

非選定行為を考慮した商品推奨アルゴリズムの一提案

吉原 大悟[†] 金久保 正明[‡] 菱沼 千明[‡]

[†]東京工科大学 工学部 情報工学科

[‡]東京工科大学 コンピュータサイエンス学部

1. はじめに

近年 POS(point of sales)システムを用いて得た大量のトランザクションデータ（店舗で一人の客が一度に買った商品のデータ）を、データマイニングの手法の一つであるバスケット分析を行い、販売促進に使える有効な情報を取り出しマーケティングに用いている。

本研究では、個々の商品全てに微小アンテナ付 IC チップを付け、商品棚にリーダー・ライターを設置することによって、個々の商品の認識を行うことができることを前提としている。

RFID 技術を用いることによって検出できる“手に取ったが戻した”という顧客の行動を非選定と定める。そして、この非選定行為を考慮した個々の顧客への推奨品提示と、非選定を含む商品の相関関係情報を抽出することを試みた。

2. 本研究の概要

従来には無い非選定という行為を考慮した個々の顧客への推奨品提示と、非選定を含む商品の相関関係情報を小売店とメーカーへ提示することが本研究の課題である。

非選定という行為をトランザクションデータに加え、大量のトランザクションデータでバスケット分析を行い、非選定を含む相関関係を導出することで解決を図る。導出された相関関係を用いて個々の顧客への推奨品の提示を行い、よく売れる商品の組合せというような小売店への情報や、非選定や競合関係といったメーカーへの情報の提示を行う。

2.1 システム構成

非選定を含むトランザクションデータを入力し、非選定を含む過去の購買履歴データを用いて相関関係を導出する。

システムの概要図を図1に示す。

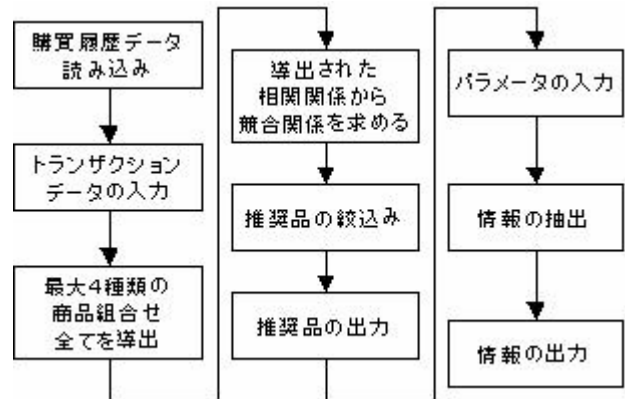


図1. システムの概要図

2.2 購買履歴データ

本研究では商品が大文字のアルファベットで表し、非選定した商品を小文字のアルファベットで表す。ランダムで発生させたアルファベットで表される文字列にある程度偏りを持たせたものを、大量のトランザクションデータに見立てて購買履歴データとした。

2.3 相関関係の導出

集合に含まれる商品の種類が増えると、その組合せの出現する頻度が低くなり、興味のある相関関係と言えなくなってしまうので、最大4種類の商品組合せの相関関係の導出とした。

2.4 競合関係の導出方法

競合関係は 2.3 で導出した非選定を含む相関関係を用いて求める。以下の(1)と(2)の相関関係で、支持度と確信度が共にほぼ等しい時、(1)と(2)は競合関係といえる。

$$f(A; b) \tag{1}$$

$$f(a; B) \tag{2}$$

A と B は共に 1 個以上のアイテム集合である。また、(1)は A を購入する人は B を非選定するという関係であり、(2)は A を非選定する人は B を購入するという関係である。

A modified basket analysis for RFID shop

Daigo Yoshihara [†], Masaaki Kanakubo [‡], Chiaki Hishinuma [‡]

[†]Department of Information Technology, Tokyo University of Technology

[‡]School of Computer Science, Tokyo University of Technology

2.5 推奨品の選定アルゴリズム

導出した相関関係を(3)で表される併買データと(4)で表される否併買データごとに管理する。

$$f(G_1 \dots G_n; K_i) > \alpha_i \text{ かつ } K_i \notin \{G_1 \dots G_n\} \quad (3)$$

$$f(G_1 \dots G_n, g_s \dots g_t; K_i) > \beta_i \text{ かつ } K_i \notin \{G_1 \dots G_n, g_s \dots g_t\} \quad (4)$$

$$k_i \in \{g_s \dots g_t\} \quad (5)$$

(3)は従来のバスケット分析の併買データの条件式であり、Gは入力トランザクションデータを表している。パラメータとして与えられる支持度と確信度の下限値である α_i 以上であれば K_i を推奨品として提示する。また(4)と(5)は非選定を考慮した否併買の条件式であり、下限値である β_i 以上であれば K_i を推奨品として提示するが、 K_i が(5)式のとき推奨されない。

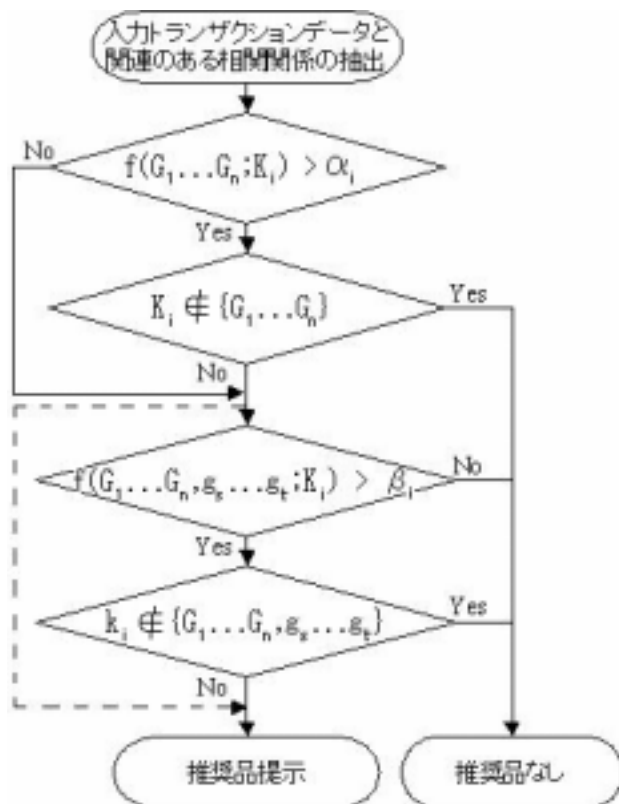


図2. 推奨品選定

導出した相関関係の内、入力したトランザクションデータと関連のある相関関係を抽出し、その相関関係の示す商品を推奨品候補とする。点線で示した流れは従来のバスケット分析である。入力したトランザクションデータから推奨品が定まらなかった場合は、支持度の高い商品

を推奨候補品として提示する。

2.6 情報の抽出

入力したトランザクションデータと関連があり、パラメータを上回った相関関係を情報として抽出する。情報の抽出の直前でパラメータの入力を行えるようにしたことで、ある程度任意でデータの抽出を行えるようにした。

出力結果を以下の図3に示す。

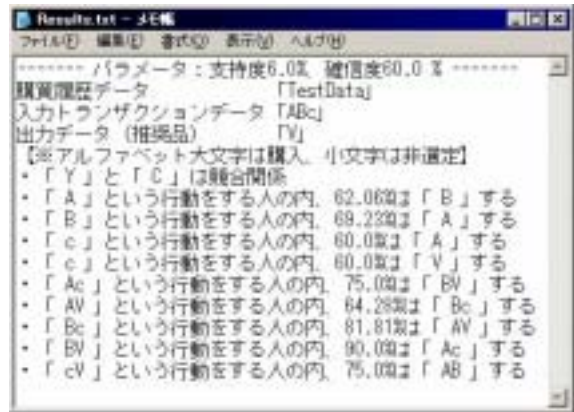


図3. 出力結果

3. 評価

従来のバスケット分析は非選定が無いため、非選定した商品を推奨品として提示してしまう。しかし、本研究のアルゴリズムでは推奨品として提示されない。顧客の非選定行為を含む購買行動から相関関係を用いて推奨品提示を行っているので、課題として挙げた個々の顧客への推奨品提示は実現できたといえる。また、従来は不確かであった競合関係が明らかになり、小売店やメーカーに対しての競合関係や非選定を含む相関関係の情報を抽出することができた。

4. まとめ

RFIDを用いて「手に取ったが戻した」という非選定行為を検出しているが、その顧客動向の意図は図れない。非選定や非選定を含む相関関係の示す意味の解釈が重要であると考えられる。

参考文献

- [1] 鷲尾隆:「バスケット分析」,日本ファジィ学会誌 Vol. 13 Num. 1 pp. 54 (2001. 02)
- [2] 松浦大樹、鷲尾隆、元田浩:「極値集合を用いたバスケット分析の電話回線データへの適用」,情報処理学会研究報告. ICS, 知能と複雑系, Vol. 97, Num. 81, pp. 45-50 (1997. 08)