

有効な操作履歴の評価によるテキストマイニングスキルの伝達支援

Skill Transmission Support for Text Mining by Evaluating Effective Operation Histories

中江 剛士*¹

Tsuyoshi Nakae

砂山 渡*²

Wataru Sunayama

畑中 裕司*²

Yuji Hatanaka

小郷原 一智*²

Kazunori Ogohara

*¹滋賀県立大学大学院工学研究科

Graduate School of Engineering, The University of Shiga Prefecture

*²滋賀県立大学工学部

School of Engineering, The University of Shiga Prefecture

In these days, the text mining skill that analyzes a review article and questionnaire data is demanded for development and the improvement of the product. However, it is difficult to get know-how of the data analysis by the self-education in present conditions. A framework to get a skill of text mining has been suggested by the frame which shared an operation history at the time of the text mining by operation of other person. Therefore, purposes of this study is what perform skill acquisition support of the text mining by making it easy to look for an effective history from the operation history of many people. As a result of evaluation experiment, a tendency to increase to the number of interpretation to be provided was seen from a text. Thus, it is effective to make it easy to look for an effective history from the operation history of many people.

1. はじめに

近年、非常に多くの新製品や新アプリケーションソフトウェアが登場し競争が行われている。これにより新製品や新アプリケーションソフトウェアを開発する企業などには自社の製品やアプリケーションソフトウェアのレビュー記事、及びアンケートデータの分析を行うためのテキストマイニングのスキルが求められている。しかし、データ分析のノウハウを獲得することは、独学では困難な現状がある。これまでもテキストマイニングのツールにおける文章要約や単語抽出などの解析ツールの操作履歴を共有する枠組みにより、他人の操作からスキルを獲得する枠組みが提案されてきた。

運転が上手な人の運転を見て学ぶことで、自分の運転技術を上げることができるといったように他人の操作を見て真似することはスキルの獲得に有効である。テキストマイニングにおいてもツールの操作を見ることで操作を参考にし、テキストマイニングのスキルを獲得することができると考えられる。しかし、保存される履歴数が増加するほど、参考となる履歴の選択が困難になるという点が課題となっている。

そこで本研究では、対象をテキストマイニングを行ったことがない初心者に限定し、TETDM 上の機能の一つとして、保存された履歴を評価する機能、保存された履歴の絞り込み機能を実装し、多くの人の操作履歴の中から有効な履歴を探しやすくする。テキストマイニングを行う際の発想を支援し、利用者のテキストマイニングのスキル獲得支援を行うことを目的とする。

2. 関連研究

2.1 履歴の参照によるスキル伝達に関する研究

他者の学習成果物、成績、学習日記、学習履歴などを蓄積して公開・共有することで学びを誘発する e ポートフォリオ推薦システムを提案した研究がある [植野 14]。この研究では、対象と類似の学習履歴を持つ学習者の e ポートフォリオを可能な限り多様に推薦する支援を行ったが、本研究では、テキスト

連絡先: 中江剛士, 滋賀県立大学大学院工学研究科電子システム工学専攻, 〒 522-8533 滋賀県彦根市八坂町 2500, oh23tnakae@ec.usp.ac.jp

マイニング時の操作内容を保存した操作履歴を評価し、対象に見てもらった操作履歴を絞り込むことでテキストマイニングのスキル伝達を支援することを目的とする。

自己及び他者の視線行動を可視化・共有できる仮想空間を構築することにより、作業支援を行う研究がある [藤本 04]。

この研究では、視線履歴から注視点を算出し、HMD (ヘッドマウントディスプレイ) に出力することで、ユーザに作業支援を行うが、本研究では、TETDM の操作履歴を再現することでスキルの伝達支援を行う。

2.2 有効な情報の絞り込みに関する研究

ネットワーク上に分散する複数のデータベースやコンテンツを横断的に検索し、検索された情報を地図上にマッピングする情報ナビゲーションシステムを提案した研究がある [北角 01]。この研究では、筆者らの提案しているディレクトリ・ビューという機構により効率的な情報の絞り込みを行っているが、本研究では、多数の操作者が共通して行う操作により表示する操作履歴を絞り込むことでテキストマイニングのスキル伝達を支援することを目的とする。

マルチメディア情報を対象にした情報検索システムの枠組みと、それを実現する 3 次元情報検索インタフェースに関する研究がある [上本 96]。この研究では、3 次元情報空間のブラウジング検索に検索対象の拡大・絞り込みを実現するキーワード検索、対象指定による類似検索などのハンティング検索を統合したインタフェースに関する研究であるが、本研究では、TETDM の操作履歴を内容によって絞り込む機能を開発し操作履歴選択を促すことで、ユーザへテキストマイニングのスキルの伝達支援を行う。

2.3 テキストマイニングのスキル伝達支援に関する研究

テキストマイニングスキルの伝達支援に関する研究がある [西原 15][中垣内 14][金谷 16]。この研究では、スキル獲得を支援するチュートリアルの開発により、ユーザにテキストマイニングの課題を基礎的なものから応用的なものへと順に解かせることでスキル伝達を促す。この研究と本研究では、TETDM を用いてテキストマイニングのスキルの伝達支援を行うという点で共通している。しかし、本研究では他人の操作を評価し、操作者の意図を伝えることでスキル伝達の支援するという点で異なる。

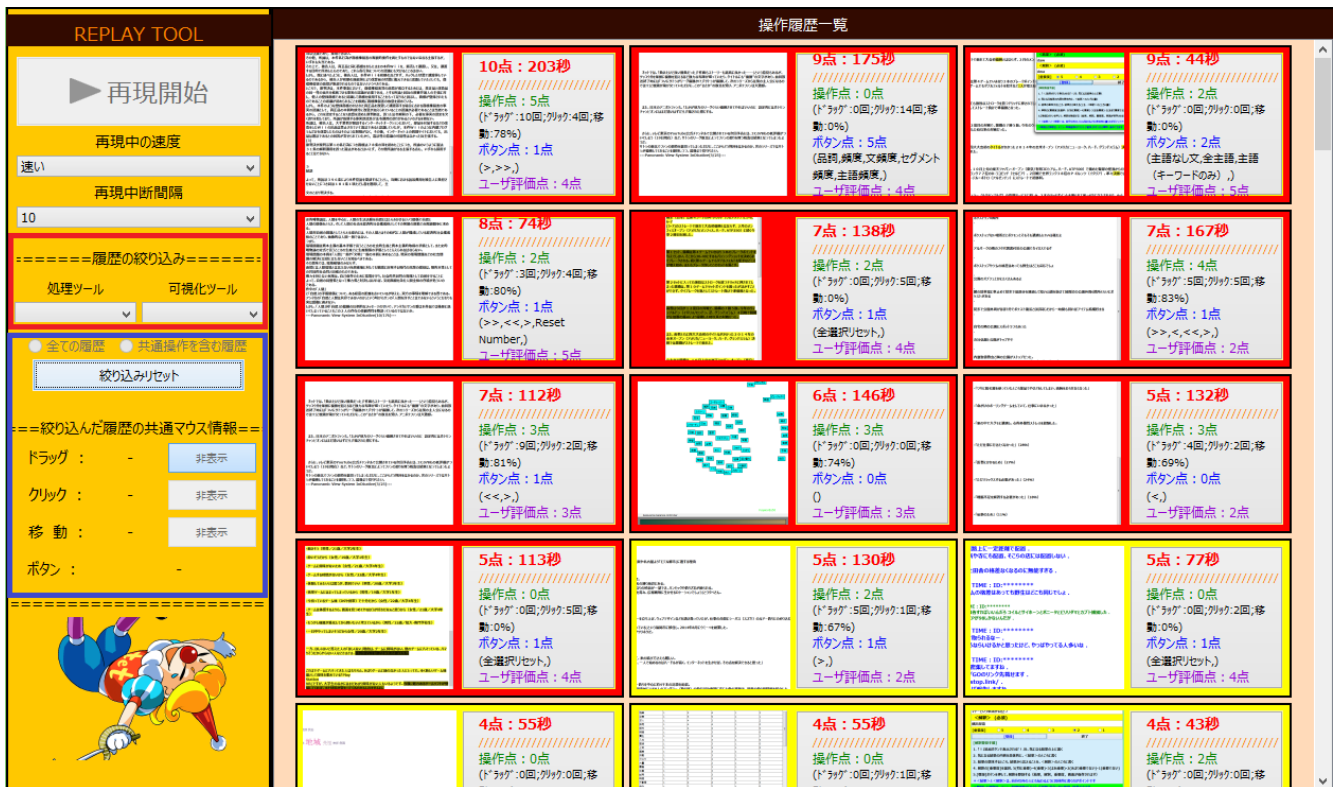


図 2: 再生する操作履歴を選択するインターフェース

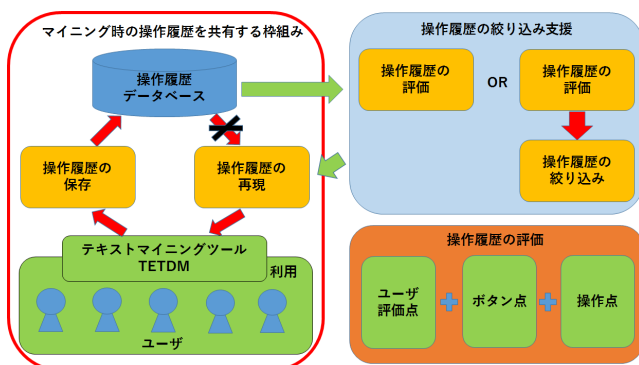


図 1: マイニングスキル伝達支援の枠組み

3. スキルの伝達のためのツールの操作履歴の絞り込み支援

図 1 に本研究が提案する支援の枠組みを示す．本研究では、マイニング時の操作履歴を共有する枠組み [金谷 16] の操作履歴の再現前に、保存された操作履歴を評価し、ユーザに操作履歴を絞り込んでもらうことで、ユーザが有効な操作履歴を探しやすくする．

操作履歴の再現インターフェースに追加する形で操作履歴の絞り込み機能を追加した．図 2 は再生する操作履歴を選択するインターフェースである．左に操作履歴の再現機能と絞り込み機能のインターフェース、右に操作履歴の一覧が表示される．

3.1 操作履歴の評価

1 章でも述べたように、本研究ではスキル伝達の支援の対象をテキストマイニング初心者と限定しているため、ユーザに基本操作を覚えてもらうことがテキストマイニングスキルの伝

達につながるとしている．基本操作とは多数のユーザが行う操作、共通操作に含まれるものと考えられるので、多数のユーザが行う共通の操作に高得点を与える評価とする．

TETDM の起動時に処理ツールごとにデータベース内の操作履歴を分類し、その分類ごとにユーザ評価・共通ボタン操作・共通マウス操作の 3 つの評価項目で操作履歴を評価する．評価項目 1 つを各 5 点で採点し、15 点満点でその操作履歴の評価点を決定する．

3.1.1 ユーザ評価点

TETDM の機能に分析の結果や解釈を登録する機能があるが、登録する際にユーザが決める 5 点満点の重要度をそのまま評価点として利用する．ユーザが主観で決める点数であるが、高得点の操作履歴ほど登録したユーザが重要であると考えた解釈を探した操作履歴であるので、参考になると考えられる．

3.1.2 ボタン点

テキストマイニング時に押したボタンに関する点数評価である．ボタンにおける共通操作は、多数の人が使用したボタンの操作が共通な操作であると考えられる．同じ処理ツール内で押されたボタン名をすべて取得し、共通だとみなせると考えられる「結果と解釈」登録者の半数の人が使用したボタンを含む操作履歴には 1 点を加算するものとし、最大 5 点を加算するものとする．多数の操作履歴で押されたボタンということは、それだけ重要な操作、または基本操作であると考えられる．

3.1.3 操作点

テキストマイニング時に動かしたマウスの動作に関する点数評価である．マウスの動作はドラッグ、クリック、マウス移動の 3 つの動作を評価する．マウス操作に関しての操作点の配点を表 1 に示す．高得点であるほど、ユーザのテキストマイニングに対する積極性が高いといえる．

表 1: 操作点の配点

マウス操作	評価方法
ドラッグ	上位半分で1点, 上位4分の1で2点
クリック	上位半分で1点, 上位4分の1で2点
移動量	移動面積がディスプレイの5割以上で1点

3.1.4 操作履歴の評価結果の反映

図 2 右の操作履歴一覧の様子、再現機能の操作履歴ボタンに表示することで操作履歴の評価結果を反映させた。ボタン左側には解釈の登録時の画面の画像を表示している。また、右側には1番上に、操作点、ボタン点、ユーザ評価点をまとめた総合点とその操作履歴を保存したユーザのツールセットを行ってから解釈の登録までの操作時間が表示される。その下には総合点の内訳と評価基準となった操作内容が表示される。また、操作履歴ボタンの上にマウスを乗せると使用したツール情報、解析の目的が表示される。

3.1.5 共通マウス操作

図 2 赤枠内のインタフェースを用いツールにより操作履歴を絞り込んだ後、共通マウス操作の有無によって更に絞り込む機能を開発した。図 2 青枠内下部は、絞り込んだ履歴の共通マウス情報である。ツールにより絞り込んだ操作履歴群において、各操作を含む操作履歴が多数を占めるならば「有」、含まない操作履歴が多数を占めるならば「無」が表示される。また、右の非表示ボタンにて共通操作を含む操作履歴で絞り込むことができる。

共通マウス操作で操作履歴を絞り込むことによって、普通ではあまりしないような独自の操作を省くことができる。

3.1.6 共通ボタン操作

ツールにより操作履歴を絞り込んだ後、操作履歴に使用されているボタンによって更に絞り込む機能を開発した。ツールにより絞り込んだ操作履歴群において、5割の操作履歴に含まれるボタンの操作をその操作履歴群において共通ボタン操作であると判定し表示する。また、図 2 青枠内上部の共通操作で絞り込むを選択することで共通ボタン操作を含む操作履歴を絞り込むことができる。

共通ボタン操作で操作履歴を絞り込むことによって、多数の人が使用するツールの機能を使用することができる。

4. 操作履歴の評価及び絞り込み機能の評価実験

4.1 実験目的

本研究で開発した TETDM のマイニング時の操作履歴の評価および操作履歴の絞り込み機能を利用することで、テキストマイニングスキルの伝達に役立てられるかをについて評価するため、評価実験を行った。以下に実験について述べる。

4.2 実験方法

一度もテキストマイニングを行っていない 16 名を対象として評価実験を行った。実験では 2 日間を利用して、1 日目に全員を対象に登録忘れが無いよう、これ以上「結果と解釈」が思いつかなくなるまでテキストマイニングを行ってもらい操作履歴を収集する。2 日目に被験者を半分に分け、操作履歴の評価、および操作履歴の絞り込み機能を含む TETDM(提案)を用いる群と含まない TETDM(比較)を用いる群でグループ分けをし、それぞれの TETDM にてテキストマイニングを行ってもらった。このグループ分けにより、操作履歴の評価、および操作履歴の絞り込みの検証を行う。

マイニングに利用するテキストにはポータルサイト Amazon [Amazon] で販売されるマウスコンピューター製タブレットノートパソコンのレビューを収集し使用した。

被験者には、「被験者がとあるパソコン会社の製品開発部の社員である。」という仮定のもと、次に開発するパソコンのアイデアを考えてもらう」という課題を出題し、上記のテキストを用いて TETDM によるテキストマイニングを行ってもらった。

1 日目の実験手順を以下に述べる。

1. TETDM でレビュー記事をテキストマイニングし「結果と解釈」を登録する。「結果と解釈」は 10 個を目標に、最低でも 5 件の登録を行う。ツールの選択はアプリケーションツールによるセットで選択する方法に限定する。ただし「単語抽出」は選択不可とする(制限時間は最大 40 分)
2. 知識創発インタフェースを利用し、タブレットノートパソコンの特徴の「結果と解釈」を 1 つの解釈にまとめる。
- 2 でまとめた解釈から次に開発するパソコンのアイデアを考える(制限時間は 2, 3 で 20 分)

2 日目の実験手順を以下に述べる。

1. TETDM でレビュー記事をテキストマイニングし 1 日目のものに追加する形で「結果と解釈」を登録する。「結果と解釈」は可能な限り登録してもらおう。また再現機能で 1 日目の他の被験者の操作を閲覧することができ、分析に取り入れることができる。ここでもツールの選択はアプリケーションツールによるセットで選択する方法に限定する。ただし「単語抽出」は選択不可とする(制限時間は最大 40 分)
2. 知識創発インタフェースを利用し、1 日目のものを含めたタブレットノートパソコンの特徴の「結果と解釈」を 1 つの解釈にまとめる。
- 2 でまとめた解釈から次に開発するパソコンのアイデアを再度考える(制限時間は 2, 3 で 20 分)
4. アンケートに回答する(制限時間 5 分)

5. 結果と考察

5.1 操作履歴の閲覧から得られた「結果と解釈」の比較

「結果と解釈」の登録数の増加が再現機能によって発生したものであることを検証するため、再現機能を利用して登録された「結果と解釈」の登録数の集計を行った。結果を図 3 に示す。

ここで、被験者 A-H が提案群、被験者 I-P が比較群である。また、以降のように扱うものとする。グループ別に登録数の伸びの平均を比較すると、提案群が約 4.4 個、比較群が約 3.6 個となりあまり差は現れなかった。また、操作履歴の再現による登録数の伸びの平均を比較すると、提案群が約 3.9 個、比較群が約 2.8 個となり、より登録数の伸びの差が顕著となった。これは履歴の評価機能と絞り込み機能により、有効な履歴を探しやすくなったためだと考えられる。

5.2 再現によって得られた発想

再現を見た後で、どのような発想を得ることができたかを検証するため、2 日目に得られたログデータから 1 日目には使用しなかったが、2 日目に新しく使用して「結果と解釈」を登録したツール数、1 日目に使用したが、2 日目に新しく行った操作数をそれぞれ数えた。ここで、1 日目には使用しなかったが

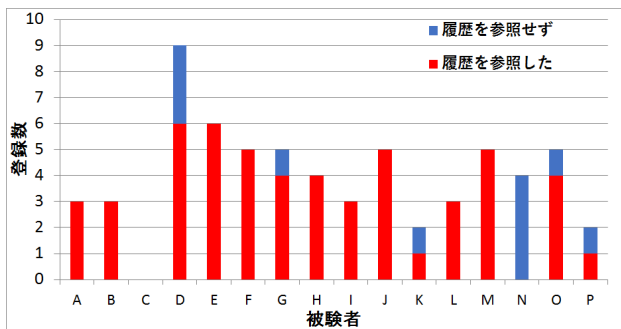


図 3: 「結果と解釈」の登録数

表 2: 操作履歴の再現により得られた発想の数 (被験者平均)

	新たに使用した ツール数の平均	新たに行った 操作数の平均	発想数 平均
提案	0.8	1.4	2.1
比較	1.3	0.1	1.4

2 日目に新しく使用して「結果と解釈」を登録したツールおよび 1 日目に使用したが 2 日目に新しく行った操作を再現によって得られた発想と定義する。表 2 に再現によって得られた被験者平均の操作履歴の再現により得られた発想の数を示す。新しい発想を得ることができ、スキルを習得できたと考えられる。

5.3 アンケート結果からの分析

2 日目の実験終了後、被験者にアンケートに回答してもらい、操作履歴の評価、及び絞り込みがどのような理由でマイニングに影響を与えたのかを調査した。

Q4「再生された操作履歴は、新たな操作方法を知るきっかけになったか」という問いに対し、提案組では 4 人の被験者が 4 または 5 点、1 人の被験者が 1 または 2 点と回答した。しかし、比較組のグループでは 2 人の被験者が 4 または 5 点、3 人の被験者が 1 または 2 点と回答した。操作履歴の評価、及び絞り込みは新たな操作方法を知るきっかけとして有効であったという結果が得られ、テキストマイニングのスキル伝達に役立ったといえる。

Q3-1～Q3-4「参考になる操作履歴を探す際に、総合点、操作点、ボタン点、ユーザ評価点の表示は役に立ったか」という問いに対し、総合点とユーザ評価点に関しては、総合点では 6 人の被験者が 4 または 5 点、1 人の被験者が 1 または 2 点と回答した。また、ユーザ評価点では 5 人の被験者が 4 または 5 点、2 人の被験者が 1 または 2 点と回答した。

操作履歴を探す際に総合点、ユーザ評価点が操作履歴を再生する指標となったと考えられる。しかし、操作点、ボタン点に関しては、操作点では 3 人の被験者が 4 または 5 点、3 人の被験者が 1 または 2 点と回答した。また、ボタン点では 4 または 5 点と回答した被験者はおらず、2 人の被験者が 1 または 2 点と回答した。記述内容には、「評価基準が分からない」「メリットが分からない」といった意見が散見され、総合点、ユーザ評価点と比べると操作履歴選択への貢献度は低かったのだと考えられる。この課題を克服する方法として、操作回数や移動面積、使用したボタンの表示を大きくし、評価点の決定因子を強調することにより、点数の意味をユーザに分かりやすく伝えることが考えられる。

6. 結論

TETDM を利用してテキストマイニングのスキルの伝達支援を行うため、TETDM の機能として保存された操作履歴の評価、絞り込み機能を開発し、有効な操作履歴を探しやすくする枠組みを構築した。

提案する機能が、初心者スキル獲得につながるかを検証する実験を行った。タブレットノートパソコンのレビュー記事についてテキストマイニングを行い、商品の特徴を発見するという課題を提案する機能を実装してこなしてもらったところ、操作の参考になる操作履歴を探しやすくなり、操作履歴を閲覧することで新しいツールを使いテキストの分析を行うテキストマイニングスキルの伝達を行えた。

今後の課題として、点数の意味をより強調するなど、ユーザに分かりやすいインターフェースにすることによって、初心者ユーザへのテキストマイニングスキルの伝達をより効率的に支援することを目標としていきたい。

参考文献

- [植野 14] 植野真臣: 過去の学習者履歴データを利用した e ポートフォリオ・システム, 情報知識学会誌, Vol.24, No.47, pp.414-423 (2014)
- [藤本 04] 藤本武司, 砂山渡, 山口智浩, 谷内田正彦: 視線行動の可視化による着眼スキル伝達支援, 人工知能学会論文誌, Vol.19, No.3B, pp.174-183 (2004)
- [北角 01] 北角智洋, 池田哲夫, 田辺弘実, 星隆司: 分散した地理情報を統合利用する情報ナビゲーションシステム, 情報処理学会論文誌データベース (TOD), Vol.42, No.SIG15(TOD12), pp.30-39 (2001)
- [上本 96] 上本俊博, 上原祐介: 3 次元情報検索インターフェース, 情報処理学会研究報告情報学基礎 (FI), Vol.1996, No.35, pp.45-52 (1996)
- [西原 15] 西原陽子, 中垣内李菜, 川本佳代, 砂山渡: TETDM を用いたテキストマイニングのスキル獲得を支援するためのチュートリアルシステムの開発, 知能と情報, Vol.27, No.5, pp.771-783 (2015)
- [中垣内 14] 中垣内李菜, 川本佳代, 砂山渡: 統合環境 TETDM を用いたテキストマイニング初心者のスキル獲得支援, 第 28 回人工知能学会全国大会, 1H5-NFC-01b-3 (2014)
- [金谷 16] 金谷直哉, 砂山渡: ツールの操作履歴の共有によるテキストマイニング時の発想支援, 人工知能学会第 12 回インタラクティブ情報アクセス可視化マイニング研究会, SIG-AM-12-05, pp.29-34 (2016)
- [砂山 15] 砂山渡, 高間康史, 西原陽子, 徳永秀和, 串間宗夫, 阿部秀尚, 梶並知記, 松下光範, Bollegala Danushka, 佐賀亮介, 河原吉伸, 川本佳代: テキストデータマイニングのための統合環境 TETDM, 第 29 回人工知能学会全国大会, 2E3-NFC-01a-1 (2015)
- [TETDM] Total Environment for Text Data Mining (テキストデータマイニングのための統合環境), (URL)http://tetdm.jp
- [Amazon] Amazon.com, Inc.,(URL) www.amazon.co.jp