

Q&A コーパス学習に基づく Web 検索のための意味類似クエリの発見

福地 大助[†] 山本 岳洋[†] 田中 克己[†]

[†] 京都大学大学院情報学研究科 〒 606-8501 京都市左京区吉田本町

E-mail: †{fukuchi,tyamamot,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし Web 検索において、ユーザは自らの検索意図に沿ったクエリを選択しているつもりでも、適切な検索結果が得られないことがある。本論文では、質問応答コーパスとして Q&A コーパスを位置づけ、ユーザの質問に類似した、適切な語彙を用いている質問、またはその回答から知識を抽出し適切なクエリを生成する手法を提案する。提案手法では、まず Q&A コーパスを用い、単語の分散表現を学習する。次に、ユーザの情報要求のクエリと、その中でどれが変換すべき語であるかを示す変換対象キーワードを入力として取る、次に、初期クエリに類似した質問とその回答を取得し、そこから変換対象キーワードに類似する語を変換候補としてを抽出する。そして、初期クエリを変換候補で変換したクエリが Web 検索に有用であるかを評価し、ユーザに提示する。

キーワード 情報検索, クエリ推薦, 自然言語処理

1. はじめに

近年、Web 検索エンジンの急速な発展により、多くの人々が検索エンジンを使用して情報を獲得している。PageRank や learning to rank に代表される様々なランキング手法により、Web 検索の精度は大きく向上してきた。現在では、ユーザは自らの検索意図に沿ったクエリを検索エンジンに投入するだけで、多くの事柄を求める情報を探ることが容易になっている。

しかし、ユーザが自らの検索意図に沿ったクエリを選択しているつもりでも、適切な検索結果が得られない場合がある。特に、ユーザの語彙力が低く、口語的な語や汎用性の高くて様々な意味に捉えられる語を含んでいるようなクエリは、ユーザが検索意図を良く表していると考えているクエリでも、異なる検索意図に関する検索結果しか得られないことがある。

たとえば、あるユーザが「宇宙は真空であるのに日光で暖まるのはなぜ」という検索意図で「日光 真空 暖まる」というクエリを実行したとする。この場合、この検索意図に適合する検索結果ではなく、「日光と月光」や「日光に対する住宅」に関するページが上位にランキングされてしまう。また、「真空なのに日光で暖まる理由」というクエリでフレーズ検索を行ったとしても、「真空なのに日光で暖まる理由」というフレーズを含んでいるページは存在せず、適合する検索結果を得ることができない。このように、ユーザが実行した「日光 真空 暖まる」というクエリは、検索意図を適切に表していると考えられるものの、現在の検索エンジンでは求める情報を得ることができない。

意図に沿った検索結果が得られない要因として、Web ページで使われている語彙とユーザの用いる語彙が一致しないこと、つまり、ユーザの語彙空間とページ作成者の語彙空間に乖離が存在することが考えられる。ユーザが「日光」というキーワードをクエリに用いていたとしても、ユーザの情報要求を満たす Web ページで使われている語は「太陽光」や「熱伝導」であることがある。一方、人と人との間の質問応答の場合、質問者の語彙力が低くても、回答者が質問者の情報要求意図を理解し、

質問者の語彙を正しく言い換えたり、質問内容を言い換えて確認をするインタラクションがあることで、要求を満たす結果を返すことが可能になっている。本研究では、質問応答コーパスとして Q&A コーパスを位置づけ、ユーザのクエリに類似して適切な語彙を用いている質問、またはユーザのクエリに酷似している質問に対応する回答から語彙の知識を抽出し、それを用いて適切なクエリを生成する手法を提案する。

たとえば、Q&A コーパス内に「なぜ宇宙は真空なのに日光で暖まるのか？」や「なぜ宇宙は真空なのに太陽光で暖まるのか？」という質問があり、「熱伝導には 3 種類の現象がある」という回答があれば、これらの質問と回答から「日光 真空 暖まる」というクエリを「太陽光 真空 暖まる」や「熱伝導 真空 暖まる」というクエリに変換し、検索意図に適合したページを取得することができる。このように、システムが Q&A コーパスから自動的にこうした質問または回答を発見し、代替クエリをユーザに提示することで、初期クエリで意図する検索結果が得られなかった場合に、提示されたクエリで適合するページを得ることができるようになると考えられる。

提案手法は Q&A コーパスから語彙の質問応答関係を学習し、それに基づいて検索意図をより適切に表現した語を発見することで、クエリを生成する。提案手法ではまず、ユーザが情報要求を満たすと考える初期クエリと、その初期クエリ中で変換すべきであると考えられる変換対象キーワードを入力する。次に、初期クエリに類似する質問とそれに対応する回答を取得し、質問集合と回答集合を生成する。さらに、それらの集合の要素を構成している文のうち、初期クエリに類似するものから変換候補を取得する。そして、得られた候補を基にクエリを生成し、Web 検索に有用であるかどうかを評価するクエリ妥当性に基づきランキングしユーザに提示する。

本稿の構成は以下の通りである。2 節では、関連研究について述べる。3 節では、提案手法について詳細に述べる。4 節では、提案手法に関する実験の概要とその評価結果について述べる。5 節では、まとめと今後の課題について述べる。

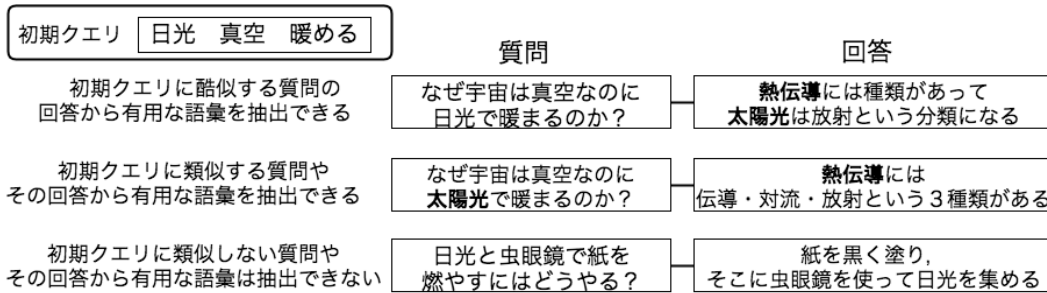


図 1 Q&A から語彙を抽出するアイデア

2. 関連研究

ユーザのクエリを修正したり、他の有用なクエリを発見してユーザに提示することで Web 検索を改善させる手法が多く提案されてきた。本節では、本研究に関連する研究について述べる。

2.1 クエリ修正

本研究は、ユーザが投入したクエリ中の単語を別の単語に置換することでクエリ修正を行う研究の 1 種である。単語の置換によりクエリ修正を行う研究としては、Boldi らの研究がある [1]。Boldi らはクエリ修正のタイプとして詳細化、並列化、エラー修正といったカテゴリを提唱し、そのようなカテゴリに応じてクエリを修正する手法を提案している。Mandala らは複数種類のシソーラスを組み合わせることで、重要な単語を発見し、クエリ拡張として利用する手法を提案している [7]。また、Xu らは初期クエリで上位にランキングされる検索対象からクエリ拡張に用いる語を取得する方法を提案している [11]。また、Xu らの研究では、クエリ拡張の候補として名詞を想定していたのに対して、本研究ではクエリ変換の候補として名詞、動詞、形容詞または副詞のいずれかを対象とし、クエリ変換の候補として扱う語の品詞の種類が大きく異なる。

また、クエリ修正に関する他の研究としては、クエリ推薦を多様化する研究 [6]、初期クエリに対して別の語を付加することでクエリ修正を行う研究など、クエリ修正に関する研究は多く存在し [9] [5]、クエリログ [13] [8] などを用いてクエリ修正を取得する研究も多く存在する。また、クリックログ [4] を用いて推薦する研究もされており、これらの研究は、ユーザの入力履歴などの URL をクリックしているかの履歴を用いてユーザの検索意図に適合するクエリを推薦するというものである。

4 章で述べる実験では、単純に Word2Vec を用いた類義語による出力、Bing によるクエリ推薦のそれぞれと提案手法を比較することで提案手法の有用性を検証する。

2.2 Q&A コーパスを用いた検索支援

Q&A コーパスを用いて検索支援を行う研究は多数存在する。Zhou らは質問文を入力として、その質問文内の 2 語以上からなるフレーズを別の表現に変換することで、質問検索や回答検索の精度を向上させる手法を提案している [3]。Xue らは質問内で用いられている語が回答内ではどのような表現で用いられているかをモデル化することで、語の変換を行い、質問検索や回答検索に活用する手法を提案している [12]。

これらの研究は、Q&A コーパスから知識を抽出することで質問検索や回答検索を支援するものであるが、本研究は Q&A コーパスから抽出した語彙を用いて検索エンジンによる Web 検索を支援しようとする点が大きく異なる。

3. 提案手法

本節ではまず、本研究で用いる Q&A コーパス、変換対象キーワードの変換候補を取得するアイデアについて順に説明する。その後、手法の概要を示し、各手法の詳細について述べる。

3.1 Q&A コーパス

Q&A コーパスには Yahoo!知恵袋を用いた。通常の Q&A では、1 つの質問に対して複数の回答がつくこともあり、Yahoo!知恵袋もそれに倣う。しかし Yahoo!知恵袋において質問者は、複数ある回答の中で質問の情報要求を最も満たしている回答を 1 つ選ぶことができ、これをベストアンサーと呼ぶ。この回答はその質問者が知りたい内容の答えとして最も良い回答であると人手で判断したものであるため、本研究ではこの回答を質問に 1 対 1 対応するものとみなす。

Q&A コーパスを D とすると、定義は以下のとおりである。

$$D = \{(q_1, a_{q_1}), \dots, (q_n, a_{q_n})\}$$

q_i は質問であり、 a_{q_i} は q_i に付与されたベストアンサーである。また、 q_i は複数の文で構成されるものもあり、本研究ではこの 1 つの文のことを質問文と呼ぶ。

3.2 研究のアイデア

1. 節で述べたように、ユーザの意図に沿う検索結果が得られない要因として、ユーザが用いる語彙とページ作成者が用いる語彙に乖離が存在することが考えられ、ユーザが“日光”という語をクエリに用いていたとしても、ユーザの情報要求を満たす Web ページで使われている語は“太陽光”や“熱伝導”であることがある。一方、人と人の中での質問応答の場合、質問者が適切な語彙を用いていなくても、回答者が質問者の情報要求意図を理解し、質問者の語彙を正しく言い換えたり、質問内容を言い換えて確認したりすることで、要求を満たす結果を返すことが可能になっている。そこで、本研究では、質問応答が行われる Q&A コーパスを用いて、ユーザのクエリに類似している質問やそれに対応する回答から適切な語彙の知識を抽出し、クエリ変換を行う。Q&A から変換候補として有用な語を抽出するアイデアを図 1 に示す。

初期クエリを $q = \text{“日光 真空 暖まる”}$ とすると、「なぜ宇宙は

真空なのに日光で暖まるのか？」という質問は q に酷似しているため、有用な変換候補が含まれていない。しかし、そのような質問が存在すれば、その質問に対応する回答から質問の言い換えや答えとなる語が得られ、Web 検索に有用な語を抽出できると考えられる。

「なぜ宇宙は真空なのに太陽光で暖まるのか？」という質問は q に類似しており、この文脈において“太陽光”が“日光”の言い換えになっている可能性があり、変換候補として“太陽光”を抽出することは有益である。またその質問が q に類似しているならば、上述と同様に、それに対応する回答も Web 検索に有用な語を含んでいると考えられる。

一方、初期クエリに全く関係のない質問や、「日光と虫眼鏡で紙を燃やすにはどうしたらいい？」という初期クエリとは異なるトピックの質問、またその回答からは初期クエリを変換するための有用な語を抽出できないと考えられる。

3.3 提案手法の流れ

提案手法は、ユーザによって与えられたクエリ $q = \{k_1, \dots, k_n\}$ と、 q に含まれる変換対象キーワード k^c を入力とし、 k^c を変換することで得られるクエリ集合を、ユーザの意図に沿う順でランキングした上位 8 件を出力する。提案手法の流れを図 2 に示す。

3.4 Q&A コーパスを用いた Word2Vec の学習

Word2Vec を用いて、後述するキーワード集合間の類似度計算や変換必要性キーワードの変換候補の取得を行うことによって、ユーザが初期クエリに対して抱いていた情報要求に沿う文や変換候補として有用な語を選択できると考えられる。

Word2Vec とは、Mikolov らが提案した手法で、単語をベクトル化して表現する定量化手法である。その各単語は D 次元の空間内におけるベクトルとして表現することができる。

本手法では、Q&A コーパスに用いて Word2Vec に言葉の意味を学習させることで、質問応答における語彙空間における単語の意味を学習することが可能である。また、1 節でユーザの検索意図に適合する検索結果が得られない要因の 1 つに、ユーザが使う語彙とページ作成者が使う語彙に乖離が存在する問題をあげたが、このモデルを生成することで、Q&A コーパスにおいてユーザが使う適切でない語彙に意味が類似した語を発見することができるため、この問題に対処する手がかりとなる。学習用のコーパスには、Q&A コーパスから 2004 年 4 月 2 日から 2014 年 12 月 16 日の期間に行われた質問または回答を無作為に 5,000,000 件を用いた。

3.5 文集合の生成

Yahoo 知恵袋などの Q&A サイトで質問されている質問や回答の長さは短いものもあれば長いものもあり、その中には質問や回答の本質とは言えない部分も多く含まれている。たとえば、「今の日本の内閣総理大臣って誰ですか？」という質問は質問意図として本質的な部分を端的に表しており、これだけでなにを情報要求としているのか理解できる。一方、「太平洋戦争をはじめのきっかけを作った総理大臣は誰ですか？やはり東條英機でしょうか？太平洋戦争の主因の一つになった日本による中国侵略については、近衛文麿の罪は大きいのですが…（以下略）」という質問は、冒頭で質問意図を明確に表しているが、その後

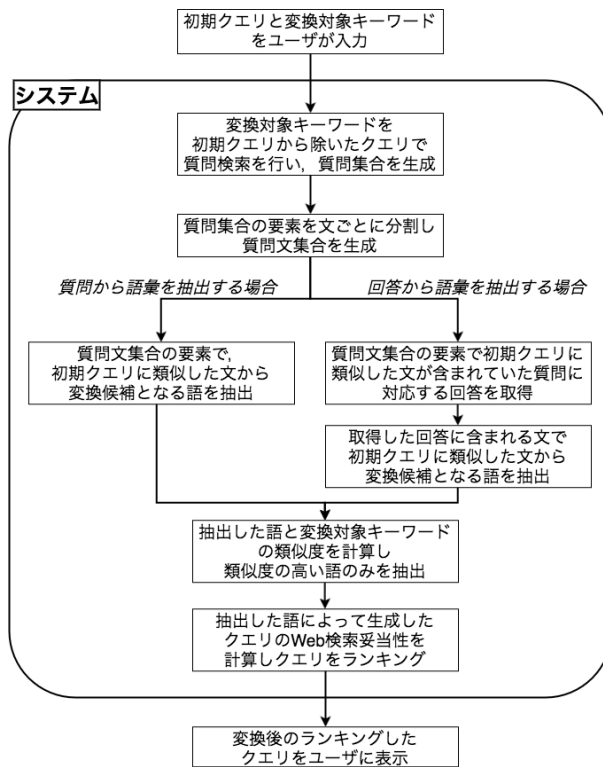


図 2 提案手法の流れ

の文章は質問の本質的な部分とは言い難く、この質問全体を同時に処理すると変換候補を取得する上でノイズになる懸念がある。そこで質問や回答は複数の文から構成されていると考え、文の区切りで分割し、ノイズとなる文の除去や本質的な文を抽出する手法を適用する。

本手法は、文の区切りが“。”と“！”，“？”のいずれかであるとして質問や回答を文に分割することで、初期クエリの意図に類似した、質問や回答として本質的な文の抽出を行うための前処理を行う。たとえば、前者の質問は「今の日本の内閣総理大臣って誰ですか？」という 1 つの文で構成されているとみなすことができ、後者の質問は「太平洋戦争をはじめのきっかけを作った総理大臣は誰ですか？やはり東條英機でしょうか？」「太平洋戦争の主因の一つになった日本による中国侵略については、近衛文麿の罪は大きいのですが…（以下略）」という 3 文以上で構成されているとみなすことができる。

3.6 クエリと文の類似度計算

3.5 節において質問や回答を複数の文に分割する手法をみてきた。変換候補の取得の手がかりとして、これらの文の中からノイズとなる文と情報要求として本質的な文を区別し、後者から変換対象キーワードの変換候補として有用な語を抽出する必要がある。そこで、クエリと文の類似度を計算することで、それがクエリの情報要求を満たす文であるかどうかを判定する。

本手法は、文とクエリをともにキーワード集合であるとみなして、キーワード集合間の類似度を計算することで文とクエリの類似度を算出し、文にスコアを付与する。キーワード集合を $X = \{x_1, \dots, x_n\}$ 、 $Y = \{y_1, \dots, y_m\}$ とすると、キーワード集合間の類似度の計算には、Kajiwara と Komachi ら [10] による、

以下の式を用いる．

$$\text{Sim}_{\text{asym}}(X, Y) = \frac{1}{|X|} \sum_{i=1}^{|X|} \max_j \phi(x_i, y_j)$$

$$\text{Sim}_{\text{sym}}(X, Y) = \frac{1}{2} \text{Sim}_{\text{asym}}(X, Y) + \text{Sim}_{\text{asym}}(Y, X)$$

ここで、 X, Y はキーワードの分散表現集合であり、 x_i, y_j はその集合の要素である分散表現、 $\phi(x_i, y_j)$ は x_i と y_j の類似度を計算する関数、本研究ではコサイン類似度を用いる． X が Y に類似することと Y が X に類似することは同値ではないため、上式は X が Y にどれほど類似しているか計算し、また Y が X にどれほど類似しているかをそれぞれ計算してその平均を取ることで、 X と Y の類似度を計算している．この $\text{Sim}_{\text{sym}}(X, Y)$ を初期クエリ q と文 t_i に用いることで、 q が意味する意図と t_i が表している意図がどれほど似ているか評価することができる．

3.7 変換対象キーワードの変換候補の取得

1 節で例示した“日光 真空 暖まる”という初期クエリの場合、変換必要性キーワード“日光”を“太陽光”に変換すれば、検索意図に適合したページを取得することができる．本節では、こうした変換対象キーワードの適切な変換候補を発見し、変換候補集合 $K^c = \{k_1^c, \dots, k_n^c\}$ を生成する手法について述べる．

3.6 節においてクエリと文をそれぞれキーワード集合とみなして類似度を計算する手法をみてきた．本手法では、変換対象キーワードの変換候補として有用な語は初期クエリ q の情報要求を満たす文、つまり q に類似する文に含まれると考えられるため、その類似度が大きい文を抽出する．また、3.2 節で述べたように、質問応答が行われる Q&A コーパスでは、質問として q と情報要求は同じであるが、情報要求文として q より適切な語彙を用いたものが含まれると考えられる．また回答として、質問者の情報要求意図を理解して語彙を正しく言い換えたり、質問内容を言い換えて確認したりするものも Q&A コーパスには含まれると考えられるため、質問内の質問文と回答内の回答文のそれぞれから変換対象キーワードの変換候補として有用な語を取得できると考えられる．以降では、質問文と回答文のそれぞれから変換候補を取得する手法について述べる．

3.7.1 質問文からの取得

質問の中には、初期クエリに似た情報要求を持っており、また初期クエリよりも適切な語彙が用いられている質問があると考えられる．そして、その質問を構成する複数の文から、情報要求を本質的に表す文を抽出できれば、そこから変換対象キーワードの変換候補として有用な語を抽出できる．しかし、3.5 節の手法によって生成された質問文集合 $Q^t = \{t_1^t, \dots, t_n^t\}$ にはノイズが多く含まれるため、それを除去する必要がある．

本手法では、3.6 節の手法で初期クエリ q と $t_i^t \in Q^t$ の類似度 s_i を計算する．そして閾値 θ とし、 $s_i > \theta$ となる t_i^t だけからなる質問文集合 $Q_{>}^t$ を生成する．そして、 $t_i^t \in Q_{>}^t$ を有用な文であるとして、 t_i^t に含まれている語 $w \notin q$ を変換対象キーワードの変換候補として抽出する． w は名詞・動詞・形容詞・副詞のいずれかである．今回、 $\theta = 0.5$ に設定した．

例えば、 $q = \text{“日光 真空 暖まる”}$ 、 $t_1^t = \text{“日光が宇宙を通じて$

地球を暖めるのはなぜ?”、 $t_2^t = \text{“太陽光が真空中でも伝わって暖まるのはなぜ?”}$ 、 $t_3^t = \text{“日光を虫眼鏡で集めて紙を燃やすには?”}$ として、類似度がそれぞれ $s_1^t = 0.7$ 、 $s_2^t = 0.6$ 、 $s_3^t = 0.4$ とすると、 t_1^t からは“宇宙”、“通じる”、“地球”を、 t_2^t からは“太陽光”、“伝わる”を変換候補として抽出する． $s_3 > \theta$ は満たされないため、 t_3^t から変換候補の抽出はしない．

3.7.2 回答文からの取得

回答の中には、質問者の情報要求意図を理解することで、質問者の語彙を正しく言い換えている回答があると考えられる．また、その回答を構成する複数の文から、情報要求を本質的に表す文や質問を正しく言い換えている文を抽出できれば、そこから変換対象キーワードの変換候補として有用な語を抽出できる．しかし、初期クエリ q に類似していない質問に対応する回答には、 q のクエリ変換として有用な変換候補が含まれている可能性が低い．また、 q に類似している質問に対応する回答である場合も、3.7.1 節と同様に、生成した回答文集合にはノイズが多く含まれる．そこで、 q に類似する質問に対応する回答を抽出し、またそこに含まれたノイズを除去する必要がある．

本手法では、まず、3.7.1 節において生成した $t_i^t \in Q_{>}^t$ に対し、 t_i^t を含んでいた質問 q_x に対応する回答 a_{q_x} を取得し、回答文集合 $A^t = \{t_1^t, \dots, t_m^t\}$ を生成する．次に、3.7.1 節と同様に、初期クエリ q と $t_i^t \in A^t$ の類似度 s_i^t を計算する．そして、 $s_i^t > \theta$ となる t_i^t だけからなる回答文集合 $A_{>}^t$ を生成する．そして、 $t_i^t \in A_{>}^t$ 有用な文であるとして、 t_i^t に含まれている語 $w \notin q$ を変換対象キーワードの変換候補として抽出する．

3.8 最適な変換候補の選択

3.7 節の手法によって取得される変換必要性キーワードの変換候補集合 $K^c = \{k_1^c, \dots, k_n^c\}$ の中にはノイズが含まれるため、すべての変換候補が有用であるとは限らない．そこで我々は、各変換候補の有用性を推定し、その有用性の高いものを用いてクエリ変換を行う．

本手法の目的は、初期クエリ中の変換対象キーワードと意味が類似し、変換後のクエリが Web 検索が行ううえで妥当なクエリのような変換候補を発見することである．そのために以下の 2 つの指標を順番に用いて、変換候補の有用性スコアを計算する．具体的には、第 1 の指標によるスコアで変換候補をランキングして上位 30 件を抽出するフィルタリングを行う．そして、その上位 30 件それぞれに対して第 2 の指標によるスコアを計算し、そのスコアを変換候補の有用性スコアとして付与する．

(1) 変換対象キーワードと変換候補の類似度

(2) Web 検索におけるクエリ妥当性

この指標によって得られた有用性スコアの降順にクエリ変換結果をランキングし、その上位 10 件を出力する．以降では、上述の各指標の計算手法について述べる．

3.8.1 変換必要性キーワードとその変換候補の類似度

ユーザの検索意図に沿ったクエリ変換を行うためには、変換対象キーワードに類似する変換候補が必要となる．そこで、変換対象キーワード k_i^c とその変換候補 $k_j^c \in K^c$ の類似度を、

- k_i^c を抽出した質問文 (回答文) と、それに対応する回答文 (質問文) において出現したかどうか

- k^c と k_i^c の分散表現のコサイン類似度

から計算する．変換候補を w ，初期クエリを q ，および w を抽出した文とその文に対応する質問文または回答文のペアを (t, t^a) とすると， w と k^c の類似度 $S(w, k^c)$ は以下で表される．

$$S(w, k^c) = (1 - \alpha) \cdot f(w, t) + \alpha \cdot f(w, t^a)$$

$$\alpha = \frac{\text{Sim}_{\text{sym}}(q, t) - 1}{1 - 1}$$

$$f(w, s) = \begin{cases} 0 & (w \notin s) \\ \text{Cos}(w, k^c) & (w \in s) \end{cases}$$

$\text{Sim}_{\text{sym}}(q, t)$ は 3.6 節で述べたクエリと文の類似度計算であり， $\text{Cos}(w, k^c)$ は， w と k^c それぞれの分散表現のコサイン類似度である．また， f は 3.7.1 節で定義したものである．

3.7.2 節における変換候補取得の方法から， w を t^A から抽出した場合，質問文と回答文のペア (t, t^A) が得られるが， w を t から抽出した場合， t^A が得られない場合がある．その場合は，

$$S(w, k^c) = (1 - \alpha) \cdot f(w, t)$$

として類似度を計算する．

3.8.2 Web 検索におけるクエリ妥当性

初期クエリから変換対象キーワードを取り除いた $q \setminus \{k^c\}$ と変換候補 $k_i^c \in K^c$ の関連性が低い場合，それらを用いて変換されたクエリは Web 検索を行ううえで妥当とは言えず，ユーザの意図する検索結果を取得することは困難であると考えられる．そこで我々は， $q \setminus \{k^c\}$ と k_i^c の関連の強さを，両者の共起度を測ることで，変換されたクエリが Web 検索を行うクエリとして妥当であるかを計算する．

語の共起度の計算には Bollegala らによって提案された WebPMI [2] がある．WebPMI は，語 p, q の共起度を検索エンジンを用いて測る尺度であり，その値は次式で計算される．

$$\text{WebPMI}(p, q) = \log_2 \left(\frac{\frac{\text{DF}(p \cap q)}{N}}{\frac{\text{DF}(p)}{N} \cdot \frac{\text{DF}(q)}{N}} \right)$$

$\text{DF}(x)$ は検索エンジンから返ってくるキーワード x によるクエリの検索結果文書数である．ここで， $x \cap y$ はキーワード x と y を用いた AND 検索， N は検索エンジンにインデックスされている全てのページ数を表す．我々は Bollegala [2] に従い， $N = 10^{10}$ と設定した．

変換対象キーワードを k^c ，変換候補を $k_i^c \in K^c$ ，初期クエリを q とすると，我々は WebPMI を適用した

$$\text{WebPMI}(q \setminus \{k^c\}, k_i^c) = \log_2 \left(\frac{\frac{\text{DF}((q \setminus \{k^c\}) \cap k_i^c)}{N}}{\frac{\text{DF}(q \setminus \{k^c\})}{N} \cdot \frac{\text{DF}(k_i^c)}{N}} \right)$$

により， $q \setminus \{k^c\}$ と k_i^c の Web 上での共起度を計算する．この $\text{WebPMI}(q \setminus \{k^c\}, k_i^c)$ を用いることで， k^c を k_i^c で変換したクエリの Web 検索におけるクエリ妥当性を計算できる．

4. 実 験

本節では，提案手法の有効性を検証するために用いた行った評価実験について述べ，実験結果を基に考察を行う．

4.1 概 要

4.1.1 実験に用いる提案手法

提案手法の有効性を評価するために，質問文集合や回答文集合から，各評価クエリに対するクエリ変換候補を取得した．3.7.1 節，3.7.2 節で述べたように，質問文集合には初期クエリよりも適切な語彙が用いられている質問があると考えられ，回答文集合には質問者の語彙を正しく言い換えている回答があると考えられる．これらから取得した語が有効に機能するケースや限界点を明らかにするために，取得した語によるクエリ変換結果を評価した．変換候補を取得する際にはコミュニティ Q&A サービスである Yahoo!知恵袋を用いた．また，3.8.2 節において変換後クエリを評価する際には Bing Search API を用いた．

4.1.2 ベースライン手法

提案手法と比較するベースライン手法として，2 つの手法を用意した．1 つ目の手法として，Bing の検索エンジンが提示するクエリ推薦を採用した．同検索エンジンに，“日光 真空 暖まる”というクエリを入力した場合，

- 暖まる
- 体が暖まる
- 温まる 暖まる
- 暖まる 英語
- 暖まる 温まる 違い
- 心が暖まる
- 体が温まる 暖まる
- 真空波動研

といった 8 件の推薦クエリが検索結果ページの下部に表示される．これらの推薦クエリは，元の検索クエリの結果に不満のあるユーザに対して，その問題を解決するために検索エンジンが提示したものとみなせるため，提案手法との妥当な比較対象である．以降，本手法を Bing クエリ推薦手法と呼ぶ．

2 つ目の手法として，Word2vec で学習されたモデルを用いて変換対象キーワードに類似する語を得る手法を採用した．Word2Vec に対し，“日光”という語を入力した場合，“紅葉”，“日陰”，“直射日光”，“箱根”，“日差し”，“函館”，“展望台”，“海水浴”という出力を得ることができる．これらは，質問応答空間において変換対象キーワードと意味が類似しており，キーワードの言い換えを実現しようとする提案手法との妥当な比較対象である．以降，本手法を Word2Vec 手法と呼ぶ．

4.2 評価クエリ

実験にあたり，被験者を用いてクエリ作成を行い，10 件の評価クエリを用意した．本節では，被験者を用いた評価クエリ作成実験について述べる．

実際のユーザの検索意図と，その検索意図を基に用いられる可能性のあるクエリで実験を行うために，Yahoo!知恵袋を用いて 10 件のクエリを作成した．クエリを用意するにあたり，Yahoo!知恵袋から，質問者の情報要求を満たすための質問文として多様な表現が考えられるような質問を著者が 10 件選択した．表 1 に選択した 10 件の質問を示す．ここでは，3.4 節で学習した Q&A コーパスには含まれない質問を選択した．その後，5 名の被験者に対して，それぞれの質問文を見せ，「あなたが質問者の立場であるとして，この質問の情報要求を満たすにはどのようなクエリを作成すると思いますか？好きな語彙を用

表 1 Yahoo!知恵袋の質問文を基に被験者が作成した、評価に用いる初期クエリとそのクエリの検索意図、および当該質問の Yahoo!知恵袋内での質問 ID

初期クエリ	変換対象キーワード	検索意図	質問 ID
黒人 男性 坊主 髪型	坊主	黒人男性に坊主が多いのはなぜか	q10169983700
昆虫 翅 4 枚	翅	昆虫の翅が 2 枚ではなく 4 枚なのはなぜか	q11168763831
旨い 美味しい 語彙 違い	語彙	「旨い」と「美味しい」の違いはなにか	q10167788640
動物 自然死 姿 いない	いない	自然死した動物の姿をほとんど見かけないのはなぜか	q1128064321
日本 西暦 元号 統一	統一	日本で未だに元号が使われているのはなぜか	q10169092715
世界 言語 統一 可能	可能	世界の言語を統一しないのはなぜか	q12168082112
マニュアル オートマ 性能 違い	性能	自動車のマニュアルはオートマより燃費がいいのか	q11156983556
日光 真空 暖まる	日光	宇宙は真空であるのに日光で暖まるのはなぜか	q10167188345
ワイン 蓋 コルク 起源	起源	ワインの蓋するのにコルクが使われるようになったのはなぜか	q13169713648
タバコ 税金 値上げ たびたび	たびたび	たばこ税ばかり値上げされるのはなぜか	q13167378548

い、思いつくクエリを最大 10 個記述してください」という指示の下、被験者にクエリを作成してもらった。5 名の被験者はいずれも情報学を専門とする学生であり、男性であった。表 1 中の“検索意図”は、質問が表すと考えられる検索意図を被験者に記述してもらったものを表している。

4.3 評価方法および評価尺度

表 1 に示した合計 10 件の評価クエリそれぞれに対して、提案手法、Bing クエリ推薦手法および Word2Vec 手法の精度を評価した。本節では、それぞれの手法の評価方法、および評価尺度について述べる。4.1 節で述べた各手法で取得された 8 件のクエリ変換候補のランキングを「想定意図適合性」という観点から評価した。これは、表 1 に示されるクエリの背後に隠れたユーザの本来の検索意図に適合する検索結果を返すクエリを正解とみなし、そうしたクエリが上位 k 件のクエリ変換候補の中にどの程度存在するかを評価する。各クエリの正解判定時には、Google の上位 10 件の検索結果を評価した。

例として、初期クエリ“日光 真空 暖まる”にたいして、

- (1) 太陽光 真空 暖まる (2) 日差し 真空 暖まる
(3) 昼過ぎ 真空 暖まる (4) 真上 真空 暖まる

というクエリ推薦のランキングを取得したとする。表 1 に示すように、初期クエリの背後に存在する本来の検索意図は「宇宙は真空であるのに日光で暖まるのはなぜか」である。上記のランキングのうち、“太陽光 真空 暖まる”および“日差し 真空 暖まる”というクエリからはこの意図に適合する検索結果が得られるため、正解であるとみなす。その他のクエリでは、初期クエリと異なる検索結果が得られるが、初期クエリに関連する検索意図とは言えないため、不正解とみなす。評価実験では、これらのクエリの正解判定 (適合・不適合の 2 値) を人手で行った。以降、適合と判定されたクエリのことを正解クエリと呼ぶ。

各手法によって生成されたクエリ変換候補のランキングの評価に用いた尺度について述べる。初期クエリの本来の検索意図に関する適合性を評価するために、 $MRR@k$ 、 $Success@k$ 、 $Max@k$ 、 $SDCG@k$ という 4 種類の尺度を用いた。

$MRR@k$ は、上位 k 件のランキングの中で正解クエリの順位を考慮した評価尺度である。 T を評価クエリ集合であるとすると、 $MRR@k$ の計算式は以下で定義される。

$$MRR@k = \frac{1}{|T|} \sum_{t \in T} RR(t)$$

ここで、 $RR(t)$ は評価クエリ t のクエリ変換候補のランキング

の上位 k 件の中で最初に出現する正解クエリの順位の逆数として計算される。上位 k 件中に正解クエリが存在しない場合は、 $RR(t) = 0$ とする。多くの評価クエリで正解クエリがランキングの高順位に存在する場合に $MRR@k$ の値は 1 に近づく。

$Success@k$ は、正解クエリの順位によらない尺度であり、上位 k 件以内に正解クエリを 1 個でも含む評価クエリの割合を評価する。 $Success@k$ の具体的な定義式は、 T を評価クエリ集合とした時に以下で表される。

$$Success@k = \frac{1}{|T|} \sum_{t \in T} E(t)$$

ここで、 $E(t)$ は評価クエリ t に対する上位 k 件のランキング中に 1 個でも正解が含まれる場合に 1 を、それ以外の場合に 0 を返す関数である。

$Max@k$ は、 $Success@k$ 同様に、正解クエリの順位によらない尺度であり、 T を評価クエリ集合とした時、評価クエリ $t \in T$ に対する上位 k 件のランキングのクエリ t_i に対し、 $DCG_{t_i}@10$ を計算し、その中の最大値を t のスコアとし、その平均を取ることによって計算される。 $Max@k$ の具体的な定義式は、 T を評価クエリ集合、 $DCG_{t_i}@10$ を評価クエリ t に対する i 番目のクエリ推薦 t_i を用いて得られた値とした時に以下で表される。

$$Max@k = \frac{1}{|T|} \sum_{t \in T} \arg \max_{1 \leq i \leq k} DCG_{t_i}@10$$

$SDCG@k$ は、現れる正解クエリの最大スコアを用いる $Max@k$ に対し、上位 k 件のランキングの中で正解クエリの順位を考慮した評価尺度である。 T を評価クエリ集合、 $DCG_{t_i}@10$ を評価クエリ t に対する i 番目のクエリ推薦 t_i を用いて得られた値とすると、以下で表される。

$$SDCG@k = \frac{1}{|T|} \sum_{t \in T} \sum_{i=1}^k \frac{DCG_{t_i}@10}{\log_2(i+1)}$$

実験では、 $k = 8$ とし $MRR@8$ 、 $Success@8$ 、 $Max@8$ および $SDCG@8$ を実際の評価尺度に用いた。 $Success@8$ および $Max@8$ は、ユーザにとって有用なクエリを出力しているかというクエリ推薦の精度を測る尺度である。一方、 $MRR@8$ および $SDCG@8$ は、推薦されたクエリのランキングも評価に加えた上で精度を測る尺度である。

4.4 実験結果

本節では、提案手法およびベースライン手法 (Bing クエリ

表 2 評価結果

	Bing クエリ推薦手法	Word2Vec 手法	提案手法
Success@8	0.400	0.500	0.800
MRR@8	0.263	0.099	0.300
Max@8	0.276	0.368	0.801
SDCG@8	0.404	0.184	0.616

推薦手法と Word2Vec 手法)によるクエリ変換の実験結果を報告する。まず、提案手法、Word2Vec 手法および Bing クエリ推薦手法によるクエリ変換候補のランキングを 4.3 節で述べた想定意図適合性に基づき、MRR@8, Success@8, Max@8 および SDCG@8 を用いて評価した。その結果を表 2 に示す。

まず、Bing クエリ推薦手法と Word2Vec 手法とを比較すると、Word2Vec 手法の方が Success@8 と Max@8 において高い精度を得ており、Bing クエリ推薦手法よりも高い精度で正解クエリを得ることができることが分かった。たとえば、Word2Vec 手法では、Success@8 の値が 0.500 となっており、全 10 件の評価クエリのうち 5 件のクエリについて同手法が出力した上位 8 件のランキングの中に適切なクエリ変換が含まれていたことを示している。一方、MRR@8 と SDCG@8 においては Bing クエリ推薦手法が高い精度を示す結果となった。

次に、Word2Vec 手法と提案手法とを比較すると、4 つの評価尺度において、提案手法が Word2Vec 手法に対して非常に高い精度を示す結果となった。提案手法の中でも Word2Vec による語の類似度計算を行う処理があったが、精度に大きな差が生じた考察については、4.5.1 節にて議論する。

最後に、Bing クエリ推薦手法と提案手法とを比較すると、4 つすべての評価尺度においては提案手法が高い精度を示す結果となった。提案手法が Success@8 と Max@8 において非常に高い精度を示した一方で、MRR@8 と SDCG@8 においては少し高い精度を示す結果となった。ベースライン手法である Bing クエリ推薦手法と提案手法の比較によるそれぞれの有効性の考察については、4.5.1 節にて議論する。

4.5 考察

本節では、提案手法の有効性と推薦されたクエリの妥当性に関して議論をすることで、提案手法の有効性および課題に関する考察を行う。

4.5.1 提案手法の有効性

手法ごとの実際のクエリ変換について、評価クエリ内で提案手法が有効に働いたものの例を表 3、あまり有効に働かなかったものの例を表 4 に示した。

はじめに、Word2Vec 手法に対して提案手法が非常に高い精度を示した要因について述べる。まず、提案手法では、

- (1) 初期クエリに関連した質問とその回答から変換候補を抽出
- (2) 変換候補を Word2Vec による類似度でフィルタリング
- (3) フィルタリングした語を用いたクエリの妥当性評価

という 3 つの処理でクエリ変換を行い、ユーザの意図する文脈に関連した変換候補を取得してからその変換候補を評価するという手法を用いている。一方、Word2Vec 手法では、上記の(2)の処理しか行われておらず、変換対象キーワード以外の初

表 3 提案手法が有効に働いた例

初期クエリ	Bing クエリ推薦	Word2Vec	提案手法
	1. 暖まる	紅葉 真空 暖まる	昼過ぎ 真空 暖まる
	2. 温まる 暖まる	日陰 真空 暖まる	真上 真空 暖まる
日光 真空 暖まる	3. 暖まる 温まる 違い	直射日光 真空 暖まる	地面 真空 暖まる
	4. 体が温まる 暖まる	箱根 真空 暖まる	日差し 真空 暖まる
	5. 体が暖まる	日差し 真空 暖まる	水 真空 暖まる

表 4 提案手法が有効に働かなかった例

初期クエリ	Bing クエリ推薦	Word2Vec	提案手法
	1. 昆虫の森	昆虫 黒体 4 枚	昆虫 ゴンタ 4 枚
	2. 昆虫図鑑	昆虫 常体 4 枚	昆虫 海獣 4 枚
昆虫 翅 4 枚	3. 昆虫標本	昆虫 痕跡 4 枚	昆虫 鬼面 4 枚
	4. 昆虫採集	昆虫 交雑 4 枚	昆虫 触覚 4 枚
	5. 昆虫 イラスト	昆虫 劣性 4 枚	昆虫 水生 4 枚

期クエリに含まれる語を無視している。また、Word2Vec はある語の周辺語の分布を用いて意味を学習するため、最も類似する語を取得するという単純な手法ではユーザの検索意図に沿う単語が得られないと考えられる。たとえば、表 1 に示すように、「日光」という語は「真空でも日光によって暖まる」という科学に関する文脈で使われているが、「日光浴、毎朝日光を浴びよう」という健康に関する文脈などの他の文脈使われることもあり、単純に Word2Vec を適用して得られた語がユーザの意図に沿う変換ができるとは考えにくい。

次に、Bing クエリ推薦手法より提案手法が有効に働いた要因について述べる。上述したように、提案手法ではユーザの検索意図に関連した変換候補を取得して、それを評価することでクエリ推薦をしていた。一方、Bing クエリ推薦手法では、初期クエリが表す意図から大幅にずれてしまうようなクエリが推薦される場合がある。たとえば、初期クエリが「日光 真空 暖まる」のとき、Bing クエリ推薦手法は表 3 に示すようなクエリを推薦し、これは初期クエリが表す検索意図からずれている。これらの要因によって、4 つの評価尺度において提案手法が Bing クエリ推薦手法を上回る結果になったと考えられる。

最後に、Bing クエリ推薦手法に対して提案手法や Word2Vec 手法が Success@8 と Max@8 において高い精度を示した一方で、MRR@8 と SDCG@8 においては提案手法が少し高い精度を示す結果、Word2Vec 手法は低い精度を示す結果となった要因について述べる。Success@8 はと Max@8 は正解クエリの順位を考慮しない評価尺度である。一方、MRR@8 と SDCG@8 は、正解クエリの有無や正解クエリが返す検索結果の精度のみではなく、さらにそれらのランキングまで考慮した評価尺度である。各手法に対し、正解クエリのみに着目してそれぞれ MRR@8 と SDCG@8 を計算した結果を表 5 に示す。表 5 から、正解クエリが得られる場合、3 つの手法の中で、Bing クエリ推薦手法は正解クエリを上位にランキングする精度が最も高いことがわかる。表 2 と表 5 から、提案手法は Bing クエリ推薦手法と比較して、正解クエリを出力する精度は高いが、正解クエリのランキングは出力された中では低い。提案手法におけるクエリ推薦ランキングの精度向上が必要であると考えられる。

表 5 正解クエリのみに着目した評価結果

	Bing クエリ推薦手法	Word2Vec 手法	提案手法
MRR@8	0.658	0.198	0.375
SDCG@8	1.010	0.368	0.770

4.5.2 推薦されたクエリの妥当性

各手法によって変換されて提示されたクエリ推薦の妥当性について考察する。提案手法ではクエリ推薦のランキングするために、Web 検索におけるクエリとしての妥当性を評価する手法を用いた。また、4.4 節の結果から、ベースライン手法より提案手法が「想定意図適合性」という評価方法から有効に働くことを確認した。しかし、その推薦されたクエリ自体をユーザが本当に使うかどうかの妥当性に関して言及していない。例えば、表 3 にあるように、提案手法は「地面 真空 暖まる」というクエリを推薦している。このクエリを提案手法が Web 検索クエリとして妥当であると判断し、また「想定意図適合性」という観点から正解クエリであるが、実際にユーザがこのクエリを選択するかどうかは疑問が残る。一方、Bing クエリ推薦手法では、「想定意図適合性」という観点からは不正解のクエリではあるが、表 3、表 4 とともにユーザが用いるクエリとしては妥当なものを推薦している。このように、提案手法において、ユーザが実際に用いるものとして妥当なクエリを推薦することによる改善も必要であると考えられる。

5. おわりに

本研究では、質問応答コーパスとして Q&A コーパスを位置づけた。この Q&A コーパスにはユーザの検索意図と類似した質問または回答の中に有用な語があると考え、ユーザの意図をより適切に表現した語を発見することで、クエリを生成する手法を提案した。提案手法では、ユーザのクエリに類似する質問文とその回答文を取得し、その中から変換対象キーワードに類似する語を変換候補として抽出した。その語によってユーザのクエリを変換し、その変換されたクエリが Web 検索において妥当なクエリであるかに基づきランキングしたものをクエリ推薦として提示した。10 件のクエリによる評価実験を行った結果、既存の検索エンジンのクエリ推薦と比較して、提案手法は非常に高い Success@8 と Max@8、より高い MRR@8 と SDCG@8 を達成することが分かった。また、実験結果を分析することで、既存の検索エンジンのクエリ推薦より精度が高い一方で、提案手法のクエリ推薦ランキング手法に課題があることを確認した。

今後は、クエリ推薦のランキング手法およびクエリ推薦がユーザに使われる妥当性を改善・向上させることに取り組む予定である。提案手法の 1 つの問題点は、正解クエリを上位にランキングすることができていないということである。正解クエリを出力に含む精度は高い一方で、それを上位にランキングできていないため、精度が落ちている。したがって、提案手法のランキング手法を海戦することが必要である。また、2 つ目の問題点は、推薦されたクエリがユーザに使われるクエリとして妥当であるかということである。本研究では言及しなかったが、推薦されたクエリが実際のユーザが選択するものであるかどう

かは、将来的に検索エンジンのクエリ推薦で用いる手法とするためには必要であると考えられる。今後は、実際のクエリログを分析することで、ユーザに使われる妥当性という観点からクエリ推薦の精度を評価し、精度を正確に計測するためにより多様なクエリで提案手法の有効性を検証したい。

謝辞 本研究の一部は、JSPS 科研費 15H01718, 16H02906, 16K16156 によるものです。ここに記して謝意を表します。

文 献

- [1] P. Boldi, F. Bonchi, C. Castillo, and S. Vigna. From “dango” to “japanese cakes”: Query reformulation models and patterns. In *Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*, pages 183–190, 2009.
- [2] D. Bollegala, Y. Matsuo, and M. Ishizuka. Measuring semantic similarity between words using web search engines. In *Proceedings of the 16th International World Wide Web Conference*, pages 757–766, 2007.
- [3] Z. Guangyou, C. Li, Z. Jun, and L. Kang. Phrase-based translation model for question retrieval in community question answer archives. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 653–662, 2011.
- [4] C. Huanhuan, J. Daxin, P. Jian, H. Qi, L. Zhen, C. Enhong, and L. Hang. Contextaware query suggestion by mining click-through and session data. In *the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 875–883, 2008.
- [5] R. Kraft and J. Zien. Mining anchor text for query refinement. In *Proceedings of the 13th international conference on World Wide Web*, pages 666–674, 2004.
- [6] H. Ma, M. R. Lyu, and I. King. Diversifying query suggestion results. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 1399–1404, 2010.
- [7] R. Mandala, T. Tokunaga, and H. Tanaka. Combining multiple evidence from different types of thesaurus for query expansion. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 191–197, 1999.
- [8] B. Y. Ricardo, H. Carlos, and M. Marcelo. Query recommendation using query logs in search engines. *Current Trends in Database Technology-EDBT 2004 Workshops*, pages 588–596, 2005.
- [9] J. Rosie, R. Benjamin, M. Omid, and G. Wiley. Generating Query Substitutions. In *Proceedings of the 15th International Conference on World Wide Web*, pages 387–396, 2006.
- [10] K. Tomoyuki and K. Mamoru. Building a monolingual parallel corpus for text simplification using sentence similarity based on alignment between word embeddings. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1147–1158, 2016.
- [11] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 4–11, 1996.
- [12] X. Xue, J. Jeon, and W. B. Croft. Retrieval models for question and answer archives. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 475–482, 2008.
- [13] 山口雅史, 大島裕明, 小山聡, and 田中克己. サーチエンジンのクエリログを利用した同位語の発見. *日本データベース学会 Letters*, 5(2):17–20, 2006.