

統合環境TETDMを用いたテキストマイニングにおける

初心者のためのツール選択支援

Interface for Beginners to Set Text Mining Tools and Visualization Tools on TETDM Appropriately and Quickly

中垣内 李菜
Rina Nakagochi

川本 佳代
Kayo Kawamoto

砂山 渡
Wataru Sunayama

広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

TETDM is an integrated environment which flexibly combines all sorts of mining tools with various visualization tools. When beginners try to analyze data with TETDM, it is hard for them to set proper text mining tools and appropriate visualization tools. Hence in this paper, we propose an intuitive graphical interface which effectively assists beginners to get expected knowledge. We showed that our interface is effective for beginners by reporting some experimental results.

1. はじめに

近年、安価で大容量の記憶装置や、高性能の演算処理装置の普及、ネットワーク環境が整備されたことにより、膨大な量のデータが取り扱われるようになり、「データマイニング」が注目されるようになった。テキストデータマイニングを行うツールの一つに統合環境TETDM[砂山 2013]がある。統合環境であるTETDMは、テキストマイニング技術を提供している複数の処理ツールと、画面への出力を行う複数の可視化ツールを柔軟に組み合わせて、ひとつのソフトウェア内で動作させることが可能な統合環境である。

しかし、用意されているツールが多いため、初心者がすべてのツールを把握することや、目的に応じて適切にツールを選び、組み合わせることが困難である。

そこで本論文では、初心者のツール選択支援に有効なインタフェースを提案し、テキストマイニング初心者が迅速にツールを選択して、期待する結果が得られるようにすることを目的とする。

2. 関連研究

2.1 統合環境 TETDM

TETDM は、複数のテキストマイニング技術を柔軟に組み合わせて使用可能な統合環境を構築し、社会的創造的活動を支援できる環境の提供を目指している。TETDM に関して、これまで処理ツールや可視化ツールの開発に関する様々な研究[砂山 2001][山手 2012][梶並 2012]が行われてきたが、本研究では、TETDM を初めて使う人を対象とし、TETDM の機能をより容易に、より有効に使えるようにするために、新しいインタフェースの提案を行う。

2.2 情報可視化に関する研究

情報の見せ方の工夫によりユーザの理解を支援する研究として、3次元グラフィックスを用いて情報の視覚化を行っている研

究[塩澤 1997][牧田 2006]がある。これらの研究では、画面上に表示された情報(ノード)の中から、任意にピックアップしたいものをユーザが選択し、持ち上げ操作を行うことで関連のあるノードが順次引き上げられる形を取り、選択した情報が他の情報とどのような関連を持っているかを多角度から把握することができる。しかし、従来研究では、与えられた情報の全てを提示しているため、ユーザに、大量の情報からピックアップしたい情報を見つけ出すという負担がかかってしまう。本研究の提案システムでは、ユーザの選択肢を減らして提示することにより、ユーザの負担の軽減を行う。

2.3 インタフェースを用いたユーザの支援

学習に先立って学習内容、順序を決定する学習パスプランニング支援環境を提供し、学習活動を支援している研究がある[鈴木 2000]。この環境では、学習者自身が膨大な学習リソースを比較、参照しながら学習プランを立て、最後に完成したプランに従い学習を行う。しかし、従来研究では、複数存在する探究可能なパスから、学習者自身が学習リソースの検討を行うため、効果的に学習リソースを探究することが難しくなっている。また、プランニングを終えなければ学習を始められないという難点もある。本研究では、あらかじめ学習者が学習に使用するツールの候補をシステムにより減らしておくこと、ツールを設定した直後にツールの使用が可能ということが目的であるという点で異なる。

3. システム構成

本システムの構成を図1に示す。システムは、TETDM で用意されている各ツールに関する情報を入力とし、入力された情報をもとに処理/可視化ツール同士の組み合わせの可否のチェックを行う。チェック時に取得した情報により並び替えたツールをインタフェース上に描画し、ユーザがインタフェース上で各パネルにツールをセットしたら、インタフェース上に変更状況に従い再描画した後、TETDM へ向けて各パネルとツールの情報を出力する。出力された

連絡先: 中垣内李菜, 川本佳代, 砂山渡, 広島市立大学大学院情報科学研究科システム工学専攻, 広島市安佐南区大塚東三丁目4番1号, {nakago,kayo,sunayama}@sys.info, hirosima-cu.ac.jp

情報をもとに TETDM の各パネルとツールを変更する。以下に各モジュールについて説明する。

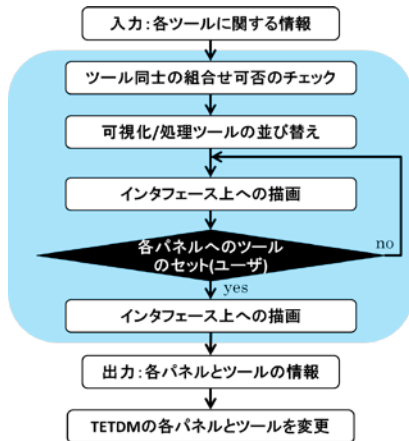


図 1: システムの構成

3.1 ツール同士の組合せ可否のチェック

各処理ツールにおいては、組み合わせることができる可視化ツールがあらかじめいくつか指定されている。そこで、各処理ツールと、それぞれ対応している可視化ツールにリンク付けを行い、各可視化ツールに付けられたリンクの数を数える。

3.2 可視化/処理ツールの並び替え

3.2.節にて付けられたリンクの数が多い順に可視化ツールを並び替える。その際、各可視化ツールに対応している処理ツールも並び替える。

3.3 インタフェース上への描画

3.3 節にて並び替えた処理ツールと可視化ツールをインタフェース上に描画する。図 2 に本インタフェースの表示例を示す。このとき、(1)インタフェースの概要、(2)直感的にツールをセットする機能、(3)ユーザが着目したツールの組み合わせと、ツールの説明の表示機能は以下の通りである。

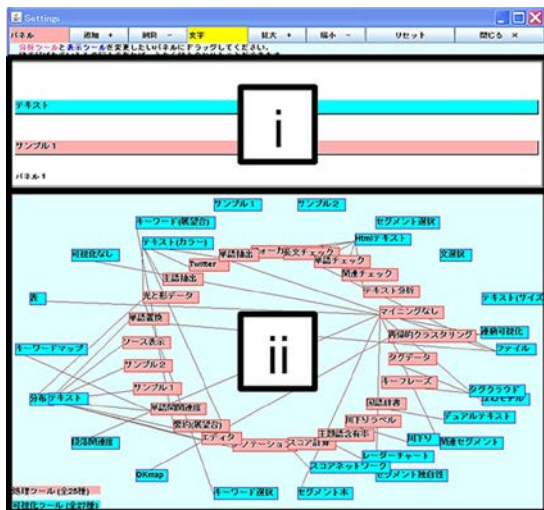


図 2: インタフェースの表示例

(1) インタフェースの概要

本インタフェースの画面は、上下 2 つの大きな長方形に分けられている。上部の領域 (i) は「セットエリア」、下部の領域 (ii) を「ツール表示エリア」である。

セットエリアでは、各パネルと、それぞれのパネルにセットされた処理ツール及び可視化ツールの名前が表示される。ツール表示エリアでは、TETDM で使用可能な処理ツールと可視化ツールの一覧が表示される。すべてのツール名は黒字で示されるが、各ツール名を囲んだ四角形の色で処理ツールと可視化ツールの判別を可能にした。また、処理ツールと可視化ツールの組み合わせが可能な場合、ツール同士を線で結び組み合わせ可能であることを示した。

(2) 直感的にツールをセットする機能

ユーザは、ツール表示エリアから、使用したいツールをセットエリアにドラッグアンドドロップすることにより、直感的に移動先のパネルへのツールのセットを行う。ユーザがセットしたツールは連動して TETDM 上で実行されるため、ウィンドウを閉じることなく、ツールをセットするだけでテキストマイニングを行うことが可能である。

(3) ユーザが着目したツールの組み合わせと、ツールの説明の表示機能

本インタフェースは、初期状態ではすべての処理ツールと可視化ツール、及び組み合わせが可能なことを示す線を表示することを先に述べた。しかし、常にすべての情報が表示されていた場合、情報量が多すぎて、すべてのツールを理解し、ツール同士の組み合わせを判断することは難しい。そこで、ユーザがツールに着目 (画面上のオブジェクトにマウスマウスカーソルを重ねること) したときに、そのツールと組み合わせ可能なツールのみを表示することにより、情報量の削減を行なった。このとき、削減後も元のツールの位置関係がわかるように、削減対象を処理ツールまたは可視化ツールのどちらか一方に限定した。

また、各ツールに着目した時に、本インタフェースのツール表示エリアの中央部分に、対応するツールの説明を表示できるようにした。ここで、説明を中央に表示したのは、ツール名が隠れないようにするだけでなく、各ツールからの目線の移動量を少なくし、ユーザの負担を軽減するためである。

3.4 各パネルへのツールのセット(ユーザ)

ユーザが、ツールの一覧から、使用したい処理ツールと可視化ツールをパネルに移動させることにより、ツールのセットを行う。

3.5 出力:各パネルとツールの情報

システムは、3.4 節にて、各パネルにセットした処理/可視化ツールのペアを統合環境 TETDM に向けて出力する。統合環境 TETDM は出力された情報をもとにパネルを作成し、それぞれのパネルで指定された処理ツールと可視化ツールをセットする。

4. ツール選択支援インタフェースの評価実験

4.1 実験目的

提案したツール選択支援インタフェースが初心者のツール選択の支援に有効であるか、また、使いやすいものであるかを検証することを目的とする。

4.2 実験方法

被験者は TETDM を使用したことのない情報系・国際系の学部・研究科に属する大学生・大学院生の男女 24 名である。

アプリケーションに関する事前テストの結果により被験者を、提案システムを用いる群と、比較システムを用いる群に分けた。比較システムは、各パネルの上部に備え付けられている「ツール選択」ボタンを押すと出現するポップアップウィンドウの、ポッ

プアップ内に格子状に並べられたツール名の一覧から任意のツールをクリックすると、そのツールをセットするものであり、ツールの検討を行う際に、その都度システムの起動とウィンドウの開閉を行う必要がある。

まず、両グループに TETDM とインタフェースの使い方について簡単な説明を行った。次に、両群に TETDM を用いて、制限時間を 10 分として、テキストマイニングを行う課題を 12 問解いてもらった。課題の内容は、文章中の文の数や主題などの情報を問うものや、文章の要約や誤字の修正を行うものである。このとき、解答の正誤とともに両群がそれぞれ使用したシステムによる「課題の達成速度の違い」を明らかにするために、時間を計測した。

最後に、両群がそれぞれ使用したシステムの「使いやすさ」について主観の評価を得るために、各群の被験者に、事後アンケートを行った。

4.3 実験結果と考察

(1) 課題の得点

各課題について、被験者の平均得点(1 課題 1 点として 12 点満点)について t 検定の結果を表 1 に、群ごとの課題別の平均得点を図 3 に示す。

表 1: 課題の平均得点(t 検定)

	提案システム群	比較システム群
平均	9.90	8.92
検定結果	t(22)=2.26 p<.01	

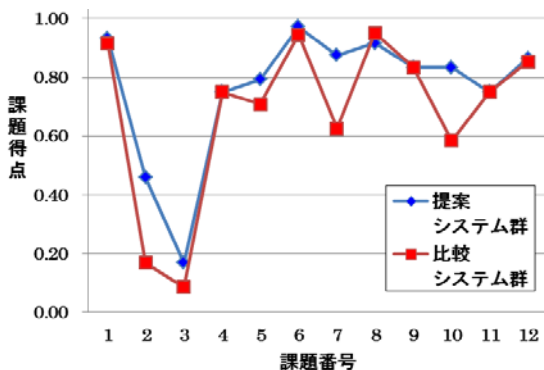


図 3: 課題別の平均得点

表 1 より、両群の平均に有意差があった。よって、提案システム群の方が、的確に課題を遂行するのに有効であるといえる。

次に、図 3 より、課題 2, 7, 10 で、提案システム群の方が比較システム群よりも得点が高かった。これらは、パネルを複数使用するような課題設定をしており、提案システム群ではひとつのウィンドウで複数のパネルへのセットができたため、効率よくツールをセットでき、結果として制限時間内に適切なツールのセットができたと考えられる。

一方、課題 8 では、比較システム群の方が提案システム群よりも得点が高かった。この課題は、一度行った課題を少し変えて出題しており、難易度を上げるために画面上に表示される言葉と異なる言葉を課題文に使用した。すべての被験者が適切なツールをセットすることができていたため、ここで生じた差は、言葉の認識の差であり、システムによる差ではないと考えられる。

(2) 各課題の必須ツールセットの所要時間

各課題を開始してから、解答に必須のツールを全てセットするまでの平均時間について、t 検定の結果を表 2 に示した。また、

課題別の必須ツールを全てセットするまでの所要時間を群ごとに示したものを図 4 に示す。なお、課題 12 はツールのセットが必須ではなかったため省略した。

表 2: 必須ツールを全てセットするまでの時間(sec)(t 検定)

	提案システム群	比較システム群
平均	123.16	194.97
検定結果	t(22)=4.30 p<.01	

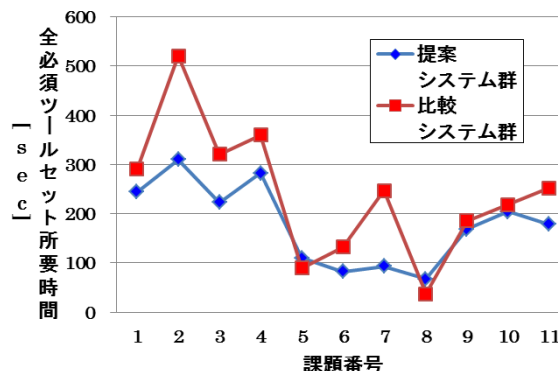


図 4: 課題の必須ツールを全てセットするまでの時間(sec)

表 2 より、両群の課題の必須ツールを全てセットするまでにかかった時間について、両群の平均に有意差があった。よって、提案システム群の方が、迅速に解答を行えたといえる。

次に、図 4 より、各課題において、概ね提案システム群の方が比較システム群よりも所要時間が短かった。

特に、課題 7 は端的である(p<.01 以下のものでも有意差があるもののみ)。これは、実際にツールのセットを行って適切なツールを探すような課題であったため、ウィンドウの開閉を繰り返す必要がある比較システムの方が、より多くの時間がかかったと考えられる。

また、比較システムのウィンドウのサイズを変更して使用した被験者からは、「ウィンドウのサイズを変えると、ウィンドウで表示されるツール名の場所が変わってしまい、その都度探すのに時間がかかった」という意見をいただいた。比較システムでは、ウィンドウがパネルを覆う大きさで表示される。ここで、ウィンドウを使用中に文章を見ようとすると、ウィンドウの大きさや位置を変えなければならず、ウィンドウのサイズと位置を変えてさらにツールを探しなおす必要があったため、より多くの時間がかかってしまったと考えられる。一方提案システムでは、ウィンドウがパネルよりも小さいサイズで表示されるため、ウィンドウのサイズの変更が少ない回数で済み、さらにツールの位置が変わらないため、比較システムよりも所要時間が短かったと考えられる。

この他、比較システムを使用した被験者から、「マウスが指しているツールが、実際にどう動くかがパネルに反映されるとツール選びに時間がかからなくていいと思った」という意見をいただいた。提案システムでは、ウィンドウをその都度閉じる必要がなく、ツールの内容の確認がしやすくなっており、結果として所要時間を短縮できたと考えられる。

(3) 提案/比較システム起動・ツールセット回数

各被験者のシステム起動回数(実験中に、被験者がそれぞれ使用したシステムを 1 課題あたり何回起動したか)について、t 検定の結果を表 3 に、群ごとの課題別のシステム起動回数を図 5 に示す。また、1 課題あたりシステムを用いてツールをセットした回数について、t 検定の結果を表 4 に示す。

表 3: システム起動回数(t 検定)

	提案システム群	比較システム群
平均	1.26	14.12
検定結果	t(22)=11.29 p<.01	

表 4: 課題の平均ツールセット回数(t 検定)

	提案システム群	比較システム群
平均	10.36	13.92
検定結果	t(22)=2.20 p<.01	

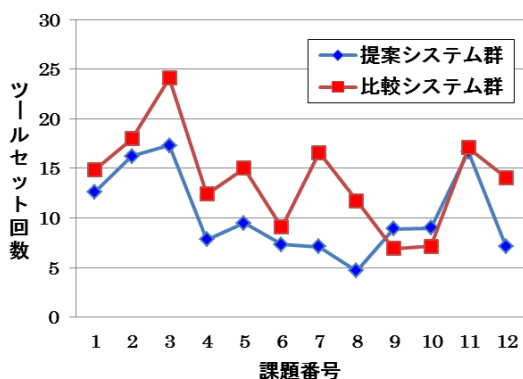


図 5: 課題の平均ツールセット回数

表 3 より, 両群のシステム起動回数について, 両群の平均に有意差があった。よって, ツールをセットするたびに起動を行う比較システムよりも提案システムの方が, ウィンドウの立ち上げ回数を大きく減らせ, 被験者の負担を軽減できたといえる。

次に, 表 4 より, 両群のツールのセット回数について, 両群の平均の差は有意であった。よって, 提案システムの方が, 少ないツールセット回数で課題に解答できたといえる。また, 図 5 より, ほとんどの課題において, 提案システム群の方が比較システム群よりもツールセット回数が上回った。

全課題終了後に, 両群の被験者に「処理ツールと可視化ツールの組み合わせをパネルにセットして探すのは難しかったか」という質問をしたところ, 提案システム群と比較システム群それぞれについて「簡単だった」と回答した人数と, 「難しかった」と回答した人数は, 5:7 と 3:9 であった。提案システムでは, 組み合わせを線で結んで表示しているが, 被験者から「ツール同士の組み合わせを示す線が重なっていると, 関連性がわかりにくい」という意見をいただいたが, 一方で「カーソルを合わせると対象のツールのみが浮かび上がるので組み合わせがわかりやすかった」という意見もいただいた。比較システムは, ウィンドウ内で各ツール名を囲んだ四角形の色を変更することにより組み合わせ対象のツールを示すが, 被験者から「ツールをクリックしたときにペアが変わるのがよくわからなかった」「処理ツールと可視化ツールの境目がはっきり表示されていないので, ツールと一緒に見える」という意見をいただいた。以上より, ツールの適切な組み合わせを行うために繰り返しツールをセットする必要があった課題においては, 組み合わせや処理ツールと可視化ツールの違いがわかりにくかった比較システムよりも, 見やすさでは個人差がみられたものの組み合わせを適切に示せた提案システムの方が, ツールセット回数の削減において, 有効であったと考えられる。

(4) 事後アンケート

実験終了後に, 両群の被験者に「使用する『処理ツール』もしくは『可視化ツール』をパネルにセットするのは簡単でしたか？」

というアンケートを行ったところ, 提案システム群と比較システム群それぞれについて「簡単だった」と回答した人数と, 「難しかった」と回答した人数は, 10:2 と 7:5 であった。

その結果について, 群ごとに二項検定を行った結果, 提案システム群でのみ有意傾向がみられた。よって, 提案システム群では, ツールのセットが簡単と感じる傾向があったといえる。これは, 提案システムのツール表示エリアにあるグラフィカルに並べられたツールから適切なツールを選択し, ドラッグアンドドロップでセットエリアにセットするという操作が, 比較システムの羅列されたツール名から適切なツールをクリックしてセットするという操作に対し, 直感的で操作しやすかったためと考えられる。実際, 提案システムを使用した被験者から, 「ツールをドラッグしてセットするのが楽」という意見をいただいた。

5. 結論

テキストマイニング初心者のツールの選択を支援するために, TETDM と連動して複数のパネルにツールをセットできるインタフェースを提案した。さらに, 比較実験を行うことにより, このインタフェースの効果を検証した。その結果, 本インタフェースは, 迅速に適切なツールのセットを行う上で有効であることを確認した。

今後は, インタフェース上からツール名を探しやすいツール一覧の表示方法のさらなる追究を行う。さらに, 本実験において, 用意されているツール名を見るだけでは, どのような処理が行われるのかわかりにくい, それぞれのツールに付属している説明を読んでもわかりにくいという意見をいただいたため, 初心者がより直感的で簡単に, たくさんのツールの中から目的に応じたツールを選択できるようなインタフェースを提案し, 本システムの実用化を目指したい。

参考文献

- [砂山 2013] 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: テキストデータマイニングのための統合環境 TETDM の開発, 人工知能学会論文誌, 28(1),1-12, 2013
- [砂山 2001] 砂山渡, 谷内田正彦: 展望台システムによる複数文書の要約と Web ページ集合への適用, 一般社団法人情報処理学会, 2001(86), 57-62, 2001
- [山手 2012] 山手砂都美, 砂山渡: 文章の話の組み立てと展開速度による段落間関係の評価, 第 26 回人工知能学会全国大会, 3K2-NFC-3-4, 2012
- [梶並 2012] 梶並知記: TETDM を用いた関連 Tweet 探索の一手法, 第 26 回人工知能学会全国大会, 3K2-NFC-3-7, 2012
- [塩澤 1997] 塩澤秀和, 西山晴彦, 松下温: 「納豆ビュー」の対話的な情報視覚化における位置づけ, 情報処理学会論文誌 38(11), 2331-2342, 1997
- [牧田 2006] 牧田裕喜, 川上拓郎, クアンプ, 佐々木整: Project Looking Glass を用いた情報視覚化ツールの開発, 電子情報通信学会技術研究報告. ET, 教育工学 106(437), 15-20, 2006
- [鈴木 2000] 鈴木亮一, 長谷川忍他: WWW 上における学習パスプランニング支援環境の構築とその評価, 教育工学関連学協会連合全国大会講演論文集, 6(1), 255-258, 2000