

## 共起語を用いた学習情報検索結果の主観的評価

3Z-05

鈴木 雅実<sup>†</sup> 松本 一則<sup>‡</sup> 井ノ上 直己<sup>†‡</sup> 中山 実\* 清水 康敬\*\*通信・放送機構<sup>†</sup> KDDI研究所<sup>‡</sup> 東京工業大学\* 国立教育政策研究所\*\*

## 1. はじめに

これまで本研究では、学習分野毎内で検索語と同時に出現する確率の高い語を連想共起語として追加する方法について、その効果を検索実験にて確認した。すなわち、学習分野別に抽出した共起語を追加した類似文書検索が、学習情報を含む文書の効率的な検索に寄与し得ることを、後述する新聞記事を用いた検索実験により示した [1]。一方、学習情報の検索の観点からは、教育関係者による検索結果に対する有用度の判断が、検索精度の高低と相関することを確認しておく必要がある。そこで、共起語追加条件の異なる検索結果を示した際の、個々の文書に対する主観評価を実施して、情報の適合度と、学習への利用可能性の2つの尺度について、評価値の比較分析を行なった。

## 2. 共起語情報を追加する検索手法と評価実験

## 2.1 検索手順

本研究の学習文書検索の手順は次の通りである。

- (1) 検索要求の入力と学習分野の指定  
検索要求としての検索語1語と、検索対象とする分野(理科の7分野の一つ)を指定する。
- (2) 共起語情報を参照した検索語の追加  
指定された分野において、入力された検索語との共起度の強い順に、検索語を追加する。共起度の強さは、同一のWebページ内に該当する2語が同時に出現する分布の頻度を対数尤度で表したものである [1]。また、この共起語情報を抽出するために、7分野合わせて約4,200ページの学習向けWeb文書を人手で選択した。
- (3) 検索の実行  
追加(拡張)された検索要求に基づいて、類似文書検索処理(確率モデルに基づく文書間類似度 [2]を用いる手法)を実行する。

## 2.2 新聞記事を用いた検索実験結果

本実験では、検索実験の対象とする文書集合については、新聞記事1年分のうち約2万件を用いた。各学習分野毎の出現頻度の比較的高い語群から任意に検索語を3語づつ選び(計21語)、これらに対応する正解集合として、学習利用可能な情報を含むと判断される記事を用意した。新聞記事を選んだ理由は、比較的多数の文書集合として利用可能であり、均質性が高いことや、検索語は含んでいても検索目的と合致しないもの等を包含しており、検索実験の対象として適していることによる。さらに、新聞記事自体が教育目的で利用される場合も多いことも考慮した。

実験結果から以下の点が明らかとなった。

- (1) 学習分野限定および全学習分野のどちらの共起情報を参照した場合も、共起語を2~3語追加した際に最も検索精度が向上した。また、精度の向上は、再現率が低い状態(検索結果の上位部分)で、より顕著であった。すなわち、検索語に対して共起語を2~3語追加することにより、検索結果の上位部分で、追加しない場合と比較して精度が最大7割程度向上することが明らかとなった。このことは、共起語を2語程度追加した場合の検索結果を利用者に示すことにより、検索の効率化が期待できることを意味する。
- (2) 学習分野別の共起語情報を用いた方が、全学習分野(理科)から抽出した共起語情報を参照した場合よりも、高い検索精度が得られた。

## 3. 主観評価

## 3.1 評価者

主観評価を行なうに当たって、実際に教育関係の分野に携わっている、教員養成系大学の学部生83名、および教育指導主事を中心とする教員60名の2つの異なる集団に、評価回答を依頼した。後述するように、評価者は5つのグループに分けられ、それぞれ異なる検索結果を提示された。

## 3.2 評価対象

2.2で述べた検索実験との比較のため、新聞記事を対象とした検索結果を評価に用いた。検索実験で使用した検索語のうち、理科の7つの分野から1語づつ(地学のみ2語)選び、計8語の検索結果を提示した。各検索結果はA4用紙に印字し、1つの検索語に

ついて、1ページ当たり5件、計20件分を、記事タイトルと本文(先頭から300字までの範囲)の対で提示した。また、検索条件に対応して、次のような5種類の異なる検索結果を用意し、評価者はA～Eのグループに分かれて、同一条件の検索結果を判定した。

- A: 検索語のみで検索
- B: 検索語に分野固有の共起語を2語追加
- C: 検索語に分野固有の共起語を4語追加
- D: 検索語に分野全体の共起語を2語追加
- E: 検索語に分野全体の共起語を4語追加

### 3.3 評価尺度

各評価者は、検索結果中の各記事1に対し、次の2つの尺度についての4段階評価を行なった。

尺度1: 検索語に関する情報がどれだけ含まれているか(情報の適合度)

尺度2: 検索語に関する学習を助ける参考情報となり得るか(学習利用可能性)

### 3.4 主観評価の結果

上記の4段階評価の低い順に1, 2, 3, 4の数値を与えて集計し、グループ毎に平均を取った結果を表1に示す。

表1 各検索条件毎の平均評価値の比較

グループ (検索条件)	学部生83名		教員60名	
	尺度1	尺度2	尺度1	尺度2
A	1.79	1.56	1.91	1.84
B	2.14	1.90	2.15	2.00
C	1.94	1.83	2.07	1.92
D	1.84	1.68	1.96	1.91
E	1.56	1.47	1.63	1.58
全平均	1.85 (2.43)	1.69 (2.19)	1.94 (2.45)	1.85 (2.33)

(注)括弧内は、検索実験で事前に正解とみなした記事だけについての、評価値の平均を示す。

これらの比較分析に基づいて、これまでに次のような傾向が観察された。

- ・全体的に学生よりも教員の方が評価値が高い。二元配置法による検定結果では、両評価者集合間および両尺度間に有意差が生じた。
- ・尺度1に対する尺度2の評価値は、学生の場合、A～Eの平均で9%程度低い値であったのに対し、教員の場合、5%程度の低下に留まっている。
- ・グループ(検索設定条件)間の比較では、両尺度とも共通して、 $B > C > D > A > E$ の順で評価値が下がる。この順序は、学生・教員集合とも同様である。この結果は、2.2の検索実験結果とほぼ一致する。すなわち、事前に選定した正解記事の含有率は、B, C, D, E, Aの順に減少する。ただし、グループAとEの関係は逆転している(図1参照)。

- ・各グループ間で共通して提示された記事のみについての比較では、学生・教員両集合とも各グループ間で評価値のレベルに差は見られなかった。
- ・事前に正解と判定した記事についてのみの各グループの評価者の平均スコアは、どのグループについても4段階評価値の中央(2.5)付近に分布しており、非正解記事より、平均1段階以上の評価値を得た。
- ・各評価者の専門分野(理数系とそれ以外)と評価値の間には意味のある相関は見られなかった。

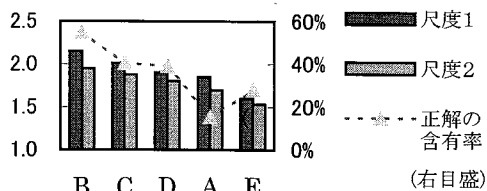


図1 各グループの平均評価値と正解含有率

### 3.5 考察

グループAとEについては、Eの方が正解認定記事が多いにもかかわらず、平均してAよりも低い評価値となっている(表示された記事間の重複は僅か)。この理由として、分野全体の共起語を4語加えた条件で検索した結果を提示したEでは、あまり適切でない共起語が同時に検索語に追加されたことにより、全体として正解記事集合との類似度が低下したものと推定できる。このような現象も考慮すると、分野毎に抽出された共起語を少数加えて、正解文書を上位に絞り込むことが、検索の効率化に役立つことが明らかである。その上で、他の共起語を検索者自身が選択追加するような手順が望ましいと考えられる。

### 4. まとめ

学習情報に関する新聞記事検索結果の主観評価を行なった結果、元の検索語に対して共起語を異なる条件で追加した場合の検索精度の変化と評価値が、ほぼ対応することが確認された。今後、2種類の評価尺度の詳細な比較や、Web上の学習情報検索について、さらに検討する予定である。

本研究は、通信・放送機構(TAO)の直轄研究「学校における複合アクセス網活用型インターネットに関する研究開発」の一環として実施しているものである。関係各位の支援と助言に感謝する。

### 参考文献

- [1] 鈴木・松本・井ノ上・橋本・中山・清水: "学習分野別の共起語情報を用いた学習情報の検索手法の検討", 情報処理学会研究報告 NL144, pp.67-74, 2001.
- [2] 岩山・徳永: "確率的クラスタリングを用いた文書連想検索", 自然言語処理, Vol.5 No.1, pp.101-118, 1999.