

# 内容の独自性を視覚化するレポート評価支援システム

## Report Evaluation Support System by Visualizing Contents Originality

川口 俊明<sup>\*1</sup>  
Toshiaki Kawaguchi

砂山 渡<sup>\*1</sup>  
Wataru Sunayama

<sup>\*1</sup> 広島市立大学大学院情報科学研究科  
Graduate School of Information Sciences, Hiroshima City University

In recent years, much information such as an email, a chat or a report came to be exchanged electronically by information diversification. As the electronic information can be easily edited by using the copy-paste function, texts similar contents increase that. We paid attention to a report text in this study. I suggest a system that evaluates originality of a report to grasp a characteristic or a tendency of the whole report by measuring objective relations among reports, and suggest an interface that we can evaluate visually each report originality. We plan reduction of load of a report rater.

### 1. 序論

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た。このような電子情報はパーソナルコンピュータや携帯電話でコピーやペーストなど簡単に編集することが可能である。そしてその情報を送受信したり補助記憶装置の普及によりデータの持ち運びも容易になった。それに従って類似する電子情報が増加している。

電子情報の中でも特にレポートは大学などの授業の課題の提出方法としてよく用いられている。しかし、グループ間などでレポートを写したり、写させたりすることが容易になってきたことから類似するレポートが増えてきた。そのため、レポート課題の出題者は評価することが困難になってきた。類似したレポートが多い場合、レポートの評価方法として、内容の独創性で図る方法がある。つまり、他人のレポートをコピーしたような内容は低い評価をし、他のレポートにはない内容だと高い評価をつける。しかし、この評価は評価者の目視で図ることは難しい。このため類似したレポートの関係を調べる指標が必要である。

本研究では、レポート全体の特徴や傾向を把握するために、客観的なレポート間の関係を測ることで、レポートの独自性を評価し、評価者に指標を与えることで評価の支援を行う。これにより、レポート評価者の負荷の軽減を図る。

### 2. 関連研究

本文グラフ分析を利用し、文書集合を対象とするテキストマイニングを行う方法として「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」[1]がある。これは文書をノード、文書間の類似度をエッジで表す文書集合グラフ中の各ノードの中心性を求めることで、文書集合中の話題構造をマイニングし、文書群の閲覧や分析を行うものであるが、本研究では一つのレポート課題についての複数文書の相違を分析し、グラフ中ではノード間の関係をエッジなどを用いて視覚的に表示するのではなく、各文書間の類似度を色分けなどによって視覚的に理解できることを目的としているところが大きく異なる。

他にもレポートの評価支援システムとして「模倣レポート判定支援システムの開発」[2]がある。これは文章の類似度をもとにコピーレポートを判別するものであるが、本研究ではコピーレポートだけにポイント置くのではなく、類似度をもとにすべてのレポートの評価の指針を与える点や、評価をインタフェースを伴って視覚的に表示できるようにする点が大きく異なる。

### 3. 内容の独自性を視覚化するレポート評価支援システムの構成

本システムでは与えられたレポート集合の独自性に関する評価を視覚化する。本章では本システムの構成とその詳細を述べる。

#### 3.1 システム構成

本研究ではレポート文章の集合を入力したときに、「各レポートから名詞を抽出」「レポート間の cos 類似度を計算」「各レポートの評価値を計算」「インタフェース上に同じ評価値のレポートを同心円状に配置」の4ステップを経て、出力として評価値の視覚化を得る。以下に各モジュールについての説明を述べる。3章ではこれらのシステムの構成とその詳細について述べる。

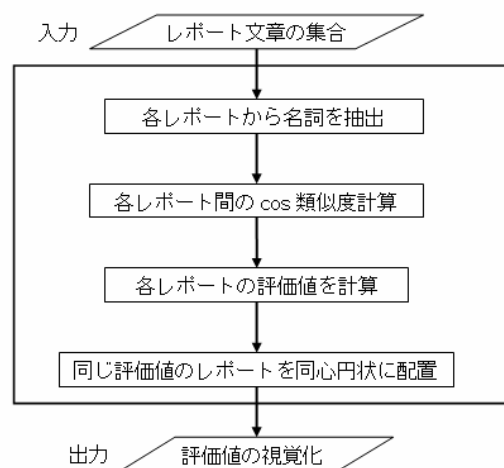


図 3.1: 提案システム構成

#### 3.2 レポート集合の入力

本システムへの入力、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。

#### 3.3 各レポートから名詞を抽出

品詞の中でも使用頻度が高く、表現がたくさんある名詞を抽出することで文章の独自性、または模倣を図れると考えられる。これより各レポートの文章から形態素解析 chasen[3]を用いて名詞のみを抽出する。

### 3.4 各レポート間の cos 類似度をもとに評価値を計算

文書間の類似度を図る方法の一つとして cos 類似度がある。cos 類似度はそれぞれのレポートの特徴を表すベクトルの内積をベクトルの長さで正規化することで文書の長さによらない類似度を求めることができる。レポート  $i$  と他のレポート  $j$  の関係の強さを式(3.1)の cos 類似度  $\cos(i,j)$  で測る。

$$\cos(i, j) = \frac{\text{レポート}i\text{とレポート}j\text{の共起名詞数}}{\sqrt{\text{レポート}i\text{の名詞数} \times \text{レポート}j\text{の名詞数}}} \quad (3.1)$$

各レポートには独創性に基づいた評価をつけるので、他のどのレポートとも類似していないレポートほど高い評価を与えたい。このことから式(3.1)を用いてレポート  $i$  の独創性の評価値  $\text{Score}(i)$  を式(3.2)で評価する。

$$\text{Score}(i) = \min_{1 \leq j \leq \text{レポート数}} \{1 - \cos(i, j)\} \times 100 \quad (3.2)$$

これは、レポート  $i$  と他のレポート  $j$  の cos 類似度  $\cos(i,j)$  のうち最大値をレポート  $i$  の独創性として評価する。この式(3.2)により、他のどのレポートとも類似していないレポートに高い評価が与えられる。

### 3.5 同じ評価値のレポートを同心円状に配置

まず、インタフェース上に各レポートを表すノードを配置する。このとき、ノードの初期位置をばねモデル[4]をもとに計算することで最適な位置に配置する。このばねモデルはレポート間の類似度が大きいほど弾性力が高くなるものであり、それを全ノード間で計算することによって各ノードの位置を決めるものである。図 3.2 にばねモデルでノードの初期位置計算し、配置したものを示す。また、ノード間を、与えた閾値以上の類似度を持つもの同士を結び、無向グラフで表す。

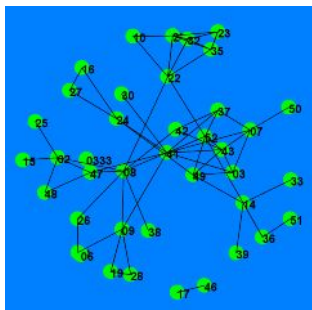


図 3.2:ばねモデルにより配置したノード

この初期配置をもとに、同じ評価値のレポートを同心円状に再配置し、最低点のレポートをパネルの中心に、以下評価値が高くなるにつれて外側に配置する。パネルの中心に置かれたレポート  $i$  (ノード  $i$ ) を点  $O_i$  としその他のレポート  $j$  (ノード  $j$ ) を  $P_j$  としたとき、 $P_j$  の座標は以下の式(3.3)のよって  $P'_j$  に更新される。

$$\overrightarrow{O_i P'_j} = \frac{|\text{Score}(i) - \text{Score}(j)|}{|\overrightarrow{O_i P_j}|} \overrightarrow{O_i P_j} \quad (3.3)$$

また、評価値の高低を直感的に理解するために、ノードの分布エリアを色で塗り分けた。色の塗り分けの基準を以下の表 3.1 に示す。

表 3.1:色の塗り分け基準とその解釈

評価値の分類基準	色	解釈
$\text{Score}(i) < 40$	赤色	独創性が低い可能性がある
$40 \leq \text{Score}(i) < 50$	黄色	独創性が中ぐらいの可能性はある
$\text{Score}(i) \geq 50$	青色	独創性が高い可能性がある

今回扱ったレポート集合の各レポートの評価値を計算したところ、評価値の最大値は 60 程度、最小値は 30 程度であった。このことを踏まえてレポートの内容と評価値を見比べてから評価値の分類基準を表 3.1 のように決めた。

### 3.6 出力

出力として 3.4 によって求めた評価値を 3.5 によって決めた分類基準で視覚化したものを図 3.3 に示す。

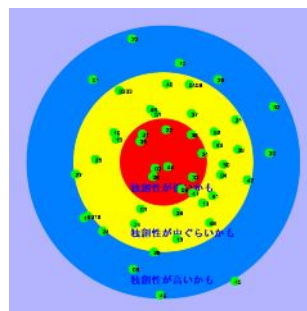


図 3.3:評価値をもとに再配置されたノード

緑色のノードは各レポートを指している。各ノードの配置は円の外側にいるほど評価値が高い。また、3.5 で述べたばねモデルでは各ノードの関係を線でつないでいたが、インタフェース上には評価値を視覚的に把握できればいいので線は削除することにした。

## 4. 内容の独自性を視覚化するレポート評価支援システムの使用法

本章では第 3 章で提案したシステムを実際に用いる際にユーザインタフェースを考慮して設計した「内容の独自性を視覚化するレポート評価支援インタフェース」について述べる。以下で、操作の方法について本インタフェースを用いたシステムを順に説明する。

### 4.1 レポート集合の入力

本システムの初期画面として、左上パネルにはノード表示枠が、中パネルにはレポートの内容表示枠、右端には評価値入力用の表が並べられている。

### 4.2 レポートの独創性評価

ユーザが右端の表の参照したいレポートの名前が書いてある行をクリックすると、そのレポートの内容が中パネルに表示される。次に、レポートの評価値を右端の表の“1回目”と書いてある列に入力していく。

### 4.3 提案システムの評価値の視覚化

右端の表の“1列目”の欄をすべて入力し終えたら表の最下部の“1”のボタンを押し、左下パネルの“blind off”のボタンを押すと、提案システムの評価値を視覚化したものが左上パネルに表示される。なお、ユーザが入力した評価値  $\text{Score}2(i)$  と提案システムの評価値  $\text{Score}(i)$  の差に比例してノード  $i$  の名前のフォン

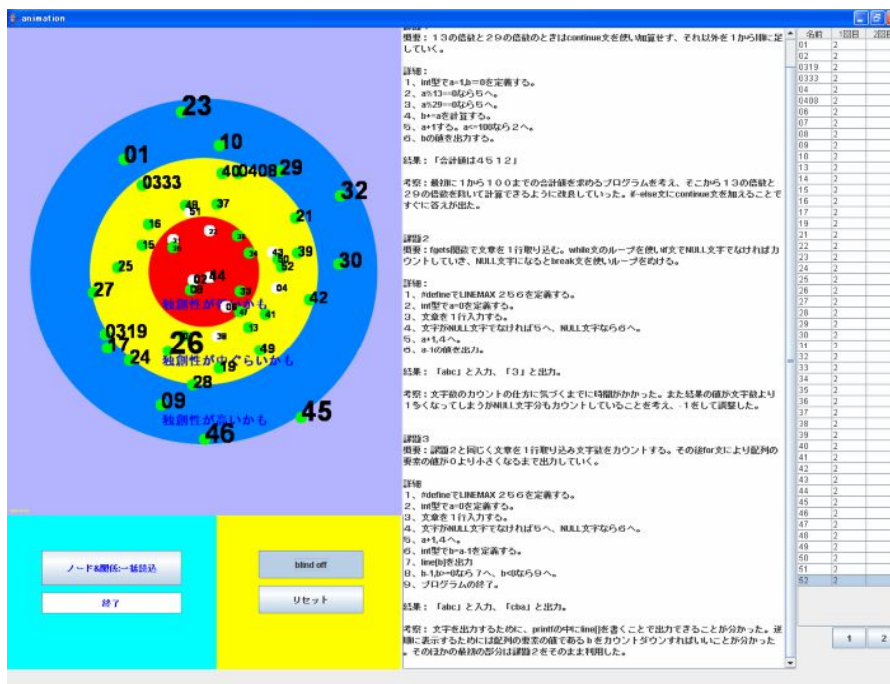


図 4.1: 提案システムとユーザの評価値の比較画面

トサイズ  $fontsize(i)$  を変える(4.1 式)ことで、違いを明らかにし、評価の再確認を促す。その様子を図 4.1 に示す。再確認した後、ユーザの評価に修正が必要だと判断した場合は、右端の表の“2回目”の列に修正した評価値を入力する。

$$fontsize(i) = \text{フォントサイズ初期値} + |\text{Score}(i) - \{\text{Score}(2i) - 1\} \times 25| \div 2 \quad (4.1)$$

## 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 使用データ、被験者

実験に使用するデータとして、本学のプログラミング演習について書かれたレポート集合を2つ(49人分と51人分)用意した。被験者は男女大学生16名とした。以後、49人分のレポート集合と51人分のレポート集合をそれぞれレポート集合1、レポート集合2と呼ぶことにする。レポートの記述形式と、被験者に評価してもらったレポート集合1とレポート集合2の課題を以下に示す。

各レポートは、3つのプログラミング課題に対する解答を以下の形式で記述している。

1. 概要: 作成したプログラムの基本的な考え方
2. 詳細: プログラムの詳細な流れ(アルゴリズム)  
 <アルゴリズムの記述形式>
  1. ○○を入力する
  2. ○○をもとに××を計算する
  3. ××が△△なら4へ、そうでないなら5へ
  4. ××の値を□□倍して3へ
  5. ××を出力する
3. 結果: 作成したプログラムの動作結果
4. 考察: プログラムを作成してわかったこと、工夫したところ、改良点、問題点

レポート提出者にはこの記述形式を指定した。さらに、詳細(アルゴリズム)には細かい記述形式を指定し、提出者はこれに

倣ってレポートを記述したため、被験者には詳細(アルゴリズム)を見る必要がないと指示した。

### 5.2 実験手順

被験者16名にレポート集合1、レポート集合2の各レポートそれぞれの独創性を評価してもらった。実験の手順を、順を追って説明する。

**Step1** レポート集合1の内容を参照しながら各レポートの独創性を5段階評価により点数をつけてもらった。

**Step2** インタフェースを使用して被験者の評価値を再確認し、評価値に修正があれば修正してもらった。

**Step3** レポート集合2の内容を参照しながら各レポートの独創性を5段階評価により点数をつけてもらった。

**Step4** インタフェースを使用せずに被験者の評価値を再確認し、評価値に修正があれば修正してもらった。

被験者のうち、半数の8人には Step2 と Step4 を入れ替えて行った。

### 5.3 実験結果

評価実験で得られた結果と考察を以下に示す。

#### (1) 所要時間

評価と再確認にかかった時間をインタフェースの使用の有無についてまとめたものを表 5.1 に示す。

表 5.1: 評価時間と再確認時間の中央値

		評価時間(分)	再確認時間(分)	合計(分)
レポート集合1	インタフェース使用	14.5	10	23.5
	インタフェース非使用	18.5	5	25
レポート集合2	インタフェース使用	20	17.5	35
	インタフェース非使用	16	6.5	26

表 5.1 より、インタフェースを使用しながら再確認を行ったほうがインタフェースを使用せずに再確認を行うより時間がかかることがわかった。また、インタフェース非使用のときだけ再確認時間が0分の被験者がいた。

これは自分の評価とは違う、システムの評価なのでレポートの内容を確認する必要があるためだと考えられる。インタフェースを利用したほうが時間がかかっているのに、しっかり確認は行われており、ただシステムに盲従して修正しているわけではないと推測される。また、インタフェース非使用のときだけ再確認時間が0分の被験者がいたことについては再確認をする際に修正意欲が湧くかどうかに関係あると考えられる。

(2) 再確認後における評価値の修正率

再確認する際に修正箇所割合、つまり修正率の中央値をまとめたものを表 5.2 に示す。

表 5.2:再確認後の修正率の中央値

	レポート集合1	レポート集合2
インタフェース使用	19%	12%
インタフェース非使用	21%	31%

表 5.2 よりインタフェースを使用したときのほうが、インタフェース非使用のときより修正率が低くなった。

まず、5.4(1)よりインタフェースを使用したほうが再確認に時間がかかっているのに修正はしっかり行われていると考えられる。ただし、修正箇所としては、インタフェースが表示する自分の評価とは違うシステムの評価値の必要十分な場所だけを見るので、数は減ると推測される。これに反してインタフェース非使用の場合には、どのレポートを修正すべきなのか指針が存在せず、あれやこれやとレポートを探っては修正する事になるので、インタフェース使用のときのほうが修正箇所に無駄がなく修正率が低くなったと推測される。

(3) システムと被験者の評価値における評価後と再確認後の相関

再確認を行った際に評価者がシステムの評価値に近い修正を行ったかを図るために、システムと被験者の各レポートの評価値における評価後と再確認後で相関係数をとった。その相関係数の中央値をインタフェースの有無についてまとめたものを表 5.3 に示す。

表 5.3:システムと被験者の評価値の相関係数の中央値

		評価後	再確認後
レポート集合1	インタフェース使用	0.10	0.37
	インタフェース非使用	0.06	0.12
レポート集合2	インタフェース使用	0.05	0.32
	インタフェース非使用	0.02	0.09

表 5.3 より、インタフェースを使用して再確認を行ったほうが再確認後の相関係数が増加した。また、レポート課題の出題者にも事前に評価実験を行った際に、出題者とシステムの評価値の相関係数は0.23であった。

これは、再確認をしっかり行った後での結果なので、システムの評価値が被験者にとってある程度正しいと判断されたと考えられる。出題者とシステムの評価値の相関係数 0.23 と比較して、インタフェースを使った被験者の評価値がこれに近い結果が得られた事が分かる。しかし、被験者とシステムの評価値の相関係数が出題者とシステムの評価値の相関係数よりやや高目の相関が得られており、システムの提示に合わせてしまう傾向も少しありと推測される。

(4) 再確認後の評価値の変化

各レポート集合において、インタフェースを使用して再確認を行った被験者とインタフェースを使用せずに再確認を行った被験者は8名ずつである。この8名が、各レポートに与えた5段階評価の評価値のうち、各評価値ごとのレポート数を求めた。この際、評価時と再確認後で各評価値の人数の変化をまとめたものを表 5.4 に示す。

表 5.5:被験者 8 人中の評価値の変化

			1点	2点	3点	4点	5点
			(人)	(人)	(人)	(人)	(人)
レポート集合1	インタフェース使用	評価時	49	116	129	70	28
		再確認後	35	113	150	72	22
	差		-14	-3	21	2	-6
	インタフェース非使用	評価時	26	123	155	57	31
		再確認後	28	118	149	67	30
	差		2	-5	-6	10	-1
レポート集合2	インタフェース使用	評価時	25	94	193	67	29
		再確認後	25	105	193	66	19
	差		±0	11	±0	-1	-10
	インタフェース非使用	評価時	24	97	158	92	37
		再確認後	20	116	119	114	39
	差		-4	19	-39	22	2

表 5.5 より、インタフェースを使用して再確認を行った被験者は1, 5点の極端な値の修正を行っているという特徴があり、インタフェースを使用せずに再確認を行った被験者は2~4点の中央の値の修正を行っているという特徴があることがわかった。

インタフェースを使用せずに修正した場合、2~4点の中央の値の移動が見られたが、これは修正の指針がなく、修正したとしても微調整しかできなかったためこのような結果になったと考えられる。また、表 5.5 より、相関にあまり変化がなかったことも、修正の指針がないことから微調整しかできなかったことが原因だと推測される。インタフェースを使用して再確認を行った際、修正の前後で1, 5点の人数が減っているが、評価時の人数から再確認後に3割程度の減少なので、むやみに1, 5が減っているわけではなく、きちんと見直した上で、不適切なもののみが修正されていると考えられる。このことからシステムの評価値から外れた被験者の評価値を修正するという効果があったと推測される。

6. 結論

本論文では客観的なレポート間の関係により、レポートの独自性を評価し、視覚的に確認できるインタフェースを作成した。本インタフェースがレポート評価の支援に有効であることを確認した。

今後は、システムの精度の向上のために類似度、評価値に重みをつけていくことや、インタフェース上にレポートの関係をより直感的に捉えることができるように改良したり、今回はプログラミング課題のレポートについて実験を行ったが、それ以外の内容の電子テキストでも実験を行うなどの課題が挙げられる。

参考文献

[1] 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, Vol.90 No.2, pp292-310(2007)  
 [2] 太田, 増山: 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).  
 [3] 日本語形態素解析システム「chasen」:  
 (URL) <http://chasen.naist.jp/hiki/ChaSen/>  
 [4] 杉山公造: グラフ自動描画法とその応用- ビジュアルヒューマンインターフェース, 計測自動制御学会, コロナ社(1993)