

音声会話型検索エージェントにおける多様な情報獲得を促すための 観点推薦タイミングに関する研究

飛岡 憲[†] 山本 岳洋[†]

† 兵庫県立大学 大学院情報科学研究科 〒 651-2197 兵庫県神戸市西区学園西町 8-2-1

E-mail: †ad22h042@gsis.u-hyogo.ac.jp, ††t.yamamoto@sis.u-hyogo.ac.jp

あらまし 本研究の目的は、(1) 観点推薦が多様な情報獲得の支援に貢献するのか、(2) 観点推薦のタイミングが、多様な情報獲得をするユーザの検索行動にどのような影響を与えるか、という2つのリサーチクエッションを明らかにすることである。具体的には、3つのトピックについての情報検索タスクを実験参加者に与え、Wizard-Of-Oz法を用いて、それぞれのトピックに対して「観点推薦なしのインターフェース」、「クエリ応答後即時に観点推薦するインターフェース」、「ユーザが悩んだタイミングで観点推薦するインターフェース」の3つのインターフェースで観点推薦を行ったり、観点推薦タイミングを操作したりすることで実験参加者に与える影響を分析した。実験参加者の、あらかじめ用意された観点に占める獲得した観点情報の割合や、自発的なクエリ数、多様な情報を獲得できたかを問うアンケート結果、半構造化インタビューから得られたデータを分析した結果、観点推薦するシステムのほうが、より多様な情報が獲得できることが分かった。また、観点推薦のタイミングに間があることによって、システムから得た回答を整理する時間ができるものの、自発的な情報検索を必ずしも促さず、むしろシステムからの観点推薦を待つような行動変容が起こることが分かった。

キーワード 対話型検索, 情報検索, 多様性

1 はじめに

Amazon Echo や Google Home, Siri をはじめとする音声会話型検索エージェントを日常生活の中で利用する機会が増えてきている。市場も拡大傾向にあり、2024年には8.4億台が市場に出回ると予測されている¹。また、Amazon Echo Showのように画面が存在するデバイスも一部あるが、多くは音声のみのデバイスであり、音声のみでの情報検索がWebサーチの約30%になるといわれている²。現状の利用用途の上位を占めているのは、音楽をかけること、天気を確認すること、アラームやリマインドを設定することである[1]。つまり、単純な質問やクエリに対して正確に情報を提供することが可能な場面に利用されているといえる[9]。このように現状は限定的な利用用途であるものの、市場の拡大や技術革新に伴い、音声によって学習や意思決定などの際に行われる探索的検索にも利用範囲が拡大していく可能性がある。

探索的検索は、多面的で制限のない反復的な情報検索であるという特徴をもつ[6]。そのため、ユーザが必要となる観点の情報を網羅的に検索する必要がある。デスクトップでのウェブ検索では、一つのクエリを入力するだけでも、様々な観点の情報が出力結果として返ってくるため探索的検索が行いやすい。例えば、「ダークチョコレート 健康」とクエリを入力した場合、

SERP上のスニペットに、「血圧低下」、「ダイエット」、「成分」、「摂取量」など様々な観点が表示されることによって、これらの情報が次のクエリの候補となるため、ユーザの探索的検索が進むと考えられる。

さらに、既存手法として、あるトピックの観点を推薦する手法や、網羅的に検索できるようにするためのクエリ推薦手法が存在する[11][10][7]。これらの手法はデスクトップにおける検索を前提にしているが、多様な情報を獲得するために、音声のみでの情報検索においても重要だと考えられる。音声会話型検索において、単純にユーザのクエリに対して回答するだけでは、情報を探索するための手掛かりが少ない。そのため観点の推薦が有効だと考えられるが、多様な情報を獲得するための対話設計の議論が十分になされていないため、観点の推薦が効果的かどうかは明らかになっていない。

また、観点を推薦するタイミングも、多様な情報を獲得するうえで重要な要素であると考えられる。ユーザとシステムの対話において、既存の研究では、コスト、便益、不確実性を考慮したうえで、適切なタイミングで主体性が混在するインタラクションが望ましいとされている[4]。観点推薦における混在型のインタラクションのアプローチとして、ユーザの初回クエリへの回答後、システムが主導権を握り、即座に観点推薦する即時介入型と、基本的には、ユーザが主導権を握り、クエリを思い浮かず困ったタイミングで観点推薦するという人手介入型があると考えられる。しかし、多様な情報を獲得するうえで、どのような混在型のインタラクションが望ましいかは明らかになっていない。

以上から、本研究で取り組む研究課題は以下のとおりである：

1 : <https://www.businesswire.com/news/home/20200427005609/en/Juniper-Research-Number-Voice-Assistant-Devices-Overtake>

2 : <https://www.gartner.com/smarterwithgartner/gartner-predicts-a-virtual-world-of-exponential-change>

RQ1: 音声での情報検索において、観点推薦が多様な情報獲得の支援に貢献するのか。

RQ2: 音声での情報検索において、観点推薦のタイミングが、多様な情報獲得をするユーザの検索行動にどのような影響を与えるか。

これらの研究課題を明らかにするため、Wizard-Of-Oz 法による被験者内実験を行った。実験を通して実験者から得られたデータを分析する。

2 関連研究

2.1 ウェブ検索を支援するインターフェース

Umemoto らはウェブ検索中にクエリ推薦を行う際に、これまでに得た情報量とそのクエリによって得られる情報量を表示するシステムを提案している [11]。提案システムは、ユーザに得られる情報量を示すことで探索行動を促すものである。Saito らは、情報の出展があいまいな表現をハイライト表示して情報探索を促す手法を提案している [10]。提案手法によってウェブブラウジング中に訪れるページ数の増加や、ブラウジングにかかる時間の増加が見られた。Yamamoto らは、認知心理学の知見に基づくクエリ推薦システム「クエリプライミング」を提案している [12]。「根拠」、「比較」といった批判的思考を喚起するキーワードをクエリ補完やクエリ推薦時に提示することで、ユーザに注意深い情報探索を暗黙的に促すものである。Srishti らは、情報の探索とユーザの知覚メモリに断絶があるため認知的負荷がかかっていることに注目し、検索スペース横にワーキングスペースを用意し、その中で文脈に応じた検索キーワードをサジェストするシステム「インターウィーブ」を提案している [7]。能動的な探索を生み出すのみならず、新しく発見した知識を既存の知識と結びつけることができることを明らかにした。

いずれの研究もデスクトップでの検索を前提とした話であるため、音声のみの会話型検索において観点推薦と多様な情報獲得を促すための研究はない。

2.2 明確化質問による情報検索の支援

会話型検索エージェントにおいて、ユーザの情報検索を支援するインタラクションの一つに明確化質問がある。明確化とは、「何かを明確にしたり、理解しやすくするための説明や詳細」³と定義される。会話型検索エージェントは、曖昧さの解消や、潜在的なエラーを防止し、ユーザのクエリや応答を明確にするために明確化質問を行うことが可能である。Aliannejadi らは、会話型検索において、クエリの意図を特定するために明確化質問を使用した [2]。システムが次の質問を選択する際に、元のクエリと以前の質問と回答のやり取りを考慮した明確化質問を用いると、ユーザの検索パフォーマンスの向上につながった。明確化質問は、ユーザが知りたい情報が定まっているときは、効率的に情報獲得ができるため有効であると考えられるが、知り

たいことが明確に定まっていなかったり、様々な情報を網羅的に獲得したいという探索的検索が求められる場合には有効ではないと考えられる。

2.3 会話型検索における混在型のインタラクション

人とシステム間のインタラクションのアプローチは、ユーザ主導かシステム主導かの大きく2種類に分けられる。現在の検索エンジンは、ユーザが常にクエリ発行し、検索エンジンはその結果を回答するというインタラクションであることからユーザ主導のシステムであるといえる。一方、会話型検索システムは、エージェントがユーザのニーズをくみ取って回答する中で、ユーザとエージェントが状況に応じて会話を主導していく混在型のインタラクションが求められるシステムであると定義されている [8]。会話型検索システムの定義に表れているように、混在型のインタラクションが重要な要素であるとされているものの、多様な情報検索を獲得するうえでどのような混在型のインタラクションが望ましいかを示した研究はない。

3 実験デザイン

本章ではまず、本研究における観点と情報の多様性の定義について述べる。そのあと、本研究で検証する観点推薦と観点推薦のタイミングの検証方法について述べる。

3.1 観点と多様な情報の定義

本研究における観点とは、調べたいことをトピックと定義した際、そのトピックを構成する一つ一つの情報のことを示す。例えば、「ダークチョコレートの健康について」をトピックとして、クエリを入力した際に、出力結果として返ってくる関連するキーワード「ダイエット」、「食べるタイミング」、「味」等が観点である。

本研究における多様な情報とは、観点の集合体のことを示す。したがって、観点が多ければ多いほど、多様性が高いといえる。

3.2 観点的推薦とタイミングの検証方法

観点的推薦は、多様な情報獲得の支援に貢献するという仮説のもと、観点推薦するインターフェースと観点推薦しないインターフェースを比較する。ユーザとシステムのクエリの合計数や、あらかじめ実験者の用意した観点到占める実験参加者の獲得した観点的情報の割合を比較することによって、観点推薦が多様な情報獲得の支援に貢献したかを検証する。また、アンケートによって多様な情報を獲得できたと感じたかを聞き取り、インターフェース間で比較を行う。

観点推薦のタイミングについては、常にシステムが主導権を握って観点を即時に推薦するよりも、ユーザがクエリを思いつかず困っているタイミングで観点推薦するほうが、得た情報を整理したり、一方的ではないインタラクティブな対話によって刺激されたりすることによって、新たな観点を思いつくなどユーザの能動的な検索行動を促すという仮説をもっている。そのため、即時に観点推薦するインターフェースと、困っているタイミングで観点推薦するインターフェースを比較し、多様な

3 : <https://dictionary.cambridge.org/us/dictionary/english/clarification>

情報獲得をする検索行動が見られたかを検証する。具体的には、ユーザが発したクエリの数を比較することによって、観点推薦のタイミングによって多様な情報を獲得するための検索行動の違いがみられたかを検証する。さらに、観点の即時推薦は、システムからの一方的なコミュニケーションであるため不自然なインタラクションであるという仮説のもと、ユーザビリティに関するアンケートを実施し、インターフェース間で検証を行う。

4 実験

本章では実施したユーザ実験の詳細を述べる。まず実験計画について述べ、その後具体的な実施手順について説明する。最後に、用いたシステムと実験参加者の募集方法について述べる。

4.1 実験計画

本研究では被験者内実験計画を用いた。このとき、実験参加者が3つの検索タスクを行う順序については、実験条件およびタスクの両者を考慮した Graeco-Latin 方格 [5] に従い実験参加者を割り当て、実験条件およびタスクに関する順序効果を最小限に抑えた。

4.2 トピックとシナリオの作成

本研究では、実世界における様々な情報検索の場面やトピックを想定し、3つのトピックと、それぞれに対して最終的に調べたことをまとめてもらうためのシナリオを作成した。

以下がトピックとシナリオである：

トピック 1: ダークチョコレートの健康への効果について

シナリオ: ダークチョコレートの健康効果について、テレビで特集が組まれた結果 SNS でバズっています。興味をもったあなたは、最新のアレクサを使って調べることにしました。調べた後は、最近健康食にはまりだした友人にダークチョコレートの健康への効果を伝えようと思っています。調べたことをまとめてください。

トピック 2: 風力発電について

シナリオ: 学校の授業で SDGs のことについて学んだあなた。持続可能な開発の中で自然エネルギーに関して、調べてくるよう先生から宿題が課されました。自然エネルギーに関して調べていく中で風力発電に興味を持ちました。風力発電に関して調べて、調べたことを書いてください。

トピック 3: スティーブンスピルバーグ監督について

シナリオ: 最近映画にはまり、毎週のように金曜ロードショーで映画を見ている。さらにNetflixも契約し、毎日のように映画をみている。良い映画がないかなと調べていると、スティーブンスピルバーグ監督の映画が評価がよいことがわかりました。気になったあなたは、スティーブンスピルバーグ監督について調べようと思いました。調べたことを検索終了後まとめてください。

上記の3つのトピックそれぞれに関して、事前にプレ実験を行い、クエリとしてよく出現した観点を整理し、あらかじめクエリに対する回答を作成した。

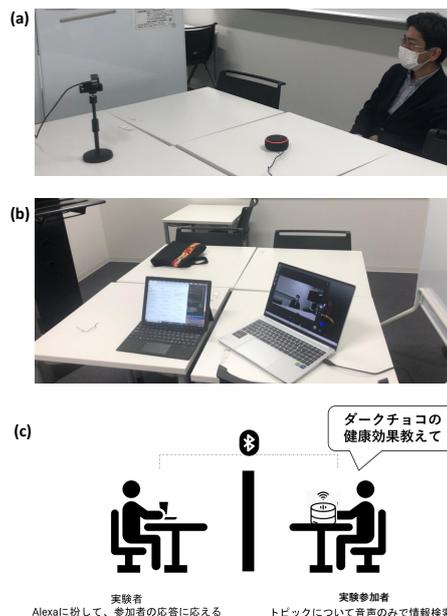


図 1 実験環境の設営

4.3 実験環境

兵庫県立大学情報科学研究棟のセミナー室で実験環境を設営した。実験に用いるスマートスピーカーはテーブルの中央に設置し、実験参加者と実験者がスマートスピーカーに命令するためのパソコンは、実験参加者と実験参加者をホワイトボードで仕切ることによって、疑似的に別室を作った (図 1 (a))。なお、実験者はビデオカメラを通じて実験参加者をモニタリングできるようにした (図 1 (b))。

図 1 (c) で示したように、Wizard-Of-Oz 法を用いて、実験参加者のクエリに応じて、実験者はスマートスピーカーに命令をする。

4.4 実験手順

まず、実験参加者に対して検索タスクの概要、収集するデータ、収集したデータの研究目的への利用などについて説明した。これらについて同意した実験参加者のみが実際の実験タスクへ進み、実験を行った。

つぎに、実際の実験タスクに入る前に、これから行う検索タスクへのイメージを深め、使用する検索システムに慣れるために、訓練タスクを行ってもらった。訓練タスクでは、「兵庫県立大学について」をトピックに設定した。

訓練タスクでは、クエリに対する回答について、さらに深める質問ができることや、別の観点で検索できること、検索結果が聞き取れなかった際に聞き直すことができること、観点の推薦があった際に必ずしも推薦に従わなくてもよい旨を伝えた。また、実際の実験タスクに入る前に、最大検索時間は 10 分とすること、情報が十分に集まったと感じたらいつでも検索を終了してもよいこと、実験中パソコンやスマートフォンでの検索をしてはいけないこと、タスク終了後は再度検索できないこと

を指示した。

訓練タスク終了後は、実験参加者に3つの検索タスクを実施してもらった。各検索タスクは以下の手順で実施した。

はじめに、4.2節で示したトピックとシナリオを実験参加者に提示した。トピックとシナリオについて理解してもらった後、1つ目の検索タスクを実行してもらった。実験参加者が1つ目の検索タスクを終了した後、システムに関するアンケートに回答してもらった。2つ目、3つ目の検索タスクも同様の手順を踏んで実施した。最後に、インタビューを実施し、プロフィールに関する質問に回答してもらった。

4.5 タスク毎のアンケート

Google Form を用いてアンケートを作成した。アンケートは、検索で調べた内容を自由記述で回答する項目と、System Usability Scales [3] と観点の推薦が多様な情報検索に貢献できたのかを問う質問で構成した。

アンケート項目は以下のとおりである：

- Q1 検索エージェントで調べたことをまとめてください。
- Q2 この検索エージェントを頻繁に使うと思う。
- Q3 この検索エージェントが不必要に複雑に感じた。
- Q4 この検索エージェントが使うことが簡単だと思った。
- Q5 この検索エージェントを使うために、技術に精通した人のサポートが必要だ。
- Q6 この検索エージェントは様々な機能が統合されていることがわかった。
- Q7 この検索エージェントは応答に一貫性がないと思った。
- Q8 たいていの人はこの検索エージェントの使い方をすぐに学ぶことができると想像できる。
- Q9 この検索エージェントの使い勝手が悪いと感じた。
- Q10 この検索エージェントを安心して使えると思った。
- Q11 この検索エージェントを使い始める前に、多くのことを学ぶ必要があると感じた。
- Q12 この検索エージェントとの対話は自然であったと感じた。
- Q13 様々な観点から情報を収集できたと感じた。
- Q14 この検索エージェントは様々な情報を獲得するうえで役に立ったと感じた。
- Q15 様々な情報を獲得するうえでシステムの観点のおすすめは有用だと感じた。
- Q16 意見をまとめるうえで必要な情報が十分に集められたと感じた。

Q1を除く15項目について、5段階のリッカート尺度(1:まったくあてはまらない, 2:どちらかというにあてはまらない, 3:どちらともいえない, 4:どちらかというにあてはまる, 5:とてもよくあてはまる)を設定して回答してもらった。

4.6 インタビュー内容

すべての検索タスク終了後に、観点推薦や観点推薦のタイミングに関してアンケートの補完的に半構造化インタビューを行った。インタビュー項目は以下のとおりである：

- Q1 観点推薦は有用だと感じたか。



図2 実験参加者が情報検索するために用いたスマートスピーカー

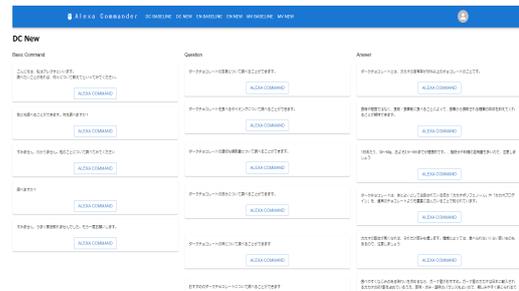


図3 実験で用いたシステムのスクリーンショット。画面上のボタンを押すことによって、あらかじめ用意した回答を Alexa が発話する。

- Q2 観点推薦のタイミングについてどう感じたか。
- Q3 音声で情報検索をしてみてどう感じたか。

上記に加え、検索中のユーザの行動で気になった検索行動に関して適宜質問した。

4.7 音声対話システム

実験では、実験参加者が情報検索に使うデバイスとして、純粋に音声での検索をするために、画面付きデバイスを避け、図2で示した Echo Dot (エコードット) 第3世代 - スマートスピーカー with Alexa を選択した。Alexa を選定した理由は、カスタマイズするためのドキュメントが充実していることと、自然な日本語を読み上げることができるためである。

次のとおり3つのインターフェースを準備した。

観点推薦なし ユーザのクエリへの回答後に、観点推薦しないインターフェース。

観点推薦あり(即時介入) ユーザのクエリへの回答後に、即座に観点推薦するインターフェース。

観点推薦あり(人手介入) 提案手法。ユーザがクエリに悩んだタイミングで、観点推薦するインターフェース。

図3に実験で用いた実験者側の音声システムのユーザインターフェースを示す。本システムは、実験者が実験参加者のクエリの結果を返したり、観点推薦したりする際に用いる。Amazon Alexa にコマンドラインを用いて文章を読み込ませ、発話できるように amazon-remote-control⁴ を利用した。また、実験参加者のクエリに、ボタン一つで即座に反応できるように、フロントエンドは React.js, バックエンドは Express のフレームワークを用いた Node.js, データベースは MongoDB を用いて

4: <https://github.com/thorsten-gehrig/alexa-remote-control>

表 1 観点推薦なし、観点推薦あり（即時介入）、観点推薦あり（人手介入）の情報獲得割合の平均。表中の括弧は、標準偏差を表す。太字上の数字は有意差 ($p < 0.05$) が確認されたインターフェースを示す（1：観点推薦なし、2：観点推薦あり（即時介入）、3：観点推薦あり（人手介入））。

	観点推薦なし	観点推薦あり（即時介入）	観点推薦あり（人手介入）
情報獲得割合	0.35(0.18)	0.82 (0.18) ¹	0.52(0.15)

ウェブアプリケーションを構築した。画面左には、検索開始の合図や検索を促すコマンドを表示している（例：「私は Alexa といひます。調べたいことがあれば、何々について教えてといひてみてください。」、「ほかにも調べることができます。何を調べますか？」）。画面中央には、プレ実験をした際に出現した観点をもとにした観点推薦をするコマンドを表示している（例：「ダークチョコレートの定義について調べることができます。」）。画面右には、クエリに対する回答を表示している（例：「ダークチョコレートとは、カカオの含有率が 55%以上のチョコレートのことです。」）。なお、回答できないクエリに対しては、インターフェース間で統一して「すみません、わかりません。ほかのことについて調べてみてください。」というコマンドを用意した。

4.8 実験参加者

兵庫県立大学社会情報学部の Microsoft Teams のチャンネルで募集をし、2022 年 11 月 21 日から 2022 年 12 月 26 日の間、計 9 名（うち、男性 7 名、女性は 2 名）が実験に参加した。実験の開始から終了までに最大 90 分の時間をとった。実験を最後まで完了した 9 名の実験参加者には 1,455 円を支払った。どの実験参加者も問題なく実験をやり終えたため、すべての実験参加者の検索タスクに関するデータを得て分析に用いた。

5 分 析

本章では、1 章で述べた 2 つの研究課題それぞれについて分析した結果を報告する。

5.1 RQ1: 観点推薦が多様な情報獲得の支援に貢献するのか

5.1.1 分析方法と分析対象とした指標

観点推薦が多様な情報獲得の支援に有効であれば、観点推薦あり（即時介入）と観点推薦あり（人手介入）が観点推薦なしよりも情報獲得割合が高くなることが期待される。そこで、3 つのインターフェース間で事前に準備した情報をどれだけ獲得できたかについて、情報獲得割合の平均値で比較した。

本実験で用いたシステムの限界として、回答を準備していないクエリに関しては、クエリを返答することができない。したがって、ユーザのクエリに全て答えられた場合の情報獲得数の差を見るために、合計クエリ投入数の平均値についても比較した。情報獲得割合と同様に、観点推薦が多様な情報検索に有効であれば、観点推薦あり（即時介入）と観点推薦あり（人手介入）が観点推薦なしよりも合計クエリ数が増えることが期待される。

また、観点推薦による多様な情報獲得への貢献度を問うアンケート結果の Q13～Q16 を用いて、ユーザの体感として観点推薦が有効であったかを分析する。もし有効であれば、これまでと同様に、観点推薦あり（即時介入）と観点推薦あり（人手介入）が観点推薦なしよりも各項目の平均値が上回ることが期待される。

以上の動機から、次の指標をインターフェースごとに測定し、分析に用いた。

情報獲得割合 実験者がトピック毎に用意した観点のうち、実験参加者が獲得できた観点情報の割合。

合計クエリ投入数 実験参加者がシステムからの観点推薦に頼らずにクエリ投入した数とシステムの観点推薦によるクエリ投入数の合計。

なお、トピックごとに事前に用意した情報の数が異なるため、獲得割合の指標を用いた。

5.1.2 結 果

まず、3 つのインターフェース間の情報獲得割合の平均値について、Tukey HSD 検定を用いてインターフェース間の情報獲得割合の平均について比較を行った。その結果、観点推薦なしと観点推薦あり（即時介入） ($p = 0.0$) について有意差が認められた。表 1 にこれらの指標をインターフェースごとにまとめたものを示す。観点推薦あり（即時介入）が多様な情報を貢献する上で有効であったことが分かった。

次に、合計クエリ投入数についても同様に、Tukey HSD 検定によって、3 つのインターフェース間の合計クエリ投入数の平均値に有意な差がみられるかを検定した。その結果、インターフェース間の情報獲得割合には有意差が認められなかった。表 3 にこれらの指標をインターフェースごとにまとめたものを示す。システムがユーザの全てのクエリに回答できていた場合、獲得できる情報量の差はなかったといえる。

最後に、アンケート結果それぞれについても同様に、Tukey HSD 検定によって、3 つのインターフェース間の合計クエリ投入数の平均値に有意な差がみられるかを検定した。表 2 にアンケート結果をインターフェースごとにまとめたものを示す。

Q13 では、情報検索をするうえで、様々な情報を収集できたと感じたかについて調査を行った。Tukey HSD 検定を用いてインターフェース間の平均について比較を行った結果、観点推薦なしと観点推薦あり（即時介入） ($p = 0.0$)、観点推薦なしと観点推薦あり（人手介入） ($p = 0.0$) について有意差が認められた。観点推薦によって、ユーザは様々な情報を獲得できたと感じていることが分かった。

Q14 では、様々な情報を獲得するうえでシステムが有用であったかについて調査を行った。Tukey HSD 検定を用いてインターフェース間の平均について比較を行った結果、観点推薦なしと観点推薦あり（即時介入） ($p = 0.0$) について有意差が認められた。情報獲得割合の結果と同様の結果であることから、観点推薦によって得られる情報が多いほうが有用であることが分かった。

Q15 では、情報検索をするうえで、観点推薦の有用性について調査を行った。Tukey HSD 検定を用いてインターフェース

表 2 観点推薦なし、観点推薦あり（即時介入）、観点推薦あり（人手介入）の多様な情報獲得への貢献度を問うアンケート結果の平均。表中の括弧は、標準偏差を表す。太字上の数字は有意差 ($p < 0.05$) が確認されたインターフェースを示す（1：観点推薦なし，2：観点推薦あり（即時介入），3：観点推薦あり（人手介入））。

質問	観点推薦なし	観点推薦あり（即時介入）	観点推薦あり（人手介入）
Q12 この検索エージェントとの対話は自然であったと感じた。	3.22(1.31)	3.67(1.25)	3.56(0.68)
Q13 様々な観点から情報を収集できたと感じた。	1.63(0.70)	4.63 (0.48) ¹	4.13 (0.71) ¹
Q14 この検索エージェントは様々な情報を獲得するうえで役に立ったと感じた。	3.11(1.20)	4.78 (0.42) ¹	4.11(1.05)
Q15 様々な情報を獲得するうえでシステムの観点のおすすめは有用だと感じた。	3.22(1.23)	4.56 (0.50) ¹	4.56 (1.41) ¹
Q16 意見をまとめるうえで必要な情報が十分に集められたと感じた。	2.89(1.20)	4.44 (0.96) ¹	3.89(2.01)

表 3 観点推薦なし、観点推薦あり（即時介入）、観点推薦あり（人手介入）の平均クエリ投入数。表中の括弧は、標準偏差を表す。太字上の数字は有意差 ($p < 0.05$) が確認されたインターフェースを示す（1：観点推薦なし，2：観点推薦あり（即時介入），3：観点推薦あり（人手介入））。

	観点推薦なし	観点推薦あり（即時介入）	観点推薦あり（人手介入）
自発クエリ投入数	11.56 (3.83) ²	3.78(1.03)	8.00 (3.33) ²
推薦クエリ投入数	0.00(0.00)	10.33 (2.05) ^{1,3}	4.44 (1.95) ¹
合計クエリ投入数	11.56(3.83)	14.11(2.02)	12.44(1.71)

間の平均について比較を行った結果、観点推薦なしと観点推薦あり（即時介入）($p = 0.0$)、観点推薦なしと観点推薦あり（人手介入）($p = 0.0$) について有意差が認められた。観点推薦なしよりも、観点推薦あり（即時介入）と観点推薦あり（人手介入）の平均が有意に高かったことから、観点推薦が有用であることが分かった。

Q16 では、情報検索をするうえで、様々な情報を収集できたと感じたかについて調査を行った。Tukey HSD 検定を用いてインターフェース間の平均について比較を行った結果、観点推薦なしと観点推薦あり（即時介入）($p = 0.015$) について有意差が認められた。観点推薦が意見形成するうえでも有用であることが分かった。

5.1.3 考察

本節の分析で得られた知見をまとめると以下ようになる。システムがユーザからのクエリに対して回答できていた場合、情報獲得量は変わらない可能性がある。しかし、表 1 より、観点推薦あり（即時介入）において、あらかじめ用意していた情報のおよそ 8 割の情報を獲得できていたことが分かる。この結果から、システムからの推薦クエリを断ることができたにも関わらず、そうせずに観点推薦を受け入れていることが読み取れるので、ユーザはシステムからの観点推薦を有効だと判断した可能性がある。

Q13 と Q15 のアンケート結果からも、様々な情報を獲得するうえで、会話型の情報検索において観点推薦が有用であったことが分かった。

タスク終了後の観点推薦の有用性に関するインタビューにおいて、「自分が思いつかない視点からの提案だったので、役に立った。」、「自分の思い浮かばない観点をおすすめしてもらって調べやすい。」といった声があった。ユーザが何を調べたらよいか分からない場合に観点推薦は有用であることが分かる。さらに、「観点推薦によって、ほかにも調べようと思った。」、「賞

について調べたときに、その評価はどうなんだろうって幅が広がった。」といった声もあった。具体的な検索行動として、システムがスピルバーグ監督の有名な映画を回答した際に、例外なくユーザは「ジュラシックパークってどんな映画？」とクエリを発行していた。また、ダークチョコレートのトピックにおいても、システムがダイエット効果があることを回答した際に、「体重 BMI がどれくらい減るが教えて。」というようなクエリを発行しているケースも見られた。このことから、観点推薦から得られた情報を深堀りしていくような検索行動が誘発されるが分かった。

以上の考察から、会話型検索エージェントによる観点推薦は多様な情報獲得の支援に貢献すると考えられる。

5.2 RQ2: 観点推薦のタイミングが、多様な情報獲得をするユーザの検索行動にどのような影響を与えるか

5.2.1 分析方法と分析対象とした指標

ユーザがクエリに悩んでいるタイミングでのシステムからの観点推薦は、ユーザにシステムからの回答を整理させたり、インタラクティブな会話によって刺激を与えたりする可能性があることから、多様な情報を獲得するために、ユーザの自発的なクエリ発行を誘発するという仮説をもっている。すなわち、観点推薦あり（人手介入）が観点推薦あり（即時介入）よりも、ユーザがシステムの観点推薦に頼らずに、自発的に投入するクエリ数が多くなることが期待される。さらに、タイミングによって自発的なクエリが発行されているならば、観点推薦あり（人手介入）は、観点推薦なしよりも自発的に投入するクエリ数が多くなることを期待できる。そこで、3つのインターフェース間で自発クエリ投入数の平均値を比較した。

また、観点推薦あり（即時介入）のほうが、観点推薦あり（人手介入）よりも、誘導的であり、ユーザとシステム双方のインタラクティブな情報検索ではないため、違和感を感じるものが期待される。そこで、Q12 でシステムとの対話に違和感がないかをシステム間で比較した。また、ユーザビリティの点においても、観点推薦あり（人手介入）が最も高くなると期待されるので、Q2～Q11 で SUS を測定することにした。

以上の動機から、次の指標をインターフェースごとに測定し、分析に用いた。

自発クエリ投入数 実験参加者がシステムからの観点推薦に頼らずにクエリ投入した数。

なお、言い換え表現や同じ質問は重複して計測しない、「もう

一度教えて」をカウントしない、システムから推薦されたものに対して「調べて」という対応は計測しない、というルールのもと自発クエリ投入数を計測した。

5.2.2 結 果

まずは、自発クエリ投入数について、Tukey HSD 検定を用いてインターフェース間で比較を行った。その結果、観点推薦なしと観点推薦あり（即時介入）（ $p = 0.0$ ）、観点推薦あり（即時介入）と観点推薦あり（人手介入）（ $p = 0.025$ ）について有意差が認められた。表3にこれらの指標をインターフェースごとにまとめたものを示す。即座に観点推薦あり（即時介入）よりも観点推薦あり（人手介入）のほうが自発クエリ投入数が多いことが分かった。しかし、観点推薦なしと観点推薦あり（人手介入）には有意差がみられないことから、観点推薦あり（人手介入）によって自発クエリ投入数が増加したとはいえない。

次に、アンケート結果 Q12 についても同様に、Tukey HSD 検定によって、3つのインターフェース間の平均値に有意な差がみられるかを検定した。

Q12 では、情報検索をするうえで、システムとの対話に違和感がないかについて調査を行った。Tukey HSD 検定の結果、インターフェース間に有意差が認められなかった。会話型の情報検索において観点推薦のタイミングが異なっているにもかかわらず、インターフェース間で対話の自然さに違いがないことが分かった。

最後に、SUS を用いて、インターフェース間のユーザビリティに関して比較を行った。肯定的な質問である Q2, Q4, Q6, Q8, Q10 の各スコアから 1 を引き、否定的な質問である Q3, Q5, Q7, Q9, Q11 の各スコアを 5 から引く。これらの総計に対して、2.5 を乗じたものが SUS のスコアである。その結果、観点推薦なしは 64.4 点、観点推薦あり（即時介入）は 72.6 点、観点推薦あり（人手介入）は 70.6 点となった。一般的に 80% 以上が非常に優れたユーザビリティを提供しているといえる。観点推薦あり（即時介入）が最もユーザビリティが高いが、どのインターフェースも改善の余地がある。

5.2.3 考 察

本節の分析で得られた知見をまとめると以下ようになる。自発的なクエリ投入数において、ユーザがクエリを悩んでいるタイミングで観点を推薦する観点推薦あり（人手介入）は、即時に観点推薦する観点推薦あり（即時介入）よりも多いが、観点推薦なしと比較すると有意差がないことから、ユーザの自発的検索行動を促せるわけではなかった。

タスク終了後に観点推薦のタイミングについてインタビューしたが、そもそもタイミングに気づいていない実験参加者が 2 人いた。気づいている実験参加者からは、「あまり聞きたいことが出てこなかったの、もうちょっとはやく提案してほしい。」、「普通の人との会話だったら、ポンポンしてくれるが、間があると不安に感じた。」というような観点を即座に推薦してくれるほうがよいという声があった。

一方で、「タイミングであれば、頭で理解できる時間があるので待ってもらったほうがよい。」、「遅いほうがよかった。情報を整理している間にくるのでうっとうしい。」、「口をはさみにくい。」というように、情報を整理する時間がなかったり、聞きた

いことがあっても聞きづらかったりするの、間があるほうがよいといった声もあった。

観点推薦あり（人手介入）について、実験者からみるとクエリに悩んでいると思われた時間に、何を考えていたかを聞くと、「待っているときは、次の提案を予測しているみたいな感じ。どういことを提案してくれるのか？を考えていた。」、「観点推薦を待っていた。」、「ダークチョコの時は返事を待っていた。」といった声があり、クエリに悩んでいるのではなく、システムからの観点推薦を待つ時間となっており、逆に自発的クエリ投入を妨げてしまう場合があることが分かった。

以上の考察から、即時でない観点推薦は、ユーザに情報を整理する時間を与えられるものの、自発的な情報検索を促すわけではなく、逆に阻害してしまう場合があることが分かった。

6 議 論

本章では、前章の分析で得られた知見を基に、会話型エージェントを用いた情報検索において、ユーザの多様な情報獲得を促すための方法について議論する。本研究の大きな知見は、会話型検索エージェントを用いた情報検索において、観点推薦が多様な情報の獲得に貢献するという点と、また、観点推薦をする場合、ユーザにクエリを考えさせる時間を与えることが、自発的なクエリ投入を促すのではなく、逆にそれを妨げる場合があることである。

今回の実験では、システムから一つずつ観点を推薦した。前章で述べたように、ユーザは推薦された観点を深掘する検索行動をとる場合がある。一度に提示する観点の情報量を増やすことによって、ユーザの視点を広げて、興味を喚起し、深掘できる情報を拡大できる可能性がある。結果として、より様々な情報を獲得するために、自発的な検索行動をとるのではないかと考えている。

ただし、単純に観点を推薦をすればよいわけではない。「観点の中でも、話題から外れているものについて調べたが、調べなくてもよかった。」、「味とか健康への効果だったので途中でいらなと思った。」という声があったように、文脈に合わなかったり、聞きたいと思ったことでない観点を推薦したりすると、システムへの信頼が低下しかねないため、推薦する観点の内容も重要である。

また、今回の実験では、回答できないクエリに対して、インターフェース間で統一して「すみません、わかりません。ほかのことについて調べてみてください。」と返答した。しかし、自発的なクエリ投入を妨げた要因の一つに、すべてのクエリに回答できなかったことの影響も少なからずあったと考えられる。「どのような情報をもっているかを知りたかったので、システムからの提案を待っていた。」という声があったように、クエリを投げても答えられないのであれば、システムが分かることを把握して、それについて調べたほうが確実に情報を獲得できるので、自発的なクエリ投入を避けるという行動を選択した可能性がある。

すべてのクエリに対して適切な回答を用意することは、情報

を漏れなく収集する必要があり困難である。また、仮に収集できたとしても、適切な回答をシステムや Wizard-Of-Oz 法でも困難である。しかし、一つの解決方法として、2022 年 11 月に OpenAI がリリースした ChatGPT を利用する方法がある。ChatGPT は、クエリに対して、これまでの文脈を保ちつつ、正確な回答を提示してくれる。ユーザが、システムがクエリに適切に返してくれるということを理解すると、システムからの提案を待つということはなくなる可能性がある。ChatGPT は現状、クエリに対して回答するのみであるが、観点推薦を加えることによって、正確に回答し、かつ、よりユーザとシステムのインタラクティブな会話による情報検索を実現できる可能性がある。

7 ま と め

本稿では、会話型検索エージェントにおける多様な情報を獲得するための観点推薦システムの提案と、観点推薦のタイミングによる情報検索行動への影響に関して、リサーチクエションを明らかにするための実験を行い分析した。実験の結果、会話型エージェントにおける情報検索において、観点をシステムから推薦することは多様な情報獲得に貢献することが分かった。また、観点推薦のタイミングによって、自発的な情報検索の行動を促すことはできず、場合によっては、システムからの観点推薦を待つ行動を誘発することから、自発的な情報検索行動を妨げてしまうことが分かった。今後は、ユーザからのクエリに全て回答できるシステムを用いて、一度に提示する観点情報量を増やすことによる情報検索行動への影響について分析したい。

謝辞 本研究は JSPS 科学研究費助成事業 JP21H03904, JP21H03775, JP22H03905, による助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] Adobe. State of voice. *Adobe Digital Insights 2019*, page 7, 2019.
- [2] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W Bruce Croft. Asking clarifying questions in open-domain information-seeking conversations. In *Proceedings of the 42nd international acm sigir conference on research and development in information retrieval*, pages 475–484, 2019.
- [3] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [4] Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 159–166, 1999.
- [5] Diane Kelly et al. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, 3(1–2):1–224, 2009.
- [6] Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.
- [7] Srishti Palani, Yingyi Zhou, Sheldon Zhu, and Steven P Dow. Interweave: Presenting search suggestions in context scaffolds information search and synthesis. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–16, 2022.
- [8] Filip Radlinski and Nick Craswell. A theoretical framework for conversational search. In *Proceedings of the 2017 con-*

ference on conference human information interaction and retrieval, pages 117–126, 2017.

- [9] Gary Ren, Xiaochuan Ni, Manish Malik, and Qifa Ke. Conversational query understanding using sequence to sequence modeling. In *Proceedings of the 2018 World Wide Web Conference*, pages 1715–1724, 2018.
- [10] Fumiaki Saito, Yoshiyuki Shoji, and Yusuke Yamamoto. Highlighting weasel sentences for promoting critical information seeking on the web. In *International Conference on Web Information Systems Engineering*, pages 424–440. Springer, 2020.
- [11] Kazutoshi Umemoto, Takehiro Yamamoto, and Katsumi Tanaka. Scentbar: A query suggestion interface visualizing the amount of missed relevant information for intrinsically diverse search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 405–414, 2016.
- [12] Yusuke Yamamoto and Takehiro Yamamoto. Query priming for promoting critical thinking in web search. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 12–21, 2018.