

プロダクトおよびプロセス分析による剽窃検出法の比較と検討

心理学 宮本 淳, 物理学 仙石昌也, 外国語 山森孝彦, 外国語 久留友紀子, 数学 橋本貴宏

医学部初年次チュートリアル教育において Google ドキュメントを用いたレポート課題を課し、そのレポートファイルから取得した変更履歴を辿ることで剽窃行為を中心にレポートの作業過程の分析を続けている。変更履歴からレポート作成プロセスを視覚的に把握することができるとともに、直前の履歴からの文字数の大幅な増加及び不自然な改行や書式の違い等から剽窃行為と判断することができた。本研究では、同一レポートに対して、変更履歴によるプロセス分析と剽窃チェックソフトによるプロダクトの剽窃判定を行い比較した。その結果、剽窃チェックソフトから検出されるコピー率と文字数増加から検出されるコピー回数との間には正の相関が見られた。両手法の検出結果が一致しないレポートを調査することにより、それぞれの手法の特徴が明らかになった。剽窃の検出には両手法を組み合わせることにより、より精度が高まり、詳細な確認が可能になると考えられる。

1. 背景

(1) プロダクト分析による剽窃の検知

大学生がレポートを作成する際に Web 上に存在する文章をコピー&ペースト（以下、コピー）し、それをあたかも自分の意見であるかのようにしてしまう、「コピー剽窃問題」が社会的問題として取り上げられている。筆者らは、2008 年度から初年次教育の中で剽窃行為についての調査を行っている。調査初年度は、学生のレジメの構成と引用文献に関する調査と分析を行った（仙石ほか 2008）。その結果、レポートのほぼすべてがコピーだけで構成されたものが 27.4 % も見られ、学生のレポートの多くがインターネット上の文章のコピーの寄せ集めであることが明らかになった。同様の傾向、すなわち大学生がレポートを書く際に一定の割合でコピー剽窃を常習的に行っていることは杉光(2010)など、いくつかの実態調査研究からも指摘されている。

光原(2011)では、Web ページ上の情報をそのままコピーして、情報を吟味しなくても、容易に体裁の整ったレポートを作成できてしまうことを、知識構築に至らない“非生産的コピー”としている。自ら考えることなく、コピー（剽窃）に頼ったレポート作成は大学生の能動的学習・課題探求能力・問題解決能力育成の妨げに繋がる大きな要因の一つであろう。

このようなコピー問題の対策の一つとして、金沢工業大学の杉光らによって、レポート内のコピー行為を発見する剽窃チェックソフト（コピーペルナー）が開発され話題になった。これを受けて、藤本・川合・志村（2009）、花川（2013）など各大学独自のコピー検索システムの開発について論文報告されるようになってきている。

剽窃チェックソフトによるプロダクト分析のメリットとしては、完成したプロダクトだけを判定するため、初期設定を行うだけで分析にかかる労力が少ないこと、コピーが疑われる箇所だけでなくコピー元のホームページや文献も自動的に検出する機能を有していること、などが挙げられる。

デメリットとしては、ソフトが検出したキーワードに基づいて類似した文章を検索するシステムであり、文章の一部を加工したようなコピーには対応していないため、インターネットからのコピーについては検出精度が必ずしも高くないという点が挙げられる。例えば、近藤(2014)は、学生に「できる限り意図的に Web 上からのコピーを駆使して執筆する」ように義務づけたレポートを課し、コピーペルナーの検出精度をみた。その結果、平均 74.1 % のコピーをしているはずのレポートに対して 24.2 % だけしかコピーは検出されず、チェック精度は必ずしも高くないことを報告している。筆者らも 2009 年から検索ソフト（コピーペルナー）を用いたコピーの実態調査を行ってきたが、この報告と同様、“非生産的コピー”をしているレポートであるにもかかわらず、コピーペルナーでは一部しか検出されず、コピー率が低く判定されたものが確認された。個々のレポートの内容を精査していくと、レポート全体の分量に対して過半数がコピーで構成されていると判断できるレポートは、コピーペルナーのコピー率が 20 % 以上で増加する傾向が見られた。（宮本ほか 2015）

(2) 編集履歴を用いたプロセス分析による剽窃の検知

上述の経緯から筆者らは、提出された学生のレポートというプロダクトではなくて、レポート作成プロセスか

ら剽窃にアプローチする手法に着目した。

編集履歴を用いたレポートのプロセス評価についての先行研究では、例えば、永井・中村(2010)は独自の Web 上のアプリケーションで作成されたエディタを用いて、「コピー」「ペースト」が行われる毎に履歴情報を取得することや一定量の文字入力削除動作を伴わずに行われた場合に、剽窃の可能性としてチェックする機能を備えたシステムの開発過程を報告している。また山口・永井・横山(2009)は xyzzzy というテキストエディタに編集履歴保存機能を付与することで、編集履歴からレポート作成プロセスにアプローチしている。

ただ、このような各大学や研究者によって作成された独自のシステムを用いることでレポート作成プロセスにアプローチする報告はあるものの、既存の汎用性の高いエディタを用いた研究報告はほとんど見られない。

これらに対して筆者らは、オンラインストレージサービスの一つである Google ドキュメントを用いたレポート課題を課し、そのレポートファイルから取得した変更履歴を辿ることでコピペに注目してレポート作成プロセスを調査した(宮本ほか2013)。

Google ドキュメントでは各変更履歴で文字数増加が異なる色で表示されることから、レポート作成プロセスを目視で把握する事が可能であった。コピペという点から作成プロセスを見ると、削除や訂正を伴わずに大学生の平均入力速度を大幅に超える文字が一度に入力されるなど、履歴間で増加した文字数及び不自然な改行や書式の違い等の入力状況に注目することで、これらの中に剽窃行為(コピペ)と判断できるものが多く見られた。

プロセス分析のメリットは、履歴間の大幅な文字数増加だけに注目することにより、作成プロセスにおけるコピペの状況を確実に検知できることにある。さらにプロダクト分析だけでは見えてこなかった、「コピペした部分をどのように加工しているのか」、「複数のウェブページの文章をどのようにまとめているのか」など、インターネットを用いたレポート作成における特徴などの情報を把握することができる。

Google ドキュメントを用いたレポート作成は学生側のメリットも大きい。様々な OS で動作し、PC やタブレット、スマートフォンなど様々なデバイスにも対応し、ネット環境につながっていれば、時間的、空間的な制約なく編集作業を行うことが可能である。

デメリットとしては、プロセス分析はレポート作成過程のすべての変更履歴を分析対象とするため、現状ではその分析に多くの時間と労力を必要とすること、コピペ剽窃の検出において、履歴間での大幅な増加文字数を「コ

ピペ剽窃の可能性のある文字の集まり」として検知するだけであり、その内容や出典については個別に確認する必要があることが挙げられる。

2. 研究の目的

上述のような研究背景において、先行研究では図表などが含まれる一般的なレポートに対して、汎用性の高いシステムを用いたプロセス分析による剽窃検知に関する研究がほとんど見られないこともあり、剽窃チェックソフトを用いたプロダクト分析と変更履歴を用いたプロセス分析による剽窃の判定結果を詳細に比較した研究は行われてこなかった。

そこで、本研究では、Google ドキュメントを用いて、履歴間における文字数増加を見ることにより、同一レポートに対して、変更履歴のプロセス分析と剽窃チェックソフトによるプロダクトの剽窃判定を行った。

両分析の結果を比較検討することで、両分析の特徴及び関連を明らかにすることを目的とする。

3. 方法

(1) 対象および分析対象レポート

初年次対象の基礎科学プレチュートリアルを受講した平成 26 年度学生 113 名のレポートを対象とした。

分析対象レポートは上述の授業で実施した個人研究発表の際に使用するレジメである。個人研究発表では、学生がそれぞれ自由にテーマを設定し、それまでのグループワークで培った経験をもとにして、個人で調査・発表を行う。レジメは発表の際の補助資料として作成させているが、一般的なレポートに近い内容になっている。

書き方については、Google のオンラインストレージサービスである Google ドライブの共有ファイル内に予めレポートのテンプレートを Google ドキュメントで作成しておき、そちらに各自で書き込んで提出するよう指示した。図 1 にテンプレートファイルを、図 2 に Google ドキュメントの変更履歴表示画面を示す。テンプレートファイルのタイトルがグループワークとなっているが、構成や体裁は個人研究と同じものである。

(2) プロセス分析方法(変更履歴の分析)

図 2 の画面右側のように、Google ドキュメントでは概ね 1~3 分の変更履歴が自動的に記録される。各変更履歴では、履歴間で入力された文字やコピペされた部分は違う色で表示され(図では色が薄くなっているのが追加・変更箇所)、削除された文字は取り消し線が加えられる。そのため変更履歴を辿ることで作業過程を視覚的に把握することができる。

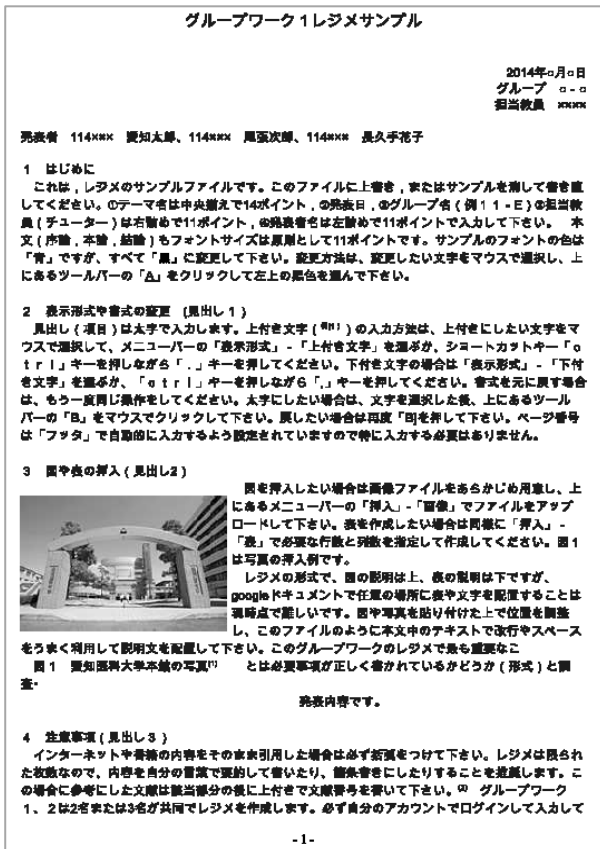


図1 レジメテンプレート



図2 Google ドキュメント変更履歴表示画面(右側は変更履歴と記録時間, 編集者名)

筆者らの研究(宮本ほか2013)では、レポートの作成開始から完成までに保存された全ての履歴を一つずつ表示させて、Excelのセルに貼りつけ、時間や文字数の差を計算した。この中から履歴間で手入力が不可能と思わ

れる増加文字数の履歴を抽出し、それらがコピー(剽窃)かどうかを判定したところ、1履歴あたり150文字以上増加したものについては、その多くが表1に示した判断基準によって「コピー(剽窃)」と判定された。

本研究では、この分析方法の改良版による増加文字数の抽出を行った。履歴情報をより正確に反映させるため、履歴をスタイル形式も含めてExcelのシートに貼り付け、フォントカラーの違いによる「削除」「入力」文字数をExcel VBAプログラムにより計算し、各履歴間での増加文字数を求めた。このようにして、まず150文字以上の文字数増加を伴う履歴を抽出し、ファイル内移動(カット&ペースト)など明らかにコピーではないものを除いた後、表1の判断基準を基に、複数の評定者にて「コピー(剽窃)」と判定される履歴を検出した。

なお、本研究で「コピー(剽窃)」履歴とは、杉光(2010)を参考に、「コピー&ペースト機能を用い、他人の文章等を写して自分の文章等と詐称する行為」という定義に該当する履歴とした。

表1 大幅な文字数増加に対するコピーの判断理由

| |
|--|
| 1. コピーであると判断した理由 |
| (1) 削除・訂正を伴わない文章が一気に増える |
| ・1分あたりの文字数増加が訂正なく150字以上 |
| ・段落としてのかたまりで増えている |
| ・文章の途中からいきなり完成している |
| (2) 不自然さ |
| ・文体(言い回し)の変化が見られる |
| ・改行が多い・不自然な改行になっている |
| (メールからのコピーの可能性が高い) |
| ・フォントの種類・サイズが異なる |
| 2. 実際に入力されたと判断した理由 |
| (1) 実際に文章を作成している様子 |
| ・履歴を辿ると少しずつ文章作成を進めている |
| (2) レポートに合った段落構成 |
| ・1履歴の中で複数の箇所を書いている |
| ・本文に合わせた改行になっている |
| (3) 手入力特有の現象 |
| ・入力ミス・変換ミスがある |
| ・文末が入力途中のアルファベットで終わっている |
| 3. コピーかどうか分からないもの |
| ・大幅に文字数は増えているが、前の履歴との時間が大きく空いているため判定できない |
| ・入力した可能性は高いが判断は難しい |

(3) プロダクト分析方法(剽窃チェックソフトによる分析)

プロダクト分析として、同一レポートについてコピペルナーV3 (v.3.0.3) にかけて、コピペ判定結果で得られたコピペ率を分析結果として使用した。コピペルナーの設定項目は表2のようにした。この設定は、コピペルナーを用いて、インターネットからのコピペ状況を調べる「インターネット検索のみの実行」のデフォルト設定から、検索ページ数を増やし、先行研究における検索頻度を参考に、特定ドメインリストを追加したものである。

表2 コピペルナーV3の設定項目

| | |
|--------------|-------------------|
| キーワード自動設定 | : 有効 |
| HTML 検索ページ数 | : 100 |
| PDF 検索ページ数 | : 20 |
| 特定ドメイン検索ページ数 | : 20 |
| 特定ドメインリスト | |
| Wikipedia | ja.wikipedia.org |
| 教えて! goo | oshiete.goo.ne.jp |
| allabout | allabout.co.jp |
| J-castNews | www.j-cast.com |
| コトバンク | kotobank.jp |
| 総務省 | www.soumu.go.jp |
| 厚生労働省 | www.mhlw.go.jp |
| 経済産業省 | www.meti.go.jp |
| CiNii | ci.nii.ac.jp |

4. 結果

(1) 分析対象

履歴数が極端に少ないなど、他のレポートと著しく異なる特徴を有する5部を除いた108部のレポートを分析対象とした。各レポートは図表も含めてA4用紙3枚程度のものである。対象とした履歴総数は20,530で、1レポートあたりの平均変更履歴回数は190.1回であった ($N = 108, SD = 81.6$)。

直前の履歴から150文字以上増えているものは209回で、履歴総数の1.0%であった。ファイル内移動(カット&ペースト)など明らかにコピペでないものを省き、そのうち、表1の判断基準による評定で「コピペ(剽窃)」を1回以上行ったと判断されたレポート数は34部、1レポートあたりの平均は1.07回、延べ116回検出された ($N = 108, SD = 2.18$)。1レポートあたりのコピペ回数の分布は0~10回であった。

コピペルナーによる判定結果から得られたコピペ率の平均は17.1% ($N = 108, SD = 10.0$) であった。コピペ率の範囲は2.5~54.2%であった。

(2) 履歴からみたコピペと剽窃チェックソフトによるコピペ判定結果の関係

同一レポートにおいて、「コピペ(剽窃)」と判定した履歴回数(以下、コピペ回数)と剽窃チェックソフトによるコピペ率の判定結果との相関係数を算出したところ、比較的強い正の相関がみられた ($r = .578, p < .01$)。対象となった108部のレポートについて、コピペ回数とコピペルナーによる判定結果の分布を図3に示す。縦軸はコピペ回数、横軸はコピペルナーによるコピペ率を表している。

(3) コピペ(剽窃)の識別について

本研究では、プロダクト分析からコピペが行われたと判定する基準として、背景で述べた理由からコピペルナーでのコピペ率が20%以上のレポートとした。一方、プロセス分析では、履歴間の増加文字数に着目し、150文字以上増加した変更履歴が1回以上検出された場合をコピペが行われたレポートと判定した。

この2つの観点で分類したコピペ判定法による結果を表3に示す。このデータに対して χ^2 検定を行ったところ、独立性の仮説は棄却された ($\chi^2(2, N = 108) = 16.40, p < 0.01$)。レポートからコピペの有無を識別する際に、コピペルナーの判定結果と150文字以上の文字数増加の変更履歴有無には関連があると言える。次に、直前の履歴から150文字以上増加して、「コピペ(剽窃)」と判定された116回の履歴について、増加文字数別に4群に分類し、コピペルナーによるコピペ率との関連を調べた。その結果を図4に示す。コピペ率が20%未満の群では、150~300文字増加のコピペが多く、一度に500文字以上コピペしているレポートはほとんどない。一方コピペ率が20%以上になると、300~500文字以上のコピペの増加が目立ち、30%を超える群では一度に500~1,000文字以上の大きなかたまりでコピペをする傾向が見られた。

表3 編集履歴から見たコピペとコピペ率から見たコピペ

| | 20%以上 | 20%未満 | 合計 |
|------------|-------|-------|-----|
| 150文字以上増加有 | 19 | 15 | 34 |
| 150文字以上増加無 | 13 | 61 | 74 |
| 合計 | 32 | 76 | 108 |

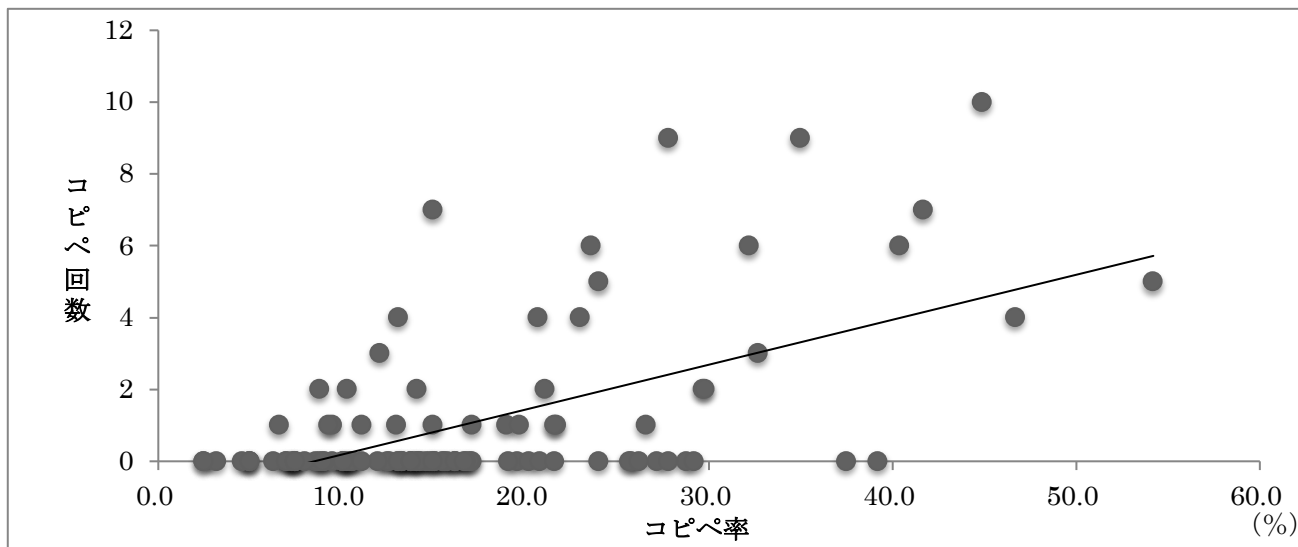


図3 履歴から見たコピー回数と剽窃チェックソフトによるコピー判定結果の関係

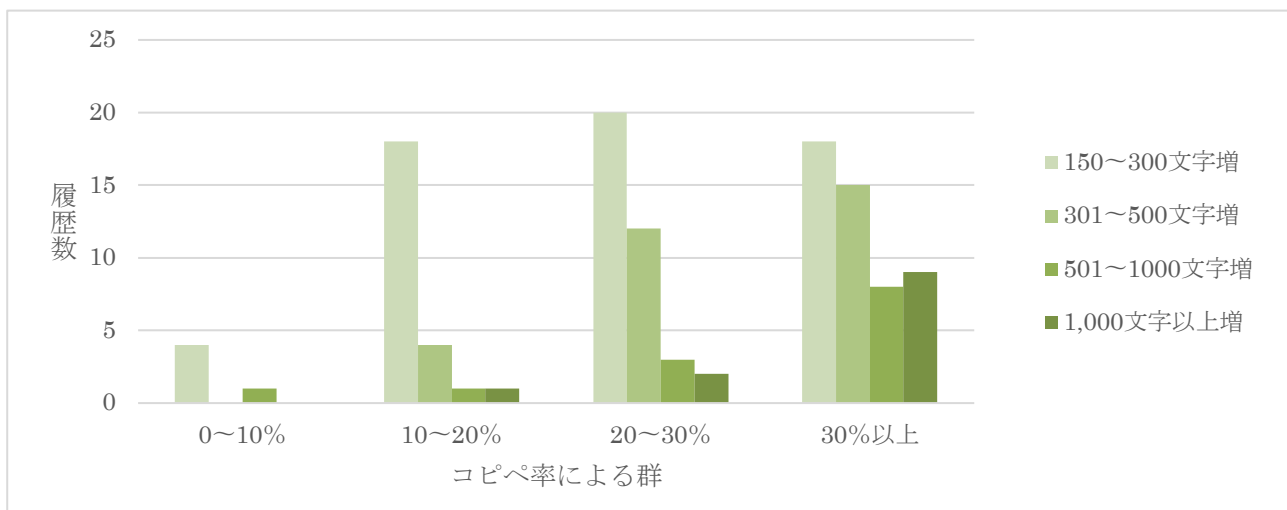


図4 コピー率によって分類したときの増加文字数別の履歴回数

5. 考察

(1) コピペルナーと変更履歴による剽窃判定の比較

同一レポートに対して剽窃チェックソフトによる剽窃判定（プロダクト分析）と変更履歴から見たコピー回数（プロセス分析）には、図3のように正の相関があった。両手法によるコピーレポートの検出結果が一致するのは、表3において、コピーを行ったと判定された「コピー率20%以上」、「150文字以上増加有」の19部のレポートと、コピーを行わなかったと判定された「コピー率20%未満」、「150文字以上増加無」の61部のレポートであり、全体の74.1%が該当した。

図4から20%以上のコピー率のレポートには、300文字以上の増加が確認される履歴が急激に増加している。これらの結果から、一度に300文字以上の文章を貼り付けるというプロセスを経て作成しているレポートは、多くの文章がコピー元と同じになり、コピペルナーによるコピーの検出が容易になったのではないかと考えられる。

次に、両手法の結果が一致しなかったレポートについて、検討する。まず、「コピー率20%以上」、「150文字以上増加なし」のレポートは13部であった。これは、コピペルナーでは検出できたが、履歴からの分析では検出できなかったと考えられるレポートである。該当するレポ

ートを変更履歴とコピペルナーの検出結果画面から再調査した結果、次のような要因が考えられた。

(1)対象となったレポートが発表用資料のため、箇条書きや図表が多くなって文字情報が減ることにより、結果として単語をコピペと認識したために相対的にコピペ率が高くなった。

(2)実際に、出典元と同じ文章だが、コピペは行わずに文章を見ながら自分で入力した、あるいは150字以内のコピペを繰り返して少しずつ作成した。

(1)に関しては、本来であればコピペと判定されないレポートであり、(2)の場合は、目視ではコピペと判定されるがプロセス分析においては150文字以上増加という条件では検出できないレポートである。

一方、「コピペ率20%未満」、「150文字以上増加有」のレポートは15部であった。これは、コピペルナーではコピペを検出できなかったが、履歴からの分析では検出できたレポートである。同様に、該当するレポートを変更履歴とコピペルナーの検出結果画面から再調査した結果、次のような要因が考えられた。

(1)コピペを行ったが、最終的なレポートには使用しなかった、または完成したレポートには一部だけしか使用されなかった。

(2)コピペをした後、文章の順序の入替や表現を一部だけ変更する事により、コピペの検出精度が下がり、コピペ率が減少した

(1)に関しては、本来であればコピペと判定されないレポートであり、(2)の場合は、目視ではコピペと判定されるがプロダクト分析において20%以上という条件ではコピペと判定されないレポートである。

このように、どちらかの手法でしかコピペ検出できないレポート、及び、本来であればコピペと判定されないレポートが検出されることが確認された。

(2) コピペルナーと編集履歴による剽窃検出の特徴

レポートの剽窃について、剽窃チェックソフトやシステムを使う目的は、一般的にコピペが疑われるレポートを効率的にピックアップすることである。その観点から考えると、背景でも述べたとおり、時間的な効率性は、コピペルナーによるプロダクト分析の方がすべての履歴を取得して調査する必要があるプロセス分析よりも有効である。

一方、履歴を利用したプロセス分析では、コピペが疑われるレポートの詳細な質的分析が可能であり、今回の分析には用いなかったが、コピペルナーでは検出されない図表や画像のコピペや、コピペした部分を部分的に加工して引用文献を偽った意図的な剽窃レポートも確認す

ることができる。

剽窃検出の感度については、コピペルナーによるプロダクト分析であればキーワードを自動設定にせず、レポートを通読してより適切なキーワードを手動で設定するか、コピペが疑われるコピペ率の基準を本研究で行った20%から下げると高くなる。一方、編集履歴によるプロセス分析であれば、履歴の増加文字数を150字から減らすことにより感度を上げることができる。この場合、両手法とも誤検出の比率が高くなり、確認が必要となるレポート数が増加し効率は低下する。

今回の条件では両手法でも約3/4のレポートについて一致した結果が得られたことから、単独でもある程度の検出精度があり、両手法を組み合わせることで互いの弱点を補ってコピペ剽窃の検出の精度をさらに高めることができると考えられる。

6. おわりに

Googleドキュメントを用いることにより、Wordなどの文章作成ソフトに匹敵する汎用性の高いクラウドサービスを用いることで変更履歴からレポートの作成過程にアプローチすることが可能となり、同一レポートに対して、変更履歴の分析と剽窃チェックソフトによる剽窃判定を行うことができた。その結果、両分析の関連及び特徴を明らかにできた。

Ryan(2009)は剽窃対策には予防 (prevention)、検出 (detection)、罰則 (penalty) のバランスが重要であると指摘している。剽窃レポートの予防について、検知システムを事前に学生に周知させること、検知に関しては異なるアプローチの2種類の検知手法を用いることにより対策の効果を高めることができる。

剽窃を含むレポートの評価・指導については、レポート作成中の編集情報も伴ったプロセス分析の方が、質的分析にも応用可能で有効性が高いと言える。個々のレポートを詳しく調査すれば、剽窃のチェックだけでなく、インターネットを用いたレポート作成におけるプロセスからの問題点の分析も可能となる。

本調査では56.5%のレポートが両手法で剽窃の対象外となった。しかし、単に両方に引っかからないことが必ずしも良いことではない。学生が非生産的なコピペレポートを作成しない、すなわち、インターネットの情報を十分に取捨選択し、それらを有機的に組み合わせる新たな知識体系を創造できるように促すシステム・教育についてさらに研究を積み重ねていく必要がある。

付記

本稿は宮本ほか(2013),(2015)が日本教育工学会研究会において発表した研究に加筆修正を加えてまとめたものである。

謝辞

本研究の一部は、文部科学省科学研究費助成事業（基盤研究(C)「大学生のレポート作成における情報リテラシー：剽窃行為の把握と対策」課題番号25330421）による助成により行われている。

文献

藤本貴之・川合博之・志村敦史 (2009) 「大学生のカンニング/不正行為の傾向分析とその抑止のシステムの提案」, 『情報処理学会研究報告』, 32, 17-22.

花川典子 (2013) 「コピペ対策の実践-コピペ検出システム」 関西地区FD連絡協議会京都大学高等教育研究開発推進センター編『思考し表現する学生を育てるライティング指導のヒント』 ミネルヴァ書房, pp.221-230.

近藤暁夫 (2014) 「地理学の卒業論文における学生のコピペ問題と対応-愛知大学地理学専攻の事例から-」 人文地理学会大会 ポスター発表.

(<http://taweb.aichi-u.ac.jp/geogr/2014jinbunchirigakkaiP03.pdf>)(2021年9月1日)

永井昌寛・中村隆義 (2010) 「作成編集履歴を用いたレポート評価支援システムの提案と開発課題」 『日本教育工学会論文誌』 34, 9-12.

Noriko Hanakawa, Masaki Obana(2012).A Plagiarism detection system for reports based on a large-scale distribution environment using idle computers, *proceeding of the 15th IASTED International Conference on Computers and Advanced Technology in Education* CATE 12-19.

光原弘幸 (2011) 「Web を情報源とするレポート作成のためのコピー・アンド・ペースト制限とリフレクション支援ボード入力能力と大学進学以前の情報教育の関連性について」 電子情報通信学会技術研究報告.

宮本淳・仙石昌也・山森孝彦・久留友紀子・橋本貴宏(2013) 「Googleドキュメントを利用したレポート作成過程の分析」 『日本教育工学会研究報告集』 13(4), 1-6.

宮本淳・仙石昌也・山森孝彦・久留友紀子・橋本貴宏(2015) 「プロダクトとプロセスからみた剽窃の分析」 『日本教育工学会研究報告集』 15(1), 21-26.

Ryan, G et al(2009) Undergraduate and Postgraduate Pharmacy Students' Perceptions of Plagiarism and Academic Honesty, *AMERICAN JOURNAL OF*

PHARMACEUTICAL EDUCATION, 73(6) 105

仙石昌也・宮本淳・小島貞男・山森孝彦 (2008) 「チュートリアル教育におけるレジメの構成と引用文献に関する調査と分析」 『愛知医科大学基礎科学紀要』 35, 1-9.

仙石昌也・宮本淳・小島貞男・山森孝彦 (2009) 「チュートリアル教育におけるレジメの構成と引用文献に関する調査と分析(2)」 『愛知医科大学基礎科学紀要』 36, 1-12.

仙石昌也・宮本淳・山森孝彦・久留友紀子・橋本貴宏 (2013) 「クラウドシステムのグループ学習への導入 -編集履歴から見る作業状況について-」 『愛知医科大学基礎科学紀要』 40, 1-8.

杉光一成(2010) 「大学等における「コピペ」問題の現状と対策及びその課題」 『PC Conference 論文集』 243-246.

山口榮作・永井昌寛・横山祐司 (2009) 「文書編集履歴を考慮したレポート評価支援システムの試作」 『大学情報システム環境研究』 12, 19-26.

米沢誠(2009) 「レポート作成におけるコピペ防止策 コピペを超えるライティング授業デザイン」 『情報管理』 52(5), 276-285.

SUMMARY

In our tutorial education for first-year medical students, the students were given a task of writing a report online, using Google Docs. In our previous studies, by tracing the work history obtained from the online report file, we analyzed their process of writing, especially focusing on the process of copy-and-paste plagiarism. Analyzing the change log of written reports enabled us to visualize the students' writing process. Sudden increase in total letter (or character) counts or other unnatural behavior demonstrated the possibility of plagiarism. In this study, each report was scanned with a plagiarism detection software, and its work history was analyzed by checking sudden increase in letter counts. As a result, a positive correlation was found between the amount of copy and paste, and copy-and-paste action frequencies. In addition, even when the product analysis did not detect plagiarism in a report, the process analysis was able to reveal hidden plagiarism.