4択クイズを連想問題として解く

外池昌嗣　佐藤理史　宇津呂武仁

京都大学大学院 情報学研究科
tonoike@pine.kuee.kyoto-u.ac.jp, {sato, utsuro}@i.kyoto-u.ac.jp

1. はじめに

近年、自然言語で与えられた質問に明確に答える、質問応答 (QA) と呼ばれるタスクが注目されている。例えば、「世界で一番高い山は何ですか？」という質問があれば、QA では「エベレスト」と答えることが求められる。

これまでの QA を分類すると大きく 2 つのタイプに分けることができる。1 つは文の深い理解に基づく方法 [1] であり、もう 1 つはウェブなどの大量文書から得られる単語の頻度に基づく方法 [2] である。

しかしながら、これまでの QA ではどのような問題なら深い理解が必要で、どのような問題なら浅い理解でもよい、といったことが考えられていない。問題の難易度をある尺度で測定することもされていない。

そこで、本研究では連想問題という問題のクラスを定義し、その難しさと効率的な解き方を示す。連想で解くとは、文書中からその質問の答えだとわかるような記述を探すのではなく、問題文中の重要語句に関係の深い語を探すことを意味する。

しかし、QA の枠組みでは問題が大きすぎるので、この枠組みを大量文書から解候補を探すステップと、解候補の中から解を選ぶステップに分割し、後者のみを対象とする。この制限された枠組みには、解候補はあらかじめ与えられている 4 択クイズが当てはまる。

本研究では、「クイズ＄ミリオネア」☆[3] を題材として、4 択クイズを連想問題として解くことを試みる。情報源としては、膨大なデータがあり、関係の深い語が共起する傾向があって、誰もが使いやすいウェブを用いる。そして問題文中のキーワードとサーチエンジン goo☆☆の返すヒット数だけを使って連想が実現できることを示す。この枠組みは、適切なキーワードを選ぶ問題と、ヒット数から解を決定する問題に分割できる。以下では、この小問題に対する分析と解決法について述べ、2 つの手法を実装する。そして、この方法で「クイズ＄ミリオネア」の問題の 8 割が解けることを示す。

2. 問題文中のキーワードによる解の連想

ここに「クイズ＄ミリオネア」の例がある。

表 1　ヒット数

選択肢	ヒット数
ジョージ・ルーカス	349
スティーブン・スピルバーグ	120
フランシス・フォード・コッポラ	86
黒澤明	72

Q1: 「アメリカン・グラフィティ」の監督は誰？
a: ジョージ・ルーカス
b: スティーブン・スピルバーグ
c: フランシス・フォード・コッポラ
d: 黒澤明

あなたはこの問題の正解を知らず、ウェブのサーチエンジンを使ってその正解を探そうとしているとしよう。最も簡単な方法は、キーワード「アメリカン・グラフィティ」をサーチエンジンに与えて、検索結果のページから 4 人の監督の一人を探すというものである。この戦略は、キーワード「アメリカン・グラフィティ」を含むウェブページ上に正解が現れることを仮定している。

もう少し賢い方法は、選択肢の一つとキーワードの両方を含むページ数を調べるというものである。「アメリカン・グラフィティ」と「ジョージ・ルーカス」の 2 つの語でアンド検索することで、このページ数はサーチエンジンのヒット数から推測できる。この仮定に基づくと、最大のヒット数を持つ解候補を正解であると決定するのは理にかなっている。

上記の例であれば、この戦略は成功する。それぞれの解候補をキーワードと組み合わせたクエリーのヒット数を表 1 に示す。正解である「ジョージ・ルーカス」と組み合わせたクエリーが、最大のヒット数を返している。

上記の戦略は問題文から選んだキーワードのみを使うので、クイズは以下のような形式に変換することができる。

Q1': 「アメリカン・グラフィティ」
a: ジョージ・ルーカス
b: スティーブン・スピルバーグ
c: フランシス・フォード・コッポラ
d: 黒澤明

この変換された形式は、キーワードと選択肢の結び付きや関係の程度を問うものである。言い換えれば、「次のうち与えられたキーワードと最も関係の深いものを選べ」ということであり、これが連想問題である。

☆ Celador 社がオリジナルを製作し、フジテレビが日本版を製作
☆☆ http://www.goo.ne.jp

表 2　ヒット数の例

キーワード	選択肢	選択肢のみ	組み合わせ
カツオ	ノリ	165000	2240
	コンブ	12400	882
	ワカメ	29100	6060
	サンゴ	52900	976
アニメ	ノリ	165000	39400
	コンブ	12400	643
	ワカメ	29100	2210
	サンゴ	52900	600

3. ヒット数の比を用いる手法

では、このキーワードをどのようにして機械的に選べばよいか。例として、次の問題を考える。

A23 アニメ「サザエさん」で、カツオくんの妹の名前は？

A. ノリ, B. コンブ, C. ワカメ, D. サンゴ

[正解: ワカメ]

ここでは、その中からキーワードとして、以下のものを選んだ場合について考える。

- カツオ
- アニメ

ここで、選択肢自身のヒット数と、キーワードと組み合わせたときのヒット数を求めた結果を表 2 に示す。 ヒット数 1 位を選ぶことにすると、前者のキーワードは「ワカメ」を、後者は「ノリ」を支持している。どちらのキーワードがふさわしいのかを決定するために、キーワードによってどれぐらい検索結果が絞り込まれたのかを調べる。すなわち、キーワードと組み合わせたときのヒット数が 1 位の選択肢と 2 位の選択肢に注目し、選択肢自身のヒット数の比と、キーワードと組み合わせたときのヒット数の比という 2 つの比の値の比を調べる。これを式で書いたものを以下に示す。ただし X_{1st} はキーワードと組み合わせたときのヒット数 1 位の選択肢を、X_{2nd} は 2 位の選択肢を、Y はキーワードを表すものとする。

$$\frac{\frac{hits(X_{2nd} and Y)}{hits(X_{1st} and Y)}}{\frac{hits(X_{2nd})}{hits(X_{1st})}}$$

この値が、1 より小であれば、キーワードと組み合わせることによって、ヒット数の差が開いたことを意味する。それぞれのキーワードに対して、この値を求めると、「カツオ」は 0.0652、「アニメ」は 0.254 であった。これは、「カツオ」の方がより絞り込みに貢献する語であることを意味していると考えられる。そこで、すべてのキーワードの組み合せに対してこの値を計算し、この値が最少のキーワードを選ぶという方法が考えられる。そして、このキーワードと組み合わせたヒット数が最大の選択肢を選ぶ。この問題であれば、キーワードとして「カツオ」が選ばれ、解として「ワカメ」を選択することになる。

3.1 手法の実装

上記の考え方で問題を解く手法を実装した。まず、問題文を日本語形態素解析システム JUMAN 2.62 で形態素解析し、その結果を使って、名詞を取り出してキーワードとする。ここで、複合名詞は結合した状態で取り出す。ただし、疑問詞や形式名詞はキーワードとしない。次に取り出したキーワードのすべての組み合わせを生成する。例えば、4 個のキーワードが抽出された場合は $2^4 = 16$ 通りの組み合せが生成される。そして、得られたキーワードの組み合わせすべてに対して、上記のヒット数の比を計算する。最後に比が最少であるキーワードの組み合せの場合の、ヒット数 1 位の選択肢を最終解とする。ただし、1 位のヒット数が 3 未満またはヒット数の差が 2 未満であれば、信用できないので使わない。

表 3　ヒット数の比と正解率の関係

比の範囲	個数	正解数 (率)	不正解数 (率)
比 $= 0$	144	132 (91.7%)	12 (8.3%)
$0 <$ 比 ≤ 0.01	25	23 (92.0%)	2 (8.0%)
$0.01 <$ 比 ≤ 0.1	163	143 (87.7%)	20 (12.3%)
$0.1 <$ 比 ≤ 0.25	158	135 (85.4%)	23 (14.6%)
$0.25 <$ 比 ≤ 0.5	143	92 (64.3%)	51 (35.7%)
$0.5 <$ 比 ≤ 0.75	71	36 (50.7%)	35 (49.3%)
$0.75 <$ 比 ≤ 1	48	20 (41.7%)	28 (58.3%)
$1 <$ 比	53	17 (32.1%)	36 (67.9%)
失敗	3	0 (0%)	3 (100%)
合計	808	598 (74.0%)	210 (26.0%)

3.2 ヒット数の比と正解率の関係

前節で作成したシステムを使って、ヒット数の比と正解率の関係を調べる。まず、問題セットを作成する。本研究では、4 択クイズとして、トミー社から発売されているカードゲーム版の「クイズ＄ミリオネア」の問題を用いた。ゲーム版の問題は、1960 問あり、それらは、賞金額別に A(￥10,000) ~ O(￥10,000,000) の 15 クラスからなる。ここでは、それぞれのクラスを 2 つに分割し、半分 (980 問) を分析用の問題セットとして、残り半分を評価用の問題セットとして使用する。ここでは、分析用セットから、否定、比較、数値を含む問題を除いた、808 問からなる問題セットを作成した。

この問題セットを使って得られた、ヒット数の比と正解率の関係を表 3 に示す。 ここで、比が 0.25 以下の問題を数えると 490 問あり、これは全体の 60.6%を占める。そして、この範囲での正解率は 88.4%であった。すなわち、比が 0.25 以下であれば、この手法で選ばれる答えはかなり信用できるものである。一方で、この範囲にあてはまらない問題は別の解法で解くことが考えられる。

4. 言語知識とヒット数を用いる手法

4.1 キーワード抽出

2 章で検討した方法で問題を機械的に解くには、問題文から適切なキーワードを抽出することが必要となる。そこで、語の種類とヒット数により語の重要さを推測して重要度を付与し、その上位語を抽出する。語の種類で重要なのは、人名や組織名などの固有名詞である。キーワードの個数に関しては、多すぎるとサーチエンジンでヒッ

表 4　重み付けのルール

条件	重み
n 番目の文節	1 + 0.01 * n
「日本」で始まる	0.8
ストップワード	0
「」中の語句	3
人名	3
組織名	1.5
サ変名詞 + サ変動詞	0.5
関係を表す語	2
カタカナ (重要でないものを除く)	2
「賞」で終わる	2
「時代」で終わる	0.5
国名	0.5
数字	3
疑問詞より後ろ	0.2
ヒット数が 10 より小	0.3
ヒット数が 10000000 より大	0.5
ヒット数が 1000000 より大 and 1 文字	0.9
西暦 and 文中に「初」がある	2
西暦 and 文中に「初」がない	1.5
キーワードの後ろに「は」が存在 and	
「者」または「家」で終わる	0.1
ヒット数が 100000 より大	0.2
ヒット数が 10000 より小	1.1
文字数が 1	0.2
文字数が 2	0.25
文字数が 3	0.5
文字数が 4	1.1
文字数が 5 以上	1.2

トを得られなくなることを考慮して 1 個か 2 個とした。

以上の考え方でキーワードを抽出するアルゴリズムを以下に示す。日本語構文解析には KNP 2.0 を使用する。

(1) 問題文を KNP で構文解析する
(2) 文節列を先頭から順に調べる
  (a) 文節から自立語のみを切り出す
  (b) 名詞であれば候補語として抽出する
  (c) 構文情報または形態素情報で、「人名」、「地名」、「組織名」、「カタカナ」、「数字」などの特徴があれば、フラグを立てる
(3) 「」中の語句を抽出する
(4) すべての選択肢に含まれている語を削除する
(5) KNP で判別できない人名を同定する
(6) それぞれの候補語のヒット数を調べ、ヒット数が 15 未満のものを削除する
(7) 候補語に重要度を付与する
(8) 重要度の低い語から順に、個数が 2 個以下でアンド検索のヒット数が 15 以上になるまで削除する

手順 (7) では、品詞やヒット数や係り受けなどの素性を利用して、抽出された語句は重みを付ける。重み付けのルールを表 4 に示す。候補語の初期スコアは 1 とし、ルールにマッチしたらその重みを乗算する。このルールでは、「」中の語句や人名、組織名など具体的な事物を表す語を重視している。

4.2 解の決定

4.2.1 フォワードとバックワードのアソシエーション

語と語の関係の強さは、アンド検索のヒット数によって評価できる。しかし、アンド検索のヒット数は選択肢自身のヒット数の影響を受ける。選択肢のヒット数が大きければ、アンド検索のヒット数も大きくなる。そこで、正規化することが求められる。

佐藤と佐々木は、ヒット数から 2 つの語の関係の程度を評価するための 2 つの距離 [4] を提案した。

$$a(X, Y) = hits(X \text{ and } Y)/hits(X)$$
$$a(Y, X) = hits(Y \text{ and } X)/hits(Y)$$

X をキーワード、Y を選択肢とすると、我々のアプリケーションでは hits(X) は定数なので、a(X, Y) は hits(X and Y) に比例する。a(Y, X) は正規化された値である。

本稿では、1 つ目の距離をフォワードアソシエーション、2 つ目の距離をバックワードアソシエーションと呼ぶ。

$$fa(X, Y) = a(X, Y);$$
$$ba(X, Y) = a(Y, X);$$

ba が正しい答えを返すこともあるが、概して、fa の方が ba よりも信頼できる。

4.2.2 決定ルール

ここでは、fa が 1 位の選択肢か ba が 1 位の選択肢か、どちらを選ぶかを決定する。この時点までに、データとして以下のものが得られている。

- キーワードのみのヒット数
- 選択肢のみのヒット数
- 選択肢 + キーワードのヒット数

これら自身とこれらから計算できる ba の値、fa の比、ba の比を利用してどちらを選ぶかを決定する。検索結果を分析し、次のようなルールを作成した。このルールは上から順に、最初にマッチしたものが適用される。

(1) fa と ba の支持する選択肢が同じであれば、それを採用する
(2) fa が僅差 (比が 0.8 以上) なら、ba を信用する
(3) fa が大差 (比が 0.2 以下) なら、fa を信用する
(4) ba が僅差 (比が 0.8 以上) なら、fa を信用する
(5) キーワードのヒット数が大きければ (1300 以上)、ba を信用する
(6) fa がそれほど離れていない (比が 0.6 以上) なら、ba を信用する
(7) それ以外は fa を信用する

ここで、前述の 808 問の問題セットで調べると、fa と ba が一致する問題は 494 問であり、このうち 452 問が正解であった。半数以上の問題の fa と ba が一致し、このうち 9 割以上は正解であった。fa と ba が一致することは重要な判断基準となることが確かめられた。

5. 実　　験

5.1 組み合わせ手法

これまでに説明した、ヒット数の比を用いる手法と語彙知識とヒット数を用いる手法の 2 つを組み合わせた、新たな手法を設計する。ここではまず、問題をヒット数の比を用いる手法で解く。このとき、最少の比が 0.25 以下

表 5　実験の結果

	クローズテスト	オープンテスト
手法	正解率	正解率
ヒット数の比	74.0%	72.9%
言語知識とヒット数	79.2%	74.3%
組み合わせ	80.3%	77.0%

表 6　解決定ルールの比較

	クローズテスト	オープンテスト
手法	正解率	正解率
fa が 1 位	72.3%	69.0%
ba が 1 位	69.4%	67.1%
決定ルール	79.2%	74.3%

であればこの手法の解を採用する。一方、比が 0.25 より大であれば語彙知識とヒット数を用いる方法で解き直し、その結果を採用する。

5.2　クローズテストとオープンテスト

これまでに、説明した手法でどれぐらいの問題が解けるのかを確かめるために、実験を行った。これまでに、使ってきた問題によるテストはクローズテストである。一方、オープンテスト用に、評価用セットから否定、比較、数値を含む問題を除いた、840 問からなる問題セットを作成した。これらの問題セットを使って、3 種類の手法で実験を行った結果を表 5 に示す。

ヒット数の比を用いる手法では、クローズテストとオープンテストで結果に大きな違いはなかった。この手法では、問題に合わせてパラメータを調整することはしていないので、自然な結果と言える。オープンテストの結果が多少低いのは、キーワード取り出す際の、複合語を分割するルールの影響が考えられる。

一方、語彙知識とヒット数を用いる手法では、オープンテストで正解率が約 5%低下している。この手法では、キーワード抽出部及び解の決定部で様々なパラメータを調整しているので、その影響が考えられる。解の決定部が有効に機能しているかを調べるために、fa が 1 位の選択肢を答えにした場合と、ba が 1 位の選択肢を答えにした場合と比較した。その結果を表 6 に示す。 この結果、解決定ルールを使うと、単純にヒット数 1 位の選択肢を選ぶ手法よりも遥かによい結果が得られることがわかった。したがって、解決定ルールは有効であると言える。ここで、クローズテストとオープンテストの違いに注目する。すべての手法で、オープンテストの正解率が低下しているが、解決定ルールの低下が著しい。これは、キーワード抽出と解決定ルールの両方がクローズテストセット用に強く調整されてしまっている事を意味する。

組み合せ手法においては、残り 2 つの手法を単独で使うよりもよい結果となった。オープンテストではその差が顕著に現れている。この現象を説明するために、どちらの手法が使われたのかの内訳の統計を表 7 に示す。 これを見ると、ヒット数の比を用いる手法で解いた問題はどちらのテストセットでも高い精度で解けていることがわかる。このことから、比が 0.25 以下という閾値が高い正解率を保証することが確かめられた。オープンテストの結果が単独の手法よりかなりよくなったのは、多くの問題がヒット数の比を用いる手法で解かれたことによる。

表 7　組み合せ手法の内訳

	クローズテスト		オープンテスト	
内訳	占める割合	正解率	占める割合	正解率
ヒット数の比	60.6%	88.4%	62.9%	86.7%
言語知識とヒット数	39.4%	67.9%	37.1%	60.6%
合計	100%	80.3%	100%	77.0%

6. おわりに

本研究では、4 択クイズを連想問題とみなし、ヒット数を使って解く方法を 2 種類検討した。これらの手法を組み合わせることにより、クローズテストで 80.3%、オープンテストで 77.0% の正解率を得ることができた。したがって、言語知識とヒット数を組み合わせることによって、連想を実現できると言える。

本研究の手法は、一般の QA において、キーワードを抽出する部分と、何らかの方法で得られた解候補から正解を探す部分に貢献できるであろう。

関連研究として、Magnini らはウェブのヒット数を利用して与えられた解候補が正しいかを検証する方法を提案 [5] している。システムの構成は我々の語彙とヒット数を用いる方法と似ているが、本研究で成功した、ヒット数を他の解候補と比較しながらキーワードを選択する要素は含まれていない。

今後の課題としては、これらの手法で有益であるとわかった素性を利用して、機械学習でパラメータを調整することを考えている。そして、今回考慮していなかった文型情報についても検討する予定である。

参　考　文　献

1) Moldvan, D., Harabagiu, S., Girju, R., Morarescu, P. and Lacatusu, F.: LCC Tools for Question Answering, *Proceedings of the Eleventh Text REtrieval Conference (TREC 2002)* (2003).
2) Kwok, C. C. T., Etzioni, O. and Weld, D. S.: Scaling Question Answering to the Web, *Proceedings of the 10th International World Wide Web Conference*, pp. 150–161 (2001).
3) クイズ $ ミリオネア (編): 完全攻略！クイズ $ ミリオネア, フジテレビ出版 (2002).
4) Sato, S. and Sasaki, Y.: Automatic Collection of Related Terms from the Web, *Proceedings of ACL 2003*, pp. 121–124 (2003).
5) Magnini, B., Negri, M., Prevete, R. and Tanev, H.: Is It the Right Answer? Exploiting Web Redundancy for Answer Validation, *Proceedings of the 40th Annual Meeting of ACL*, pp. 425–432 (2002).