

個人適応的音声ニュース提供システムにおける関連ニュース提供戦略
 A Strategy for Providing Related News in Personalized Interactive News Systems
 三吉 達夫† 東原 智幸† 渥美 雅保†
 Tatsuo Miyoshi Tomoyuki Higashihara Masayasu Atsumi

1. はじめに

現在、我々の研究室では、音声を通じてユーザにニュースを対話的に提供してくれるシステムの開発をおこなっている。大量のニュース記事の中から関心のあるニュースを提供するには、ユーザの興味・嗜好に合わせたニュースを提供してくれる機能がもとめられる。普段テキスト文書としてニュースの続報記事を読むときには、知っている内容については読み飛ばし、その記事の中から新しい情報のみに注目して読んでいる。しかし、音声によるニュース提供の場合には、画面上の記事のような拾い読みに対応することができないため、効率的に聞きたい情報を取得できる提供方法が必要となってくる。

そこで、本システムでは、2つの機能を組み込むことでユーザの関心に合わせたニュースの提供を目指している。1つは、ユーザの興味・嗜好をモデル化し、ニュースの提供を個人適応させる機能である。もう1つは、複数の関連ニュース記事に対して、ニュースへの関心を追跡し、提供の流れをインタラクティブに制御していく機能である[1]。本稿では、この機能の中から複数関連ニュース記事上の提供の流れを制御する機能について述べる。

2. 個人適応的音声ニュース提供システム

2.1 システム構成

本システムは、ユーザからの要求を認識しニュースを音声で提供する対話管理機構、ユーザの興味構造モデルと履歴を管理するユーザ管理機構、ニュースデータ情報を管理するニュースデータ管理機構で構成される(図1)。

2.2 ユーザの興味構造モデル

ユーザ管理機構では、ユーザごとにニュースへの興味構造モデルを管理している。ユーザの興味は、TF-IDF値による単語頻度ベクトルと感性的興味ベクトルによる自己組織化関係ネットワーク[2]で表現している。図2は興味構造モデルを示しており、「面白いニュースはありますか」などの感性的興味を入力として与える。そして、感性的興味語に対応したユーザの興味ベクトルを取得できる。また、「なにかニュースはありますか」というニュースの要求をした場合には、感情ベクトルの要素を全て1にして、興味ベクトルを検索する。

3. 関心度指標の定義と提供戦略

提供するニュースには、複数の関連ニュースによりニュースが、本システムでは、あるトピックに対する一連のニュース展開のなかで、トピックに出現するニュースへの関心度と、そのニュース展開への関心度という2つの観点からニュースの関心度を与えている。本章では、このニュース関心度の計算とそれを利用した提供戦略について述べる。

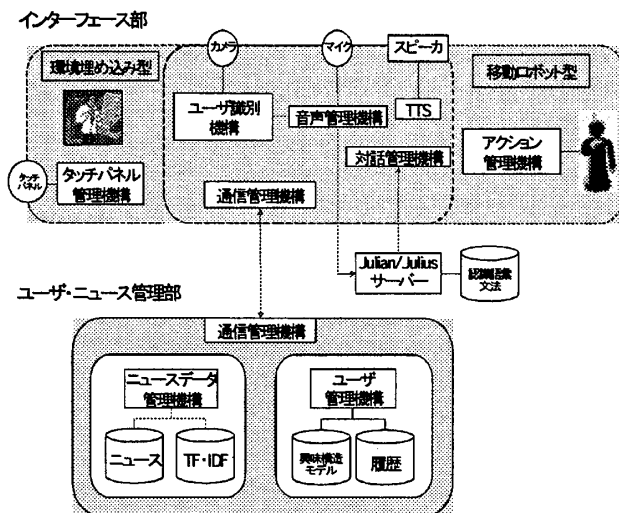


図1：個人適応的音声ニュース提供システムの構成

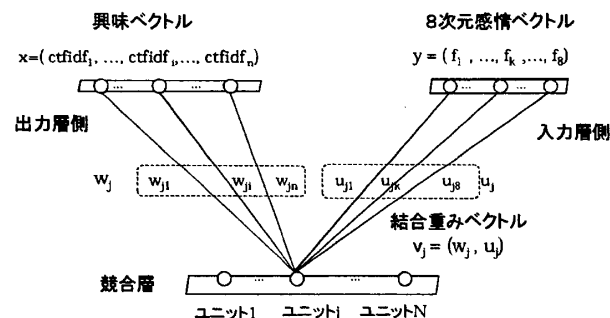


図2：ユーザの興味構造モデル

3.1 ニュースへの関心度C(A_i)とその更新

ニュース集合{A₁, A₂, ..., A_n}に対して、ニュースA_iへの関心の度合いC(A_i)を計算する。

$$C(A_i) = \frac{\sum_{j=1}^{N_p} C(A_i, p_{ij})}{N_p}$$

N_pは段落の総数、p_{ij}はニュースA_iのj番目の段落を表している。また、C(A_i, p_{ij})はニュースA_iの段落jに対する関心度、関心があるとき1、関心がないときは0になる。C(A_i, p_{ij})の初期値は、ユーザの興味単語ベクトルとの類似度により設定する。

3.2 トピック展開時系列への関心度 C(To, t)

時点t_nにおけるトピック展開を(A_i, p_{ij}, t_n)と表すとき、トピック展開時系列To={(A₁, p₁₁, t₁), (A₂, p₂₂, t₂), ..., (A_i, p_{ij}, t_n)}に対して、

†創価大学大学院工学研究科情報システム工学専攻

$$C(T_o, t) = \sum_{\tau=0}^n \gamma^{\tau} val(A_i, p_{ij}, t - \tau)$$

をトピック展開関心度として定義する (γ は割引係数)。
 $val(A_i, p_{ij}, t - \tau)$ は評価関数で, τ 回前に提供した内容 (ニュース A_i の段落 j) にユーザが関心を示したかによって, 興味が有るときは 1, 興味が無いときには -1 になる. また, トピック展開関心度 $C(T_o, t)$ とニュース関心度 $C(A_i)$ は連動している. $val(A_i, p_{ij}, t - \tau)$ に値が入ると同時に, $C(A_i, p_{ij})$ の値も 1 もしくは 0 に更新される.

$$C(A_i, p_{ij}) = \begin{cases} 1 & \dots\dots val(A_i, p_{ij}, t) = 1 \\ 0 & \dots\dots val(A_i, p_{ij}, t) = -1 \end{cases}$$

この計算式により, トピック展開への関心履歴を反映した現在の関心によって提供を制御する. $C(T_o, t)$ の値が小さくなった場合, そのトピック展開を停止する.

3.3 ニュース提供戦略

図 3 は提供戦略を利用した提供の流れを表したものである. 本システムでは, ユーザが関心を持ったトピック展開に対して, 関連ニュース探索戦略と後戻り戦略という 2 つの提供戦略を組み合わせる関連ニュースを提供する.

3.3.1 関連ニュース探索戦略

ユーザが関心を示したトピック展開については, 関連ニュースを探索することによりトピック展開を広げていく. 関連ニュースの探索には, 単語 TF-IDF ベクトルを利用することで記事間の類似度を調べている (図 4). 以下に関連ニュース探索手順を示す.

(1) ニュース全体の類似度 SN の計算

現在提供しているニュース A_m の単語 TF-IDF ベクトルとニュース A_i の単語 TF-IDF ベクトルの内積 SN を計算する. SN の値が大きいくほどニュース A_m との類似度が高いことを表している.

$$SN = TFIDF(A_m) \cdot TFIDF(A_i)$$

(2) ニュース A_m の段落との類似度 SNP の計算

直前に提供したニュース A_m の段落内容とニュース A_i との類似度 SNP を計算することで, 現在の段落内容に関連の深いニュースを探索する.

$$SNP = tfidf(A_m, p_{mj}) \cdot \sum_{j \in provided} tfidf(A_i, p_{ij})$$

$tfidf(A_m, p_i)$ は現在提供しているニュース A の段落 p の単語 TF-IDF ベクトル, $tfidf(A_i, p_j)$ はニュース A_i の段落 j の単語 TF-IDF ベクトルである. provided は提供済み段落の集合を表しており, SNP の計算から除外している.

(3) ニュース A_k の選択

SN と SNP から RNP が最も大きくなるニュースを提供する.

$$RNP = SN \times SNP$$

$$A_k = \arg \max RNP$$

(4) 類似度 SP による提供段落の選択

選択されたニュース A_k に対して, 直前に提供したニュース A_m の段落 p_{mj} を利用した SP により提供する段落 j を決定する.

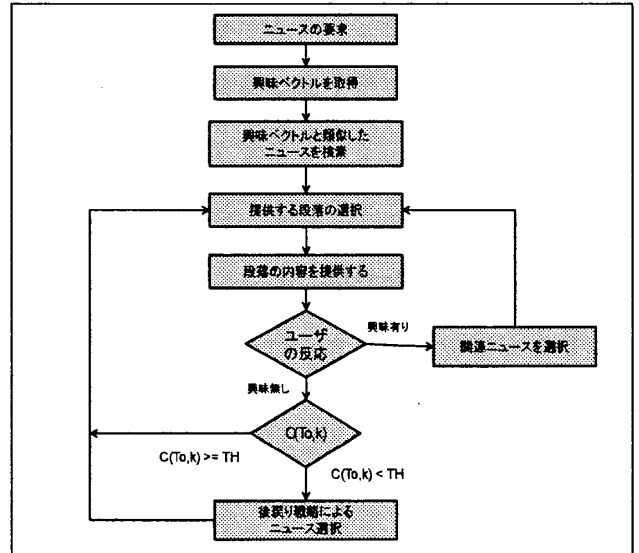


図 3: 提供戦略の流れ

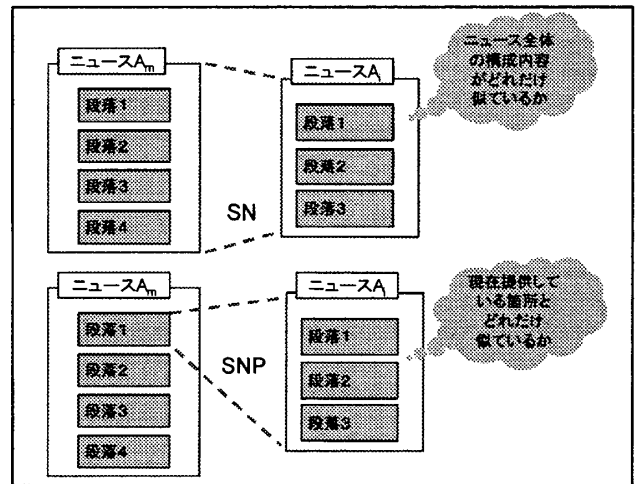


図 4: 関連ニュース探索戦略

$$SP_j = tfidf(A_m, p_{mj}) \cdot tfidf(A_k, p_{kj})$$

$$j = \arg \max_{j \in provided} SP_j$$

なお, すでに提供したことのある内容については, できる限り重複する箇所は読まないように提供をおこなう必要がある. そのため, SP の値に上限を設け, 一致度の高い段落については探索から除外する.

3.3.2 後戻り戦略

トピック展開時系列への関心度がしきい値以下になった場合, (a)~(c)のいずれかのルールによってニュース A_i を選択し, 提供を続ける. これを後戻り戦略とよぶ.

(a) トピック展開時系列内のニュースで, もっとも $C(A_i)$ が大きいニュースに戻る

(b) 1つ前のニュースに戻る

(c) 最初のニュースに戻る

そして, 後戻り戦略により選択されたニュース A_i の中で $C(A_i, p_{ij})$ がもっとも大きい段落を提供する.

4. 実験

4.1 実験内容

ユーザとの対話による複数ニュース記事間での興味追跡と読み上げの制御がおこなえているかを実験により確認する。

ニュース記事データには goo ニュースの記事を引用した。段落については手動でつけたタグをもとに段落分けをおこなった。

今回の実験では、ユーザの興味単語は固定した状態で応答の違いによる対話制御の流れを調べる。また、ユーザの応答は「はい」、「いいえ」という単語での応答に限定した。

4.2 実験結果

実験結果として本システムのニュース提供例を図 5,6 に示す。どちらの提供例もニュースA₁の段落とともに関連ニュースのA₂の段落3を提供するところまでは同じ流れで進む。しかし、図5の提供では後戻り戦略によりニュースA₁に戻り、ユーザの関心が高いと思われるニュースを追跡している。一方、図6の提供例では、A₂の内容についての提供を続行し、ニュースA₃へのトピック展開の広がりを獲得している。この違いはC(T₀,t)の割引係数 γ の値を下げた点で、ユーザの関心によるトピック展開への影響を図5の提供よりも小さくなるように設定しているためである。このようにトピック展開時系列の γ を変更することで、ユーザの関心を重視する提供戦略とユーザのニューストピックについての関心を広げる提供戦略という2つの戦略を切り替えることが考えられる。

5. まとめ

本稿では、関連ニュース探索と後戻りという2つの提供戦略を利用したニュース提供方法について述べ、対話制御の実験をおこなった。この提供戦略を組み込むことで、ユーザは自分の知りたい情報を効率的に取得することができると期待できる。

今後の課題として、システムがユーザの興味を尋ねるタイミングや尋ね方を決定する質問戦略を検討する必要がある。また、各計算式のパラメータ設定についても検討をすすめていく。

システム「ミャンマーで映像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、杏林大(東京都三鷹市)で遺体を司法解剖した。その結果、長井さんは背後から撃たれ、銃弾1発が左腰部から右上腹部に貫通していたことが判明。死因は肝臓損傷による失血死と特定した。この内容に興味ありますか。」
(ニュースA₁ 段落1)

ユーザ「はい」

システム「死因は肝臓損傷による失血死の疑いで、遺体は同日午後7時すぎ、遺族側に引き渡された。このまま読み進めますか。」
(ニュースA₂ 段落3)

ユーザ「いいえ」

システム「司法解剖の結果では、銃弾による傷は入り口の背中部分が直径約1センチだったのに対し、出口の腹部分は約1〜2センチと大きく、裂けた状態だったため、発砲距離は不明とした。」
(ニュースA₁ 段落3)

図5：ニュース提供例（関連ニュース探索＋後戻り）

システム「ミャンマーで映像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(中略)死因は肝臓損傷による失血死と特定した。この内容に興味ありますか。」
(ニュースA₁ 段落1)

ユーザ「はい」

システム「死因は肝臓損傷による失血死の疑いで、遺体は同日午後7時すぎ、遺族側に引き渡された。続きを聞きますか。」
(ニュースA₂ 段落3)

ユーザ「いいえ」

システム「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時間にわたり、遺体を司法解剖した。その結果、背後から撃たれ、銃弾1発が左腰背部から入り、右上腹部に貫通していたことが分かった。詳しく知りたいですか。」
(ニュースA₂ 段落1)

ユーザ「はい」

システム「反政府デモが続くミャンマーでAPF通信社の契約カメラマン、長井健司さん(50)が死亡した事件で、長井さんが軍部隊の兵士に至近距離から銃撃された可能性が強いことが28日、分かった。(中略)ミャンマー外務省は「流れ弾があたった」としているが、故意だった疑いも出ている。」
(ニュースA₃ 段落1)

図6：トピック展開関心度による提供の制御

参考文献

- [1] 三吉達夫, 東原智幸, 渥美雅保, “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の対話的制御”, 第70回全国大会論文集, p.2-65~p.2-66, 2008
- [2] 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, 2D-1, p.2-63~p.2-64, 2008