

論文検索における適合性フィードバックを用いたクエリ拡張支援
Query expansion using relevant feedback in academic search

出永 悠真[†] 福田 悟志[‡] 富浦 洋一[‡]
Yuma Idenaga Satoshi Fukuda Yoichi Tomiura

1. はじめに

我々が普段行う一般的な情報検索（例：カレーの作り方を調べるために、Google 検索で「カレー レシピ」というクエリで検索を行う）では、検索者の情報要求に対して適合する検索結果を一つでも得ることができれば、検索の目的が達成される場合が多い。一方、学術論文の検索では、研究者が自身の研究の新規性の確認や関連研究の調査などを目的とする場合、情報要求に適合する論文が網羅的に得られることが重要となる。しかし、論文検索では、得られた検索結果全てに対して、情報要求を満たすかを確認することは検索者にかかる負担が大きく、論文集合が大きくなると現実的ではなくなる。

そこで、論文検索において検索の網羅性を維持しながら検索の精度を向上するための方法として、クエリ拡張による検索支援の構築を本研究では目的とする。クエリ拡張では、(1)初期クエリの類義語や関連語を列挙（OR）し網羅性を高めるための拡張、(2)分野に関連の深い専門語等を追加（AND）することで精度を向上する拡張などが考えられる。これを検索者自身が行う場合、その負担は検索者が持つ情報要求への知識量に大きく依存し、分野を跨ぐ共同研究といった場合では適切な拡張クエリの発見は難しくなる。もし適切な拡張クエリの候補を自動で抽出する枠組みを構築することができれば検索者の負担を軽減することができる。本研究では AND による追加を対象としたクエリ拡張を行う。クエリ拡張では、検索者が最初に用意した初期クエリによる検索結果の一部に対して、検索者からフィードバックを受け取り、その情報を利用して拡張クエリの候補を抽出する適合性フィードバックによる手法を対象とする。

本論文では、網羅的なクエリ検索を行える検索モデルとして、LDA [1][2] によるトピック分析を利用した Wei らのランキング形式クエリ検索の研究 [3] (2 章) と、本研究と類似する、論文検索において適合性フィードバックによるクエリ拡張を行った Verberne らの研究 [4] (3 章) を関連研究として紹介する。その後、本研究で提案するクエリ拡張について説明し (4 章)、評価実験とその結果を示す (5 章)。最後に実験結果を受けての今後の課題等を述べる (6 章)。

[†]九州大学大学院システム情報科学府情報学専攻
Department of Informatics, Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan
[‡]九州大学システム情報科学研究院情報学部門 Department of Informatics, Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan

2. トピック分析を利用した検索モデル

文書検索においてブーリアン型の検索クエリを用いた完全一致による検索を行った場合、得られる検索結果の数は少なく網羅性は低い。論文検索においては検索の網羅性が高いことが重要であるので、単純なクエリの出現のみによる検索は適さないと考えられる。

Wei らは、LDA [2] によるトピック分析を取り入れることで、検索クエリの直接の出現だけでなく、同じトピックに属する共起性の高い語の出現も考慮することができる "LDA-based document model" を提案した。

LDA は文書モデルの一種であり、トピック分析の手法として利用される。LDA では、文書 D_m でトピック k の単語を確率 $\theta_k^{(m)}$ で生成し、トピック k から単語 w を確率 $\phi_w^{(k)}$ で生成する。そして、LDA によるトピック分析では、文書（単語列）の系列に対して、Gibbs Sampling を用いてパラメータ $\theta_k^{(m)}$ および $\phi_w^{(k)}$ を推定する。

Wei らの手法では、文書 D の検索クエリ Q に対する関連度は $P(Q|D_m)$ で与えられる。クエリ Q による検索結果に対して、文書 D_m を次の確率の降順にランキングする。

$$P(Q|D_m) = \prod_{q \in Q} P(q|D_m)$$

ただし、 $P(q|D_m)$ は以下のように定義される。

$$\begin{aligned} & P(w|D_m) \\ &= \lambda \left(\frac{N_d}{N_d + \mu} P_{ML}(w|D_m) \right. \\ & \quad \left. + \left(1 - \frac{N_d}{N_d + \mu} \right) P_{ML}(w|coll) \right) \\ & \quad + (1 - \lambda) P_{lda}(w|D_m) \end{aligned} \quad (1)$$

$$P_{ML}(w|D_m) = \frac{C(w; D_m)}{\sum_w C(w; D_m)}$$

$$P_{lda}(w|D_m) = \sum_{k=1}^K \theta_k^{(m)} \phi_w^{(k)}$$

$C(w; D)$ は文書 D における単語 w の出現頻度、 $P_{ML}(w|coll)$ は文書集合全体での単語 w の出現確率で、 $P_{lda}(w|D_m)$ は LDA による文書 D_m における単語 w の出現確率である。 μ と λ は 3 つの推定法の重み (混合比) に関係したメタパラメタである。 N_d は文書 D の単語数で、語数の多い文書ではその文書中での登場頻度を重視している。

3. 適合性フィードバックによるクエリ拡張

クエリ拡張を行うための方法として、初期クエリでの検索結果の一部に対して、どの文書が情報要求に適合するかというフィードバックを検索者から受け取り、その情報を利用してクエリ候補の抽出を行う適合性フィードバックによる手法がある。この方法では、初期クエリでの検索結果上位の文書を適合文書と仮定してクエリ拡張を行う疑似適合性フィードバック [5] などの方法に比べて、情報要求に関するクエリ以外の情報を検索者から直接得られるので、より適切なクエリ拡張が期待できる。

Verberne らは学術的な検索におけるクエリ拡張を適合性フィードバックによる手法で行った [4]。Verberne らの手法では、検索とクエリ追加を交互に行う。クエリ追加ステップでは、上位 100 件の検索結果のうち検索者が適合文書と判定した文書に含まれる語を追加クエリの候補としてスコア付けし、上位 10 種類を検索者に提示する。クエリ候補のスコア付けには下式で定義される、検索者が適合文書と判定した文書と一般的な英語コーパス、COCA (Corpus of Contemporary American English) [8] それぞれにおける語の出現の相違を用いる。

$$\begin{aligned} score(w) &= PR(w) \log \frac{PR(w)}{PC(w)} \\ &= \frac{C(w; R)}{N} \log \frac{C(w; R)/N}{C(w; corpus)/M} \end{aligned} \quad (2)$$

$C(w; D)$ は適合と判定された文書集合 R における語 w の出現頻度で N が総単語数、 $C(w; corpus)$ は一般的なコーパスにおける出現頻度で M が総単語数である。式から、適合文書中では出現確率が高く、一般的な文書では出現確率が低い語、つまり適合文書に特有の語が高いスコアを得ることがわかる。

Verberne らの実験では、クエリ候補中で被験者 2 名 (検索分野の知識が小さい) のうち、一人以上が関連すると判定したクエリを追加して再度検索を行い、検索とクエリ追加を繰り返している。(a)データセットの付加情報から作成された理想的なクエリ、(b)付加情報からクエリ追加を行う手法と比較を行った結果、Verberne らの手法が、手法(b)よりも改善されたことが報告されている。

4. 提案手法

本章では、本研究で提案する適合性フィードバックを用いたクエリ拡張手法を説明する。

4.1 基本方針

前章で紹介した Verberne らによる適合性フィードバックでは、検索者が情報要求を満たすと判定した適合文書の情報のみを利用したクエリ拡張を行っていた。本研究では、適合文書の情報に加えて、情報要求を満たさない非適合文書の情報も活用することでクエリ拡張の向上を図る。

検索者からフィードバックを受け取る初期クエリによる検索結果上位の文書集合は、文書中に出現する語の分布がある程度類似していると考えられる。このとき、適合文書中では高い確率で出現する一方で、非適合文書中にはあまり出現しない語が存在するならば、それが情報要求を表す重要な語 (拡張クエリ) である可能性が高いと考えられる。

4.2 クエリ候補のスコア付け

本研究で提案する適合性フィードバックによるクエリ拡張のフローを図 2 に示す。まず、検索者が最初に用意した初期クエリを使って、Wei らの "LDA-based document model" によるランキングを行い、ランク上位 100 件の論文を検索者に提示する。検索者は提示された文書集合に対して、適合文書か非適合文書かの判定を行い、それをフィードバックとしてクエリ拡張システムに与える。本手法においても、Verberne らの手法と同様に拡張クエリの候補となる語に対してスコア付けを行う。クエリ拡張システムは検索者からのフィードバックの情報を利用して、適合文書の集合 (R) 中での出現確率 (PR) が大きく、非適合文書の集合 (N) での出現確率 (PN) が小さい拡張クエリ候補の単語 w に対して高いスコアが与えられるようなスコア付けに従ってクエリ候補のランク付けを行い、その結果を検索者に提示する。得られた拡張クエリ候補のスコアとランキングを利用して検索者はクエリ拡張を行い、再度検索を行う。拡張クエリの候補となるのは、適合文書中に一度でも登場する全ての単語としている。今回の実験においては下式で表されるスコア付けを採用した。

$$score(w) = PR(w) \log \frac{PR(w)}{PN(w)} \quad (3)$$

ただし、 PR と PN はそれぞれ R と N を文書集合 D とした "LDA-based document model" によって与えられる出現確率である。これは、Verberne らのスコア付けにおける一般文書コーパスでの単語 w の出現確率 PC を、非適合文

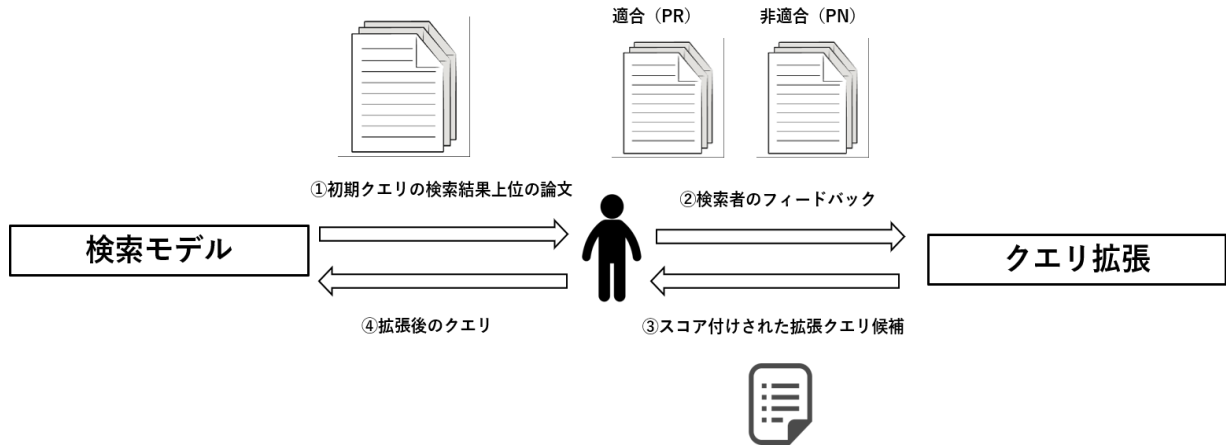


図 2 : クエリ拡張の手順

書中での出現確率 PN に置き換えた形で、適合文書と非適合文書の単語の出現確率分布の KL ダイバージェンスから w についての項を抽出した形となっている。最も単純なスコア付けとして、 PR/PN が考えられるが、一般的には適合文書に対して非適合文書の方が多いため、 PR に対して PN が極端に小さい不適切な語が上位のクエリ候補になってしまう場合が多く見られた。また、拡張クエリとしての有効性を計る特徴量として PR を重視することは自然であると考えられる。

5. 実験

提案手法によるクエリ拡張の有効性を確認するため、論文検索タスク用のデータセットによる実験を行い、クエリ拡張前後の検索性能を比較した。また、Verberne らの一般文書コーパスを使った手法との比較も行った。

5.1 実験データ

実験では、NTCIR (NII Testbeds and Community for Information access Research) により提供されている論文検索タスクのデータセット[6][7]から、27種類の検索課題を使用した。このデータセットでは、各検索課題に対して情報要求と、対応するラベル付きの論文アブストラクト集合が用意されている。ラベルについては、情報要求に「関連する」「ある程度関連する」「関連しない」の3段階で付与されており、本実験では「関連する」「ある程度関連する」のラベルが付与された論文を適合文書とした。検索課題ごとのラベル付き文書集合数の平均は2375件、適合文書数の平均は60件である。

5.2 実験手順

まず、各検索課題に対して初期クエリを人手で設定した。この初期クエリに従って、“LDA-based document model”によるランキング検索を行う。本実験では、初期クエリの検索結果上位100件に対するフィードバックを利用するとして、ラベルに従い適合文書と非適合文書集合を作成する。これらの文書集合から PR と PN を計算し、式2のスコアを拡張クエリ候補に対して付与する。スコア値によるランキング結果を用いて、次節で説明する方法に従ってクエリの追加を行い、クエリ拡張前後における検索性能の改善の比較を行った。

5.3 追加クエリの決定

現実の論文検索タスクでは、システムが拡張クエリの候補を提示し、検索者自身が最終的に追加するクエリを決定するという使用が想定される。ただし、実験においては、クエリ拡張後の検索性能は検索者のクエリ選択に大きく依存してしまうため、ロバストな実験設定として以下に挙げる2種類の方法で追加クエリの決定を行った。

スコアに対する閾値による方法

クエリ拡張によって常にクエリが改善されるとは限らない。初期クエリが情報要求を比較的良好に表現している場合や、クエリ拡張システムが適切な拡張クエリの候補を抽出していないような場合では、クエリ追加を行わない方が高い検索性能が得られることがある。そこで、クエリ追加を行うべきかの判定基準として、拡張クエリの候補に付与されたスコアに対して閾値を設定し、一定以上のスコア値を持つ語を追加する。本実験では、各検索課題に対する初期クエリで検索を行ったとき、後述する評価指標が最も高くなる検索課題には、クエリ追加が行われないような最小の値を閾値として設定した。

追加数を固定する方法

別の方法として、追加するクエリの個数を固定し、スコアによるランキング上位から順に追加する方法でも実験を行った。この方法では、適切な拡張クエリ候補がランキング上位にどの程度集まっているかを計ることを目的としている。本実験では、追加するクエリ数を1~5の5種類に設定した。

5.4 比較手法

実験の比較手法として、Verberneらの式2によるスコア付けと、PRを直接スコアとした場合のクエリ拡張との比較を行った。ただし、PRの推定は式1に従った。これによって、適合文書の情報に非適合文書の情報を加えることによる影響と、一般文書コーパスの情報を加える影響を調べた。

5.5 評価指標

本実験におけるクエリの良さを計る評価指標として、次に説明する値を採用した。図3のグラフは、ある検索課題について、論文集合に対してクエリによるランク付けを行ったとき、横軸の割合 k まで検索結果上位を参照した際の適合文書の再現率を縦軸が表している。このグラフと横軸で囲まれる面積が大きいほど良いクエリであるとして、評価値として設定しクエリ拡張による改善率を求めた。面積については、ランキング結果の上位1%から100%までの各点(1%刻み)における再現率から算出し、各検索課題のマイクロ平均を実験データ全体での評価値とした。グラフはそれぞれ、青線が初期クエリ、オレンジの線が提案手法によるクエリ拡張後のクエリ、緑の線がVerberneらの手法によるクエリ拡張後のクエリによる検索を表している。

5.6 パラメータ設定

本実験におけるパラメータとして、LDAのトピック数 K 、Dirichlet分布のハイパーパラメータ α と β 、"LDA-based document model"のパラメータ λ と μ を27分割交差検証により決定した。各検索課題に対する学習データを、

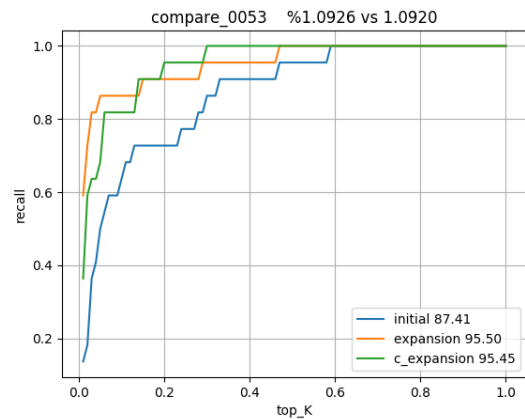


図3: k-再現率曲線によるクエリ評価

それ以外の全ての検索課題とし、初期クエリによる検索の評価指標の最大化を行っている。各パラメータの範囲は、 $\alpha = (0.05, 0.1, 0.5, 1.0)$ 、 $\beta = (0.01, 0.1, 1.0)$ 、 $K = (5, 7, 10, 20)$ 、 $\lambda = (0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$ 、 $\mu = (1, 10, 25, 50, 100, 300)$ と設定した。

5.7 実験結果と考察

表1に、拡張クエリ候補のスコア付け方法ごとの、初期クエリによる検索を基準とした評価指標の改善率(拡張後の評価値を初期クエリの評価値で割った値)の比較結果を示す。スコアの欄がPRの行は、スコアとしてPR(w)を用いた手法の改善率である。ただし、列は左から順に、クエリ追加数を1から5に固定した場合、追加数1から5の改善率の平均、スコアに対して閾値を設定した場合である。

閾値によるクエリ追加の結果を比較すると提案法が最も高い改善率が得られている。また、追加クエリ数を固定した場合は、追加数ごとに最良の手法が異なるが、平均値と最大値で比較すると共に閾値の場合と同じく、提案法、Verberneらの方法、PRの順に改善率が高い。

図4に閾値によるクエリ拡張と追加数3に固定した場合の再現率の変化を示す。閾値による方法の場合は、提案手法が他の比較手法に対して1.4%ほど改善率で上回っている。しかし、閾値による改善率は、追加数を固定する方法と比較して、改善率が全体的に小さい。

表1: 各手法ごとの評価指標の改善率の比較

スコア	1	2	3	4	5	平均	閾値
提案手法	1.0329	1.043	1.0531	1.0451	1.0417	1.0432	1.0192
Verberneらの手法	1.0327	1.0475	1.0529	1.0453	1.0362	1.0429	1.0058
PR	1.0291	1.0408	1.0382	1.0462	1.0384	1.0385	1.0048

これは、閾値で除外されてしまった適切な拡張クエリ候補の影響が、不適切な候補の影響よりも大きかったためと思われる。閾値によるクエリ追加の場合でも、クエリ追加によって評価値が下がってしまう検索課題は存在しており、閾値の設定方法については考察の余地があると考えられる。

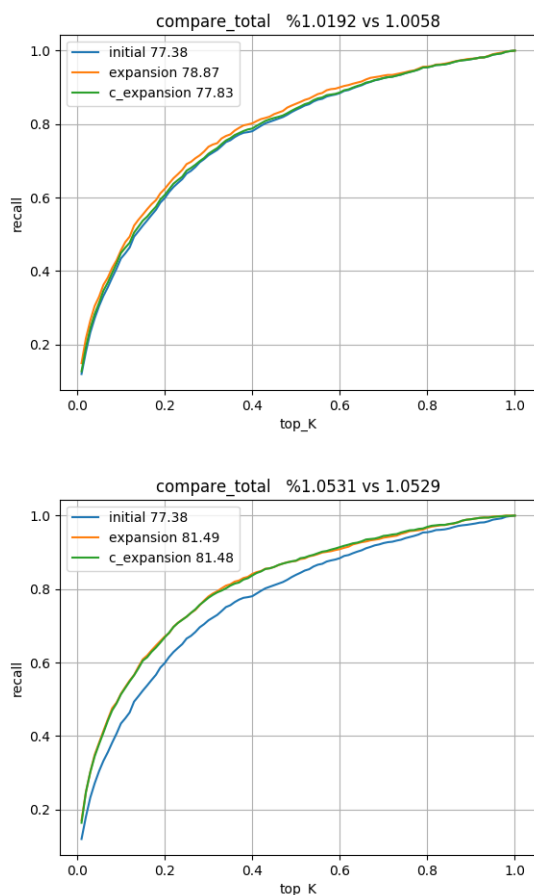


図4：閾値（上）とクエリ追加数3（下）の場合の拡張

6. おわりに

本論文では、論文検索支援において、検索者に非適合文書と判定された文書の情報を活用した適合性フィードバックによるクエリ拡張で検索性能が改善されることを確認した。今後の課題としては、より適切なスコア付け方法の考案や閾値の設定、非適合文書の情報と一般文書コーパスの情報を組み合わせたクエリ拡張の提案などに取り組むことを計画している。また、より現実での使用場面に近い実験設定として、追加クエリの決定を検索者自身が行う条件での比較実験も行いたいと考えている。

参考文献等

- [1] Blei. D. M, Ng. A. Y, Jordan. M. I, “Latent Dirichlet allocation” Journal of Machine Learning Research, 3, pp 993-1022 (2003).
- [2] Griffiths, T. and Steyvers, M. “Finding scientific topics” Proceedings of the National Academy of Science, 101:5228–5235 (2004).
- [3] Xing Wei, W. Bruce Croft, “LDA-based document models for ad-hoc retrieval” SIGIR '06, pp 178-185 (2006).
- [4] Verberne S, Sappelli M, Kraaij W, “Query Term Suggestion in Academic Search” ECIR 2014, Advances in Information Retrieval, pp 560-566 (2014).
- [5] Jinxi Xu, W. Bruce Croft, “Query Expansion Using Local and Global Document Analysis” SIGIR '96, pp 4-11 (1996).
- [6] Kando. N, Kuriyama. K, Nozue. T, Eguchi. K, Kato. H, Hidaka. S, and Adachi. J, “The NTCIR workshop: The first evaluation workshop on Japanese text retrieval and cross-lingual information retrieval” Information Retrieval with Asian Languages Workshop, pp 1–7 (1999).
- [7] Kando Noriko, “Overview of the second NTCIR workshop” NTCIR Workshop, pp 35–43 (2001).
- [8] <https://www.english-corpora.org/coca/>