

作文行動を測定・分析するための マトリックス型テキスト編集モデルの設計

山口 琢¹ 大場 みち子² 高橋 慈子³ 小林 龍生⁴

概要: 作文 (文章産出) に関するさまざまな研究や指導・学習支援の基盤とするために、ワープロなどによる作文過程を測定する技術を開発した。マトリックス型文章編集モデルと呼ぶモデルに基づいて、IT を使う作文行動をデータとして記録する。モデルを評価するために、それに基づいて編集操作を記録する「トピックライター」を使って講義を行い、データを分析して評価した。分析のために、前後して編集対象とした段落間の距離に基づく分析指標 Editing Operation Indicator (EOI) を提案した。受講生同士のピアレビュー後に文章を修正する編集について、EOI が高い編集の方が、文章の論理的構成の評点に有意な改善があると判断できた。作文行動を測定するデータモデルとして、マトリックス型文章編集モデルは有効と考える。

キーワード: 作文、行動、測定モデル、分析

A Matrix Model For Writing Activity Measurement And Analysis

TAKU YAMAGUCHI¹ MICHIKO OBA² SHIGEKO TAKAHASHI³ TATSUO KOBAYASHI⁴

Abstract: To support various studies, teaching and learning on writing (composition, text production), we developed technology to measure writing activities on word processors and other IT environment. In that technology, writing activities are recorded as data based on a model called a Matrix Model for Writing. To evaluate the model, we have developed "Topic Writer" which records editing operation based on the model, as a reference implementation. We analyzed the data and evaluated the model. To analyze the data, we proposed Editing Operation Indicator (EOI) which is calculated based on the distance between the text parts which are edited in succession. There was a significant improvement in the score for the logical composition of the texts which are edited with high EOI after peer review between the students. Thus the model is effective for measuring writing behavior.

Keywords: writing, text production, measurement model, analysis

1. はじめに

1.1 背景

近年、学習分析 (Learning Analytics) 研究の分析対象が、成績といった大きな単位からページめくり操作といった細かい単位へ移りつつある。この変化には、細粒度化と行動指向という2つの側面がある。

細粒度化とは、測定・分析・応用の対象がきめ細くなることを指す (Fig. 図 1)。学習分析は、より小さな粒度で

¹ フリー

Independent Researcher

² 公立はこだて未来大学システム情報科学部

Faculty of Systems Information Science, Future University
Hakodate

³ 株式会社ハーティネス

Heartiness Co., Ltd.

⁴ 有限会社スコレックス

Scholex Co., Ltd.

		分析・応用(の対象)の粒度		
		大	中	小
測定の粒度	大	科目の成績	ある科目の成績で、別の科目の成績を推定	講義への参画度 (engagement)
	小	ページめくり編集対象の枠	成績を予測	予習・復習・講義への参画度
				考え方 解き方
				考え方・解き方の指導

本研究のターゲット

図 1 測定と応用の粒度

Fig. 1 Granularity of measurement and application

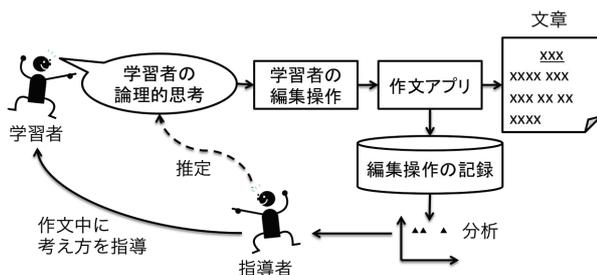


図 2 行動指向

Fig. 2 Focus on behavior rather than output

行動を測定して、より小さな粒度で効果を上げる方向に向かっている。例えば、作文アプリケーションの操作の特徴と、書かれた文章の誤字脱字や論理的構成といった細かい評価との関係を分析して、作文の指導・学習にきめ細かく役立てようとしている。

行動指向とは、従来の AI や検索の分野では「書かれたテキスト」を分析したのに対して、「そのテキストを書く過程 (process) ・ 行動 (behavior) ・ 行為 (activity)」に注目して測定・分析して活用することを指す (図 2)。

指導・学習環境に IT が導入されることで、前記のような細粒度の測定データを、アンケート分析などと比べて低コスト・低バイアスで得られる可能性が高まる。また、個別に行われた研究で得られたデータを横断的に分析する、データサイエンスの可能性も広がる。

2. 従来の研究と課題

しかし、ただデータが大量にあるだけでは、期待した分析結果を得られるとは限らない。作文の研究では、これまでパソコンでキー入力を記録して分析するキーストローク分析研究や [1]、タブレット端末でタッチペンによる手書き入力 (筆跡) を時系列データとして記録して分析するペンストローク分析研究 [2] が行われてきた。これらの研究では、授業や宿題への参画度、入力の熟練度やつまずき・停滞箇所を知ることができたものの、書き手の思考の中身により踏み込んだ分析が課題となっている。キーストローク分析では、それ以上は分からないとの報告もある [1]。知的行動の細粒度のデータから何を見いだせるか、まだ知見が少な

く、やってみなければ分からない。粒度の細かい学習分析研究の現状では、行動の傾向と結果の評価との関係に未知の部分が多い。作文であれば、行動つまり書き方と、書き手の思考内容あるいは書かれた文章の評価 - 例えば論理的構成 - との関係が、本研究の計画時点では不明であった。

また、学習環境がオンライン化することで、前記のようなデータを同時に大量に取得でき、またそれに基づいて大勢の学習者に同時に対応することができる。そのためには、測定・分析システムがコスト的にも性能的にもスケールラブルでなくてはならない。

3. 研究の目的

作文 (文章産出、文章作成) に関して、書き手の思考内容に踏み込んださまざまな研究・業務分析の基盤とするために、ワープロなどによる作文行動を測定する技術を開発する。具体的には、測定するデータを設計して有効性を評価する。

想定するデータの用途は次の通りである:

- 作文の指導・学習支援
- 指導・学習方法の評価・比較
- ワードプロセッサやアイデアプロセッサなど作文ツールの評価・比較
- 言語横断的、国際的、PC/モバイルといった機器横断的、手書き/キーボード入力といった UI 横断的、学生時代と社会人時代といったライフステージ横断的、障がい者も含めた評価・比較
- これらにおいて、企業内の業務やプライベートも含めて幅広い文章を対象とする

そこで、測定する技術の要件は次のとおりとなる:

- 行動指向のモデル:
IT を使って文章を編集する行動のモデルである。これに対して、従来は、文章のモデルといえば、章立てや組版といった静的なものだった。
- 指導・学習方法などから中立なモデル:
指導方法を比較するためには、特定の指導方法の考え方にのみ基づくのは好ましくない。
- スケールビリティ:
数人から数千、数万人規模の受講者を測定・分析できる低コストな技術であること。
- アクセシビリティ、国際化、相互運用性の高い要素技術の採用
- 機密保持・個人情報保護

4. 研究の方法

目的は測定データであるが、評価としてデータが書き手の思考内容に踏み込んだ分析の役に立つ可能性を示す。このために、なんらかの分析手法を提案し、適用例を示す。実際に測定されるデータから見いだせる傾向などは、当

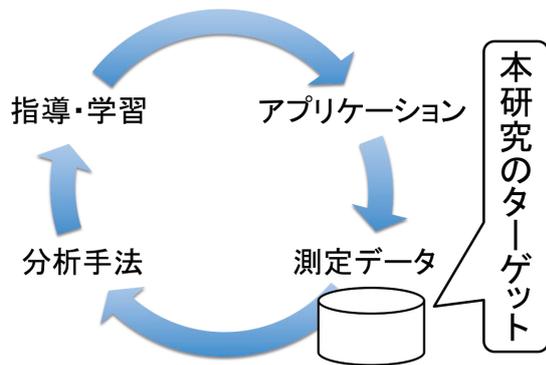


図 3 スパイラルなアプローチ
Fig. 3 Spiral approach

初は不明である。小規模なデータに対して、傾向を可視化する手法を試し、記録するデータやアプリケーションの機能へフィードバックすることを繰り返し、スパイラルに進める。次の 4.1 から 4.4 を小規模に繰り返す (図 3)。

本研究では、研究の計画、実験のデザイン、目論見、期待する成果などを、最初から公開し、研究会などで積極的に発表して議論する。

4.1 モデル設計

測定にあたっては、まず IT を使う作文行動 (writing activity) をモデル化し、そのモデルに基づいて編集操作を記録して測定データとする。

4.2 リファレンス実装

モデルに従って編集操作を測定する作文アプリケーションを開発するか、あるいは既存のアプリケーションに測定機能をアドオンする。

アプリケーションのメモリ所要量などを測定することで、モデルに基づく測定が実現可能であることを評価する。

4.3 分析手法の提案

測定データを分析する、なんらかの手法・指標の開発・提案する。書き手の意図や行動の意味を読み取れるような分析・可視化手法とする。

分析手法についても測定のモデル同様に、指導者などを想定したリファレンス実装を開発する。

4.4 適用と評価

テクニカルライティング演習などで 4.2 の作文アプリケーションを使って測定データを収集する実験をする。収集したデータを 4.3 の手法を使って分析・可視化する。

最後に結果を評価する。

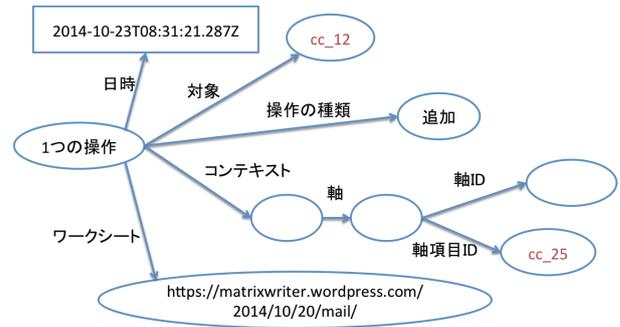


図 4 マトリックス型文章編集モデル
Fig. 4 Matrix type text editing model

5. 研究成果

5.1 モデル

開発したモデルをマトリックス型文章編集モデルと呼ぶ (図 4)。このモデルでは、段落といった文章の編集対象の部分と、いつ編集したかの時刻と、編集操作の種類と、その部分に関連づけられた他の文章部分や (文章には含まれないが) 執筆指針など (コンテキスト) と合わせて、機械的に記録する。コンテキストは複数の軸として表現できる。個々の編集操作を、複数の軸と関連付けてマトリックス型に整理するのでマトリックス型と呼ぶ。典型的には業務帳票などのワークシートを想像すると近い。ワークシートの物理的なレイアウトを捨象して論理的な構造を記録する。例えば、考慮する観点をコンテキストの軸として、会議に参加するために出張申請書を作る場面を想定すると、「理由」欄に記入するときには、オンライン会議ではダメなのかといった観点 (軸) や、出張の効果といった観点 (軸) が考えられる。文章のある部分を書くときには、こういったことを考慮しなさいという、作文指導や業務指導が行われる。これを抽象化・モデル化して作文行動を記録する。

編集操作の種類としては、ツールは PC やスマホのアプリとして動作することから、アプリケーション一般が通常持っている追加・削除・コピーといった操作をモデルに採用することで、モデルの適用範囲を広げた。

編集対象の文章そのものを記録しないことで、機密保持・個人情報保護に寄与する。

5.2 リファレンス実装

モデルに基づいて編集操作を記録する作文アプリケーション「Topic Writer」を新規に開発した。テクニカルライティングの授業で、ロジック・ツリーを教える演習で使った。演習に対応できるアプリケーションを新規に開発し、既存のアプリケーションに測定機能をアドオンする方式は取らなかった。

「Topic Writer」はクラウド型の Web アプリケーションで、スケーラブルなシステムである。Web の技術を採用す

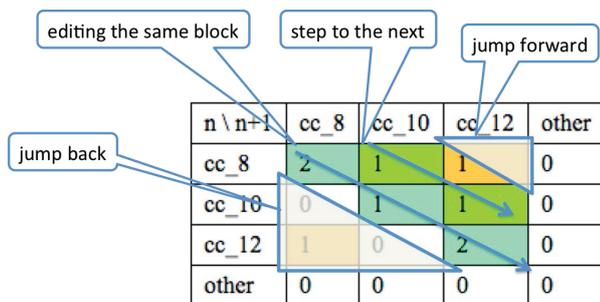


図 5 編集操作の時間的共起行列

Fig. 5 Collocation matrix for successive editing operations

ることで、アクセシビリティ・国際化・相互運用性を高めている。

評価として、開発したアプリケーションのクライアント側 (Web ブラウザ上) のメモリ所要量を測定した。作文行動を測定する機能のメモリ所要量が許容範囲内であることを確認した [3]。

また、情報系の学生でなくても、演習で導入して、その演習の時間内ですぐに使いこなすことができた。

5.3 分析手法

図 5 は、段落などが編集される時間的な前後関係に着目して、編集対象になった回数を集計した表である。n 回目 (縦軸) に編集した記入欄の次 (n+1 回目、横軸) にどの記入欄を編集したかの遷移数を行列で集計した。われわれは、これを編集操作の時間的共起行列と呼ぶ。時間的に引き続いて編集対象となる枠同士は、書き手にとって関係が深い可能性が高いと考えられる。この原理は、テキスト・マイニングで行われる、文字列や単語の共起分析と同様である。

図 5 で、cc_8 等は枠の ID である。対角線および 1 つ右のセルに集計された遷移は、単に左から右に書いたことを反映している可能性が高い。それ以外のセルに集計された遷移は、前後の枠間に特別な関係を書き手が感じている可能性が高い。

編集操作の時間的共起行列から、前後して編集対象となった段落間の遠さを積算した指標 Editing Operation Indicator (EOI) を提案した。ある段落を編集して、次に編集する段落が遠くにあるほど、EOI の値は大きくなる。書き手が文章中の 2 つの部分の整合性を取りながら作文しているとすれば、一方を編集した直後に他方を編集することが多くなるだろう。文章の中で広域の整合性を取ろうとしているならば、このような 2 つの部分の部分が遠く離れる可能性が高いとの仮説に基づいて、EOI は設計されている。

以上の可視化手法のリファレンス実装として「Writing Analytics」を開発した。「Writing Analytics」はクラウド型の Web アプリケーションである。Web の技術を採用することで、アクセシビリティ・国際化・相互運用性を高め

	全員	初期評価と EOI1	修正後評価と EOI2	改善度と EOI2
構成	①定義	0.1	0.4	0.4
	②具体的な説明	0.0	0.2	0.1
	③メリットの記述	0.1	-0.1	-0.1
表現	①読み手に合わせた表現	0.0	0.0	-0.3
	②一文一義	0.1	0.2	0.1
	③誤字脱字表現の統一	-0.1	0.1	0.0

図 6 EOI と論理的構成の相関

Fig. 6 Relationship between EOI and logical composition of the text

ている。

「Writing Analytics」は、講義・演習の講師が、特に指導を受けることなく、受講生の傾向を見るために使うことができた。

5.4 適用と評価

2 つの大学・大学院でのライティング演習でリファレンス実装を使って受講生の作文行動を測定した。ロジック・ツリーを使う作文で「Topic Writer」を使い、測定データを「Writing Analysis」で分析・可視化した [6]。

受講生同士のピアレビュー後に文章を修正する編集について、EOI が高い編集の方が、文章の論理的構成の評点に有意な改善がみられた (図 6)[4][5]。

本方式では、データの記録・可視化はコンピュータによって機械的に行われ、大規模化でき、リアルタイムに分析・可視化できる。また、「段落間に『根拠』という関係がある」といった意味付けから中立である。

5.5 まとめ

5.4 の図 6 は、作文行動の傾向と文章の論理的構成の評点との関係を示す。文章の書き方と「文章の構成を考える」という書き手の思考内容との関係を示すことができた。

5.3 に提案した分析手法と 5.4 の適用結果によって、本研究で開発したマトリックス型文章編集モデルが、書き手の思考内容に踏み込んだ分析の役に立つ可能性を示せた。これによって、本研究の目的を達成できた。これは、文章の論理的構成の指導や学習に役立つ可能性を示すものである。

5.6 今後

今後は、本方式による測定データを分析・可視化して、指導・学習に役立てる方法を明らかにしていく。このとき、他の分析手法や可視化手法を開発も試みる。

また、「Topic Writer」以外の作文アプリケーションも開発して、データを取得し、分析手法を開発して指導・学習に役立てる方法を検討する。すでに、文でジグソーパズルを行うアプリケーション「ジグソー・テキスト」を提案し、それによる編集過程の測定・分析について考察して、研究会で発表し議論し、またテクニカル・ライティングの授業で使う予定である [8]。

謝辞 本論文は科研費 (JP17K01085) の助成を受けたものである。

参考文献

- [1] LEIJTEN, M., VAN WAES, L. "Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes". *Written Communication*, 30(3): 358 - 392. (2013)
- [2] 飯山将晃、中塚智尋、森村吉貴、橋本敦史、村上正行、美濃導彦、ペンストロークの時間間隔を用いた解答停滞箇所の検出、*教育システム情報学会誌* 34(2), 166-171, 2017
- [3] 山口琢、高橋慈子、小林龍生、大場みち子、作文行動を測定・分析するためのマトリックス型テキスト編集モデルの設計、*研究報告コンピュータと教育 (CE)*、2017-03-11、津田塾大学小平キャンパス (東京都小平市)
- [4] 大場みち子、山口琢、高橋慈子、小林龍生、文章作成における文章評価と編集操作との関係分析、*研究報告コンピュータと教育 (CE)*、2016-12-03、長崎県立大学シーボルト校 (長崎県西彼杵郡長与町)
- [5] Oba Michiko, Yamaguchi Taku, Takahashi Shigeko, Kobayashi Tatsu, Analysis of Relationship Between Text Editing process and Evaluation of Written Text in Logical Writing, *IPSI-CE 141 International Conference*, 2017
- [6] 高橋慈子、大場みち子、山口琢、小林龍生、テクニカルライティング教育におけるクラウドツール活用と相互レビューの効果分析、*研究報告ドキュメントコミュニケーション (DC)*、2017-03-10、東洋大学白山キャンパス (東京都文京区)
- [7] 山口琢、高橋慈子、小林龍生、大場みち子、文章を構成する過程を測定するジグソー・テキストの開発、*研究報告コンピュータと教育 (CE)*、2016-12-03、長崎県立大学シーボルト校 (長崎県西彼杵郡長与町)
- [8] 山口琢、高橋慈子、小林龍生、大場みち子、文章編集行動の測定・分析手法の考察：手順書とジグソー・テキスト、*研究報告ドキュメントコミュニケーション (DC)*、2017-03-10、東洋大学白山キャンパス (東京都文京区)