

曖昧な情報要求に対する提案型応答の検討

Deliberation of Question Answering for Ambiguous Information Request

木村泰知† 荒木健治†
Yasutomo Kimura Kenji Araki

†小樽商科大学 商学部 社会情報学科 †北海道大学 大学院 情報科学研究科

1 まえがき

ユーザが要求する情報は、明確な正解が存在する要求とは限らない。たとえば、流行を尋ねる質問[1]や連想の支援を期待する質問は、正解を予め用意することが困難である。

そこで、本研究では、情報要求の曖昧さがある場面における、情報提供について考察する。ユーザの曖昧な要求に対して、地図、可視化などテキスト以外の表現を含めた提案型の応答を検討する。

Taylor は情報要求を次の4つに分類している[2]。

情報要求の分類

1. 直感的要求

具体的に言語化して上手く説明できない状態

2. 意識された要求

曖昧な表現でしか言語化できない状態

3. 形式化された要求

問題を具体的な言語表現で言語化できる状態

4. 調整済みの要求

問題を解決するために必要な情報の情報源が同定できるくらい問題が具体化された状態

上記の「3. 形式化された要求」、「4. 調整済みの要求」は明確な情報要求であり、ユーザが言語化できるため、情報検索のクエリを作成することが可能である。一方、「1. 直感的な要求」、「2. 意識化された要求」は、クエリを作成することが困難であり、ユーザの要求を言語化することが難しい状況である。

このような言語化が難しい場面で、ユーザを支援する方法がいくつか存在する。たとえば、連想検索は、検索キーワードから関連性の高い単語を抽出し、探し出す。他に、キーワード予測もユーザ支援である。Google サジェスト[3]のようなキーワード予測も、関連性の高い単語を提示することで、意識された要求を言語化する方法と考えることができる。

曖昧な表現でしか言語化できないユーザを助ける方法として、キーワード予測のような入力処理以外に、出力処理からのアプローチもある。本研究では、ユーザが要求していると考えられる情報の提示について考察する。

2 予備調査

曖昧な応答に対するの応答表現に関するアンケートを行った。対象者は大学生34名であり、性別は女性8名、男性26名である。具体的な質問を想定し、どのような応答表現が良いかを調査した。今回は、曖昧な情報要求として「小樽で流行っていることは何ですか」に対して、理想的な応答に必要な条件のアンケートを実施した。

●「必要な情報は何か?」(複数選択可)

1. URL
2. 文字情報
3. 音声情報
4. 写真
5. 図
6. 表
7. 動画
8. 地図
9. その他

表1 必要な情報に関するアンケート

1. URL	17/33	0.5152
2. 文字情報	26/33	0.7879
3. 音声情報	3/33	0.0909
4. 写真	29/33	0.8788
5. 図	5/33	0.1515
6. 表	2/33	0.0606
7. 動画	11/33	0.3333
8. 地図	22/33	0.6667

※有効回答数 33

●「最も重視する条件は何か?」(一つ選択)

1. 情報の質
2. 情報の量
3. クリック数などの負担軽減
4. 図、写真、表などの表現
5. デザイン
6. その他

表2 重視する条件に関するアンケート

1. 情報の質	12
2. 情報の量	3
3. クリック数などの負担軽減	0
4. 図、写真、表などの表現	18
5. デザイン	0
6. その他	1

※有効回答数 34

このアンケート結果から、図、写真、表などを含めて、簡単に判断できるような出力が望まれていることがわかる。そこで、本研究では、図、写真、表を含めた提案型応答について研究を行う。

3 提案型応答

本研究において、提案型応答とはあいまいな情報要求に対して関連ある情報(テキスト、図、写真、表、地図)を利用して応答することである。

具体的な処理

- 関連性の高い固有表現から優先的に表示
- 関連ブログを自動スクロール
- 評判情報の提供

4 処理過程

本手法の処理過程を図1に示す。

4.1 入力文解析

入力文解析は質問タイプの判定およびキーワードの抽出である。質問タイプは、入力文に含まれる単語に基づいて、予め与えた規則により判定を行う。質問タイプの種類は「人名、地名、組織名、固有物名、日付、時間、金額、割合、Why型、Definition型、How型」に加え、TREND(流行)型を付け加え、12種類としている。キーワードは、ChaSenにより形態素解析を利用し、名詞を抽出する。

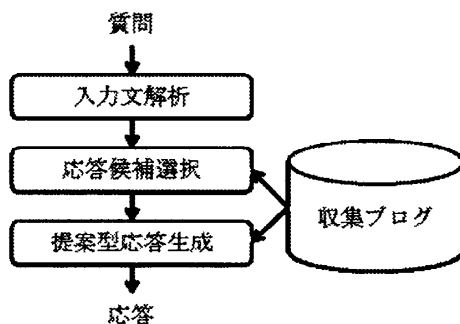


図1 処理過程

4.2 応答候補選択に利用するブログデータ

本研究では検索対象を小樽に限定する。そこで、2006年9月から「小樽」および「おたる」を含むブログ情報を1時間おきに収集しており、2007年6月30日の時点で「小樽」に関連するブログは38,000記事以上収集している。

本稿で処理するブログ情報はRSS形式で処理できる状態のものである。RSSは、サイトの更新情報をまとめたフォーマットであり、XMLによって記述されている。RSSは投稿時間を抽出可能であり、時系列の変化を考慮できる。収集ブログは、全文検索ツールNamazuを用いてインデキシングを行う。

4.3 応答候補選択

応答候補選択では、質問タイプにより利用する情報源が異なる。従来の固有表現の場合、Webを利用した情報検索を行い、Googleのsnippetを利用している。曖昧な情報要求(TREND型)の場合には、収集したブログから検索を行う。

TREND型質問は、最新のブログ1,000件を対象とする。

4.4 提案型応答生成

提案型応答生成では、ユーザの満足する情報を提供するためにテキスト以外にも地図、評判情報の可視化、自動スクロールなどの技術を利用する。

たとえば、「カフェプロジェクトはどこにありますか」と場所を尋ねる質問の場合、住所に加えて地図を提示することがユーザの満足度を高めると考えられる。そこで、曖昧な情報要求に対しても、関連情報を提供することで、ユーザの要求を満足させ

る。本研究では次の処理を行っている。

● 関連情報の表示

➢ 関連性の高い固有表現の表示

質問文に含まれるキーワードを含む文だけを抽出し、抽出された文書から固有表現抽出を行う。タイプ(人名、地名など)ごとに分類し、出現頻度の高い上位3単語を表示する。

➢ 電子地図の利用

地図はGoogle MAPS APIを利用している。

● 関連ブログの表示

➢ キーワードを含むブログタイトルとリンクを表示

➢ 自動スクロールによりスペースを有効に利用

● 評判情報の提供

➢ 形容詞に基づく評判情報の可視化

➢ 形容詞を5段階に分類し、gnuplotを用いてグラフを自動的に生成する。

5 本システムの実装と検討

「最近、流行っていることは何ですか」の質問に対する実装例を図2に示す。対象を小樽に限定しているため、応答は「小樽ドリームビーチ、ダブルフロマージュ、ルタオ、小樽観光、おたる水族館」が流行っています。」と答えており、季節を反映した海に関する内容がある。

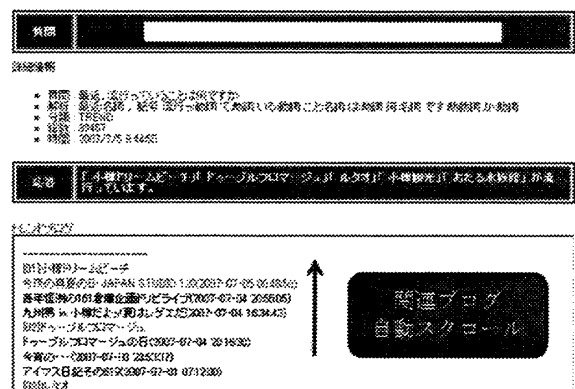


図2 実装例

提案型応答の満足度は、ユーザの要求に依存するため、ユーザごとの評価が必要である。また、ユーザの要求を理解するには、1回の質問から判断することは困難である。そこで、対話システムとして、ユーザの曖昧な要求を解決するタスクを想定し、解決するまでの労力を軽減する視点の評価を予定している。

6 今後の予定

提案型応答の応答表現の評価実験を行う。また、提案型応答を組み込んだ対話システムの構築を行う予定である。

参考文献

- [1] 木村泰知, 荒木健治, “ブログの投稿時間を利用した地図付きトレンド型質問応答システムの提案,” 言語処理学会第13回年次大会, pp.298-301, 2007.
- [2] 徳永健伸, “情報検索と言語処理,” 東京大学出版会, 1999.
- [3] <http://www.google.co.jp/webhp?complete=1&hl=ja>