

ルール生成に基づくデータマイニングのLSI不良解析への適用

2C-6

牧 秀行* 前田 章* 内田明久† 中島康博‡

(株)日立製作所システム開発研究所* (株)日立製作所デバイス開発センタ†
日立コンピュータエンジニアリング(株)‡

1 はじめに

計算機の処理、通信速度の向上、記憶容量の増大により、大量のデータの収集、蓄積が可能となってきている。しかし、データを活用する技術が不十分で、蓄積されたデータが死蔵されている状況にある。そこで、大量のデータから特徴や知識を抽出するデータマイニング技術が注目されている。

本稿ではデータマイニングの一手法として、データの特徴を抽出、記述するルール生成手法について説明し、その適用事例として、LSI不良解析への適用について述べる。

2 ルールによる特徴の記述

複数の属性を持つ事例の集合を対象データとし、データの特徴をif-thenルール形式で記述する。if部、then部に記述されるのは「属性に関する条件」で、データの特徴とは属性に関する条件間の関係である。以下に、ルール生成の詳細を説明する。

2.1 カテゴリ化と条件の定義

まず必要なのは「属性に関する条件」の定義である。対象データの属性の型はいろいろ考えられるが、ここでは最も一般的と思われる数値属性を中心に述べる。数値属性の場合、最も簡単な条件定義は、値の範囲に応じて「大」「中」「小」「3.0以上」「3.0未満」などのカテゴリに分類する方法である。これにより、例えば「測定電圧が大」というような条件が定義される。非数値属性の場合、カテゴリ化の方法は異なるが、適当な方法で属性に関する条件を定義すれば、あとは数値属性の場合と同様に扱うことができる。

An Application of Data Mining Based on Rule Generation to LSI Failure Analysis

Hideyuki MAKI*, Akira MAEDA*, Akihisa UCHIDA†, Yasuhiro NAKAJIMA‡

Systems Development Laboratory, Hitachi, Ltd.*

Device Development Center, Hitachi, Ltd.†

Hitachi Computer Engineering, Co., Ltd.‡

2.2 ルールの評価尺度

上で定義された属性に関する条件の組合せからif-thenルールを作る(図1)。ここで、ルールの良さを評価するための尺度が必要となる。

「if A then B」というルールを考えたとき、その評価尺度 μ を次のように定義する。

$$\mu = P(A)^\alpha \log \frac{P(B|A)}{P(B)}$$

$P(A)$ 、 $P(B)$ はそれぞれ、ある事例が条件A、Bを満たす確率、 $P(B|A)$ は事例が条件Aを満たすという条件の下で条件Bを満たす確率である。また、 α は0~1の範囲の実数値で、一般性パラメータと呼ぶ。ルールの評価値は、ルールがどの程度の数の事例を説明しているかという「カバー率」と、属性間の関係をどの程度正確に説明しているかという「精度」のトレードオフである。一般性パラメータはこのトレードオフを調節するもので、値が大きいほどカバー率を重視した評価尺度となる。

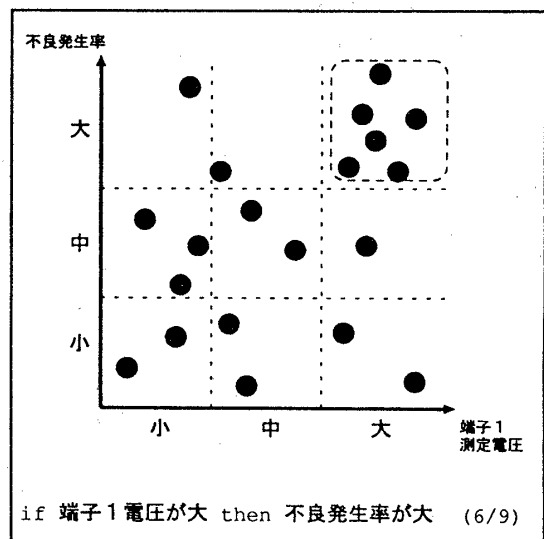


図1: データの分布とルールの例

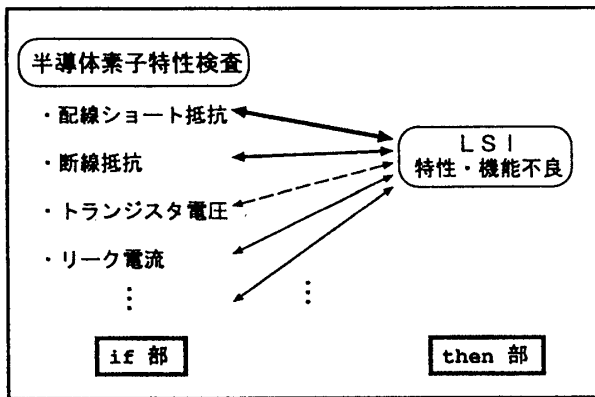


図 2: 関連の大きい項目の探索

2.3 条件の指定とルールの生成

データ解析では、何か着目している事象があり、その事象に関するデータ上の特徴を知りたいという場合が多い。この着目する事象を then 部条件として指定する。また、着目する事象との関連を調べたい種々の項目を if 部条件として指定する (図 2)。ルール生成処理では、指定された種々の if 部条件と、then 部条件の組み合わせについてルールを生成し、その評価値を算出する。評価値の大きいルールほど、データの特徴をよく表していると見なす。実際の処理では、大量データに対応するために評価値やカバー率の小さいルールを切り捨てるなどの枝刈りを行う場合もある。

3 LSI 不良解析への適用

以上で説明したルール生成手法を LSI 不良解析に適用した。LSI は非常に微細で複雑な回路が作り込まれており、構成要素の 1 つ 1 つを検査することはできないので、製品に不良が発生した場合、その原因は電気特性などから間接的にしか知ることができない。したがって、不良原因を特定するためには、まず不良発生と関連の強い電気特性検査項目を探索する事から始める。この項目の探索にルール生成手法を適用した。

適用例として、LSI の電氣的機能不良の解析結果を示す。各項目は全て数値データで、大、中、小の 3 つのカテゴリに分けた。着目するのは「機能不良率が大きい」という事象で、これを then 部条件に指定した。ルール生成の結果、評価値の大きい順に、表 1 に示すようなルールが得られた。n1 層、n2 層などは、配線層の呼び名である。1 番目のルールを例にとると、if 部条件を満たす事例数が 37、そのうち、さらに then 部条件をも満たす事例数が 31 である (表 1 中の「事例数」欄)。全データ数は 323 であり、したがって、カバー率は $37/323 = 11.46\%$ 、精度は $31/37 = 83.78\%$

表 1: 生成されたルール

	ルール	精度	事例数
1	if n2 層配線ショート抵抗が小 then 機能不良率が	83.78%	31/37
2	if n1 層配線ショート抵抗が小 then 機能不良率が	70.97%	22/31
3	if n3 層配線オープン抵抗が大 then 機能不良率が	61.33%	157/256
4	if n3 層配線ショート抵抗が小 then 機能不良率が	80.00%	4/5
...

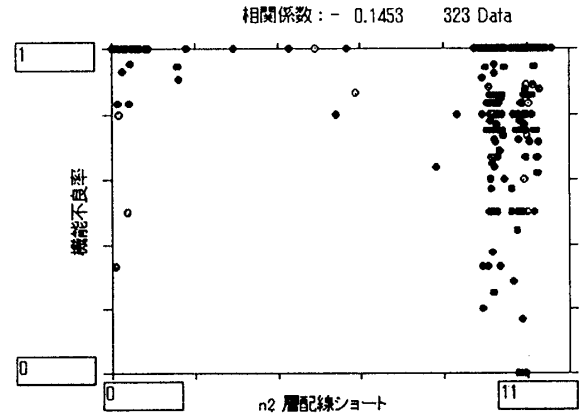


図 3: 機能不良率と n2 層配線ショート抵抗の関係

(表 1 中の「精度」欄) となる。

ルール生成によって不良発生と関連の強い検査項目の候補が挙がったら、次にこれらの検査項目と不良率の関係をグラフを用いてさらに詳しく調べる。図 3 は、n2 層配線ショート抵抗と機能不良率をプロットした散布図である。これを見ると、ルールにあるように、配線ショート抵抗が小さいほど不良率の大きい事例の占める割合が大きくなっており、n2 層配線ショート抵抗と機能不良率の関連が強いことがわかる。

以上から、機能不良の原因は n2 層配線にありそうだと推測できる。こうして不良原因の候補をある程度絞り込んだら、これを手がかりにして LSI の実際の物理的の不良箇所を詳しく観察する。電子顕微鏡による観察を行ったところ、当該配線層でショート箇所が発見され、ルールの指摘が正しいことが確認された。

4 おわりに

大量データを対象とし、データの特徴を抽出する手法としてルール生成に基づくデータマイニング手法を開発した。これを LSI 不良解析に適用したところ、不良原因の発見、対策に有効であるとの評価を得た。

参考文献

[1] 芦田、前田、高橋：
データマイニングにおける特徴的ルール生成方式、
情報処理学会第 50 回全国大会 3-19