

関連ルールの視覚化に関する一考察

1 Z-5

吉田万里子 飯塚哲也 磯部成二
NTT情報通信研究所

1. はじめに

大量のデータに隠れている意思決定に有効なルールやパターンを発見するデータマイニングの技術が注目されている。関連ルール（以下、単にルールという）は、その一手法として、POSデータの分析などによく利用されている。しかし、自動的に生成されるルールは大量であり、そこから有効な知識を発見したり、知識に基づく分析をすることは困難である。そこで我々は、直観的に理解しやすい形でルールを視覚化したり、支持度・確信度や、関連情報、ルール同士の関係等を同時に視覚化する方法を検討している [1]。

本稿では、[1]の視覚化形式を発展させた視覚化例を示し、そこから発見できる知識について考察し、視覚化形式の比較・評価を行う。

2. 関連ルール

トランザクションの集合Dに対し、X、Yをアイテムの集合とすると、関連ルールは $X \Rightarrow Y$ のように表され、Xをルールの本体(body)、Yをルールの頭部(head)、XとYに含まれるアイテム数の和をパスという。ルールは、Xを含むトランザクションの割合に対して、X、Yをともに含むトランザクションの割合を表す確信度と、Dに対して、X、Yをともに含むトランザクションの割合を表す支持度をパラメータとして抽出される。

3. 関連ルールの視覚化

関連ルールは一般にテキスト形式でユーザーに提示されるが、有向グラフを用いる手法も検討されている [2]。

しかし、ルールからの有効な知識獲得を支援するためには、関連情報も含め、抽出されたルールを直観的に理解しやすい形で視覚化することが重要である。

我々は、この視覚化に視覚的多次元データ分析ツール INFOVISER [3] を適用している。INFOVISER は、データ属性をノード型やライン型の図形属性にマッピングして視覚化し、ユーザーが直観的にデータのパターンを発見するのを支援するためのツールである。以下に、実データから抽出したルールを INFOVISER によって視覚化した例を示す。

4. 視覚化例

5ヵ月間、1611件のECデータをもとに、最小支持度 0.004、最小確信度 0.25 でルールを抽出し、視覚化を行った。このデータには、トランザクション ID、年月日、顧客 ID、商品 ID が含まれる。

- 網状図 (図1): 丸図形が head の商品、菱形図形が body の商品、横軸が商品 ID、縦軸が商品の売上個数、ラインの色がルール ID、幅がルールの確信度を表している。
- 包含図 (図2): head が同一商品であるルールについてまとめたものである。大図形が head の商品、中図形がルール、小図形が body の商品を、各図形の色は、それぞれ head の商品の売上、ルールの確信度、body の商品の売上を表している。
- 円状図 (図3): ある商品を head とするルールについてまとめたものである。中心図形が head の商品、周辺図形が body の商品、各図形の大きさが商品の売上、ラインの太さがルールの確信度を表している。
- 時系列図 (図4): head が同一商品であるルールについて、商品の月々の売上と併売関係を表現したものである。丸図形系列が head の商品の売上、四角図形系列が body の商品の売上、四角図形の大きさが月毎のルールの確信度を表している。

5. 考察

ルールの視覚化形式の比較を表1にまとめる。

網状図は、「商品 A と商品 B は双方向に高い併売関係がある」のように、2つの商品間に相互に高い併売関係がある時、この関係を表現できる (評価例1)。一方、ルールが多くなると各ルールを識別しにくい、パスが3以上では表示が複雑になるなどの問題点がある。

包含図は、長いパスのルールが表現可能で、head、body の商品に注目して、ルールの確信度、併売商品の売上を比較できるが、異なるパスのルールに同一商品が出現する場合、それを表現できない (評価例2)。

円状図は、包含図同様、注目商品について複数ルールを表現できる。さらに、「商品 A を買った人が商品 B を買った場合、商品 C を買う確率が大幅に高くなる」のように、パスが長くなることで、確信度が上昇する場合、この関係を表現することができる (評価例2)。この関係は、包含図では表現できなかった。

時系列図では、「商品 A の売上は月によってばらつきはあるが、商品 A を買った人は、一定の確率で商品 B を買っている」のように、ルールの時系列変化を表現できる（評価例 3）。

このように、視覚化によって、注目商品について、同一バス、異なるバス、時系列等でルールの支持度や確信度の比較ができ、さらに有益な知識を得ることが可能となる。

2つの商品で、相互に高い併売関係がある場合は、網状図で表現することが出来るが、3つ以上の商品集合グループで高い併売関係がある場合（評価例 4）、今回示した視覚化例ではこのような関係を表現するのは難しく、今後、さらに視覚化形式を検討する必要がある。

表 1：ルールの視覚化形式の比較

評価項目(評価例番号)	網状図	包含図	円状図	時系列
大量ルールの表示	△	○	△	△
バスが長いルールの表示	△	○	○	△
併売商品ペアの表示 (1)	○	×	○	×
異なるバスのルール間の関係 (2)	△	×	○	×
時系列表示 (3)	×	×	×	○
複数併売商品グループの表示 (4)	△	×	△	×

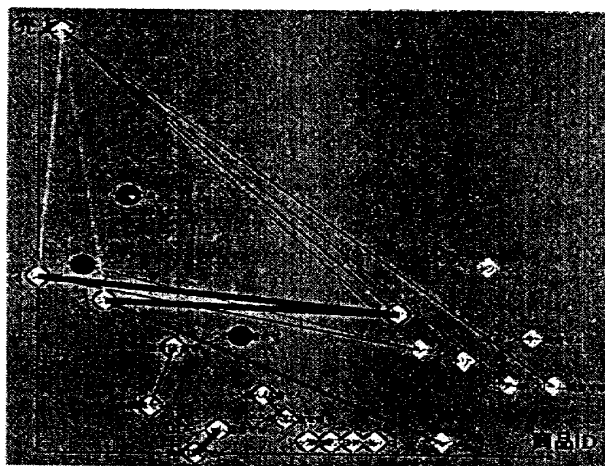


図 1：網状図

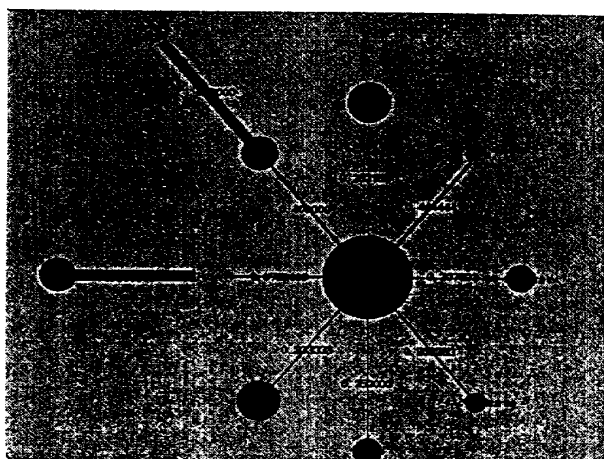


図 3：円状図

6. まとめ

実データから抽出した複数のルールを INFOVISER によって視覚化し、それぞれの視覚化形式について比較・評価した。これらの視覚化パターンをいくつか同時に提示することで、有効な知識獲得を支援することができると考えられる。

今後は、他の有効な視覚化形式、視覚化形式の組み合わせパターンについても検討していきたい。

さらに、支持度や確信度、注目商品をインタラクティブに切り替えるためのインターフェースや、ルールの記述形式等の検討も必要である。

参考文献

- [1] 飯塚他: “相関ルールの視覚化によるデータマイニング支援方式”, 信学技法, AI97-35, Vol.97, No.415, pp.7-12, 1997.
- [2] 福田他: “相関ルールの視覚化について”, 信学技法, DE95-6, Vol.95, No.02, pp.41-48, 1995.
- [3] 黒川他: “情報可視化のためのデータビジュアル化モデル”, 情報処理学会研究報告 96-HI-65, Vol.96, No.21, pp.51-56, 1996.

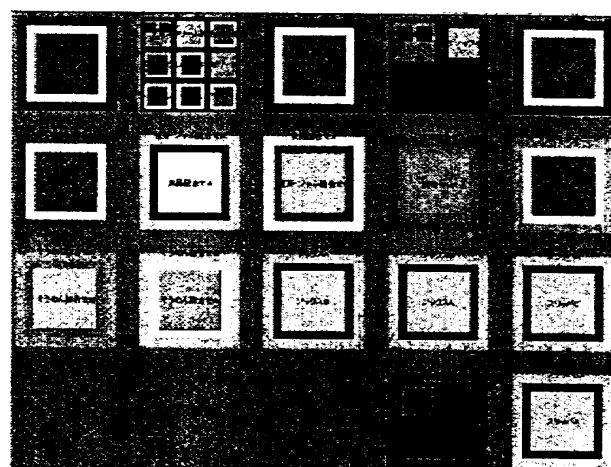


図 2：包含図

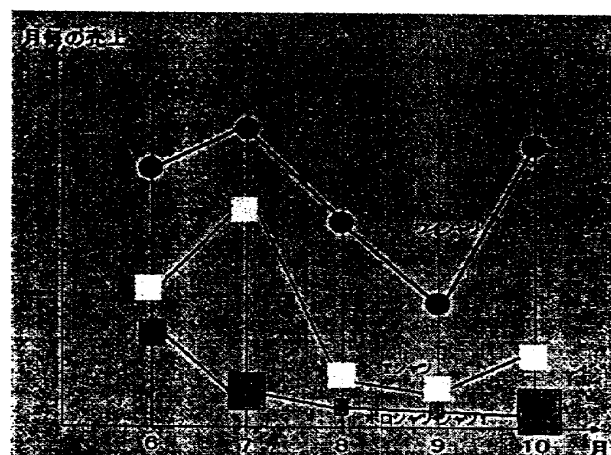


図 4：時系列図