

酒井 哲也 小山 誠 鈴木 優 真鍋 俊彦
(株) 東芝 研究開発センター 知識メディアラボラトリー

1 はじめに

情報検索において検索要求文および各文書を単なる語の集合 (bag of words) として扱うことの限界は従来から指摘されている。現在主流であるベクトル空間モデルや確率検索モデルといった検索手法では、例えば「AのためのB」「BのためのA」「AによるB」などの意味的に異なる検索要求に対して、ターム A および B の統計量を元に算出した同一の文書スコアを返すことしかできない。上記の問題を解決するには、例えば松村ら [4] のようにターム間の係り受けを考慮した検索が必要となると考えられるが、残念ながらこのようなアプローチは既存の確率検索モデルなどに比して有意な検索有効性向上を示してはならず、現状ではその労力がペイしているとは言い難い。このことは TREC や NTCIR [2] といった「コンテスト形式」の学会において、依然として “bag-of-words” アプローチが主流であることから伺える。また、小船ら [3] のように予め構文解析を行いタームの依存関係をベクトルの要素とするアプローチもあるが、これには計算量を抑える工夫が必要であり、その検索有効性も明らかになってはいない。

我々が上記 “bag-of-words” 問題に対してとったアプローチは、意味役割解析 (Semantic Role Analysis; SRA) により既存の検索モデルのスコアを補正するというものである。三池ら [5] は、日本語文書の各文に対して「目的」「背景」「結論」などの意味役割を付与し、非ランキング型検索を行った。我々はこれまでにこれを発展させ、日本語文書を単位文と呼ばれる処理単位に分割し、各単位文の形態素解析結果に対して SRA を行い、これをもとにベクトル空間モデルによる文書スコアを補正する検索システム KIDS (Knowledge Information on Demand System) を開発した [12]。現在我々は、KIDS を日本語・英語間の言語横断検索に対応させたシステム BRIDGE (Bidirectional Retriever / Information Distiller for Japanese and English) の開発を進めている。BRIDGE は TREC などでも実績のある確率検索モデル [7] をベースにしており、英文に対しても SRA を行う機能を有している。言語横断検索を行う際には、検索要求文を機械翻訳により検索対象の言語に翻訳する [8, 9]。さらに、機械翻訳辞書にない語を扱う transliteration 機能 [11]、検索条件の自動拡張手法である PRF (Pseudo-Relevance Feedback) 機能 [9] も備えている。

本稿では、英語文書検索における BRIDGE の SRA 機能の有効性を検証する。特に、NTCIR-2 テストコレクションの多段階適合性判定情報を活用し、高適合 (highly relevant) 文書の検索に着目した評価を行う。従来 TREC などが提供してきた適合文書集合には “partially relevant” なものが多数含まれていたが [10]、近年 NTCIR が多段階適合性判定を導入したこともあり、高適合文書検索の重要性は見直され始めている。そこで我々は、高適合文書検索の観点から、BRIDGE

の SRA が英語単言語検索および日英言語横断検索に対して統計的に有意な検索有効性向上をもたらすことを示す。

2 英語の前置詞に着目した意味役割解析

BRIDGE における文書登録の流れは以下の通りである。

1. 各文書に対して part-of-speech (POS) tagging および stemming を行う。
 2. 上記結果を元に単位文分割を行い、人手で作成したルールを用いて各単位文に意味役割を付与する。
 3. 各タームを意味役割とともにインデックスに登録する。
- 一方、検索の流れは以下の通りである。

1. 検索要求文に対して、文書の場合と同様に POS tagging/stemming および SRA を行う。
2. 抽出された各検索タームおよび各文書に対して、確率検索モデル “BM25” に基づく重み [7, 9] を計算する。
3. 上記重みを SRA に基づき補正する。例えば、あるタームに関して検索要求文中の意味役割と文書中の意味役割とが一致した場合にはこのタームの重みを大きくする。
4. 上記ターム重みの総和により各文書スコアを算出し、検索結果を出力する。

BRIDGE では、用途に応じてさまざまな意味役割を定義できる。例えば、“What is ...?” といった質問に対する答えを含む可能性がある文書を検索結果の上位に押し上げるように、“... is a kind of” や “... is defined as” といった特定の表現を含む文に対して DESCRIPTION/DEFINITION という意味役割を付与することが可能である。また、三池ら [5] が日本語文書に対して行ったように論文の結論部分を強調するといった使い方も可能である。ここでは、英語文書検索のための SRA の簡単な例として、以下を考えた。

1. 文書および検索要求文を “for” や “with” などの前置詞の直前で分割し単位文とする。
2. 各単位文に対し、PURPOSE, MEANS, UNDETERMINED いずれかの意味役割を付与する。例えば、“A for B with C” という文は 3つの単位文に分解され、“A” には UNDETERMINED が、“for B” には PURPOSE が、“with C” には MEANS が付与される。

上記ルールにより、例えばターム “B” に関して検索要求文中の意味役割と文書中の意味役割が共に PURPOSE であった場合はこの重みを大きくする。

なお、上記ルールは、英語文書検索において検索要求中に目的や手段が示される一般のケースを想定したものであり、本稿の実験に特化したチューニングは行っていない。また第一筆者が上記ルール作成に要した時間は 30 分程度である。

表 1: NTCIR-2 英語文書検索結果

S 正解のみを用いた評価結果 (PRF なし)		
	英英検索	日英検索
SRA なし	0.3507	0.2683
SRA あり	0.3775 (+8%)*	0.2869 (+7%)**
検索要求内訳	13↑3↓	15↑2↓
S 正解および A 正解を用いた評価結果 (PRF なし)		
	英英検索	日英検索
SRA なし	0.2505	0.2049
SRA あり	0.2536 (+1%)	0.2034 (-1%)
検索要求内訳	26↑16↓	26↑19↓

3 実験

3.1 実験方法

検索有効性評価の題材としては NTCIR-2 テストコレクションの英英検索および日英検索タスクを用いた [2]. 検索要求数は 49, 文書数は 322,058 である. 評価指標としては TREC や NTCIR で認知されている Mean Average Precision (MAP) を用い, 統計的検定には符号検定を用いた. NTCIR-2 では S(高適合), A(適合), B(部分適合) の 3 段階の適合性判定情報が提供されているが, 本稿では高適合文書検索の観点から主に S 正解のみを用いた評価を行った.

BRIDGE における確率検索モデルのパラメタにはデフォルト値 [9] を用いた. また, PRF [9] は多くの A 正解および B 正解を上位に押し上げる性質があり, 高適合文書検索には必ずしも有効でないことが予備実験において確認されたため, 本稿では適用しなかった. 日英検索は, これまでの我々の研究 [8, 9] と同様に日本語検索要求の機械翻訳により実現した.

3.2 実験結果

表 1 に実験結果をまとめる. 「SRA なし」は BRIDGE において従来の確率検索モデルのみを用いた場合, 「SRA あり」は上記を SRA により補正した場合をあらわしている. ここでは比較のために, S 正解による評価結果に加え S 正解および A 正解による評価結果も示した. 例えば「13 ↑ 3 ↓」は 49 の検索要求のうち 13 件については SRA が有効に働き, 3 件については検索有効性が低下したことを表す. また「*」および「**」が示すように, S 正解による評価では SRA の効果は英英検索, 日英検索ともに統計的に有意 (それぞれ $\alpha = 0.05, 0.01$) であり, 平均的にはそれぞれ 8%, 7% の向上が得られている. 一方, A 正解も含めた評価では SRA の効果は有意ではないが, この場合でも有効に働くケースのほうが多いことがわかる.

以上のことから, SRA は広い意味での正解を多く上位に押し上げるタスクよりも, 特定の表現を含む高適合文書を実際に上位に押し上げるタスクに向いていることがわかる.

なお, NTCIR-2 においては S 正解の 5 倍以上の数の A 正解が存在するため, Järvelin らが最近提案した多段階適合性判定情報を直接利用する検索有効性評価方法 [1] では SRA の

効果が確認できなかった. また, 高適合文書検索とは似て異なるタスクである high-precision retrieval [6] の観点 (例えば上位 10 件の精度) から SRA の効果は確認できなかった.

4 まとめ

英語の前置詞に着目した簡単な意味役割解析により, 英英検索および日英検索において, 高適合文書を検索する場合に 7-8% の統計的に有意な検索有効性向上が得られることがわかった. 情報検索界では「自然言語処理ベースのアプローチは検索有効性の向上にさほど寄与しない」という見方が多いように思われるが, 我々は, 本研究で高適合文書検索を取り上げたように適切な問題設定および評価方法をもってすれば, 自然言語処理ベースのアプローチの利点がより明らかになるのではないかと考えている. 今後は, 英語検索に対する意味役割解析の適用範囲を拡張するとともに, 問題に応じた適切な検索戦略の選択方法について考えて行きたい.

参考文献

- [1] Järvelin, K. and Kekäläinen, J.: IR Evaluation Methods for Retrieving Highly Relevant Documents, *ACM SIGIR 2000 Proceedings*, pp. 41–48 (2000).
- [2] Kando, N.: Overview of Japanese and English Information Retrieval Tasks (JEIR) at the Second NTCIR Workshop, *NTCIR-2 Proceedings*, pp. 73–96 (2001).
- [3] 小島ほか: 意味的情報検索, 人工知能学会研究会資料 SIG-J-A101, pp. 33–38 (2001).
- [4] 松村, 高須, 安達: 情報検索における単語間の関係の効果, 情報処理学会研究報告 2001-DBS-125, pp. 257–264 (2001).
- [5] 三池, 住田: 文の意味役割解析に基づく全文検索, 情報処理学会研究報告 94-FI-34, pp. 17–24 (1994).
- [6] Mitra, M. et al.: An Analysis of Statistical and Syntactic Phrases, *RIAO '97 Proceedings*, pp. 200–214 (1997).
- [7] Robertson, S. E. and Sparck Jones, K.: Simple, Proven Approaches to Text Retrieval, Computer Laboratory, University of Cambridge (1997).
- [8] 酒井ほか: 機械翻訳を用いた英日・日英言語横断検索に関する一考察, 情報処理学会論文誌 Vol. 40, No. 11, pp. 4075–4086 (1999).
- [9] Sakai, T.: Japanese-English Cross-Language Information Retrieval Using Machine Translation and Pseudo-Relevance Feedback, *IJCPOL*, Vol. 14, No. 2, pp. 83–107 (2001).
- [10] Sakai, T. and Sparck Jones, K.: Generic Summaries for Indexing in Information Retrieval, *ACM SIGIR 2001 Proceedings*, 2001, pp. 190–198 (2001).
- [11] Sakai, T., Kumano, A. and Manabe, T.: Generating Transliteration Rules for Cross-Language Information Retrieval from Machine Translation Dictionaries, *IEEE SMC 2002 Proceedings*, to appear (2002).
- [12] 鈴木ほか: 知識情報共有システム (KIDS) によるヘルプデスク業務支援- 分野適応型検索の実業務への適用 -, 「知識発見のための自然言語処理シンポジウム」 (1999).