ユーザーの要求に応じた 情報フィルタリングシステム NEAT のプロファイル生成

酒井哲也 Gareth J. F. Jones 梶浦正浩 住田一男

(株) 東芝 研究開発センター

1 はじめに

近年，日々発生する膨大な量の情報の中から個人の欲しい情報のみを選出する情報フィルタリングの技術が注目を集めている．我々は，新聞社・雑誌社により提供される電子化記事から，個々のユーザーの興味に合ったものを選出し電子メイルなどで配信する情報フィルタリングシステム NEAT を開発し，1996 年 4 月に運用を開始した [1]．我々はこれまでに，自然言語文で表現されたユーザーの興味を，人手により [2]，そして自動的に [3]，プロファイルとよばれる NEAT が処理可能な表現形式に変換する実験について報告している．これらのプロファイルは，ブール式や各検索語の文書構成要素における位置・文書内密度・文書内分布などの多様な検索条件ベクトルの併用により文書に対して加点を行うもので，このような試みは英語・日本語対象を問わずこれまでに他に報告されていない．

質のよい情報フィルタリングサービスを実現するには，不要な記事を送ってしまうこと・必要な記事を見落としてしまうことを個々のユーザーに対して最小限にする必要がある．たとえ同一のトピックに興味があると言っても，個々の記事が有用であるかどうかはユーザー毎に異なる．また，同じユーザーでも時がたつとともに興味が移り変わっていくかも知れない．このような問題に対処するには，ユーザーとのインタラクションを通してプロファイルを修正していく必要がある．情報検索の分野では，このようなインタラクション・修正を relevance feedback と呼ぶ．本論文では，与えられた有用な記事の事例を利用して前述のプロファイルに対する relevance feedback を行ったうえで，未知の記事に対して情報フィルタリングを行ったときの精度を客観的・定量的に評価し，我々の手法の有効性を示す．

2 本研究の位置づけ

2.1 検索条件の付加による relevance feedback

情報検索における relevance feedback の研究は 1960 年代から行われている．これらの殆どは英語を対象としているか，あるいは英語による研究成果をそのまま他の言語に流用したもので，基本的にベクトル空間モデルを前提としている [4]．

ベクトル空間モデルは，与えられた文書集合から個々の文書の識別に有効な t 個の語を選出し，これに対応する t 次元ベクトルにより文書およびプロファイルを表現し，ベクトル演算により両者の類似度を計算する．これをそのまま情報フィルタリングサービスに適用するには以下のような問題点がある．

1. 蓄積型の情報検索とは異なり，日々内容や語彙が移り変わる情報フィルタリングにおいては，有効な t 個の語を予め定めることは困難である．
2. t 次元ベクトルというフラットな表現では，例えば新聞記事などのもっている見出しや段落といった文書構造情報を反映させるなど，多様な観点から総合的に文書の有用度を決定することが難しい．

WWW 上のパーソナル電子新聞を目指した [5][6] などでは relevance feedback が試みられているが，これらもベクトル空間法に基づくものであるため上記の問題を抱えている．NEAT は上記の問題を解決するため，任意の次元の多様な検索条件ベクトルをプロファイル内に複数記述することを可能とした．従来の relevance feedback が本質的には t 次元ベクトルの各要素の値を調整することであるのに対し，本研究は，有用な記事の事例から抽出した検索語に関する新たな検索条件ベクトルを初期プロファイルに付加するものである．本手法によれば，初期プロファイルに対してどのような検索条件および検索語が feedback されたかを容易に把握できるため，プロファイルの管理が容易になる．従来，このような feedback 過程の可視性についての議論はあまり見られなかったが，relevance feedback の実用化が進むにつれ，より重要になっていくと考えられる．

ベクトル空間モデル以外では，inference network という構造をもったプロファイルに対する relevance feedback の研究などが報告されている [7]．しかし，これは本研究のような多様な検索条件ベクトルを扱うものではなく，relevance feedback 自体は従来法に準じたものにとどまっている．

2.2 年代の異なるテストコレクションを用いた評価

Relevance feedback の評価方法としては，rank freezing 法，residual collection 法，test and control 法が代表的である [8]．ある文書集合中の事例を利用して feedback を行ったプロファイルの精度を，同じ文書集合に対して評価するという単純な手法では，その文書集合のみに対して最適化を行うことにより，どのような手法を用いても初期プロファイルに比べ大きく精度が向上するように見えてしまう．上記 3 手法はこの問題を解決しようとするものである．Rank freezing 法および residual collection 法は，単一の文書集合を用い，修正後のプロファイルを評価する際に既に検索された文書を除外して考えようとするものである．一方 test and control 法は，文書集合を学習用と評価用のふたつに分割するものである．

情報フィルタリングにおける relevance feedback では，過去に到着した記事の事例を利用して，新たに到着する記事の検索精度を向上させることが望まれる．記事中の語彙は時代とともに移りかわるため，これは蓄積型の文書検索における relevance feedback よりも難しいタスク

Profile Generation at the User's Request for the NEAT Information Filtering System

Tetsuya SAKAI, Gareth J.F. Jones, Masahiro KAJIURA, Kazuo SUMITA

Toshiba Corporation

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

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

e-mail:{tets1,jones,kajiura,sumita}@eel.rdc.toshiba.co.jp

であると考えられる．前述の 3 つの評価方法では，feedback 前後で均質な文書を用いるため，このタスクの評価方法としては必ずしも十分とは言えない．

本論文ではこのことを踏まえて，検索要求文は共通だが記事の年代が異なる 2 つのテストコレクションを利用して評価を行う．一方は検索システムの客観的評価のために一般に公開されている BMIR-J1[9] [1]，もう一方は自前で作成したもので，一方を学習用に，もう一方を評価用に用いる．これは究極的には，年代を超えた一般的知識が relevance feedback により得られるかを問うものであり，前述の評価手法に比べかなり厳しいものであるといえる．

また，現状では日本のテストコレクションの規模は欧米に比べてかなり小さく，統計的に有意な評価結果を示すことが難しい．そこで本論文では，学習用・評価用テストコレクションを入れ替えて feedback 効果の一貫性を確認し，評価結果の信頼性を高める．

3 NEAT の主な検索条件

NEAT のプロファイル中で使用される様々な種類の検索条件のうち一部について簡単に説明する．文書の得点は bool 条件以外の条件の得点を加重平均した値である [1][3]．

bool 条件 AND,OR,NOT からなる通常のブール式．bool 条件を利用した場合，これを満足する文書のみが他の検索条件による類似度計算の対象になる[2]．

text 条件 文書中に特定の語が出現した場合に得点を与える．

head/line(1)/para(1) 条件 文書の見出し・第一文・第一段落に特定の語が出現した場合に得点を与える．

ldens 条件 特定の語の文書中における密度 (何文中何文に出現しているか) に応じて得点を与える．文書内で語が占める割合が高いほど高得点がつく．

表 1: テストコレクション

	BMIR-J1	TCIR-N1
検索対象	日本経済新聞 1993 年経済面 600 記事	日本経済新聞 CD-ROM 版 1995 年 7 月 1 ~ 10 日 5048 記事
検索要求	60 件	BMIR-J1 の検索要求から 下記の方法により 正解が見つからなかった 4 件を除いた 56 件
正解集合	全数調査 平均 10.1 記事	簡単な bool 検索にマッチ した記事のみについて BMIR-J1 の基準に従い判定 平均 15.1 記事

[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)．

[2]bool 条件を利用しない場合，プロファイル中の語 (あるいはその展開語) をひとつ以上含む文書が類似度計算の対象になる．

4 relevance feedback の実験

4.1 方法

表 1に学習用および評価用に用いたテストコレクションの概要を示す．検索要求は両者に共通な 56 件を利用した．まず，TCIR-N1 の正解記事情報を用いて relevance feedback を行い，修正されたプロファイルの BMIR-J1 に対する再現率・適合率曲線および 11 点平均適合率 [10] を求めた．次に，2節で述べたように TCIR-N1 と BMIR-J1 を入れ替えて同様の実験を行った．

初期プロファイルには以下の 3 種類を用いた．

automatic 検索要求文から検索語を抽出し，bool,text,ldens 条件を用いて自動生成したプロファイル [3]．

auto-nobool 上記自動生成プロファイルから bool 条件を除いたもの．上記に比べ検索精度は向上するが，一般に類似度計算対象となる記事数が多くなるため処理速度は低下する．

manual 検索要求文およびコメントを読んで人手で作成したプロファイル [2]．

以下の手順で Relevance feedback を実現した．

1. 正解記事から以下のいずれかの尺度により n 個の名詞を抽出する [4]．
   (a) $rtf =$ 名詞の文書内頻度 (tf) の，正解記事集合に関する和
   (b) $rntf =$ 上記で tf を文書長で正規化したもの
   (c) $idf = \log($記事総数 / 名詞が出現する記事数$)$
   (d) $rtf * idf$
   (e) $rntf * idf$
2. 上記に対応する n 個 $(n = 5, 10, \ldots, 30)$ の 1 次元 text 条件を初期プロファイルに付加する[3]．text 条件の条件重みは $0.1, 0.2, \ldots, 1.0$ と変化させる．つまり，上記の各尺度のもとで 60 通りの組合せを試してみる．

4.2 結果および考察

表 2の上半分が TCIR-N1 を学習用に，BMIR-J1 を評価用に用いた結果であり，下半分がその逆である．4.1節の各尺度のもとで，検索語数 n と条件重みの 60 通りの組合せのうち 11 点平均適合率が最も高かったものについて結果を示している．“initial” は初期プロファイルの精度である．**automatic**，**auto-nobool**，**manual** において最も効果があったものと initial の再現率・適合率曲線を図 1 ~ 6に示す．

表 2から，**automatic**，**auto-nobool**，**manual** のいずれに対しても，さらにテストコレクションを入れ替えても，一貫して relevance feedback の効果が確認できる．学習用と評価用の記事の年代が 2 年ずれているにもかかわらず，11 点平均適合率は 6% ~ 16% も向上している．第 1 週の正解記事情報を利用して第 2 週のために relevance feedback を行うといった，より時間的ギャッ

[3]現状の類似度計算式のもとでは，n 次元の text 条件よりも n 個の 1 次元 text 条件のほうが有効であることがわかっている．

プの少ない実際の情報フィルタリングサービスでは，今回の実験よりも語彙の変化がはるかに少ないと考えられるので，さらに大きな効果が期待できる．また，新しい検索語の選出尺度としては，従来の情報検索研究から予想される通り [4]，$rtf*idf$ や $rntf*idf$ が有効であることがわかる[4]．さらに，**automatic** と **manual** を比較すると，initial の精度が低い前者のほうが，relevance feedback の効果が顕著である．

5 まとめ

正解記事から抽出した新たな検索語を 1 次元の text 条件のもとに記述し，これを初期プロファイルに付加する relevance feedback 手法の効果を示した．テストコレクションを入れ替え，年代の異なる未知の記事に対しても精度が向上することを確認した．また，自動生成した初期プロファイルの精度は比較的低いため，relevance feedback による向上率は最も大きかった．初期プロファイルの自動生成プラス relevance feedback という自動処理の流れは有望であると考えられる．

今後は，text 条件以外の付加，正解記事情報が少しずつ与えられた場合やユーザーの興味が変化した場合における relevance feedback などを扱っていく予定である．また，情報検索分野で主流である再現率・適合率による評価にとどまらず，情報フィルタリングサービスのユーザーの視点からの評価方法についても検討していく．

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,” 第 54 回情処学会全国大会, pp.3-299–300, 1997.

[2] 酒井ほか：“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,” 第 54 回情処学会全国大会, pp.3-301–302, 1997.

[3] 酒井ほか：“情報フィルタリングシステム NEAT のための検索要求文からのプロファイル生成,” 情処学会情報学基礎研究会 FI–47–12(NL–121–20), pp.83–88, 1997.

[4] Salton, G. et al.: “Introduction to Modern Information Retrieval,” Computer Science Series, McGraw-Hill Book Compnay, 1983.

[5] 坂上ほか：“パーソナル電子新聞 ANATAGONOMY の開発と評価,” インタラクティブシステムとソフトウェア IV, 近代科学社, pp.21–30, 1996.

[6] 菅井ほか：“WWW 上の電子新聞に対する情報フィルタリングとその評価,” 情処学会情報学基礎研究会 FI–43–13, pp.89–96, 1996.

[7] Haines, D. et al.: “Relevance Feedback and Inference Networks,” SIGIR’93, pp.2–11, 1993. In *Proceedings of the Third International Conference on Database and Expert Systems Applications*, pp.78–83, 1992.

[8] Hull, D. et al: “Using Statistical Testing in the Evaluation of Retrieval Experiments,” SIGIR’93, pp.329–338, 1993.

[9] 木本ほか: “日本語情報検索システム評価用テストコレクションの構築,” 情報学シンポジウム’98, 1998.

[10] Witten, I.H. et al.: “Managing Gigabytes: Compressing and Indexing Documents and Images,” Van Nostrand Reinhold, pp.148–151, 1994.

[4] 今回の実験では，文書長で正規化した $rntf$ が rtf より有効であるとは必ずしも言えなかった．

表 2: 11 点平均適合率および向上率

尺度	n	重み	11pt	向上率
TCIR-N1 → BMIR-J1;**automatic**				
$rtf*idf$	15	0.6	0.42	11%
$rntf*idf$	30	0.3	0.41	8%
rtf	15	0.3	0.41	8%
$rntf$	15	0.5	0.41	8%
initial	-	-	0.38	0%
idf	20	0.3	0.36	-5%
TCIR-N1 → BMIR-J1;**auto-nobool**				
$rntf*idf$	30	0.3	0.52	16%
$rtf*idf$	15	0.4	0.52	16%
rtf	10	0.3	0.51	13%
$rntf$	20	0.2	0.50	11%
idf	30	0.4	0.47	4%
initial	-	-	0.45	0%
TCIR-N1 → BMIR-J1;**manual**				
$rtf*idf$	15	0.4	0.54	6%
rtf	15	0.4	0.54	6%
$rntf*idf$	15	0.2	0.54	6%
$rntf$	15	0.4	0.54	6%
idf	15	0.6	0.52	2%
initial	-	-	0.51	0%
BMIR-J1 → TCIR-N1;**automatic**				
$rntf*idf$	25	0.4	0.34	13%
$rtf*idf$	30	0.3	0.32	7%
rtf	5	0.2	0.32	7%
idf	15	0.9	0.31	3%
$rntf$	10	0.2	0.30	0%
initial	-	-	0.30	0%
BMIR-J1 → TCIR-N1;**auto-nobool**				
$rntf*idf$	30	0.1	0.39	8%
$rtf*idf$	30	0.1	0.38	6%
rtf	5	0.1	0.38	6%
$rntf$	5	0.1	0.38	6%
idf	30	0.1	0.37	3%
initial	-	-	0.36	0%
BMIR-J1 → TCIR-N1;**manual**				
$rntf*idf$	15	0.2	0.42	8%
$rtf*idf$	10	0.1	0.42	8%
rtf	5	0.1	0.41	5%
$rntf$	5	0.1	0.40	3%
idf	15	0.4	0.40	3%
initial	-	-	0.39	0%

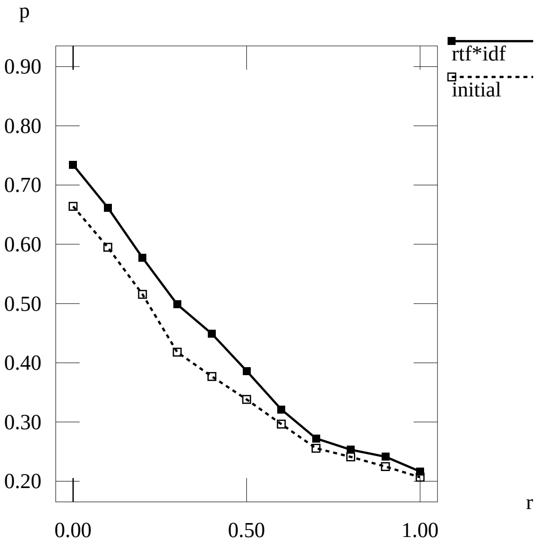

図 1: TCIR-N1 → BMIR-J1;**automatic**

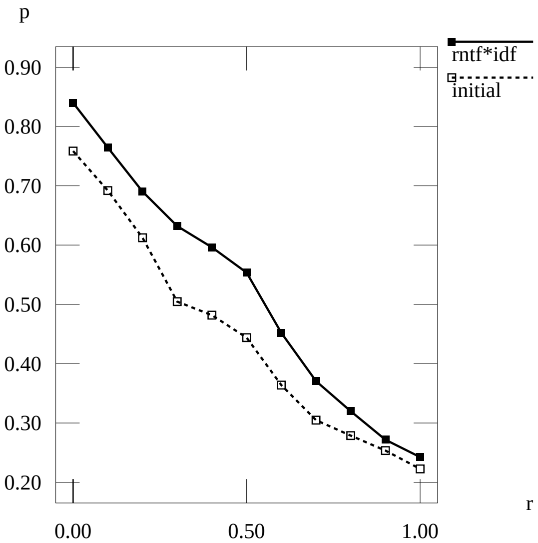

図 2: TCIR-N1 → BMIR-J1;**auto-nobool**

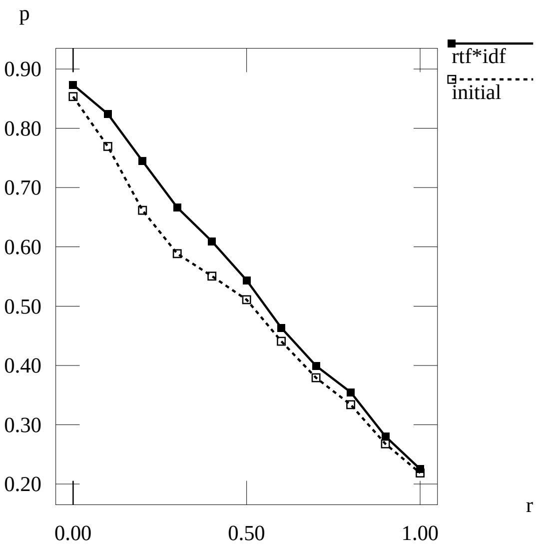

図 3: TCIR-N1 → BMIR-J1;**manual**

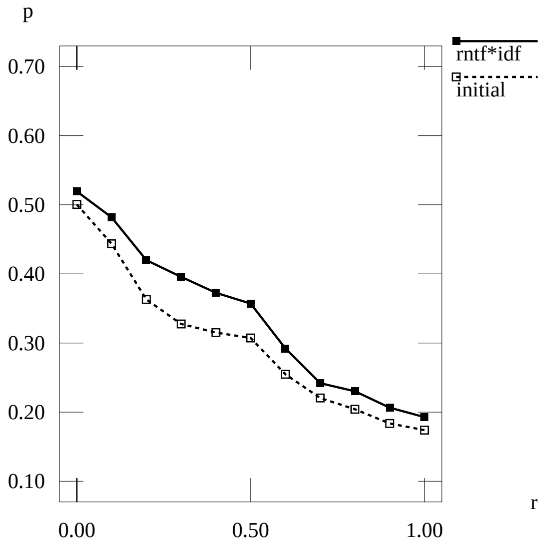

図 4: BMIR-J1 → TCIR-N1;**automatic**

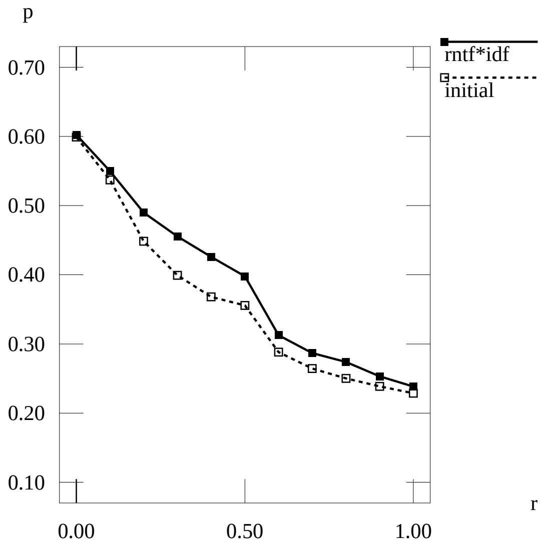

図 5: BMIR-J1 → TCIR-N1;**auto-nobool**

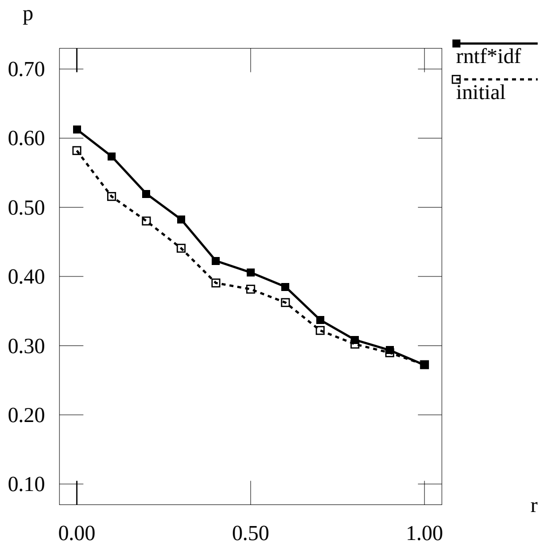

図 6: BMIR-J1 → TCIR-N1;**manual**