

WWWからの descriptor 抽出システムの開発とその評価

川村 佳史 榎井 文人 河合 敦夫 井須 尚紀
三重大学大学院工学研究科
{yosi,masui,kawai,isu}@ai.info.mie-u.ac.jp

1 はじめに

近年,日々増大しつづける電子化テキストから必要な情報を効率よく得るための技術として,「情報アクセスの為のテキスト処理」技術が注目を浴びている.この技術の基礎は,膨大なコーパスから要求された情報のみを効率よく探し出すことである.このような技術の応用例として,Web 検索が挙げられる.

Web 検索はユーザの要求に対し,大雑把に適合する文書を提示する技術である.それに対し,答えそのものを提示する技術として,質問応答技術が注目されている.質問応答技術では,ユーザの要求に対し名詞や固有表現を用いて解答を提示する factoid 型のタスクが一定の成果をあげている.[1, 2]

ところが,現在では factoid 型タスクよりも高次のタスクとして,why や how 型のタスク,定義を回答するタスクに注目が集まっている.[3]例えば,角ら[4]はWWWマイニングによって構築したオントロジーを利用した協調型合意形成支援システムを提案しているが,ユーザの誤解の解消や,未知の知識の補間などが重要であるとしている.

また,黒橋[5]は,大規模なテキスト知識ベースを利用した自動QAシステムを構築しており,システムとユーザのインタラクティブなインターフェイスについて述べているが,やはり,知らない知識に対する質問と応答の重要性を挙げている.

しかしながら,このようなタスクに対応するには,照応解析や構文処理などの複雑な処理が必要となり,これまでの factoid 型タスクでの技術の応用では困難である.このような背景から,本論文では定義を回答するタスクに対して,概念の descriptor[6](概念の記述要素,概念を描写または参照する異なる名詞句)を用いて回答を提示する手法とその評価を報告する.

2 基本的な考え方

本節では WWW からのディスクリプション情報抽出システムでの基本的な考え方について述べる.

WWW の利用

WWW は日々更新され,増え続ける大規模テキストコーパスとみなすことが出来る.このような大規模コーパスを用いることで,網羅性の確保が出来ると考える.また WWW は日々更新され,その内容も変化する.このような WWW の性質を用いることで,時事のように時間経過によって description 情報が変化する概念に対応できると考える.

descriptor の利用

抽出した知識を用いて概念の定義を示すためには,文脈処理や構文処理等の複雑かつ高度な処理が必要となってくる.このような複雑かつ技術的難易度の高い処理は計算量の増加に伴うため,動的に処理しようとした際にボトルネックになる.

川崎[7]は,プログラム理解支援システムを対象として,システムに事前知識が無い場合に,類似した知識を用いて言い換えることで,理解した場合と同等の効果が得られる事を示している.

そこで,概念に対する情報を過不足なく呈示するのではなく,概念を理解するために十分な知識を提示する必要性があると考えられる.

3 逐次型 descriptor 抽出システム Murasaki

本節では我々が開発した逐次型 descriptor 抽出システム Murasaki についての概要を述べる.Murasaki では図1のようなアルゴリズムに基づいて WWW より de-

descriptor を抽出する．処理モジュールは，(1) クエリ認識，(2) 比較表現 (A) 生成，(3) 候補抽出，(4) 比較表現 (B) 生成，(5) スコアリング，(6) 整列，(7) 可視化で，以下ではその概要を説明する．

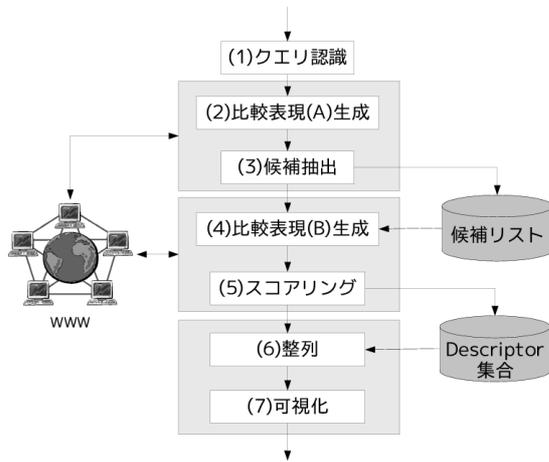


図 1: Murasaki の処理概要

- (1) では入力されたクエリを認識する．
- (2) では認識したクエリと「のような」といった統語パターンから「クエリ + 統語パターン」という WWW 検索クエリを発行する．今回は統語パターンとして，「のような」を用いた．また，検索エンジンとして `goo*1` を用いた．
- (3) 検索された Web ページのスニペット中の，「クエリ + 統語パターン + 名詞概念」という表現の中から名詞概念を抽出し，抽出した名詞概念集合を descriptor 候補とする．
- (4) descriptor 候補各々に対して「クエリ + 統語パターン + descriptor 候補」という検索クエリを生成し，この検索クエリを含むページが WWW 内に何件あるか調べる．
- (5) descriptor 候補各々の WWW 中に存在する「クエリ + 統語パターン + descriptor 候補」のページ数を頻度とし，スコアリングを施す．スコアリングには， $pf * icf$ [8] とページ数を正規化した相対頻度を用いた．ここで $pf * icf$ は (式 (1)) で示され，クエリを X ，クエリ X を表現する descriptor を w_i とする．

$$score(w_i|X) = pf(w_i|X) * icf(w_i) \quad (1)$$

- (6) スコアに基づいてソートする．
- (7) スコアを可視化して表示し，ユーザが視覚的に理解しやすい形で結果を提示する．

4 評価実験

本節では Murasaki を用いて行った，Murasaki の抽出する descriptor が妥当かどうかの評価，ベースラインとの比較評価の 2 つの評価実験について述べる．

4.1 評価環境

実験環境は次のように設定した．大学生 20 名を被験者として，Murasaki に対して自由にクエリを入力し，得られた descriptor 集合の正否判定をしてもらった．判定規準は，提示された descriptor がイメージ通りか，イメージ通りではないかの 2 値判断とした．この結果 173 組のクエリに対する判定結果が得られた．

また，対照実験として共起情報から descriptor 相当の知識を提示するベースライン手法を実装し，上述のクエリ 173 組に対し人手で正否判定をしてもらった．ベースライン手法では，クエリ X を先端記号とするサイズ n のウィンドウを設定し，その中で最も X の近傍の名詞句を descriptor 相当の共起要素として抽出する．

4.2 評価指標

提案手法によって得られた descriptor 集合の性能に関して，適合率 (Precision)，相対再現率 (Relative Recall)，MRR (Mean Reciprocal Rank) を用いて評価を行った．

・適合率による評価

まず，適合率による評価を行い，ベースラインと提案手法の比較を行った．これにより，提案手法の基本的な性能が把握できる．

・相対再現率による評価

次に，ベースラインに対する提案手法の相対再現率 [9] を求め，提案手法の網羅性を検証した．相対再現率とは本来ユーザが問題解決に必要とする正解数の何割をシステムが提示できたかを表す尺度である．Pantel ら [10] は，この尺度を 2 つの手法の処理結果に対する適合率を用いて表現できることを示し，再現率 (Recall) 計算が困難な場合に正解データの網羅性を吟味するために有効な評価尺度として，大規模な対象データ集合に対するシステムの評価に利用している．

本論文では，Pantel らと同様の考え方に立ち，前述のクエリ 173 組に対し，ベースラインに対する提案手法の

*1 <http://goo.co.jp/>

表 1 descriptor 集合の比較

クエリ	順位	pf*icf	相対頻度
ロサンゼルス	1 位	大都市	大都市
	2 位	スター	スター
	3 位	北米の都市特有の夜景	都会
	4 位	都市部の水道施設	場所
	5 位	都市の密集地域	北米の都市特有の夜景

相対再現率を用いて網羅性を吟味した。

以下に相対再現率を導出する式を示す。(式 (2)) ここで提案手法とベースラインの相対再現率を $R_{A,B}$, クエリに対するベースラインの回答数を $|A|$, 提案手法の回答数を $|B|$, ベースラインの適合率を P_A , 提案手法の適合率を P_B とする。

$$R_{A,B} = \frac{P_A \times |A|}{P_B \times |B|} \quad (2)$$

・順位尺度による評価

MRR は、順位付けされたシステム出力において最高位に正解順位の逆数の平均値として表現される。入力クエリについて提示された descriptor をスコアに基づいて順位付けして得た MRR 値は、各クエリを表現する descriptor 集合の平均的な抽出性能を評価する指標であるとみなせる。

4.3 評価結果

適合率については、ベースラインでは 0.351 に対し、提案手法では 0.531 であった。(図 2) ここで Welch の t 検定を用いて検定した結果、有意水準 0.1% で提案手法の優位性が確認できた。また、ベースラインでの n の値は $n=1 \sim 12$ とした結果を見て適切だと思われる値 $n=6$ を採用した。(図 3)

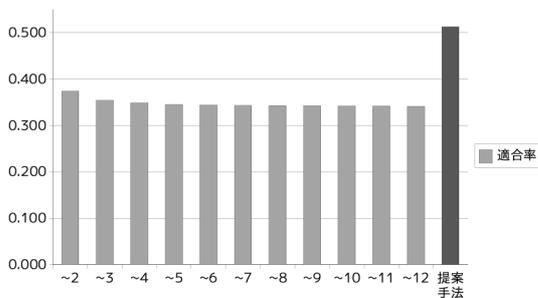


図 2: 提案手法と window size 別の適合率

また、相対再現率は 1.11 であり、提案手法の方がわずかに網羅性があることが確認された。

順位尺度では相対頻度での平均 MRR 値は 0.758 で、pf*icf で 0.742 であり、スコア付けの尺度を変化させても上位に適切な descriptor が出現することが確認された。[11]

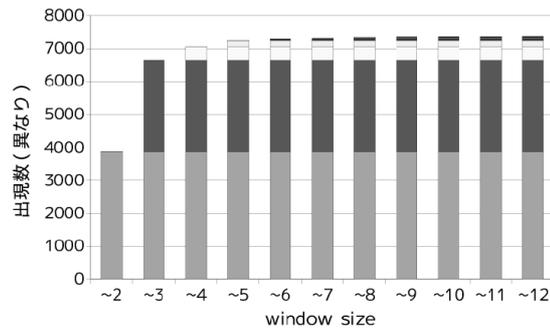


図 3: 各 window size での名詞概念出現数

5 考察

評価結果について考察を行う。適合率についていえば、ベースライン手法のように、単純な共起関係に基づきクエリと同じ文脈上に存在する名詞句を抽出するよりも、提案手法のように意味的に関係のある共起を用いた方が明らかに有効である。

また、平均 MRR 値については、相対頻度を用いた手法でも pf*icf を用いた手法でも有意な差は認められなかった。しかし、個々のデータを見たところ、提示される descriptor には、クエリに上位概念を表現するものと descriptor の特徴や具体的記述を表現するものがあり、これらの集合内における分布が異なっていることがわかった。

例えば、クエリ「ロサンゼルス」についての処理結果(表 1)を見てみると、相対頻度では「大都市」「都会」「場所」などのようにクエリに対して上位概念の関係にあるものや、抽象度の高いものが抽出される傾向が強かった。

これに対して pf*ief を用いた処理結果では、「北米の都市特有の夜景」「都市部の水道施設」「都市の密集地域」などのようなクエリの特徴や、より具体的な描写を表すものが上位に抽出される傾向が強かった。

今回の評価基準では、抽出された descriptor が妥当か妥当でないかという観点についてのみ考慮したため、評価尺度においてこれらの違いを検出することができなかった。この違いを考慮するためには、descriptor 集合を参照した時に、どちらがよりクエリ概念をイメージし易いか、というような観点からの評価を行う必要があると考えられている。

最終的にユーザに対してクエリの定義と同等の理解を促すためには、上位概念のみの提示、具体的な記述のみの提示、どちらでも十分とはいえない。両方をバランス良く提示することによって、はじめてクエリの理解が得られる。したがって、今後は抽象度の高い descriptor と具体性の高い descriptor を分類し、整理して提示するような仕組みを考える必要がある。

次に適切な descriptor 集合がうまく得られなかった例について考察する。クエリ概念が多義語であるため、descriptor 集合が複数の集合となって混在するものがあつた。例として、「AI (= 人工知能という技術, 映画のタイトル, 発音記号, 頭辞語など)」があげられる。これらは、descriptor 自体の数はある程度抽出されているので、どの意味で用いられているかが判断できれば、有効な descriptor 集合を割り当てて提示することができるはずである。この問題に対しては、descriptor をクラスタリングすることで、多義性を整理することが可能であると考えている。

6 おわりに

本論文では、情報アクセス技術への応用を目的として、比較構造を利用した WWW から descriptor 集合を抽出し、提示する手法を提案した。

提案手法を試作システム Murasaki として実装し、実験を行ったところ、適合率では 0.531、ベースラインとの相対再現率は 1.11、平均 MRR 値は 0.742 が得られ、提案手法の能力に関する有効性が確認できた。

提案手法は、質問応答における、定義を問うタスクに対応可能であり、また、逐次型のシソーラス自動構築手法にも応用できると考えている。提案手法を応用することで、従来のような複雑な処理を用いることなく、ある程度の質を保持した回答を出力するシステムの実現の可

能性ができたとと言える。

参考文献

- [1] Ellen M. Voorhees: The Evaluation of Question Answering Systems: Lessons Learned from the TREC QA Track. In Proc. of the LREC 2002 Workshop on Question Answering - Strategy and Resources, pp. 1-4, 2002.
- [2] Jun'ichi Fukumoto, Tsuneaki Kato, and Fumito Masui: Question Answering Challenge (QAC1) An Evaluation of QA Tasks at the NTCIR Workshop 3. In Papers from the 2003 AAAI Spring Symposium "New Directions in Question Answering", pp. 1-4, 2003.
- [3] 加藤恒昭, 福本淳一, 榎井文人, 神門典子: 質疑応答から対話理解へ -NTCIR QAC Task3 の提案-. 言語処理学会第 10 回年次大会発表論文集, 2004.
- [4] 角薫, 溝口理一郎: オントロジー工学と HCL を融合した協調型合意形成支援システム. Sig-fai-a201-04, 人工知能基礎論研究会, 2002.
- [5] 黒橋禎夫: 大規模テキスト知識ベースに基づく自動質問応答. NLC-2001-73, 信学技報, 2001
- [6] Nancy Chinchor: Template Element Task Object Definition Version 4.2. In MUC-7 Information Extraction Task Definition, pp. 9-21, 1998.
- [7] 川崎治夫: 言い換えのプログラム理解への応用. 言語処理学会第 7 回年次大会ワークショップ論文集, pp. 89-92, 2001.
- [8] 榎井文人, 福本淳一, 荒木健治: 比喩解釈を目的とする World Wide Web を利用した属性の適合性判定. 言語処理学会第 11 回年次大会発表論文集, pp. C2-2, 2005.
- [9] Keen, E. M.: Evaluation parameters. In G. Salton (Ed.), The SMART retrieval system-Experiments in automatic document processing. Prentice-Hall. 1971.
- [10] Pantel, P., Ravichandran, D., Hovy, E.: Towards terascale knowledge acquisition. COLING 2004, pp. 771-777, 2004
- [11] 川村佳史, 榎井文人, 河合敦夫, 井須尚紀: WWW から Descriptive 知識を抽出・提示するシステム Murasaki の試作. 言語処理学会第 12 回年次大会発表論文集, pp. P8-10, 2006.