

ユーザの興味プロフィールに基づく音声ニュースシステムにおける 複数関連ニュース提供の対話的制御

Dialog Strategy for Reading Related News in Personalized Interactive News System Using User's Interest Profile

三吉 達夫[†] 東原 智幸[†] 渥美 雅保[†]
Tatsuo Miyoshi Tomoyuki Higashihara Masayasu Atsumi

1. 背景

近年、ユーザの興味・嗜好に合わせた情報提供をおこなってくれるシステムへの関心が高まり、研究も盛んにされている。既存研究として、ユーザの記事閲覧履歴から記事中に高い頻度で出現する単語をユーザの興味語として登録し、その興味語の含まれる記事を優先的に提供する MPV Plus[1]などがある。これまで、我々は興味探索型対話方式を組み込んだニュース提供システム[2]を提案してきた。興味探索型対話方式とは、ニュース提供の途中でユーザとの簡単な対話を繰り返し、ユーザの関心の度合いがどのように変化しているのかを調べていき、関心の度合いに応じてニュース提供の流れを動的に制御する対話方式である。しかし、従来の興味探索型対話方式では、一つのニュース記事内での読み上げ順序の制御のみで、複数のニュース記事への読み上げ順序の制御はできなかった。本論文では、この対話方式にトピック展開の時系列を表現するストリームを導入する。これにより、ユーザの関心の度合いから複数の関連ニュース読み上げもできるように改良をおこなった。そして、関連ニュース読み上げの効果をいくつかの事例を通して確かめた。

2. ニュース提供システム

2.1 ニュース提供の流れ

ニュース提供におけるユーザとシステムとのやりとりについて簡単に説明する。ユーザはニュースシステムに対して「何かニュースはありますか?」という発話によりニュースを要求する。要求を受け取ったニュースシステムは、ユーザごとに用意された興味プロフィールを利用して、ユーザの興味と新着ニュース群との類似度を計算し、新着ニュース群の中から類似度の高いニュースを提供する。興味プロフィールには、ユーザが過去に読んだニュースから抽出された興味構造が表現されている。ニュースは XML 形式で記述されており、表 1 のようなタグが付けられている。一つのニュース記事は複数のフレーズによって構成され、各フレーズは複数の文でまとまった内容を表して

表 1: タグの説明

タグ名	タグの説明
<root>	xmlファイルのルートタグ
<news>	タグ内1つのニュースデータ単位
<title>	ニュースの見出し
<date>	ニュース記事の日付
<source>	ニュース記事の提供元
<category>	ニュースの分類
<description>	ニュース記事を説明
<phrase no="n">	ニュース内容のフレーズ(複数)を表現したもので、nには段落番号を属性値として記述する
<sentence no="n">	ニュース内容を文で表現したものを
<newsfile>	関連性のあるニュースファイルタグ内に記述する

```

<root>
<news>
<title><title>
<date></date>
<source></source>
<category></category>
<description>
<phrase no="1">
<sentence no="1"></sentence>
<sentence no="2"></sentence>
</phrase>
<phrase no="2">
<sentence no="3"></sentence>
</phrase>
</description>
</news>
<newsfile></newsfile>
<newsfile></newsfile>
</root>
    
```

図 1: ニュースのデータ構造

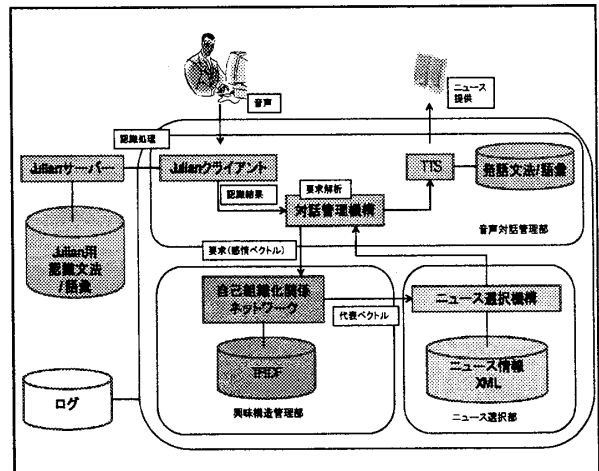


図 2: システム構成図

いる(図1)。今回、フレーズタグは手作業で付与した。また、関連ニュース先については<newsfile>タグに記述されている。

2.2 音声ニュース提供システムのシステム構成

図2は本システムのシステム構成を表している。本システムでは、ユーザの興味プロフィールを関係SOM[3]で表す。この関係SOMには、ユーザのニュースに対する興味を代表するTFIDFと、感性的興味との関連が表現されている。ニュース提供では、関係SOMの出力するTFIDFと新着ニュースのTFIDFとの類似度が計算され、類似度の高い新着ニュースが選択される。そして、後述する関連ニュース探索戦略と後戻り戦略という2種類の提供戦略を組み合わせることによりニュース読み上げを動的に制御する。また、ユーザの反応(返答)から関心の度合いが低くなった場合に、ニュース提供を終了する。関心の度合いを表す C_{nw} は次式により与えられる。

$$C_{nw} = \frac{\sum_i C(i)}{N_p}$$

ここで、 N_p はフレーズの総数、 $C(i)$ は i 番目のフレーズへの関心の有無(関心有り: 1 関心無し: 0)を表している。また、 $C(i)$ の初期値は、ユーザの興味を表現するTFIDFとの類似度として設定される。

興味探索型対話方式では、読み上げたフレーズに対してユーザが関心を示した際に、そのフレーズに関連したニュースを参照していく仕組みを取り入れている。これにより、ユーザは自分の知りたい情報に対してニュース提供の流れを変化させることができる。本システムでは、このニュース提供の流れをトピック展開時系列と呼ぶ。トピック展開時系列では、次のトピック展開関心度 C_{T0} の値をもとにして関連ニュースをどこまで読み上げるかを制御する。

$$C_{T0} = \sum_{t=0}^n \gamma^t \text{val}(t)$$

[†] 創価大学大学院工学研究科情報システム工学専攻

ここで、 $val(t)$ の値は t 回前に提供されたフレーズに対して、ユーザが関心を示した場合に1、関心を示さなかった場合には0になる。また、 γ は以前に読んだフレーズの影響を表す割引係数である。トピック展開関心度により、ユーザが関心を持った話題に対しての関心の度合いの変化がわかる。 C_{ro} がある値より低くなったときには、関連ニュースの提供を停止する。トピック展開時系列の導入により、ユーザは自分の知りたい情報を関連記事からも取得することができる。

2.3 関連ニュース探索戦略

ニュースデータベース内のニュースのTFIDFの値を利用して、現在読み上げているニュースAとそれに関連付けられたニュースBとのニュース間類似度SNと、現在読み上げているフレーズ a_i とニュースBとのニュース間類似度SNPを次式の内積計算により計算する。

$$SN = TFIDF(A) \cdot TFIDF(B)$$

$$SNP = tfidf(a_i) \cdot TFIDF(B)$$

次に、SNとSNPから合成ニュース間類似度RNPを計算する。

$$RNP = \alpha \times SN + \beta \times SNP$$

このRNPの値がもっとも大きくなるニュースを関連ニュース先として選択する。さらに、読み上げるフレーズを決定するため、次式で与えられるフレーズ間類似度SPを計算する。

$$SP = tfidf(a_i) \cdot tfidf(b_j)$$

ここで、 b_j は関連ニュースBのj番目のフレーズを表している。このSPの値がもっとも大きくなるフレーズをユーザに提供する。但し、SPの値が閾値 Th よりも高い場合には、現在の読み上げているフレーズと重複する内容を述べている可能性が高いため、探索から除外する。

2.4 後戻り戦略

後戻り戦略は次の3つの戦略から構成される。

- トピック展開時系列内のニュースの中で、もっとも C_{NW} の値が大きいニュースに戻る。
- 1つ前のニュースに戻る。
- 最初のニュースに戻る。

トピック展開関心度 C_{ro} の値が低くなり、関連ニュースの提供を停止した場合には、(a)~(c)のいずれかのルールに従って関連先から戻り、ニュース提供を続けていく。

3. ニュース提供例

本稿で提案している興味探索型対話方式と、フレーズを順番に読んでいく逐次読み上げでどのような違いが生まれるのかをニュース提供例を通して比較をおこなう。なお、ニュース記事はgooニュースより引用した。

図3と図4は、逐次読み上げと本手法のニュース提供例を示している。逐次読み上げでは、現在読んでいるニュースの提供が終了してからでないと関連ニュースを読み上げることができない。そのため、特定の内容のみを詳しく知りたい場合に逐次読み上げではユーザの要求に対して柔軟に対応することができない。一方、本手法の場合には、フレーズごとの関心の有無によって情報を関連ニュースからも取得することができ、ニュース情報の補足することが可能である。この提供例では、どちらもニュースAと関連ニュースBをユーザに提供しているが、興味探索型対話方式の場合にはニュースAからフレーズの一つ読んだ後すぐにニュースBを読み上げていることがわかる。また、そのトピック展開時系列を履歴として保存することで、次回からのニュース提供に反映させることが可能になる。

4. まとめ

本論文では、ニュース提供システムにおける複数関連ニュース読み上げ制御のための提供戦略を提案した。この提供戦略により、ユーザの関心の度合いの変化に応じて、複数の関連ニュースからも情報を取得することができる。ユーザがニュース提供システムを利用していくなかで、ニュースを詳しく知りたい場合やニュースの概要のみを知りたい場合、特定の箇所のみ関心がある場合など、様々な利用形態が考えられる。今後は、本システムがユーザの利用形態に対応しながら、提供の流れを変化させていくことによる有用性を検証していきたい。

```

U:「何かニュースはありますか？」
ニュース提供を開始する(ニュースA)
S:「野捕獲の失血死 長井さん、背後から射殺」についてのニュースです。
ミャンマーで鉄道ジャーナリストの長井健司さん(50)が取材中に射殺された事件で、警視庁は4日、各林大(東京都三鷹市)で遺体を司法解剖した。」……(フレーズ1)
S:「海外で殺人や誘拐などの重要犯罪に巻き込まれた場合、日本の警察が捜査できる司法の国外犯規定に基づき、警視庁は長井さんの住居地を管轄する中野署に捜査本部を設置。」……(フレーズ2)
S:「司法解剖の結果では、銃弾による傷は入り口の背中部分に直径約1センチだったのに対し、出口の腹部分に約1〜2センチと大きく、開けた状態だったため、発砲距離は不明だった。」……(フレーズ3)
S:「ただ、先に解剖した埋地医師は、長井さんが契約していたAFP通信社(東京都港区)の山田善社長(46)に「距離は1ヤード(約90センチ)程度」と説明したという。長井さんが路上に倒れる瞬間をとらえたビデオ映像などからも、至近距離で射殺された可能性が高い。関連するニュースも読みますか？」……(フレーズ4)
U:「はい」
関連するニュースを提供(ニュースB)
S:「ミャンマーで射殺されたジャーナリスト長井健司さん(50)について、警視庁は4日午後、都内の病院で、約4時間にわたって、遺体を司法解剖した。その結果、背後から撃たれ、銃弾が肋骨から入り、右上半部に貫通していたことが分かった。」……(フレーズ1)
S:「死因は銃弾による失血死の疑いで、遺体は同日午後7時すぎ、遺族前に引き渡された。」……(フレーズ2)
S:「調べによると、左肋骨部の銃弾による傷跡は直径約1センチで、右上半部は直径2〜3センチ。銃弾は体内を縦やかな上向き角度で貫通した。」……(フレーズ3)
(中略)
S:「関連するニュースも読みますか？」
U:「はい」
関連するニュースを提供(ニュースC)

```

図3：逐次読み上げによるニュース提供例

```

U:「何かニュースはありますか？」
ニュース提供を開始する(ニュースA)
S:「野捕獲の失血死 長井さん、背後から射殺」についてのニュースです。
ミャンマーで鉄道ジャーナリストの長井健司さん(50)が取材中に射殺された事件で、警視庁は4日、各林大で遺体を司法解剖した。興味ありますか？」……(ニュースA、フレーズ1)
U:「はい」
ニュースAの関心の度合いを1にする
(関連ニュースBを探検してニュースBのフレーズを読み上げる)
S:「調べによると、左肋骨部の銃弾による傷跡は直径約1センチで、右上半部は直径2〜3センチ。銃弾は体内を縦やかな上向き角度で貫通した。詳しく知りたい？」……(ニュースB、フレーズ2)
U:「はい」
(後戻り戦略に従って、前に読んでいたニュースに戻る)
S:「海外で殺人や誘拐などの重要犯罪に巻き込まれた場合、日本の警察が捜査できる司法の国外犯規定に基づき、警視庁は長井さんの住居地を管轄する中野署に捜査本部を設置。面白いですか？」……(ニュースA、フレーズ2)
U:「はい」
ニュースAの関心の度合いを3にする
(関連するニュースを探検してニュースA内で次に関心の高いフレーズを読み上げる)
S:「司法解剖の結果では、銃弾による傷は入り口の背中部分に直径約1センチだったのに対し、出口の腹部分に約1〜2センチと大きく、開けた状態だったため、発砲距離は不明だった。興味ありますか？」……(ニュースA、フレーズ3)
U:「はい」
ニュースAの関心の度合いを3にする
S:「先に解剖した埋地医師は、長井さんが契約していたAFP通信社(東京都港区)の山田善社長(46)に「距離は1ヤード程度」と説明したという。長井さんが路上に倒れる瞬間をとらえたビデオ映像などからも、至近距離で射殺された可能性が高い。もっと聞きたい？」……(ニュースA、フレーズ4)
U:「はい」
関連するニュースAの関心の度合いを4にする。
(関連ニュースCを探検する)
S:「29日深夜のフジテレビの放送では、経費のいるヤンゴンや心部の路上に、軍のトラックが停車。遺体の軍服を着た兵士が次々に飛び降り、新築に向かって走り出した。そのうちの1人が、半ズボン姿の長井さんの前に立ち止まり、1〜2メートルの至近距離で発砲。長井さんははたき飛ばされるように倒れた。兵士は長井さんを介抱しようともせず、銃聲のあとを去っていった。もっと知りたいですか？」……(ニュースC、フレーズ2)
(関連ニュースDを探検する)

```

図4：興味探索型対話方式によるニュース提供例

参考文献

- 河合由起子, 熊本忠彦, 田中克己, “印象と興味に基づくユーザ選好のモデル化手法の提案とニュースサイトへの応用”, 日本知能情報ファジィ学会誌, vol.18 No.2, pp.173-183, 2006
- 三吉達夫, 東原智幸, 渥美雅保, “個人適応的音声ニュース提供システムにおける対話戦略”, FIT2007, E-065, pp299-300
- 東原智幸, 三吉達夫, 渥美雅保, “ウェブニュース提供のための自己組織化関係ネットワークと格重み付き単語頻度ベクトルを用いたユーザの興味構造表現”, 第70回全国大会論文集掲載予定