

マトリックス型テキスト編集モデルによる 編集操作ログの実際と分析方針

山口琢¹ 大場みち子² 高橋修²

概要：作文学習・指導や文書レビューの手がかりとすることを目標に、われわれは、作文をモニタする仕組みを研究している。これまでに、マトリックス型編集モデルによるテキスト・エディタを試作し、解釈可能な編集操作ログを得ることができている。今回、編集操作の種類と時刻を軸にとった散布図を、自動作成する仕組みをシステムに付加した。この散布図によって、特定の書き手の複数の作文について、編集操作ログを比較した。その結果、ゼロから書き始めてシステム上で完結させられる作文では、完了よりも手前でマトリックス編集操作がなくなるといったパターンが見られた。そこで、複数の書き手による短い作文でも、本システム上で完結させることができれば同様のパターンが見られるという、次のサイクルの仮説を得た。そのために、マトリックス編集操作の分布・頻度を強調して可視化する仕組みの開発が必要である。次の段階では、その可視化の仕組みを試作して、複数人に対する作文指導実験を行う。

Matrix type model for text editing: Editing Logs and Analysis

Abstract: We research and developing a monitoring system for teaching writing. We have developed a text editing system by Matrix Type Model. We can get editing log which can be interpreted in some way. This time, we added a mechanism of automatic scatterplot creation which took to the time axis and type of editing operation. And we compare these scatterplots for some writing. As a result, we have got an idea that it will be helpful to emphasize the frequency and distribution of matrix editing operation.

1. 背景

MOOC's など、オンラインでのデジタル教育が成長してきた。このオンラインでの学習活動をとらえて、教育分野にもビッグデータ分析を適用しようという動きがある。オンライン教育の主要システムベンダーなどが参画して、このための標準化が進んでいる。例えば IMS Global Learning Consortium では、このような標準仕様で構成する枠組みを Caliper と呼んでいる [1]。

標準化の目的は、異なる複数の学習システムをまたがって学習活動を測定すること (measurement) である。これによって単に大規模なだけでなく、指導方法、言語や国をまたがり、障がい者も含めたデータ分析 (analytics) を実現

できる。それによって、学習者の成長が比較できたり、指導方法を比較できると期待されている。また、システムの UI や機能の優劣を比較できると、われわれは期待している。標準化の対象は、学習活動の情報モデル (information model) である。そこで検討されているのは、個々の学習システムの下に横たわるモデルであって、個別の学習システムの UI などではない。本稿の検討対象も、同様にモデルであって、システムの UI などではないことに留意されたい。

われわれは、作文指導の手がかりとすることを目標に、作文活動をモニタする仕組みを研究してきた。Caliper の学習活動ジャンルには、読解 (reading) は含まれているが、作文 (writing) は含まれていない。そこで標準化に対応可能な、作文活動の情報モデルを提案したい。

¹ 公立はこだて未来大学大学院
Graduate School of Systems Information Science, Future University Hakodate

² 公立はこだて未来大学
Future University Hakodate

2. 先行する取組

2.1 標準化による読解の測定

読解指導の研究はさまざまに行われてきた。それに対する学習者支援システムは、学習者の深い理解をとらえようとしている。

それに比べて、読解に対する Caliper の活動測定項目は、annotations, page/block use, media use, lookup というシンプルなものである。これらは、現行の市販製品の機能になぞらえると、ハイライト、ページめくり/図表拡大、動画再生、検索などである。既に、製品で広く実装されている。標準化によって、製品をまたがってデータを収集することができ、「〇〇の講義で、よくハイライトされる箇所」などを分析することができる。

別の観点からは、これら活動項目は、文章データの構成要素と比べると、粒度が大きい・粗い、上位レベルのものである。デジタル読書のための標準データフォーマットである EPUB は、文章のデータフォーマットとして HTML5 を採用している。HTML5 には、ページよりも小さな単位、例えば段落や見出しが含まれる。読者が段落や、段落中の単語を追う活動をとらえられれば、読解指導の役に立つと考えられる。現行の汎用製品には、それらに原理的に該当する機能はない。

活動項目は、拡張可能である。標準化されるのは、基本的かつ幅広いシステム間で共有可能な項目と、それらを拡張するための仕組みであると考えられる。

2.2 作文の測定

Caliper には存在しないが、作文を機械的に測定して分析する研究は、古くからある。館野らの最近の研究 [3] は、作文を機械的に記録 (測定) し、再生して人が分析するものである。

InputLog システムによる LEIJTEN らの研究は、書き手のキーストローク (活動) を測定して、機械的に分析するものである [4]。「や」「ま」「ぐ」「ち」「た」「く」といったキー操作の一つ一つを測定するもので、原理的には、幅広いワードプロセッサやテキストエディタに適用可能である。前に述べた、Caliper における読解に比べると、測定の粒度が細かい。キーストローク分析における作文の情報モデルは、作文行動をキー操作の連続ととらえるものである。文章が2段組であることや、章節構造の深さや、論旨の破綻などは捨象されている。

LEIJTEN らによれば、これらの分析によって、優れた書き手と初学者とでは中断 (pause) のパターンが異なることが分かった。しかし、何をしようとしてこれらキー入力しようとしているのか、例えば「氏名を入力しようとしている」ことは分らない。

2.3 まとめ

大規模で広範囲に学習活動を測定するためには、機械的に測定・分析できることと、標準化が必要である。採用される標準となるためには、既存のシステムで実装済みであるか、あるいは紙ベースの運用に原理的に含まれるものを、情報モデルとする必要がある。キーストロークは、そのような情報モデルのひとつである。しかし、キーストローク分析では「氏名を入力しようとしている」といったことが分らない。

われわれは、もっと粒度の大きな測定を目指したい。読解におけるハイライトや検索といった粒度のものである。ただし、それが何であるのか、具体的なニーズや要求仕様が明確化されているわけではない。

3. 着想: マトリックス型モデル

マトリックス型テキスト編集モデルとは、帳票、フォーム、作文ワークシートなどを抽象化したものである。作文指導や企業の業務で昔から広く使われさまざまなものが考案されている。用途に応じて、さまざまな UI が考案され、使われている。それら UI は、紙のワークシートであったり [6]、申請のためのワープロテンプレートであったりする。

ワークシートは、記入欄が設けられているので、何を書こうとして入力したのかが明確である。例えば「氏名」欄に「北海道函館市」と入力されれば、それは「氏名」を書こうとしたと見なすことができる。キーストローク分析や、形態素解析による分析では、これは分からず、むしろ「住所」を書こうとしたと判定されるだろう。行動分析の観点からは、「氏名」欄を何度も書き直す傾向のある群と、いちど記入して直すことのない群があることが分かったりするだろう。この違いは、人によるのかもしれないし、システムの UI によるのかもしれない。

「住所」「氏名」であればワークシートは1次元である。他方で、就職活動における「自己紹介書」などでは、「得意科目」など自分の特長と、「職種」など相手企業の特徴を併せて考慮して書くように求められ、多次元となる。そのため、これらを抽象化したモデルはマトリックス構造を持ち、アウトラインプロセッサなどに見られるツリーとは構造が異なる。

4. アプローチ

われわれはマトリックス型のテキスト編集モデルを出発点に、作文活動を測定するための情報モデルを開発する。そのために、先行研究の検討をふまえて、清水らが指摘する「テクノロジー・プッシュ」のアプローチをとる [5]。本研究全体では、小さな目標・仮説を立て、新しいテクノロジーを開発し、目標・仮説を修正または破棄して次のサイクルに進む。次の仮説を得ることが、各サイクルの目標で

抄録(SUMMARY) x Call for Paper マトリックス		
	Call for Paper ⇒	
論文の構成 ⇒	各自が端末を ⇒	学習履歴の活用 ⇒
背景 ⇒	1人1台端末の時代をふまえて、 ⇒	
研究の動機 ⇒		
目的と研究課題 ⇒		
研究方法 ⇒		
結果 ⇒	実行・運用環境はクロス・プラットフォームに対応し、データ形式はWeb標準形式を採用した。 ⇒	
知見 ⇒		
先行研究 ⇒		
考察 ⇒		作文状況を大規模に長期に観察できるようになり。 ⇒

われわれは、マトリックス型のテキスト編集モデルを提案する。テキスト編集システムにマトリックス型モデルを採用することによって、作文状況を大規模に長期に観察できるようになり、内省や指導方法の改善に役立つと考えている。マトリックス型の編集モードは、利用者による編集操作の意図を推測する手がかりとなる。通常の文章レイアウトのモードと相互にリンクさせることができ、それによって作文プロセスの試行錯誤に対応できると考えている。1人1台端末の時代をふまえて、**実行・運用環境はクロス・プラットフォームに対応し、データ形式はWeb標準形式を採用した。**

現段階での評価として、モデルが実際のシステムとして実装できることを確認した。プロトタイプ・システムを開発して、研究室で使用し、本論文もそれで書かれている。編集操作ログから編集操作の意図が読み取れること、アプリケーション・システムとして許容される性能であることを確認した。

テキスト編集、ワードプロセッシング、作文、ソフトウェアプロトタイピング、可視化

1. はじめに

図 1 マトリックス型モデルによるテキスト・エディタ
Fig. 1 Example of a text editor adopting matrix model

ある。このような進め方を、事業開発分野ではリーン開発などとも呼び、システム開発の手法としてアジャイルなどが用いられる。

4.1 ここまでの成果

われわれの最初の目標は、マトリックス型モデルが、アプリケーションとして実装可能であり、特定の書き手が特定の文章を実際にかくことができ、そのときの特定の作文行動記録から何らかの解釈が得られることであった。これらは、確認することができた。

4.1.1 マトリックス型テキストエディタ

われわれはこれまでに、マトリックス型編集モデルによるテキスト・エディタを試作した [2]。試作システムは、「マトリックス型編集モデルによって作文活動を測定する」というコンセプトを実証する (proof of concept, POC) ためのものである。使い勝手は最低限のものであり、実際、使いこなせるのはコンセプトを理解している研究チームのメンバーに限られる。システムは、マトリックス型編集モデルという情報モデルを理解するための参照実装 (reference implementation) でもある。試作システムは、実際に投稿論文を執筆に使うことができ、Web アプリケーションとして許容される性能で動作する。システムが出力する編集操作ログが、何らかの解釈が可能なのであることを確認した。ここまでは、最初の POC の範囲であった。

図 1 で、左側の表が表示されている領域がマトリックス領域、右側がプレビュー領域である。マトリックス領域は、帳票・フォーム・ワークシートなどに該当し、これらを抽象化したものである。プレビュー領域の文章が、作文のアウトプットとしての文章であり、印刷や Web ブラウザーでの表示など、読者が読むときの見え方に対応している。この文章で書き手は、「各自が端末を所有するように

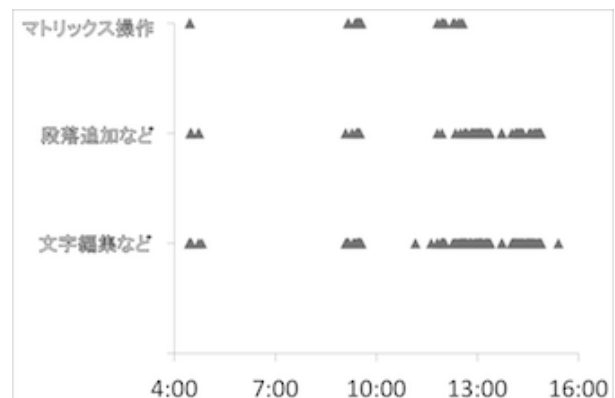


図 2 編集操作ログ
Fig. 2 Text editing log

なってきた」とことと「学習履歴の活用」という2つの視点から、論点を整理しようとしている。それらを、「背景」、「研究の動機」といった一般的な論文の構成にあてはめて、文章を整理しようとしている。

帳票と異なり、作文指導のワークシートでは、フォームに入力して終わりではなく、フォームの内容と対応付ながら実際の文章を作成していく。「実際の文章」は右側のプレビュー領域に表示され、マトリックス領域のテキストと対応づけられている。マトリックスの中で、枠線がハイライトされている部分が、現在、選択されているテキスト、右側のプレビューで白黒反転表示されている部分が、アウトプット中でそれに対応するテキストである。このように、アウトプットの文章に採用されているマトリックス側のテキストは、末尾に「⇒」マークが付いている。「⇒」マークが付いていないテキストは、アウトプットの文章に採用されていない。

4.1.2 ログの解釈

編集操作ログから解釈を導き出すために、手作業でグラ

フ化を行った。まず、システムからログを CSV 形式で取り出し、加工して、スプレッド・シート・アプリケーションに読み込んで、さまざまなグラフを試すという作業を繰り返した。

図 2 は、プロトタイプを使ってある研究会へ投稿したときのログを、編集操作の種類で大分類してプロットしたものである。▲記号が、1つの編集操作に該当する。「マトリックス操作」と分類した操作は、図 1 で示した、マトリックスに対する編集操作である。図 2 から、この書き手は、編集の最後の頃には、マトリックス操作を行っていないことが分かる。これは、もっともなことだろう、✓切り間際になって、論点を変えろといった大胆な見直しはなかなか行わないだろう。

特定の書き手による特定の文章について、「✓切り間際になって、論点を変えろといった大胆な見直しを行わなかった」という、妥当な解釈をすることができた。

ここまでで、研究・開発サイクルの第 1 段階に達することができたので、次の段階に進む。

5. 今回の目標と課題

今回の目標は、前回までの成果を元に、測定対象とする特定の作文活動のシナリオと、その分析で着目するポイントを具体化することである。このために、特定の書き手による「複数」の作文について、編集操作ログを可視化して比較する必要がある。

そのためには、分析グラフを、前回で得た解釈をより強調するものとし、前回までの手作業を脱して、グラフ化を自動化することが課題となる。編集操作の種類を縦軸に、編集時刻を横軸にとった散布図を自動表示する仕組みを開発する。そして、さまざまな作文に対する編集操作ログを見比べて、特徴的な点をリストアップする。

検討対象とする作文については、試作システムをよく分かっている書き手によるものを採用する。次の段階以降いづれかの段階で、さまざまな書き手による、さまざまな作文を対象にすべきである。今回は、違うパターンが見いだせるのであれば、よしとする。

6. 結果

3通りの、典型的な編集操作ログのパターンを得られた。

グラフの見方は図 3 の通りである。縦軸が編集操作の種類、横軸が時間経過に対応する。編集操作の種類には、今回は、内部的な実装上のメソッド名がそのまま表示されているが、上の赤や黄色にプロットされているのが段落や見出しの並べ替えなどの操作、緑にプロットされているのが図 1 左側のマトリックスに対する操作、紫やすみれ色にプロットされているのが通常の段落や文字列の編集操作である。

中段のマトリックス操作に着目すると、パターンが見えて

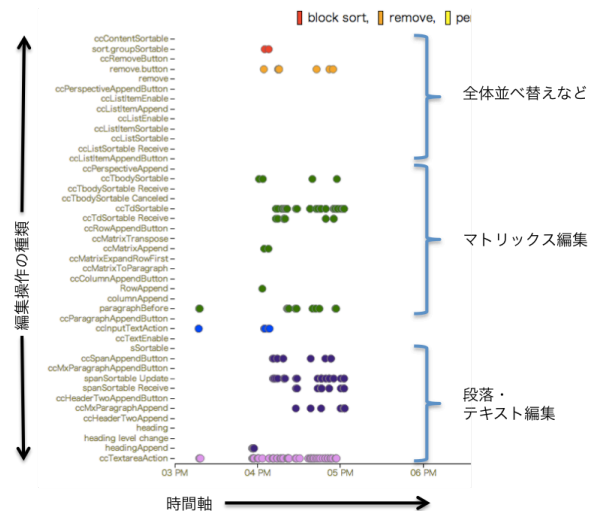


図 3 グラフの見方

Fig. 3 Legend

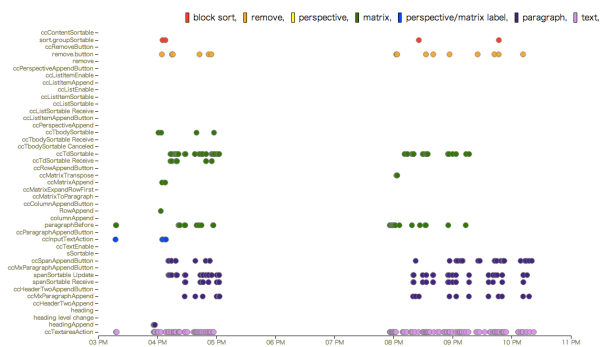


図 4 編集操作パターン#1

Fig. 4 Editing pattern #1

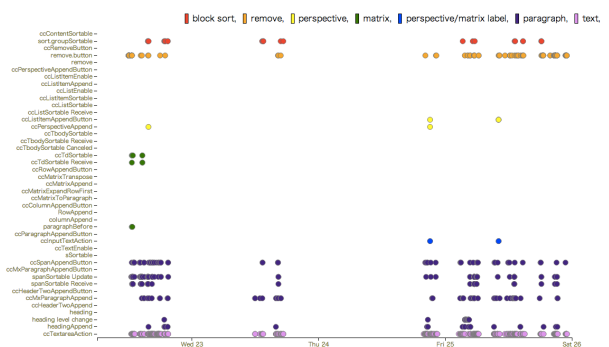


図 5 編集操作パターン#2

Fig. 5 Editing pattern #2

くる。

図 4 のパターン#1 では、編集完了、すなわち投稿時点よりも前に、中段のマトリックス操作が止んでいる。

図 5 のパターン#2 では、マトリックス操作が、かなり早い段階で止んでいるか、あるいは全く存在しない。

図 6 のパターン#3 は、マトリックス編集の現れ方が、#1 や#2 のパターンにあてはまらないもの全体である。

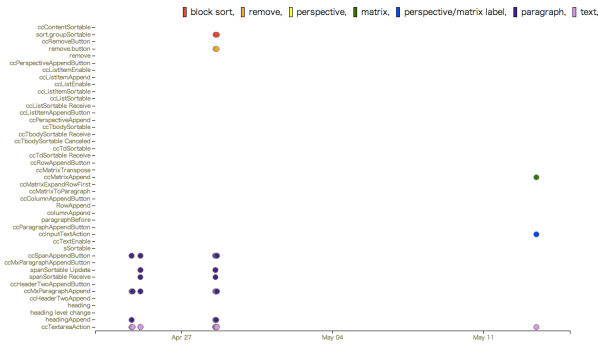


図 6 編集操作パターン#3

Fig. 6 Editing pattern #3

7. 考察と結論

マトリックス編集の分布に特徴が見られる。これらの作文がどのようなものであったか、書き手は思い出すことができる。

パターン#2は、既存の文章を修正する編集をしている場合に見られるパターンである。すなわち、内容を大きく組み替えるなどの編集作業を行っていない。

パターン#3は、このシステムでは実現できないレイアウト(2段組の1段目に脚注を入れる、など)が要求されていて、途中でこのシステム上での編集を取りやめた、などの場合である。これは、将来的には、エディタ操作の習熟度などを勘案するときの参考になると考えられる。

以上を考慮すると、次回のサイクルでは、分析においてはマトリックス編集の分布・頻度に着目し、POCシステム上で最後まで書き上げられる文章内容を想定したシナリオとすればよい。これによって、複数の書き手による場合に#1のようなパターンが見られるかどうかを確認する。このサイクルでは、論文執筆で必要となる図の挿入や章節単位の移動機能は不要である。

8. 今後

今回は、まず、マトリックス編集の分布・頻度を強調した分析表示機能を開発する。それを、複数の書き手(学習者)に対する作文指導の講義に導入して実験を行う。複数の書き手から#1のパターンが見られれば、その次のサイクルでは、これを指導に活用する仕組みを考案して実験する。

参考文献

- [1] IMS Global Learning Consortium, "Caliper Analytics", <http://www.imsglobal.org/caliper/>, 2014年07月16日閲覧
- [2] "マトリックス型モデルによるテキストエディターと作文過程の可視化", 日本教育工学会第29回全国大会講演論文集, 19-22 (2013)
- [3] 館野泰一, 大川内隆朗, 平野 智紀 "レポート執筆プロセスの可視化システム「レポレコ」の開発: チューターによる正課課程外の指導場面に着目して". 日本教育工学会論文

誌 37(3): 241-254 (2013)

- [4] 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)
- [5] 清水康敬, 中山実, 向後千春 "教育工学研究の方法". ミネルヴァ書房, 東京 (2012)
- [6] 秋田喜代美 "読む心・書く心一文書の心理学入門 (心理学ジュニアライブラリ)". 北大路書房, 京都 (2009)