電子投票のための意思決定支援システムにおける アンケート作成支援システムの試み

藤村 春輝 凌 暁萍 ‡

†神奈川工科大学大学院 情報工学専攻 〒243-0292 神奈川県厚木市下荻野 1030 ‡神奈川工科大学 情報学部 〒243-0292 神奈川県厚木市下荻野 1030

E-mail: † s1585013@cce.kanagawa-it.ac.jp, ‡ ling@nw.kanagawa-it.ac.jp

あらまし 公職選挙法の改正により選挙の候補者はウェブサービスを利用した選挙運動が可能となった.このことから本研究チームは候補者情報が乏しい有権者のために役立つ意思決定支援システムを提案・試作した.だが,その一機能で扱うアンケート文の内容を作るのは人であるため,内容の客観性が損なわれる可能性がある.そこで本論文では自動的に質問文を生成するために必要な文章要素を生成する方法を提案し,人の手でアンケートを作成することを支援するアンケート作成支援システムを試作する.また,アンケート作成実験の結果を示し,考察することで今後の課題を導き出す.

キーワード テキスト分析, 意思決定支援, 電子投票

1. はじめに

候補者の情報が乏しい有権者のために役立つ意思 決定支援システムの試作・実験・評価を行ったが[1][2], このシステムが備える候補者マッチング機能で用いる アンケート文の内容は人の手で作られるため、国政選 挙への適用に当たり、これが適切であるか、特定の候 補者に有利になるような内容になっていないか、など の問題がある.

ここで,アンケート(調査)とは,質問紙調査とも 言い、各個人の意見、態度、価値観、感情、性格、行 動傾向などといった様々な事柄を,質問文とそれに対 する回答という言語活動から測定する方法である. ア ンケートは、「あなたは、春夏秋冬のうちどの季節が好 きですか?」といった具体的な質問から、「歩きスマホ をしている人に、あなたはどのように注意しますか?」 というような, 仮想状況を設定した質問まで取り扱う ことができるため、比較的広範な問題を扱うことがで きるという長所を持つ.一方で,回答者の言語的理解 力により,正しい回答が得られない,嘘の回答を行う という可能性があるという短所も持っている. アンケ ートの回答方法は、大きく分けて次の4種類に分類で きる.一つ目は、回答者に文章で回答してもらう自由 回答法.二つ目は、質問項目に対し複数の選択肢を用 意し,回答者にはその選択肢の中から選んで回答して もらう多肢選択法. 三つ目は、複数の項目について順 位をつけてもらう順位法. 4 つ目は, 例えば,「1. 非 常に好き~5. まったく好きでない」の中から選んでく ださい. のような程度や頻度などについていくつかの 段階を設定し、その中からひとつ選択してもらう評定 法である[3]. また, アンケートを作成するためには, 何を調べたいのかを明確にする必要がある.その上で, 質問項目が事実, もしくは行動, あるいは意識につい て尋ねるものかにも留意して作成する必要がある.

本研究でのアンケートとは、利用者を特に特定せず、質問文も政治から、趣味まで、幅広い問題を取り扱う.また、例えば、「憲法改正についてあなたはどう思いますか」や、「あなたが好きな乗り物は何ですか」、といったようにアンケートの質問文についてのみ取り扱う、本研究では質問文のみを取り扱うため、回答方法についてはユーザが決定権を持つが、例に挙げた質問文のような多肢選択法、もしくは評定法を選定することを想定している.

上記の点を踏まえ、人の手を介さずに文章を作成すれば、少なくとも故意に特定の候補者に有利になるような文章にはならないという考えの基、自動アンケート作成システムを考案した。ここでいうアンケート作成とは、質問文を自動的に作成するという意味である。これは、質問文さえ決まれば、回答方法や選択肢も自然に決まるだろうという考えからである。また、今回扱う言語は日本語に限定する。

本論文では、第2章で意思決定支援システムにおける意思決定支援の定義づけをし、第3章で課題を解決するための手法を提案する。第4章で提案システムの設計と構成を解説し、第5章でシステムの動作テストについて述べる。第6章で考察をまとめ、第7章で結論を結ぶ。

2. 意思決定支援と意思決定干渉

意思決定支援システムにおける「意思決定支援」が エキスパートシステム等の分野でどの位置に当たるか を改めて定義する.

意思決定者の意思決定に干渉する人を「支援者」. 支援者が意思決定者の意思決定にどの程度干渉してよいかの許容範囲を「干渉許容度」(以降 IFTTR と記す。) と定義すると、IFTR は 0~8 の全 9 レベルに分けることができる. IFTR の各レベルの定義とその事例を表 1 に示す.

また、IFTR は、「意思決定への干渉が許されない」グループ、「意思決定への干渉が許される」グループ、「条件付きで意思決定への干渉が一部許される」グループの3つのグループに大別できる。条件付きで意思決定への干渉が一部許されるグループではさらに、「生情報の提供のみが許される」グループと「分析・加工した情報の提供が許される」グループの2つのグループに細分化できる。各 IFTR と意思決定支援システムの3つの機能の位置関係を図1に示す。

表 1 IFTR の定義および事例

IFTR	表 1 IFTKの定義お。 定義	事例
0	支援者は,意思決定 者の意思決定にい。 してはなら、意思 しては場合,意見 定者の利益を う。	国政選挙
1	支援者は基本的に意思決定者のは決定に干渉してはなられい。 干渉しては場合, 調査結果の信憑性が 損なわれる。	アンケート調査
2	意思決定者の意思に より指定された生情 報の提供は許され る.	書籍検索システム
3	意思決定者の参考と なる生情報の提供は 許可される.	UpToDate ^[4]
4	意思決定者の意思に より指定された生情 報に対し分析・加工 をした結果を提供す ることは許される.	scrobo ^[5]
5	意思決定者の参考と なる生情報に対け分析・加工をした結果 を提供することは許 される.	OpinionReader ^[6]
6	意思行分野変素を 定事時では を専門するの をでいる をでいる をでいる をでいる をでいる でいる でいる でいる でいる でいる でいる でいる	エキスパート システム
7	意思決定者の不利益 にならないならば, 支援者は意思決定者 の意思決定に干渉す ることが許される.	Google 検索
8	支援者は意思決定者の意思決定に干渉することが許される.	CM, 広告

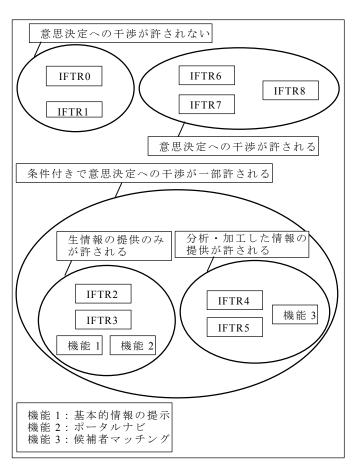


図 1 IFTR と意思決定支援システムの位置づけ

意思決定支援システムの候補者マッチング機能は、アンケートを利用する.ここで最も心配されることは、アンケートの内容により有権者の投票意思が誘導されてしまうこと、すなわち有権者の意思決定に干アンケートの客観性を保つため、アンケートの作成が自動的に生成する方法や、有権者の意思により作成する方法や、有権者の意思により作成する方法や、有権者の意思により作成するために参考となるキーワード文を文章要素としてザスすることを提案する.

3. 課題解決手法の提案

アンケートを自動的に作成する方法について、坂本らによる研究[7]や、佐藤らによる研究[8]などがあるが、ウェブを利用するという点では、同じような考え方だが、どれもアンケートそのものの作成を自動化する、もしくはアンケート作成の支援を行うというものであり、本研究で考えるアンケートの質問文の作成を自動化するという目的には沿わない。質問文の自動生成についても、國近ら[9]や百々ら[10]などの研究があるが、英文の自動生成である、想定に近いものではあるが、

エキスパートシステムのようになっているため、応用させにくいなど、今回の目的には沿わない。また、ウェブ上に CREATIVE SURVEY[11]や Questant[12]など、アンケート作成支援サービスが存在するが、これらはアンケートフォームの作成や結果の集計を支援するものであり、本研究で考えるアンケート文の自動生成ないしは作成の支援をするという目的からは外れている。

以上のように、既存研究やサービスでは、本研究の目的とする、特定の意見に誘導するような内容になっていない日本語のアンケートを生成する仕組みは整えられない。よって、本研究では特定の人物や団体の思想が反映されていない、もしくは反映されにくい文章群として、Twitterや FaceBookのような一般の人たちの利用者数が多い SNS を利用し、特定の意見に誘導するような内容になっていない日本語のアンケートを生成するシステムを考案した。本研究では、利用する SNSとして Twitter を採用する.

具体的には、次の4プロセス方式を提案する.

【プロセス1】共起語導出

<目的> アンケート文に使うキーワードに関するツイートを検索する. 検索結果に含まれる無関係なツイートを排除するために, 検索結果に共通する単語である共起語を導き出す.

<方法> Twitter API を利用し、ツイートを検索する. 検索結果を形態素解析することで共起語を導き出す. 本研究では、検索結果は最大 1000 ツイート、検索対象 は日本人アカウントのみとする. 共起語の検索範囲は、 キーワードの前後 5 単語とし、共起語が複数存在する 場合は最大 10 単語を導出結果とする.

【プロセス 2】共起語関連ツイート抽出

<目的> 検索結果に含まれる明らかに無関係なツイートを排除する.

<方法> プロセス1の検索結果のうち、同じくプロセス1により導出された共起語が文章内にあるツイートのみを検索結果のツイート群から取り出す. プロセス1で導出された共起語が複数ある場合は、ツイート内に1語以上の共起語があれば、そのツイートを検索結果から取り出す.

【プロセス3】分かち書き

<目的> 文章要素作成で採用されたマルコフ連鎖の ための前処理をする.

<方法> プロセス2で取り出したツイート文を形態素解析し、3つの形態素を一組とする分かち書きにする。

【プロセス4】文章要素作成

<目的> 大量にあるツイート文の要約をする. アンケート文を作成するための参考になる文章要素を作り出す.

<方法> プロセス3の結果に対し、マルコフ連鎖を 適用することで新たな文章要素を作り出す. 新規文章 要素内にはツイート検索に用いたキーワードが必ず含 まれるようにする. 複数の文章要素を作成する際は、 同じ文章要素が結果として提示されないようにする. ここで、マルコフ連鎖について簡単に説明する.マルコフ連鎖とは、未来の状態は現在の状態のみによって決まり、過去の状態には関係しないという確立の状態のことである.応用例の一つとして文章の要約が挙げられる.本研究では、複数のツイート文の要約結果を新たな文章要素として提示することになる.

マルコフ連鎖以外にも、自動文章作成技術の応用例として、言語処理分野への応用は研究段階ではあるが、人工知能技術の一種であるディープラーニングが挙習られる。ここでは詳しい説明を避けるが、大量の学習ボータを入力することで人工知能が自動的に特徴。今回とかりできる。しかし、学習がより人間らしい文章を結果として出力することが期待できる。しかし、学習デーとして出力することが期待できる。しかし、学習が上との数や学習時間がどれほど必要なのか予測が立たないことや、コストパフォーマンスの観点からみて本当にから、今回はディープラーニングの採用を見送った。

本研究では、自動的に質問文を生成するために必要な文章要素を生成する方法として先に述べた4プロセス方式を提案した.この4プロセス方式を利用することで、人の手でアンケートを作成することを支援する。本研究では、アンケート作成支援システムを試作する。本研究では、アンケート作成支援システムの試作後、システムを用いた体験コーナーを文化祭で開催することで、動作テストを行い、参加者からいただいた感想や意見から、各システムの問題点や改善案を発見し、考察することで今後の課題を導き出す.

4. 設計·構成

4.1. 開発環境

表 2 開発環境

OS	Windows 7 Enterprize
CPU	Intel Core i5-4460
メモリ	8.00GB
ウェブサーバ	Apache 2.4.17
データベース	MySQL 5.0.11
使用言語	HTML, PHP 5.6.23, R 3.3.1
API ほか	Twitter API, abraham/twitteroauth, Yahoo!Japan Web API, MeCab, RMeCab
統合開発環境	Eclipse 4.3 Kepler plaiades all in one

アンケート作成支援システムの試作に利用した開発環境を説明する.まず、開発環境をまとめたものを表2に示す.

ウェブサーバ,データベース,PHPに関しては,管理の手軽さからXAMPPを用いた.

今回は、プログラム上からツイッターを利用するため、Twitter API を利用し、OAuth 認証用のライブラリとして、情報が多く存在する abraham/twitteroauth を選

択した. 日本語形態素解析ツールは、 MeCab および RMeCab を利用した. また、作成した文章要素の文法 に誤りが存在する場合に、ユーザに指摘するために Yahoo!Japan Web API を用いた.

4.2. アンケート作成支援システムの構成

アンケート作成支援システムは,3 つの処理プログラムから構成される.

- 1. ツイート検索プログラム
- 2. 共起語導出と分かち書きのための形態素解析プログラム
- 3. マルコフ連鎖による文章要素生成プログラム

また,2つのユーザーインターフェース(表示画面) が設けられている.

- 1. Twitter 用検索フォーム
- 2. 生成文章要素の表示

全体の処理の流れを図2に示す.

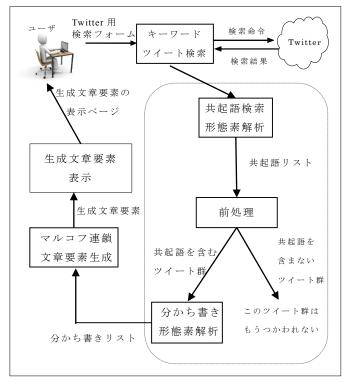


図 2 アンケート作成支援システムの概念図

4.3. Twitter 用検索フォーム

まず、ユーザはこの画面で、ツイート検索のキーワード、共起語検索のキーワード、1 リクエストあたりのツイート数、リクエスト数、作成する文章要素数を入力する.また、オプションとして、ツイート文を整形する際に用いる正規表現や、特定の文字列が存在するツイートを除外する設定、形態素解析の際に、形態素を終止形に変換するかどうかも指定する.

ツイート検索と、共起語検索のキーワードを別に指定するのは、例えば、「アメリカ大統領選挙」と検索し

ようとすると、共起語を導き出す段階で「アメリカ」、「大統領」、「選挙」と分かれてしまい、「アメリカ大統領選挙」という単語は存在しないと判断され、正常に動作しなかったためである.

検索ツイート数は,「1 リクエストあたりのツイート数×リクエスト数」である. 初期設定では, 1 リクエストあたりのツイート数は最大値である 100, リクエスト数については最大値 10 の半分である 5 としている.

作成する文章要素数の初期値は 10 としている. 文章要素数を多くしても,完成するとは限らず,サンプル数が少ないのもまた問題であるため 10 とした.

正規表現については、初期設定では、リツイートを示す「RT @~:」、ユーザのアカウント名、リンク URL、ハッシュタグ、絵文字をそれぞれ削除するという指定になっている.

ツイートの除外設定は、まとめサイトやニュースサイトなどの引用や共有ツイートが多数存在するときなど、文章要素の生成に支障が出る場合に用いる.

最後の形態素解析の際の変換については、終止形に変換すると、文章要素数が増えやすい代わりに、普段使う日本語からは離れた文章になる傾向がある。変換しない場合は、元の文章のままであるため、日本語らしい文章になりやすいが出来上がる文章数が少なくなる可能性がある。

4.4. ツイート検索プログラム

ツイート検索プログラムでは、abraham/twitteroauth を利用し、ツイート検索を行う、検索されたツイートは、正規表現に一致した部分が削除されるなどの処理が施されてから、テキストファイルに書き込まれる.また、このとき、共起語検索用キーワードも別のテキストファイルに書き込まれる.

4.5. 形態素解析プログラム

形態素解析プログラムは、機能別に共起語解析プログラムと、分かち書き生成プログラムの 2 つに分かれている.

検索された前処理後のツイートには、キーワードが 含まれているだけで、想定している話題とは無関係で あるなど不要なツイートが存在する. そこで, 共起語 というものを利用する. この共起語は, あるキーワー ドの周辺によく出現する単語のことをいう. つまり、 ツイート検索の際に用いたキーワードの共起語をもつ ツイート同士は何かしらの関連性を持つといえる. す なわち、共起語検索プログラムは、無関係であると思 われるツイートを除外し,より関連性の高いツイート を用いて文章要素の生成を行うようにするためのプロ グラムである. また, 今回共起語を検索する範囲は, キーワードの前後5単語とした.理由としては、ツイ ート文は長くても120字という短文であるため、キー ワードの前後5単語以内に、十分共通の語が出現する であろうという考えからである. RMeCab では、統計 解析でもよく使われ、共起頻度が有意であるかどうか を判断する指標となる T 値も計算される. コーパス言語学では, T 値が 1.65 以上であれば十分に有意であるとされているため $^{[14]}$, 本研究では T 値が 1.65 以上である共起語のみを使用することにした.

実際の動作内容としては、Rで共起語を検索した後、PHP側で、共起語を含むツイートのみを抜き出し、結果をテキストファイルに書き込むという動きとなる.

分かち書き生成プログラムは 4.6 節で説明する文章 要素生成プログラムでマルコフ連鎖による文文を素生 成の際に用いる分かち書きリストを生成するために用いる.このプログラムでは,表3に示す「私は増税に 反対する」という文章での例のように,3 単語ずつの 分かち書きにするが,マルコフ連鎖による文かち書きにするが,形態素ごとに分けた分かち書き でよい.3 単語ずつの分かち書きにしたのは,2 単語でつ,3 単語ずつ,と1 組の単語の数を増やすごととでいて、文章要素生成の結果が,より元の文章に近づき,不能 まごとに分けた分かち書き,つまり1単語を1組とする分かち書きでは,でたらめな文章が結果として得られることになる.

実際の動作内容としては、R でツイートを形態素ごとに分けた後、PHP 側で、3 単語ずつの分かち書きにする. また、この分かち書き生成プログラムで終止形に変換してから分かち書きをするのか、変換せずに分かち書きをするのかにより、最終的な完成文章要素の形も変わってくる.

表 3	分か	ち書	き	の例
3 X 3	JJ 14"	ソヨ	_	V D1

私	は	増税
は	増税	K
増税	K	反対
に	反対	する

4.6. 文章要素生成プログラム

文章要素生成プログラムでは、4.5 節で説明した分き 書き生成プログラムにより生成された分き リストを用いて、マルコフ連鎖を適用することでザボオートを用いている大変を作成する上での参考となる文章を表を提出である。大変を表している。大変を表している。大変を表したが、新たに文章ではなったが、新たに文章ではなったが、大変を表したが、大変を表したが、最低ではこのでは、一次ではこのでは、一次ではこの文章要素とがある。大変を表しては、大変の動作人では、大変の動作人では、大変の動作人では、大変の動作人では、大変の動作人では、大変の動作人では、大変の動作人では、大変の動作人では、大変の動作人では、大変の動作人では、マルコフ連鎖を利用することを表しては、マルコフ連鎖を利用することを表しては、マルコフ連鎖を利用することを表しては、マルコフ連鎖を利用することを表しては、マルコフ連鎖を利用することを表しては、マルコフ連鎖を利用することを表しては、マルコフ連鎖を利用することを表しては、マルコフ連鎖を利用することを表しては、マルコフェーを表しては、マルコフェーを表しては、ファルコフェーを表しては、ファルコフェーを表しては、ファルコフェーを表しては、ファルコフェーを表している。

とで文章要素を生成していく. ひとつの文章要素生成が終わると, 生成済みの文章要素と同一のものでないかのチェックが入り, 問題がなければダミー文字列の削除などが行われ, 作成済み文章要素リストに格納される. また同時に, Yahoo!Japan API の文章校正機能により校正対象があれば, 指摘文章が文章校正リストに格納される.

入力ページで指定された個数の文章要素が作成されるまで繰り返される. ただし,無限ループに陥るのを回避するための処理が組み込まれているため,必ず指定個数の文章要素が作成されるわけではない.

4.7. 生成文章要素の表示

この生成文章要素の表示インターフェースでは、プログラムが生成した文章要素とその文章要素の校正結果を本システムの実行結果として表示する.また、ユーザが文章作成する際の参考として、共起語検索プログラムで使用した共起語も共に画面に表示する.

各文章要素の左隣には、チェックボックスが用意してあり、初期状態ではチェックが入っているが、このチェックをはずすとページ下部に設けられたテキストボックスから対応した文章要素の表示が消える。このテキストボックスは、ユーザが作成した文章要素をコピーするために利用することを目的として用意されている。このテキストボックスの隣には、文章要素の校正結果をテキストボックスに表示するかどうかを切り替えるためのラジオボタンが用意されている。

5. 動作テスト

5.1. 実験手法

2016年に大学で開催された文化祭にて、本システムを体験するコーナーを開設した.体験コーナーでは一般客にアンケート作成体験をしてもらい、その後、こちらが用意したアンケートに答えてもらうという方法を取った.実験の詳しい流れは以下の通り.

- 1. 実験のため、こちらが事前に用意した設定で文章要素の作成を行う. 今回の検索キーワードは、「歩きスマホ」、「小池知事」、「東京オリンピック」、「トランプ氏」、「クリントン氏」の5つから選択する形とした. 実験用アンケート作成支援システムのトップページ、すなわち、Twitter 検索用フォームのページを図3に示す. 図4には、質問文作成のための参考となる文章要素の生成結果の一例を示す.
- 2. 作成した文章要素から、実際に質問文を 1~3 個作成してもらう. さらに、「歩きスマホと歩きタバコは法律で禁止すべきだ.」という、事前に本システムを用いて作成したアンケートに回答してもらう. 図 5 に実施の様子を示す.
- 3. 改めて、システムに対する意見や感想を問うアンケートに回答してもらう. これも図 5 にその様子を示す. 図 6 は図 5 に示す実験ページと同じページに表示されている、作成された質問文作成のための参考文章要素と導出された共起語を示している.

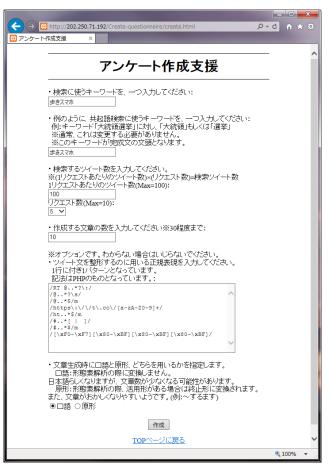


図 3 Twitter 検索用フォーム

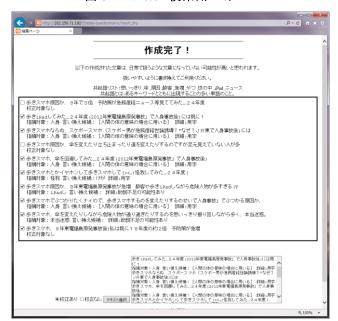


図4 生成文章要素の表示

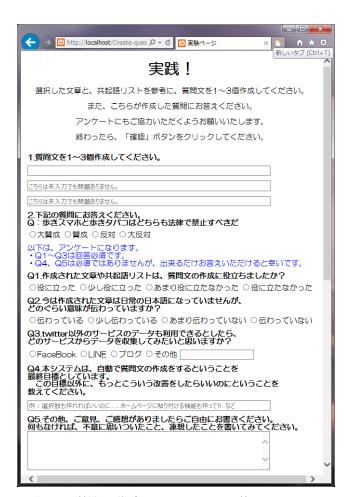


図 5 質問文作成とアンケート回答フォーム

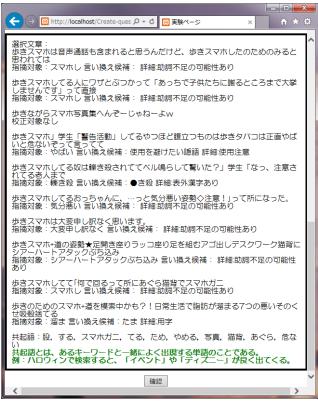


図 6 質問文作成のための参考文章要素と共起語

5.2. 実験結果

アンケート作成体験には、31 人のお客様にご参加いただいた。まずは 5.1 節で述べたように、図 $3\sim$ 図 6 に示した通り動作確認はできた。

そして、システムを評価するために表 4 に示す設問を設けたアンケートを参加者に回答してもらった.

表 4 アンケート作成体験のアンケート内容

	女生 アングードド成体験のアングード的各
Q1	作成された文章や共起語リストは、質問文の 作成に役立ちましたか?
Q2	今は作成された文章は日常の日本語になって いませんが、どのぐらい意味が伝わっていま すか?
Q3	twitter 以外のサービスのデータも利用できる としたら、どのサービスからデータを収集し てみたいと思いますか?
Q4	本システムは、自動で質問文の作成をするということを最終目標としています. この目標以外に、もっとこういう改善をしたらいいのにということを教えてください.
Q5	その他, ご意見, ご感想がありましたらご自由にお書きください. 何もなければ, 不意に思いついたこと, 連想したことを書いてみてください.

※設問中の作成された文章とは、文章要素を指す.

アンケートの実施結果を次に述べる.

図7にアンケートの $Q1\sim Q3$ の回答結果をまとめたグラフをそれぞれ示す. これらの円グラフから次のことが読み取れる.

Q1, Q2 に対しては合計 8 割以上の人が「役に立った」、「少し役に立った」および「伝わっている」、「少し伝わっている」と回答しており、いい評価と悪い評価が同程度かまたは少し悪い評価だという予想よりも良い結果であった.

また、Q3 の結果は、 FaceBook とブログは予想より 少し多く、LINE は予想よりも少し低いという結果となった. これは、LINE はプライベートなコミュニケーションを行うためのツールであり、会話を見られたくないという人が一定数いるからだと考えられる.

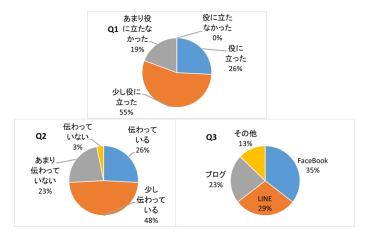


図7 Q1~Q3の回答結果

Q4, Q5 の自由回答欄の意見をまとめると次のようになる.

- 文章の羅列になっており、やるべきことが直感的 にわかりづらい.
- 現段階では文章にするよりも、共起語リストのような単語のほうが良い.
- 情報を分類・カテゴリ分けするほうが良い.
- イラスト(写真)もあるとわかりやすい.
- 例文や、定型文を用意して、簡単にアンケートを 作れる方式も良いのではないか.

6. 考察

実験の結果により、参考文章要素や、共起語リストが予想よりも高評価であったことから、システムは役に立っており、文章要素作成の精度が上がればより良いものになると期待が高まった.

また、実験の参加者からいただいた意見から、初めての人やコンピュータ初心者向けのチュートリアルの提供や、イラスト等を利用した親しみやすくかつ分かりやすい画面デザインへの変更など、システムには改善の余地がある。これらの改善によりさらに一般の人たちの実験への参加が気軽になり、研究を進める助けになることが期待できる。

文章要素作成の精度が低い要因の一つとして,提供する参考文が質問文型ではないことが挙じられる.実験では参加者のアンケート作成の参考として自動的入文章を生成するように実装したが,文型を考慮に入れていなかった.文章を質問文に成型するためには,質問文型の構成を分析し,構成要素や品詞の組み合わせ方などを定義するなど,別の課題も必要となる.本研究で着目したのはアンケートの作成を補助することをあるため,今回は文章を質問文に成型する手法の提案は取り扱わなかった.

文章要素作成の精度や上記に上げた今後の改善点以外にも課題が確認された。例えば、今回利用したTwitter のツイートには実際に人が投稿するツイート以外に、コンピュータが自動でツイートを投稿するようにプログラミングされた、ボットプログラムによる

ツイートや, ブログ, ニュースサイトなどの共有目的 で投稿されたツイートが多く含まれる. 生身の人間の 意見を収集したい本システムでは, ボットプログラム のツイートを除外したいのは当然のことであるが, 共 有ツイートは生身の人間の意見として捉えることが出 来る. しかし、ニュースサイトの共有ツイートなどは 同一時間帯に同一内容のツイートが投稿される場合が 多く, 時間帯によっては文章要素の作成結果がほぼ同 一内容のものが並んでしまうということがありうる. これらの対策として、ボットプログラムのアカウント は、「@~」で示されるアカウント名に「~bot」と付 いていることが多いため,アカウント名に「bot」が含 まれるアカウントからのツイートは除外することで対 策が可能である. 共有ツイートについては, ツイート の一部ないしは全体が同じ内容であるツイートが複数 回出現した場合,2回目以降のツイートは検索結果か ら除外するという方法が考えられる. また, Twitter は 拡散性が高いが故に、偽情報も急速に拡散するなど、 不正確な情報も多いため, どのようなデータが適切で あるかなど,情報の正確性をどのようにして担保する のかについても今後検討すべき課題となる.

上記の他に、共起語の検索範囲について今回はキーワードの前後5単語で十分あろうと判断したが、範囲を広げた場合の変化を調査する必要がある。また、分かち書き生成プログラムでは、今回は3単語ずつの分かち書きにしたが、5単語を一組とするほうが最も当てはまりが良いなど、一組あたりの単語数も検討の余地がある。

7. まとめ

本研究では、意思決定支援システムの一機能で扱うアンケート文が人の手により作られる場合に特定の候補者に有利になるなど内容の客観性が損なわれる可能性がある問題に対する解決方法として、自動的に質問文を生成するために必要な文章要素を生成する方法を提案し、人の手でアンケートを作成することを支援するアンケート作成支援システムを試作した.

実験の結果から、システムの有効性が確認できたが、 直感的なわかりやすさの追求や、作成文章要素の精度 向上であることが今後の課題であることが分かった. この課題の要因として、画面デザインが文字中心であ り、作成された文章も文字の羅列を並べているだけな ど、参考文章としては見にくいものになってしまって いることが挙げられる.他にも、マルコフ連鎖だけで はなく、辞書を利用する等、自動文章作成手法に検討 の余地が多いことも要因の一つである.

参考文献

[1] 藤村春輝,凌暁萍,"意思決定支援機能を備えた Web-DB型電子投票システム",情報科学技術フォ ーラム講演論文集,vol.14, No.4, pp.527-532,愛 媛大学,2015年8月24日.

- [2] 藤村春輝,凌暁萍,"意思決定支援機能を備えた Web-DB 型電子投票システムの評価および考察", 第 78 回全国大会講演論文集, Vol.2016, No.1, pp.1021-1022, 2016 年 3 月 10 日.
- [3] 田名場忍,"研究タイプによる質問紙調査の質問項目作成について",弘前大学教育学部附属教育実践総合センター研究員紀要,No.5,pp.143-148,2007年3月.
- [4] 製品 | UpToDate, http://www.uptodate.com/ja/home/product, 2016.
- [5] scorobo | 製品 | テクノスデータサイエンス・マーケティング株式会社, https://www.tecnos-dsm.co.jp/products/scorobo, 2016
- [6] 佐々木 千晴,藤井 敦,石川 徹也, "意思決定支援のための主観情報マイニング",言語処理学会第12回年次大会発表論文集,pp.77-80,2006.
- [7] 坂本尚子,森康真,北上始,"階層的管理機能を持つアンケート実施支援システムの構築",情報処理学会研究報告コンピュータと教育,vol.75,No.6,pp.41-48,2004年7月3日.
- [8] 佐藤由一, 椋田實, "Web を利用したアンケート の自動生成プログラムの開発", 情報処理学会第 65 回全国大会講演論文集, pp.441-442, 2003 年.
- [9] 國近秀信,花多山知希,平嶋宗,竹内章,"英語 長文読解学習のための質問文自動生成機能の実 現とその評価",電子情報通信学会論文誌, Vol.J83-D-I, No.6, pp.702-709, 2000 年 6 月 25 日.
- [10] 百々健人,清木康,"対話型レコメンド検索を対象とした問い合わせ自動生成システム",交通運輸情報プロジェクトレビュー,No.23,pp.56-60,2014年.
- [11] WEB アンケート作成ツール【クリエイティブサーベイ】 セルフアンケート ASP, https://creativesurvey.com/, 2016.
- [12] 無料セルフアンケート ASP『Questant (クエスタント)』 -MACROMILL-, https://questant.jp/, 2016.
- [13] 今さら聞けないディープラーニングの基本、機械 学習とは何が違うのか 野村総合研究所 古明地 正俊氏が解説 | ビジネス+IT, http://www.sbbit.jp/article/cont1/32033, 2016.
- [14] K Church, W Gale, P Hanks, D Hindel, "Using Statistics in Lexical Analysis." Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon, pp.115-164, 1991.