

言語事例推論型ヘルプデスク支援システム

8 L - 0 4

鈴木 克志 相川 勇之 高山 泰博

三菱電機株式会社 情報技術総合研究所

1. はじめに

事例推論に基づくヘルプデスク支援システムでは、過去の問題と回答を事例として事例ベースに格納しておき、検索することによって専門家の判断に匹敵する推論効果を得ることを目的とする。しかし、事例ベース構築が業務に依存し作業負荷が大きいことや、事例の照合の有効性が不明確で導入効果が算出しにくい等の課題があり、実際の業務で広く使用されるに至っていない。

この課題の解決に向けて言語事例推論方式を開発中である。本報告では基本原理を説明する。事例ベースの構築を容易に行なうために、オペレータが業務中に蓄積するテキストそのものを言語事例として階層的に問題解決木の形で自動分類し、推論に利用する。質問入力により問題解決木を探索し、対話的に事例を絞り込めるのが特徴である。

2. ヘルプデスク支援システムの課題

多くの事例ベース推論ツールでは、事例を属性と属性値のペアの集合で記述し、事例間の類似度の定義により、最近隣法にもとづく事例の類似照合を行なう⁽¹⁾。この場合、属性間の重み付けが事例推論結果に大きな影響を与えるので、現実の問題に対し属性を如何に記述し事例ベースを構築するかが課題となる。

テキストにより記述した事例を用いる種類の事例推論システムに、Remind⁽²⁾、CBR-Express、岡本⁽¹⁶⁾がある。Remind はテキストから構成した診断決定木を入力テキスト中の単語で検索する。単語単位の索引のため語順の取り扱いがなされておらず、わずかな語順の違いにより診断結果が異なる事例が多く含まれる場合、検索精度悪化の恐れがある。CBR-Express は事例に対して質問と回答を記述しユーザの質問選択と回答の操作を反復することで対話的に事例推論を行なうが、事例に対して注意深く質問と回答を記述する必要があり、事例構築の負荷が高い。岡本⁽³⁾では、事例テキストと意味カテゴリとの対応から学習したキーワードの重みを類似度計算に用いる。意味カテゴリを付

加するだけなので事例構築負荷が小さいが、個々の事例テキストと類似度計算を行なうので、事例の大規模化が難しい。また、一括検索でありオペレータの対話的使用を考慮していない。更にExpertGuide⁽⁴⁾のような検索型システムもあるが、知識を目次の形で構成する必要がある。

3. 言語事例推論の方式

3.1 概要

提案する言語事例推論方式の概要を図1に示す。

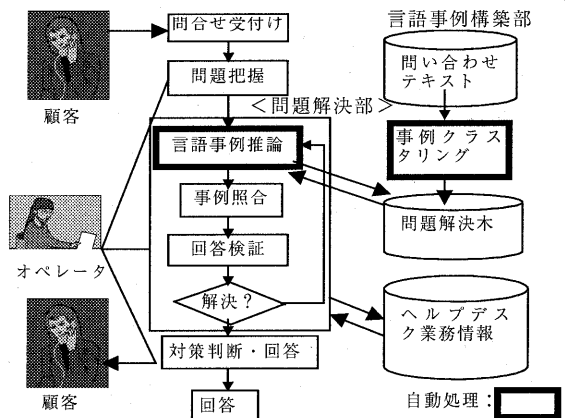


図1. 言語事例推論に基づくヘルプデスク支援システム

3.2 問題解決木の自動構築

事例クラスタリング(図2:次頁)は、同じ意味を表わすグループに分類した問い合わせテキストのサンプルを学習し、問題解決木を自動構成する。

順に以下の処理を行なう。

(1) 意味カテゴリの付与

事例テキストを手手で分類し、分類ごとに意味カテゴリを付与する。この部分は応用に依存する。

(2) 意味カテゴリ階層分類の作成

まず、意味カテゴリに対してキーワードベクトルを作成する。キーワードベクトルは、事例テキスト集合の各文を文解析して求めたキーワードから、意味カテゴリごとに出現頻度の χ^2 乗値を求め、ある閾値(Th1)以上の値を持つキーワードのみを選んで要素としたベクトルとする。キーワードは名詞、動詞等の自立語、及び「ない(否定)」など意味カテゴリの分別精度に貢献する可能性のある付属語とする。全出現単語からベクトルを作成すると事例数増大にともないベクトルが

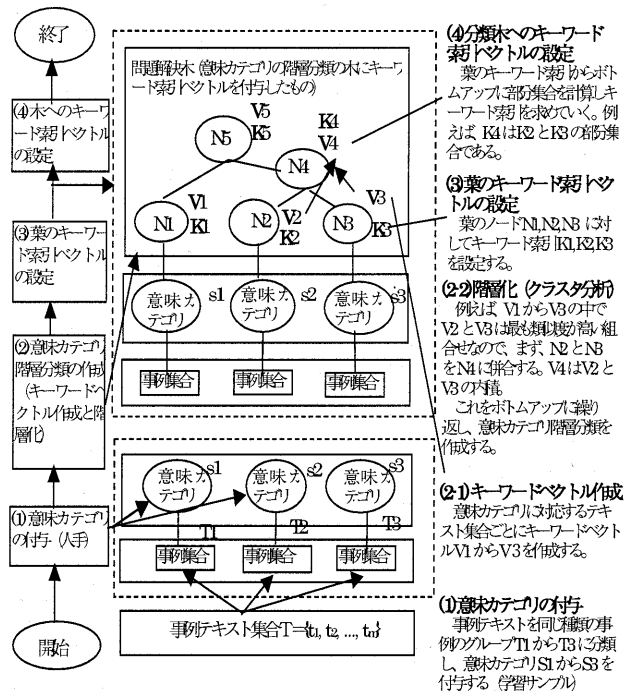


図2. 問題解決木の自動構築 (事例クラスタリング)

大きくなるので、閾値 $Th1$ により精度を比較的落とさずにベクトルを小さくする効果を狙う。

次に、キーワードベクトル間の類似度計算を用いて、もっとも類似する意味カテゴリ同士を一つのカテゴリにまとめ、同様の処理を繰り返し、意味カテゴリを階層的に分類する。キーワードベクトルの類似度は内積により定義する。

(3) 葉に対するキーワード索引ベクトルの設定

さらに、意味カテゴリ階層分類の木をボトムアップになぞりながら、キーワード索引ベクトルをノードに設定していきキーワード分類木を作成する。最初に木の末端にある葉ノードに対してキーワード索引ベクトルを求める。

まず、葉 i のキーワードベクトル $V_i = \{v_{i1}, v_{i2}, \dots, v_{ik}\}$ の各要素と同じ文内に出現 (共起) し係受け関係にあるキーワード共起をすべて抽出する。共起ごとに意味カテゴリに対する出現頻度を求め、 χ^2 乗値がある閾値 ($Th2$ とする) 以上の共起からなる共起ベクトル $\{c_1, \dots, c_m\}$ を作成する。そして、葉のキーワードベクトルと共起ベクトルの合成 $K_i = \{v_{i1}, v_{i2}, \dots, v_{ik}, v_{ik+1}, \dots, v_{ik+m}\}$ を葉 i のキーワード索引ベクトルとする。ただし、 $v_{ik+1}, \dots, v_{ik+m}$ はそれぞれキーワード共起 c_1, \dots, c_m の意味カテゴリ i に対する出現頻度の χ^2 乗値である。 $Th2$ は $Th1$ と同様にベクトルの要素を選択するときのパラメータであるが、 $Th1$ はキーワード単独の出現をベクトルの要素として選択するときの調整パラメータであるのに対し、 $Th2$ はキーワード共起の選択の調整役を果たす。

(4) キーワード索引ベクトルの設定

同じ親を持つ葉ノードのキーワード索引ベクトル間の共通要素を親のキーワード索引ベクトルとする。さらに木の上位に向けて操作を繰り返す。

3. 3 言語事例推論方式

一括推論と対話推論の2つの方式がある。

(1) 一括推論

オペレータが入力した新規事例テキストから構文解析によりキーワードとキーワード共起を抽出しキーワード索引ベクトルの対応する要素が入力テキスト中に出現すれば1、さもなければ0を取る入力ベクトルを作成する。入力ベクトルと高い類似度 (ベクトル間の角度の余弦値) を持つ意味カテゴリを問題解決木から求め、推論結果とする。

従来はキーワードの集合を入力し事例テキストと照合し最も類似度の高い事例テキストの属する意味カテゴリを推論結果とする方式⁽¹⁶⁾であった。従って、事例ごとに類似度計算が必要である。一方、本方式では意味カテゴリのキーワード索引ベクトルとの類似度のみ計算するので効率が良い。

(2) 対話推論

対話推論の動作は、オペレータが入力したテキストから作成した入力ベクトルと類似度の高いキーワード索引ベクトルを持つノードを検索する。一括推論と異なり、検索対象を葉に限定しない。検索されたノードに対して、ガイダンスメッセージを表示しオペレータが子供ノードを選択可能とする。この操作を繰り返し、オペレータが意味カテゴリに到達するのを支援する。検索結果ノードが複数ある場合は、システムは類似度の高いノードを選んでガイダンスを行なう。これによりオペレータが少数のキーワード入力で木の上位から回答を絞り込んでいく対話的支援機能が実現できた。

4. まとめ

以下の特徴を持つ言語事例推論方式を提案した。

- (1) 事例ベース (問題解決木) の自動構築
- (2) 共起のベクトル化により照合精度を向上
- (3) 問題解決木によるオペレータの対話的支援

今後は実用化を進め、業務効果を確認していく。

参考文献

(1) I.D. Watson: Applying Case-Based Reasoning : Techniques for Enterprise Systems, Morgan Kaufman Publishers(1997).
 (2) Mark Kriegsman, Ralph Barletta : Building a Case-Based Help Desk Application, IEEE Expert, December 1993.
 (3) 岡本青史, 佐藤健, 塙順吉, 松尾勝美: 類似事例検索システムー通信ソフト故障診断問題への適用ー, 情報処理学会第51回全国大会, 7J-4(1996).
 (4) 柴田, 他: 診断型 WWW 情報検索システム Expert Guide, 1998 年度人工知能学会全国大会予稿集(1998).