

スムーズなテキスト分析を支援する音声ナビゲーション

Voice Navigation to Support Smooth Text Mining

佐藤 允哉^{1*} 砂山 渡²
Masaya Sato¹ Wataru Sunayama²

¹ 滋賀県立大学大学院 工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学 工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: In recent years, the sophistication of information and communication technologies has led to a continuous accumulation of data, but the training of data scientists to perform data mining has not kept pace. In addition, to learn the flow of data mining, it is necessary to understand the mining procedure, but it is currently difficult to learn this by oneself. In this study, we propose a voice navigation system using TETDM, a text mining tool, to assist those who are not familiar with data mining in smoothly executing the mining procedure, from determining the purpose of the mining to obtaining knowledge. The experimental results show that the proposed voice navigation system is effective in supporting the smooth execution of the text mining procedure.

1 はじめに

近年、IT 技術の発展によってデータのやり取りは紙媒体ではなくデジタルデータとして取り扱われることが多くなっている。その影響でビッグデータと呼ばれる膨大な量のデータが蓄積され続けている。それらのデータの中には、顧客アンケートの結果や、SNS に書き込まれた製品の評価といった企業戦略や商品開発に使えるような重要なデータも含まれている。

また、これらのデータには数値データやテキストデータと様々な種類のデータが含まれており、分析を行う際にはそれらのデータの種類に応じた分析手法をとる必要がある。中でも、テキストデータの分析は、膨大な量のテキストデータを読まなくてはならず、尚且つテキストデータの種類によっても分析の手法が変わってくるため解析の目的にあったデータ取得方法を考える必要がある [1]。

これまでにも、テキスト分析のツールの使い方や分析の手順のチュートリアルによって、テキスト分析の基本的な知識を身につけさせるシステムが提案されてきた。しかしながら、チュートリアルだけでは、テキスト分析の手順やツールの操作方法について理解しにくいことが存在し、それらのつまづきによってテキスト分析がスムーズに行われない可能性がある。

そこで本研究では、テキスト分析を行うための統合環境である TETDM に新たな機能として、音声ナビゲーションによるテキスト分析支援を行う機能を実装する。この音声ナビゲーション機能に従いながらテキスト分析を進めることで、ツールの使い方や次に行う動作が理解しやすくなり、これらによってテキストデータから有益な情報をスムーズに取り出すことを目的としている。

2 関連研究

2.1 データサイエンスのスキル獲得支援の研究

データサイエンスのスキル獲得を支援する研究 [2] がある。この研究では、初心者ユーザを対象に、テキスト分析スキルの獲得を目的としたチュートリアルシステムを開発し、それらによって基礎的なスキルから応用的なスキルまで順にテキスト分析課題を解かせることで、ユーザの一定のテキスト分析のスキル向上により支援を行った。しかしながら、これらの研究ではチュートリアル用としてあらかじめ用意されたテキストで行うので、ユーザが分析したいテキストとは同じように進められない可能性がある。そこで本研究では、実際に分析を行いながら音声ナビゲーションによるテキスト分析支援を行うことで、手順やツールの操作方法を

*連絡先：滋賀県立大学大学院工学研究科 電子システム工学専攻
佐藤 允哉
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: os23msato@ec.usp.ac.jp

習得してもらうことを目的とした。

2.2 音声ナビゲーションによる操作支援

テキストや事前の説明だけでは理解が容易ではない事項を効率的に学習させるために、音声ナビゲーションを用いた電子教材を制作することを提案した研究 [3] がある。この研究では、音声ナビゲーションとアニメーション、つまり音声情報と映像情報の 2 つを用いて通信工学入門科目の電子教材の内容の理解の支援を促すとともに、音声情報の違いによる学習者の反応を測定するために約半分の教材はシミュレーションに関する技術的解説とシミュレーション操作方法の両方を音声でナビゲーションを行った。

また、残りの半分の教材は、シミュレーション操作方法のみを音声でナビゲーションを行った。本研究では、映像情報は用いていないが音声ナビゲーションを用い理解を促すと言う点で共通している。

3 テキスト分析

3.1 テキスト分析のための統合環境 TETDM

レビューやアンケート調査の記述部分、SNS の書き込み、ネット掲示板の書き込みといった様々なテキストを分析するための分析ツールを提供し、これらを統合的に扱えるようにすることで分析に没入し新たな発想を得られる環境である TETDM(Total Environment for Text Data Mining)[4][5] がある。TETDM では、レビューやアンケート調査の記述部分、SNS の書き込み、ネット掲示板の書き込みといった様々なテキストを分析するための分析ツールを提供し、これらを統合的に扱えるようにすることで分析に没入し新たな発想を得られる環境の構築を目指している。

3.2 テキスト分析の手順と流れの概要

本節では、テキスト分析の手順と流れの概要について述べる。テキスト分析の流れの図を図 1 に記す。

テキスト分析手順にはデータ分析プロセスと、知識創発プロセスがある。

データ分析プロセスは、データを用意するところからデータから得られた事実を集めるまでの流れを表す。「0. 蓄積データ」で手元に蓄積されているデータから、「1. 分析目的の決定」で分析の目的を決定する。決定した目的に対して、必要なデータが蓄積データだけでは不足している場合、「2. データ収集」で分析に必要なデータを集める。そして、集めたデータを「3. データ整形」で分析ツールに入力できる形式に整形を行い、

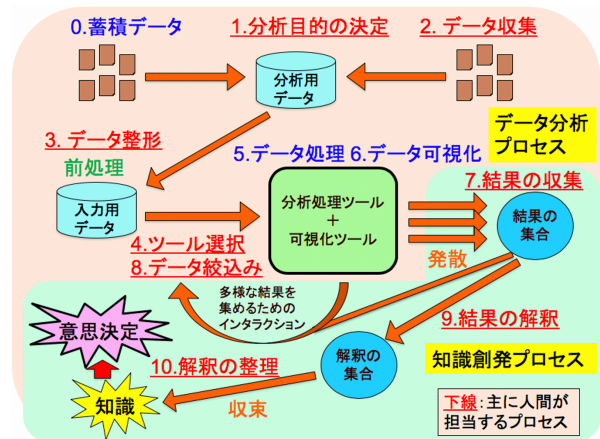


図 1: テキスト分析の流れ

「4. ツール選択」で分析に用いるツールを選択し、「5. データ処理」と「6. データ可視化」で、データ分析処理とその結果の出力の表示を行う。「7. 結果の収集」で分析結果の中で注目する結果を集め、「8. データ絞り込み」は、より多様な結果を集めるために行う。

知識創発プロセスは、データから得られた事実を、意思決定に活用できる知識としてまとめるまでの流れを表す。「7. 結果の収集」で結果を集め、「9. 結果の解釈」で結果から分かる事実を解釈として取り出す。「10. 解釈の整理」で集めた解釈を整理して知識としてまとめる。

4 テキスト分析のための音声ナビゲーション

本章では、音声ナビゲーション機能の概要と前節で示したテキスト分析手順のそれぞれ手順における音声ナビゲーションの実装の方針を述べる。この音声ナビゲーションに従ってテキスト分析を進めていくことでユーザはテキスト分析中につまづきにくくなり、スムーズにテキスト分析を進められると共に分析手順の理解を支援すること目的として音声を準備した。

4.1 音声ナビゲーションの発話条件

本節では、音声ナビゲーションの発話条件について述べる。図 2 に本システムの構成図を記す。発話条件としては、前章で述べたテキスト分析手順の順番で分析が進んで行くものとして、それぞれの手順に必要なボタン操作やテキストの入力が行われることで、図 2 のようにその場面に沿った音声ナビゲーションが発話されるようになっている。例えば、マイニングツールをセットした場合には、セットしたツールの使い方の説明や着目すると良いポイントをナビゲーションとし

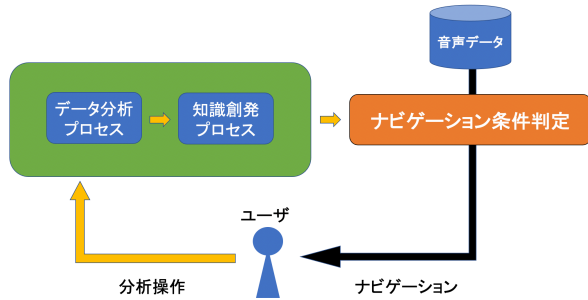


図 2: 音声ナビゲーション機能のシステム構成図

分析手順	つまづきやすいポイント	つまづきを解消するナビゲーション
1.分析目的の決定	分析の目的が決められずテキスト分析を始められない	分析を試しながら目的を決める感じでもいいと思います。
2.データ収集	テキストデータの <input/> 入力方法がわからない	メニューウィンドウの <input/> ファイルボタンから、テキストを入力して下さい。
3.データ整形	入力されたテキストが <input/> 分析に適した整形がされておらず良い結果が得られない	文の数が少ないように思います。テキストデータに句点を追加するか、 <input/> キーワード設定で、 <input/> 句点の記号を設定して下さい。
4.ツール選択	最初どのツールを使ったら良いかわからない	処理ツール「テキスト評価アプリケーション」の「 <input/> 単語情報」ボタンを押して、ツールをセットしました。
5.データ処理・6.データ可視化	選んだツールの使い方、見方がわからない	選んだツールの説明を行う。
7.結果の収集	結果の見つけ方がわからない	結果を見つけるコツは予想と異なる出力に着目することです。
8.データ絞り込み	データ絞り込みを行う <input/> タイミングが分からない	データの絞り込みを行っていませんか？
9.結果の解釈	結果の解釈の <input/> 書き方が分からない	結果の解釈は、結果と解釈の登録ウィンドウで、 <input/> 入力した結果から考えられることを解釈として入力して行います。
10.解釈の整理	解釈の統合の際に <input/> 着目しながら進めていくか分からない	意味が似ているなど、何らかの共通点をもつ解釈同士を結合させていきます。薄い線で結ばれた解釈には、同じ単語が使われているので参考にして下さい。

図 3: テキスト分析の各手順におけるつまづきやすいポイント

て発話する。つまづきやすいポイントとそれに対する音声ナビゲーションの例を図 3 に記す。

4.2 音声ナビゲーションの内容

本節では、それぞれのテキスト分析手順毎に音声ナビゲーションの内容を述べる。

4.2.1 分析目的の決定

「分析目的の決定」では、分析の目的を決めていないユーザーに対して「分析対象から知りたいことは何か、を考えると決めやすいですよ。また、分析を試しながら目的を決める感じでもいいと思います。」とナビゲーションすることで分析の目的を決めにくい分析初心者に対して目的決定の促進と決定方法の説明を行う。

4.2.2 データ収集

「データ収集」では、テキスト分析に用いるデータが足りない場合に、データを集めてくるように促せるように意図してナビゲーションを行う。例えば、「ではまず、分析に必要なデータを集めてきてね」とナビゲー

ションを行うことでデータ収集を促す。また、TETDMのテキストエディタに文章が入力されていない時はテキストの入力方法を説明した上でテキスト入力を促す音声ナビゲーションを行う。

4.2.3 データ整形

「データ整形」では、入力された文章の文の数やセグメントの数の評価を行うことを意図してナビゲーションを行う。評価方法は、文の評価は、文の文字数の平均が 100 文字以上の場合、文ごとの文字数が多すぎるので「テキストデータに句点を追加するか、キーワード設定で、句点の記号を設定してください」とナビゲーションを行うことで文ごとの文字数を減らし、文ごとの特徴をより顕著に表れやすくする。

4.2.4 ツール選択

「ツール選択」では、表 1 の音声ナビゲーションを各発話条件に従って行う。これらの音声は、分析に用いるマイニングツールの選択で初心者が行き詰まらないように意図してナビゲーションを行う。具体的には、「目的に応じたツールを、いろいろと選んで試して下さい」とナビゲーションを行うことでユーザーの分析の目的に沿ったものを簡単にパネルにセットできる。これらのツールセットは、TETDM に実装された処理ツールと可視化ツールを選択した目的に応じて最適な組み合わせで自動でセットを行う。

4.2.5 データ処理・データ可視化

「データ処理・データ可視化」では、これらの音声は、マイニングツールを選ぶことができたが、その使い方がわからず行き詰まってしまうように意図してナビゲーションを行う。具体的には、主題関連語評価が選択された場合「文章中の単語の主題との関連を評価します。ボタン操作で段落ごとに評価することができます」とナビゲーションを行うことでツールの簡単な使い方や見方の説明を行う。

4.2.6 結果の収集

「結果の収集」では、結果の見つけ方がわからず、初心者が行き詰まってしまうように意図してナビゲーションを行う。具体的には、何か気になる結果があった時は、メニューウィンドウの知識創発の結果ボタンを押して、結果と解釈を書き込みましょう。結果を見つけるコツは、予想とは異なる出力に着目することです」とナビゲーションを行うことで、結果の見つけ方を説

表 1: 「ツール選択」における音声ナビゲーションの内容

項番	音声ナビゲーションの内容	音声ナビゲーションの発話条件
4-1	それでは、分析のためのツールをセットしましょう。	ツール初回セット時
4-2	分析では、特徴的な、単語、文またはセグメントに着目します。	4-1 の音声が終わった
4-3	単語に着目しますか？(Yes or No)	4-2 の音声が終わった
4-4	処理ツール「テキスト評価アプリケーション」の「単語情報」ボタンを押して、ツールをセットしました。	4-3 の選択肢で「Yes」が選択された
4-5	文やセグメントに着目しますか？	4-3 の選択肢で「No」が選択された
4-6	処理ツール「テキスト評価アプリケーション」の「文・セグメント情報」ボタンを押して、ツールをセットしました。	4-5 の選択肢で「Yes」が選択された
4-7	起動モードをライトモードにして、メニューウィンドウの「目的」ボタンを押すと、目的に応じたツールを選びやすくなります。	4-5 の選択肢で「No」が選択された
4-8	ライトモードに変更しますか？(Yes or No)	4-7 の音声が終わった
4-9	まずは、処理ツール「テキスト評価アプリケーション」を使って、いろいろなツールを試してみてくださいね。	4-5 の選択肢で「No」が選択された
4-10	どんな分析をしたいですか？	4-5 の選択肢で「No」が選択されたかつモードがライトモード以降
4-11	目的に応じたツールを、いろいろと選んで試してみてください。	4-10 の音声が終わった

明を行う。また、「結果と解釈の入力だけでなく、その結果を得られたパネル番号と重要度の選択も忘れずに行いましょう」とナビゲーションを行うことで忘れやすい操作を忘れにくくする。

4.2.7 データ絞り込み

「データ絞り込み」では、データ絞り込みを行うタイミングがわからない、または絞り込み機能を忘れている場合に絞り込み機能を使うことを促進する意図でナビゲーションを行う。具体的には、「データの絞り込みを行ってみませんか？データの絞り込みを行って、一部のデータに対する出力を見ることでも、新しい結果を集めることができますよ」とナビゲーションすることで絞り込み機能の存在と使うタイミングを明示する。

4.2.8 結果の解釈

「結果の解釈」では、結果からどのように解釈をすれば良いかわからず初心者が行き詰まらないよう意図してナビゲーションを行う。具体的には、「結果の解釈は、結果と解釈の登録ウィンドウで、入力した結果から考えられることを解釈として入力して行います」とナビゲーションを行うことでユーザ自身が感じたことをそのまま記入すればいいと分かるように説明する。

4.2.9 解釈の整理

「解釈の整理」では、結果の解釈を統合させるにあたって、どこに着目して統合させたら良いかわからず行き詰まらないよう意図してナビゲーションを行う。具体的には、「意味が似ているなど、何らかの共通点をもつ解釈同士を結合させていきます。薄い線で結ばれた解釈には、同じ単語が使われているので参考にしてください」とナビゲーションを行うことで統合させるときの着目点を説明する。

5 音声ナビゲーション機能によるテキスト分析支援の検証実験

5.1 実験設定

音声ナビゲーション機能によってテキスト分析をスムーズに行えるかどうかを評価するために、テキスト分析に不慣れな理系の大学生 18 名に、音声ナビゲーション機能を利用しながらレビューテキストの分析を行ってもらった。

まず、TETDM のチュートリアルを一通り被験者に行ってもらい、TETDM の基本的な使い方や最低限のツールの使い方、結果と解釈の書き方を理解してもらう。このチュートリアルでは、あらかじめ用意されたテキストデータを使って、指示された通りに TETDM の操作を行っていくことで、テキスト分析の手順それぞれの基礎的な知識を身につけることができる。

チュートリアル終了後、チュートリアルに要した時間が均等に分かれるように被験者をグループ A、B の 2 グループに分ける。その後 3 つのレビューデータ [6] でテキスト分析を行ってもらった。グループの振り分けとテキストと音声ナビゲーションの使用の有無の振り分けを表 2 に記す。

グループ A、B 共に練習として改善案のアイデアの提案を分析の目的としてワイヤレスイヤホンのレビューデータのテキスト分析を音声ナビゲーション機能なしで行ってもらった。この際に、集めてもらう結果と解釈の個数は 10 個以上、実験時間は 30 分から 1 時間を目安に行った。

2 つ目はスマートスピーカーのレビューデータのテキスト分析を行ってもらった。この際に音声ナビゲーション機能の有効性を調べるためにグループ A は音声ナビゲーション機能なしで分析を行い、グループ B は音声ナビゲーション機能ありで行った。

3 つ目はタブレット端末のレビューデータのテキスト分析を行ってもらった。今回は、グループ A は音声

ナビゲーション機能ありで分析を行い、グループ B は音声ナビゲーション機能なしで分析を行った。

表 2: 実験のグループ振り分け

グループ	手順 1 イヤホン (練習)	手順 2 スマート スピーカー	手順 3 タブレット 端末
A	ナビなし	ナビなし	ナビあり
B	ナビなし	ナビあり	ナビなし

5.2 音声ナビゲーション機能によるテキスト分析のスムーズさの評価

TETDM を操作している時間を除いたテキスト分析全体にかかったグループ毎の平均時間を図 4 に示す。図 4 から、練習テキストでの分析時間平均は、グループ A では 55 分、グループ B では 62 分という結果になった。また、実験時間はどちらのグループでも同様にナビゲーション機能ありで実験を行ったところ 35 分、ナビゲーション機能なしの場合が 42 分という結果が得られた。

この結果から、ナビゲーション機能がある場合の方が、なしの場合よりも平均 7 分、分析を早く終えることができていることがわかるので、この結果から音声ナビゲーション機能は、入力されるテキストの種類によらずスムーズなテキスト分析を支援できていることがわかった。

一方でグループ B のナビゲーション機能なしの時間がナビゲーション機能ありの後にも関わらずグループ A のナビゲーション機能ありの場合と同じ時間になった。しかし、練習テキストの時間をグループ A, B で比べると B の方が時間が長くテキスト分析が苦手な人が多いと考えられるので、グループ B でも音声ナビゲーションの効果によりテキスト分析の手順理解に役立ったと考えられる。

5.3 音声ナビゲーション機能の効果

分析をスムーズに行えることがわかったが、実際にテキスト分析のどこの部分でナビゲーション機能の効果が大きかったのかを調べるために、一番実験時間に影響が大きいと考えられるツール選択のプロセスに着目して、結果と解釈の登録に関係のないマイニングツールが使われた時間を調べ、音声ナビゲーション機能がある場合となしの場合で分けて結果を昇順に並べたものを図 5 に示す。図 5 の真ん中より左側の結果登録に関係のないツールが使われた時間が少ない場合では、ナビゲーション機能がなしの場合、0 が多い結果が得られた。

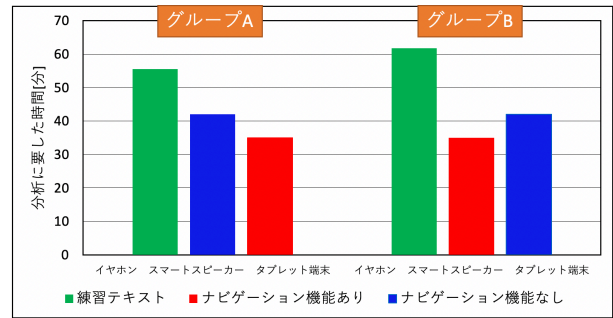


図 4: グループ毎のテキスト分析にかかった平均時間

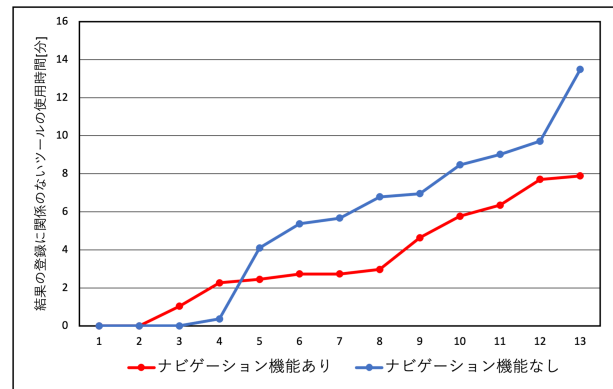


図 5: 結果登録に関係のないツールが使われた時間 (時間順にソート)

これは、分析操作に慣れていて、あまりツールを使わないためナビゲーションの効果小さかった体と考えられる。一方で図 5 の真ん中より右側では、結果登録に関係のないツールが使われた回数は、ナビゲーション機能がある場合がなしの場合よりも少ない結果が得られた。主に結果登録に関係のないツールを、多く使うのは、テキスト分析に慣れていない初心者だと考えられるので、本研究の音声ナビゲーション機能は、テキスト分析初心者に対して特に効果的であることがわかった。また、ツールを多く使うユーザに効果が大きく見られたので、音声ナビゲーション機能で特に効果的であったテキスト分析プロセスは、ツール選択であることがわかった。

6 おわりに

テキスト分析において、スムーズに分析を行えるように音声ナビゲーションを用いて分析の支援を行った。本研究のナビゲーション機能が、テキスト分析を行う際に分析をスムーズに行えるかを検証するために、音声ナビゲーション機能を用いて実際にテキスト分析を行い検証実験を行った。実験の結果から、音声ナビゲーション機能を用いた場合、なしの時に比べて分析にかかる時間が短くなる傾向が得られ、スムーズにテキス

ト分析を行うことができた。また、今回実装した音声ナビゲーションの中で最も効果が大きかったテキスト分析プロセスはツール選択であることがわかった。

今後の課題として、テキスト分析初心者だけでなく基本的な知識を持っているユーザに対しても有益なナビゲーションの実装を検討していきたい。

参考文献

- [1] 保田 明夫: テキスト・マイニングの概要, 電気学会論文誌 C (電気・情報・システム部門誌), Vol.125, No. 5, pp. 682–689(2005)
- [2] 西原 陽子, 中垣内 李葉, 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)
- [3] 今澤 達哉, 稲垣 貴俊, 野須 潔: 音声ナビゲーション付きシミュレーション型通信工学入門電子教材の制作とユーザビリティ評価, 映像情報メディア学会誌, Vol.59, No.1, pp. 65–68 (2005)
- [4] 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)<https://tetdm.jp>
- [5] 砂山渡, 高間康史, 徳永秀和, 串間宗夫, 西村和則, 松下光範, 北村侑也: 統合環境 TETDM を用いた社会実践, 人工知能学会論文誌, Vol.32, No.1, NFC-A, pp.1–12 (2017)
- [6] Amazon.com, Inc.,(URL)www.amazon.co.jp