

## ニュースコンテンツ提示ロボットにおける ユーザ興味を考慮したコンテンツ選択手法

佐竹 聡† 川島 英之‡ 今井 倫太†§

†慶應義塾大学大学院 理工学研究科 開放環境科学専攻

‡慶應義塾大学 理工学部 情報工学科

§ 科学技術振興機構 さきがけプログラム

E-mail: † satake,kawasima,michita@ayu.ics.keio.ac.jp

本論文はニュースコンテンツ提示ロボットで必要となるコンテンツ選択手法を提案する。ニュースコンテンツ提示ロボットにおけるコンテンツ選択では、以下2つの機能が要求される。(1)時間経過でうつりかわるユーザの興味を追跡する興味追跡と、(2)ユーザの興味を引けない場合に、よりユーザの興味を引くニュースコンテンツを選択する話題転換である。本論文は、興味追跡と話題転換を実現するためにTC-TfIdfベクトルを提案する。TC-TfIdfベクトルは、最近提示したニュースコンテンツに高い重みを与える時間減衰項を導入することで興味追跡を実現する。またユーザの興味を引けなかったニュースコンテンツに負の重みを与える興味認識項を導入することで話題転換を実現する。評価実験により、TC-TfIdfベクトルは興味追跡と話題転換を実現することを示す。

キーワード: TC-TfIdfベクトル, ニュースコンテンツ提示ロボット, IRIOS

### Contents Selection Method Based on a User Interest on the Interactive News Announcer Robot

Satoru SATAKE† Hideyuki KAWASHIMA‡ Michita IMAI †§

†School of Science for OPEN and Environmental Systems,  
Faculty of Science and Technology, Keio University.

‡Department of Information and Computer Science, Faculty of Science and Technology,  
Keio University.

§PRESTO, JST

E-mail: † satake,kawasima,michita@ayu.ics.keio.ac.jp

This paper presents a contents selection method on a interactive news announcer robot. The contents selection should satisfy the interest trend and the topic change. The interest trend is to detect a capricious user interest by time. The topic change is actively changing the current topic to attract a user if the current news isn't attractive for the user. This paper introduces TC-TfIdf vector to satisfy these two requirements. To satisfy the interest trend, TC-TfIdf vector uses a time decreasing factor which prioritizes recent presented news. To satisfy the topic change, TC-TfIdf vector uses a interest recognition factor which gives negative weight to unattractive news. The result of the evaluation experiments showed TC-TfIdf vector satisfied both two requirements.

keyword : TC-TfIdf Vector , the Interactive News Announcer Robot , IRIOS

# 1 はじめに

近年コミュニケーションロボット [12, 13, 7] はジェスチャや音声対話を利用した新しいメディアとして注目されている。しかしコミュニケーションロボットが日常的なものとして人間に使用されるためには飽きられやすさの問題 [8] を解決する必要がある。飽きられやすさの一因は、ロボットプログラマにより作成された提示用の情報にバリエーションが少ないことである。一方、豊富な情報を持つ Web 上のニュースコンテンツを人間に提示するニュースコンテンツ提示ロボットが実現されれば、情報のバリエーションが増加すると思われる。本論文は、ニュースコンテンツ提示ロボットで必要となるコンテンツ選択手法を提案する。

ニュースコンテンツ提示ロボットにはユーザの興味を引くニュースコンテンツを選択し、提示する機能が要求される。よってニュースコンテンツ選択手法には、ユーザの興味を推測し表現する方法とニュースコンテンツに対するユーザの興味を見積もる方法が必要とされる。

文献 [16, 2, 3, 4, 6] に登場するエージェントシステムは、ユーザ興味を単語の集合や単語から構成された特徴ベクトルとして扱っている。これらのエージェントシステムは Web ブラウザ上で行われるユーザの操作を監視し、興味対象となる単語や単語に対する重みづけを行う。具体的にはユーザの閲覧した Web コンテンツの閲覧時間やコンテンツに含まれる単語の TfIdf 値が使用される。そしてユーザがクエリを投げたとき、ユーザの興味と一致するコンテンツの抜きだしやユーザの興味と関連するリンク先のハイライトを行う。各コンテンツに対してユーザが持つ興味の大きさは、システムで推測した特徴ベクトルと各 Web コンテンツの特徴ベクトルの内積である。

しかし従来のエージェントシステムは下記 2 つの機能を持たないため、ニュースコンテンツ提示ロボット用コンテンツ選択機構に不十分である。

1. 興味追跡: 会話をしている人間の興味は短期的でありうつりかわりやすい。例えば 1 時間前に知人の近況に興味を向けていたが、現在はノート PC に興味を向けている事態が発生する。つまりニュースコンテンツ選択機構には、現在ユーザが持つ興味を追跡し推測する**興味追跡**が必要となる。

一方、従来のエージェントシステムでは、市場調査やサーベイと言った特定の目的

をもつユーザを想定している。特定の目的に対するユーザの興味はタスク実行中に変化しない長期的な興味であり、ユーザの興味を示す特徴ベクトルは固定値へと収束する。このためエージェントシステムは同一タスクにおける興味対象のうつりかわりを扱っておらず、興味追跡を実現できない。

2. 話題転換: 著者の目指すニュースコンテンツ提示ロボットはロボット主導で提示するニュースコンテンツを決定する。このため提示したニュースコンテンツでユーザの興味を引けないことがある。この場合、ユーザの興味を引けなかったコンテンツとは類似しないが、興味を引けたコンテンツとは類似するものを新たに選択する**話題転換**が必要となる。

一方、従来のエージェントシステムでは提示するコンテンツの決定はユーザによって行われる。ユーザは目的に沿ったコンテンツの選択と収集を行うため、話題転換は必要とされない。したがってエージェントシステムは話題転換を実現できない。

本論文は、うつりかわるユーザの興味を考慮したニュースコンテンツ選択手法を提案する。我々の手法は、TfIdf ベクトルに時間減衰項と興味認識項を拡張した TC-TfIdf ベクトルでユーザの興味を表現する。

興味追跡を実現するため TC-TfIdf ベクトルは、提示したニュースコンテンツの TfIdf ベクトルに時間減衰項を掛け加算する。時間減衰項は、最後にユーザの興味を引いたニュースコンテンツほど現時点におけるユーザの興味に関連するという考えに従い重みづけを行なう。

話題転換を実現するため TC-TfIdf ベクトルは、提示したニュースコンテンツの TfIdf ベクトルに興味認識項を掛け加算する。興味認識項は、ユーザの興味を引けなかったニュースコンテンツにのみ多く含まれる単語に対して負の重みを与える。

本論文の構成をのべる。まず 2 節でコンテンツ選択手法で必要となる機能を明確化する。3 節でコンテンツ選択時に使用される TC-TfIdf ベクトルを提案する。4 節で TC-TfIdf ベクトルの実装をのべ、5 節で評価を行なう。6 節でコンテンツ選択手法に対する議論を行なったのちに、7 節で本論文の結論をのべる。

## 2 ニュースコンテンツ提示ロボット

本論文の目的は、ニュースコンテンツ提示ロボットで使用するニュースコンテンツ選択手法の開発である。本章ではコンテンツ選択手法に必要な機能をまとめる。

本章の構成は以下の通りである。まず2.1節でニュースコンテンツ選択に必要な機能をまとめる。2.2節でユーザの興味を表現する特徴ベクトルとユーザの興味を見積る方法を説明する。2.3節で著者の開発したニュースコンテンツ提示ロボットシステム IRIOS をのべる。2.4節で本論文で扱う問題をまとめる。

### 2.1 ニュースコンテンツ選択手法

ニュースコンテンツ提示ロボットでは、ロボットと対話をしているユーザの興味を引くことが望まれる。そのためニュースコンテンツ選択手法には、興味追跡と話題転換の2機能が必要となる。

#### 2.1.1 コミュニケーション相手の興味追跡

人間同士の会話は、時間経過と共に忘れられていく短期的な興味から次の話題を選択する。話題を選択する場合、現在の話題と関連性を持ったものが次に続くことは自然である。あるノートPCに対して話した後に、そのノートPCメーカーの話が続くことは自然な話題選択の例である。

ユーザの興味を引いた話題を利用することで、より聞き手の興味を引きやすい話題選択が可能である。例えば、会社Cの株主に興味を持たない人間Hがいたとする。Cの株主に対する話題をHに突然振った場合、Hの興味を引きにくい。しかし、Hが所有するC製品の話題で興味を引いた直後に、Cの株主に対する話題を振った場合、Hの興味を引きやすくなる。つまり興味を引いたニュースコンテンツと関連したものを選択することは、ユーザの興味を引くために有効である。

一方、会話中に人間の持つ興味は、短期的であり時間経過とともにうつりかわる性質を持つ。ゆえに時間が経過すればするほど、興味を向けられる対象は変わってく。例えば、1時間前では知人の近況に対して興味を向けていたが、現在ではノートPCに対して興味を向けている事態が発生しうる。

ロボットによる自然なニュースコンテンツ選択には現在ユーザの持っている短期的な興味を推測し、そしてそれに関連したニュースコンテンツの選択が必要となる。よってニュースコンテンツ選択システムには、現在ユーザが持っている短期的な興味を追跡し推測する興味追跡の機能が必要となる。

#### 2.1.2 ロボットからの話題転換

ニュースコンテンツ選択ロボットでは、提示するニュースコンテンツをロボットが決定する。しかしロボット主導のコンテンツ選択では、ユーザの興味を引けない場合がある。

興味を引けなかったニュースコンテンツに類似したものを続けて提示した場合、ユーザの興味を引くとは考えにくい。例えば、IBM PC 部門売却に興味を持たなかった人間がIBM PC 部門の展望に対して興味を持つとは考えにくい。つまりユーザの興味を引けなかった場合、提示したコンテンツと類似しないものを選択する必要がある。加えて選択されたコンテンツは、今までユーザの興味を引いたコンテンツと類似するほどユーザの興味を引きやすい。本論文ではこの機能を話題転換とよぶ。

### 2.2 コンテンツに対するユーザの興味

ニュースコンテンツ選択を行なうシステムは、各ニュースコンテンツに対してユーザが持つ興味の高さを見積もらなければならない。本論文ではニュースコンテンツに対してユーザが持つ興味の高さに、ユーザの興味を推測した特徴ベクトル  $vec_{user}$  と、ニュースコンテンツから求められた特徴ベクトル  $vec_{news}$  の内積値を使用する。

本論文では、特徴ベクトルの1手法であるTfidfベクトルを拡張しユーザの興味を表現する。全ニュースコンテンツで使用される単語を  $n_w$  種とすると、Tfidfベクトルは  $n_w$  次元のベクトルとなる。ベクトルの各要素は単語に対する重みである。

ユーザが持つ興味の高さは式1で見積もられ、 $Eval$  が大きいほど、そのコンテンツに対してユーザが大きな興味を持つことを示す。 $vec_{user}$  と  $vec_{news}$  が類似し高い関連性を示すほど  $Eval$  は大きくなる。

$$Eval = vec_{user} \cdot vec_{news} \quad (1)$$

ニュースコンテンツ  $N$  に対応したTfidfベクトル  $vec(N)$  の算出方法を説明する。単語

$word_i (1 \leq i \leq n_w)$  の重み  $w_i$  は式 2 で求められる。ただし、 $TermFreq_i(N)$  は  $N$  内で出現する  $word_i$  の個数を示し、 $DocFreq_i$  は全ニュースコンテンツ中で  $word_i$  を含むニュースコンテンツ数を示す。 $C_{AllNews}$  は全ニュースコンテンツの数である。

$$w_i = TermFreq_i(N) \times \log \left( \frac{C_{AllNews}}{DocFreq_i} \right) \quad (2)$$

式 2 では  $N$  内で頻繁に使用される  $word_i$  ほど高い  $w_i$  が与えられる。また少数の  $N$  でしか使用されない  $word_i$  ほど高い  $w_i$  が与えられる。

## 2.3 IRIOS

本論文の目的は、ニュースコンテンツ提示ロボットにおけるコンテンツ選択手法の開発である。そのために我々は、Web 上のニュースサイトからニュースコンテンツを取得し人間に提示するシステム IRIOS を開発した。

IRIOS は、サーバマシン上で動作する IRIOS サーバとコミュニケーションロボット上で動作する IRIOS クライアントから構成される。IRIOS サーバは提示するニュースコンテンツを決定し、ロボットのモーションを付加する。IRIOS クライアントは、ユーザにニュースコンテンツを提示して、コンテンツに対するユーザの興味を検出する。

IRIOS で使用する記号を導入する。 $n$  個のニュースコンテンツがすでに提示されており、 $i$  番目に提示されたニュースコンテンツを  $N_i$  と表記する。 $N_n$  は、現時点で最後に提示されたニュースコンテンツとなる。また  $N_i$  がユーザの興味を引いたか否かの検出結果を  $i(N_i)$  と表記する。 $N_i$  がユーザの興味を引いた場合を  $i(N_i) = true$  と表記し、引けなかった場合を  $i(N_i) = false$  と表記する。 $i(N_i) = true$  となる  $N_i$  を  $N_t$  と表記し、 $i(N_i) = false$  となる  $N_i$  を  $N_f$  と表記する。つまり  $N_t$  はユーザの興味を引いた  $N_i$  であり、 $N_f$  はユーザの興味を引けなかった  $N_i$  である。

IRIOS によるニュースコンテンツ提示手順は以下の通りである。まず IRIOS クライアントはセンサ入力から  $i(N_n)$  を決定する。次にクライアントは  $N_{n+1}$  決定要求を IRIOS サーバに送る。このとき  $i(N_n)$  も同時に IRIOS サーバに送信される。IRIOS サーバは  $i(N_n)$  から  $N_{n+1}$  を決定する。続いて IRIOS サーバは、決定した  $N_{n+1}$  にモーションを付加する。付加できるモーションは興味ある、驚く、残念がるの 3 つである。モーションを付加した後、IRIOS サー

バは  $N_{n+1}$  を IRIOS クライアントに送信する。最後に IRIOS クライアントは  $N_{n+1}$  を実行し、 $i(N_{n+1})$  を検出する。

IRIOS で用いられるニュースコンテンツ選択は、 $i(N_n)$  を入力とし  $N_{n+1}$  を出力する。つまり IRIOS サーバは  $NewsSel$  と定式化できる (式 3)。

$$N_{n+1} = NewsSel(i(N_n)) \quad (3)$$

また、 $N_i$  を出力した  $NewsSel$  が開始された時刻を  $t(N_i)$  と表記する。 $t(N_{n+1})$  は、式 3 の実行を開始した時刻である。

本論文では IRIOS クライアントに Robovie[9, 1] を使用した。そして  $i(N_n)$  の認識を簡単にするため、 $i(N_n) = true$  時のみ Robovie の肩にあるタッチセンサに触れてもらう。

## 2.4 本論文で扱う問題

本論文の目的は、 $NewsSel(i(N_n))$  に興味追跡と話題転換の 2 機能を組み込むことである。興味追跡は  $NewsSel(i(N_n))$  ごとに実行され、話題転換は  $i(N_n) = false$  ごとに実行される。

## 3 TC-TfIdf ベクトル

TC-TfIdf ベクトルは、 $vec_{user}$  に使用する。TC-TfIdf ベクトルは、すでに提示されたニュースコンテンツ  $N_i$  に対して時間減衰項  $TD(N_i)$  (式 4) と興味認識項  $IR(N_i)$  (式 5) を掛け加算することで算出される。ただし計  $n$  個の  $N_i$  がすでに提示されたとする。

$$TD(N_i) = TimeDec(t(N_{n+1}) - t(N_j)) \quad (4)$$

$TimeDec(t)$  は下記 3 条件を満たす単調減少関数である。

$$\begin{cases} TimeDec(0) = 1 \\ TimeDec(t) \geq 0 \\ t < t' \text{ ならば } TimeDec(t) > TimeDec(t') \end{cases}$$

$$IR(N_i) = \begin{cases} 1 & \text{if } i(N_i) = true \\ -1 & \text{else } i(N_i) = false \end{cases} \quad (5)$$

TC-TfIdf ベクトル  $vec_{TC}(n)$  は式 6 で定義される。

$$\begin{aligned} &vec_{TC}(n) \\ &= \sum_{i=1}^n IR(N_i) \cdot TD(N_i) \cdot vec(N_i) \quad (6) \end{aligned}$$

### 3.1 興味追跡の実現

興味追跡を実現するために TC-TfIdf ベクトルは  $TD(N_i)$  を導入した。  $TD(N_i)$  による重みづけは、最近の  $N_t$  に対してユーザは高い興味を持つという考えに基づく。最近提示された  $N_t$  に出現しない  $word_k$  の重み  $w_k$  は  $TD(N_i)$  により減少する。また最近の  $N_t$  によく出現する  $word_k$  が持つ重み  $w_k$  は増加する。つまり高い  $w_k$  を持つ  $word_k$  は、最近の  $N_t$  に多く含まれる。

興味追跡は、短期的でうつろいやすいユーザの興味を推測する機能である。  $TD(N_i)$  は、最近の  $N_t$  に含まれる  $word_k$  にのみ高い  $w_k$  を与えることで、興味を持つ短期性とうつろいやすさを再現する。よって TC-TfIdf ベクトルは、興味追跡を実現する。

### 3.2 話題転換の実現

話題転換を実現するため TC-TfIdf ベクトルは  $IR(N_i)$  を導入した。  $IR(N_i)$  は、  $N_f$  とは類似せず、  $N_t$  にのみ類似した  $N_{n+1}$  を選択する。  $IR(N_i)$  によって  $N_f$  にのみ多く含まれる  $word_i$  の  $w_k$  は負である。一方、  $N_t$  にのみ多く含まれる  $word_l$  の  $w_l$  は大きな値を示す。よって  $N_{n+1}$  は  $word_k$  をほとんど含まず、  $word_l$  を多く含む。

話題転換は、ユーザの興味を引けなかった場合に提示したコンテンツと類似せず、かつ興味を引けたコンテンツに類似したものを選択する機能である。一方  $IR(N_i)$  は  $word_k$  を含まないが、  $word_l$  を多く含む  $N_{n+1}$  を選択する。つまり、 TC-TfIdf ベクトルより話題転換が実現される。

## 4 TC-TfIdf ベクトルの実装

式 6 では過去の  $vec(N_i)$  を全て保持しなければならない。なぜならば  $TimeDec(t)$  には多くの関数が当てはまるため、  $vec_{TC}(n)$  から  $vec_{TC}(n+1)$  を直接求められないからである。  $vec(N_i)$  を全て保持することなく TC-TfIdf ベクトルを計算するために、式 7 の性質を  $TimeDec(t)$  が持てばよい。すなわち  $TimeDec(t) = e^{-\alpha t}$  である。

$$\begin{aligned} TimeDec(t) &= e^{-\alpha t} = e^{-\alpha(t-t')} \cdot e^{-\alpha t'} \\ &= TimeDec(t-t') \cdot TimeDec(t') \quad (7) \end{aligned}$$

式 8 により  $vec_{TC}(n)$  から  $vec_{TC}(n+1)$  を算

出できる。

$$\begin{aligned} vec_{TC}(n+1) &= IR(N_{n+1}) \cdot vec(N_{n+1}) + \\ &TimeDec(t(N_{n+1}) - t(N_n)) \times \\ &vec_{TC}(n) \quad (8) \end{aligned}$$

TC-TfIdf ベクトルでは  $TimeDec(t(N_{n+1}) - t(N_n))$  を全  $w_i$  に乗じるため、  $N_n$  に含まれない  $word_i$  の  $w_i$  は減少する。よってコンテンツ決定に影響する  $w_i$  は全  $word_i$  と比較し少数である。一方メモリ領域は有限であるため、全  $w_i$  をメモリに確保するのは無駄である。コンテンツ決定に影響ある  $w_i$  のみをメモリに格納すればよい。

以上を踏まえて著者の使用したアルゴリズムを図 1 に示す。ただし変数  $pN_n$  と  $TC-TfIdf$ 、  $T_p$  は静的変数である。図 1 中 3 行目と 4 行目において式 8 による TC-TfIdf ベクトル計算を行う。ただし  $T_p$  には以前 TC-TfIdf-cal を呼び出した時刻が入っている。5 行目において、コンテンツ決定に影響しない  $word_i$  と  $w_i$  を TC-TfIdf ベクトルから削除する。具体的には  $w_i$  の絶対値がしきい値以下の場合、  $word_i$  と  $w_i$  を  $TC-TfIdf$  から削除する。6 行目で  $T_p$  を変更し、次の呼出しにそなえる。最後の 7 行目で結果を返す。

```

1: TC-TfIdf-cal( $i_n$ ) {
2:    $T_n :=$  現在の時刻
3:    $TfIdf := IR(i_n) \times TfIdf-cal(pN_n)$ 
4:    $TC-TfIdf := TfIdf + TimeDec(T_n - T_p) \times TC-TfIdf$ 
5:   RemoveUneffectiveWord( $TC-TfIdf$ )
6:    $T_p := T_n$ 
7:   return  $TC-TfIdf$ 
8: }
```

$i_n$  = 最後に提示した  $N_n$  の興味認識結果  
 $pN_n$  = 最後に提示した  $N_n$  へのポインタ  
 $T_p$ ,  $T_n$  = 時刻を代入する変数  
 $TfIdf$ ,  $TC-TfIdf$  =  $word_i$  と  $w_i$  からなる集合。

図 1: TC-TfIdf 計算アルゴリズム

TC-TfIdf ベクトルの動きを図 2 を用いて説明する。図 2 では左側が  $vec_{TC}(n)$  を示し、右側が  $vec(N_{n+1})$  を示す。まず  $vec_{TC}(n)$  (図 2(A) 左) から  $N_{n+1}$  が選出される。つづいて  $vec_{TC}(n)$  に  $TimeDec(t(N_{n+1}) - t(N_n))$  が乗じられる (図 2(B) 左)。一方  $IR(N_{n+1})$  が  $vec(N_{n+1})$  に

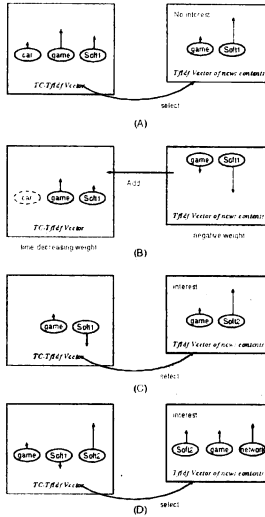


図 2: TC-Tfidf ベクトル動作例

対して乗じられる (図 2(B) 右). 図 2(B) 右では  $i(N_{n+1}) = false$  であるため,  $IR(N_{n+1}) = -1$  である. 加算により car に対する重みがしきい値を下回ったため,  $vec_{TC}(n+1)$  から削除される (図 2(C) 左). 同様に  $vec_{TC}(n+1)$  (図 2(C) 左) から  $N_{n+2}$  が選択される. そして  $N_{n+2}$  に存在する単語 Soft2 が,  $vec_{TC}(n+2)$  (図 2(D) 左) に追加される.

NewsSel は  $vec_{TC}(n)$  と内積値最大となる  $N_{n+1}$  を選択する. 単純な実装では  $N_{n+1} = N_n$  となり, 同じ  $N_n$  を続けて選択することがある. こうした  $N_n$  の連続選択を避けるため,  $N_i (1 \leq i \leq n)$  の URL アドレスをリストに保存しておき, 選択候補から除外する.

## 5 評価

TC-Tfidf ベクトルを評価するため, 次の実装環境で評価実験を行なった. IRIOS システムは C++ 言語で実装され, IRIOS クライアントは Robovie 上で, そして IRIOS サーバは PC Linux 上で動作する. IRIOS サーバが動作する PC Linux マシンは CPU に Pentium4 3.0GHz を採用し, 4GB のメモリをもつ. また OS はカーネル 2.6.8.1, ディストリビューションは RedHat Enterprise である.

以下に示す 3 戦略を比較することで, 興味追跡と話題転換が実現されるかを評価する. 1 つ目の戦略は, TC-Tfidf ベクトルを使用した

コンテンツ選択 TCR(all TC-Tfidf Record) である. 2 つ目の戦略は, 最後に提示した  $N_n$  の  $vec_{TC}(N_n)$  を使用したコンテンツ選択 LTR (Last Tfidf Record) である. LTR では  $vec_{user} = IR(N_n) \cdot vec_{TC}(N_n)$  であり,  $i(N_n) = true$  ならば  $N_n$  と最も類似した  $N_{n+1}$  を選択する. また  $i(N_n) = false$  ならば  $N_n$  と最も類似しない  $N_{n+1}$  を選択する. 3 つ目の戦略は, 全ての Tfidf ベクトルを使用したコンテンツ戦略 TR (all Tfidf Record) である. TR は適合性フィードバック [10] の一種であり, 関連研究で使用される. 具体的には  $vec_{user} = \sum_{i=1}^n IR(N_i) \cdot vec_{TC}(N_i)$  である.

ユーザと Robovie 間のインタラクションを各戦略ごとに 2 回行ない評価をとった. 各インタラクションは (1), (2) を順に実行する. (1) ユーザは最初 3 つの  $N_{1,2,3}$  に興味を示す. (2) 続く  $N_4$  に興味を示さない. 評価値として,  $N_5$  と  $N_{4,3,2,1}$  それぞれの類似度  $sim(5, n) = vec_{TI}(N_5) \cdot vec_{TI}(N_n)$  を算出した.

興味追跡の評価では,  $i(N_{i-1}) = true$  と  $i(N_i) = true$  が成立するとき,  $sim(5, i) > sim(5, i-1)$  が成立するかを調べればよい. なぜならば最近ユーザの興味を引けた  $N_i$  ほど, 現在ユーザが持つ興味に関連するからである. つまり興味追跡では,  $sim(5, 3)$  が最大値を持ち,  $sim(5, 2) > sim(5, 1)$  となるほど良い. 話題転換の評価では,  $sim(5, 4)$  が小さければ小さいほど, また  $sim(5, 3)$ ,  $sim(5, 2)$ , そして  $sim(5, 1)$  が大きければ大きいほどよい. なぜならば,  $N_j$  には類似せず,  $N_i$  にのみ類似したコンテンツ選択を実現するからである.

結果を図 3 に掲載する. 図 3 では x 軸が  $n$  を示し, y 軸が  $sim(5, n)$  を示す. また図 3 の上段, 中段, 下段がそれぞれ TCR, LTR, TR の結果である.

TCR では, どちらのインタラクションにおいても  $sim(5, 3)$  が最大値を示した. またインタラクション 1 では  $sim(5, 2) > sim(5, 1)$  が成立したが, インタラクション 2 では成立しなかった. しかしインタラクション 2 において,  $sim(5, 2)$  と  $sim(5, 1)$  の差は 10 以下と非常に小さいため, TCR は興味追跡をほとんど実現すると思われる. 一方  $sim(5, 4)$  は, いずれのインタラクションにおいても最小値を示した. ここから TCR では話題転換が実現されることがわかった. 以上により, TCR は興味追跡と話題転換を実現できる.

LT と TR は, 興味追跡と話題転換を同時に実現できない. LT では全  $sim(5, n)$  は非常に小さく,  $N_5$  は  $N_{4,3,2,1}$  のいずれとも類似しない.

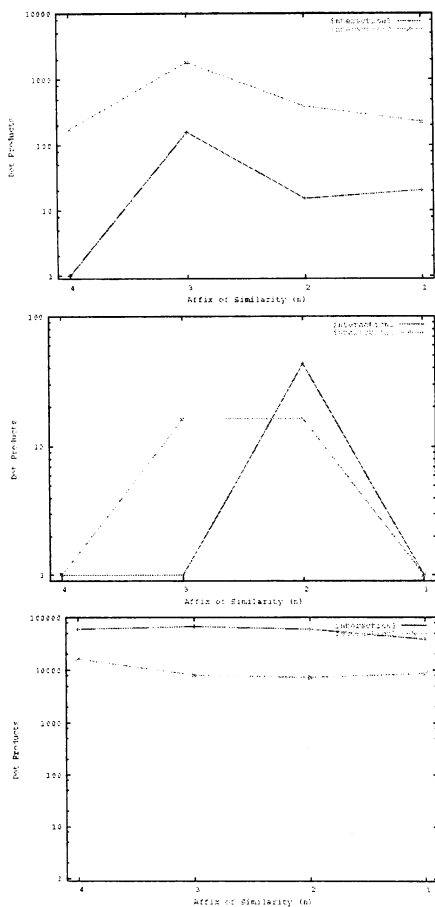


図 3:  $sim(5, n)$  と TCR(上), LTR(中), TR(下) の関係

つまり LT は興味追跡と話題転換のいずれも実現できない。TR では、 $sim(5, 4)$  と  $sim(5, 3)$  がほぼ同じである。従って、TR は話題転換を実現できない。

評価実験の結果、TCR は興味追跡と話題転換の実現に最も効果的であることが示せた。他手法では興味追跡と話題転換を同時に実現出来なかった。

## 6 議論

以下 3 点について議論を行なう。1 点目は、TC-Tfidf を使用できる他のアプリケーションに対する考察である。2 点目は、今後必要となる実験についての考察である。3 点目は、ユーザの興味を特徴ベクトルで表現した理由である。

TC-Tfidf を使用した *NewsSel* は、デスクトップ上の対話型エージェントにも適用できる。なぜならば対話型エージェントと人間の対話は、興味の短期性とうつりかわりやすさを持つからである。同様に TC-Tfidf ベクトルは、興味の短期性とうつりかわりやすさを扱うアプリケーションに必要な手法である。

TC-Tfidf ベクトルの効果を詳細に解析するためには、多くの実験が必要となる。(1) 式 7 上のパラメータ  $\alpha$  をどのようにとつたらいいのか? また状況によって  $\alpha$  を動的に変更するべきなのか? 特に Interaction 2 では、 $sim(5, 2) < sim(5, 1)$  が成立した。この原因は  $\alpha$  にある。従って  $\alpha$  変更によってコンテンツ選択がどう変化するかを調べる必要がある。(2) 心理実験の必要性。コンテンツ選択がどう行なわれているのかのみを我々は評価した。しかしニュースコンテンツ提示ロボットではコンテンツ選択による心理的影響も重要であり、調べる必要がある。

最後に、今回ユーザの興味を表現する手法として Tfidf ベクトルを使用した。適用できる手法は他にないかという疑問が挙がる。しかし実際にユーザの興味を引いている単語を検出することは困難である。このためニュースコンテンツを特徴付ける単語がユーザの興味を引くとし、ユーザの興味を推測する手法を我々はとつた。Tfidf を使用した理由は 2 つある。1 つ目は、ニュースコンテンツを特徴付ける手法として実績をもつことである。2 つ目は、計算に必要な前処理が少ないことである。

しかしコンテンツに対する特徴ベクトルは Tfidf だけではない。例として文書クラスタリングで使用される単語間共起による重み付け [15]、KeyGraph による重み付け [11, 14]、そして LSI による重み付け [5] が挙げられる。それぞれ高い重みを与える単語、次元数そして計算時間が異なる。今後、使用する特徴ベクトルについての比較検討が必要である。

## 7 結論

本論文は、ニュースコンテンツ提示ロボットで必要となるコンテンツ選択手法を提案した。コンテンツ選択手法では以下 2 つの機能が必要とされる。1 つ目は、時間経過でうつりかわるユーザの興味を追跡する興味追跡である。2 つ目は、ユーザの興味を引けない場合に、よりユーザの興味を引くニュースコンテンツを選択する話題転換である。本論文では、TC-Tfidf ベクトルを使用することで上記 2 つの機能を実現

した。興味追跡を実現するため、TC-TfIdf ベクトルに時間減衰項を導入した。話題転換を実現するため、TC-TfIdf ベクトルに興味認識項を導入した。評価実験の結果、TC-TfIdf ベクトルは興味追跡と話題転換を実現したと結論できる。

## 参考文献

- [1] Robovie - 基本仕様 - .  
<http://www.revast.co.jp/Revast%20Homepage-Dw/root/robovie/spec.html>.
- [2] Travis Bauer and David B. Leake. Real time user context modeling for information retrieval agents. In *Tenth International Conference on Information and Knowledge Management (CIKM)*, pp. 568-570. ACM Press, 2001.
- [3] J.C. Bottraud, G. Bisson, and M.F. Bruander. An adaptive information research personal assistant. In *Proc. of Workshop AI2IA (Artificial Intelligence, Information Access and Mobile Computing) IJCAI*, August 2003.
- [4] Liren Chen and Katia Sycara. WebMate: A personal agent for browsing and searching. In Katia P. Sycara and Michael Wooldridge, editors, *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)*, pp. 132-139, New York, 9-13, 1998. ACM Press.
- [5] Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, Vol. 41, No. 6, pp. 391-407, 1990.
- [6] Thorsten Joachims, Dayne Freitag, and Tom M. Mitchell. Web Watcher: A tour guide for the world wide web. In *Proc. of IJCAI97*, pp. 770-777, 1997.
- [7] 神田崇行, 石黒浩, 小野哲雄, 今井倫太. 人間と相互作用する自律型ロボット Robovie の評価. 日本ロボット学会誌, Vol. 44, No. 11, pp. 2699-2709, 2002.
- [8] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. Person identification and interaction of social robots. In *Proc. 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, pp. 1657-1664, Las Vegas Nevada USA, October 2003.
- [9] Takayuki Kanda, Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, and Ryohei Nakatsu. Development and Evaluation of an Interactive Humanoid Robot "Robovie". In *IEEE International Conference on Robotics and Automation (ICRA 2002)*, pp. 1848-1855, May 2002.
- [10] 北研二, 津田和彦, 獅々堀正幹. 情報検索アルゴリズム, 第 4.3 章, pp. 64-65. 共立出版, 2002.
- [11] Yukio Ohsawa, Nels E. Benson, and Masahiko Yachida. KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor. In *Proc. of IEEE Advanced Digital Library Conference*, pp. 12-18, Santa Barbara, California, April 1998.
- [12] Sony Global - QRIO.  
<http://www.sony.net/SonyInfo/QRIO/top.html>.
- [13] wakamaru.  
<http://www.sdia.or.jp/mhikobe-e/products/etc/robot.html>.
- [14] 山田誠二, 大澤幸生. Www における概念理解のためのナビゲーションプランニング. 人工知能学会誌, Vol. 14, No. 6, pp. 1125-1133, 1999.
- [15] 湯浅夏樹, 上田徹, 外川文雄. 大量文書データ中の単語間共起を利用した文書分類. 情報処理学会論文誌, Vol. 36, No. 8, pp. 1819-1827, 1995.
- [16] Sandra Zabala, Gábor Loerincs, Yubesli Bello, and Victor Dias. Calvin: A personalized web-search agent based on monitoring user actions. In *GI Jahrestagung (1)*, pp. 353-357, 2001.