

事例に基づく定性的知識から定量的知識への変換手法の検討

2N-5

松本 一則 清水 徹 横田 英俊 山口 典男 橋本 和夫
国際電信電話株式会社

1. はじめに

障害診断では、領域専門家の知識が「もし属性 X の値が大きく、属性 Y の値が普通ならば、故障 A が考えられる」等のように、定性的に表現され蓄積されている場合が多い。この定性的知識を直接利用して診断を行なうためのシステムは、一般に図 1 に示す構成をとる。定量的知識は、診断対象からの定量値を定性値に変換する写像関数と、与えられた定性値から故障原因を出力する定性的知識とを直列に組み合わせて構成する。

図 1 の構成の下で、診断事例に基づいて定量的知識を最適化する方式としては、以下の方式が考えられる。

- (1) 写像関数を固定し、定性的知識を最適化する
- (2) 定性的知識を固定し、写像関数を最適化する

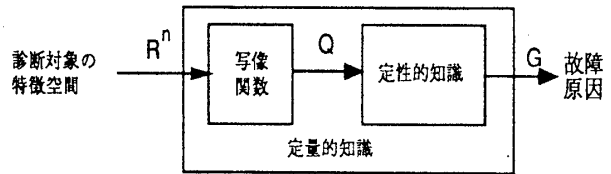
(1) は、観測値が定常確率過程となるプロセス診断のように、一旦適切な写像関数を求めた後はその修正の必要がない場合に適した方式であり、これに属する具体的手法としては、定性シミュレーションの結果と現実が合致するよう定性モデルを修正する手法^[1]等が提案されている。

(2) は、観測値が非定常確率過程となる通信路の診断のように、一旦信頼性の高い定性的知識を得た後はその知識を修正する必要がない場合に適した方式である。しかし、定量値を定性的知識が必要とする粒度の定性値へ変換する写像関数を最適化するための具体的手法はこれまでほとんど検討されておらず、応用事例についても報告されていない。

(2) の診断事例に基づく写像関数の最適化方式は、写像関数が診断事例に矛盾しないための制約条件を導出する問題と、導出した制約条件下での写像関数の最適設計問題とに分けることができる。本稿では、対象とする写像関数について述べ、写像関数の制約条件の導出問題を定式化する。次に、同制約条件下で写像関数を設計するため、エントロピーを用いた評価関数を提案する。

2. 対象とする写像関数

文献^[2]では、特徴空間の分割の方法を以下のように分類している。



R^n : 実数 (定量値) Q : 記号もしくは記号列 (定性値) G : 記号

図 1: 定性的知識を利用する診断システムの構成

分割法 1(境界標による分割) 特徴空間を各座標軸に垂直な超平面で分割する。

分割法 2(超平面による分割) 特徴空間を任意の超平面によって分割する。

分割法 3(非線形な制約による分割) 非線形の制約によって特徴空間を分割する。

分割法 1、分割法 2、分割法 3 の順に特徴空間の分割能力は高くなるが、適切な分割を得るために調整しなければいけないパラメータの数も順に増え、調整に必要な事例の数も増える。

本稿では、少ない事例から写像関数を最適化する場合に適しているという観点から、分割法 1(境界標による分割) に基づく写像関数を扱うこととする。

3. 写像関数の制約条件の導出

診断事例から写像関数が満たすべき制約条件を導出する問題を定式化する。以下では、記法とモデルを説明し、同問題を解くプログラムの入力、変数間の制約、出力内容を示す。

3.1 記法およびモデル

事例は定量的な n 個の属性で表現する。事例の総数は m とし、個々の事例を、 C_1, C_2, \dots, C_m で表わす。事例 C_j の i 番目の定量的な属性の値を v_i^j ($1 \leq i \leq n, 1 \leq j \leq m$) で表わす。

診断の結果得られる故障原因の集合を G で表わし、事例 C_j が診断されるべきクラスの名前を g_j ($g_j \in G$) で表わす。事例 C_j に対して、クラス名 g_j は一意に定まるとする。

i 番目の定量的属性がとり得る値域を l_i 個の境界標 $\theta_1^i, \theta_2^i, \dots, \theta_{l_i}^i$ で分割する。この時、 $\theta_{s-1}^i < \theta_s^i$ ($1 \leq s \leq l_i$) であるものとし、分割でできた領域を $q_1^i, q_2^i, \dots, q_{l_i}^i$ ($\theta_{s-1}^i \leq x < \theta_s^i$ ならば、 x は q_s^i に属する) で表わす。表記の利便上、 $\theta_0^i = -\infty$ ($1 \leq i \leq n$) とする。これらの領

“A Proposal for Case-Based Transformation from Qualitative to Quantitative Knowledge”, Kazunori MATSUMOTO, Tohru SHIMIZU, Hidetoshi YOKOTA, Norio YAMAGUCHI and Kazuo HASHIMOTO: Kokusai Den-shin Denwa Co. Ltd.

域を表わす記号を、「 i 番目の属性に対する定性的カテゴリ」と呼び、 i 番目の属性に対する定性的カテゴリの集合を Q_i とする。

定性的知識は、 $Q_1 \times Q_2 \times \dots \times Q_n$ から G への写像関数であり、定性的知識によってクラス $g (g \in G)$ に分類される $Q_1 \times Q_2 \times \dots \times Q_n$ 上の点の集合を $Q(g)$ で表わす。また、 $Q_1 \times Q_2 \times \dots \times Q_n$ 上の点 ϕ の i 番目の属性に対する定性的カテゴリが、 $q_{cat_i(\phi)}^i$ で表現できるような関数 cat_i を定義する。

3.2 制約条件の導出プログラム

以下に示す入力、変数間の制約、出力を持つプログラムが、写像関数の制約条件導出問題を解くプログラムとなる。

入力 (事例と診断結果)

全ての $i, j (1 \leq i \leq n, 1 \leq j \leq m)$ に対し、 v_i^j, g_j を入力する。

これは全事例に対する全属性の値 v_i^j とその事例の正しい診断結果のクラス g_j を与えることに相当する。

変数間の制約 (事例と診断結果と境界標が満たすべき関係)

全ての $i, j (1 \leq i \leq n, 1 \leq j \leq m)$ に対し、ある $\phi \in Q(g_j)$ が存在し、 $\theta_{cat_i(\phi)-1}^i \leq v_i^j < \theta_{cat_i(\phi)}^i$ が成り立つ。

これは、定性的知識への入力候補となる記号列 ϕ がうまく定量化されるための条件である。

出力 (写像関数の制約条件)

写像関数の制約条件を、境界標

$\theta_1^1, \dots, \theta_{l_1}^1, \theta_1^2, \dots, \theta_{l_2}^2, \dots, \theta_1^n, \dots, \theta_{l_n}^n$ に関する $\sum_{i=1}^n l_i$ 元連立一次不等式として出力する。

4. エントロピーを用いた評価関数

前節 3.2 で得られる写像関数の制約条件だけでは一意に写像関数を決定することができない。そこで、定量的な属性を定量化する際に失う情報量 (エントロピー) が最小になる写像関数が最適であると定義し、「各属性が独立で属性間に相関がない」という仮定の下で、エントロピーを用いた評価関数を提案する。

i 番目の属性の s 番目の領域 q_s^i において、診断結果のクラスが g となる領域 q_s^i 内の事例数を $r(q_s^i, g)$ とし、 q_s^i 中の事例の診断結果が g となる確率 $P(q_s^i, g)$ 、および q_s^i のエントロピー $H(q_s^i)$ を、それぞれ式 (1)、(2) により定義する。

$$P(q_s^i, g) = \frac{r(q_s^i, g)}{\sum_{g' \in G} r(q_s^i, g')} \quad (1)$$

$$H(q_s^i) = - \sum_{g \in G} P(q_s^i, g) \log P(q_s^i, g) \quad (2)$$

式 (2) を用いて、写像関数の評価関数 H を式 (3) で定める。

$$H = \sum_{i=1}^n \sum_{s=1}^{l_i} H(q_s^i) \quad (3)$$

式 (3) で定義した H を用いることで、定量化で失う情報量を最小にする分割を各属性毎に独立に見つけることができる。

属性間に相関がないとする仮定は、一般的には成り立たないため、式 (3) の評価関数の有効性には限界がある。しかし、通信網管理での回線群 (circuits group) の使用状況監視に応用する場合、通常状態と初期障害の状態において、部分的な設備障害と疎通要求等の属性間に相関がほとんどないため、式 (3) の評価関数は有効である。

5. おわりに

本稿では、事例に基づく定量的知識の最適化方式として十分に検討されていなかった「定性的知識を固定し、写像関数を最適化する」方式について検討し、以下の結果を得た。

- 特徴空間を境界標により分割するものと仮定した上で、写像関数の境界標に関する制約条件を導出する問題を定式化した。この仮定の下では、写像関数の境界標に関する制約条件は、 $\sum_{i=1}^n l_i$ 元連立一次不等式として得られる。
- 上記で求めた制約を満たす写像関数は一意に決定できない。そこで、属性間に相関が無い場合を仮定し、定量化によって失われる情報量を最小にするためのエントロピーを用いた評価関数を提案した。

本稿で提案した評価関数は、通信網管理の回線群の使用状況監視等の適用領域において有効である。

最後に、日頃御指導頂く KDD 研究所浦野義頼所長、眞家健次次長、御討論頂いた浅見徹ネットワークエンジニアリング支援グループリーダーならびに小花貞夫網管理グループリーダーに感謝します。

参考文献

- [1] 石田好輝: 定性推論を利用したプロセス診断知識獲得-定性的シミュレーションによる診断ルールの生成-, 電子情報通信学会論文誌, Vol. J72-D-II No.3, pp.416-426 (1989)
- [2] 西田豊明: 定性推論の基礎. 人工知能学会誌, Vol.4, No.5, pp.522-527 (1989).