

情報検索システムにおける高齢者への適法手法 Application Method to the Elderly People for a Information Retrieval System

井ノ上 直己¹ 西澤 健史² 米山 正秀²
Naomi Inoue Takeshi Nishizawa Masahide Yoneyama
KDDI 研究所¹ 東洋大学²
KDDI R&D Laboratories, Inc Toyo University

1. はじめに

インターネットが普及し、様々な情報が手に入るようになり、利用者が検索できる情報の数も日々増えている。

インターネットの普及率は高齢者よりも若者の方が高いと言えるだろう。もしも高齢者に若者と同じくらい普及していった時、高齢者と若者とは求めている情報に違いがあると考えられるため、高齢者が求めている情報を検索することが可能となる検索手法が必要になってくると考えられる。

本研究では、高齢者と若者の求めている情報の違いを調査し、高齢者が求める情報を提供することが可能となる手法を提案する。

2. 高齢者向け単語の抽出

2.1 高齢者向け記事の選択実験

この実験では、高齢者向け単語を抽出するために、新聞記事のどのような内容を高齢者が読んでいるのかを調べる。

2.1.1 実験内容

高齢者と若者、それぞれ 30 人（男女各 15 人ずつ）に新聞を読んでもらう。新聞 2 日分を読んでもらい、ある制限時間内に読んでもらう。今回は 35 分に設定した。

読んだ記事と読んだ記事の順番をアンケートに答えてもらう。読む時間に制限があるため、読む順番についてアンケートをとり、アンケート結果に参考とした。

アンケートの結果から、 χ^2 乗検定法^{[1][2]}を用いて違いがあるのかを調べる。

2.1.2 実験結果

アンケート結果の例として、1 ページ目の記事内容と共に以下に示す。

1 ページの記事内容	
1	田中知事失職、出直し選
2	宮路副大臣が辞任
3	審判再び（長野県の県政）
4	白馬村長、福島氏 3 選
5	けさの一句
6	斜面（長野県議会）
7	知事退職金

次に記事ごとの高齢者と若者の読んだ人数、 χ^2 乗値、どちらが好んで読むのかを図に示す。

記事	高齢者	若者	数値	基準値	判定
1	23	19	1.270	3.841	
2	19	4	15.864	3.841	高齢者
3	12	7	1.926	3.841	
4	7	0	7.925	3.841	高齢者
5	4	7	1.002	3.841	
6	20	11	5.406	3.841	高齢者
7	2	1	0.351	3.841	

表 1：1 ページの χ^2 乗値

基準値は有意水準 5 %

2.2 高齢者向け単語の抽出

高齢者がどのような新聞記事を好むのか判定ができた。次に、高齢者が好む記事から高齢者向け単語を抽出する。

2.2.1 抽出手順

1. 先の実験を元に、記事を高齢者向け、若者向け、どちらでもない記事に分類する。
2. 高齢者向け記事に出てきた単語を第 1 次高齢者向け単語とする。高齢者向け以外の記事でも同様にする。
3. 高齢者向け単語と若者向け単語の両方に出現する単語を消去する。
4. 全ての単語に対して χ^2 乗検定法を行い、 χ^2 乗値をもとに高齢者向け単語を抽出する。
5. 1 から 4 を全ての記事で行い、単語を抽出する。
6. 1 から 5 までの方法で抽出した単語では数が多い、ユーザからのキーワードの影響力が薄れてしまうので、抽出された単語に対して更に様々な判定を行い高齢者向け単語を厳選していく。

2.2.2 抽出結果

- (1) 記事からの抽出結果
第 1 次高齢者向け単語の抽出単語数：1506
第 1 次若者向け単語の抽出単語数：555
高齢者と若者の共通単語：162
高齢者と若者の抽出単語数の違い
- ・ 向け記事とされた記事数に差がある。
 - ・ 記事の大きさ（単語数）に差がある。
- (2) 単語を厳選するための判定方法

- 有意水準 5 %かつ 1 つ以上の記事に出現。
- 有意水準 1 %かつ 2 つ以上の記事に出現。
- 有意水準 5 %かつ 1 つ以上の記事に出現。
- 有意水準 1 %かつ 2 つ以上の記事に出現。

高齢者	252	159	252	159
若者	352	250	31	30

図 1：向け単語の数

単語数が一番多いという理由から、条件（有意水準 5 %かつ 1 つ以上の記事に出現）のものを利用し、実験を行った。

(3) 高齢者向け単語の例

2 乗値の高いものから 10 個の単語を例とする。

決議	県民	改革	知事	県政
選挙	不信任	対応	表明	意思

政治・経済に関する単語が多い。

(4) 若者向け単語の例

2 乗値の高いものから 10 個の単語を例とする。

リーグ	デビュー	練習	サッカー	撮影
得点	挑戦	試合	活躍	付近

スポーツに関する単語が多い。

3.3 評価実験

3.3.1 正解データ

適合フィードバック検索サーバプログラムを使用し情報検索をする。

1. ある 12 テーマに対して類似検索をする。
2. 検索結果から類似度の高い上位 30 記事を取り出し、アンケートをとる。
3. この 30 記事から高齢者と若者 30 人ずつに、30 記事から興味のある記事を選んでもらう。
4. 人数の多い順番に並び替えたものを正解データとする。

3.3.2 検索手法

以下の手順で検索を行う。

- (1) 入力キーワードに対して高齢者向け単語を追加して類似検索を行う。
- (2) 検索結果を高齢者向けに次に示す値が高い順に並び替える。
 (記事中の高齢者向け単語数) * (単語の重み)
 (記事中の高齢者向け単語数) * (単語の重み)
 - (若者向け単語数) * (単語の重み)

3.3.3 評価方法

比較方法として順位相関係数を用いる。順位相関係数は 1 から -1 の値をとり 1 に近いほど相関を強く示す。

$$\text{順位相関係数} = 1 - \frac{6 \sum (\text{順位の差})^2}{n(n^2 - 1)}$$

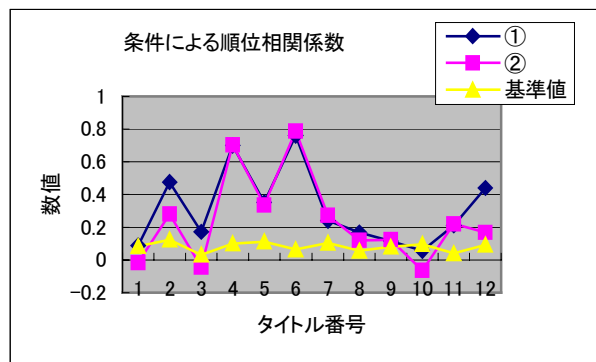
1. アンケート結果から高齢者と若者が選んだ人数

の多い順に並びかえる。

2. 向け単語を用いた検索サーバの結果からアンケート記事と同じ記事を抽出する。
3. 2 つの記事の順番に対して順位相関係数を求める。
4. 元の検索サーバの順番とアンケート結果との順位相関係数を求める。
5. 2 つの順位相関係数を比較し、単語を用いた検索サーバの結果の方が 1 に近ければ成果があったと取れる。

3.3.3 評価結果

次のグラフは条件によって並びかえられた記事と高齢者が求めている順番との順位相関係数のグラフである。基準値は高齢者向けに適應せず単純に類似検索を行った結果との順位相関係数の値である。



12 タイトル中、9 タイトルについて順位相関係数が基準値と比較して、高くなっている。条件 ① と ② では ① のほうが数値が高いことから、高齢者に関するの情報だけを付加する方が、今回の結果からは良いという結果が得られた。

4. まとめ

抽出された高齢者向け単語は政治・経済に関するものであり、若者向け単語はスポーツに関するものであった。このことは一般的に考えられていることと同じ傾向を示しているといえることから、抽出方法は妥当であったと考えられる。

本研究での提案手法により高齢者向けに適應できることを確認した

今後は単語の厳選条件を変えたり、検索テーマ数多くして実験を行いたい。

なお、本研究は通信・放送機構(TAO)からの受託研究「高齢者・障害者のための機能代行・支援システム技術の研究開発」の一環として実施した。

参考文献

- [1] 長尾 真：“言語工学”
- [2] 北川 敏男：“推測統計学”