

インタラクティブデータマイニング手法の提案

- アンケート調査解析への適用 -

山田和明[†] 山本恭裕[†] 中小路久美代[†] 上田完次[‡]

[†] 東京大学先端科学技術研究センター

[‡] 東京大学人工物工学研究センター

1 はじめに

情報技術の発達により顧客の購買履歴, Web の遷移履歴, オンライン教育における学習履歴など, 新しいビジネスチャンスを広げる有益なユーザ情報が大量に蓄積できるようになり, マーケティングなどへの利用が期待されている [1].

これらのユーザ情報は, 大規模かつ重要度の異なる情報が混在し, 流行に左右される購買履歴のように時間の経過とともに情報の意味や重要性が容易に変化する. また, データ解析を通して現状を把握すれば新しい商品開発や販売戦略が決定できるという単純な問題ではない. したがって, 意思決定者は仮説の形成と検証を繰り返しながら試行錯誤的に意思決定を行う必要がある.

そのため, 上述のような大規模で動的に変化するデータを客観的に解析することは困難であり, むしろ, エキスパートが持つ問題領域に依存したドメイン知識や経験則を利用して解析を進める必要がある. そこで本研究では, エキスパートがデータの解析結果を参照しながら対話的に仮説形成から仮説の検証までを一貫して行える意思決定支援システム環境の構築を目的とする.

本稿では, まず, 問題領域の特徴について述べ, 提案するインタラクティブデータマイニングの概要を説明する. そして, 支援システム環境を構成する「仮説形成支援システム」と「仮説検証支援システム」の説明を行う.

2 問題領域の特徴

本研究が対象とする問題領域の特徴として次のようなものがある.

Interactive Data Mining - Application to Questionnaire Analysis -

Kazuaki YAMADA[†], Yasuhiro YAMAMOTO[†], Kumiyo NAKAKOJI[†] and Kanji UEDA[‡]

[†] RCAST, The University of Tokyo 153-8904, Tokyo, Japan

[‡] RACE, The University of Tokyo 153-8904, Tokyo, Japan

{yamada, yxy, kumiyo}@kid.rcast.u-tokyo.ac.jp

ueda@race.u-tokyo.ac.jp

- (1) マーケティングにおける意思決定過程は, 現状把握だけでは目標や戦略を一意に決定することができない. そのため「現状把握」「目標設定」「戦略規定」がそれぞれ相互に作用するために試行錯誤的な意思決定過程である.
- (2) ユーザ情報から有益なルールを発見するためには, 問題領域固有のドメイン知識を含めた解析が必要であり, それら全ての情報を入力することは困難である.

しかし, 従来のデータ解析では, 背景知識を持つエキスパートと解析者が異なっていたため, データ解析の各フェーズにおいてエキスパートは思考を中断させられ, また, 思いついた仮説を直ちにデータ解析に反映させることが困難であった.

3 取り扱うアンケート調査データ

次に, 本研究で取り扱う WebPAC データの特徴とその解析手法の概要について述べる.

WebPAC データは, 約 6000 人のインターネット利用者に対し実施した「アンケート調査」, および, 個々の利用者の一ヶ月間の Web サイトへの「アクセス状況」が記録されている. アンケート調査には年齢, 性別から所有商品情報といった各種のプロフィールを含む利用者の情報が記録され, Web サイトへのアクセス状況には接続した Web サイトとアクセス時間などが記録されている. アンケート調査の属性データは多岐に渡るため, 従来のデータマイニング手法をそのまま適用したのでは膨大な数のルールが生成され, エキスパートの思考を阻害する恐れがある.

4 提案手法の概要

提案するインタラクティブデータマイニング環境の概念図を図 1 に示す. 提案するシステムは,

- エキスパートが解析結果を参照しながらシステムと対話的に解析が進められる仮説形成支援システム

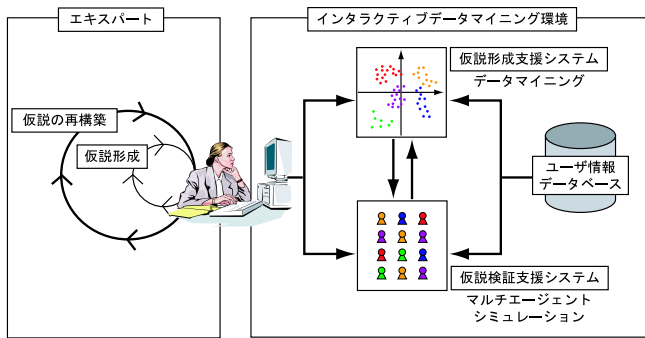


図 1: インタラクティブデータマイニングの概念図

- 対象問題をモデル化したシミュレーションを利用した仮説検証支援システム

の 2 つのシステムから構成され、次のように仮説の形成と検証が行われる。

(1) まず、仮説形成支援システムにより「アンケート調査データ」を解析し、インターネット利用者の特徴を抽出する。(2) 次に、「Web サイトへのアクセス状況」から個々のインターネット利用者の行動パターンを抽出する。(3) そして、これらの解析結果から個々のインターネット利用者の意思決定モデルを構築し、マルチエージェントシミュレーションを作成する。(4) シミュレーション結果から仮説の不具合を発見し、不具合の原因となった解析ステップに戻り、仮説の再構築を行う。

このようにエキスパートが直接システムと対話的に解析を進めることで、思考の中断を抑え、シミュレーション環境を利用することで、何通りもの仮説を短時間で検証することが期待できる。

4.1 仮説形成支援システム

ここでは、仮説形成支援システムについて説明する。

WebPAC データは、年齢、性別といった重要度の高い情報から所有商品や好きなブランドまで、重要度の異なる情報が混在しているため、全ての属性について解析すると計算量が爆発する。

これを回避するには、問題領域の背景知識や解析者の主観を利用することが考えられる。そのため、少なくとも (1) タスクに対して明らかに重要性が低い属性を取捨選択できるユーザインターフェースを備え、(2) インタラクティブにデータ解析することで生じる負担を減らし、インタラクティブに解析することで新しい発見を促すように、リアルタイム処理、バッチ処理すべき計算内容の区分ができていなければならない必要がある。

例えば、データの正規化や欠損データの穴埋めはバツ

チ処理すべきであり、属性の重要度を变化させたときのクラスタの変化過程はリアルタイムに示す方が、解析者の問題意識を活性化することが期待できる。

4.2 仮説検証支援システム

ここでは、仮説検証支援システムについて説明する。インターネット利用者の意思決定は掲示版など様々な情報や他者の行動に左右され、その挙動は複雑である。近年、このような複雑な現象をトップダウン的に解析するのではなく、ボトムアップ的に解析する手法としてマルチエージェントシミュレーション [2] がある。

仮説検証支援システムでは、インターネット利用者の特性調査データの解析を通して、利用者の意思決定モデルを構築し、構築した意思決定モデルを組み込んだマルチエージェントシミュレーションによって仮説の検証、および、広告など、新しい情報を提示したときの利用者の挙動予測に用いる。

インターネット利用者の意思決定モデルはゲーム理論で用いられる利得表などでは表現しきれないため、Web サイトへのアクセス状況のデータをニューラルネットワークに学習させることで構築する。

5 おわりに

本稿では、マーケティングにおける意思決定支援として、インタラクティブデータマイニング環境を提案し、支援システム環境を構成する「仮説形成システム」と「仮説検証システム」に必要な要件について述べた。本稿で提案した手法は、オンライン教育サービスで記録される「学習履歴」や「ユーザ情報」に適用することで学習コンテンツの開発支援を行うことができると期待される。

6 謝辞

研究の一部は科学技術振興事業団プロジェクト「計算科学技術活用型特定研究開発推進事業 (ACT-JST)」の研究助成、および、2003-2005 年度、科学研究費補助金 (若手研究 (B) 課題番号 15700504) の助成を受けて実施された。また、数値データを提供して頂いた (株) 電通メディアマーケティング局メディアリサーチ 1 部 松永久氏、および森田喜文氏に感謝の意を表す。

参考文献

- [1] 大澤 幸生: チャンス発見, 東京電気大学出版局, (2003).
- [2] 車谷 浩一: マルチエージェント社会シミュレーション展望, システム / 制御 / 情報, システム制御情報学会誌, Vol.46, No.9, pp.518-523 (2002).