

## 連関分類器の予測過程に対するインタラクティブ可視化ツール

池田武史<sup>†</sup> 前田健太郎<sup>†</sup> 亀谷由隆<sup>†</sup>

<sup>†</sup>名城大学 理工学研究科 情報工学専攻

### 1 はじめに

機械学習手法を説明責任が伴う問題に使用する場合、モデルや予測結果に解釈・説明可能性が求められる。そのため、パターン発見手法に基づく連関分類器 (associative classifier) などの If-Then ルール形式のホワイトボックス的な手法が注目を集めている。しかし、個々の事例に対してどのルールが条件を満たし、どのように分類が行われるかを知るのそれはほど容易ではない。本研究では、発展を続ける Web フロントエンド技術を活用し、識別パターンを用いた連関分類器による予測過程をインタラクティブに可視化するツールを提案する。

### 2 研究の背景

本研究では、離散値・連続値を同時に扱える識別パターン発見手法である ECHO (Exhaustive Covering in Hybrid dOmain)[1] を用いた連関分類器に対する分類過程の可視化を行うツールを提案する。そしてその有用性を示す使用例として、誤分類したトランザクションについて、インタラクティブに可視化し、エラー分析を行った場合を示す。

#### 2.1 ECHO

従来、識別パターン発見においては、離散値からの識別パターン発見しか行えないことが多かった。そのため、連続値は各属性を離散化させる必要があるが、その際に情報損失が発生するという問題があった。ECHO は著名な頻出パターン発見手法である FP-growth を拡張することで連続値を含むデータからの識別パターン発見を可能とする手法である。また、決定木を含めた初期のルール形式の分類器は連続値・離散値を同時に扱うことができるが、これらの手法の多くはヒューリスティクスを用いた貪欲的な手法であり、識別パターンに比べ網羅性が低く、解釈・説明可能性の観点からは不十分であると考えている。

#### 2.2 ECHO を用いた連関分類器

ECHO によって得られた識別パターンを用いて、連関分類器を構築できる。提案手法では、識別パターンがカバーする事例数の積集合を使用する手法 [2] を簡易化したものを使用する。この手法では正負クラスから得られるパターンを両方使用し、トランザクションを以下に示すケースに当てはめることで分類を行う。本実験における分類手順の概要図を図 1 に示す。

- Case 0** カバーするパターンが存在しない場合 (多数派クラス)
- Case 1** カバーするパターンが 1 つ存在する場合 (カバーしたクラス)
- Case 2** カバーするパターンが 2 つ以上存在し、それらをか  
カバーするクラスが競合していない場合 (カバーしたクラス)

- Case 3** カバーするパターンが競合し、積集合に含まれる事例  
が 1 個のみの場合 (積事例が示すクラス)
- Case 4** カバーするパターンが競合し、積集合に含まれる事例  
が 2 個以上の場合 (クラス比を考慮して決定)
- Case 5** カバーするパターンが競合し、積集合に含まれる事例  
が 0 個の場合 (多数派クラス)

Case 4 に当てはまったトランザクションは、積集合に含まれる事例数に全体のクラス比を重み付けし、少数派クラスの希少性に考慮した分類を行う。

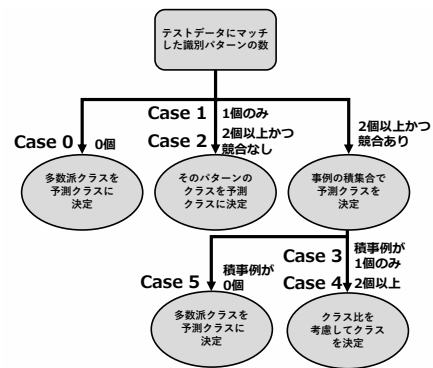


図 1: ECHO を用いた連関分類器の分類手順

### 3 提案ツール

可視化ツールは JavaScript (ユーザインターフェース構築に React を使用) で実装され、Web ブラウザ上で動作する。その有用性を示すために、以下ではエラー分析における使用例を示す。そこでは多くの属性を持つトランザクションでの使用を想定しているが、紙面の都合上 Iris データの誤分類の場合を扱う。Iris データセットは 3 種類のあやめのデータを含むものであるが、本研究では問題の単純化のために Iris-Versicolor とそれ以外 (Other) の 2 クラスに限定し、2 値分類データセットとして扱う。このデータを訓練データ 105 件とテストデータ 45 件に分割し、訓練データから得た識別パターンを用いてテストデータに対する分類を行う。本研究で提案する可

#### 3.1 可視化画面の構成

提案手法を用いて分類した結果の混同行列は図 4 の画面に表示され、ここで誤分類したトランザクションの番号を確認できる。また、全てのテストデータがどのケースに当てはめられて分類されたのかを可視化した画面を図 3 に示す。これによって、それぞれのケースに当てはめられたトランザクションのカバーしたパターンの数及び、事例の積集合の数の確認が容易になる。図 2 は対象としたトランザクションの分類過程の詳細を確認する画面である。対象としたトランザクションは図 2 の下部に常に定位置で表示されるため、クラス等を確認しながらパターンとの比較を行うことができる。図 2 の左側に得られた識別パターンが表示され、対象となる識別パターンをカバーしているパターンはテーブルのセルの色が変わることで容易に確認

An Interactive Tool that Visualizes the Process of Associative Classification

<sup>†</sup> Takeshi Ikeda

<sup>†</sup> Kentaro Maeda

<sup>†</sup> Yoshitaka Kameya

Division of Information Engineering, Graduate School of Science and Technology, Meijo University (<sup>†</sup>)

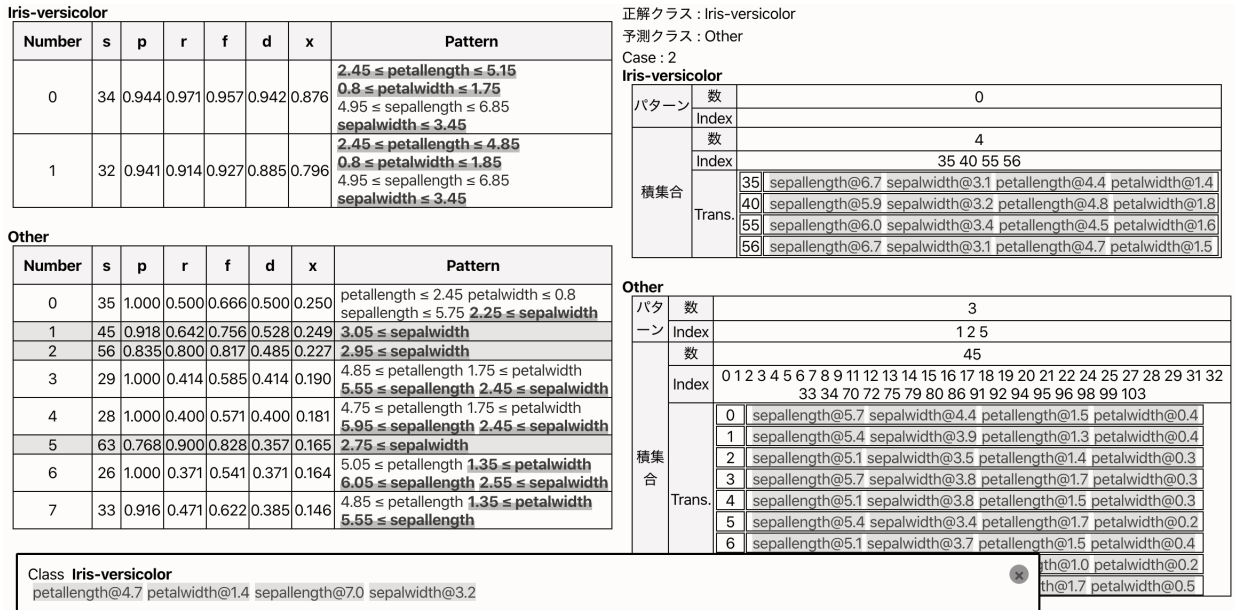


図 2: 誤分類したトランザクションの可視化過程

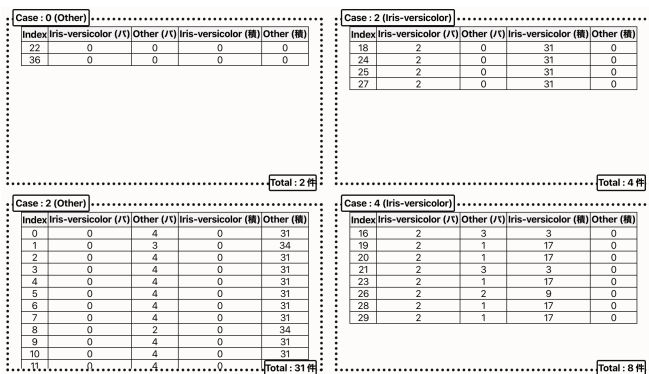


図 3: テストデータが該当するケースの表示の様子

することができるようになってきている。また、パターンの条件の一部を満たす場合、その条件がマーカ等で強調され、トランザクションとの比較を容易にしている。図 2 の右側では、分類結果と各クラスのカバーしたパターン数・積集合の事例数とそのトランザクションを確認できる。

### 3.2 エラー分析

図 4 の混同行列のセルをマウスでホバーすることにより、15,17,22 番のデータが Iris-Versicolor のデータに対して Other と誤分類していることがわかる。今回は 15 番に対してエラー分析を行うため、この中から 15 番のデータを選択する。図 2 は選択した 15 番のトランザクションの分類過程を可視化したものである。これによって、対象のトランザクションが Iris-Versicolor のパターンを 1 つもカバーせず、Other のパターンを 3 つカバーし、Case 2 の事例に当てはまり Other と誤分類していることがわかる。ここで、対象のトランザクションをカバーした Other のパターンに注目する。これらのパターンは全てパターン長が 1 であり、他のパターンと比較し、再現率は高いものの



図 4: 混同行列から誤分類したトランザクションをポップアップ表示にて特定する様子

適合率が低いパターンであることがわかる。そのため、このパターンは他パターンと比べ信頼性が低いと推測され、図 4 で確認した他の誤分類した番号のトランザクションとパターンを比較することでそれを確認することができる。以上のように識別パターンを用いた連関分類器の対話的なエラー分析が容易となり、信頼性の高い分類器の構築が可能となると考えられる。

### 4 おわりに

本研究では、ECHO を用いた連関分類器を使用し、予測過程を対話的に可視化するツールの提案を行った。このツールによって、ホワイトボックス的な分類手法の分類過程をより理解しやすいものとした。今後は、アニメーション等を用いることで使用者の体験を改善し、より可視化過程を理解しやすい UI の作成を目指したい。また、提案手法を用いて後から人手でパターン集合を修正できるような仕組みも検討していきたい。

### 参考文献

- [1] 亀谷由隆: 数値を含むデータからの効率的なパターン発見に向けて, FIT-18 予稿集 (2018).
- [2] K. Maeda and Y. Kameya: Associative classification using Common Instances among Conflicting Discriminative Patterns, Proc. of the 2019 Int'l Conf. on Technologies and Applications of Artificial Intelligence (TAAI-19), 2019.