



HOKKAIDO UNIVERSITY

Title	対話システムにおける対話履歴要約の有効性について
Author(s)	福田, 彩子; 荒木, 健治; ジェプカ, ラファウ
Description	ここに掲載した著作物の利用に関する注意 本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。
Citation	情報処理学会研究報告. 情報学基礎研究会報告 (FI) , 2010(11), 1-6
Issue Date	2010-01-21
Doc URL	https://hdl.handle.net/2115/63780
Type	journal article
File Information	IPSJ-FI10097011.pdf



対話システムにおける 対話履歴要約の有効性について

福田彩子[†] 荒木健治[†] ジェプカラファウ[†]

近年、非タスク指向型の対話システムは様々な場面で利用されている。しかし、それらのシステムはユーザーに飽きられやすいという問題がある。よりシステムを実用的なものとするためには、話し手としての応答だけでなく、聞き手としての応答を適切に行うこともまた重要となってくると考えられる。そこで我々は、聞き手としてより良い応答を返すシステムを実現するために、過去の対話内容を要約して応答に用いる手法を提案する。本稿では、さらに本手法の有効性を確認するために行った実験について述べる。

Effectiveness of Non-task-oriented Dialogue System Using Abstract of Chat Logs

AYAKO FUKUDA[†] KENJI ARAKI[†]
RAFAL RZEPKA[†]

Recently, non-task-oriented dialogue systems are used in various fields. However, almost all users lose interest in communicating with these systems if they use it every day. In order to realize a practical system, a suitable response not only as a speaker is important, but also as a listener is important. It is an excellent system that the system can memorize the previous mentioned matter, and talk with it. In order to realize such kind of system, our proposed method uses an abstract of chat logs for the generation of responses. In this paper, we did questionnaires to confirm effectiveness of our system, and asked participants to evaluate the responses that the system generated.

1. はじめに

近年、子供の情操教育や、老人への癒しを目的として、また人々を楽しませるために、非タスク指向型対話システムが利用されている[1][2]。これらの対話システムは日常的に繰り返しユーザーと対話する事が前提となるが、実情としては、システムの目新しさがなくなると飽きられてしまう。これはシステムが話し手としてユーザーの興味を引くような情報を提供し続けるだけでは、新たな情報を提供出来なくなった時点で限界が見えてしまうためである。そこでこのようなシステムにおいては、時に聞き手として振舞うことでユーザーの発話や話題提供を促し、システムによる情報提供以外の方向からも会話を促進させることが、長期的利用を目指す上で有効な手段の一つとなると考えられる。

本稿では、聞き手としての応答において対話履歴要約の利用が有効であると仮定し、ユーザーが過去に話した対話内容と関連のある発言を再び行った際に、対話履歴からユーザー発話の要約を用いて応答するシステムを提案する。チャットログの要約においては、松尾ら[3]の電子掲示板の要約における継承、新情報、影響の3つの指標を用いたモデルをチャットに適用し、ユーザ発話に関連するチャットログの部分抽出や応答文生成に利用する。

以下に、第2章でアンケートによる仮定の検証について、第3章で対話履歴の要約に用いる範囲抽出について、第4章で要約手法について、第5章で評価実験について、第6章でまとめについて述べる。

2. 予備実験

本研究の重要性を確認するために、アンケートを行った。被験者に話題提供者と聞き手による短い対話文と、話題提供者が後日同じ話題にふれた際の聞き手の応答文を読ませ、その評価結果について考察する。

2.1 アンケートの概要

アンケートに用いる対話文は2文用意した。うち1つは実際にチャットで話された対話内容から抜粋したもの、1つは第一著者が創作したものである。聞き手の応答文は、第一著者が独自に考案した7種類の応答タイプに対応したものを、各対話文につき1文ずつ用意した。応答タイプのA, B, C, Dは第一著者が主観的に実際の会話においてよく現れる応答と判断したもの、タイプE, Fは従来のチャットボットによく見られる応答、タイプGが本手法の仮定により有効であると考えられる応答である。よって、この7種で起きうる全ての応答を網羅しているわけではない。アンケートに用いた対話文のうち第一著者が創作したものを図1に、聞き手の7種の応答タイプと

その例を表 1 に示す。

■会話例 2

ある日の CさんとDさんの会話

C：最近、私の家の近所に***っていうケーキ屋が出来たんだよ！

D：また！？ Cさんの近所はケーキ屋が一杯ですね。

C：でも、いくらお店が多くてもお気に入りがないのが悩みだよ。だから、今度のお店には期待してるんだ。

D：あれ、ということは、まだ行ってないんですか？

C：うん。お土産で貰ったマカロンを食べただけ。すごく美味しかった！だから、ケーキはまだ食べてないんだよね。

D：それじゃ、そのお店のケーキの味までは分かんないですよね。

C：でしょ？今度行くから、オススメの店かどうか報告するね。

後日 2人が話していて、その話題の事をCさんが思い出します

C：あ、***なんだよね。

図 1 対話の例文

Figure 1 Short Conversation Examples.

表 1 7種の応答と例文

Table 1 Seven Types of Response.

Type A	相槌 e.g. はい。
Type B	覚えていない反応 e.g. あれ？...何でしたっけ？
Type C	覚えている反応 e.g. あっ、この前の？
Type D	結果について促す応答 e.g. どうでしたか？
Type E	全く脈絡のない応答 e.g. 明日台風が来るらしいですよ、怖いなあ。
Type F	関連のある話題だが脈絡のない応答 e.g. ケーキといえば、今度の母の日どうしようかな？
Type G	過去の発言から代表的な部分を要約した応答 e.g. マカロンが美味しいっていう、近所に来たお店ですね。

評価は全 7 段階で行い、応答の印象が非常に悪い場合を 1、非常に良い場合を 7 とした。また、被験者には 7 種の応答について評価を行う前に、自分が聞き手だった場合の標準的な応答文と、自分が話題提供者だった場合の理想的な応答文を書くことの

依頼を行った。被験者は 10 代～50 代までの計 29 名で、平均年齢は 23.5 歳、うち男性が 9 名、女性が 20 名である。理系学生 15 名、文系学生 4 名、理系会社員 1 名、文系会社員 4 名、その他 5 名で構成される。

2.2 結果と考察

評価結果を表 2 に示す。アンケートの評価結果より、過去の対話内容に基いた応答のタイプ G が最も評価が高く、次いで結果を促す応答のタイプ D、覚えていることを示す応答のタイプ C の印象が良いことが明らかとなった。

表 2 7 段階評価による応答の種類別平均点

Table 2 Result average of evaluates on a scale of 7.

	Type A	Type B	Type C	Type D	Type E	Type F	Type G
対話の例文 1	3.76	2.13	5.02	5.39	1.77	2.92	5.77
対話の例文 2	3.71	2.27	4.92	5.40	1.85	2.95	5.83
平均評価点	3.73	2.20	4.97	5.39	1.81	2.93	5.80

被験者の回答により得られた応答の例文は 96 文であった。うち標準的な応答文は 58 文、理想的な応答文が 38 文である。これらの回答文を第一著者の主観に基づき 7 種の応答に分類した。回答文と応答タイプの分類の例を表 3 に、回答文に現れる応答タイプ 7 種の出現率を表 4 に示す。ただしタイプ B については、相手の情報の受容を表す“ああ”や“あー”などのフィラーのみの場合もタイプ B であるとみなす。またタイプ G については要約に限定せず、回答文中に 1 語でも対話文と関連する語が存在すれば、タイプ G とみなすとした。なお、得られた 96 文はテンプレートに編集し、システムが生成した応答文の前後に付加する。

表 3 回答文の分類の例

Table 3 Examples of Classified Seven Types of Response.

被験者による回答文	タイプの分類結果
ケーキ食べたの？美味しかった？	D, G
Cさんのご近所のケーキ屋さんですね！	G
ああ、あれね。行ってみた？	C, D, G

表 4 回答文中の応答 7 種の出現率

Table 3 Incidence Rates of Seven Type of Responses.

	Type A	Type B	Type C	Type D	Type E	Type F	Type G
標準的な応答文	0.21	0	0.38	0.81	0	0	0.67
理想的な応答文	0.13	0	0.21	0.82	0	0	0.87

分類の結果、複数のタイプを複合した回答文が多くみられ、その平均使用タイプ数は 2.05 個であった。最も多かった組み合わせは、タイプ D、G で、全体における出現率は 0.66 であった。次いでタイプ C、D が多く、出現率は 0.22 であった。3 種類の組み合わせではタイプ C、D、G が最も多く、出現率は 0.12 であった。単独での出現率については、タイプ D、G が並んで高く、覚えていない応答のタイプ B や、脈絡のない応答であるタイプ E とタイプ F は全く現れなかった。また標準的な応答文と理想的な応答文を比較すると、タイプ G の応答の出現率は上昇し、逆にタイプ A、C は出現率の減少がみられた。よってタイプ D、G の組み合わせを用いた応答をした場合に良い印象が得られると考えられる。以上のことから、仮定どおり過去の対話内容の要約利用が、応答において有効であることの確認を行った。

3. 話題区間の抽出

対話履歴の要約をするには、まず過去の対話履歴中から現在の話題と関連がある区間を自動抽出しなくてはならない。そこで、分析のためにチャットログを収集し、人手により話題転換点の抽出と分類を行った。この結果を正解データとして、システムが話題区間の抽出に用いるパラメータを定める。

3.1 チャットログの収集

分析に用いるための対話履歴は、被験者 2 名による 1 対 1 の対話をチャットにて行い、1 回 30 分の対話を数日おきに 5 回にわたって収録する。これを 1 組分行い、合計 1,025 発話、単語数にして 13,881 語の対話履歴を得た。

3.2 人手による話題転換点の分析

話題転換点の抽出と分類は、藤本ら[4]の手法に基づき、対話をした被験者本人が主観により行った。ここではきわめて関連性の強い話題への転換は考慮しないものとし、元の話題と中程度の関連性を持つ話題へ転換する発話を TypeA、関連性のない話題へ転換する発話を TypeB と分類する。2 つの対話履歴につきそれぞれ 2 人、計 4 人の被験者が話題転換点を抽出、分類し、対話相手の話題転換点を正解データとして、どの程度相手の話題転換点と一致するかを分析した。全員の結果を平均したものを表 4 に示す。

表 5 人手による話題転換点の分析結果

Table 5 Manually Classified Topic Changes

	Type A	Type B
話題区間の平均発話数	11.37	21.32
話題区間の最小発話数	3.00	4.00
相手の話題転換点との一致率	0.13	0.66
話題転換点のずれの平均発話数	1.86	0.76

表 4 より、対話相手との話題転換点の一致率は、関連性のない話題への話題転換点であるタイプ B でも 0.66 と高い値ではなかった。また中程度の関連性を持つ話題への話題転換点であるタイプ A に至っては、相手との一致率が 0.13 と低く、個人による判断の揺れが大きいことが判明した。しかし、最も近い相手の話題転換点までの発話数の平均が 1.86 と小さい値であったため、話題転換点が完全に一致しない場合でも話題区間の位置の判断は対話相手と概ね一致していると考えられる。そこで、1.86 を切り上げた 2 発話を許容値として、前後 2 発話までに相手の話題転換点が存在すれば正解とした場合、タイプ A の一致率は 0.29、タイプ B の一致率は 0.88 まで上昇した。

3.3 話題区間の抽出アルゴリズム

現在の話題と関連のある区間を抽出するために、ユーザー発話から得られたキーワードを用いることとした。対話履歴からキーワードが話題となっている話題区間を抽出する。まず対話履歴から最もそのキーワードが話題として取り上げられている発話を特定し、次にその発話と関連のある発話の範囲を前後から抽出する。

そのために、ある発話中出现する語が、前後の発話において話題となっている割合を主題度と定義し、対話履歴の各発話中の単語について分析する。対話履歴の形態素解析には、MeCab[6]を用いた。主題度の算出には、松尾ら[3]の単語の継承度と影響度という指標をチャットに適用したものを用いる。ここでいう継承度とは、ある発話中出现する語が、直前の発話中にどの程度出現しているかを表す指標である。また、直前の発話中出现しない語を新情報と呼ぶ。影響度とは、新情報である語が直後の発話中にどの程度出現しているかを表す指標である。主題度は仮に継承度または影響度の内、取り得るどちらかとし、これにより現時点では実質的に継承度と影響度を区別しないこととする。主題度は以下の式(1)と式(2)により算出する。なお式(2)の、前後の情報より $tf \cdot idf$ 値を算出する手法は、松尾ら[3]の論文から引用したものである。

$$Topicalization(\omega, u(i)) = tfidf(\omega, S(\omega, u(i))) \quad (1)$$

$$S(\omega, u(i)) = \begin{cases} P(u(i)) & \text{if } \omega \in P(u(i)) \\ F(u(i)) & \text{otherwise} \end{cases} \quad (2)$$

$u(i)$ とは対話履歴中の i 番目の発話であり、 $P(u(i))$ は発話 $u(i)$ の直前の発話、 $F(u(i))$ は発話 $u(i)$ の直後の発話である。従って語 ω の主題度は、直前又は直後の発話中の $tf \cdot idf$ 値によって表される。ただし、直前又は直後の発話とは、発話 $u(i)$ より数えて 279 語目までの前後の発話を指す。これはチャットにおけるユーザー発話中の語数に個人差が出ることを考慮したもので、収集したチャットログの 1 発話中に含まれる平均語数である 13.00 語に、タイプ B の話題区間の平均発話数である 21.31 発話をかけたものである。キーワードの主題度が最も高い発話を対話履歴から検出し、その発話を基

点として周辺から関連性のある話題区間の抽出を行う。

まず、基点となる発話 $u(i)$ 中における各単語の重要度を計算する。語 ω の発話 $u(i)$ 中における重要度は、発話 $u(i)$ が含む語の主題度を足し合わせた総スコアにおける、語 ω の主題度が占める割合で表す。重要度は式(3)により算出する。

$$Weight(\omega, u(i)) = \frac{Topicalization(\omega, u(i))}{\sum_{\omega' \in \mathcal{P}(u)} Topicalization(\omega', u(i))} \quad (3)$$

次に、前後のある発話の話題が、基点の発話 $u(i)$ の話題と関連している度合いを話題関連度と定義し、周辺の各発話について調査する。重要度を重み係数とした話題関連度の式(4)に示す。

$$Relatedness(u(j)) = \sum_{\omega \in \mathcal{P}(u)} e(j, Begin(\omega), End(\omega)) \times Weight(\omega, u(i)) \quad (4)$$

$Begin(\omega)$ とは、語 ω が対話履歴中で話題となっているいくつかの区間のうち、発話 $u(i)$ を含む話題区間の開始位置であり、 $End(\omega)$ はその終了位置である。具体的に語 ω の話題区間とは、語 ω が主題度を持って現れる発話が連続している区間のことで、語 ω が主題度を持つ発話と発話の間が 21 発話以上離れた場合は、別の話題区間であるとして区切る。なお 21 発話とは、分析により得られたタイプ B の話題区間の平均発話数に基いたものである。よって、発話 $u(j)$ の発話 $u(i)$ に対する話題関連度は、発話 $u(i)$ に含まれる主題度を持つ各語について、語の話題区間が発話 $u(j)$ を含んでいた場合の重みを足し合わせたものであり、0 から 1 の間で表される。

3.4 名詞による発話検出の拡張

主題度を用いた話題区間の検出方法では、1 度しか出現しない語の話題区間を抽出することはできない。しかし、収録した対話履歴には、1 度しか出現しない語でも直後の話題に大きく影響する例がみられた。実例としては、あるアーティストの名前を含む発話を受けて、直後にそのアーティストの作品が次々と話題にあがったが、そのアーティストの名前自体は、最初の 1 回しか出現しなかった場合が挙げられる。

そこで、キーワードが名詞の場合に限り、対話履歴中で主題度を持たなくとも過去に話題にあがった重要な語であるとみなし、キーワードが最も多く出現する回の最初の出現位置を基点とすることとした。また、これにより基点となる発話が主題度を持つ単語を含まないことがあるため、その場合は直後の発話を順次調査して行き、最初に主題度を持つ単語を含んだ発話を用いて話題区間を抽出する。ただし、抽出された話題区間が基点の発話を含まない場合は、関連のない話題区間を抽出したとみなし、抽出結果を返さないこととする。

3.5 話題区間の自動抽出結果と人手による正解データとの照合

システムの話題区間の抽出にはキーワードを必要とするため、対話履歴の収録の回毎に、主要な話題となった語をいくつか挙げるよう、被験者に依頼した。それにより得られたいくつかの語のうち、被験者両名が共通して挙げた語をキーワードとして用いた。それぞれ各回に付き 2 語、計 10 語のキーワードをシステムに入力し、システムが抽出した話題区間を人手による正解データと照合する。これを 2 組、被験者 4 名分を行い、その結果を平均する。なお、人手による話題転換点の分析結果から、正解の条件については以下のように定めた。抽出した話題区間が最低 3 発話以上の長さで、タイプ B の話題転換点をまたがず、かつ話題区間の開始位置と終了位置の両方が人手による話題転換点と一致した場合を正解、それ以外を不正解とした。ただし、一致には 2 発話までのずれを許容する。これはタイプ B の話題転換点の一致率が高く、タイプ A の一致率が低いことに基く。さらに正解率を高めるために、システムには 2 つのパラメータを設定する。話題関連度の切捨て値による再現度、精度、F 値の平均値の推移を図 2 に、主題度の切捨て値毎の話題関連度の切捨てによる F 値の推移を図 3 に示す。

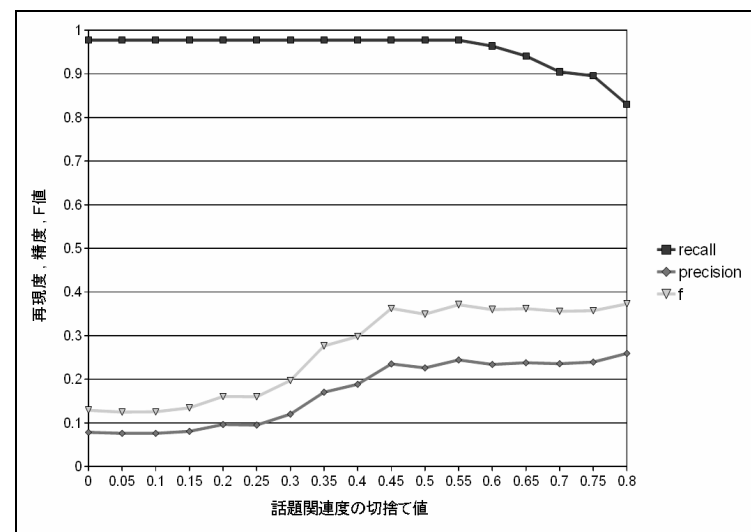


図 2 再現度、精度、F 値の推移

Figure 2 Transition of Recall, Precision and f.

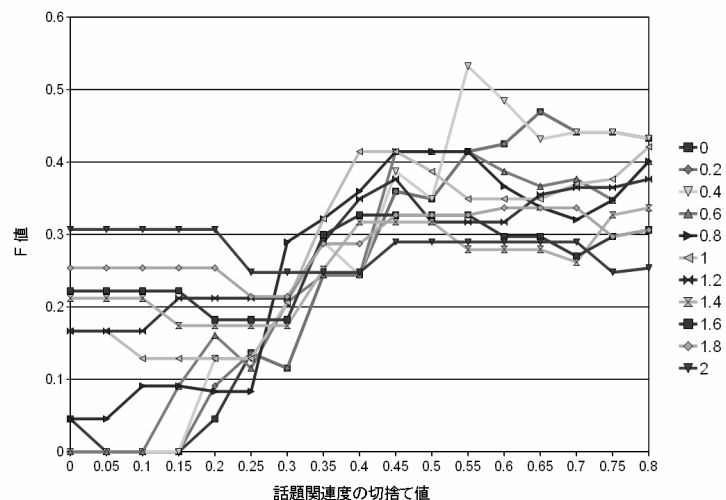


図 3 主題度別の F 値の推移

Figure 3 Transition of f with Topicalization.

パラメータの 1 つは話題区間を話題関連度によって切捨てるもので、値に満たない区間は抽出する話題区間に含まないものとする。これは広範囲にわたって出現する一般的な語や、別の話題区間との繋がりが強い語によるノイズを除去するためである。

もう 1 つは重み係数に利用する主題度を持つ語を切捨てるもので、値に満たない語は重み係数の算出に用いないものとする。これは前者同様に一般的な語を除去するためである。図 2 より、話題関連度の切捨て値の引き上げに伴って F 値が上昇することから、話題関連度の切捨て値が有効であると考えられる。また図 3 より主題度の切捨て値の引き上げは、話題関連度の切捨て値に抛らず F 値を一定に均す効果がみられるが、全体的に F 値の上昇がみられる 0.35 から 0.7 までの話題関連度の切捨て値の区間も下げてしまうことが明らかとなった。ただし、各主要度の切捨て値について F 値を平均したところ、切捨て値の引き上げにより、ごくゆるやかな F 値の上昇がみられた。したがって図 3 より、最も F 値の高い結果が得られた話題関連度の切捨て値 0.55 と、主題度の切捨て値 0.4 を、印象評価実験の際のパラメータとして用いることとした。

4. 話題区間中の発話文の要約

発話文の要約には、まず複数の語からなる短いフレーズを抜き出し、次にそれらを繋げていくこととする。フレーズのスコアは式(5)と式(6)により算出する。

$$Phase\ Score(p) = \prod_{\omega \in \mathcal{P}(p)} Topicalization(\omega, u(i)) \times Relatedness(u(j)) \times A(u) \quad (5)$$

$$A(u) = \frac{\sum_{\omega \in \mathcal{P}(u)} 1}{\sum_{\omega \in \mathcal{P}(u)} idf(\omega)} \quad (6)$$

フレーズ p のスコアは、主題度、発話 $u(i)$ に対する話題関連度、話題区間中の idf 値の平均値の逆数をかけるという計算を、フレーズ中に含まれる全ての語について行ったものである。話題区間中の idf 値の逆数をかけることにより、平均的な語よりも重要な語についてはより大きく、重要でない語についてはより小さな値となる。このフレーズのスコアが 1 を越えたものについて、2 階マルコフ辞書を用いて可能ならば各フレーズを繋げる。次に、以下の式(7)に基づき最もスコアの高い語を順次繋げて文を生成する。なお、 $j(p)$ の初期値は基点となる発話 $u(i)$ の位置を示す i とする。フレーズが含む語数が 13.00 語を超えるか、次に繋げるべきフレーズのスコアが 1 よりも小さくなった場合に、フレーズの接続を停止する。収集した対話履歴をもとに生成された文の例を図 4 に示す。

$$Sentence\ Score(u(i)) = \sum_{p \in S(u(i))} \left(\frac{Current(p(i)) - Last(p(i))}{End(u(i)) - Begin(u(i))} \times Phase\ Score(p) \right) \quad (7)$$

飲んでも意外と体が冷えるので温かいコーヒー飲んでるんです
雪降ったねえ積もるかな
株主優待券の使い方がウマイなーとあれって半額な

図 4 生成文の例

Figure 3 Examples of Sentence.

5. 印象評価実験

本手法に基づき開発した実験システムは、ユーザーが過去に話した対話内容と関連のある発言を再び行った際に、対話履歴からユーザー発話の要約を用いて応答するため、ユーザー発話に対話履歴と関連のある語が含まれる際にしか応答を行うことができない。そこで若原ら[6]の非タスク志向型対話システムをベースシステムとして用いることとした。若原らのシステムをベースとして用いた理由については、若原らが非同期による発話タイミングについて研究しているため、システム自体はチャットボツ

トにおいて基礎的な仕組みであり、ベースシステムが持つ特殊な機能を評価実験において考慮する必要がないためである。また、若原らのシステムには従来型の1対1で交互に発話するモードがあるため、そのモードをベースシステムとして利用する。印象評価実験は、ベースシステムのみシステムAと、ベースシステムに本手法に基づき開発した実験システムを搭載したシステムBとの比較によって行う。被験者は10代～20代までの計4名で、うち男性が2名、女性が2名である。理系学生3名、文系会社員1名で構成される。被験者のうちチャットを日常的に使う人が2名、過去に数回使用したことがある人が2名である。またチャットボットについては過去に数回使用したことがある人が1名、知っているが使用したことがない人が1名、知らない人が2名であった。システムAの評価結果を表6に、システムBの評価結果を表7に、システムA、Bのどちらが優れているかを直接的に比較評価した結果を表8に示す。

表6 システムAの評価結果

Table 6 Evaluation Result of SystemA

質問の項目	被験者1	被験者2	被験者3	被験者4	平均
賢さ	2	1	1	2	1.50
面白さ	4	2	6	4	4.00
人間らしさ	2	1	5	3	2.75
被験者の話を聞いているか	2	2	1	2	1.75

表7 システムBの評価結果

Table 7 Evaluation Result of SystemB

質問の項目	被験者1	被験者2	被験者3	被験者4	平均
賢さ	2	2	2	3	2.25
面白さ	3	2	5	3	3.25
人間らしさ	3	3	3	4	3.25
被験者の話を聞いているか	3	3	5	5	4.00

表8 システムAとシステムBを比較した評価結果

Table 8 Evaluation Result Between SystemA and SystemB

質問の項目	被験者1	被験者2	被験者3	被験者4	平均
賢さ	5	5	6	5	5.25
面白さ	4	5	1	3	3.25
人間らしさ	4	5	2	4	3.75
被験者の話を聞いているか	5	6	4	5	5.00
今後どちらと話したいか	5	6	5	6	5.50

対話実験は、時間において3回ずつ、1回8分の対話を、両システムに対して行う。被験者には最後に両システムとの対話履歴を読ませ、印象評価を行った。対話した順番による評価の揺れを防ぐため、被験者によりシステムAとシステムBの順番を入れ替えた。評価は7段階で行い、最もネガティブな評価を1、最もポジティブな評価を7とした。システムAとBの直接的な比較については、システムAが非常に優れている場合を1、システムBが非常に優れている場合を7とし、中立の場合を4とする。

5.1 考察

表6と表7より、システムAとシステムBの単独の評価において、被験者の話を聞いているかという項目で、本システムを搭載したシステムBがシステムAに対し倍以上の高い評価を得た。またシステムAとBの直接的な比較について、今後どちらと話したいかという項目においても、Bに対しポジティブである6点近い評価を得た。この結果により本手法が有効であると確認された。しかし面白さ、人間らしさについてはベースシステムのみシステムAが優れている評価を得た。これはベースシステムによる脈絡のない応答が被験者にとって興味を引いたことと、ベースシステムは学習した人間の文を崩さず自然な応答文を生成するのに対し、本手法に基づくシステムは要約を行うため、不自然な文の繋ぎ方が多くみられたことが影響していると考えられる。

6. まとめ

聞き手としての応答において対話履歴要約の利用が有効であることを確認する為に、実験システムを開発し評価実験を行い、その有効性を確認した。評価実験においては人間同士による1対1の対話と違い、1発話中に含まれる語数が少ないことや、話題転換が頻繁に行われることから、本手法の実験システムが生成する要約文が短く、ただ抜粋をしたのみとみられる応答が多く現れた。今後はよりシステム対人間の対話履歴を要約するのに即したパラメータの検討をするために、人間対人間の対話履歴ではなく、人間対システムによる対話履歴を用いて分析を行う予定である。

参考文献

- 1) 株式会社タカラトミー: <http://www.takaratomy.co.jp/products/healingpartner/index.html>
- 2) Alice: <http://alice.fishbone.jp/>
- 3) 松尾豊, 大澤幸生, 石塚満: 電子掲示板における会話からのハイライト部分の抽出, 第47回人工知能基礎論研究会 特別セッション(予兆・チャンス発見とその応用), (2002).
- 4) 藤本英輝, 高梨克也, 河野泰之, 木戸出正継: 概念的関連性に基く談話の話題転換点分析, 人工知能学会全国大会, 2004-2G3-01, (2004)
- 5) Mecab: <http://mecab.sourceforge.net/>
- 6) 若原基, ジェブカラファウ, 荒木健治: 非同期型チャットアプリケーションに対応した雑談システムの構築と評価, 第33回ことば工学研究会, (2009)