

マーケットバスケット分析

はじめに

みなさん、こんにちは。今回は「マーケットバスケット分析」についてご紹介します。マーケットバスケット分析 とは「アソシエーション分析」の一種で、購買データを分析することで、同時に購入されやすい商品の組み 合わせを発見したい場合などに使われます。

有名な例として、「おむつとビール」があります。販売データを分析した結果、おむつとビールを同時購入する頻度が高いことが判明し、2 つを並べて陳列したところ売り上げが上昇したという事例です。

マーケットバスケット分析では、販促につながる有効なルールを見つけることによって、店舗レイアウト設計の示唆や購入者へのリコメンドなどを効果的に行うことができます。本記事では、マーケットバスケット分析 について、取引データサンプルを用いて、ハンズオン形式でご説明します。ぜひお手元の RapidMiner で 再現してみてください。※本資料は、macOS、RapidMiner Studio Version9.10を使用しています。



プロセスの概要



1. 使用データ

今回使用するデータセットは、RapidMiner のリポジトリの中に備え付けられている「Transactions」です。 取引 ID、商品 ID、注文回数、販売価格の属性を持つ購買に関わるデータセットです。



- 2. データは「Data Prep」サブプロセスに移されます。データ準備に、サブプロセス内で次のことを行います。
 - a. 各取引が1行で表されるようにピポット処理 一つの取引で同じ商品が複数存在することを考慮し、合計値で集計
 - b. 取引 ID に、ID 属性の役割を設定
 - c. 属性名(列名)を変換
 - d. 欠損値の処理
 - e. 注文回数(数値)を商品購入の真偽(二値カテゴリ true/false)に変換
- 3. データの 75%が「FP-Growth」オペレータと「Create Association Rules」オペレータに、 25%が「Apply Association Rules」オペレータに送られるように、データを分割します。
- 4. 「FP-Growth」オペレータで、どの商品が最も頻繁に同時購入されているかを計算します。
- 5. 「Create Association Rules」オペレータで、頻繁に購入されている商品の組み合わせに基づき、 ルールを作成します(ルールの確信度に応じて商品の推奨に使用可能)
- 6. 「Multiply」オペレータでデータを複製し、5で作成したルールを出力します。
- 「Apply Association Rules」オペレータで、作成したルールをデータに適用します。
 (現在顧客のバスケットに入っているアイテムを考慮して、推奨されるアイテムを返します)

では、プロセスの概要 2~7を順番に見ていきましょう。



2. データ準備

以下の画像は、処理前と処理後の取引データです。

元のデータには、取引 ID(Invoice)、商品 ID(product 1)、注文回数(Orders)、販売価格(Sales value)が含まれています。 処理後のデータでは、「Invoice」属性に「id」という役割が与えられています。 その他全ての列は個々の商品を表し、商品が各取引で購入されたかどうかは、"true "または"false "で 表示されています。

処理前

| Row No. | Invoice | product 1 | Orders | Sales value |
|---------|---------|------------|--------|-------------|
| 1 | 131506 | Product 20 | 1 | 40 |
| 2 | 131506 | Product 21 | 1 | 80 |
| 3 | 131507 | Product 11 | 1 | 80 |
| 4 | 131508 | Product 19 | 1 | 32 |
| 5 | 131509 | Product 31 | 1 | 9 |
| 6 | 131510 | Product 11 | 1 | 80 |
| 7 | 131510 | Product 20 | 2 | 40 |
| 8 | 131510 | Product 20 | 1 | 40 |
| 9 | 131519 | Product 11 | 1 | 80 |
| 10 | 131541 | Product 11 | 1 | 80 |

処理後

| Row No. | Invoice | Product 10 | Product 11 | Product 12 | Product 13 | Product 14 | Product 15 | Product 16 |
|---------|---------|------------|------------|------------|------------|------------|------------|------------|
| 1 | 131506 | false |
| 2 | 131507 | false | true | false | false | false | false | false |
| 3 | 131508 | false |
| 4 | 131509 | false |
| 5 | 131510 | false | true | false | false | false | false | false |
| 6 | 131519 | false | true | false | false | false | false | false |
| 7 | 131541 | false | true | false | false | false | false | false |
| 8 | 131542 | false | true | false | false | false | false | false |
| 9 | 131543 | false |
| 10 | 131559 | false |

さて、この変換がどのように行われたかを見てみましょう。(プロセスの概要 2. a~e)



a. 各取引が1行で表されるようにピポット処理、合計値で集計

| パラメータ × | 環境 | | | |
|-------------------------|-------|------------|--------------|-----|
| Pivot | | | | |
| group by attributes | | | 入属性選択 | 1 |
| column grouping attribu | produ | ict 1 | • | |
| aggregation attributes | | i i | Edit List (1 |) 🗊 |
| use default aggrega | ation | | | 1 |
| | | | | |

「Pivot」オペレータを使用して、各取引が 1 行で表示されるように列と行のグループ化を行います。また、 一つの取引で同じ商品が複数存在することを考慮し、合計値での集計を行います。どの属性をグループ 化するか、どの属性で集計するかを選択する必要があります。次のボタンをクリックして設定します。 ・行のグループ化:「group by attributes」パラメータの「属性選択」(赤枠) ・列のグループ化:「column grouping attribute」パラメータの「 ▼」(赤枠) ・指定属性の集計:「aggregation attributes」パラメータの「Edit List」(青枠)

▼Invoice(取引 ID)をグループ化し、取引 ID につき 1 行となるように設定

| 属性を選択: group by attributes Attributes that groups the examples which form | one row after pivoting. |
|--|---------------------------|
| 属性 | 選択した属性 |
| 校索 🗙 | 使索 🕒 🗙 |
| 🚴 Invoice | |
| # Orders | |
| ₿ product 1 | |
| # Sales value | |
| | (適用 🗶 キャンセル |
| | |
| 属性を選択: group by attributes Attributes that groups the examples which form | n one row after pivoting. |
| 属性 | 選択した属性 |
| 検索 | 秋 奈 〇 × |
| # Orders | 🚴 Invoice |
| 🚓 product 1 🕒 🕞 | |
| # Sales value | |
| | 🥥 適用 🗙 キャンセル |

グループ化したい Invoice (取引 ID)を指定します。 左側の 属性 列から Invoice 属性をクリック し、右矢印ボタンをクリックすると、右側の 選択した属性 の方に移動します。



▼product 1(商品 ID)をグループ化し、商品 ID 1 つにつき 1 列となるように設定

| × | |
|--|---|
| | |
| 🖹 属性選択 | Í |
| product 1 | |
| InvoiceOrders | (1) |
| B product 1 # Sales value | (i) |
| | × 属性選択 product 1 象 Invoice # Orders 象 product 1 # Sales value |

「column grouping attribute」パラメータは、「product 1」を指定します。

▼Orders(注文回数)を集計し、合計値が表示されるように設定

| | パラメータリストを編集: aggregation attributes The attributes which should be aggregated. | | | | | |
|------------|--|-------------------------|-----------------------------|--|--|--|
| aggregatio | n attribute | aggregation function | | | | |
| Orders | | Sum | $\overline{\mathbf{\cdot}}$ | | | |
| | דארד | リを追加 🛛 🖳 エントリを削除 🛛 🥑 適用 | ★ キャンセル | | | |

一つの取引で同じ商品が複数存在することがわかるように、合計値での集計を行います。
 「aggregation attribute」=合計値で集約する必要のある属性 は、「Orders」
 「aggregation functions」=集計方法 は、「Sum」を赤枠の▼から選択します。

適用後、データは次のようになります。

| Row No. | Invoice | sum(Orders)_Product 10 | sum(Orders)_Product 11 | sum(Orders)_Product 12 | sum(Orders)_Product 13 | sum(Orders)_Product 14 | sum(0 |
|---------|---------|------------------------|------------------------|------------------------|------------------------|------------------------|-------|
| 1 | 131506 | ? | ? | ? | ? | ? | ? |
| 2 | 131507 | ? | 1 | ? | ? | ? | ? |
| 3 | 131508 | ? | ? | ? | ? | ? | ? |
| 4 | 131509 | ? | ? | ? | ? | ? | ? |
| 5 | 131510 | ? | 1 | ? | ? | ? | ? |
| 6 | 131519 | ? | 1 | ? | ? | ? | ? |
| 7 | 131541 | ? | 1 | ? | ? | ? | ? |
| 8 | 131542 | ? | 1 | ? | ? | ? | ? |
| 9 | 131543 | ? | ? | ? | ? | ? | ? |
| 10 | 131559 | ? | ? | ? | ? | ? | ? |



b. データの役割を設定

| パラメータ × | 環境 |
|----------------------|---------------|
| 😼 Set Role | |
| attribute name | Invoice 🔻 🕄 |
| target role | id 🔻 🕄 |
| set additional roles | Edit List (0) |

「Set Role」オペレータを使用し、「Invoice」属性(赤枠)に、「id」の役割(青枠)を割り当てます。 この設定により、Invoice 属性はデータの準備とモデリング段階の両方において、計算に影響を与えない 属性になります。

c. 正規表現を使って属性名を商品名のみに変換

| パラメータ × | 環境 |
|-----------------------|----------------|
| 😼 Rename by Replaci | ng |
| attribute filter type | all 🔻 |
| invert selection | Œ |
| include special attri | butes (|
| replace what | sum\(Orders\)_ |
| replace by | Ĩ |

「Rename by Replacing」オペレータを使用し、属性名を商品名のみに変換します。

各属性は「sum(Orders)_Product 10」のような名前になっているので、「sum(Orders)_」の部分を 削除する必要があります。「replace what」パラメータには置き換えたい部分(正規表現)を設定し、 「replace by」パラメータを空白にすることで、正規表現にマッチするテキストを削除することができます。

適切な正規表現を作成するには、「replace what」パラメータ(青枠)ボタンをクリックし設定します。 クリックすると、以下のウィンドウが表示されます。



| 正規表現を編集: A regular expression defining what should be replaced in the attribute names. | | | | | | |
|--|----------------------|-------|---------|--|--|--|
| 正規表現 | | | | | | |
| sum\(Orders\)_ | - | | - × | | | |
| 장 正常な正規表 | 現 | | | | | |
| Replacement (va | alue for 'replace by | ") | | | | |
| | | | | | | |
| | | | | | | |
| インライン検索 | 結果リスト (1) | オプション | | | | |
| テキスト | | | | | | |
| sum(Orders)_product10 | | | | | | |
| プレビュー | | | | | | |
| product10 | | | | | | |
| | | | | | | |
| | | 🕑 適用 | 🗙 キャンセル | | | |

正規表現が各属性の対象テキストとマッチしているかどうかをテストするために、「テキスト」(赤枠)に属 性名を入力します。マッチした部分は黄色に表示され、結果のテキストが「プレビュー」(青枠)に表示 されます。正規表現の詳しい情報は、<u>こちらの記事</u>(英語)をご覧ください。

d~e. 欠損値を 0(数値)に置き換え、全ての数値を「True/False」値に変換

| パラメータ × | 環境 | |
|-----------------------|-----------------|--------------|
| 👫 Replace Missing Va | lues | |
| create view | | 1 |
| attribute filter type | all | • |
| invert selection | | (1) |
| include special attri | butes | (1) |
| default | zero | • (1) |
| columns | 🛒 Edit List (0) | 1 |



| パラメータ × | 環境 × | |
|------------------------|-------|---|
| 🍄 Numerical to Binom | inal | |
| attribute filter type | all | • |
| invert selection | | ٦ |
| include special attrib | outes | ١ |
| min | 0.0 | ٦ |
| max | 0.0 | ٦ |

- d. まず、「Replace Missing Values」オペレータを使用して、欠損値を 0(zero)に変換し全てのデー タを数値にします。「default」パラメータ(赤枠)は、デフォルトでは「average」に設定されています が、「zero」に変更することで、次のステップが非常に簡単になります。
- e. 次に、「Numerical to Binominal」オペレータを使用し、全ての数値を二値カテゴリ「True/False」 値に変換します。パラメータは全てデフォルト値のままにします。「min」と「max」の値が 0 に設定さ れているため(赤枠)、データ内の 0 の値は"false "に、それ以外の値は全て"true "になります。
- データは以下のようになり、この後に接続する アソシエーション オペレータで使用できます。

| Row No. | Invoice | Product 10 | Product 11 | Product 12 | Product 13 | Product 14 | Product 15 | Product 16 | Product 17 | Product |
|---------|---------|------------|------------|------------|------------|------------|------------|------------|------------|---------|
| 1 | 131506 | false | false |
| 2 | 131507 | false | true | false | false | false | false | false | false | false |
| 3 | 131508 | false | false |
| 4 | 131509 | false | false |
| 5 | 131510 | false | true | false | false | false | false | false | false | false |
| 6 | 131519 | false | true | false | false | false | false | false | false | false |
| 7 | 131541 | false | true | false | false | false | false | false | false | false |
| 8 | 131542 | false | true | false | false | false | false | false | false | false |
| 9 | 131543 | false | false |
| 10 | 131559 | false | false |

3. データの分割

| パラメータ | × | 環境 | | | | | | | |
|---------------|--------|-----------|--------|------------|---|-------------|-------|--|--------|
| 🍸 Split Data | | | | | | | | パラメータリストを編集: partitions | |
| partitions | Г | 쀍 Edit I | Enumer | ation (2). | 🗊 | | | The partitions that should be created. | |
| | | | | | | | ratio | | |
| sampling type | lir | near samp | ling | | • | | 0.75 | | ^ |
| | | | | | | | 0.25 | | \sim |
| use local ran | aom se | eea | | | Í | $ \rangle$ | עד 📕 | トリを追加 | < |



「FP-Growth」オペレータと「Apply Association Rules」オペレータに入力するデータが準備できました。 「Apply Association Rules」オペレータで使用するデータが、ルール作成に使用するデータと異なるよう に、データを分割する必要があります。データの分割には、「Split Data」オペレータを使用します。

今回は、上の画像のように、sampling type」パラメータは「linear sampling」を選択し、線形サンプ リングを行いました(青枠)。分割の比率を設定するには、「partitions」パラメータの「Edit Enumeration」ボタンをクリックします(赤枠)。ここでは、データの75%を「FP-Growth」オペレータ に、25%を「Apply Association Rules」オペレータに送るように設定しましたが、この比率は必要に応 じて調整ができます。

4~7. アソシエーション オペレータ

| パラメータ × 環境 | \times | |
|-------------------------------|------------------------------|---|
| FP-Growth | | |
| input format | items in dummy coded columns | • |
| positive value | | i |
| min requirement | support | • |
| min support | 0.95 | Œ |
| min items per itemset | 1 | C |
| max items per itemset | 0 | C |
| max number of itemsets | 1000000 | Œ |
| ✓ find min number of itemsets | | Œ |
| min number of itemsets | 100 | C |
| max number of retries | 15 | C |
| requirement decrease factor | 0.45 | ſ |
| must contain list | Edit Enumeration (0) | 0 |
| must contain regexp | | Œ |

断

るためには、まず、「FP-Growth」オペレータを使 用し、データセットで頻繁に発生するアイテムセットを効率的に計算します。 「FP-Growth」オペレータで最も重要なパラメータは、「input format」パラメータ(赤枠)と、 「min requirement」パラメータ および「min support」パラメータ(青枠)です。

「input format」パラメータは、3 種類の入力形式をサポートしています。



【「input format」3種類の入力形式とは】

\Box item list in a column:

一つの取引に属する全ての商品は1列に表示され、CSV形式のように商品が区切られています。

| Row No. | Invoice | Products |
|---------|---------|--|
| 1 | 1306797 | Product 10 |
| 2 | 1306799 | Product 10 Product 10 |
| 3 | 1306800 | Product 11 Product 12 Product 12 |
| 4 | 1306824 | Product 13 Product 14 |
| 5 | 1306825 | Product 15 |
| 6 | 1306835 | Product 16 |
| 7 | 1306845 | Product 17 Product 17 |
| 8 | 1306872 | Product 18 |
| 9 | 1306873 | Product 19 Product 19 |
| 10 | 1306874 | Product 11 Product 15 Product 15 Product 15 Product 20 |

□ items in separate columns :

一つの取引に属する全ての商品は別々の列に表示され、最初の商品名が1列目、2番目の商品

名が2列目となるように表示されます。

| Row No. | Invoice | ltem_1 | ltem_2 | ltem_3 | ltem_4 | ltem_5 | ltem_6 | ltem_7 |
|---------|---------|------------|------------|------------|------------|------------|--------|--------|
| 1 | 1306797 | Product 10 | ? | ? | ? | ? | ? | ? |
| 2 | 1306799 | Product 10 | Product 10 | ? | ? | ? | ? | ? |
| 3 | 1306800 | Product 11 | Product 12 | Product 12 | ? | ? | ? | ? |
| 4 | 1306824 | Product 13 | Product 14 | ? | ? | ? | ? | ? |
| 5 | 1306825 | Product 15 | ? | ? | ? | ? | ? | ? |
| 6 | 1306835 | Product 16 | ? | ? | ? | ? | ? | ? |
| 7 | 1306845 | Product 17 | Product 17 | ? | ? | ? | ? | ? |
| 8 | 1306872 | Product 18 | ? | ? | ? | ? | ? | ? |
| 9 | 1306873 | Product 19 | Product 19 | ? | ? | ? | ? | ? |
| 10 | 1306874 | Product 11 | Product 15 | Product 15 | Product 15 | Product 20 | ? | ? |

□ items in dummy coded columns: ※今回選択する形式

全ての商品は各々列があり、商品名が列名になります。各取引に対して、

二値カテゴリ(true/false)は、列名の商品が取引されているかどうかを示します。

| Row No. | Invoice | Product 10 | Product 11 | Product 12 | Product 13 | Product 14 | Product 15 | Product 16 |
|---------|---------|------------|------------|------------|------------|------------|------------|------------|
| 1 | 131506 | false |
| 2 | 131507 | false | true | false | false | false | false | false |
| 3 | 131508 | false |
| 4 | 131509 | false |
| 5 | 131510 | false | true | false | false | false | false | false |
| 6 | 131519 | false | true | false | false | false | false | false |
| 7 | 131541 | false | true | false | false | false | false | false |
| 8 | 131542 | false | true | false | false | false | false | false |
| 9 | 131543 | false |
| 10 | 131559 | false |

各形式の入力データの準備は「FP-Growth」オペレータ ヘルプ>jump to Tutorial Process からご確認いただけます。



「input format」パラメータは、「items in dummy coded columns」の形式を選択しました。 選択理由は、後に接続する「Apply Association Rules」オペレータが必要とする形式だからです。 「min requirement」パラメータでは、「frequency」(データセットに商品が出現する回数)またはデフォ ルト値の「support」が選択できます。「support」(支持度)は、 frequency と同様にデータセットに 出現する頻度を示し、商品の出現頻度をデータセットの全取引で割って算出されます。

$support(itemsetX) = \frac{frequency(itemsetX)}{total transactions}$

support の値を変更するには、手動で変更する方法と、「find min number of item sets」にチェッ クし(デフォルト)「requirement decrease factor」(p.9 緑枠)で変更する方法があります。デ フォルトの設定では、0.95(min support)に0.9(requirement decrease factor)を最大 15 回(max number of retries)かけて、100 個以上のアイテムセット(min number of itemsets) が返されるようになっています。しかし、デフォルト値のままでは、0 個のアイテムセットしか返されませんでし た。そこで、恣意的に「requirement decrease factor」の値を半分の 0.45 にしてみたところ、 138 個のアイテムセットが返ってきました。

| 🛒 Frequentitem | Sets (FP–Growth) | × | | |
|--------------------|------------------|---------|------------|------------|
| No. of Sets: 138 | Size | Support | Item 1 | Item 2 |
| Total Max. Size: 5 | 1 | 0.149 | Product 12 | |
| Min. Size: 1 | 1 | 0.135 | Product 11 | |
| Max Size: 5 | 1 | 0.100 | Product 20 | |
| Contains Item: | 1 | 0.081 | Product 10 | |
| | 1 | 0.081 | Product 18 | |
| Lindata Viaw | 1 | 0.070 | Product 23 | |
| Opulate view | 1 | 0.070 | Product 26 | |
| | 1 | 0.068 | Product 15 | |
| | 1 | 0.059 | Product 13 | |
| | 1 | 0.057 | Product 21 | |
| | 1 | 0.057 | product 1 | |
| | 1 | 0.054 | Product 24 | |
| | 1 | 0.049 | Product 19 | |
| | 1 | 0.046 | Product 14 | |
| | 1 | 0.043 | Product 16 | |
| | 1 | 0.041 | Product 29 | |
| | 1 | 0.030 | Product 25 | |
| | 1 | 0.030 | Product 27 | |
| | 1 | 0.027 | Product 22 | |
| | 1 | 0.027 | Product 31 | |
| | 1 | 0.019 | Product 28 | |
| | 1 | 0.011 | Product 17 | |
| | 1 | 0.011 | Product 30 | |
| | 2 | 0.008 | Product 12 | Product 11 |
| | | | | |



5. 頻繁に購入されている商品の組み合わせに基づいてルールを作成

| パラメータ × | 環境 |
|---------------------|----------------|
| 🛒 Create Associatio | on Rules |
| criterion | confidence 🔻 🗊 |
| min confidence | 0.8 |
| gain theta | 2.0 |
| laplace k | 1.0 |

「Create Association Rules」オペレータを使用し、頻出アイテムセットから関連ルールのセットを作成します。最も重要なパラメータは、「criterion」(基準)と「min confidence」です。

「criterion」には、confidence、lift、conviction、ps、gain、laplace などがありますが、最も一般的なのはデフォルト値の「confidence」です。「confidence」(確信度)は、作成した関連ルールがどのくらい真であると判断されたか(範囲は 0~1)を示し、以下のように計算されます。

 $confidence(itemsetX implies itemsetY) = \frac{support(itemsetX \cup itemsetY)}{support(itemsetX)}$

簡単に言うと、「商品 X が購入されると、商品 Y が購入される」という確信度は、X と Y が一緒に購入される取引の出現率を、商品 X の出現率で割ったものになります。確信度が高ければ高いほど、その関連ルールの確実性は高くなります。

注:今回のデータセットは、ルールの確信度が高いほど、データセット内でのルールに対する支持度が低くなっています。 ご自身の分析を行う際は、ルールの確信度と支持度の二つのバランスがどの程度必要なのかを考えてみてください。

6. 5で作成したルールを複製し、出力

「Multiply」オペレータを使用して、5 で作成した関連ルールを複製します。複製することによって、ルールの出力とデータへの適用を行うことができるようになります。出力すると以下のルールセットが返されました。

| Show rules matching | No. | Premises | Conclusion | Support | Confidence | LaPlace | Gain | p-s | Lift | Conviction |
|-------------------------------------|-----|------------------------|------------|---------|------------|---------|--------|-------|-------|------------|
| all of these conclusions: \bullet | 1 | Product 11, Product 27 | Product 12 | 0.003 | 1 | 1 | -0.003 | 0.002 | 6.727 | 00 |
| Product 12 | 2 | Product 12, Product 26 | Product 20 | 0.003 | 1 | 1 | -0.003 | 0.002 | 10 | 00 |
| Product 11 Product 20 | 3 | Product 20, Product 16 | Product 12 | 0.003 | 1 | 1 | -0.003 | 0.002 | 6.727 | 00 |
| Product 10 | 4 | Product 12, Product 27 | Product 20 | 0.008 | 1 | 1 | -0.008 | 0.007 | 10 | 00 |
| Product 15 | 5 | Product 20, Product 22 | Product 12 | 0.003 | 1 | 1 | -0.003 | 0.002 | 6.727 | 00 |
| Product 13 | 6 | Product 20, Product 31 | Product 12 | 0.003 | 1 | 1 | -0.003 | 0.002 | 6.727 | 00 |
| Product 19 | 7 | Product 18, Product 15 | Product 12 | 0.005 | 1 | 1 | -0.005 | 0.005 | 6.727 | 00 |
| Product 16 Product 29 | 8 | Product 18, Product 22 | Product 12 | 0.003 | 1 | 1 | -0.003 | 0.002 | 6.727 | 00 |
| Product 27 | 9 | Product 15, Product 21 | Product 12 | 0.003 | 1 | 1 | -0.003 | 0.002 | 6.727 | 00 |
| Product 22 | 10 | Product 15, Product 22 | Product 12 | 0.003 | 1 | 1 | -0.003 | 0.002 | 6.727 | 00 |

Section Rules (Create Association Rules)



7. 作成したルールをデータに適用



「Apply Association Rules」オペレータを使用して、作成したルールに、分割してある 25%のデータを 適用させます。「confidence aggregation method」パラメータでは、確信度の集約方法を選択しま す。「binary」を選択すると、ルールに従って推奨されたアイテムには"1 "、推奨されていないアイテムには "0 "と表示されます。パラメータの他の選択肢には、aggregated confidence, aggregated conviction, aggregated LaPlace, aggregated gain, aggregated lift があります。これらのい ずれかが選択された場合は、正の値が選択した集約結果の値になり、負の値には "0 "が表示されます。

| Examp | oleSet (Apply A | ssociation Rules) | × | FrequentItemSets (FP-Grov | wth) $	imes$ | 🛒 Associatio | nRules (Create A | ssociation Rule | s) × |
|---------|-----------------|-------------------|-------------|---------------------------|--------------|--------------|------------------|-----------------|------------------------|
| 開く TI | urbo Prep | Auto Model | | | | | | フィルタ (123 / 1 | 23 行): all |
| Row No. | Invoice | confidence(| confidence(| confidence(Product 12) | confidence(| confidence(| confidence(| confidence(| confidence(Product 19) |
| 20 | 647868 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | 647869 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 647870 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 647871 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 647873 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 647874 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 26 | 647875 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27 | 647876 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28 | 647877 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29 | 647878 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30 | 647879 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 31 | 647880 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 32 | 647881 | 0 | 0 | <u>_</u> | 0 | 0 | 0 | 0 | 0 |
| 33 | 647882 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 34 | 647889 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 35 | 647978 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 36 | 647979 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

以下のデータセットが得られました。(正の値は赤枠)

正の値がほとんど返されていませんが、これはルールの元々の確信度が非常に高く設定されていたためで す。ルールの確信度(Confidence)は全て "1 "で、支持度(Support)は最大でも "0.008 " でした。「Create Association Rules」オペレーターの「min confidence」パラメータ値を下げることで、 提案される商品の数が増える可能性があります。



まとめ

RapidMiner プロセスでは、まず読み込んだ取引データを分析するためのデータ準備を行いました。次に データを分割し、頻繁に同時購入されている商品の組み合わせルールを作成した後、作成したルールに データを適用させました。今回の実行結果の考察の通り、業務視点での確信度と支持率のバランスの検 討が、マーケットバスケット分析活用のためのポイントとなります。

マーケットバスケット分析(アソシエーション分析)は、蓄積された顧客毎の取引データを分析することに よって、思いもよらなかった関連性を見つける可能性があります。RapidMiner は、分析するためのデータ 準備から実行までを Excel やプログラミング言語に比べて容易に行うことができるため、分析結果から施 策を導き出し、素早く PPDAC サイクル(問題解決方法)を回すことができます。

RapidMinerのマテリアルページでは、データ分析の基本的な内容から発展的な内容までご紹介しております。RapidMiner・機械学習についてさらに学習したいという方は、実務課題解決のための動画トレーニングコースをご提供しておりますのでぜひご活用ください。

▼Learning サブスクリプションは<u>こちら</u>