

次に、論文のような専門家の文章では、再現率が低くなった。この原因として考えられるのは、被験者が選んだ文末表現に「今後の課題である」「必要がある」「必要である」「検討すべきである」などの「である」「がある」で終わる文が多かったことである。しかしながら、この「である」「がある」を含む文末表現は、レポート集合の中にはほとんど存在しなかったため、レポートに含まれるべき意見文の特徴には必要はないと考える。高い適合率を得ることができたので、論文のような専門家の文章でも、システムの意見文抽出方法が示された。

このことから、システムは一定の意見文抽出が行えることがわかった。

4.2 人間が与えるレポートの評価値の傾向を検証する実験

本節では、人間がつけるレポートの評価値の傾向を検証した実験について述べる。

4.2.1 実験内容

実験は Web 上でを行い、大学の C 言語による「プロ

表 5 提案システム、文字数と被験者の評価値の相関 (中央値)

評価方法	プログラミング演習	学生実験
提案システム	0.68	0.59
文字数評価値	0.80	0.92

プログラミング演習」, 学生実験の「ロボット作成実験」の授業で実際に提出された学生レポートの考察部分, 合計2種類のテキスト集合を用いて, 各レポートの情報量を5段階評価してもらった. この2つのレポート集合は4.1節の実験で用いたレポートと同じである. 被験者には, 情報科学を専攻する大学院生10名であり, テキストの内容を理解するための素養を有していた.

被験者には採点基準を「レポート課題に対する結果, および結果に対する意見の量」として, 情報量を評価するように指示した. また, 各レポート集合に対して, 被験者に与えた実験手順は以下の通りとした.

- 1 最初に, 表示される全てのレポートに目を通す
- 2 レポートを一つずつ読み, 「情報量」を評価して, 1点から5点を与える
- 3 全レポートの判断が終了したらチェックしたデータを送信

一人の被験者に対して3種類全てのテキスト集合の情報量を評価してもらった.

4.2.2 実験結果

提案システムが出力する評価値, 各レポートの文字数を5段階評価にしたものと被験者の与えた評価値の中央値との相関を表5に示す.

4.2.3 考察

表5より, 情報量の採点基準を与えたにもかかわらず, 被験者が与えた評価値は文字数評価値との相関のほうが高いことがわかった. この原因として考えられるのが, 被験者は大学院生であり, レポートの評価に精通していない素人であるためと考えられる. この結果から, レポートの評価に精通していない素人がレポートの評価を行う際, 評価基準が与えられても, ただ文章を読むだけでは, 長く書かれたレポートが良いレポートだと判断される傾向にあることがわかる. したがって, 文章量は多いが内容のない悪いレポート, 文章量は少ないが内容のしっかりした良いレポートをうまく発見できていない可能性がある.

したがって, 本システムの出力インタフェースを用いた採点によって, これらのレポートの発見ができることが期待される.

5. 評価実験

本章では, 提案システムの有効性を確認するために以下の要領で行った実験について述べる.

5.1 実験内容

実験は, 出力インタフェースの同心円上に, 1点から5点のエリア出力されるレポートに対して, その評

表 6 被験者の採点結果と文字数との相関 (中央値)

	集合 A	集合 B	集合 C	集合 D
提案システム	0.63	0.71	0.58	0.52
比較システム	0.90	0.89	0.84	0.80

表 7 比較システムを用いたユーザが提案システムより高く評価したレポート

レポート	意見文数 (割合)	テーマ 関連語数	独自性 単語数	平均文字数 に対する割合
A	0(0.00)	11	13	3.48
B	0(0.00)	46	11	2.58
C	0(0.00)	0	14	2.22
D	0(0.00)	17	3	1.56
E	0(0.00)	13	0	1.43
F	2(0.40)	29	0	1.34
G	1(0.20)	17	4	1.33
H	1(0.20)	19	3	1.00
I	0(0.00)	11	7	0.97

価値を修正してもらうことで行った.

被験者には採点基準を「レポート課題に対する結果, および結果に対する意見の量」として, 情報量を評価するように指示した. また, 各レポート集合に対して, 被験者に与えた実験手順は以下の通りとした.

- 1 インタフェース上に表示されるレポートを読み, 5段階で評価
- 2 システムが出力する点数と異なる点数を与える際には, インタフェース上でレポートをドラッグ, 再配置
- 3 全レポートの判断が終了したら座標を保存し実験終了

なお, 比較システムとして, レポートの文字数から5段階評価して配置した出力図を用意した. ただし, 比較システムでは, インタフェース右側のレポート本文表示部では意見文を大きく表示しない.

レポートは, プログラミング演習レポート2セット, 学生実験レポート2セット用意した. 実験はプログラミング演習レポートセット2つ, 学生実験レポートセット2つの順番で採点を行い, 提案システムと比較システムを用いた採点が均等になるように割り当てた. 各レポート集合の詳細を以下に示す. 被験者には, 情報科学を専攻する大学生, 大学院生20名であり, テキストの内容を理解するための素養を有していた.

5.2 実験結果

各被験者の採点結果と文字数評価値との相関を表6に示す.

次に, 各レポートについて, 各レポートの平均点が提案システムと比較システムで標準偏差以上差が出たレポート, すなわち, 比較システムを用いたユーザが提案システムより高く評価したレポート, 提案システムを用いたユーザが比較システムより高く評価したレポートについてそれぞれ表7, 表8に示す.

表 8 提案システムを用いたユーザが比較システムより高く評価したレポート

レポート	意見文数 (割合)	テーマ 関連語数	独自性 単語数	平均文字数 に対する割合
J	3(0.43)	6	9	1.55
K	1(0.17)	10	11	1.36
L	2(0.50)	8	8	1.27
M	1(0.20)	13	6	1.26
N	1(0.20)	10	10	1.24
O	2(0.67)	15	8	1.20
P	2(0.50)	14	9	1.18
Q	2(0.40)	15	14	1.08
R	2(0.50)	10	6	1.08
S	2(0.50)	6	10	0.96
T	2(0.50)	13	13	0.88
U	2(0.50)	12	10	0.88
V	1(0.33)	10	12	0.70
W	1(0.50)	4	4	0.67

5.3 考 察

表6について、比較システムと文字数との相関は高い値となっており、文字数に強く依存した採点が行われたことがわかる。4.2節で述べたように、レポートの評価に精通していない素人による採点結果であるため、このような結果になったと考えられる。学生実験レポートに比べてプログラミング演習レポートのほうがより高い相関値を取っているが、これは4.1節でも述べたように、プログラミング演習レポートはレポートの程度が低く、2、3行で終わるレポートの数は少なくない。このことから、文字数との相関はより高い値を取った可能性が高い。

提案システムでは、文字数との相関はある程度あるものの、比較システムに比べて、相関は低い値となっていることがわかる。このことから、被験者は提案システムを用いることで、文字数以外の内容を踏まえた採点が行われた可能性が高いと考えられる。提案システムでも、プログラミング演習レポートのほうがある程度高い相関値を取っているが、これも、レポートの程度という問題が指摘される。

また、表7、表8に示すように、各レポートの採点結果について、提案システムを用いた被験者の平均点と、比較システムを用いた被験者の平均点とで、標準偏差以上の開きがあるレポートの内容を確認した。

比較システムを用いたユーザが提案システムより高く評価したレポートは表7に示すように、意見文がないものが多かった。また、その他の特徴として、テーマ関連語、独自性単語の含まれる割合が極端であったり、レポートA、B、Cのように文章量が非常に多いものがあつた。比較システムでは、文章の長さに応じた採点の傾向があり、レポートA、B、Cのようなレポートが高く評価されたと考えられるが、そのうちレポートA、Bは意見文や、独自性単語が少なく、ただ当たり前のことを書き連ねていた。レポートCは、それに対して、独自性単語しか含まれておらず、さらに意

見文がないことから、意味の分からない文を書き連ねていた。その他のレポートに関しても、全体的に意見文が少なく、テーマ関連語が含まれる割合が高いことから、比較システムを用いたユーザはこのような、レポートとして情報量の少ないレポートは発見することが困難であったと考えられる。しかし、提案システムではこのようなレポートに低い評価を与えることができ、適切な評価ができたと考えられる。

提案システムを用いたユーザが比較システムより高く評価したレポートは表8に示すように、意見文の含まれる割合が5割前後であるものが多かったり、テーマ関連語と独自性単語をバランスよく含み、さらに、文少量が波であるものが多かった。これらのレポートは本研究が定義した情報量に沿ったレポートが多く、内容の正しさは無視したとして、筆者の意見をしっかりと述べている文が含まれていた。提案システムでは、意見文のフォントサイズを大きく強調したので、このような文を発見し、適切な評価ができたと考えられる。また、全体としては文章量は長くはないが、「これは、実機の左右のモーターパワーの微妙なズレ、キャスターの位置や転がり・滑らかさ、更には土俵の摩擦などの問題が考えられる」「これは、回転角の応答の場合、姿勢計測によるロボットの向きが不安定な動作をしているからだと考えられる。」「そのため、偏差が残り、オフセットを0にすることができないP制御では完全にストップすることはできないため、PI制御の方が適していると言える。」という意見として有効で、評価されるべき文が含まれていたことから、適切に評価されており、文字数に依存しない評価を与えていた。

したがって、提案システムを用いることで、文字数に依存しなく、適切な意見文を発見できることがわかった。このことから、レポートに適切な情報量による評価が与えられ、システムとしては妥当な評価値を出力できたと考えられる。

6. おわりに

本論文ではレポートのテーマ関連度と意見文を特定することで、レポートの情報量を評価し、視覚的に確認できるインタフェースを作成した。実験により、適切な意見文の発見や、本インタフェースが見た目によらない、レポートの情報量の評価支援に役立てられることを確認した。

今後は、インタフェース画面において、中心からの距離だけでなく、意見文数や、テーマ関連語、独自性単語の数を組み合わせ、総合評価からノードの座標を計算するように改良していきたいと筆者らは考えている。

参 考 文 献

- [1] 田上恒大, 阿部公輝: 比較的大きなプログラミング課題のための自動採点システム, 情報処理学会研究会報告, Vol.2006 No.16, pp135-140,(2006).
- [2] Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., and Harris, M. D. : Automated scoring using a hybrid feature identification technique, In Proc. Of the 17th annual meeting of Association of Computational Linguistics, pp. 206 - 210, (1998).
- [3] 石岡恒憲, 亀田雅之: コンピュータによる小論文の自動採点システム Jess の試作, 計算機統計学, Vol.16, No.1, pp3-18, (2003).
- [4] 砂山渡, 川口俊明: 内容の独自性の視覚化によるレポートの独自性評価支援システム, 人工知能学会論文誌, Vol. 23, No. 6, pp392 - 401, (2008).
- [5] 小川貴博, 岩堀祐之, 岩田彰: “情報メディア教育における類似レポート判定システムの構築”, 平成 13 年度電気関係学会東海支部連合大会講演論文集, No. 604, p. 304, (2001).
- [6] 太田貴久, 増山繁: 模倣レポート判定支援システムの開発, 言語処理学会第 11 回年次大会, pp. 293 - 296, (2005).
- [7] 小高知宏, 村田哲也, 高建斌, 諏訪いずみ, 白井治彦, 高橋勇, 黒岩丈介, 小倉久和: n-gram を用いた学生レポート評価手法の提案, 電子情報通信学会論文誌, Vol. J86-D1, No.9, pp. 702 - 705, (2003).
- [8] 太田貴久, 増山繁: 学生レポート採点支援のためのレポート類似部分発見手法, 電子情報通信学会技術研究報告, NLC2005, pp.37-42, (2006).
- [9] H.Yu, V.Hatzivassiloglou: Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, EMNLP, pp. 129 - 136, (2003).
- [10] 川口敏広, 松井藤五郎, 大和田勇人: 二段階アプローチによる Weblog からの意見文抽出, 電子情報通信学会研究報告, Vol. 106, No. 473, pp. 49 - 54, (2007).
- [11] 峠泰成, 大橋一輝, 山本和英: 繰り返し学習を用いた話題に順応する意見文抽出, 情報処理学会研究報告, Vol.2004, No.119, pp.43-50, (2004).
- [12] 廣嶋伸章, 山田節夫, 古瀬蔵, 片岡良治: 記述の重点に基づく意見文抽出手法, 言語処理学会第 12 回年次大会ワークショップ「感情・評価・態度と言語」, (2006).
- [13] K. Knight, and D. Marcu, Summarization beyond Sentence Extraction: A Probabilistic Approach to Sentence Compression, Artificial Intelligence, Vol.139, No.1, pp.91-107, (2002).
- [14] 大竹清敬, 岡本大吾, 児玉充, 増山繁, 自由作成要約に対応した新聞記事要約システム YEL-LOW, 情報処理学会論文誌「データベース」, Vol.43, No.SIG2(TOD13), pp.37-43, (2002).
- [15] 砂山渡・谷内田正彦, 文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム, 電子情報通信学会論文誌, Vol.J84-D-I, No.2, pp.146-154, (2001).
- [16] 吉見毅彦, 奥西稔幸, 山路孝浩, 福持陽士, 表題へのつながりに基づく文の重要度評価, 自然言語処理, Vol.6, No.1, pp.43-57, (1999).
- [17] 川谷隆彦: 多文書間の共通性分析に基づく文書クラスタリング, 情報処理学会論文誌, Vol47, No.6, pp.1903-1917, (2006).
- [18] 戸田浩之, 北川博之, 藤村考, 片岡良治, 奥雅博: グラフ分析を利用した文書集合からの話題構造マイニング, 電子情報通信学会論文誌 D, Vol.J90-D, No.2, pp.292-310, (2007).
- [19] 椿本弥生, 赤堀侃司: 主観的レポート評価の系列効果を軽減するツールの開発と評価, 日本教育工学会論文誌, Vol.30(4), pp.275-282, (2007).
- [20] Tsubakimoto, M., Yanagisawa, M., Akahori, K. : The Development and Evaluation of the Term Paper Grading Assistance Map Visualizing Contents and its Scores, Supporting Learning Flow through Integrative Technologies, T.Hirashima et al. (Eds.) IOS Press,, pp.197-204,(2007).
- [21] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸: 形態素解析システム『茶筌』, Version 2.2.9, 使用説明書, (2002).
- [22] 人工知能学会論文誌: <http://www.jstage.jst.go.jp/browse/tjsai/-char/ja>.