E-002

大規模な障害事例を用いた質問応答システム
Question and answering system using a large amount of trouble shooting logs

瀬川 修 † **村上 一彦** ‡ **古里 宗寛** ‡
Osamu Segawa Kazuhiko Murakami Munehiro Furusato

1. はじめに

我々は、社内 PC や IT システムの障害対応の支援技術の開発を行なっている。本稿では、約 9 万件にのぼる大量の障害事例を用いた質問応答システムについて述べる。障害対応を担当するヘルプデスクでは、電話で受け付けた各事象について「現象内容」と「対応内容」の記録をテキストで蓄積しており、これらの大規模かつ非構造的な記録は障害対応において重要な知識源であると言える。しかしながら、現状の対応業務においては過去事例の蓄積はなされるものの、知識として有効活用が十分になされていない。そこで、本研究では大規模な知識源に基づく、利便性の高い自然言語による「質問応答」という手段を提供し、障害対応業務の支援技術の実現と有効性評価を行う。

2. 障害対応を支援する質問応答技術

自然言語で入力された質問に対し適切な回答を提示する質問応答システム [1][2] は、蓄積された非構造的な知識源を有効に活用する技術である。

設備の障害対応においては予め知識化された FAQ があれば大変有用であるが、知識の整理とメンテナンスに多大なコストを要し、また事象に対するカバー率も十分ではない。そこで、構造化されていない障害記録のテキストを知識源としてそのまま活用することができれば、知識管理のコストや事象のカバー率の観点から多大なメリットがある。今回開発した質問応答システムでは、蓄積された過去事例を情報源として用い、自然言語による利便性の高い知識検索の実現を目指す。

今回対象とするタスクの障害記録の例 (電話応対のログ) を表 1 に示す。

表 1: 障害記録の例

(現象内容) ファイルの圧縮方法がわからない。 社外の方にファイルをメールで送りたい。
(対応内容) 下記のコンテンツを案内。 「ヘルプデスク＞操作マニュアル＞ ...(中略)... 　＞ｚｉｐ形式にファイルを圧縮する)」 ファイルを圧縮し容量が５ＭＢ以上超えるようなら、分けて送付頂くようお願い。

図 1 に本システムの基本構成を示す。モジュールは「質問解析」、「事例検索」、「回答候補評価」、「回答生成」から成る。以下、各モジュールの処理内容詳細について述べる。

†中部電力 (株) エネルギー応用研究所
‡(株) 中電シーティーアイ

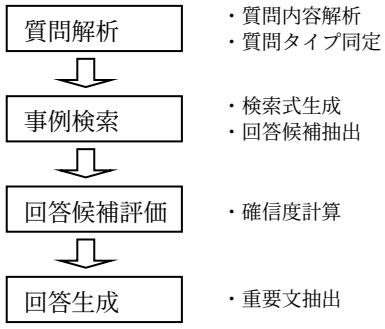

図 1: システムの基本構成

2.1 質問解析

ここでは、入力された質問の文字列 (フレーズまたは単語系列) に対し形態素解析を行い (chasen を使用)、キーワードと品詞情報を得る。同時に品詞と「手掛かり表現」から質問タイプを同定する。質問タイプとしては、How_to 型、Who 型、Where 型、When 型を定義した。

2.2 事例検索

質問解析で得られたキーワード群から検索クエリを生成し、蓄積された事例の検索を行う。また同義語や表記揺れに対応するためシソーラス (独自作成) によるクエリ拡張も行っている。

検索は基本的に「文字列マッチング」により行うが、初期候補の再現率向上のため以下の 3 段階の手順により実行する。

(1) フレーズ検索: 入力質問の文字列を分割しないでフレーズのままマッチングを行う。

(2) 文節検索: 上記 (1) で候補が無い場合は、入力質問の文字列をチャンキングにより文節単位に分割してマッチングを行う。

(3) キーワード検索: 上記 (1)、(2) で候補が無い場合は、入力質問の文字列を形態素解析により単語に分割し、特定品詞のキーワードから検索クエリを生成し事例検索を行う。

2.3 回答候補評価

回答候補評価では、事例検索で得られた回答候補の適合性を「確信度」として定量的に評価する。確信度の算定には以下の各要因を考慮し、総合的な評価値を求める。

(1) 質問との類似性: 入力質問と事例のキーワードレベルでの類似性 (キーワードの TF-IDF の累積値で、事例中の形態素数で正規化) を評価。

(2) 質問タイプとの整合性: 入力された質問のタイプと事例のタイプ (前処理で判定) の整合性を評価。

(3) キーワード共起性: 質問に含まれるキーワードの事例内での共起、すなわち観測窓内 (文単位) での同時出現数を評価。

2.3.1 ロジスティック回帰分析による学習

ここでは、評価式の学習にロジスティック回帰分析 [7] を用いる。ロジスティック回帰分析は、目的変数や説明変数に「ある」、「ない」のような質的変数を扱うことが可能で、さらに目的変数を確率値のように 0〜1 の範囲の実数スコアとして出力できることから、今回の「確信度」の算出に適している。

確信度 p は次式で求められる。

$$p = \frac{1}{1 + e^{(-Z)}}$$

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

ここで、x_1 は「質問との類似性」(キーワードの TF-IDF の累積値)、x_2 は「質問タイプとの整合性」(0 or 1 の 2 値)、x_3 は「キーワード共起性」(観測窓内での最大共起数)、β_0 はバイアス項、$\beta_1, \cdots, \beta_3$ は各変数の回帰係数である。

2.4 回答生成

回答生成では適切と判定された事例から重要文を抽出することにより、簡易な要約という形で情報提示する。重要文の判定は以下の基準を用いた。

- 質問に含まれるキーワード (名詞) とそのシソーラス語が含まれている文。
- 「対応内容」のテキストで、手掛かり表現「回答」が含まれている文。
- 「対応内容」のテキストで、対応に関わる手順 (例えば、メニュー階層など) が含まれている文。
- 「対応内容」のテキストで、手掛かり表現「お伝え」or 「お願い」 or 「案内」が含まれている文。

3. システム実装

本システムは ASP.NET の Web アプリケーションとして実装し、ブラウザをインタフェースとして用いる。ユーザ管理はログイン認証を設け、個々の操作ログの収集を可能としている。

システムの入力画面例を図 2 に示す。ユーザは参照する事例の期間と分野を指定することができる。また、システムの実行例を図 3 に示す。

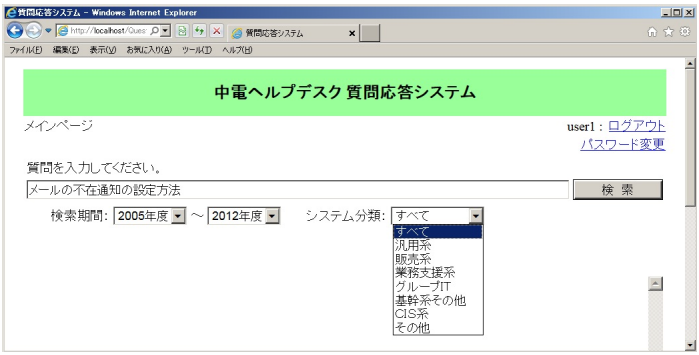

図 2: システムの入力画面例

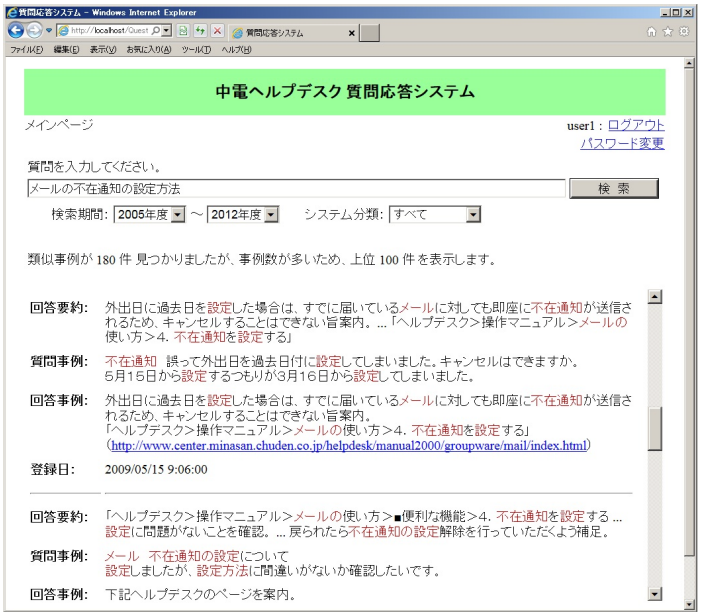

図 3: システムの実行例

4. 評価実験

4.1 事例データ

評価では 2005 年度から 2012 年度までに蓄積された障害事例約 9 万件を知識源として用いた。

4.2 テストセット

テストセットは、確信度の評価式の学習に用いていない事例からサンプリングして作成した質問と回答のペア 50 セットを用いた。質問の例を表 2 に示す。

表 2: 評価に用いた質問の例

メールの添付ファイルが開けません
使用者不注意の処理表の提出先を教えて欲しい
従業員証を自宅に忘れました
印刷物に黒い線が入る
セキュリティパッチ 再起動 繰り返し

4.3 評価方法

評価では以下の 3 手法で実装したシステムの比較を行う。

(1) ベースライン 1

- 事例検索は 2.2 節の「キーワード検索」のみ実行。
- 確信度の評価は次式を用いた。ここで、x_1 と x_2 は、それぞれ 2.3 節の「質問との類似性」(キーワードの TF-IDF の累積値) 及び「質問タイプとの整合性」(0 or 1 の 2 値) である。重み係数 α_1, α_2 は実験的に調整。

$$p = \alpha_1 x_1 (1 + \alpha_2 x_2)$$

(2) ベースライン 2 (改良法)

ベースライン 1 に以下の改良を行った。

- 事例検索において 2.2 節の「フレーズ検索」と「文節検索」を追加。
- 確信度の評価式に 2.3 節の「キーワード共起性」(観測窓内での最大共起数) の変数 x_3 を追加。重み係数 $\alpha_1, \cdots, \alpha_3$ は実験的に調整。

$$p = \alpha_1 x_1 (1 + \alpha_2 x_2) + \alpha_3 x_3$$

(3) ロジスティック回帰分析

事例検索については上記 (2) と同条件。確信度の評価式の学習にロジスティック回帰分析を用いる。

回帰係数の学習では、学習データの事例 N 個をランダムに抽出し、各事例の「現象内容」と「対応内容」のペアを正解 (正例) とする。そして「現象内容」に対応しない他の事例の「対応内容」とのペア ($N-1$ 個) を不正解 (負例) とした。ここでは、20 の事例をランダムにサンプリングし、$20 \times 20 = 400$ の組み合せを 1 つのデータセットとする。このデータセットを 5 組用意し、合計 2000 の学習データとしてロジスティック回帰分析を行った。

4.4 評価結果

前述のテストセットを用いてシステムの性能評価を行った。評価指標には、MRR(Mean Reciprocal Rank: ランク逆数の平均) 及び Top-N(上位 N まで考慮した正解率) を用いた。評価結果を表 3 に示す。

表 3: 評価結果 (N 位の数値は当該順位での正解数)

手法	1 位	2 位	3 位	4 位	5 位	MRR	Top-5
ベースライン 1	36	2	3	3	2	0.783	0.920
ベースライン 2 (改良法)	41	1	2	2	2	0.861	0.960
ロジスティック回帰分析	41	1	4	1	1	0.866	0.960

4.5 考察

- ベースライン 1 の手法に対し、「フレーズ検索」と「文節検索」を追加し、また確信度に「キーワード共起性」の要因を加えることによって性能向上が見られた。
- さらに、ロジスティック回帰分析による手法では、ベースライン 2 の改良法と遜色がない結果が得られており、自動学習の効果が現れている。
- ロジスティック回帰分析の学習データについては、さらに種々の組み合せ方法を検討していく。

5. 関連研究

ヘルプデスクを対象とした質問応答システムの研究としては、清田らのダイアログナビ [3] がある。本システムでは、質問の曖昧性に対応するために同義語辞書や上位・下位語辞書を用いて表記揺れに対応している。また、ユーザ支援機能として対話システム的なインタフェースを設け、質問の曖昧性を解消するための「聞き返し」を可能としている。知識ベースとしては、マイクロソフトが公開しているテキストデータ (用語集 4,707 件、ヘルプ集 11,306 件、サポート技術情報 23,323 件) を用いてる。

この他、ヘルプデスクを指向した研究事例としては、三原らのシステム [4] がある。本システムでは Web 検索エンジンを利用して質問に関連するテキストの収集を行い、その中で質問に関連する「行動表現」を抽出している。著者らの定義する行動表現とは、名詞+助詞+動詞という係り受け構造である。また同じ研究グループの佐々木らのシステム [5] では、Why 型 QA のタスクを取り上げ、「取るべき行動と理由」を併せて提示する手法を提案している。知識ベースは Web 検索エンジンを用い、上記システムと同様の「行動表現」の近傍に出現する理由の手掛かり表現 (理由語) を含む文を解答候補としている。

さらに、岡本らのシステム [6] では、回答候補のランキングに事例のカテゴリ情報を利用し、CCFI(Cross Category Feature Importance) と TF-IDF を融合した類似度関数を用いている。本システムでは知識ベースとしてパソコンに関する QA 事例 21,000 件を用いている。

6. まとめ

本稿では、大量の障害事例を用いた質問応答システムの構築について述べた。本システムでは、入力質問に適合した回答を評価するために侯補事例の確信度を算定する枠組みを用いた。また、確信度の評価式の学習にロジスティック回帰分析を適用し、その有効性を確認した。今後は実フィールドにおいて試行評価を行い、ユーザの質問ログを収集しながら自然言語インタフェースの知見を得ると共にシステムの改良を進めていく。

参考文献

[1] 磯崎, 東中, 永田, 加藤, “質問応答システム”, コロナ社, 2009.

[2] E.M.Voorhees and D.K.Harman, “TREC: Experiment and Evaluation in Information Retrieval”, MIT Press, 2005.

[3] 清田, 木戸, 黒橋, “大規模テキスト知識ベースに基づく自動質問応答: ダイアログナビ”, 自然言語処理, Vol.10, No.4, pp.145-175, 2003.

[4] 三原, 藤井, 石川, “Web を用いたヘルプデスク指向の質問応答システム”, 言語処理学会第 11 回年次大会論文集, pp.1096-1099, 2005.

[5] 佐々木, 藤井, “取るべき行動と理由を提示するヘルプデスク指向の質問応答システム”, DEIM2010 A8-5, 2010.

[6] 岡本, 関口, 三末, 西野, “カスタマーセンター支援システム”, 人工知能学会誌, Vol.15, No.6, pp.1027-1034, 2000.

[7] S.Menard, “Applied Logistic Regression Analysis (Quantitative Application in the Social Sciences)”, Sage Publication, 2001.