

# マトリックスモデルによるテキスト編集の共起分析

Analyzing Temporal Collocation of Actions by Matrix Type Text Editing Model

山口琢、大場みち子(はこだて未来大)、高橋慈子(ハーティネス)、  
小林龍生(スコレックス)、高橋修(はこだて未来大)

Taku Yamaguchi, Michiko Oba (Future University Hakodate), Shigeko Takahashi (Heartiness Co.,Ltd.),  
KOBAYASHI Tatsuo (Scholex Co., Ltd.), Osamu Takahashi (Future University Hakodate)

## 1. はじめに

本稿では、作文(writing、text production)行為に着目する自然言語処理を提案する(図1)。作文行為(作文過程、writing process)を分析する手法は、キーストローク分析など、いくつか研究されている。われわれは、マトリックス型テキスト編集モデルに基づくテキスト編集操作の測定を提案している[01]。このモデル構築が主たる研究内容である。このモデルの有効性は、最終的には、これに基づく(測定結果に対する)分析の有効性によって間接的に示される。分析にはさまざまなものが考えられる。今回は、共起分析の可能性を示すことを課題とする。

まず、このモデルに基づいたテキスト編集操作の測定を共起分析の観点から説明する。続いて、測定結果(編集操作ログ)と、共起行列の例を示す。最後に、テキスト編集の出力としてのテキスト(完成文章)に着目する自然言語処理との関係について考察する。

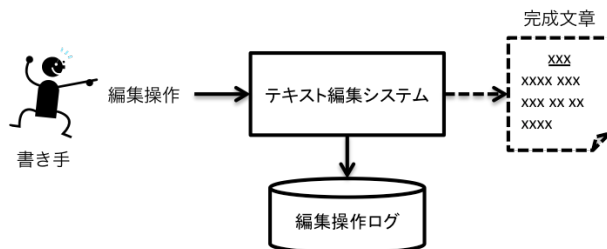


図1 テキスト編集操作ログの分析

## 2. マトリックス型テキスト編集モデル

「編集操作の共起分析」の前提は、「近くで編集されたテキスト要素同士は、文字列上やページ上や章節構造上では遠く離れていても、書き手にとって関係が深い。」というものである[02]。ここでの「近さ」とは、時間的な近さ、時間的な隣接関係(N-gram)、編集画面上の配置の近さ、同じ観点に基づくという近さ、などが考えられる。

本モデルでは、編集「操作」を記録し、編集された具体的なテキストは記録しない。ただし、後述する「共有されたラベル」すなわちワークシート中のラベルは分析者が参照できる。

### 2.1 マトリックス: ラベル付きテキスト・グループ

本モデルでは、テキストはあるグループに属している。そのグループは、例えば段落や帳票の記入欄などである。グループは、IDによって識別される。また、グループには人間(書き手)が読めるラベルをつけることができる。電子メールの「件名」、就活のエントリーシートの「志望動機」などがラベルの例である。グループとIDの関係は1対1だが、グループとラベルの関係は多対多である。書き手は、個々の作文にあたって任意にラベルを作成してつけることもできるし、既定のラベルを採用することもできる。書き手間で共有されている既定のラベルを採用するとき、それをワークシートと呼ぶ。前記のラベルの例は、一般にワークシートとして利用される。

テキストが所属するグループには、複数のラベルをつけることができる。また、ラベルをグルーピングすることもできる。この2点がマトリックスというネーミングの由来だが、本稿ではこれ以上触れない。

### 2.2 アプローチ: 共起する要素と区間

「共起」とは、「ある区間の中で2つの要素どうしが同時に観察されること」と説明される[02]。「編集操作の共起分析」では、「テキストのグループに対する編集操作」を共起する要素と考えることができる。要素が共起する区間としては、いくつか考えられる。時間に着目すると、ある時間間隔という区間、(時間的隔たりは捨てて)操作が前後隣接するという区間が考えられる。また、テキスト・グループのラベルに着目すると、同じラベルという区間などが考えられる。

本稿では「操作が前後隣接する」という区間について検討を進める。

## 3. 実験: 編集操作ログの取得

ロジカル・ライティングの講義で、「トピック・ライター」というテキスト編集ツールを使う。トピック・ライターは本モデルに従って編集操作を記録する。講義では課題に応じて、複数のワークシートを使う。

講義中で受講生が課題として行ったプレゼンについて、自他の受講生が評価するという課題を出し、そこで「プレゼン評価」というワークシートを使う。「プレゼン評価」ワークシートは、3つのラベル付きテキスト・グループを提示する。そのラベルは「ロジカルなプレゼンテーションとは何か?」、「他のグループのプレゼンを

聞いてのコメント」、「自分のグループのプレゼンテーションの企画・実施に関するコメント」の3つである。図2は「プレゼン評価」ワークシートに含まれるテキスト・グループについて、ID、ラベルと画面配置を示している。「cc\_8」などがグループのIDである。このIDを手がかりに、続く散布図や共起行列を見比べる。

ワークシート

プレゼン評価		
ロジカルなプレゼンテーションとは何か? cc_7	他のグループのプレゼンについてのコメント cc_9	自分のグループのプレゼンテーションの企画・実施に関するコメント cc_11
cc_8	cc_10	cc_12

図2 「プレゼン評価」ワークシートでのテキスト・グループのID、ラベル、編集画面上の配置

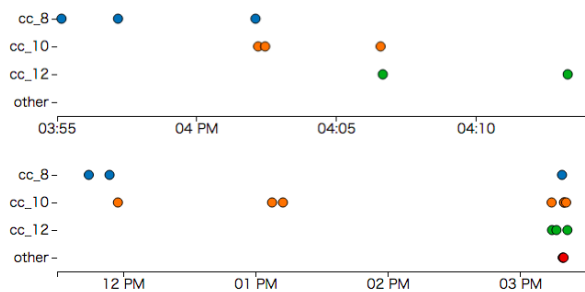


図3 編集操作の散布図

このワークシートを使った受講生の作文について、編集操作を記録して、「操作が前後隣接する」共起行列を作成して分析する。

#### 4. 結果: 編集操作ログと共起行列

図3は、編集操作について、横軸に時間、縦軸に編集操作対象のテキスト・グループのIDをプロットした散布図である。1つの散布図が1つの作文行為に対応する。散布図中の点が、その時刻にそのテキスト・グループに対して編集操作が行われたことを示している。

個々の作文について、あるグループがn番目の編集操作対象だったときに、その次(n+1番目)に編集操作対象となるグループの頻度を行列に表したものが図4である。縦軸がn番目の編集操作対象、横軸がn+1番目の編集操作対象である。縦軸は上から下へ、横軸では左から右への順番が、トピック・ライター画面上でのテキスト・グループの左から右への並び順に一致している。

#### 5. 考察

行列の、左上から右下への対角線上のセルは、同じテキスト・グループに対して編集操作を続けたことに対応する。その1つ右隣のセルは、画面上右隣のテキスト・グ

ループに編集対象が移ったことに対応している。この2つの対角線上のセルは、ワークシートに従って左から右へ書き込んでいく流れに対応する。

他方で、この対角線から右上や左下にずれたセルは、この流れから外れた編集操作が行われたことを示している。これは例えば、他グループのプレゼンについてコメントを書いた後で、あらためて「ロジカルなプレゼンテーションとは何か？」考え直したのかもしれない。

対角線から外れる/外れないという傾向が、良いか悪いかは別の問題である。一方で、対角線から外れない、すなわち左から右へ記入していけば良いということは、ワークシートのできが良いことを示すと考えられる。他方で、書き手の推敲が足りないことを示すとも考えられる。

n \ n+1	cc_8	cc_10	cc_12	other
cc_8	2	1		
cc_10		2	1	
cc_12			1	
other				

n \ n+1	cc_8	cc_10	cc_12	other
cc_8	1	1		1
cc_10		5	2	
cc_12	1		1	
other		1		2

図4 編集操作の共起行列

#### 6. 結論

以上の結果と考察から、「操作が前後隣接する」という区間に対する編集操作の共起行列は、作文行為の特徴を表している可能性がある。

統計処理によって、対角線から外れた編集操作が、対応するラベル間あるいは対応するテキスト・グループ間の何らかの関係を示唆する場合は考えられる。そのとき、ラベルやテキスト・グループ内の文字列に対する自然言語処理と組み合わせることで、自然言語理解が深まると期待できる。

#### 参考文献

- [01] 山口琢, 大場みち子, 高橋修 "マトリックス型テキスト編集モデルによる学習過程の可視化". 教育システム情報学会研究報告 28(7), pp.101-108 (2014-03)
- [02] 相澤彰子, "共起に基づく類似性尺度", 社団法人日本オペレーションズ・リサーチ学会, オペレーションズ・リサーチ: 経営の科学 52(11), 706-712, 2007-11-01