

遺伝的アルゴリズムを用いた文脈処理による質問応答処理手法の提案

木村 泰知†

荒木 健治‡

† 小樽商科大学 ‡ 北海道大学大学院情報科学研究科

1 はじめに

質問応答は、自然言語の質問に対して、適切な範囲で応答することを目的としており、TREC QA Task 及び、NTCIR の QAC などにおいて、研究が活発に行われている[1][2][3]。最近では、文脈処理を必要とする、シリーズ型の質問が注目されている。つまり、照応解析などを行う必要があり、対話処理を想定とした質問となっている[3]。

質問応答では、Factoid と呼ばれる事実に基づいた質問を対象としており、人名、組織名、日付などを応答するものである。Why 質問、How 質問を対象としていないため、簡単な問題設定と指摘されることもあるが、基礎となる研究であり、精度向上が望まれている。さらに、文脈処理を必要とするシリーズ型の質問は、一問一答型の質問とは異なり、代名詞、省略の補完が必要となる困難なタスクである。そのため、自然言語処理技術の応用が期待されている。

そこで、従来研究の文脈処理技術を質問応答に適用し、代名詞及び、省略単語の補完を行う手法が福本らにより提案されている[1]。省略された単語補完した質問文で、一問一答の質問応答の処理を行う。しかしながら、直前の応答を利用する質問も存在するため、正解を持っていないければ、完全に補完できない場面も存在する。また、村田らの研究[2]では、特に補完することを行わず、今までの質問文からクエリを生成し、適切なスコア計算により応答を選択し、高い精度を達成している。しかし、それまでに入力された質問文からクエリを生成するだけでは、正解を探せない下記のような場合も存在する。

- 正しく補完しても正しい検索が行われない場合
- 質問文の情報だけでは、適切なクエリ生成をできない場合

これらの問題を解決するために、我々は文脈処理を必要とする質問に対して、「文脈を必要としない質問文において検索された文書の回答候補付近に存在する単語は、次の質問のクエリを構成する単語となる可能性が高い」と仮定する。

文脈処理に利用するキーワードの獲得の流れを

下記に示す。

- 文脈処理の必要ない第一の質問文からクエリ生成する。
- 検索された文書候補から入力文を構成する名詞を含んだ文を抽出する。
- それらの文に含まれる名詞を抽出した「キーワード候補」とする。

これらの処理は、2 次キーワードを利用した文書検索と同様に、関連したキーワードの拡張により、クエリを拡張する。しかしながら、キーワード候補が多くなるため、最適なキーワードの組合せを高速に行う必要がある。そこで、我々は、最適な組合せを決定する機能を有する遺伝的アルゴリズム[6] を用いた文脈処理による質問応答処理手法を提案する。本稿では、遺伝的アルゴリズム (GA) を適用する方法とその性能評価実験の結果について述べる。

2 基本的な考え方

評価関数を実現する必要があることから、言語処理に GA を適用することは困難とされている。たとえば、GA が得意とする「ナップザック問題」において、評価関数はナップザックに入れる物の値を計算するが、言語処理では正解を評価することが困難である。そこで、本研究では、単語、文、文脈、応答候補を考慮することで、正解を含む可能性を評価する。もし、正解を判断できるのであれば、GA を適用する必要がなく、評価関数だけを利用すれば良いことになる。しかしながら、本研究で扱う対象は、質問に対する正しい応答を GA によって求めるものではなく、GA を適用する目的はクエリの生成であり、キーワードの最適な組合せをみつけることである。一問一答のように、質問文を構成する、限られた単語から最適な組合せをみつける場合には GA の必要性は少ないが、関連性が高い何十～何百の単語から最適なキーワードの組合せを求める問題設定にした場合、GA の適用は最適であると考えられる。

3 処理過程

3.1 概要

処理の概要を図 1 に示す。

- 入力文処理

- 入力文から質問タイプを決定する.
- クエリ生成処理
 - シリーズ型の第一質問の場合，質問文に含まれる名詞を組合せた，クエリパターンを生成する．第二質問以降，遺伝的アルゴリズムを用いることで最適な組合せを生成する.
- 文書検索処理
 - クエリを構成する単語の IDF を合計し，合計の大きいクエリから検索する.
 - 検索エンジンは Namazu である [5].
- 応答抽出処理
 - 質問タイプを利用して，検索された文書候補から固有表現 (Named Entity) の抽出を行う．本研究では NExT を用いた [4].

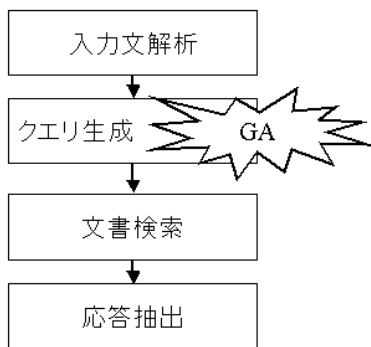


図 1 処理概要

4 GA の適用方法

従来の文脈処理を行う質問応答では，補完することを目的としているが，補完した場合にも，正解が得られるとは限らない．つまり，人間により適切と考えられるクエリを作成しても，文書候補が存在しないことがある．また，文書候補が存在したとしても正しい文書候補を検索できない場合がある．

そのため，関連性のあるキーワードからクエリ生成を試みるのが有効と考えられる．つまり，クエリを生成するキーワード候補は存在するが，最適な組合せを生成することが難しい状況であるため，GA を利用することが適している．GA で求める解は，最適なクエリである．最適な解は，質問文と類似性の高い文の周辺に存在すると仮定して，クエリを生成する．

GA の処理過程は一般的に，以下のような処理から構成される．

- 初期集団の生成
 - 適応度の評価
 - 選択
 - 交叉
 - 突然変異
- 世代

次に，これらの処理を文脈処理に適用する方法について述べる．

4. 1 初期集団の生成

質問文に含まれる単語，関連ある直前の質問に含まれる単語に加えて，前回の入力文と類似性の高い文に含まれる単語を初期集団のクエリを構成する候補とする．クエリは単語候補からランダムに組み合わせて生成する．クエリを構成するキーワードの数は，2~3 単語とする．これは，NTCIR の QAC3 の質問 360 問 [7] に対して，適切と考えられるクエリを人手によって生成する予備実験をした結果，平均 3.2 単語で構成されていたためである．

初期集団の生成手順を下記に示す．

- 文脈処理を必要としない質問文を対象とする．
- 質問文からクエリを生成する．
- 検索された文書候補から質問文の単語を含む文を抽出する．
- 抽出された文に含まれる単語を初期集団のクエリを生成する単語候補とする．

初期集団の生成する例を図 2 に示す．初期集団のキーワードには，正解が含まれていることも少なくない．さらに，文書検索が成功したクエリに正解が含まれていることが多いため，固有表現抽出が失敗した場合には，文書検索が成功したクエリから，入力文を構成する単語を除いた単語を解答候補とする．

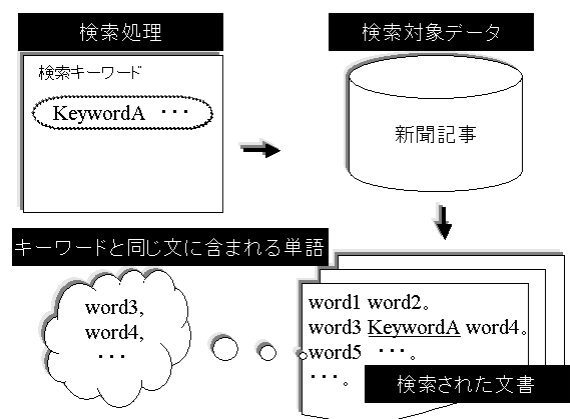


図 2 初期集団生成の例

4. 2 適応度の評価

適応度は，一般にその個体が問題の環境にどの程度適応しているかを測る尺度であり，解の評価の高さを示す．適応度の評価は GA の中枢部分といえるが，言語処理としては設定が困難である．

つまり、一般に評価関数は正解の評価であるため、評価可能ならば、GAを用いる必要がないことになる。しかしながら、今回の適用箇所はクエリ生成であり、GAによる応答抽出や応答生成ではない。適切なクエリは、検索結果により評価される。もし、正しいクエリを探すならば、キーワードの組合せをすべて試す必要があり、時間的に困難である。文脈処理の場合、適切なクエリを生成するためのキーワードが、関連ある第一番目の質問の回答付近に存在する可能性が高いと考えられる。これは、一問一答の組合せよりも、多くの組合せを考える必要があり、組合せの問題と考えられる。そのため、文脈を用いた質問応答では、一問一答の質問と異なり、GAを利用する必要がある。

評価関数に利用する基準を下記に示す。

- キーワード
- 関連度
- 検索の可能性

ランダムにキーワードが2~3個の組合せを生成する。生成されたキーワードの組合せから、IDFの合計に基づいて降順に検索を行い、文書検索が可能な10個を抽出する。10個の各クエリで検索された最高30の文書から、キーワードが含まれる文を抽出する。キーワードの種類は以下のような3つが考えられる。

- 文脈キーワード
 - 直前の関連ある質問に含まれる単語
- 質問キーワード
 - 今回の質問に含まれる単語
- クエリ構成キーワード
 - クエリを構成している単語

これらのキーワードを含む文を文書候補から抽出し、計算することにより、質問文との関連性が高い文を抽出する。式(1)に関連性の評価式を示す。式(2)は文脈キーワード、式(3)は質問キーワード、式(4)はクエリ構成キーワードに関する評価式である。

$$\text{関連性} = \text{Con} + \text{Inp} + 0.5 \cdot \text{Qry} \cdots (1)$$

$$\text{Con} = \frac{\text{一致した文脈キーワードのIDF合計}}{\text{文脈キーワード数}} \cdots (2)$$

$$\text{Inp} = \frac{\text{一致した質問キーワードのIDF合計}}{\text{質問キーワード数}} \cdots (3)$$

$$\text{Qry} = \text{一致したクエリ構成キーワードのIDF合計} \quad (4)$$

4.3 選択

選択は、集団の中から優秀なものを選び出すことである。本手法における、集団とはクエリ集合であり、1つのクエリが個体となる。優秀な個体を決定するために、評価関数があり、その値に基づいて、選択される。選択の絶対条件として、文書検索が行えたクエリを利用することとした。文書検索が成功しない場合、応答候補抽出ができないため、文書検索が可能という要素は大きい。検索可能なクエリで、IDFの大きい順に成功したクエリを10個選択する。

4.4 交叉

交叉は2つの個体から、新しい個体を作り出す処理である。本手法では、適応度の高い個体(クエリ)から2つのクエリを選び、親クエリとする。親クエリのペアから一様交叉を行う。一様交叉は、交叉する部分をランダムに決定する交叉であり、新しい子クエリが2つ作られる。子クエリが今まで試していないキーワードの組合せであれば、そのクエリを利用して文書検索を行う。

4.5 突然変異

突然変異は、個体をランダムに改変するものであり、別の個体に変えてしまうことである。初期集団に発生している遺伝子(キーワード)の中で、現在までに利用されていないキーワードをランダムに交換する。ランダムに変換する割合である突然変異率は2%とする。

4.6 世代

世代は、交叉・突然変異・選択の1サイクルのことであり、1世代毎に各個体を進化させることによって最適な解を見つける。世代が進むにつれて個体もより進化するが、世代数が増加すると応答処理時間が長くなるため、GAの利点である短時間で最適な解が求められなくなる。そこで、本手法の世代数は3世代とした。

5 性能評価実験

5.1 実験の目的

本実験では遺伝的アルゴリズム用いた文脈処理に対する性能評価実験を行う。

5.2 実験方法

質問データにNTCIR5-QAC3の質問セット360問[7]を利用する。本実験ではQAC3の質問セットを利用する。たとえば、次のような質問が存在する。

- 「日光東照宮はいつ創建されましたか。」
- 「誰が造営しましたか。」
- 「誰を祭っていますか。」

この質問セットは50シリーズからなり、情報収

集するための質問である Gathering の質問が 35 シリーズ，意の赴くままに質問するブラウジングを想定した質問が 15 シリーズの質問セットである [3]．検索する対象データは，4 年分の新聞記事データ（2000 年，2001 年の毎日新聞，読売新聞）である．

5. 3 評価方法

文書検索の評価方法の 1 つである MRR(Mean Reciprocal Rank)を用いて評価を行う．MRR は，質問応答で利用されている評価方法であり，最初に出現した正解の順位の逆数を求め，それらを全文にわたって平均する．MRR を式(4)に示す．正解の順位の逆数を求める RR_i を式(5)に示す．

$$MRR = \frac{\sum_i^n RR_i}{n} \dots(4)$$

$$RR_i = \frac{1}{Rank_i} \dots(5)$$

5. 4 実験結果

MRR は 20 位までを対象として，20 位以内に含まれなかった場合，21 番として計算を行った．その結果を表 1 に示す．

表 1.MRR の結果

RRi	33.20
MRR	0.092

質問 360 問中，上位 20 位以内に正解が存在する応答が 46 応答存在した．46 応答中，1 位から 5 位までに存在した正解応答数の分布を表 2 に示す．

表 2. 5 位以内の正解応答数

順位	応答数
1	10
2	5
3	5
4	5
5	6

5. 5 考察

表 2 の 46 応答中 31 応答(67%=31/46)が上位 5 位までに含まれていた．この基準を広げても，ほとんど正解が増加しないことがわかる．これは，正解が候補上位に計算されなかったのではなく，候補としても選択されていないことがわかる．この原因として，固有表現抽出部分における未抽出などがあげられる．

下記に，文脈処理を必要とする質問における正応答の例を示す．

Q1 「日光東照宮はいつ創建されましたか」

■ 「1617年」・・・第1候補

Q2 「誰が造営しましたか」

■ 「秀忠」・・・第4候補

Q3 「誰を祭っていますか」

■ 「家康」・・・存在するが，21位以下

上記のように，応答可能な場合も存在するが Q3 の「誰を祭っていますか。」のように，質問文中にキーワードとなる名詞を含まない質問が存在する．同様な質問に，「どこで起きましたか」がある．これらの質問からさらに多くの情報を抽出するために，動詞から名詞へ変換したキーワードを作成することも必要である．たとえば，「起きる」を「発生」とすれば，検索可能な場面も存在した．

6 まとめ

本稿では，文脈処理を必要とする関連ある連続の質問に対して，遺伝的アルゴリズムの適用について述べた．

質問が連続される場合，省略される単語は，第一番目の質問の応答候補付近に存在すると仮定して，GA により最適なキーワードの組合せを求めた．NTCIR 5 - QAC 3 の質問応答セットを利用した結果，MRR は 0.092 であった．照応解析が必要である質問に対しても応答可能であることを示した．

今後の課題として，固有表現抽出の精度向上がある．また，キーワードとなる名詞が存在しない場合の対処及び，質問文と応答候補文の類似性を考慮することも追加する予定である．

文 献

- [1] Megumi Matsuda and Jun'ichi Fukumoto, "Answering Questions of IAD Task using Reference Resolution of Follow-up Questions," NTCIR-5 QAC3. In Proceedings of the Fifth NTCIR Workshop Meeting, pp.414-421, Tokyo, Japan, Dec 2005.
- [2] Masaki Murata, Masao Utiyama and Hitoshi Isahara, "Japanese Question-Answering System using Decreased Adding with Multiple Answers at NTCIR 5," NTCIR-5 QAC3. In Proceedings of the Fifth NTCIR Workshop Meeting, pp.380-385, Tokyo, Japan, Dec 2005.
- [3] 加藤恒昭，福本淳一，榊井文人，神門典子，"質問応答技術は情報アクセス対話を実現できるか，" 情報処理学会研究報告，2004-NL-162，pp.145-150，2004.
- [4] 渡邊一郎，榊井文人，福本淳一: "固有表現抽出ツール N E x T の精緻化とユーザビリティの向上"，言処理学会第 10 回年次大会発表論文, pp.413-415 ,2004.
- [5] Namazu <http://www.namazu.org/>
- [6] D.E.Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning.," Addison-Wesley, 1989
- [7] NTCIR <http://research.nii.ac.jp/ntcir/index-ja.html>