

超高齢社会における高齢者のための情報想起支援システム

石垣 達也[†] You Jingyi[‡] 瀧本 洋喜[‡] 奥村 学[†]

[†] 東京工業大学科学技術創成研究院 [‡] 東京工業大学大学院

{ishigaki, takimoto, youjy}@lr.pi.titech.ac.jp oku@pi.titech.ac.jp

1 はじめに

“物忘れ (Memory loss)” [2] は円滑な会話の妨げとなる。例えば、図1に示すように、過去に訪れた場所の具体的な名称が思い出せず、会話が中断するといった状況は、とくに高齢者による会話中に多くみられる。この問題は超高齢社会に向かう諸国において深刻である。本研究では、このような状況のユーザを支援する情報想起支援システム [1] を構築し性能を評価する。

構築システムは3段階の処理によりユーザの情報想起を支援する。システムは1) 会話をモニタリングし記憶を想起する必要のある状況を生み出す発話 (以後、情報想起発話と呼ぶ。) を同定し、2) 知識ベースから情報想起支援に必要な情報を検索した上で、3) ボットの発話としてユーザに提示する。ここで、知識ベースには例えば、過去の新聞記事データやユーザから提供されたライフログ、ソーシャルネットワークサービスへの投稿等を想定する。本研究では特に新聞記事データを用いる。

本稿で述べる情報想起支援システムは情報想起発話同定器および記事検索器の2つの部品からなる。前者は会話をモニタリングし、発話が情報想起発話であるか否かの判定を二値分類問題として解く。具体的には規則および機械学習に基づくベースライン手法を用いて実現する。記事検索器は情報想起発話をクエリとし、情報想起に役立つ記事を新聞記事データベースからキーワード一致により検索する。

本研究の貢献は以下である。1) 情報想起発話の特徴分析、同定器の構築および評価のための公開可能なデータセットを作成した。分析より、情報想起発話の特徴は過去の経験や属性について尋ねる質問文を含む発話もしくは、本人以外の現在の属性について尋ねる質問文を含む発話であることが分かった。2) データセットを用いた定量評価実験より、情報想起発話同定器は再現率の観点からは質問文を情報想起発話と同定する単純なベースラインでも0.96の高い再現率を示すが、精度の観点からは性能向上の余地があることが分かった。3) 記事検索器を評価するための被験者実験から、キーワード一致によるベースライン手法では正しく検索できない発話が存在し、発話をクエリとした検索タスクの難しさが明らかになった。



図1: “物忘れ” とシステムによる情報想起支援の例。

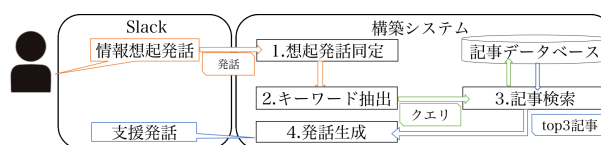


図2: 情報想起支援システムの構成。

2 システム概要

提案システムの構成を図2に示す。ユーザはテキストベースのチャットシステムであるSlack上で会話すると想定する。提案システムはボットとしてSlackに接続され、ユーザ同士の会話をモニタリングし、必要なタイミングでボットとして会話に参加する。システムはまず、想起発話同定器を用いてユーザの発話から情報想起が必要な状況であるか判定する。情報想起が必要であると同定されれば情報想起発話からキーワード抽出を行う。キーワードは記事検索器に渡され、新聞記事データベースから関連する記事群を取り出す。具体的には検索エンジンの出力する重要度スコアが高い上位3記事までを検索結果として返す。これらの記事から発話生成器がボットの発話を生成し支援対象ユーザに提示する。以下に本システムの主要な構成部品である情報想起発話同定器および記事検索器について順に説明する。

2.1 情報想起発話同定器

情報想起発話を同定する課題はこれまでに組み立てられておらず、どのような手がかりを用いてこのような発話を同定可能であるか必ずしも明らかではない。そこで、本研究では、まず情報想起発話を含むデータセッ

	質問文あり	質問文なし
全体	73.6% (81/110)	26.3% (29/110)
- 情報想起が必要	- 74.1% (60/81)	- 6.9% (2/29)
- 情報想起が不要	- 25.9% (21/81)	- 93.1% (27/29)

表 1: インタビュワーによる発話から無作為に抽出した 110 発話が情報想起発話であるか、質問であるか否か分類した結果。

ト作成し、情報想起発話の特徴を分析した。また、このデータセットを用い情報想起発話同定器の定量的な評価も行った。データセット作成、分析、同定器の定量評価について順に説明する。

データセット作成: 情報想起発話を収集するために、30 歳代から 70 歳代の 13 名に作業を依頼した。作業者には 2 人ペアを組み、Slack 上で会話するよう指示した。ペアのうち片方は主に質問をする役割とし、もう片方は主にそれに答える役割とした。質問をする役割の作業者には、相手の記憶を想起する必要がある発話をするよう指示した。質問をする役割の作業者の発話はすべて質問である必要はなく、自然で多様な情報想起発話を収集できるよう配慮した。なお、質問をする役割の作業者の年齢が低くなるようペアを作成し、高齢者が情報想起を必要とする場面を再現するよう努めた。結果、情報想起発話の特徴を分析するために 99 対話からなるデータセットを得た。99 対話には質問をする役割の作業者の発話が 1,081 が含まれ、これらの発話に対し作業者自身が情報想起発話であるか否かを表す二値ラベルを付与した。結果、1,081 発話のうち 549 発話が情報想起発話であるとラベル付けされた。

情報想起発話の分析: 次に、情報想起発話の特徴を明らかにするため 1,081 発話のうち 110 発話を無作為に抽出し分析した。結果を表 1 に示す。また、分析対象とした発話の例を表 3 に示す。110 発話のうち 81 発話は質問文を含み、29 発話は質問文を含まない。質問文を含む 81 発話のうち 60 発話は情報想起発話であるのに対し、質問文を含まない 29 発話のうち情報想起発話は 2 発話のみであり、質問文を含むか否かは情報想起発話を同定するための手がかりであることが分かる。質問文を含み情報想起発話である事例を表 3 の 1) から 4) に例示する。

質問文を含む 81 発話をさらに、時制および質問対象となるエンティティに着目し分類した結果を表 2 に示す。過去の経験などについて尋ねる質問を含む発話の場合、質問対象が作業者本人に関する質問 (41/81) であっても、本人以外に関する質問 (14/81) であっても、情報想起が必要である。このような発話の例を表 3 の 1) および 3) に例示する。一方、現在の属性を尋ねる質問の場合、本人に関する質問を含む発話であれば情報想起発話でない事例が多い (19/21)。本人以外に

	過去	現在
本人に関する質問	41/81	21/81
- 情報想起が必要	- 100.0%(41)	- 9.5%(2)
- 情報想起が不要	- 0.0%(0)	- 90.5%(19)
本人以外に関する質問	14/81	5/81
- 情報想起が必要	- 100.0%(14)	- 60.0%(3)
- 情報想起が不要	- 0.0%(0)	- 40.0%(2)

表 2: 質問を含むインタビュワーによる 81 発話を本人に関する質問か他人に関する質問か、過去のことを尋ねる質問か、現在のことを尋ねる質問かという観点から分類した結果。

に関する質問であれば、現在の属性を尋ねる質問であっても情報想起発話である割合が高い (3/5)。これらの分析より、過去の経験や属性について尋ねる質問もしくは、本人以外の現在の属性について尋ねる質問を含む発話は情報想起発話であることが多いと考えられる。**情報想起発話の自動同定手法:** 作成したデータセットを用いて、3 つのベースライン手法を定量的に比較する。POSITIVE はすべての発話を情報想起発話とするベースラインである。RULE は規則により質問文を同定し、質問文を含む発話を情報想起発話と判定する。規則には文末表現が“か。”や“?”である事例を情報想起発話とした。SVM は Support Vector Machine (SVM) を用いて情報想起発話を同定する。学習データには前節で述べた作業者による発話とアノテーションを用いる。素性ベクトルは n-gram、単語分散表現および質問文の有無を用いて作成した。n-gram 素性は Mecab により形態素に分割した発話を二値素性に変換した。単語分散表現は発話中の形態素に対応する Glove による事前学習済み埋め込み表現ベクトルを全形態素で平均した。質問文か否かを表現する素性は RULE により正例と判定されたか否かを二値で表現した。これらの手法により、発話が情報想起発話であり、ユーザの情報想起が必要な状況を生み出すと判定されれば、以下の記事検索器による情報検索が行われる。

2.2 記事検索器

情報想起発話が同定され、記憶想起が必要な状況であると判定されると記事検索器に情報想起発話が渡される。記事検索器は過去の新聞記事から情報想起に有益な記事を検索する。検索された記事はボットの発話として情報想起が必要なユーザにのみ提示される。記事検索器はキーワード抽出、キーワードによる記事検索、発話生成の 3 つの処理を順に行う。キーワード抽出は、情報想起発話から名詞、動詞、形容詞および副詞を抽出する。単語分割および品詞タグ付けには

質問 ⊃ 情報想起発話	1) 若いころは主にどんな映画を見ていましたか？ジャンルとか。 2) ご両親の趣味はなんですか？ 3) お父さんは大学教授だったのですね。お父さんは何を教えていらしたのですか？ 4) 好きな俳優さんの誕生日はいつでしたっけ？
質問 ⊃ 情報想起発話でない	5) 普段から料理はされますか？
質問でない ⊃ 情報想起発話	6) 子どもたちも夏休みの実家での滞在を楽しめたでしょうね。
質問でない ⊃ 情報想起発話でない	7) わたしもよく料理するんですよ。

表 3: 質問する役割の作業者の発話例。上は質問文を含む発話、下は含まない発話。

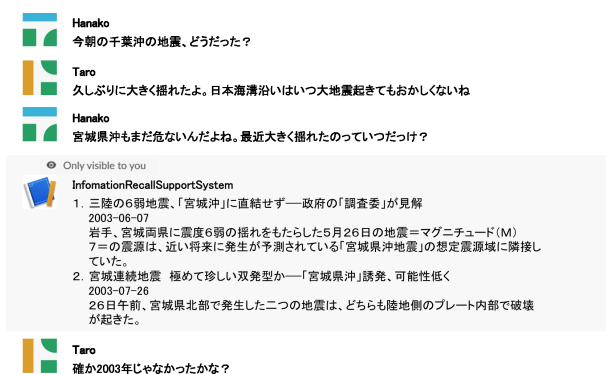


図 3: 情報想起支援システムの動作例。情報想起を喚起する発話を同定し、関連する新聞記事を一覧する。

Janome¹を用いる。抽出されたキーワードは検索エンジン Whoosh²に渡され、1991 年から 2005 年までに発行された毎日新聞記事データから関連記事を OR 検索する。最終的に重要度スコアの上位 3 記事を検索結果とし発話生成する。発話は記事タイトル、記事の先頭 1 文および発行年月日を図 3 のように列挙する。

3 評価

3.1 情報想起発話同定器

表 4 に情報想起発話同定器の各比較手法の精度，再現率および F 値を示す。SVM のハイパーパラメータ C は 5 分割交差検定により開発セットの F 値が最大になるよう調整した。すべての比較手法が再現率の観点からは高いスコアを得た。POSITIVE は低い精度 (0.53) を得た。SVM や RULE では改善 (それぞれ 0.68 と 0.67) し、これらは同様の F 値 (0.79) を得た。一方、SVM が精度ではわずかに良い性能を示し、再現率では RULE が 3 ポイントほど高い。これらの 2 手法は精度が重視される状況，再現率が重視される状況によって使い分けることが望ましい。

¹<https://github.com/mocobeta/janome>
²<https://bitbucket.org/mchaput/whoosh>

	精度	再現率	F 値
POSITIVE	0.53	1.00	0.68
SVM	0.68	0.93	0.79
RULE	0.67	0.96	0.79

表 4: 情報想起発話同定器の定量評価。

3.2 記事検索器

検索器の性能評価では、4 人の被験者に情報想起発話を投げかけ、ボットによる発話により正しく記憶が想起されたか判定させた。具体的には情報想起発話とそれに対するボットの発話を 1 ターンと考え、ボットの発話が情報想起に役立ったか判定した。ボットが発話しない場合は、情報想起発話のみを 1 ターンとした。再現率を重視し、ボットの発話タイミングは RULE 手法により決定した。合計 78 ターンの対話を行い、うち 30 ターンにおいて情報想起を必要であった。残りの 48 ターンは挨拶や相槌などが 26 ターン、情報想起発話であるが被験者がもともと正しい返答に必要な知識を持っておりボットの助けが必要なかった 22 ターンである。表 5 に情報想起が必要であった 30 ターンに対し、被験者がボットの発話が役立ったと判定したターン数を示す。結果、30 ターン中 12 ターンが想起に役立ったと判定された。一方、18 ターンは役立たなかったと判定された。このことから、情報想起を支援する発話生成のための情報検索は難しいタスクであることがわかる。被験者が役に立ったと判定した 12 ターンのうち、2 ターンはボットが間違っ情報を提示しているにも関わらず被験者はその情報が正しいものだと思い、役立ったと判定している。情報想起の役に立たなかった理由は、情報検索に失敗し関連しない情報を提示した事例が 15 ターンと多数を占める。これら 15 ターンについて正しい検索結果が得られない理由としては、1) 対話の文脈が必要、2) 対話での表層表現と新聞での表現が異なる、などが挙げられる。前者については、例えば“この年は大きな地震が他にもありましたよね？”という情報想起発話は“この年”が指し示す年を文脈から理解する必要があり、正

想起に役立った	12 ターン
- ボットが正しい情報を提示	10 ターン
- ボットが誤った情報を提示した	2 ターン
想起に役立たなかった	18 ターン
- ボットが関連する情報を提示しなかった	15 ターン
- ボットが発話しなかった	3 ターン

表 5: 情報想起が必要な状況で情報想起支援システムが役に立っていると被験者が判定した数と理由の内訳。

しく検索されない。後者については“2004 年は小泉首相が北朝鮮に行ったようです。誰と会ったんですか？”という情報想起発話に対し、“会った”という口語表現は新聞記事では“会談した”という表現で出現する。よって、単純なキーワード一致での検索が難しい。また、“新潟中越地震”などのように自然災害や事件の発生日よりもあとに名称が付与されることがあり、このような事例はキーワード一致による検索を難しくしている一因である。

3.3 システム全体の定性評価

最後に、前述した被験者にシステムの使用感に関する感想を自由に述べてもらい、定性的な評価を行った。ボットが正しく情報提示した場合は役立ったとの意見を得た。ボットの発話が情報想起には役に立たない場合であっても、会話中の話題に関連する補足的な情報が提示される、会話を中断させずに話題を広げることが出来たりと、スムーズな会話を継続させるために役立ったという意見を得た。情報想起システムが提示する情報により、情報想起以外の観点からも会話をより円滑にする可能性が示唆されることは、興味深い。

4 関連研究

情報想起発話同定は発話分類課題の一つである。発話分類課題としては対話行為推定 [6]、発話意図推定 [5] などが代表的な課題として挙げられる。本研究はこれらとは異なり、情報想起発話であるか否かを分類する課題を新たに提案する。多くの情報想起発話は質問を含むことを考えると、質問を対象とした分類問題とも関連する。Tamura ら [7] は複数文から構成される質問からもっとも重要な 1 文を同定する二値分類問題に取り組み、質問応答システムの性能が向上することを報告している。Ishigaki [4] らは複数文から構成される質問を 1 文に要約する課題を分類問題として扱った。これらは質問文を分類する問題として類似するが、情報想起を喚起するか否かを分類する本研究での設定とは着眼点異なる。

情報想起を支援するための新聞記事検索は情報探索課題の一つである。従来の研究の多くで、入力として与えられるクエリは単語集合である。一方、本研究では発話がクエリとなりキーワード抽出する点異なる。対話の文脈を考慮したり口語から文語への変換など従来手法では対応できないキーワード抽出手法や探索手法そのものの発展が必要となる。

情報想起発話の特徴分析に用いたデータセットは対話コーパスの一つと捉えられる。対話コーパスには非タスク指向型、タスク指向型の分類が存在するが我々のデータセットは前者である。また、モダリティの観点では我々のデータセットはテキストを用いる対話に着目しており、音声対話とは異なる。参加人数の観点からは人間同士の 1 対 1 対話を想定しており、マルチパーティ対話 [8] や人間と機械による対話コーパス [3] とは異なる。既存コーパスにおいて、対話者の年齢を考慮するコーパスは我々の知る限り存在しない。

5 今後の展望

本研究では、会話中の情報想起発話を同定し、関連する新聞記事を提示することで情報想起支援を行うシステムを構築し評価した。情報想起発話の特徴分析、同定器の自動評価のためのデータセットを新たに提案した。情報想起発話の同定では再現率の評価において概ね良い性能を得ているが、精度に向上の余地がある。時制や対話文脈を考慮した素性設計などの方向性が考えられる。記事検索器の構築は従来の検索課題と比較し文脈や口語への対応など追加的に考慮すべき点があることを述べた。文脈を考慮する検索システムや表層ではなく意味ベクトルを用いた検索システムへの拡張などが今後の方針として考えられる。また、構築したシステムは過去の新聞記事を検索対象としているが、ライフログやソーシャルメディアへの投稿などより個人に紐付いたデータへも検索対象を広げたい。

参考文献

- [1] Hsin-Hsi Chen and Manabu Okumura. Information recall support for elderly people in hyper aged societies. In *Proceedings of the The Web Conference 2018*, pp. 431–432, 2018.
- [2] Fergus IM Craik, Mark Byrd, and James M Swanson. Patterns of memory loss in three elderly samples. *Psychology and aging*, Vol. 2, No. 1, p. 79, 1987.
- [3] Ryuichiro Higashinaka, Kotaro Funakoshi, Michimasa Inaba, Yuiko Tsunomori, Tetsuro Takahashi, and Nobuhiro Kaji. Overview of dialogue breakdown detection challenge 3. Vol. 6, 2017.
- [4] Tatsuya Ishigaki, Hiroya Takamura, and Manabu Okumura. Summarizing lengthy questions. In *Proceedings of IJCNLP2017*, Vol. 1, pp. 792–800, 2017.
- [5] Tsuneko Kato, Atsushi Nagai, Naoki Noda, Ryosuke Sumitomo, Jianming Wu, and Seiichi Yamamoto. Utterance intent classification of a spoken dialogue system with efficiently untied recursive autoencoders. In *Proceedings of the 18th Annual SIGDial Meeting on Discourse and Dialogue (SIGDIAL2017)*, pp. 60–64, August 2017.
- [6] Vipul Raheja and Joel Tetreault. Dialogue Act Classification with Context-Aware Self-Attention. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL2019)*, pp. 3727–3733, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [7] Akihiro Tamura, Hiroya Takamura, and Manabu Okumura. Classification of multiple-sentence questions. In *Proceedings of IJCNLP2005*, pp. 426–437, 2005.
- [8] David C Uthus and David W Aha. The ubuntu chat corpus for multiparticipant chat analysis. In *Proceedings of 2013 AAAI Spring Symposium Series*, 2013.