

ヘルプデスク支援システムにおける言語事例検索 -全体構成-

1 P-4

高山 泰博 相川 勇之 鈴木 克志

三菱電機株式会社 情報技術総合研究所

1. はじめに

問い合わせ相談窓口業務に使う現在のヘルプデスク支援システムは、全文検索やデータベース検索等を提供しており、単純な検索や集計はできるが、テキストの分析・再利用には不十分である。単語の統計処理による文書分類・検索の手法では、文の意味を的確に捕らえられない。

我々は、対象領域や業務内容に依存した知識であるオントロジーを利用した言語解析や類似文照合により、文の構造の違いを意識した事例分類や推論の実現を目指している。言語事例分類により、業務や応用に依存したテキスト情報の事例全体の関係をあらかじめ問題解決木として表現しておく。オペレータが顧客対応する場合には、この問題解決木に対して自然言語で類似文検索を行なう。

2. 従来のヘルプデスクの問題解決機能の課題

従来システムは問い合わせの蓄積・管理中心で、問題解決機能は一部しか実装されていない。島津ら⁽¹⁾は問題解決機能を次の4つに分類している。

- (1) DB検索型（ユーザ主導のSQL検索）
- (2) 文書検索型（ユーザ主導の全文検索、概念検索）
- (3) 診断エキスパートシステム(ES)型(システム主導の決定木)
- (4) 事例検索型（双方向の事例索引検索）

現状のシステムは(1)、(2)を提供する。(3)の診断ES型は属性-属性値の形のデータを対象とする。類似する事例を利用して問題解決を図る事例ベース推論(Case-based Reasoning: CBR)を基礎とする(4)の事例検索型の従来システムに下記がある。

・CBR Express⁽²⁾：事例に付与した質問-回答対からシステムが質問を提示する。質問に回答する形で事例を絞り込めるが、大規模な事例では質問の選択が煩雑になり、質問-回答対の構築負荷が高い。

・HelpDesk Builder/Expert Guide⁽¹⁾：事例をペー

Natural Language Case-based Retrieval in Helpdesk Support Systems - Framework -
Yasuhiro TAKAYAMA, Takeyuki AIKAWA, Katsushi SUZUKI
Mitsubishi Electric Corporation.

5-1-1 Ofuna, Kamakura, Kanagawa 247-8501, JAPAN

ジという単位で蓄積し、階層的な目次と全文検索でたどる。生事例からページに変換するときの支援機能が課題である。

・ReMind⁽³⁾：自然言語の入力を、階層を定義した概念記号に置き換えてから2分決定木により事例を絞り込んだ後、最近隣法で類似度を計算するが、語順の違いに対応できないなど照合能力が弱い。

3. 言語事例推論による問題解決

前節の課題に対処するため、言語事例推論による問題解決機能を提供する。対象領域や業務依存の文の類似性判定には、否定や推量などの様相表現を処理する必要がある⁽⁴⁾。領域・業務依存の用語と用語間の関係の知識（オントロジー）を用いた類似文検索により事例の中から問題を特定する過程を言語事例推論と呼ぶ。我々のヘルプデスク支援システムの構成を図1に示す。

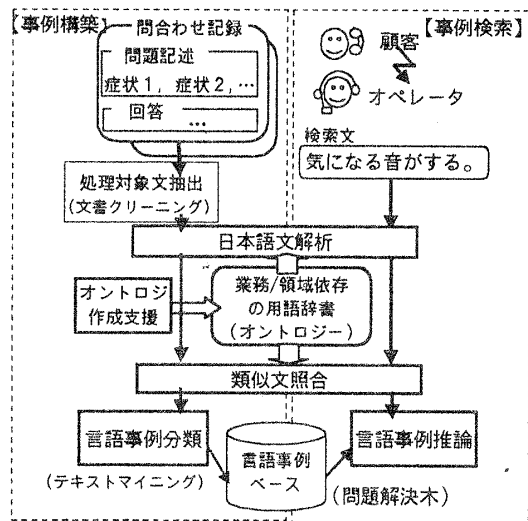


図1 ヘルプデスク支援システムの全体構成

図1の左側は事例構築処理である。処理対象文抽出部で問合わせ記録から文を抽出し、言語事例分類部で事例中の文をクラスタリングして言語事例データベース(DB)に格納する。図1の右側は事例検索処理である。検索文の解析結果と言語事例DB中の各文との類似文照合により事例を検索する。

3.1 言語事例ベースと問題解決木

(1) 文書クリーニング

問い合わせ記録の項目には、顧客名称などの書誌的な情報を含む定型フィールドと、質問や回答といった複数の自由記述文を含む非定型フィールドがある。処理対象文抽出(文書クリーニング)は、非定型フィールドのテキストを文ごとに分割し、文の種別ごとに【症状】、【原因】等の文タグと呼ぶラベルを付与して、分類が未決定の状態と言語事例ベースに格納する。

(2) 言語事例分類による問題解決木の構築

言語事例分類は、言語事例 DB 中の各文が属するクラスタを決定する。問題解決木は、注目した文タグを持つ文をキーワード・クラスタリングした後、更に類似文照合により一つのクラスタにまとめたものである。各クラスタは、関連付けられた事例の文、クラスタの表示ラベルからなる。図2は症状の文タグを持つ文をクラスタリングした問題解決木の例である。「気にかかる音がする」、「気になる音が聞こえる」が表示ラベル「気になる音」のクラスタにまとめられている。

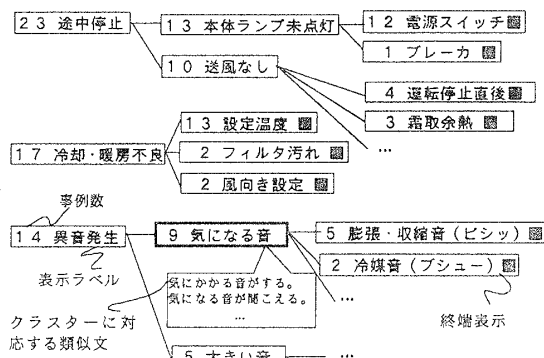


図2 問題解決木の例 (内容は説明用に作成したもの)

オペレータは GUI 上で問題解決木を検索し、指定したクラスタを展開/縮退しながらたどることで、対応中の問い合わせと類似した事例を探し出す。

3.2 言語事例推論に用いるオントロジー

ここでは、シソーラス(上位-下位関係を記述した IS-A 関係や部分-全体関係を記述した HAS-A 関係の記述)との違いでオントロジーを捉える。シソーラスは、全文検索の検索要求の展開などに使用される。これに対し、我々のオントロジーは、多様な言語表現の類似性を判定するため、シソーラスが持つ知識に加え、問題依存の格記述(用言と体言の間の関係を中心とする概念間の関係)辞

書、異なる依存関係に対応付ける言換え辞書(文脈依存で言換え可能な表現【例】「電話をかける」⇔「通話中」)を含む。オントロジーは、問い合わせ記録から複合語、特定の接尾辞との連続、係り受け用例等を抽出し、機械支援により作成する。

3.3 日本語文解析と類似文照合

言語事例推論や言語事例分類の中心は、オントロジーを用いた日本語文解析と類似文照合である。日本文解析は下記の処理で文の依存構造(係り受け構造)を作成する。

- ・記号列処理
- ・形態素解析
- ・未知語範囲推定
- ・文節処理(様相表現処理を含む)
- ・係り受け解析

形態素解析は、基本語辞書とオントロジーへのポイントを含む用語辞書を用いる。係り受け解析は、類似文照合に必要な文節間の関係として、一意な依存構造を求める。

類似文照合は、依存構造の類似性を判定する。類似度は完全に同一の表現同士を1とし、0から1の値をとる。類似度は、ノード間の類似性と構造の類似性と2つの観点から、ノード間および構造間の「異なり度合い」を計算する。ノード間の類似性には、表層の自立語の文字の一致、付属語の一致、IS-A、HAS-A階層上の距離、様相表現などの属性を用いる。構造の類似性は、係り受けの類似度⁽⁵⁾により求める。

4. まとめ

対象領域や業務ごとの問い合わせ記録から詳細な情報を取り出すために、領域・業務依存の用語と用語間の関係の知識であるオントロジーを利用した言語解析や類似文照合で言語事例推論を行ない、問題解決を支援するアプローチを提案した。このアプローチでは、テキスト中の類似した文をクラスタリングして作成した問題解決木に対して、自然言語による類似文検索を行なう。

参考文献

- (1)島津秀雄,伊藤慎:"ヘルプデスク支援システムの最新動向",情報処理, Vol.39, No.9, pp.912-917(1998).
- (2)Kriegsman, M, Barletta, R. : "Building a Case-based Help Desk Application, IEEE EXPERT, pp.18-26(1993).
- (3)黒川高光:"CBR機能を組み込んだART-IMとドメインシエール CBR Express",日経AI別冊 1992 冬号, pp.130-139 (1992).
- (4)乾裕子,他:"モダリティ分析に基づく自由回答アンケートの分類",言語処理学会第4回年次大会, pp.540-543 (1998).
- (5)相川勇之,他:"ヘルプデスク支援システムにおける言語事例検索-類似文照合-",第59回情報処理学会全国大会, 1P-05(1999).