

# 事物間の多様な状態に基づくファクトイド型質問応答

田上 諒† 木村 輔† 宮森 恒†

† 京都産業大学大学院先端情報学研究科 〒 603-8555 京都府京都市北区上賀茂本山

E-mail: †{i1788124,i1658047,miya}@cc.kyoto-su.ac.jp

**あらまし** 本稿では、事物間の多様な状態を表すグラフを用いたファクトイド型質問応答のための柔軟な自動回答手法を提案する。ファクトイド型質問応答は、これまで広く研究されてきており、多くの従来研究では、文書検索に基づく手法や、ナレッジベースに基づく手法が採用されている。しかし、それぞれの手法には問題点もあり、多様な質問への対応が難しい。本稿では、これら2つのアプローチの利点と欠点をふまえ、事物間の様々な状態を表現可能なグラフを導入し、これにより多様な質問に対して事物間の関係を考慮した柔軟な自動回答が実現されることを目指す。実験では、大学入試二次試験の世界史語句解答問題を対象として、提案手法が正答率にどのような影響を及ぼすかについて明らかにする。

**キーワード** 自然言語処理, ファクトイド型質問応答, 自動回答, 事物間状態グラフ

## 1. はじめに

近年、ユーザからの多様な情報要求を満たす技術として、質問応答などの自動回答技術が注目されている。大量に存在する情報源の中からユーザが必要な情報を得る手段としては、関連するキーワードをクエリとして文書検索を行い、検索結果となる複数文書から得たい情報を探し出す方法が一般的である。しかし、この方法は、クエリ生成の過程や、複数文書の中から要求を満たす情報を選択する過程を、ユーザ自身で行わなければならない。それに対し質問応答は、ユーザ自身の情報要求を自然言語で入力し、情報源から1つの回答を出力する技術である。ユーザの得たい情報を身近な言語で伝えることができ、かつ、複数の情報を比較する必要がない点が特徴といえる。

質問応答で扱う質問は、人名や地域名など、短い語句による事実を回答すれば良いファクトイド型と、ある語句の定義や手順などを説明する必要があるノンファクトイド型に大別することができる。また、システムに対しては、特定の分野に関するクローズドドメインな質問を対象とするか、分野を限定しないオープンドメインな質問を対象とするかでも分類される。

ファクトイド型質問応答に関しては広く研究されてきており、様々な手法が提案されているが、それらは基本的に図1に示す2つのアプローチのいずれかに分類される。特に、知識源に対する文書検索に基づいた手法は、多くの従来研究で採用されている。しかし、この手法は、文書内に存在する複数の事物間の関係を捉えることが容易ではなく、同種の複数回答候補を混同して誤答してしまう場合があることが問題点として挙げられる。また近年では、オープンドメインなオントロジーデータベースの出現に伴い、事物の関係をより明確にしたナレッジベースを用いた手法も注目されているが、文書に比べ表現力が低く、多様な質問への対応が難しい。

クローズドドメインを対象とする質問応答は、オープンドメインのものと比較すると、質問対象が特定分野に限定されるため、より扱いやすいという側面はあるものの、現実世界で想定

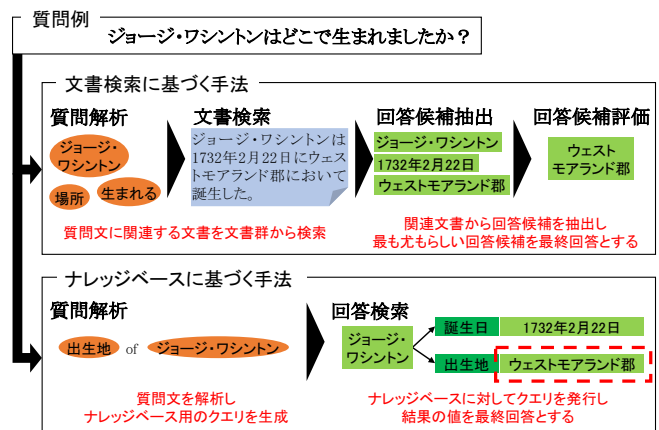


図1 ファクトイド型質問応答のアプローチ

されるより多様な質問に対して適切に回答するのは必ずしも容易ではない。なぜなら、多様な質問方法に対して適切に回答を導くための十分な知識源を用意したうえで、質問内容をより正確に解析し、回答を探し出す必要があるためである。実際、オープンドメインのシステムで用いられる手法を、そのままクローズドドメインのシステムに適用しても適切な回答は導くのは難しい。評価型ワークショップである NTCIR<sup>(注1)</sup> で実施されている QA Lab タスク [1] [2] [3] では、大学入試二次試験の世界史問題という特定ドメインを対象として、より多様な質問に対する適切な回答が可能な質問応答技術の確立を目指しているが、記述式の語句解答問題に分類されるファクトイド型問題の正答率は十分とはいえない。

本稿では、事物間の様々な状態を表現可能なグラフをナレッジベースとした、ファクトイド型質問応答のための柔軟な自動回答手法を提案する。具体的には、先に述べた2つの手法の利点および欠点をふまえ、事物間の状態の関係を表すグラフ構造のナレッジベースを構築する。また、回答する際には、このナ

(注1) : NTCIR: <http://research.nii.ac.jp/ntcir/>

レッジベースを用いて、質問の意図する回答を導出する。

実験では、世界史の語句解答問題を対象として、グラフにおける事物間の関係性、および自動回答の正答率を調査し、提案手法の有用性を確認する。

## 2. 関連研究

ファクトイド型質問応答システムに関する研究は、これまでに数多く行われている。それらは、用いる手法によって、非構造化データである文書への検索に基づくシステムと、構造化されたデータであるナレッジベースに基づくシステムの、2種類に分類される[4]。

Kupiec [5] は、1993年に文書検索に基づく手法を発表し、後の質問応答システムの研究に大きな影響を与えている。基本的な処理手順として、まず、質問の内容を解析した後、その質問に関連した文書が知識源より検索される。次に、得られた文書から、回答候補となる単語群が抽出され、先の質問解析の結果等に基づいて、最も尤もらしい候補が最終的な回答として選択される。ただし、知識源に含まれる文書は、原則として非構造化データであるため、事物間の関係が必ずしも明確に記述されているわけではない。もし、文書内に同種の回答候補が複数存在した場合、質問内容とそれぞれとの関係性を把握できなければ、誤答へとつながることも考えられる。よって、適切な回答を常に選択することは容易ではない。

一方、ナレッジベースに基づく手法も歴史は深く、1961年には、Greenら[6]が、野球の試合結果などのデータを、独自のスキーマを用いて格納することで、自然言語の質問に対応するBASEBALLというシステムを開発している。近年では、Wikidata [7]などのオープンドメインなオントロジーデータベースの出現により、それらを質問応答のナレッジベースとする手法が目ざされている。自然言語で記述された非構造化データである文書集合とは異なり、事物間の関係をエッジで結ぶグラフで記述された構造化データであるため、ユーザの質問を適切なクエリに変換することができれば、直接的に正しい回答に到達することができる。しかし、自然言語で書かれた文書と比較すると、データの格納形式が厳密であるナレッジベースは、表現できる内容に制限があるため、多様な質問への対応が難しいという問題がある。

倉田ら[8]は、文書検索とナレッジベースを組み合わせたハイブリッド型のアプローチを導入し、質問に関連する文書を検索によって得た後、文書の係り受け構造を解析してグラフを生成し、回答を選択する手法を提案している。この手法により、単語間の関係がより明確化され、単純な文書検索に基づく手法と比べ、適切な回答が選択されやすいと報告されている。しかし、事物間の関係を直接示すグラフ構造では必ずしもないため、文書長が長くなるにつれて、質問に関連した事物関係を捉えることが困難となる。

近年では、クローズドドメインの対象に対し、多種多様な質問に対する適切な応答を目指す研究も活発である。NTCIR QA Labタスクにおいて、Sakamotoら[9]は、大学入試の記述式語句解答問題に分類されるファクトイド型問題の自動回答にお

いて、それらの問題に取り組んだチームのうち、最も良い成績を残している。この手法は、文書検索に基づく手法を採用しており、問題文を解析する際、「人名を聞いているのか」「地名を聞いているのか」などのカテゴリや、「王の名前を答えるべきか」「大統領の名前を答えるべきか」などの焦点を特定することによって、回答候補を選択している。しかし、正答率自体は、2～4割程度に留まっており、十分な成績を残していると言えない。

本稿では、教科書等の非構造化文書から事物間の様々な状態を表現するグラフを構築し活用するファクトイド型自動回答手法を提案する。本手法は、豊かな内容を含む非構造化文書の利点と、事物間の関係を端的に扱えるナレッジベースの手法の利点を共に活かすことを目指したものである。本手法により、クローズドドメインを対象としたシステムにおいても、より適当な知識源が確保され、多種多様な質問により的確に対応できるようになることが期待される。

## 3. 提案手法

本節では、提案するファクトイド型質問応答手法について述べる。ファクトイド型質問応答では、十分な知識源と正確な質問解析が重要である。3.1節では、知識源となる事物間状態グラフについて述べ、3.2節では、質問解析を含めた自動回答手法について述べる。

### 3.1 事物間状態グラフ

提案手法では、知識源として、事物間の状態の関係性を表現するグラフ構造のナレッジベースを構築する。本稿では、この知識源に格納されるグラフを、**事物間状態グラフ**と呼ぶ。このグラフは、非構造化文書から生成され、構造化されたデータベースに格納されるため、非構造化文書の知識源と比べ、より事物間の関係を直接的に捉えやすい知識源として活用が期待できる。

#### 3.1.1 構造

事物間状態グラフは、人名や地名、出来事名などの事物を表すノードと、あるノードから別ノードへの動作を表すエッジから構成される。

#### 3.1.2 構築手法

事物間状態グラフの構築手順を図2に示す。教科書等の知識源に含まれるテキストをあらかじめ句点で分割し、得られた各文に対し、図2の処理を実行する。

- (1) 不要な空白や記号等を除去する。
- (2) 形態素解析及び構文解析を行い、1文を基本句<sup>(注2)</sup>に分割する。各基本句は、原則として用言または体言に分類される。
- (3) 各基本句内に存在する形態素のうち、副詞、判定詞、助動詞、接続詞、指示詞、助詞、接尾辞、特殊、未定義語のいずれかに該当するものを除去する。
- (4) 体言として判定された基本句を、ノードとして生成す

(注2)：本稿では、自立語となる1つの形態素と、それに続く付属語となる形態素で構成される表現要素を、基本句と定義する。



図 2 事物間状態グラフの構築処理

る。当該基本句が、固有表現だった場合は、同時にカテゴリ情報も付与する。カテゴリは、形態素解析エンジンの解析結果をもとに判定する。

(5) 構文解析の結果を順次走査し、(5-A)～(5-C)のいずれかに該当する部分が存在した場合、当該処理を実行する。

(5-A) 基本句間の係り受け関係のうち、体言から体言に係る関係が存在した場合、係り元の体言ノードから係り先の体言ノードに向け、有向エッジを生成する。

(5-B) 用言として判定された基本句の自立語が形容詞であり、体言の基本句に係る場合、係り先の体言ノードの属性値として、係り元の基本句を追加する。

(5-C) 用言として判定された基本句の自立語が動詞である場合、この基本句と格関係にある体言<sup>(注3)</sup>に該当するノード間に、一定の規則に基づき有向エッジを生成する。

(6) 次数0のノードを消去する。

なお、(5-C)における、どの向きにエッジを生成するかを定める規則としては、例えば、AからBへエッジを張ることを $A < B$ で表すとき、{ガ格 < フ格, ガ格 < ニ格, フ格 < ニ格}として与えることができる。

各文で生成されたグラフは、同一のデータベースに格納されるため、原則として、同一名のノードはマージされる。実際に、事物間状態グラフを自動回答の知識源として用いる際は、あら

かじめ用意した非構造化文書の全文を、1つのグラフに集約した上で、使用することとなる。ただし、「こと」や「もの」など、元の文書に頻出する体言の基本句を、原則に従ってマージしてしまうと、関連のない事物ノード間の距離が、過剰に短くなる可能性がある。よって、グラフ構築処理を開始する前に、元の文書の基本句の出現頻度を求め、一定以上の頻度で出現した基本句については、例外的にマージさせないようにする。

また、事物によっては、「ソ連」と「ソビエト連邦」のように、同義語が存在する場合があるが、それらは同一ノードとしてマージされる。

### 3.2 自動回答手法

自動回答時は、図3に示すように、3つの処理を経て、回答を出力する。

#### 3.2.1 質問解析

質問解析では、まず、質問文が「人物」を聞いているのか「場所」を聞いているのかななどの、回答カテゴリを推測する。予め定めた規則に基づいて、質問文から質問の焦点を特定し、その焦点からカテゴリを判定する。本稿では、Sakamotoら[10]の手法を参考に、全18種類のカテゴリに分類した。なお、この規則でカテゴリが推定できなかった場合、推測カテゴリ不明という扱いで、後続の処理を進める。

次に、質問文を形態素解析し、固有表現を抽出する。

以上で得られたカテゴリ、固有表現は、後に知識源を探索する際に使用する。

#### 3.2.2 関連エッジ・回答候補ノード探索

事物間状態グラフでは、質問文中に含まれる複数の固有表現が、お互いに近い位置関係になっており、かつ、回答もその付近に存在すると考えられる。この処理では、この仮説に基づ

(注3)：本稿では、ある1つの動詞に対して、動作主や場所などを表す表層格にあたる体言を、当該動詞と格関係をもった体言と定義する。日本語では、動作主などを表すガ格や、場所・時間などを表すニ格などの格関係が存在する。

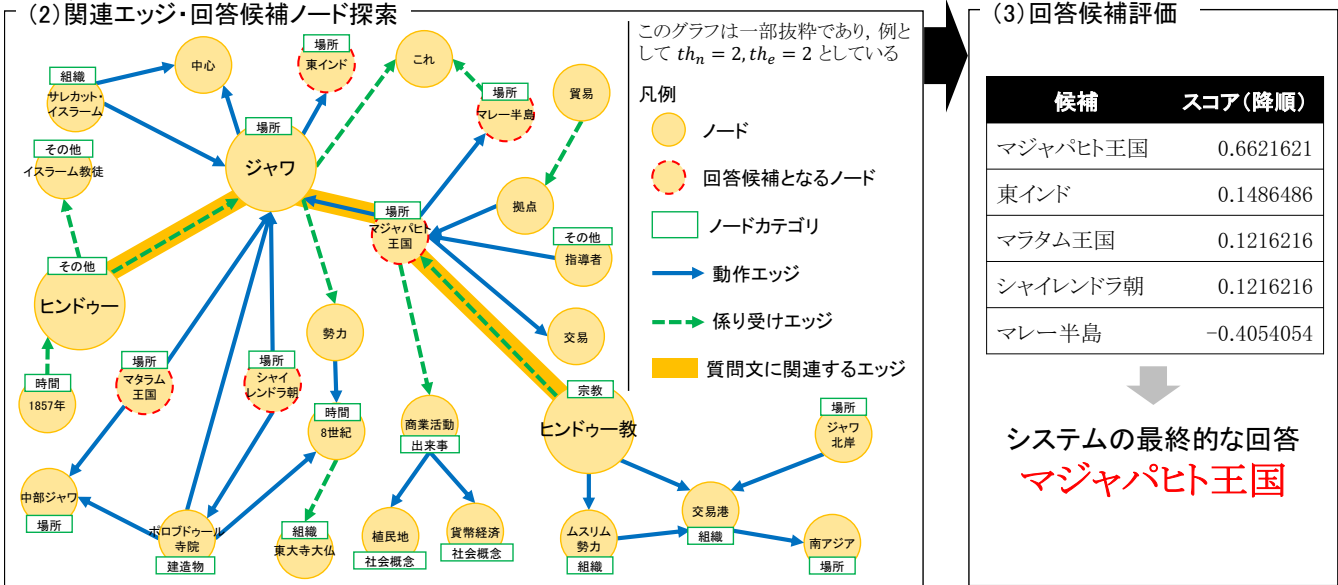
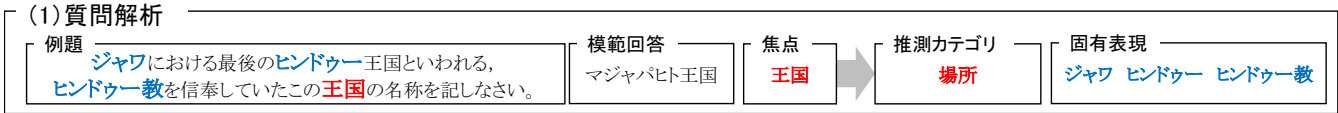


図 3 自動回答処理

いて、質問文に関連するエッジ、及び、回答候補ノードを探索する。

まず、知識源として用意しているグラフから、質問文と関連すると思われるエッジを探索する。質問解析で得た固有表現の集合より、2つの固有表現の組み合わせをすべて抜き出し、それぞれの組み合わせで、グラフ上の、2つの固有表現を表すノードの最小距離を求める。その際、最小距離が  $th_n$  以下であった場合、該当する経路上に存在するエッジを、質問文に関連するエッジとして記録しておく。最終的に、すべての組み合わせで、1回でも記録されたエッジを、すべて質問文と関連するエッジとする。なお、同じエッジが、違う固有表現の組み合わせで複数記録された場合、より質問文との関連度が高いエッジとして記録し、最終的に、各エッジの関連度を  $0 \sim 1$  のエッジ関連度スコア  $S_e$  として算出しておく。

次に、関連するエッジの周辺から、回答候補となるノードを収集し、回答候補集合を得る。エッジごとに、以下の処理を実施し、ノードを探索する。

- (1) エッジの両端に存在するノードを、エッジからの距離が1であるノードとみなし、エッジからの距離  $d_e$  が  $th_e$  以内であるノードをすべて収集し、一時的候補集合に格納する。
- (2) 一時的候補集合に存在するノードのうち、すでに質問文中に存在する固有表現を表すノードを除外する。
- (3) 一時的候補集合に存在するノードのうち、質問解析時に推測されたカテゴリと違うカテゴリであるノードを除外する。ただし、推測カテゴリが不明の場合、この処理は行わない。
- (4) この時点で、一時的候補集合に存在するノードをすべ

て、回答候補集合に追加する。追加時には、ノードごとに以下の情報も併せて保持しておく。なお、すでに回答候補集合に存在するノードが、再度追加される場合もある。

- 当該ノードが収集される要因となったエッジのエッジ関連度スコア  $S_e$  (値が大きいほど回答らしいとみなす)
- 当該ノードが収集される要因となったエッジとの距離スコア  $S_d$  (ただし  $S_d = \frac{d_e}{th_e}$  とし、値が低いほど回答らしいとみなす)
- 同じノードが回答候補集合に何度追加されたかを示すノード選出スコア  $S_c$  (値が高いほど回答らしいとみなす)

すでに回答候補集合に存在するノードを、再度追加する時、 $S_e$  および  $S_d$  の値がより回答らしいものであった場合、その値に上書きする。

ノードの収集が完了した後、回答候補集合内に存在するすべてのノードの  $S_c$  の値を、もっとも追加された回数の多いノードの値が1となるように正規化する。

3.2.3 回答候補評価

この処理では、先の処理で得られた回答候補集合から、最終的な回答を出力する。この時点で、回答候補集合に存在する各解答候補  $w$  に付与されているスコア  $a, b$  および  $c$  は、すべて  $0 \sim 1$  の値に正規化されている。よって、 $w$  ごとに、評価スコア  $Score(w)$  を式 (1) の通り計算する。

$$Score(w) = S_e - S_d + S_c \tag{1}$$

すべての  $w$  について計算完了後、最も  $Score(w)$  の値が高かった  $w$  を、最終的なシステムの回答として出力する。

## 4. 実 験

質問対象を、大学入試二次試験の世界史語句解答問題として、2つの実験を行う。

### 4.1 質問のデータセット

質問のデータセットとして、以下に示す大学入試問題を使用する。これらのデータは、NTCIR-13 QA Lab-3 タスクで提供された世界史問題のうち、記述式の語句解答問題となる全 128 問である。

- 2000 年度 東京大学 (全 23 問)
- 2001 年度 東京大学 (全 15 問)
- 2002 年度 東京大学 (全 11 問)
- 2004 年度 東京大学 (全 11 問)
- 2006 年度 東京大学 (全 9 問)
- 2008 年度 東京大学 (全 11 問)
- 2010 年度 東京大学 (全 15 問)
- 2012 年度 東京大学 (全 13 問)
- 2013 年度 東京大学 (全 9 問)
- 2014 年度 東京大学 (全 11 問)

なお、タスクで提供された語句解答問題 (全 145 問) のうち、以下に該当する問題は除外している。

- 穴埋め型問題 (全 6 問)
- 入試問題特有の出題形式であり、問題文脈部だけでは回答出来かねると我々が判断した問題 (全 11 問)

世界史の語句解答問題は、大学や出題年度によって多少異なるものの、基本的に図 4 に示すとおり、問題指示部と問題文脈部から構成される。

#### 問題指示部

書物の文化は、製作方法の改良や識字率の高まりなどの影響で、時代とともに大きく変化してきた。このような書物の文化の歴史に関連する以下の設問(1)~(10)に答えよ。解答は、設問ごとに行を改め、冒頭に(1)~(10)の番号を付して記せ。

#### 問題文脈部

宗教の発展には典籍による教義の研究が大きな役割を果たしている。仏典を求めて、インドへ渡った玄奘らがその目的を果たしたのは、教義を研究する僧院がそこにあったからである。この僧院の名称を記せ。

図 4 大学入試二次試験の世界史語句解答問題の例  
(2004 年度 東京大学)

問題指示部は、回答方法を指示する 1 つ以上の文からなる。問題文脈部は、実際に回答させたい質問を述べた 1 つ以上の文からなる。問題によっては、1 つの問題文脈部で 2 つの質問を同時に出题する場合もあり、そのような問題は 2 問分として換算する。本稿の実験では、データセットに含まれる情報のうち、問題文脈部と模範解答のみを使用する。

### 4.2 ツール

実験で使用するツールは、次のとおりである。

形態素解析エンジンとしては、JUMAN++<sup>(注4)</sup>を使用し、我々が独自に作成した世界史分野に特化した固有名詞を含むユーザ

辞書 [11] を追加した。ただし、単語のカテゴリを判定する際には、同時に Sakamoto ら [10] が構築したユーザ辞書も使用した。構文解析エンジンには、KNP<sup>(注5)</sup>を使用した。

グラフデータベースは Neo4j<sup>(注6)</sup>で構築し、ノード生成時の同義語の処理に関しては、我々が独自に作成した同義語辞書 [11] を用いて対応した。

### 4.3 知識源構築のための文書

実験の際に使用する知識源は、3.1 節で述べた手法で構築された、事物間状態グラフとする。グラフは、以下の教科書の文書をもとに構築した。これらの教科書についても、QA Lab-3 タスクで提供されたデータである。

- 詳説 世界史 (山川出版社, 2008 年)
- 世界史 B (東京書籍, 2007 年)
- 新選世界史 B (東京書籍, 2007 年)
- 世界史 A (東京書籍, 2008 年)

### 4.4 実験 1: 事物間状態グラフにおける事物間距離の特性

#### 4.4.1 目 的

事物間状態グラフにおいて、関連のある 2 つの事物、関連のない 2 つの事物が各々どのような距離に位置するかを調べる。

#### 4.4.2 方 法

まず、準備として、関連のある事物の対の集合、および、関連のない事物の対の集合を用意する。

関連のある事物の対の集合は、データセットより用意する。基本的に、ある質問の模範解答は、質問文中に存在する固有表現と関連があると考えられる。よって、データセットの質問文(問題文脈部)内に存在する固有表現の 1 つと、模範解答を、1 つの対として収集する。4.1 節で用意したデータセットからは、この処理によって、全 577 件の関連のある事物の対を取得した。

関連のない事物の対の集合は、ランダムに生成する。知識源上に存在する事物を表すノード群から、ランダムに 2 ノードを選択し、1 つの対として収集する。この処理によって、全 1 万件の関連のない事物の対を取得した。

以上で得られた全ての対に対して、知識源上での最小距離を求める。最小距離は、当該対の 2 単語を表すそれぞれのノードが、最短でいくつのエッジを経由すれば繋がるかどうかで求めることができる。よって、最小距離が取り得る最小値は 1 となる。また、同じ事物を表すノードが複数存在した場合は、すべての組み合わせで距離を求め、それらのうちの最小値を、当該対の最小距離とする。

ただし、以下の理由で最小距離を求めることが出来ない場合もある。これらに該当した対は、度数分布より除外する。

- 対の一方、または、対の両方の事物を表すノードが、知識源上に存在しない場合
- お互いのノードの間に経路が存在しない場合

関連のある対と、関連のない対で、距離の度数分布の傾向にどのような違いがあるのか比較する。

(注4) : JUMAN++: <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

(注5) : KNP: <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

(注6) : Neo4j: <https://neo4j.com>



### 4.4.3 結果

関連の有無による、事象間距離の相対度数分布を、図5に示す。関連のある事象は、距離4がピークで、全体の8割が距離5以内に収まっているのに対し、関連のない事象は、距離6がピークで、全体の8割が距離8以内に収まっている。また、関連のある事象は、距離1～2での相対度数も高い。

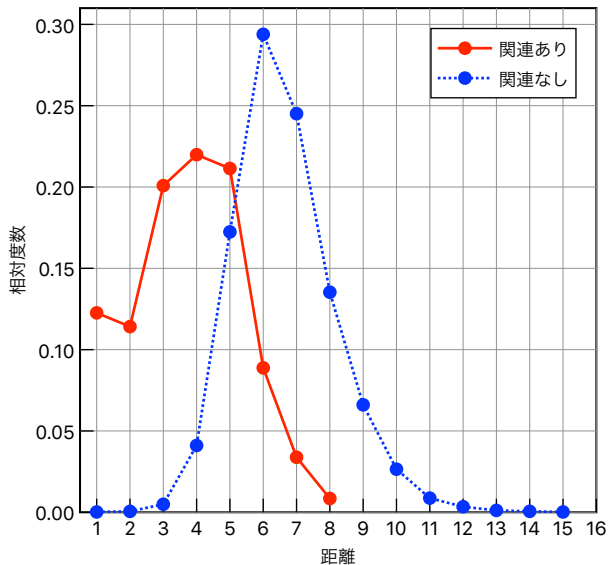


図5 関連の有無による事象間距離の相対度数分布

### 4.5 実験2：提案手法による自動回答の正答率

#### 4.5.1 目的

用意したデータセットに対して、提案した自動回答手法がどの程度正答できるかを調べる。

#### 4.5.2 方法

データセット128問の各々について、問題文脈部を入力として、3.2節の自動回答処理を実行し、正答数および誤答数を調べる。なお、システムが誤答する原因として、以下の状況が考えられるため、それぞれの内訳も併せて求める。

- 正解となるべきノードが知識源上に存在しない場合
- 質問解析時に誤ったカテゴリを推測した場合（推測カテゴリ不明である場合を除く）
- 質問文に関連するエッジが1件も得られなかった場合
- 得られた回答候補集合内に正解となるべきノードが含まれていない場合（1件も回答候補が収集できなかった場合を含む）
- 回答候補集合内の正解となるべきノードの  $Score(w)$  が、集合内で最も高い値ではなかった場合

本実験では、自動解答処理のパラメータの値は、 $th_n = 2, th_e = 2$  とする。

#### 4.5.3 結果

全128問に対する自動回答の結果を表1に示す。ただし、正答数に関しては、 $Score(w_i)$  の値が回答候補集合内で最も高かったものの、他にも同一の値であった別の回答候補  $w_j$  が存在した場合、該当する質問を同率正答と分類し、それ以外の質

問を単独正答と分類している。よって、実質的にシステムが正解できた質問数は、単独正答と分類されたもののみである。

表1 自動回答の結果（カッコ内は該当質問数/全問題数）

大分類	小分類	内訳
正答	単独正答	0.04 (5/128)
	同率正答	0.02 (3/128)
誤答	原因 (a)	0.11 (14/128)
	原因 (b)	0.08 (10/128)
	原因 (c)	0.26 (33/128)
	原因 (d)	0.42 (54/128)
	原因 (e)	0.07 (9/128)

## 5. 考察

4章の各実験について考察する。

### 5.1 考察1：事象間状態グラフにおける事象間距離の特性

実験1では、知識源である事象間状態グラフ上で、世界史分野において関連のある2単語間がどのような位置関係で配置されるかを確認した。

関連の有無による相対度数分布を比較すると、関連のある事象間のほうが、グラフ上でより近い位置関係にあることが分かる。つまり、グラフ上で、関連する事象は互いに近くに密集し、無関係な事象は、それよりも離れた位置に配置される傾向にあるといえる。世界史分野における事象は、主に時代や地域によって分類されることが多いため、同じ時代や国で起きた出来事や活躍した人物などが、グラフ上で互いに近くに配置されていると考えられる。

関連のある事象間の距離の度数は、距離4がピークであるものの、距離3以内に4割以上の度数が分布していることも重要である。これは、自動回答処理において、どの範囲まで探索すれば回答が見つかるかどうかの指針となる。

4.4.2節でも述べたように、本実験では、2つの事象を表すノードが、グラフ上にそれぞれ存在するものの、経路が存在しなかった場合もあり、それらは、図5の結果から除外している。関連のある対の集合、および、関連のない対の集合のそれぞれで、経路が存在しなかった対の件数を調査すると、関連のある対では、全体の6%であったのに対し、関連のない対では、全体の47%に及ぶことが分かった。これにより、関連のない対は、関連のある対に対して、関係が結ばれにくいことがわかる。

### 5.2 考察2：提案手法による自動回答の正答率

実験2では、提案した自動回答手法で、各質問に対してどの程度正答できるかを確認した。実験の結果、現状では9割以上の質問に対して適切に回答できていない。誤答となる各原因について考察する。

まず、原因(a)に分類される質問は、知識源上に正解となるノードが存在しなかったものである。これは、事象間状態グラフを構築する元となったテキストに、正解に関連する情報が無かった場合、または、形態素解析エンジンや構文解析エンジンが固有表現を正しく認識できず、グラフ構築の時点で意図しないノードが生成されてしまった場合などが、原因として考えら

れる。よって、知識源や固有表現辞書の拡充によって、改善が期待できる。

原因 (b) に分類される質問は、誤ったカテゴリが推測されたものである。3.2.1 節で述べているように、現状では、カテゴリ推定はルールベースで行われている。これにより、想定していなかった聞き方や、曖昧な聞き方をするような質問に対して、適切なカテゴリを推定できていない。よって、より精度の高いカテゴリ推定手法を導入するか、カテゴリ推定が不要となるように自動回答手法を改良することが必要であると考えられる。

原因 (c) に分類される質問は、質問文に関連すると推測されるエッジが、知識源上に 1 件も存在しなかったものである。まず、質問解析において、固有表現が 1 件しか取得できなかった場合、または全く取得できなかった場合、必然的に、関連するエッジは 1 件も取得できなくなる。よって、質問文中から取得できる固有表現の件数が限られる場合の代替手法か、グラフ上で質問文に関連するおおよその範囲を推測する別の手段を考える必要がある。また、2 件以上取得できたとしても、 $th_n$  の値が小さいほど、取得できるエッジは少なくなる。ただし、 $th_n$  の値が大きすぎると、質問文と関連のないエッジまで取得される可能性が高くなるため、適切な  $th_n$  の値を見つける必要がある。

原因 (d) に分類される質問は、質問文に関連すると推測されたエッジの周辺に、正解となるノードが存在しなかったものである。つまり、回答候補集合に正解が含まれていなかったことになる。これは、 $th_e$  の値を大きくすることによって、回答候補集合に含まれやすくなると考えられるが、正解である回答候補  $w$  の評価スコア  $Score(w)$  の値は小さくなる可能性が高いため、実質的に改善にはつながらない。また、原因 (e) に分類される質問は、 $Score(w)$  の値が回答候補集合内で最大値でなかったものであり、言い換えれば、正解である  $w$  の  $Score(w)$  の値が小さいものである。よって、質問文が何を聞いているのかをよりの確に解析に、回答候補を絞り込むことができれば、 $th_e$  の値を大きくしたとしても、正解する可能性は高くなると考えられる。

表 2 は、同じデータセットを用いて自動回答を行った従来手法と、提案手法との比較である。ここで、ベースラインとは、我々が NTCIR QA Lab-3 タスク参加時に提案した文書検索に基づく手法 [12] である。また、Sakamoto ら [9] の手法も、文書検索に基づく手法であり、同タスクに取り組んだチーム中、最も良い成績を残している。なお、正答率は、本実験で一部除外した問題を考慮して比較するため、QA Lab-3 タスクで公表された各手法の問題別正答結果に基づき、再計算している。実験で用いたデータセットは、QA Lab-3 タスクの Phase-1 と Phase-2 で配布されたデータセットを統合したものであるため、ベースラインについては、Phase-1 RUN1 と Phase-2 RUN3 の結果を統合し、Sakamoto らの手法については、Phase-1 RUN1 と Phase-2 RUN1 の結果を統合している。表 2 より、提案手法の正答率は、従来手法と比べ著しく低いことがわかる。

以上より、提案した事物間状態グラフの構築手法、および自動回答手法には、まだ課題が多く残されており、今後、有用性を高めるための改善が必要がある。グラフ構築に関しては、形

表 2 従来手法との正答率の比較 (カッコ内は該当質問数/全問題数)

手法	正答率	
ベースライン	0.32	(41/128)
Sakamoto らの手法	0.41	(53/128)
提案手法	0.04	(5/128)

態素解析や構文解析だけでなく、意味役割解析 [13] などを行うことによって、元となる文の深層格を取得することができ、より事物の状態を適切に表したグラフが構築されることが期待できる。また、現状の自動回答手法では、グラフ上の動作エッジに付与された動作の情報を、一切使用できていないため、回答候補を絞り込むための手掛かりとして使用することができれば、正答率が向上する可能性がある。

## 6. ま と め

本稿では、事物間の様々な状態を表現可能なグラフをナレッジベースとする、ファクトイド型質問応答のための柔軟な自動回答手法を提案した。具体的には、非構造化文書をもとに、構造化されたナレッジベースである事物間状態グラフを構築し、それを自動解答に取り入れることで、多様な質問への対応を目指した。

実験では、世界史の語句解答問題を対象として、グラフにおける事物間の位置関係の傾向、および自動回答の正答率を調査した。結果として、構築されたグラフ上では、関連のある事物同士が、関連のない事物よりも、互いの距離が近い関係にあることがわかった。しかし、現時点では、提案した自動回答手法による、正答率の向上は見られなかった。

今後は、グラフ構築手法や自動回答手法の見直しを予定している。

## 文 献

- [1] Hideyuki Shibuki, Kotaro Sakamoto, Yoshinobu Kano, Teruko Mitamura, Madoka Ishioroshi, Kelly Y. Itakura, Di Wang, Tatsunori Mori, and Noriko Kando. Overview of the ntcir-11 qa-lab task. In *Proceedings of the 11th NTCIR Conference*, pp. 518–529, December 2014.
- [2] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, and Noriko Kando. Overview of the ntcir-12 qa lab-2 task. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 392–408, June 2016.
- [3] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, and Noriko Kando. Overview of the ntcir-13 qa lab-3 task. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 112–128, December 2017.
- [4] Poonam Gupta and Vishal Gupta. A survey of text question answering techniques. *International Journal of Computer Applications*, Vol. 53, No. 4, September 2012.
- [5] Julian Kupiec. Murax: a robust linguistic approach for question answering using an on-line encyclopedia. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '93, pp. 181–190, New York, NY, USA,

1993. ACM.

- [6] Bert F. Green, Jr., Alice K. Wolf, Carol Chomsky, and Kenneth Laughery. Baseball: An automatic question-answerer. In *Papers Presented at the May 9-11, 1961, Western Joint IRE-AIEE-ACM Computer Conference*, IRE-AIEE-ACM '61 (Western), pp. 219–224, New York, NY, USA, 1961. ACM.
- [7] Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledgebase. *Commun. ACM*, Vol. 57, No. 10, pp. 78–85, September 2014.
- [8] 倉田岳人, 岡崎直観, 石塚満. 係り受け関係に基づくグラフ構造を用いた質問応答システム. Technical Report 108(2003-NL-158), 東京大学大学院情報理工学系研究科, 東京大学大学院情報理工学系研究科, 東京大学大学院情報理工学系研究科, November 2003.
- [9] Kotaro Sakamoto, Madoka Ishioroshi, Yuta Fukuhara, Akihiro Iizuka, Hideyuki Shibuki, Tatsunori Mori, and Noriko Kando. Forst: Question answering system for term and essay questions at ntcir-13 qa lab-3 task. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 190–193, December 2017.
- [10] Kotaro Sakamoto, Madoka Ishioroshi, Hyogo Matsui, Takahisa Jin, Wada Fuyuki, Shu Nakayama, Hideyuki Shibuki, Tatsunori Mori, and Noriko Kando. Forst: Question answering system for second-stage examinations at ntcir-12 qa lab-2 task. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 467–472, June 2016.
- [11] Tasuku Kimura, Ryosuke Nakata, and Hisashi Miyamori. Ksu team’s multiple choice qa system at the ntcir-12 qa lab-2 task. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 437–444, June 2016.
- [12] Tasuku Kimura, Ryo Tagami, and Hisashi Miyamori. Ksu team’s qa system for world history exams at the ntcir-13 qa lab-3 task. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 135–142, December 2017.
- [13] 竹内孔一, 土山傑, 守屋将人, 森安祐樹. 類似した動作や状況を検索するための意味役割及び動詞語義付与システムの構築. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 109, No. 390, pp. 1–6, January 2010.