

1S-11

ベンチマーク BMIR-J1 を用いた 情報フィルタリングシステム NEAT の評価

酒井哲也[†] 梶浦正浩[†] 三池誠司^{††} 佐藤誠[†] 住田一男[†][†](株) 東芝 研究開発センター ^{††}Advanced-I 事業推進企画室

1. はじめに

我々は、ブール式、検索語の出現位置、検索語の文書内密度・分布などの多様な検索条件ベクトルにより文書に対して加点しランキングを行う情報フィルタリングシステム NEAT を開発した [1]。本稿では、検索システム評価用ベンチマーク BMIR-J1 [2]¹ を用いた、NEAT のブール式および検索語の出現位置情報のみを利用した場合の検索精度の評価について報告する。

2. NEAT の検索要求

NEAT における検索要求は複数の検索条件ベクトルにより構成される。検索条件の種類には、ブール式、検索語の出現位置、検索語の文書内密度・分布などがあり、各条件により文書に対して部分点を加点することができる。今回は、ブール式および検索語の出現位置情報を利用した検索条件 (bool, eachpara, head, line(1), para(1), text) の中から適宜利用した。このうち代表的な検索条件の種類について簡単に説明する。

bool 通常の AND, OR, NOT によるブール式。この検索条件が利用された場合、このブール式を満足した文書のみが他の検索条件によるランキングの対象となる。

head, line(1), para(1) 文書の見出し・第一文目・第一段落目に特定の語が出現した場合に部分点を与える。

text 文書中に特定の語が出現した場合に部分点を与える。通常のベクトル空間モデルによる点数付けに相当する。

3. BMIR-J1

検索システム評価用ベンチマーク BMIR-J1 は、日本経済新聞経済面の 600 記事と、60 の検索要求と正解記事集合により構成される。正解レベルには A, B の 2 段階が用意されており、本稿では、「～を主題とする記事」を A 正解、「～について書かれた記事」を A+B 正解と呼ぶ。検索要求は表 1 のように、基本機能 (検索語の存在判

定機能、シソーラス展開機能など)、数値・レンジ機能 (数値の大小比較機能、単位の換算機能など) を始めとする検索システムに要求される諸機能にもとづき、a～f の 6 つのグループに分類されている²。このうちグループ a および d には A 正解のない検索要求が 1 件ずつあるが、本稿ではこれらを除いた 58 件を評価に用いる。

表 1: BMIR-J1 の検索要求グループと件数

グループ	要求される検索機能	件数
グループ a	基本機能のみ	10 件
グループ b	数値・レンジ機能必要	5 件
グループ c	構文解析機能中心	6 件
グループ d	言語知識利用中心	12 件
グループ e	世界知識利用中心	10 件
グループ f	言語知識・知識処理併用	17 件
		計 60 件

4. 実験

以下の手順で実験を行った。

1. BMIR-J1 の 600 記事に対して書式解析・形態素解析・インデックス作成を行った。
2. 各検索要求について検索要求文と注釈をもとに、NEAT の検索要求記述シンタックスを熟知している筆者が検索要求を作成した。検索要求文と注釈に含まれる語以外にも有効と思われる語が想起された場合は、これも検索要求の記述に利用した。平均検索語数は 24.5 であった。
3. 各検索要求について NEAT により検索を行い、A 正解および A+B 正解のそれぞれを正解集合と見なした場合について、全体および検索要求グループ別で平均再現率・平均適合率を算出した。

図 1 に、上位 n 位 ($n=1 \sim 100$) までを検索結果とした場合の平均再現率 (r)・平均適合率 (p) を示す³。また、上位 n 位 ($n=1 \sim 20$) までを検索結果とした場合のグループ (a～f) 別平均再現率を図 2 に、平均適合率を図 3 に示す。

² 正解レベルの A, B との混同を避けるためここでは小文字で表記した。

³ NEAT は、検索要求中にブール式を記述した場合、ブール式を満足した文書のみをランキングして出力するため、検索結果が n 件に満たない検索要求は除いて平均を算出した。例えば図 1 において、 $n=1$ における値は 57 件の平均、 $n=80$ における値は 20 件の平均である。

Evaluation of the NEAT Information Filtering System Using the BMIR-J1 Benchmark

Tetsuya SAKAI, Masahiro KAJIURA, Seiji MIIKE, Makoto SATO, Kazuo SUMITA

Toshiba Corporation

1 Komukai-Toshiba-cho Saiwai-ku Kawasaki, 210, Japan

tel:044(549)2240, fax:044(520)1308

e-mail:tets1@eel.rdc.toshiba.co.jp

¹ 株式会社 日本経済新聞の協力によって、社団法人 情報処理学会・データベースシステム研究会・情報検索システム評価用データベース構築ワーキンググループが、1993 年 9 月 1 日から 12 月 31 日の日本経済新聞記事を基に構築した情報検索評価用データベース (テスト版)。

5. 考察

「A 正解の再現率 \geq A+B 正解の再現率」「A 正解の適合率 \leq A+B 正解の適合率」が各 n について成立することは、A 正解 \subseteq A+B 正解なので、A 正解における再現率の分母・適合率の分子のほうが一般に小さくなることから明らかである。特に、全ての A 正解が現われた後にその下位に B 正解が現われ始める理想的な検索結果においては、最初の B 正解が現われる直前の n 位までは「A 正解の適合率 = A+B 正解の適合率」となる。ところが図 1 では n が小さいうちから「A 正解の適合率 $<$ A+B 正解の適合率」となっている。このことから、プール式および検索語の出現位置情報のみを利用した検索では、主題検索に完全には対応できないことがわかる。とはいえ、全体的には再現率・適合率とも良好である。

図 2、3 は、グループ a の再現率・適合率が最も高くグループ f の再現率・適合率が最も低いという、直感に合致した結果を示している。基本的な検索に比べグループ b ~ f の検索が難しいことがわかる。また、グループ b の適合率が比較的高いわりに再現率が低いのは、NEAT がレンジ機能をもたないため検索要求中に記述した特定の数値がたまたま文書の字面と一致したためと思われる。

6. まとめ

検索システム評価用ベンチマーク BMIR-J1 を用いた情報フィルタリングシステム NEAT の評価結果について報告した。

- NEAT のプール式および検索語の位置情報を利用した検索により、基本的な検索に関しては高い検索精度を実現できることがわかった。
- A 正解と A+B 正解の両方を正解集合とした場合の検索性能の比較により、プール式および検索語の出現位置情報に関する検索条件のみでは主題検索に完全には対応できないことが明らかになった。
- 検索要求グループ別の検索性能の比較により、基本的な検索にくらべ、言語知識利用などを必要とする高度な検索には課題が多いことがわかった。

BMIR-J1 はテスト版であるために規模が小さく、その使い方もまちまちであるため、現時点ではシステムの客観的な比較評価は難しい。また、BMIR-J1 の検索要求には難しいものが多いが、現状ではグループ a による実験だけでも多くのシステムの比較が可能であると考えられる。ベンチマーク本格版に期待したい。

今後は、今回利用しなかった検索語の文書内密度・分布に関する検索条件を用いた検索や、クローズドテスト / オープンテストによる NEAT のチューニング能力に関する評価を行う。また検索要求の自動生成プログラムをインプリメントする予定である。

参考文献

[1] 梶浦ほか, 情報フィルタリングシステム NEAT の開発, 本大会予稿 1S-10, 1997.
 [2] 福島ほか, 日本語情報検索システム評価用テストコレクション BMIR-J1, 自然言語処理シンポジウム, 1996.

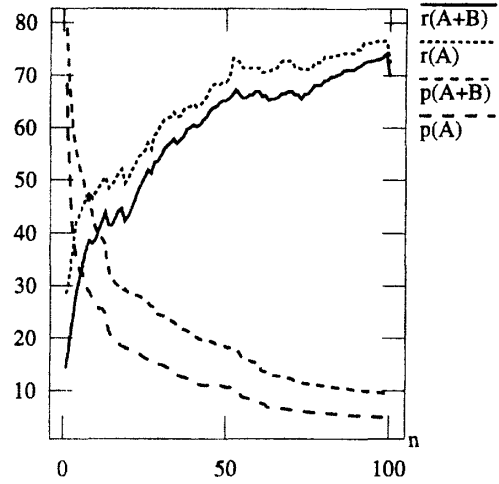


図 1: 平均再現率・適合率

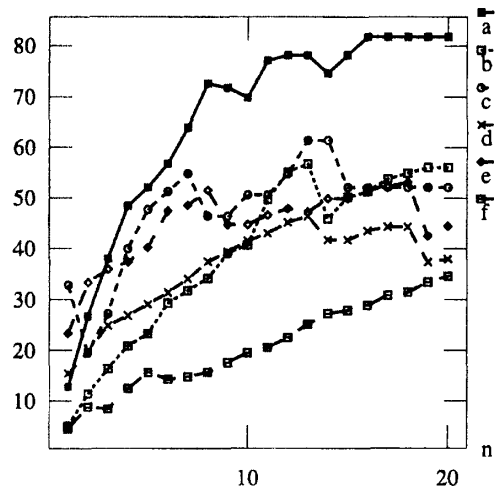


図 2: グループ別平均再現率 (A+B)

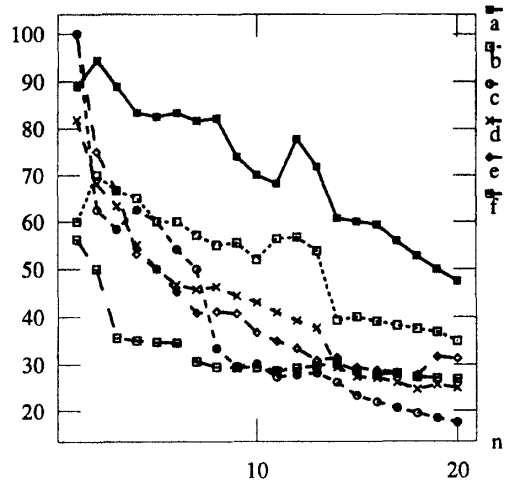


図 3: グループ別平均適合率 (A+B)