

レポート剽窃検知支援ツールの出力の効果的な読み方に関する実践的研究

江口 建¹⁾, 小島 一晃²⁾

¹⁾豊田工業大学工学部

²⁾帝京大学ラーニングテクノロジー開発室

概要

レポート課題における剽窃は教育上の重大な問題であり、剽窃の有無を適切に把握して学生に指導を与える必要がある。文書同士を比較して類似性を算出することで剽窃検知を支援するツールが存在するが、剽窃の判定自体は教員に委ねられる。レポートの性質は教示や要件に依存するため、剽窃の判定もそれに依拠して変わる。本稿ではレポート課題の添削に剽窃検知ツールを用いた際、その出力をどのように読むかを、異なる教示・要件による課題において実践的に検討した。その結果、類似性の高さ自体は必ずしも剽窃の指標にならないこと、一定水準に到達するまで繰り返し提出させる課題では受講生自身の先行する答案と類似しないことで剽窃が検出されたこと、同一のアイデアを学生自身で説明する場合には表層的な説明構造までは一致しないことが判明した。また、レポートに不慣れであると、悪意がなくとも類似性が高くなる傾向があるため、適切な作法を教えるとともに、学生が犯しやすいミス把握に基づく指導が必要である。

1. はじめに

学生自身のアイデアや取り組みの成果を文書で報告するレポート課題における剽窃は、古くから教育における重大な問題のひとつである。ICTの発展と普及により、文書を見つけてそのコピーを作成するコストが大幅に低減されたため、剽窃の実行も容易になっている。しかし、技術的な要因が剽窃を促進するとしても、それは原因の一部である。WWW上で互いの情報を共有し合うサービスを日常のコミュニケーションに利用しているため、他者の情報をコピーすることが問題となる場合に認識が及んでいないという、学生の意識も一因と考えられる[1]。そのため、情報倫理・研究倫理教育の一環として、剽窃をしないよう学生に意識付けをさせるとともに、剽窃が行われた際には学生に指導を与える必要がある。

適切な指導を行うためには、学生が剽窃を行ったこと、ならびにその内容を把握する必要がある。剽窃検知を支援するための技術的な研究は以前から行われており、学生が提出した文書と類似するWebページを検索して剽窃の疑いがある箇所を特定する機能(例えば[2])はその一例である。このような機能を持つツールは商用製品として利用できるようになってきている。しかし、こうしたツールによって剽窃検知の手間は軽減されるものの、個々の学生の文書が剽窃に相当するか否かを判定するのは教員である。レポート課題では、何がどのような教示と要件の下に課されるかによって学生が作成するものの内容は大きく変わるが、その内容に依存して剽窃のされ方も変わるため、判定の考え方も異なってくる。一例として、初歩的なプログラムコードを書かせる課題では、剽窃でなくとも他者と偶然一致する可能性が高いため、単一ではなく複数のコードで判定するといったアプローチがなされている[3]。

上述の背景に基づき、本稿ではレポート課題において剽窃検知ツールを用いた際、その出力をどのように読むかを、異なる教示・要件による課

Empirical Study on How to Effectively Interpret Output from a Tool to Support Detection of Plagiarism in Reports

Takeru Eguchi

Undergraduate School of Engineering, Toyota Technological Institute

Kazuaki Kojima

Learning Technology Laboratory, Teikyo University

題において実践的に検討する。剽窃検知支援ツールには、Turnitin 社の Feedback Studio[4]を用いた。以降、2 節において Feedback Studio の剽窃検知支援について簡潔に説明したのち、これを用いたレポート課題の結果と出力結果の読み方を 3 節で検討する。

2. Feedback Studio の剽窃検知支援

Feedback Studio は文書教育支援ツールであり、学生にレポート課題の答案としての文書ファイルを提出させ、教員が採点・フィードバックを行うことができる。その機能の 1 つに、Web 文書、学術雑誌等のレポジトリ、Feedback Studio にこれまで提出された答案のレポジトリを用いて、学生の答案から他の文書と類似する文章を検出して類似性を算出する類似性レポートがある。この類似性を算出する方法の技術的な詳細は公開されていないが、文章内の単語列の一致に基づく表層的な類似を用いていると想定される。類似性レポートは、学生の答案において他文書と類似する箇所を示し、それらが占める割合を算出することで、教員が剽窃か否かをチェックする支援を行う。ただし、剽窃の判定自体はしない。

図 1 に、類似性レポートの画面の例を示す。画面左に学生の答案が表示されており、他文書と類似する箇所がハイライトされている。画面右に類似性が表示されており、この答案の文章のうち 80%が他文書と類似していることを示している。その下に類似文章を持つ文書とその類似性の一覧が表示され、ここをクリックすることで当該の文書を参照できる。

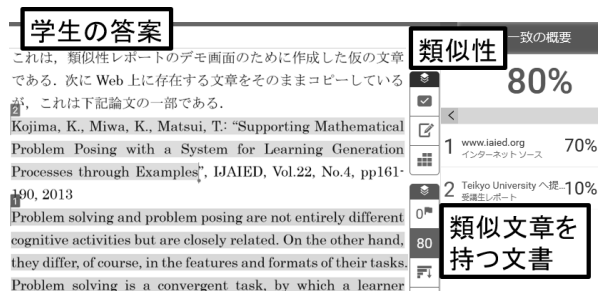


図 1 類似性レポートの画面の一部

他者の文章をそのままコピーした箇所が多い答案は類似性が高くなるため、教員が剽窃をチェックする判断基準となる。また、類似性レポートの存在があらかじめ学生に周知されると剽窃が抑止されることが期待されるが[1]、Turnitin 社も Feedback Studio の導入が剽窃を低減させたという報告を同社 Web サイトに掲載している。

3. 剽窃検知支援機能の出力結果とその読み方の検討

課題によっては、剽窃でなくとも他者のレポートと偶然に一致する可能性がある。そのため、当該のレポートが剽窃に相当するか否かを教員自身が判定する必要がある。ここで問題となるのは、判定の仕方である。どのような基準に基づいて判断するのが妥当であるかは、レポートのテーマ、課題文における指示の出し方、作成要件などによって変わってくる。以下では、いくつかの授業実践の例に基づき、剽窃検知ツールが出力した結果の読み方を検討する。

前提として、レポート課題を告知するときに、以下の内容を受講生に对面で伝えた。まず、インターネット全盛世代に見られるコピー・盗作の傾向について、実際にニュースで話題になった他大学の事例を挙げながら注意を喚起した。その上で、レポートを採点するにあたって剽窃検知支援ツールを使用することを伝え、剽窃検知支援ツールとはどのようなものか、使用するツールの性能などを説明した。あわせて、Turnitin 社の Web サイトを参照しながら[5]、「複写 (copying)」や「借用 (borrowing)」といった言葉を使うことで罪の深刻さが隠蔽されてしまう可能性があること、実際のところは「剽窃行為 (plagiarism)」とは明白に「盗み (to steal)」であり、「文書窃盗への加担 (to commit literary theft)」であり、「詐欺行為 (act of fraud)」であることを強調した。

また、不正レポートが発覚したときの取り扱いや学内の処罰規定についても説明した。さらに、匿名で書かれたフリーのオンライン百科事

典の使い方についても注意を喚起した。フリーのオンライン百科事典の特徴として、誰でも編集に携わることができ、記事の更新速度が早く、ジャンルが多様であり、多言語であるという利点がある。特に、ユーザー参加型の集合知（知の構築の共同作業）に基づいているため、学術的な権威者によって作成された特定分野の百科事典とは異なった利用法が期待できる。その一方で、情報の信憑性・公正性が保証されておらず、執筆者が匿名であるがゆえに記事の正誤に関する責任の所在が明確ではないという欠点がある。また、出版社（プロの編集者）の手が入っていないがゆえに日本語の文章作法に不安があるという難点もある。したがって、あくまでも参考程度に、取っ掛かりとして利用するのはよいが、自説を補強するためのエビデンスとしての能力は低い（それゆえ学術レポートの正式な参照・引用文献としては使用できない）ことを受講生に伝えた。

3.1 事例①：指定の素材を基に、参考文献を用いて客観的・批判的に考察させる課題

工学部の教養科目（前期開講／1年生対象）の学期末レポートとして、日本の大学生の学習意欲や基礎学力の低下を指摘する700字程度の文章（いわゆる学力低下論争）を読ませ、書かれている内容を自分なりの観点から批判的に考察し、それを踏まえた上で自分の実態について振り返り、大学での学びと自分の関係について自己分析せよ、という問題を出題した。字数は、2つの設問をあわせて1200字以上（上限なし）とし、Microsoft Wordで作成の上、オンラインで提出することとした。

【作成要件】

- ・ 参考文献として、紙媒体で出版されている書籍を必ず1冊以上用いること。
- ・ 感情的にではなく、論理的に主張を展開すること。

- ・ 主観的にではなく、客観的に記述すること。
- ・ 憶測や思い込みではなく、根拠（事実やデータ）に基づくこと。

以下では、2021年度と2022年度に実施した結果を提示する。

表 1 2021年度と2022年度に提出されたレポートの類似性の最小・最大・平均

年度	<i>n</i>	最小	最大	平均
2021	98	20%	59%	38.2% (7.5)
2022	100	22%	62%	36.6% (7.7)

※かっこ内は分散

Mann-Whitney の U 検定により 2021年度と2022年度の類似性を比較したところ、差は認められなかった($z=1.57$, *n.s.*)。また、両年度とも、明らかな不正行為と見なしうる剽窃行為は一つも発見できなかった。最大値のレポートであっても、インターネット上の記事および刊行物からの引用を拾ってしまっているがゆえの高い類似度であることが目視によって確認された。

この場合、着目すべき現象は、剽窃検知支援ツールを用いることによって剽窃を摘発できたというよりも、むしろ一定の類似性を示す数値が出ているにもかかわらず剽窃が発見されなかったということである。

ここから言えることは、以下のことである。

1. 「剽窃」ではなくとも（つまり引用の作法は守っていても）、ブロック引用の多用などで引用が多い場合は、必然的に類似度が高くなる。
2. 執筆の要件として参考文献を必ず用いるように指示した場合は、相対的に類似度は高くなる。

以上のことから、類似性の数値そのものは剽窃の指標にならないことが分かる。

3.2 事例②：思考実験を自作させ、それについて自分の思考を展開するという課題（抽象度・自由度が高い）

哲学・思想に関わる教養科目（後期開講／1年生対象）の学期末レポートとして、「道徳的ジレンマ」の構造を持つ思考実験をみずから創作し、哲学的・倫理的に分析せよ、という問題を出題した。字数については、自分の思考を展開するのに相応しい分量で書くように指示し、Microsoft Word で作成の上、オンラインで提出することとした。

【作成要件】

- ・ 少なくとも 1 つは思考実験を自分で作る。時事問題や社会問題、自分の過去の経験を基に構築してもよい。素材選びは各自に任せる。ただし、既存の思考実験をそのまま使うのは不可。思考実験を自作してはいても、「道徳的ジレンマ」の構造になっていないものは不可。
- ・ 授業中に取り上げた思考実験を「応用」しても構わない（そのまま使うのは不可）。
- ・ 必ず授業の内容と関係させること。授業の趣旨を踏まえていないものや、哲学的・倫理的に扱えていないものは不可。

出題の意図としては、正解が一つに決まる問題に回答するのではなく、みずから考えるに値する「問い」そのものを発見し、それについて筋道を立てて思考し、粘り強く分析する力がどれくらいあるかを確認することを目的としている。その上で、論理的・批判的に考える力、多面的に物事を見る力、問いを掘り下げる力、問題を創造する力、構想力、倫理的な問題に対する感性、文章で他人を説得する力などを確認している。

本稿では、比較材料として、2020 年度から 2023 年度にかけて実施した 4 年分の結果（表 2）を提示する。

剽窃行為の有無について言えば、事例①と同じく、どの年度においても明らかな剽窃と見なしうるレポートは発見されなかった。最大値のレポートであっても、インターネット辞書や省庁のサイ

トからの引用を拾った結果であることが確認できた。

また、参考文献を使いながら客観的なデータに基づいて論述することを強く推奨した事例①と比べて、自分の創造的思考を展開するという自由度の高い課題の場合、相対的に類似度が低くなることも数値の上で明らかになった。

表 2 2020 年度から 2023 年度の答案の平均と 2020 年度との比較の結果

年度	<i>n</i>	類似性	比較
2020	99	12.1% (4.7)	-
2021	97	29.7% (6.2)	11.72**
2022	92	29.6% (6.8)	11.43**
2023	102	28.9% (6.8)	11.77**

※かっこ内は分散、2020 年度との比較は U 検定の統計検定量、**は $p < .001$

しかしながら、上記の結果からは、一つの特徴的な現象が観察される。表を見れば一目瞭然のように、2020 年度だけ、類似性の値が極めて低い結果となっている。10%未満が多く見られ、20%を超えるものはほぼ 2σ の外であった（そのいずれも、ありふれた文言であるがゆえの偶然的な一致、または語句の定義のための学会 HP 等からの引用、省庁の HP 等からの引用であった。つまり、「引用」であることを明記した上での参照であり、結果として、レポート提出者 99 名のうち剽窃が疑われるレポートは一つもなかった）。

設問は全ての年度で共通であるにもかかわらず、2021 年度を境に、それ以降の年度の類似度が急に高くなっている。Mann-Whitney の U 検定により 2020 年度と 2021-2023 年度の各々を比較したところ、表 2 に示すようにいずれの間にも 1%水準での有意差が確認された。2021-2023 年度の二者間にはいずれも差が認められなかった。

このような現象が起こった理由は定かではない。2020 年度と言えば、人類史上稀にみる新型コロナウイルス感染流行 (COVID-19) が世界を席捲し、日本でも全国で一斉にオンライン授業が実施され

た年である。そのことが何らかの仕方に関係している可能性は皆無ではない。コロナ禍で、いわゆる「3密」(密閉・密集・密接)を回避するためにDXが推進され、教育現場でも宿題やレポート課題のオンライン提出が推奨された。このような文書のデジタル化がきっかけとなって、剽窃検知支援ツールのデータベースに蓄積される文書量が急増した可能性はあるだろうか。そのことを客観的に確かめる術はないが、もしも2020年のコロナ禍以降、インターネット上で参照可能な対象物が急激に増えたと仮定できるならば、そのことと2021年度を境に類似性の値が上がったこととの関連性を推測したくなるのは、不自然な頭の使い方ではなかろう。それ以外の要因として、2021年度にFeedback Studioの類似性算出のアルゴリズムが改良され、類似性検出の性能が向上した可能性なども想定できるが、現時点では推測の域を出ない。あくまでも可能性として提示しておく。

いずれにしても、この類似性検知の結果から、以下のことは言えそうである。

1. 単に類似性の値の増減だけを見ても、剽窃が増えたかどうかを判定する指標にはならない。
2. データベースに蓄積される参照可能な文書が増えれば増えるほど(比較対象の母数が増えることから)類似性の値は高くなる可能性がある。

3.3 事例③：指定の素材を基に、一定基準を満たす答案を作成するまで再提出させる課題

理工学部の専門科目(領域はデータベース)の学期末レポートとして、概念や操作を正しく理解できたかどうかを確認するために、授業で解いた問題のうち指定された7問の解法が適用可能で、指定の条件を満たす新しい問題を7問それぞれに対して作成させた。7問全てを1つの文書に記述させ、PDFかMicrosoft Wordで提出させた。

【作成要件】

- 指定の問題と同じ解法か、拡張したより複雑な解法を持つ問題を作成すること。
- 作成した問題の解法を解説すること。
- 解法に関わる名前に誤記や齟齬(例:「学生番号」と「学籍番号」が同一のものを指す)がないようにすること。
- 元の問題のテーマ(例:ショッピングサイトの注文管理)を変更すること、その際、授業に登場した問題と同一のテーマを使わないこと。

学生には、事前に上記の要件を記したループリクが提示され、これを確認して要件を満たす答案を作成するよう指示された。答案はループリクに基づき採点され、合格するまで何度でも再提出するよう求められた(この作問課題の詳細については[6]を参照されたい)。

課題の性質上、学生の答案には指定の問題と似た文章や語句が多く含まれる。たとえば指定の問題に「あるショッピングサイトの注文管理データベースのインスタンスである」という文章があるが、これに対し、ある学生が作成した問題は「あるトレーニングジムの顧客管理データベースのインスタンスである」という文章を含んでいた。そのため、どの学生の答案にも、一部の単語が異なるだけの同一文が出現する。つまり学生の答案間の類似性はある程度高くなったが、その一方で解の解説の文章については学生間でほとんど類似していないという傾向にあった。

表3に、初回から4回目に提出された答案の類似性の平均を示す。この課題においては、実施後にFeedback Studioより、Web文書、学術雑誌等のレポジトリ、他学生の答案全てと比較した際の答案の類似性(全類似性と呼ぶことにする)と、他学生の答案のみと比較した際の類似性(答案類似性)のデータを出力することが可能であったため、表にはこれらを分けて記載している。初回の一番初めに提出した学生の答案に関しては、比較可能な答案がまだ存在しなかったため、類似性は0%であったが、これを除外した40の答案の全類似性は46.6%、答案類似性は43.7%であつ

た。初回の答案が合格にならなかった学生は、修正した答案をさらに提出したが、表に示されるように 2 回目以降の類似性も初回提出とほぼ同様であった。答案のうち問題の記述は似通った文章を持つが、解の解説は全く異なっていたことから、答案類似性の平均が 40%程度になったものと思われる。

表 3 初回から 4 回目に提出された答案の全類似性と答案類似性の平均

提出回数	n	全類似性	答案類似性
初回	41	45.9% (10.7)	42.6% (12.3)
2 回目	35	46.0% (10.1)	42.7% (10.8)
3 回目	23	43.8% (10.4)	39.8% (11.1)
4 回目	12	45.9% (11.9)	41.4% (12.9)

※かつこ内は分散

この課題においては、他の学生の答案をコピーして一部を書き換えたとと思われるものが、4 回目の提出において観察された。この答案の場合、問題だけでなく解の解説もハイライトされていた。初回から 3 回目までの答案は似通っていたのに対して、3 回目と 4 回目の答案は大きく異なっていた。本人に事情を聴取したところ、合格となる問題を作る方法が分からなかったため、先に合格した友人に依頼して答案を見せてもらったことを認めた。ただし、3 回目の答案類似性が 38%であるのに対して 4 回目のコピーでは 54%と大きく変化しているものの、答案類似性が 60%を超えたり、前回より 10%以上変化したりしているケースは他にも複数観察されており、しかもそのいずれも剽窃とは見なせなかったことから、類似性の高さ自体が有効な手がかりとはなっていない。

著者の担当する他の授業でも、合格するまで再提出させるという課題を出しており、上記の例と同様の剽窃が 2 件観察されている。すなわち、複数回提出したのち、先に合格した学生の答案のコピーを使用したために、前回と答案が大きく異なっていたことが剽窃検出の手がかりとなったケースである。

3.4 事例④: グループでのアイデア立案を基に、個人で成果記録を作成させる課題 (自由度が高いがグループメンバー内で材料が共通)

経済学部の情報リテラシー科目の授業で、5 週にわたり 3~5 名の学生のグループで問題解決の演習を行わせた。各グループで自由にテーマを決めさせ、仮想のアイデア立案とその成果の口頭発表を行わせた上で、最終報告として個人単位でアイデア立案過程とアイデアを記述した文書を、Microsoft Word で作成して提出させた。

【作成要件】

- ・ 指定のフォーマットに従うこと。
- ・ テーマを表すタイトル、背景(テーマの選択理由と意義)、方法(アイデア立案の過程)、結果(アイデア)、考察を明確に説明すること。
- ・ 本文の内容と整合する要約を記述すること。
- ・ アイデア立案の過程で使用した資料は全て参考文献に記載すること。
- ・ 立案過程とアイデアはグループのメンバーと同じになってもよいが、背景と考察は自分自身の考えを述べること。

3.3 節と同様に、要件を記したルーブリックが学生に提示され、合格するまで何度でも再提出するよう求められた(詳細については[7]を参照されたい)。

出題の意図としては、問題解決を分析して記録し、将来参照可能な記録を作成するという、社会における問題解決で必要となるタスクを体験することにある。しかし、この意図を理解できず、アイデアの立案過程が具体的に説明されていなかったり、アイデアそのものを記載せず口頭発表の感想を記述したりする受講生が多い。そのため、記録と見なせる文書になるまで再提出を行わせた。

表 4 に、初回から 4 回目に提出された答案の全類似性と答案類似性の平均を示す。再提出を重ねても、類似性に変化はなかった。全類似性を高めた主な原因は、3.2 節の事例②と同様にありふれた文言の偶然の一致や、資料として用いた Web サイトのタイトル等の記載であった。

表 4 に示されるように答案類似性は低くなったが、同じグループのメンバーともほとんど類似しなかったことは予想外であった。各学生が一番初めに提出した答案を対象に、グループメンバー間の答案類似性のみを抽出したところ、平均は 4.1%であった。メンバーとの比較でハイライトされた箇所は、アイデアに登場するフレーズや参考文献の資料名といった、断片に留まっていた。方法・結果ともに各学生が自分自身の説明を作成したため、同じ内容を説明していても表層的な類似性は高くならなかったものと考えられる。

また、方法と結果が詳細に説明されておらず、メンバーと若干内容が異なる答案も見られた。グループでの演習にあまり貢献できなかつたり、演習の成果をメンバーと共有できなかつたりする学生がいたことも、類似性の低減につながった可能性がある。

表 4 初回から 4 回目に提出された答案の全類似性と答案類似性の平均

提出回数	n	全類似性	答案類似性
初回	37	33.0% (7.7)	7.3% (6.5)
2 回目	34	33.7% (7.6)	8.9% (6.7)
3 回目	26	32.8% (7.5)	6.9% (5.1)
4 回目	15	29.2% (6.1)	5.7% (4.0)

※かっこ内は分散

グループメンバー間の答案類似性の最大値は 30%で、方法・結果において先に合格したメンバーの 1 人と共通するフレーズが多く使われていた。ただし、同じフレーズがあっても文章単位で見ると一致はなく、説明の構成には違いが見られたことと、背景と考察にハイライトはなく独自の記述を行っていたことから、剽窃には相当しないものと判定した。このグループでは、別のメンバー間にも 19%の類似性が見られたが、漫才師の大会で優勝する方法というテーマを選択しており、対象の説明に長いフレーズ(例:○○とは、△△と□□が主催する日本一の若手漫才師を

決める大会)がいくつか見られた。さらに文書全体が A4 用紙 2 枚と短いため、フレーズの一致が類似性を高めたものと考えられる。

4. 考察

前節に記述した事例から、Feedback Studio の剽窃検知支援機能である類似性レポートを読む上で留意すべきは、主に次の 3 点である。まず 1 点目として、類似性の高さ自体は必ずしも剽窃の指標にならないことである。その原因のひとつは、事例①と③のように、参考文献や指定された素材の利用によって類似性が高まるためである。これらと比較すると、事例②のように自由度が高い課題では類似性がやや低くなる。しかしながら、同一課題であっても、事例②が示すように年度によって大きく類似度が変わるときがあり、剽窃の疑いが濃厚になる類似性の「閾値」のようなものは定められない。

2 点目として、一定水準に到達するまで繰り返し提出させる課題では、同一受講生の改善の具合は剽窃検知支援ツールでは可視化されないということである。すでに述べたように、繰り返し提出させる課題の剽窃では、自身の直前の答案と大きく異なるという例が観察された。「自身との類似性」が低い答案という言い方もできるが、一般に剽窃検知支援ツールは同一受講生間の比較を行わないため、この観点での剽窃検知は教員自身で行わなければならない。

3 点目として、アイデアが同じであっても、もしくは似ていても、それを学生自身で「説明」したのであれば、類似性は高くないことである。内容が似通っていたとしても、各々の学生が独立して文章を作成するのであれば、表層的な説明構造まで一致することはない。ただし無数の Web 文書等と比較するため、文言の平凡さ等の理由により多くの断片がハイライトされ、類似性がある程度高くなる場合がある。

さらに、1 年生対象の科目では、悪意のある剽窃とは違い、レポートの書き方に不慣れであるがゆえに生じたと推測できる不適切な引用が散見さ

れた。例えば、「引用／参照」と「剽窃」の違いをそもそも正確に理解していないケース(本人は、おそらく正しいと思って書いている、または自力で書いていると錯覚しているケース)や、丸ごとコピーすることを避けようとした結果、部分的な小さなコピーが蓄積され、むしろ総合的に一致率が上がってしまっているケースなどである。それらのレポートは、概して類似性の値が高くなるが、必ずしも悪質な印象を与えるケースではなく、引用や参照の作法を正しく教えることで改善が期待できるものである。類似性の値そのもの(数値の高低)とは関係なく、類似の一致結果の内実を採点者が目視で吟味することによって、初修者がどのような間違いを犯しやすいのか、受講生の傾向や不得意なスキルを見定めることが可能となる。それによって、より個別の実態に即した具体的で適切な指導を行うことができるように思われる。単純に悪意のある剽窃レポートを割り出すこと(不正レポートの受理の回避)よりも、年度ごとの受講生の引用の仕方のパターンや、昨今の学生が陥りやすい誤った参照の仕方、犯しやすいミスなどを把握することのほうが、教育的には生産性がある。悪意のない初心者に対する指導の最適化を目的とすることで、剽窃検知支援ツールは、より効果的な使い方ができるに違いない。

5. まとめ

本稿では、レポート課題の添削において剽窃検知ツールとして Feedback Studio の類似性レポートを用いた際、その出力をどのように読むかを、異なる教示・要件による課題において実践的に検討した。その結果、類似性は高くなる傾向があるものの、剽窃は見られず、各課題に応じた評価の必要性が確認された。将来の課題としては、前節で述べたように、類似性レポートから抽出される問題点のパターンの分析と、その改善のための指導法の検討が重要になると思われる。また、事例②が示すように、ある年度を境に類似度の急激な上昇が確認されたが、要因として、デ

ータベースに蓄積された参照可能な文書量が急増した可能性や、アルゴリズムの改良によりツールの類似性検出の性能が向上した可能性の他に、過年度の答案がすでに機関レポジトリにデータとして存在していることの影響がどれくらいあるのか、今後の調査研究が待たれる。

参考文献

- [1] 江口建, “剽窃検知支援ツール「Turnitin」の有効性の検証ならびに学生の「剽窃行為」から見えてくる情報教育の課題”, 帝京大学ラーニングテクノロジー開発室年報, Vol. 16, pp. 79-86, 2019
- [2] 高橋勇, 宮川勝年, 小高知宏, 白井治彦, 黒岩丈介, 小倉久和, “Web サイトからの剽窃レポート発見支援システム”, 電子情報通信学会論文誌, Vol. J90-D, No. 11, pp. 2989-2999, 2007
- [3] 新村正明, 五月女雄一, 國宗永佳, 不破泰, “レポート類似度の時系列解析に基づく学習者間の依存関係発見システムの開発”, 教育システム情報学会誌, Vol. 26, No. 1, pp. 59-67, 2009
- [4] Turnitin, LLC., Feedback Studio, <https://www.turnitin.com/products/feedback-studio/> (2024年4月1日アクセス)
- [5] Turnitin, “What is Plagiarism?”, Published May 18, 2017, <https://www.plagiarism.org/article/what-is-plagiarism> (2024年4月1日アクセス)
- [6] Kojima, K.: “Preliminary Study on Learning Assessment by Using Problem Posing in an Online Course as an Alternative Method of a Conventional Examination”, Workshop Proceedings of 28th International Conference on Computers in Education, pp. 93-98, 2020
- [7] 小島一晃, “LMS 上での文書内マークによるフィードバックが学生のレポート改善に与える効果の予備検討”, ラーニングテクノロジー開発室年報, Vol.15, pp. 73-80, 2018