あらすじのあいまいな記憶に基づく書籍検索手法

京塚　萌々†　　田島　敬史††

† 京都大学工学部情報学科　〒 606–8501 京都府京都市左京区吉田本町
†† 京都大学大学院情報学研究科　〒 606–8501 京都府京都市左京区吉田本町
E-mail: †kyozuka@dl.soc.i.kyoto-u.ac.jp, ††tajima@i.kyoto-u.ac.jp

あらまし　本研究では，書籍のタイトルが思い出せず，あらすじについてのあいまいな記憶しかない状況における書籍検索を支援する手法を提案する．ユーザの記憶に残っているあらすじの記述をクエリとして用いると，クエリ内に記憶違いによる誤った単語が含まれることがある．本研究はどのような語に記憶違いが生じやすいかの分析を行い，この分析結果に基づき記憶違いの生じやすさに基づく優先順位をつけ，ユーザの意図した書籍をランキング上位に提示する手法を提案する．

キーワード　質問緩和，質問拡張，質問修正，記憶違い

1. はじめに

図書館にはレファレンスというサービスがあり，利用者からの質問や相談を受けて資料を探す手助けをしている．その中には「昔読んだ本のあらすじはあいまいに覚えているが，タイトルを思い出せないので探してほしい.」といった質問も多く存在する．また，Yahoo!知恵袋 (注1) などの質問サイトにも同様の質問が多数寄せられている．

このような質問の答えを国立国会図書館サーチ (注2) などの検索システムを使用して見つける場合，質問者の記憶に残っているあらすじの記述をクエリとして用いると，記憶違いによる誤った単語が含まれる可能性がある．誤った単語が含まれている場合，質問者の意図と異なる書籍が提示される，または，クエリに当てはまる書籍が見つからないということが予想される．

そこで，本論文では，あらすじに関するあいまいな記述による検索を支援する手法を提案する．この手法では，質問者によるあらすじの記述内の語に記憶違いの生じやすさに基づく優先順位をつける．そして，優先順位をもとに単語を削除することでクエリを生成して検索する．クエリ中の語には質問者の提示した語を使用しているので質問者の意図をある程度反映したものになり，かつ記憶違いの可能性のある語を削除していくことで質問者の意図する書籍を提示できる可能性を高めることができると考えられる．

手法の概要は以下の通りである．あらすじの記述内の各単語を，その語が文章中で果たす役割に基づいて「主語」「述語」「目的語」「その他」の 4 種類に分類する．次に，分類した単語群から考えられるすべての組み合わせをクエリとする．最後に，生成したクエリをランキングして検索結果をクエリランキング順に連結する．このクエリランキング手法として本研究では 7 種類の手法を提案し比較する．

提案手法の性能を評価するため，質問サイトからあらすじを記述して書籍を探している質問のうち正解が見つかっているものを収集し，質問文から提案手法によってクエリを生成した．このクエリで国会図書館サーチによる検索を行いクエリと検索結果をランダムにランキングする手法と比較したところ，質問者の意図した書籍を上位にヒットさせることができたが，単純にクエリの単語数が多い順にランキングしたものを超えることはできなかった．

(注1)：https://chiebukuro.yahoo.co.jp
(注2)：http://iss.ndl.go.jp

2. 関連研究

本章では本研究と関連する研究について述べる．

クエリ推薦については多くの研究がなされている．クエリ拡張 (query expansion) は，ユーザの入力したクエリに単語を追加して検索を支援する技術であり，これまでにも多くの手法が提案されている [1]．特に検索エンジンが収集したログを基にしてクエリを推薦する手法は多数ある．Wang ら [2] は，ユーザが入力したクエリのログを解析し，効果的にクエリに単語を追加したり，クエリ中の単語を置換したりしてクエリを改良する手法を提案した．また，近藤ら [3] は，ユーザのウェブ閲覧ログからテキスト情報と閲覧時間の情報を得て，ユーザが興味を持つようなクエリを推薦している．これらはユーザの過去の検索履歴を利用してユーザの検索意図を推測するものであるが，ユーザの記憶違いが過去の検索履歴に反映されるとは考えにくく，本研究とは解決しようとしている問題が異なる．大石ら [4] は，ユーザの入力したクエリで取得した文書が検索意図に適合しているかユーザ自身に評価してもらった結果をもとに関連語を抽出してクエリに追加することでクエリ拡張する手法を提案している．この研究は，適合フィードバックによりユーザの意図を加味したクエリを生成するものであり，入力したクエリの検索結果からユーザが質問内容について新たに想起することには向いているが，ユーザが入力したクエリを変更するものではないので，記憶違いの修正には向かない．

ユーザ自身が検索意図を指定するというアプローチは多数存在している．金子ら [5] は，ユーザがクエリ中のキーワードを入力する速さにより，ユーザがどの程度各キーワードを他の

キーワードに置き換えて検索してもいいと思っているかシステムに伝える手法を提案している．また，吉田ら [6] は，ユーザが入力したクエリに対する検索結果ページ中に現れる重要語を抽出，グラフを用いて可視化し，ユーザがそのグラフを操作することでクエリを修正し再ランキングをするという手法を提案している．これらの研究は，ユーザの意図に応じてクエリ中のキーワードを置換・除去したり，キーワードの重み付けを変更したりするもので，入力されたクエリを修正してユーザの意図に近づけるものなので本研究と共通する部分があると言える．しかし，これらはユーザが自身の検索意図を自覚してクエリを入力するため，ユーザが自覚していない記憶違いを修正することには向かない．本研究では記憶違いに着目してユーザの意図する結果を得ることを目的としているため異なる．

本研究ではユーザがあらすじについて記述した文章をもとに検索を行う．しかし，このような単語数の多いクエリによる検索は単語数の少ない短いクエリに比べて適切な検索結果を実現できないという研究結果がある [7]．そこで，余分な単語を取り除きより短いクエリを生成することで適切な検索結果を実現するための研究が近年盛んに行われている．Chen ら [8] の研究では，ユーザが過去にクリックしたクエリのログをもとに，ユーザの検索意図に近く，かつ単語数の少ないクエリを生成している．この研究は，単語数の多いクエリは自然言語の文に近い形で記述されるため，従来のキーワード検索ではうまく検索できない点に着目しており，ユーザの過去の検索意図を参考に単語数の多いクエリからユーザの意図を反映していると思われる単語を抽出してより短いクエリを生成しようとしている．過去の検索履歴にユーザの記憶違いが反映されるとは考えにくいので，本研究とは想定する問題が異なる研究であると言える．Kumaran ら [9] の研究や Balasubramanian ら [10] の研究では，検索する文書中に単語が含まれている確率をもとにクエリの質を示す尺度を提案し，RankSVM [11] による学習によって適切にクエリをランキングすることで余分な単語が取り除かれたクエリを生成している．これらの研究の目的はより本研究に近いと言えるが本研究のように図書検索に特化し図書検索におけるユーザの記憶違いの特徴を利用するものではない．

あいまいな記憶のもとで情報検索を行う研究として Voorhees [12] の研究，Cucerzan ら [13] の研究，Ochiai ら [14] の研究や隅田ら [15] の研究がある．Voorhees の研究と Cucerzan らの研究はユーザが入力したクエリの表記揺れに対応しており，Voorhees の研究は同義語辞書，Cucerzan らの研究はクエリログを利用してクエリを修正する．これらは，ユーザがあらすじに含まれる単語の表記を記憶違いしている場合に対応できるが，単語それ自体を記憶違いしている場合は対応できないと考えられる．ユーザのエピソード記憶 (過去の経験に関係する記憶) に基づく情報検索の傾向をユーザ実験によって調べた Ochiai らの研究によれば，一定期間後に記憶を頼りにユーザが入力するクエリは動詞を多く含み検索性能を低下させる．どのような単語を含むと検索性能が低下するか分析しているところが本研究と類似している．ただし，この研究ではクエリ内の記憶違いを直接修正することは行っていないところが本研究と異なる．隅田らの研究の目的は，ユーザの記憶があいまいなために抽象的なクエリが入力された場合でも検索意図に応じた結果を提示する映画検索エンジンの構築であり，本研究の目的と似通う部分がある．この研究では二つの手法が提案されており，ウェブ文書の情報を用いクエリ中の抽象的な語を具体的名称に置き換え拡張するものと，2 文の類似性を調べられる文間アライメント認識 [16] を用いてクエリ文とウェブ上の映画のあらすじとの類似性によって検索するものである．この研究はあいまいな記憶下では表現が抽象的になることに着目しており，固有名詞などの具体的な語句を記憶違いしている場合にも対応できると考えられる．このように，ウェブ上の情報を用いてクエリを拡張するアプローチもあるが，本研究ではウェブ情報を用いないクエリの修正を試みる．

3. 提 案 手 法

本研究で提案する手法について述べる．

この研究の目的は，過去に読んだ本のあらすじの記述から生成したクエリをランキングし，検索結果の上位に正解の書籍を表示することである．

3.1 単 語 分 類

まず，あらすじの記述内の単語を文中の役割に応じて分類する．あらすじの記述から主要な内容であると思われる一文を取り出し，次の 4 種類の部分に分ける．

- 主語 (「〜が」「〜は」となる部分)
- 述語 (「〜する」「〜である」となる部分)
- 目的語 (「〜を」「〜で」「〜によって」となる部分)
- その他 (主語・述語・目的語に対する修飾語など)

分けた部分から「が」「を」などの助詞にあたる部分を取り除き,「主語」「述語」「目的語」「その他」の役割ごとに分類する．

3.2 クエリ生成

続いて，分類したあらすじ内の語からクエリを生成する方法を説明する．

まず，どの役割の語が実際のあらすじに含まれにくいか調べるため，予備実験を行った．

3.2.1 予 備 実 験

Yahoo!知恵袋で「うろ覚え 児童書」「うろ覚え 絵本」を検索クエリとして収集した解決済みの質問のうち，記憶に残っているあらすじから書籍を探す趣旨の質問であり，かつ回答された書籍に対して探していたものである旨の質問者のコメントがついているものを 50 件収集した．また，質問内で正解として挙げられている書籍のあらすじデータを Amazon(注3) と国立国会図書館サーチでそれぞれ収集した．

Yahoo!知恵袋で収集した質問文のあらすじに関する記述から主要と思われる部分を取り出し，3.1 で述べたように分類した．次に，分類した単語が実際のあらすじデータにも含まれるか検索を行った．質問文内のあらすじ記述に含まれる分類された全単語と，実際のあらすじデータにも含まれる単語について，各分類ごとに語数を調べた．ただし，国立国会図書館サーチにつ

(注3)：https://www.amazon.co.jp

いてはあらすじデータが収録されていない書籍が 9 冊あったので，それらについては国立国会図書館サーチのデータを利用するときは分類対象外にした．

Amazon のデータを用いた予備実験の結果を表 1，国立国会図書館サーチのデータを用いた結果を表 2 に示す．各分類について，質問に含まれるもののうちどれだけの割合であらすじデータにも含まれるかの割合をとった．質問に含まれる語のうちあらすじに含まれるものの割合が大きい順に役割の種類を並べると，Amazon では「主語」「目的語」「その他」「述語」，国立国会図書館サーチでは「目的語」「主語」「その他」「述語」の順になった．

表 1　Amazon のデータを用いた予備実験の結果

語の役割	質問に含まれる数	あらすじに含まれる数	割合
主語	48	30	0.625
述語	73	12	0.164
目的語	65	37	0.569
その他	41	18	0.439

表 2　国立国会図書館サーチのデータを用いた予備実験の結果

語の役割	質問に含まれる数	あらすじに含まれる数	割合
主語	43	19	0.442
述語	62	3	0.048
目的語	55	30	0.545
その他	34	15	0.441

3.2.2　提案するクエリランキング手法

実験を踏まえ，以下の 7 つのクエリランキング手法を提案する．1 つ目の手法では,「述語」「その他」「目的語」「主語」の順にあらすじの記述に含まれていない確率が高いという仮説のもと語を取り除いていく．2 つ目の手法では，同じ優先順位を基にして単語のリストから得られたクエリをランキングする．残りの手法では，単語のリストから得られたクエリを，予備実験で得られた数値を元にランキングする．

各クエリで検索した結果をこのランキング順に連結することで元クエリの検索結果とする．

単純除去　リスト内の語を「述語」「その他」「目的語」「主語」の順 (仮説による優先順位の低い順) に並べ，先頭から 1 語ずつ取り除いてできた順にクエリをランキングする手法．

優先順位によるランキング　リストを「述語」「その他」「目的語」「主語」の順 (仮説による優先順位の低い順) に並べる．リストの長さが n のとき，0〜$(n-1)$ 個取り除いた組み合わせを全て列挙し，取り除いた語数の少ない順にランキングする．語数が同じクエリは，優先順位の低い語が取り除かれているものほど上位に来るようにランキングする．この手法を優先順位によるランキングと呼ぶ．

重みによるランキング　リストの長さが n のとき，0〜$(n-1)$ 個取り除いた組み合わせを全て列挙する．予備実験によって算出した各タグの出現割合の平均 (主語:0.5335，述語:0.106，目的語:0.557，その他:0.44) を重みとして各クエリの重みを合計し，大きい順にクエリをランキングする．

単語数と重みによるランキング　リストの長さが n のとき，0〜$(n-1)$ 個取り除いた組み合わせを全て列挙し，各クエリを長さの順にランキングする．このとき，同じ長さのものについては先に述べた「重みによるランキング」と同様に各タグの出現割合に基づいて計算した重みの合計が大きい順にランキングする．

確率によるランキング　リストの長さが n のとき，0〜$(n-1)$ 個取り除いた組み合わせを全て列挙する．予備実験によって算出した質問に含まれる各タグの語数とあらすじに含まれる各タグの個数をもとにして，質問内に含まれるタグのうち各クエリに含まれる語のタグの組み合わせがあらすじデータ内にも出現する確率を計算し，大きい順にクエリをランキングする．

単語数と確率によるランキング　リストの長さが n のとき，0〜$(n-1)$ 個取り除いた組み合わせを全て列挙し，各クエリを長さの順に並べる．このとき，同じ長さのものについては先に述べた「確率によるランキング」と同様に各クエリに含まれる語のタグの組み合わせが出現する確率スコアを計算し，大きい順にクエリをランキングする．

期待値によるランキング　リストの長さが n のとき，0〜$(n-1)$ 個取り除いた組み合わせを全て列挙する．予備実験によって算出した各タグのあらすじへの出現確率から，クエリ中の語のタグの組み合わせがあらすじデータに出現する確率を求めることができる．$P(q)$ はクエリ q に含まれるタグの組み合わせがあらすじデータに出現する確率，$hit(q)$ はクエリ q で検索した際のヒット数とすると，クエリ q の検索結果に正解書籍が含まれる時の順位の期待値は $\frac{1}{2}hit(q)$ とでき，含まれない時の順位は $hit(q)+1$ と近似できるので，クエリ q で検索した際の正解書籍の順位の期待値 $expect(q)$ は次の式で求められる．

$$expect(q) = \frac{1}{2}P(q)hit(q) + (1-P(q))(hit(q)+1)$$

この期待値が小さい順にクエリをランキングする手法を期待値によるランキングと呼ぶ．

4. 実　　験

本研究に際して行った実験について述べる．

4.1　実験に用いたデータ

実験では，Yahoo!知恵袋より収集した書籍を探す質問と回答のデータと，国立国会図書館サーチより収集した書籍のあらすじデータを使用した．それぞれのデータの詳細を説明する．

4.1.1　質問と回答のデータ

Yahoo!知恵袋で「うろ覚え 児童書」「うろ覚え 絵本」を検索クエリとして収集した解決済みの質問のうち，記憶に残っているあらすじを記述して書籍を探すもので，かつ回答された書籍に対して質問者から探していたものであった旨のコメントがついているものを 50 件収集した．

4.1.2　書籍のあらすじデータ

4.1.1 で収集したデータで回答として挙げられている書籍名をクエリとして国立国会図書館サーチで検索を行い書籍データを収集した．ここから手作業であらすじが表示されないもの (あらすじが収録されていないもの) を削除した．また，同じ書籍

だが異なるあらすじを持つ別のデータとして収録されているものについてはそれぞれ別の本として集計した．37 冊分のデータが集まり，対応する質問は 27 件となった．

4.2 実験の手順

今回行った実験は単語の分類・クエリ生成と生成したクエリを利用した検索の 2 段階に分けられる．それぞれ説明する．

4.2.1 単語の分類・クエリ生成

4.1.1 で収集した Yahoo!知恵袋のデータの質問に記述されたあらすじから主要な内容と思われる部分を取り出し「主語」「述語」「目的語」「その他」の 4 種類の役割に分類した．役割と単語の組に対し，3.2.2 で提案した 7 種類の手法でクエリをランキングした．単語分類の例を表 3 に示す．また,「えんそくこわいぞあぶないぞ」が正解となる質問に対して 3.2.2 で提案した 7 種類の手法でランキングしたクエリの例をそれぞれ表 4〜 10 に示す (表 5〜 10 については上位 6 件のみ).

表 3　単語分類の例

書籍タイトル	あらすじ	(タグ，単語) の組
木かげの家の小人たち	少女が小人と仲良し	(主語，少女) (目的語，小人) (述語，仲良し)
えんそくこわいぞあぶないぞ	少女がおとぎ話の世界に迷い込む	(主語，少女) (その他，おとぎ話) (目的語，世界) (述語，迷い込む)

表 10　期待値によるランキング

述語	その他	目的語	主語
迷い込む	おとぎ話		
迷い込む	おとぎ話	世界	
迷い込む	おとぎ話		少女
迷い込む	おとぎ話	世界	少女
迷い込む		世界	少女
迷い込む			少女

また，比較手法として次の 4 つの方法でクエリを生成し，提案手法と比較した．

ランダム除去　タグと語の組のリストからランダムに 1 語ずつ取り除きクエリを生成する手法.

ランダムランキング　リストの長さが n のとき，1〜$(n-1)$ 個取り除いた組み合わせを全て列挙してランダムにランキングする手法.

単語数によるランキング　リストの長さが n のとき，1〜$(n-1)$ 個取り除いた組み合わせを全て列挙して除去した単語数が少ない順にランキングする手法．ただし，除去した語数が同じもの同士はランダムにランキングする．

単純 tf-idf　単語列に含まれる単語 a について次の式によって $tfidf$ を計算する．

$$tfidf_a = tf_a \times \log \frac{N}{df_a}$$

ただし，tf_a は単語列に単語 a が含まれる際に 1，含まれない際に 0 となる数値，df_a は単語 a で国立国会図書館サーチによる検索を行った際のヒット数，N は今回扱う質問データから取り出した単語全てについての df の合計である．タグづけされた元の単語列と生成した各クエリについてこの $tfidf$ を要素とするベクトルを作り，元の単語列によるベクトルと生成したクエリによるベクトルのコサイン類似度が 1 に近い順にクエリをランキングする．

4.2.2 生成クエリによる検索

4.2.1 で生成したクエリを用いて国立国会図書館サーチで検索を行い，ヒットする検索結果の数と正解書籍が最初に出てくる順位を調べた．本実験では質問データを収集する際に児童書と絵本を対象としているので,「児童書総合目録」というデータベースを対象に検索を行った．

本実験で利用する国立国会図書館サーチの複数キーワード検索は，入力されたキーワード全てを書誌情報に含むものを国立国会図書館サーチのシステム内で独自に定められた適合度順(書誌情報における検索キーワードの個数、出現頻度、データベースごとの優先度をもとに算出されている (注4)) に表示する．500 件以上ある場合は上位 500 件が検索結果リストとして表示される．

正解書籍が i 番目のクエリの検索結果全体には含まれるが上位 500 件のリストに含まれなかったときは，正解書籍の順位の期待値は 500 位からリスト末尾までの中間の順位であると言えるので，そのクエリにおける正解書籍の順位を次のように定義する．

$$rank_i = \frac{1}{2}(500 + hit_i)$$

ただし，hit_i は i 番目のクエリで検索した際のヒット数である．

質問全体についてクエリをランキングした順番にクエリの検索結果を並べたときの正解書籍の順位とするので，k 番目のクエリで正解書籍がヒットした際の正解書籍の順位 $rank$ を次のように定義する．

$$rank = \sum_{i=1}^{k-1} hit_i + rank_k$$

ただし，hit_i は i 番目のクエリのヒット数，$rank_k$ は k 番目のクエリの検索結果における正解書籍の順位である．また，正解書籍が生成した全てのクエリでヒットしなかった場合の正解書籍の順位は ∞ とする．

同じ書籍に対し異なるあらすじデータを持つ複数の書誌情報が存在する場合は，それぞれの検索結果について正解書籍の順位を計算し，最も小さいものをその質問における正解書籍の順位とする．

4.3 評 価 手 法

ユーザの意図した書籍を検索結果上位に表示することが目的であるので，各生成手法について以下の式で定義される平均逆順位 (Mean Reciprocal Rank, MRR) を計算し，比較することで評価した．

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

(注4)：http://iss.ndl.go.jp/information/faq/#B2

表 4　単純除去によるクエリランキング

述語	その他	目的語	主語
迷い込む	おとぎ話	世界	少女
	おとぎ話	世界	少女
		世界	少女
			少女

表 5　優先順位によるランキング

述語	その他	目的語	主語
迷い込む	おとぎ話	世界	少女
	おとぎ話	世界	少女
迷い込む	おとぎ話		少女
迷い込む	おとぎ話	世界	
		世界	少女
	おとぎ話		少女

表 6　重みによるランキング

述語	その他	目的語	主語
迷い込む	おとぎ話	世界	少女
	おとぎ話	世界	少女
迷い込む		世界	少女
迷い込む	おとぎ話	世界	
		世界	少女
迷い込む	おとぎ話		少女

表 7　単語数と重みによるランキング

述語	その他	目的語	主語
迷い込む	おとぎ話	世界	少女
	おとぎ話	世界	少女
迷い込む		世界	少女
迷い込む	おとぎ話	世界	
迷い込む	おとぎ話		少女
		世界	少女

表 8　確率によるランキング

述語	その他	目的語	主語
		世界	
		世界	少女
	おとぎ話	世界	
			少女
	おとぎ話		
	おとぎ話	世界	少女

表 9　単語数と確率によるランキング

述語	その他	目的語	主語
迷い込む	おとぎ話	世界	少女
	おとぎ話	世界	少女
迷い込む		世界	少女
迷い込む	おとぎ話	世界	
迷い込む	おとぎ話		少女
		世界	少女

$|Q|$ は対象とする質問の数 (本実験では 27)，$rank_i$ は i 番目の正解書籍について 4.2.2 で定義した順位を計算したものである．ただし生成したクエリ i で正解書籍がヒットしなかった場合，$\frac{1}{rank_i} = 0$ とする．

また，単純除去は元の単語列から生成可能なクエリの一部で検索を行う手法であるため，用意した質問 27 件のうちで正解書籍を発見できた件数を調べた．残りの提案手法では，発見できた正解書籍の順位を横軸，出現頻度を割合として縦軸としたヒストグラムを作り，正解書籍の順位のばらつきを調べた．

比較手法として挙げた 4 手法も，同様に平均逆順位を計算することで評価を行った．ランダム除去，ランダムランキング，単語数によるランキングについては，ランダム除去とランダムランキングでは 5 回ずつ，単語数によるランキングについては 10 回クエリを生成して MRR を計算した平均をとった．また，ランダムランキングと単語数によるランキングについては，全ての質問について正解書籍が想定しうる最低順位になったデータを作成し，最悪の場合として MRR を計算した．

また，ランダム除去は単純除去と同様に元の単語列から生成可能なクエリの一部のみについて検索を行う手法であるため，用意した質問 27 件のうちで正解書籍を発見できた件数を調べた．残りの 3 手法では，単純除去以外の提案手法と同様のヒストグラムを作り，正解書籍の順位のばらつきを調べた．ランダムランキングについては 5 回の試行と最悪の結果について，単語数によるランキングについては 10 回の試行と最悪の結果についてそれぞれヒストグラムを作成した．

4.4　結　　果

各手法について計算した MRR を表 11 に示す．また，単純除去以外の提案手法について，正解書籍の順位を横軸にとったヒストグラムをそれぞれ図 1〜 6 に示す．ランダムランキングを 5 回ずつ行った試行の累積のヒストグラムを図 7，単語数によるランキングを 10 回行った試行の累積のヒストグラムを図 9，それぞれの最悪の場合についてのヒストグラムをそれぞれ図 8，図 10 に示す．単純 tf-idf のヒストグラムを図 11 に示す．

まず，単純除去とランダム除去を比較する．これらは元の単語列から生成可能なクエリの一部のみで実際に検索を行う手法である．MRR を比較すると，単純除去がランダム除去の約 20 倍であった．また，用意した質問 27 件のうち正解書籍を発見できた件数を調べたところ，単純除去は 27 件中 10 件，ランダム除去は 5 回の試行のうち 2 回で 27 件中 1 件，3 回で 27 件中 0 件という結果になり，提案した単純除去手法の方が用意した質問に対する正解を多く見つけることができた．

表 11　MRR による比較

手法	MRR
単純除去	0.098947087
先頭優先除去	0.152101894
重みによるソート	0.148613002
単語数→重みソート	0.148610991
確率によるソート	0.044323034
単語数→確率ソート	0.149009630
期待値によるソート	0.156152319
ランダム除去 (平均)	0.004938272
ランダムソート (平均)	0.091820651
ランダムソート (最悪)	0.000421639
単語数→ランダムソート (平均)	0.164337041
単語数→ランダムソート (最悪)	0.104480127
単純 tf-idf	0.120593996

次に，残りの 6 種類の提案手法の結果を述べる．確率によるランキング以外の 5 手法の MRR については，ランダムランキング (平均)，単純 tf-idf より高く，単語数によるランキング (平均) より低い結果となった．5 手法を MRR が大きい順に並べると期待値によるランキング，優先順位によるランキング，単語数と確率によるランキング，重みによるランキング，単語数と重みによるランキングの順であった．確率によるランキングの MRR はランダムランキング (最悪) は上回っていたが，ランダムランキング (平均)，単語数によるランキング (平均)，単語数によるランキング (最悪) を大きく下回った．

正解書籍の順位の広がりについて，ランダムランキング，単語数によるランキングは最大 120000 位程度まで分布していた

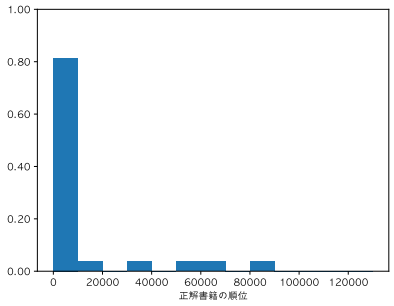

図 1　優先順位によるランキング

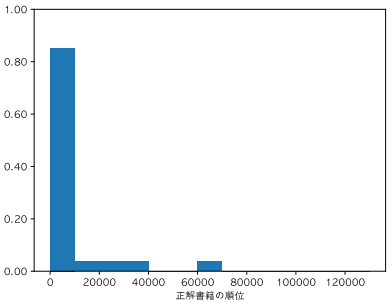

図 2　重みによるランキング

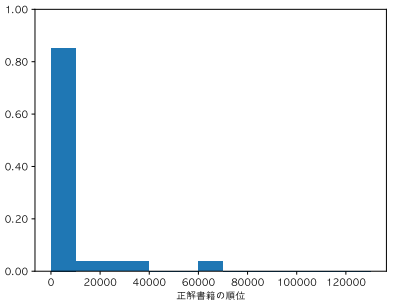

図 3　単語数と重みによるランキング

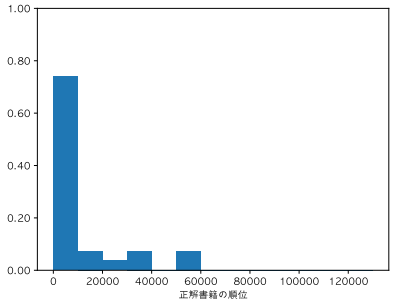

図 4　確率によるランキング

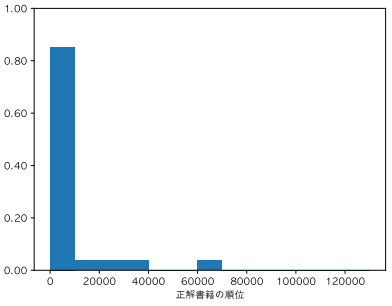

図 5　単語数と確率によるランキング

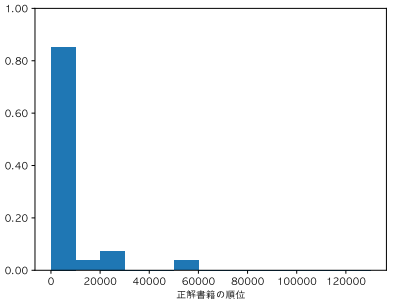

図 6　期待値によるランキング

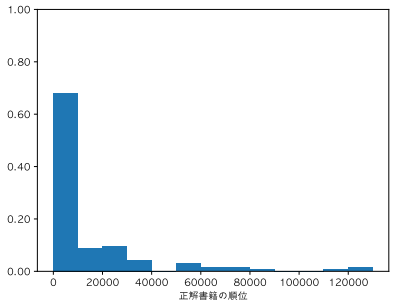

図 7　ランダムランキング (試行 5 回分の累積)

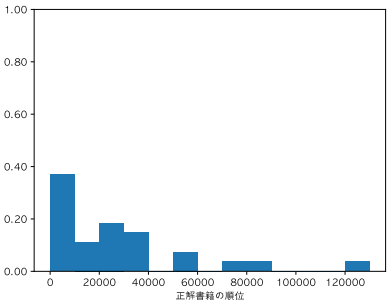

図 8　ランダムランキング (最悪の場合)

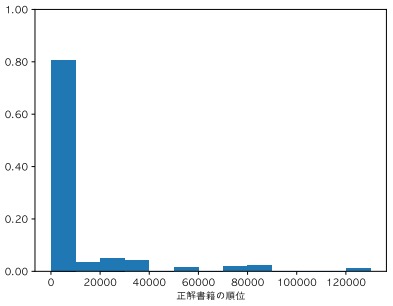

図 9　単語数によるランキング (試行 10 回分の累積)

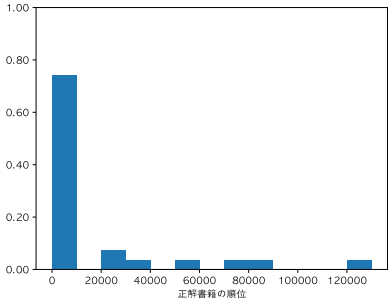

図 10　単語数によるランキング (最悪の場合)

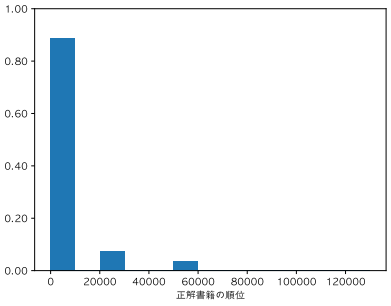

図 11　単純 tf-idf

が，先頭優先除去は 80000 位程度，残りの提案手法は 60000 位程度までの範囲に収まっていた．単純 tf-idf は 60000 位程度までに収まっており，正解書籍の順位の散らばりは比較的小さかった．

5. 考　　察

実験結果を踏まえ，各提案手法の今後の課題を考察する．

5.1　単 純 除 去

単純除去は，元の単語列から生成可能なクエリの一部のみを実際に生成する手法であり，比較対象はランダム除去となる．

単純除去の MRR はランダム除去による値の約 20 倍であった．ランダム除去で提示できた正解書籍は 0 件か 1 件であったのに対し，単純除去で質問データ 27 件中 10 件について正解書籍を提示できており，検索結果に正解が含まれるクエリを生成

できた割合が比較手法 1 より高かったため MRR が高かったと考えられる．

しかし，単純除去は 3 分の 2 程度の質問に対応できなかったとも言える．この原因として，同じ役割の語を複数含むあらすじについて，同じ役割の語は本来優先順位が同じであるのにもかかわらず，ランキングの際に無作為な順序づけを行っていたことが考えられる．ただ 1 種類の単語列を考えて先頭から除去していくのではなく，同じ役割の語をランダムに並び替えた単語列も考えるなど，より適切な手法を検討することで提案する優先順位に準拠したクエリ生成が可能になると考えられる．また，提案する優先順位が比較的高い役割の語を除去していれば書籍を見つけられるようなデータもあった．優先順位の高い語を除去していても除去語数が少なければユーザの意図に近いものとして推薦されるようにするなどの順序づけを行う方法も考えられる．以上の課題に対応するのが残りの提案手法である．

5.2 優先順位によるランキング

優先順位によるランキングの MRR は提案手法 7 種類のうち 2 番目に高かったが，ヒストグラムで比較すると，正解書籍の順位が最大 80000 位を超えており広がりが大きい．本実験で利用したデータ内には今回提案した「述語」「その他」「目的語」「主語」の役割に分類される順に検索する優先順位が低いという仮説が合うデータが多かったために MRR が高かったが，うまく当てはまらず正解書籍の順位が低くなるデータもあったためにヒストグラムで見ると正解書籍の順位の広がりが大きくなったと考えられる．単純 tf-idf は 60000 位までに正解書籍が分布しており，正解書籍が先頭優先除去よりも上位に収まっていることから考えると，利用するデータによっては tf-idf の方が良い結果を出せると考えられる．先頭優先除去の今後の課題は，本実験で記録したような比較的高い MRR がより多くのデータでも実現できるのか調べることと，よりふさわしい優先順位はないか調べることである．

5.3 重みによるランキング，単語数と重みによるランキング

重みによるランキングの MRR は提案手法 7 種類のうち 4 番目，単語数と重みによるランキングの MRR は 5 番目に高かった．重みによるランキングではクエリに含まれる単語のタグの重みの和によって検索結果を並べていたため，クエリの単語数が多いほどその検索結果が上位に表示される傾向にあり，結果的には MRR，ヒストグラム共に単語数と重みによるランキングとほぼ同じ結果となった．単語数が多いクエリは検索結果のヒット数が少なく，検索結果に正解書籍が含まれた場合相対的に正解書籍の順位が高くなる傾向にあるため，単語数の多いクエリが上位に並ぶこれらの手法はランダムランキングより良い結果であったと考えられる．2 種類の手法の比較により，重みによるランキングを単語数の加味の有無で比較する予定であったが，本実験で設定した重みではさほど差が出なかった．今後はより適切な重みを検討したい．

5.4 確率によるランキング

確率によるランキングの MRR は提案手法 7 種類のうちで一番低く，ランダムランキング，単語数によるランキングの平均 MRR よりも低かった．この手法ではクエリ内の単語につけられたタグの組み合わせのあらすじデータ内への同時出現確率が大きい順にクエリをランキングしていたが，この確率はクエリ内の単語が増えるほど低くなるため，単語数の少ないクエリの検索結果が上位に表示される傾向にあった．単語数の少ないクエリは検索結果のヒット数が多いため，正解書籍が含まれていても相対的に順位が低くなる．そのため，確率によるランキングの MRR は低かったと考えられる．しかし，ヒストグラムによって正解書籍の順位の分布を比較すると，ランダムランキングや単語数によるランキングでは正解書籍が 130000 位程度まで分布しているのに対し，確率によるランキングでは 60000 位までに収まっており，正解書籍の順位の下限は比較的高かったと言える．生成されたクエリの順番と検索結果を確認したところ，確率によるランキングでは正解書籍を含むが単語数が少ないクエリが上位となっているものが多かった．そのため，上位にランキングされたクエリの検索結果数が多くなり，相対的に正解書籍の順位が下がったと考えられる．

5.5 単語数と確率によるランキング

MRR は確率によるランキングの場合より大幅に高くなり提案手法 7 種類のうちで 3 番目に高かった．単語数の多い順にクエリをランキングすることで，正解書籍を検索結果に含むクエリのうち一番単語数の多いものによって全体の検索結果の正解書籍の順位が決まることになるので，特に質問内でタグ付けした単語が正解のあらすじに複数個含まれる場合は正解書籍が上位に持ってくることができたと考えられる．単語数によるランキングの平均 MRR が高いことからも，単語数の多い順にクエリを並べる手法に一定の効果があることがわかる．しかし，正解書籍がヒットする単語数が少ない場合は正解書籍を検索結果に含むクエリの検索結果が下位にランキングされて正解書籍の順位が下がるのが単語数が多い順にクエリを並べる手法全てに言える短所である．このため，単語数と確率によるランキングは期待値によるランキングより低い MRR にとどまったと考えられる．

5.6 期待値によるランキング

期待値によるランキングの MRR は提案手法 7 種類のうちで最も高かった．この手法では，各クエリの検索結果内の正解書籍の順位の期待値が小さい順にクエリをランキングすることで正解書籍が上位に置くことを目的とした．クエリの単語数が多いとあらすじデータへの同時出現確率が低くなるため正解が検索結果に含まれる可能性が低くなるが，ヒット数が少なくなる分正解書籍が含まれた際の順位が相対的に高くなる．正解書籍の順位の期待値を求めることで，正解が検索結果に含まれる可能性と生成したクエリによる検索結果のヒット数の両方を考慮してクエリを順序づけることが可能となり，高い MRR 値を実現できたと考えられる．ただし，ヒストグラムで比較すると，単純 tf-idf と正解書籍の順位の分布が似ており，データによってはあまり差が出ない可能性もある．今後の課題としては，より多くのデータで計算した確率を用いることや，ほかのデータに対する検索を行ってもうまくいくかどうか調査することが挙げられる．

6. 結　　論

本研究では，書籍のあらすじに関するあいまいな記憶に基づいて書籍を探す手法として，あらすじ中で語が果たす役割に基づいた優先順位があるという仮説に基づき，質問に含まれる単語に対して役割に基づいた分類を行い，その結果に基づいてクエリをランキングし検索結果を連結することでユーザの意図する書籍を上位に提示する手法を 7 種類提案した．次に，書籍のあらすじに関する記述によって書籍を探す質問データを収集し，提案した手法を適用することで提案手法の有効性の確認を試みた．その結果，クエリ中の語の役割の組み合わせがあらすじ内に出現する確率と各クエリの検索結果数を用いて正解書籍の順位の期待値を求め，小さい順にクエリをランキングする手法が最も有効とわかった．

本研究全体における今後の課題について述べる．まず，本研究ではデータを手作業で集めたため，多くのデータを集めることができず，予備実験と本実験で同じデータを用いた．提案手法の有用性を確かめるためにはより多くのデータで実験することが必要と考えられる．

また，本研究では，生成したクエリによる検索結果の表示順は国立国会図書館サーチで表示されたものをそのまま利用している．ユーザの意図をより反映したランキングにするためには，クエリ生成で除去された語に近い語を含むデータを上位に表示することが望ましいと考えられる．今後は検索結果のランキング方法についても検討し，よりユーザの意図に近づけたい．

本研究では手作業で単語の役割の分類を行ったが，従来の検索のようにキーワード群を入力とするとどの単語がどの役割かわからないという問題がある．したがって，実際に利用してもらう際にはユーザに文の形で入力してもらうなどの方法を考える必要がある．

以上より，本研究全体における今後の課題として，

- より多数のデータにより提案手法の有効性を確かめる
- クエリ生成で除去された語に近い語を含むデータを上位に表示するよう検索結果をランキングし直す
- 実際にユーザが利用する際のクエリ入力方法を考える

ことを挙げる．

謝　　辞

本研究は，JST，CREST（#JPMJCR16E3）の支援を受けたものである．

文　　献

[1] Efthimis N Efthimiadis. Query expansion. *Annual review of information science and technology (ARIST)*, Vol. 31, pp. 121–87, 1996.

[2] Xuanhui Wang and ChengXiang Zhai. Mining term association patterns from search logs for effective query reformulation. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 479–488. ACM, 2008.

[3] 近藤光正, 森田哲之, 田中明通, 内山匡. HITS に基づく wikipedia ランキングアルゴリズムとユーザ履歴を用いた個人適応型クエリ推薦. 電子情報通信学会第 19 回データ工学ワークショップ論文集, 2008.

[4] 大石哲也, 峯恒憲, 長谷川降三, 藤田博, 越村三幸. 関連単語抽出アルゴリズムを用いたクエリ拡張. In *DEIM Forum*, 2009.

[5] 金子恭史, 中村聡史, 大島裕明, 田中克己. 緩和度付き検索語の意味関連分析による検索意図推定とそのクエリ入力インタフェース. *Journal of the DBSJ*, Vol. 7, No. 1, 2008.

[6] 吉田大我, 小山聡, 中村聡史, 田中克己. Web 検索結果におけるキーワード出現相関の可視化と対話的な質問変換. 電子情報通信学会第 18 回データ工学ワークショップ (DEWS2007), 2007.

[7] Michael Bendersky and W. Bruce Croft. Analysis of long queries in a large scale search log. In *Proceedings of the 2009 Workshop on Web Search Click Data*, WSCD '09, pp. 8–14, New York, NY, USA, 2009. ACM.

[8] Yan Chen and Yan-Qing Zhang. A query substitution-search result refinement approach for long query web searches. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '09, pp. 245–251, Washington, DC, USA, 2009. IEEE Computer Society.

[9] Giridhar Kumaran and Vitor R. Carvalho. Reducing long queries using query quality predictors. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, pp. 564–571, New York, NY, USA, 2009. ACM.

[10] Niranjan Balasubramanian, Giridhar Kumaran, and Vitor R. Carvalho. Exploring reductions for long web queries. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pp. 571–578, New York, NY, USA, 2010. ACM.

[11] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 133–142. ACM, 2002.

[12] Ellen M. Voorhees. Query expansion using lexical-semantic relations. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pp. 61–69, New York, NY, USA, 1994. Springer-Verlag New York, Inc.

[13] Silviu Cucerzan and Eric Brill. Spelling correction as an iterative process that exploits the collective knowledge of web users. In *EMNLP*, Vol. 4, pp. 293–300, 2004.

[14] Shuya Ochiai, Makoto P. Kato, and Katsumi Tanaka. Recall and re-cognition in episode re-retrieval: A user study on news re-finding a fortnight later. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pp. 579–588, New York, NY, USA, 2014. ACM.

[15] 隅田飛鳥, 池田和史, 服部元, 小野智弘. 9-10 あいまいな記憶下での抽象的クエリを許容する映画検索エンジン (第 9 部門ヒューマンインフォメーション). 映像情報メディア学会年次大会講演予稿集 2012, pp. 9–10. 一般社団法人映像情報メディア学会, 2012.

[16] 水野淳太, 渡邉陽太郎, エリックニコルズ, 村上浩司, 乾健太郎, 松本裕治. 文間関係認識に基づく賛成・反対意見の俯瞰. 情報処理学会論文誌, Vol. 52, No. 12, pp. 3408–3422, dec 2011.