

## 意図性の異なる多様な情報行動を可能とする音声対話システム

## A Spoken Dialogue System for Accepting Information Behavior of Various Intention Levels

高津弘明<sup>\*1</sup> 福岡維新<sup>\*1</sup> 藤江真也<sup>\*1\*2</sup> 林良彦<sup>\*1</sup> 小林哲則<sup>\*1</sup>  
 Hiroaki Takatsu Ishin Fukuoka Shinya Fujie Yoshihiko Hayashi Tetsunori Kobayashi

<sup>\*1</sup>早稲田大学  
 Waseda University

<sup>\*2</sup>千葉工業大学  
 Chiba Institute of Technology

This paper describes a spoken dialogue system for accommodating a user's information behaviors with various levels of information need. Our system, given a set of same-topic news articles, compiles a utterance plan that consists of a primary plan for delivering main news content, and the associated subsidiary plan for supplementing the main content. A primary plan is generated by applying text summarization and style conversion techniques. The subsidiary plan are compiled by considering potential user/system interactions. To enable this, we classified user's possible passive/active behaviors and the corresponding system actions. We confirmed that our system was able to deliver the news content smoothly while dynamically adapting to the change of user's intention levels. The smoothness of a conversation can be attributed to the pre-compiled utterance plan.

## 1. はじめに

ニュース記事に代表されるまとまりのある情報から発話計画を生成し、ユーザーの意図性の变化に合わせて情報を伝達する音声対話システムを提案する。

人間の情報に関する行動のうち、情報獲得・享受に関する行動は、意図性の有無によって、意図的な情報探索と意図性のない受動的な情報行動に分類できる [1]。受動的な情報行動の典型例としてラジオの聴取が挙げられる。ラジオの聴取では、リスナーは受け身で聞いていればよいので楽に情報を取得できるという長所がある。しかし、途中で分からない単語が出てきても質問ができない、聞き逃した内容を聞き返せない、興味がない内容であっても最後まで聴き続けなければならないという短所がある。一方、能動的な情報行動の典型例として質問応答が挙げられる。質問応答は、断片的な情報を得るのには適しているが、まとまった量の情報を得ようとすると、次々と質問を考えながら質問し続けなければならないので現実的でない。そこで、本研究では、受動的な情報獲得を主体としつつも、理解や興味の状況に応じて必要十分な情報にアクセス可能とする音声対話システムを実現する。

受動的な情報獲得における快適さは、伝えられる情報の簡潔さや分かりやすさに影響される<sup>\*1</sup>。そこで、本システムでは、要約技術によりまとまりある情報の要点を明らかにし、口語化処理により会話に適した表現へ書き換える。ここで、要約をもとにした発話計画を主計画と呼ぶ。一方、能動的な情報獲得における快適さは、欲しい情報を欲しいタイミングに、簡単な手段によって取得できることで満たされる<sup>\*1</sup>。そこで、本システムでは、意識的な情報要求 (e.g. 質問) だけでなく、意図性の低い反応 (e.g. 感嘆表現) からも、ユーザーの意図をくみ取り、必要十分な情報を提供する。主計画から省かれた情報や予測される質問に対する回答を副計画として事前に用意しておくことで、ユーザーの意図性の变化に応じた円滑な情報提供を可能とする。

連絡先: 早稲田大学 理工学術院 知覚情報システム研究室

〒162-0042 東京都新宿区早稲田町 27

E-mail: takatsu@pcl.cs.waseda.ac.jp

<sup>\*1</sup> 他にも音声認識率や音声合成の質が情報享受の快適さに影響する

## 2. 主計画生成

主計画はニュースの要点に関する発話計画である。まず、同じ話題のニュース記事集合から主計画に含める文を抽出する (重要文抽出)。次に、抽出した文の提示順序を決定する (整列)。そして、重要文から不要な箇所を取り除き、文自体を短く要約する (文圧縮)。最後に、要約文を会話表現に書き換える (口語化)。

## 2.1 要約・口語化コーパス

人間がニュース記事の内容を会話で伝えるとき、どの内容を選択し、どのように話すか調査するために、要約・口語化コーパスを作成した。100 トピックについて 3 人の作業者が要約と口語化を行った。各トピックは同じ話題について記述した 2 つ以上の記事から構成され、各記事は 5 文以上からなるものを選んだ。なお、ニュース記事は毎日新聞、産経新聞、読売新聞、NHK、nifty から収集した。

要約は次の二つの観点で作成した。一つは、ニュースの見出し的な内容またはニュース記事の核となる情報を 1~4 文選択したものである (以降、要約 A)。もう一つは、ユーザーの関心を引けそうな内容または記事への理解を促進させる内容を要約 A に加えて 1~3 文選択したものである (以降、要約 B)。文選択を行った後、作業者は選択した文から不要な情報を文節単位で取り除く。次に、作業者は選んだ情報を会話表現に書き換える。口語化作業では、要約で選ばれた情報を増やしたり損ねたりしないように書き換える。また、情報の伝達相手としては友達のようなキャラクターを想定し、話し手は中性的なキャラクターとする (「~らしいわよ」(女性的)「~だぜ」(男性的)のような、特定の性別・キャラクターを想起させるような表現は使わない)。

文選択に関する作業者間の一致度を Fleiss' kappa を用いて評価したところ、 $\kappa$  係数の値は、要約 A が 0.444、要約 B が 0.357、要約 B-A (要約 B で加えられた文のみ) が 0.172 であった。この結果から見出し的な内容の選択に比べ関心を引けそうな内容の選択は作業者ごとに偏りがあることが分かる。同様に、同じ文を選択したときの文節選択に関する一致度に関しても評価したところ、 $\kappa$  係数の値は、要約 A が 0.232、要約 B が 0.220、要約 B-A が 0.227 であった。この結果から作業者の文節選択傾向は要約の種類によらないことが分かる。

表 1: 変数の定義

$x_{(i,m)}$	文 $(i, m)$ が選択されたかどうか
$y_{(i,m)(j,n)}$	文 $(i, m)$ と文 $(j, n)$ が両方選択されたかどうか
$z_l$	単語 $l$ が被覆されたかどうか
$K$	最大要約長
$a_{(i,m)l}$	文 $(i, m)$ が単語 $l$ を含むかどうか
$b_l$	単語 $l$ の重要度
$c_{(i,m)}$	文 $(i, m)$ の長さ (要約長が文数の場合は 1)
$r_{(i,m)}$	文 $(i, m)$ の主題への関連度
$s_{(i,m)(j,n)}$	文 $(i, m)$ と文 $(j, n)$ の類似度
$M_i$	記事 $i$ の文の数
$I$	記事の総数
$L$	単語の総数

$$\max. \quad \sum_{l=1}^L b_l z_l \times \left( \sum_{i=1}^I \sum_{m=1}^{M_i} r_{(i,m)} x_{(i,m)} - \sum_{(i,m) < (j,n)} s_{(i,m)(j,n)} y_{(i,m)(j,n)} \right) \quad (1)$$

s.t.

$$\forall i, j, m, n, l: \quad x_{(i,m)} \in \{0, 1\}, \quad y_{(i,m)(j,n)} \in \{0, 1\}, \quad z_l \in \{0, 1\} \quad (2)$$

$$\sum_{i=1}^I \sum_{m=1}^{M_i} c_{(i,m)} x_{(i,m)} \leq K \quad (3)$$

$$\forall l: \quad \sum_{i=1}^I \sum_{m=1}^{M_i} a_{(i,m)l} x_{(i,m)} \geq z_l \quad (4)$$

$$\sum_{i=1}^I x_{(i,1)} = 1 \quad (5)$$

$$\forall i, j, m, n: \quad y_{(i,m)(j,n)} - x_{(i,m)} \leq 0 \quad (6)$$

$$\forall i, j, m, n: \quad y_{(i,m)(j,n)} - x_{(j,n)} \leq 0 \quad (7)$$

$$\forall i, j, m, n: \quad x_{(i,m)} + x_{(j,n)} - y_{(i,m)(j,n)} \leq 1 \quad (8)$$

## 2.2 重要文抽出

要約問題を整数計画問題として定式化した。複数文書要約のモデルとしてよく用いられる最大被覆モデル (MCM) [2–4] に McDonald のモデル (関連性モデル; Rel) [5] を組み合わせたモデル (MCM+Rel) を提案する。各変数の説明を表 1 に示す。ここで、文  $(i, m)$  は記事  $i$  の  $m$  番目の文を表す。最大被覆モデルは、被覆した概念 (ユニグラムやバイグラム) の重要度の和が最大となる文集合を選択するモデルである。一方、関連性モデルは、主題への関連度は高いが、文同士の類似度が低くなるような文集合を選択するモデルである。制約として、ほぼ全ての作業者がニュース記事の一文目を要約に含めていたことから、複数記事の内いずれかの記事の一文目を必ず要約に含めるような制約 (式 5) を加える。最適化問題は分枝限定法で解く。

動詞、名詞、形容詞を対象として単語の重要度  $b_l$  を計算する (ただし、形式名詞、副詞的名詞、“する”は除いた)。単語の重みは、頻度 (Freq) や TF-IDF、RandomForest で与える。RandomForest のパラメータは次の 2 つの観点で学習した。一つは、作業者が対象単語を要約に含めたかどうかの二クラス分類問題として学習し、単語が要約に含まれる確率を単語の重みとしたもの (RFC)。もう一つは、複数の作業者が要約に含めた単語の重みが高くなるように回帰で学習したものである (RFR)。3 人の作業者が要約に含めた場合の単語の重みを 3/3 とし、2 人の作業者が要約に含めた場合の単語の重みを 2/3、1 人の作業者が要約に含めた場合の単語の重みを 1/3 とした。素性として、文書集合における単語の頻度や単語の TF-IDF 値、タイトルに含まれるかどうか、固有表現かどうかなどの情報を利用した。

表 2: 重要文抽出の ROUGE-1 による評価

	要約 A			要約 B		
	作業者 1	作業者 2	作業者 3	作業者 1	作業者 2	作業者 3
MCM (Freq)	73.7	76.8	78.7	76.2	78.6	80.4
MCM (TF-IDF)	66.2	70.7	74.5	74.5	76.0	79.8
MCM (RFC)	76.2	81.9	82.9	82.1	83.4	84.8
MCM (RFR)	77.5	81.4	84.4	82.6	83.7	85.3
Rel (BoW)	73.6	80.3	82.9	72.2	74.8	78.1
Rel (TF-IDF)	72.6	80.8	82.6	69.2	74.4	74.5
Rel (PV-DM)	71.4	76.3	81.5	69.9	74.7	76.0
MCM+Rel (RFC,BoW)	78.7	82.0	84.8	83.1	84.2	85.8
MCM+Rel (RFR,BoW)	<b>79.3</b>	<b>83.1</b>	<b>84.9</b>	<b>83.8</b>	<b>84.4</b>	<b>86.2</b>

主題への関連度  $r_{(i,m)}$  は、 $r_{(i,m)} = 1/(\text{文}(i, m) \text{ の出現位置}) + (\text{文}(i, m) \text{ と文書集合の類似度})$  で与える。文と文書集合の類似度と文間の類似度  $s_{(i,m)(j,n)}$  は、単語の Bag-of-Words (BoW)、単語の TF-IDF ベクトル (TF-IDF)、300 次元の Paragraph ベクトル (PV-DM) [6] のいずれかを用いたコサイン類似度で与える。

実験では、要約 A, B のデータセットを使用し、ROUGE-1 による評価実験を行った。結果を表 2 に示す。ただし、最大要約長  $K$  は作業者が選択した文数とする。また、MCM (RFC) と MCM (RFR)、および、MCM+Rel (RFC) と MCM+Rel (RFR) は、10 分割交差検定で単語の重みを計算し、各 ROUGE-1 の値を平均した値である。実験結果は、提案手法 (MCM+Rel) が最も高い ROUGE-1 を示した。提案手法の ROUGE-1 の値は、およそ 80% から 85% 程度であり、このことから、提案手法を用いることで、要約に含めるべき内容の 80% から 85% を抽出できることが分かった。単語の重みの与え方に関する比較では、RFR の方が良い結果を示し、関連性モデル (Rel) における類似度の与え方に関する比較では、BoW が最も良い結果を示した。

## 2.3 整理

重要文抽出で抽出された文の提示順序を決定する。文の整理問題は、主にテキスト生成の分野で取り組まれてきたが、近年では、自動要約の分野でも重要な課題として認識されている。岡崎らは記事が書かれた時間情報を利用して並び替える手法を提案した [7]。Lapata は文中での動詞や名詞とその他の係り受け関係を考慮した統計的なモデルに基づいて文の順序を決定する手法を提案した [8]。

ここでは、ある記事のある文が別の記事のどの文に該当するかをコサイン類似度に基づいて計算し、最もスコアが高くなる順序を採用するという方法をとった。

$$\text{sequence} = \max_{seq \in S} \text{Score}(seq) \quad (9)$$

$$\text{Score}(seq) = \sum_{(i,m) \rightarrow (j,n) \in seq} f((i,m), (j,n)) \quad (10)$$

$$f((i,m), (j,n)) = \begin{cases} g_{l>m}(\text{sim}((i,l), (j,n)), \text{sim}((i,m), (j,l))) \\ g(\text{sim}((i,m), \text{before}(j,n)), \text{sim}(\text{after}(i,m), (j,n))) \end{cases} \quad (11)$$

$$\begin{cases} g_{l>m}(\text{sim}((i,l), (j,n)), \text{sim}((i,m), (j,l))) \\ g(\text{sim}((i,m), \text{before}(j,n)), \text{sim}(\text{after}(i,m), (j,n))) \end{cases} \quad (12)$$

ここで、 $S$  は文の順序候補の集合、 $f((i,m), (j,n))$  は文  $(i,m)$  が文  $(j,n)$  よりも前に来うるスコア、 $g, h$  は  $\sum$  または  $\max$ 、 $\text{sim}$  は BoW のコサイン類似度を表す。また、 $\text{before}(j,n)$  は記事  $j$  の文  $n$  よりも前に出現する文の BoW を表し、 $\text{after}(i,m)$  は記事  $i$  の文  $m$  よりも後ろに出現する文の BoW を表す。制約として、いずれかの記事の一文目が最初になるように定めた。

要約 A, B のデータセットに対する実験結果を表 3 に示す。ここでは、式 13、 $g, h$  として  $\max$  を用いたが、他の設定でもさ



表 7: 「現在」か「過去」か

	精度	再現率	F 値
現在	90.7	92.2	91.4
過去	93.9	92.6	93.2
平均	92.3	92.4	92.3

表 8: 「進行」かどうか

	精度	再現率	F 値
進行である	90.7	92.2	91.4
進行でない	93.9	92.6	93.2
平均	92.3	92.4	92.3

表 9: ユーザーアクションとシステムアクション

ユーザーアクション		発話意図	システムアクション
大分類	細分類		
相槌	弱い	無意識	継続
	強い	関心	肯定・感情共有・補足説明
「え」相当表現		確認	反復・真偽判定
		驚き	反復・感情共有
		定義質問	定義型質問応答
		理由質問	why 型質問応答
反復	名詞	確認	反復・真偽判定
		驚き	反復・感情共有
		自問	肯定・継続
		定義質問	定義型質問応答
	述語 フレーズ	確認	反復・真偽判定
		驚き	反復・感情共有
		自問	肯定・継続
		理由質問	why 型質問応答
質問	明示的	質問	質問応答
	半明示的		
感想		感想	感情共有

か「過去」かの二値分類問題と「進行」かどうかの二値分類問題の識別を行った。素性として、対象文節の形態素情報 { 表層形, 品詞大分類, 品詞細分類, 活用形, 活用型 }, 係り先の文節が述語である場合その文節の形態素情報、文末文節の形態素情報、文中の時間表現の形態素情報を利用した。10 分割交差検定で評価した結果を表 7,8 に示す。これらの識別結果の組み合わせで対象述語の時制を決定する。

### 3. 副計画生成

副計画は主計画の情報を補うための発話計画である。テンポの良い会話を実現するためには、想定されるユーザーの反応に対して事前に発話計画を立てておく必要がある。ここでは、質問や反復、感想などのユーザーアクションに対して、発話意図を推定し、システムが返すべき応答を計画する。

ユーザーアクションとその発話意図に対するシステムアクションの分類を表 9 に示す。発話意図は、単語の珍しさや情報の意外性、理由や定義の有無、固有表現クラス、ユーザー発話のピッチ (F0) 変化の傾きなどの情報を用いてルールベースで判定している。

質問応答としては、ファクトイド型と定義型、why 型、真偽判定型を扱う [13]。なお、予想外の質問に対してはオープンメインの質問応答で対処する。

主計画と副計画からなる発話計画に従った会話の例を図 2 に示す。ここで、 $\epsilon$  はユーザーの反応を見ずに状態が遷移することを表す。ユーザーが受動的な態度をとり続ける限り、システムは主計画に従った情報伝達を行うが、ユーザーの理解や興味の状態に応じて、副計画に遷移し、補足説明を行う。

### 4. おわりに

快適な情報享受は、質問応答のような意図性の高い能動的な情報行動だけでも、文書閲覧・聞き流しのような意図性のない受動的な情報行動だけでも実現できず、これらの状態を自由に行き来することにより達成される。本研究では、音声対話が意

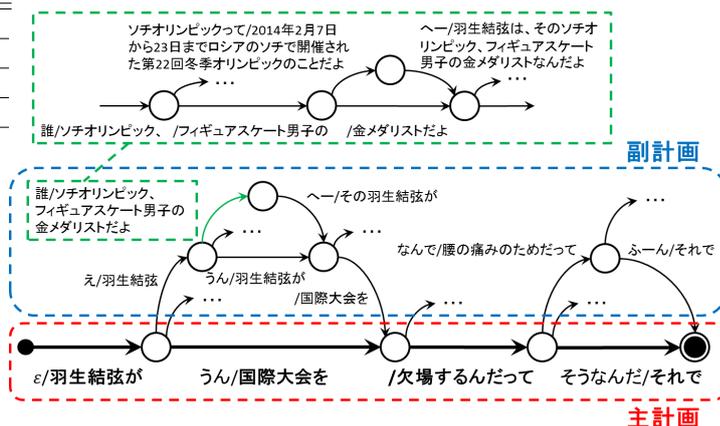


図 2: 発話計画に従った会話例

図性的変化に応じた情報行動を行うのに適したメディアであるという立場から、ニュース記事を例題としてまとまりのある情報を利用者が快適に享受するための仕組みを提案し、音声対話システム上での実現を試みた。

### 参考文献

- [1] D.O.Case: "Looking for Information", A Survey of Research on Information Seeking, Needs, and Behavior, Second Edition, Academic Press, 2007.
- [2] E.Filatova and V.Hatzivassiloglou: "A Formal Model for Information Selection in Multi-Sentence Text Extraction", in Proceedings of the 20th International Conference on Computational Linguistics, pp.397403, 2004.
- [3] W.Yih, J.Goodman, L.Vanderwend, and H.Suzuki: "Multi-Document Summarization by Maximizing Informative Content-Words", in Proceedings of the 20th International Joint Conference on Artificial Intelligence, pp.17761782, 2007.
- [4] D.Gillick, and B.Favre: "A Scalable Global Model for Summarization", in Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing, pp.1018, 2009.
- [5] R.McDonald: "A Study of Global Inference Algorithms in Multi-Document Summarization", in Proceedings of the 29th European Conference on Information Retrieval, pp.557564, 2007.
- [6] Q.Le and T.Mikolov: "Distributed Representations of Sentences and Documents", in Proceedings of The 31st International Conference on Machine Learning, pp.11881196, 2014.
- [7] N.Okazaki, Y.Matsuo and M.ishizuka: "TISS: An Integrated Summarization System for TSC-3", in Working Notes of the Fourth NTCIR Workshop Meeting, pp.436-443, 2004.
- [8] M.Lapata: "Probabilistic Text Structuring: Experiments with Sentence Ordering", in Proceedings of the 41st Meeting of the Association of Computational Linguistics, pp.545-552, 2003.
- [9] K.Knight and D.Marcu: "Summarization beyond sentence extraction: A probabilistic approach to sentence compression", Artificial Intelligence 139, pp.91-107, 2002.
- [10] T.Nomoto: "A Generic Sentence Trimmer with CRFs", in Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, Human Language Technologies, Columbus, pp.299-307, 2008.
- [11] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: "会話によるニュース記事伝達のための情報選択", 言語処理学会第 22 回年次大会発表論文集, pp.199-202, 2016.
- [12] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: "会話によるニュース記事伝達のための口語化における述語の書き換え", 言語処理学会第 22 回年次大会発表論文集, pp.1033-1036, 2016.
- [13] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: "快適な情報享受を可能とする音声対話システム", 言語処理学会第 22 回年次大会発表論文集, pp.302-305, 2016.
- [14] S.Fujie, I.Fukuoka, A.Mugita, H.Takatsu, Y.Hayashi and T.Kobayashi: "A Spoken Dialog System for Coordinating Information Consumption and Exploration", ACM SIGIR Conference on Human Information Interaction and Retrieval, pp.253-256, 2016.