

アソシエーション分析の手順

2021年 ver. 1.0
倉谷 隆博

13-1. 用途

アソシエーション分析 (association : 連想、関連) は、大量に蓄積したデータ (ビッグデータ) を使ってデータ間の相互関係などを調べて役立つ情報を見出す技術、言い換えれば有益な情報を採掘する (mine) するデータマイニング技術の一つである。

この分析の実用例として有名なのが小売りの「おむつとビール販売の関連性」の発見である。おむつとビールの両方を購入するお客様の数は少ないものの、おむつを購入した人のうち多くの人がビールを購入していたことが分かった。データを分析した結果、子供のいる家庭では母親はかさばる紙おむつを父親に買ってくるように頼み、買い物に来た父親は、なんとついでにビールを購入していたという訳だ。そしてこの分析結果を活かし、おむつとビールを並べて陳列するようにしたら売上げが増えたという事例である (この事例の真偽には諸説ある)。

お客様の買い物籠 (バスケット) の中身を分析 (バスケット分析) するにあたり、レシートの内容を分析しどの商品とどの商品と一緒に売れたかをアソシエーション分析で調べる。この分析結果を使って、キャンペーンなどの販促活動や商品陳列方法の改善などに活かすことができる。クロスセル (併売)、アップセル (より高額な商品の販売) などによる売り上げの増加や客単価の改善などが期待できる。

13-2. 分析の評価指標

アソシエーション分析では、(表1) の4つの指標 (通常は、人気度を除いた3つの指標) を使う。リフトの値が1より大きな値をとるときは併売効果があるとみて、商品 X を購入されるお客様に商品 Y もお薦めするという販促施策に活かされる。

(表1) 評価指標

指標	計算式	説明
人気度	商品 Y を購入した人数 / 全体人数 (商品 Y があるレシートの数 / 全体のレシートの数)	全体の中で、商品 Y を購入した人の割合
支持度	商品 X と商品 Y を購入した人数 / 全体人数 (商品 X と商品 Y があるレシートの数 / 全体のレシートの数)	全体の中で、商品 X と商品 Y を購入した人の割合
信頼度	商品 X と商品 Y を購入した人数 / 商品 X を購入した人数 (商品 X と商品 Y があるレシートの数 / 商品 X があるレシートの数)	商品 X を購入した人の中で、商品 Y を購入した人の割合
リフト	リフト = 信頼度 / 人気度	商品 X を購入した上で商品 Y を購入する人の割合 (分子) と商品 Y を単独で購入する人の割合 (分母) の比 リフト > 1 で、商品 Y の単独購入に比べ、商品 X と商品 Y の同時購入の比率が高いことを示す。 つまり、併売効果あるということを示す。

13-3. 分析の手順 (1/2)

レシート 10 枚、商品 10 種類の例でアソシエーション分析を行う。

1) (表2)はレシートのデータを整理した表である。レシートごとに購入された商品のカラムには数字の 1 を記入している。商品ごとの購入数は総計で確認できる。

レシート枚数 = 10 枚であり、商品ごとの総計をレシート枚数で割って支持度とする。支持度を見れば商品の購入度合いが分かる。

商品購入の相互関係を調べるアソシエーション分析では、商品の組合せの数が多いと計算負荷が膨大になる。そこで、それを避けるため、よく購入されている商品、つまり支持度の高い商品に絞って分析を行う方法が採用されている。この分析手順をアプリアリ・アルゴリズムという。

(注) a priori : 事前の、演繹的な

2) (表3)は、(表2)で支持度の高い商品に絞り込んだレシートデータである。(表2)で支持度 50% 以上の商品である。

3) (表4)は、(表2)の商品のなかから2種類の商品を同時に購入しているレシートのカラムに数字の 1 を記入した表である。あとの計算に使用する。

(表2)レシートデータ(架空データ)

レシートNo	商品名									
	A	B	C	D	E	F	G	H	I	J
1	1	0	1	0	0	1	0	1	0	1
2	1	0	0	1	0	0	0	1	0	0
3	0	0	0	1	1	1	1	0	1	0
4	1	1	1	0	0	1	0	0	0	0
5	1	1	0	0	1	0	1	0	0	0
6	0	0	1	0	0	0	0	1	0	0
7	1	0	1	0	0	1	1	1	0	0
8	0	1	0	0	0	0	1	0	1	1
9	1	0	1	1	0	1	0	0	0	1
10	1	1	1	0	0	1	1	1	0	0
総計	7	4	6	3	2	6	5	5	2	3
支持度(%)	70	40	60	30	20	60	50	50	20	30

レシートNo	商品名				
	A	C	F	G	H
1	1	1	1	0	1
2	1	0	0	0	1
3	0	0	1	1	0
4	1	1	1	0	0
5	1	0	0	1	0
6	0	1	0	0	1
7	1	1	1	1	1
8	0	0	0	1	0
9	1	1	1	0	0
10	1 ^③	1 ^①	1 ^⑥	1	1
総計	7	6	6	5	5
支持度(%)	70	60	60	50	50

(表3)支持度で絞り込んだレシートデータ

(表4)2種類の商品を同時に購入したレシートの確認

レシートNo	商品の組合せ(2種類)									
	A・C	A・F	A・G	A・H	C・F	C・G	C・H	F・G	F・H	G・H
1	1	1	0	1	1	0	1	0	1	0
2	0	0	0	1	0	0	0	0	1	0
3	0	0	0	0	0	0	0	1	1	0
4	1	0	0	0	1	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	1	0	1	0	1	0
7	1	1	1	1	0	1	1	1	1	1
8	0	0	0	0	0	0	0	0	0	0
9	1	1	0	0	1	0	0	0	0	0
10	1 ^②	0	1	1	1	1	1	1	1	1
総計	5	3	3	4	5	2	4	3	6	2

13-3. 分析の手順 (2/2)

4) (表5)は、2種類の商品を購入したときのアソシエーション分析の指標の計算結果である。A → C は商品 A を購入した上で商品 C を購入したことを表す。以下、計算式を A → C の場合を代表して記す。

商品 C 単独購入の人気度 = (表3)商品 C のレシート枚数 / レシート総数 = (表3)の ① / 10 = 6 / 10 = 0.6 (表2)の支持度と同じ

商品 A と C 同時購入の支持度 = (表4)同時購入レシート枚数 / レシート総数 = (表4)の ② / 10 = 5 / 10 = 0.5

商品 A と C 同時購入の信頼度 = (表4)同時購入レシート枚数 / (表3)商品 A のレシート枚数 = (表4)の ② / (表3)の ③ = 5 / 7 = 0.714

商品 A の購入による商品 C のリフトの比率 = (表5)信頼度 / (表5)人気度 = (表5)の ⑤ / (表5)の ④ = 0.714 / 0.6 = 1.19

(表5)2種類の商品と同時に購入ときの指標

商品組合せ	A→C	A→F	A→G	A→H	C→A	C→F	C→G	C→H	F→A	F→C	F→G	F→H	G→A	G→C	G→F	G→H	H→A	H→C	H→F	H→G
人気度	0.600	0.700	0.500	0.500	0.700	0.700	0.500	0.500	0.700	0.600	0.500	0.500	0.700	0.600	0.700	0.500	0.700	0.600	0.700	0.500
支持度	0.500	0.300	0.300	0.400	0.500	0.500	0.200	0.400	0.300	0.500	0.300	0.600	0.300	0.200	0.300	0.200	0.400	0.400	0.600	0.200
信頼度	0.714	0.429	0.429	0.571	0.833	0.833	0.333	0.667	0.429	0.714	0.429	0.857	0.600	0.400	0.600	0.400	0.800	0.800	1.200	0.400
リフト	1.190	0.612	0.857	1.143	1.190	1.190	0.667	1.333	0.612	1.190	0.857	1.714	0.857	0.667	0.857	0.800	1.143	1.333	1.714	0.800

(表5)から、支持度 50% 以上のケースに注目すると、併売によるリフトの値が大きいのは、商品 F を購入した上で商品 H を購入する場合とその逆の場合(リフトの値 1.714)であることが分かる。商品 F と商品 H との併売効果が期待できる。

5) 次いで、(表5)で支持度 50% 以上の2種類の併売品に関して、さらに3種類目の併売品を組み合わせるケースを確認する。2種類の商品 A と C を購入した上(A → C, C → A)で商品 F, 商品 G, 商品 H のいずれかを購入する場合、また、2種類として商品 C と F, および商品 F と H を購入した場合についても同様に考える。

(表6)3種類の商品と同時に購入したレシートの確認

レシートNo	商品の組合せ(3種類)						
	A・C・F	A・C・G	A・C・H	A・F・H	C・F・G	C・F・H	F・G・H
1	1	0	1	1	0	1	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0
7	1	1	1	1	1	1	1
8	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0
10	1	1	1	1	1	1	1
総計	5	2	3	3	2	4	2

これら3種類の商品を購入したときのレシートを(表6)に示す。

6) (表7)は、3種類の商品を購入したときのアソシエーション分析の指標の計算結果である。A・C → F は商品 A と C を購入の上で商品 F を購入したことを表す。計算式を以下、A・C → F の場合を代表して記す。

(表7)3種類の商品を購入するときの指標

商品組合せ	A・C→F	A・C→G	A・C→H	C・F→A	C・F→G	C・F→H	F・H→A	F・H→C	F・H→G
人気度	0.600	0.500	0.500	0.700	0.500	0.500	0.700	0.600	0.500
支持度	0.500	0.200	0.300	0.500	0.200	0.300	0.300	0.300	0.200
信頼度	1.000	0.400	0.600	1.000	0.400	0.600	0.500	0.500	0.333
リフト値	1.667	0.800	1.200	1.429	0.800	1.200	0.714	0.833	0.667

商品 F の人気度 = (表3)の ⑥ / 10 = 6 / 10 = 0.6

商品 A, C, F の支持度 = (表6)の ⑦ / 10 = 5 / 10 = 0.5

商品 A, C, F の信頼度 = (表6)の ⑦ / (表4)の ② = 5 / 5 = 1

商品 F の購入によるリフトの比率 = (表7)信頼度 / (表7)人気度 = (表7)の ⑨ / (表7)の ⑧ = 1 / 0.6 = 1.667

(表7)から、支持度 50% 以上のケースに注目すると、併売によるリフトの値が大きいのは、商品 A と C を購入した上で商品 F を購入する場合(リフトの値 1.667)と商品 C と F を購入した上で商品 A を購入する場合(リフトの値 1.429)であることが分かる。商品 A と商品 C と商品 F との併売効果が期待できる。

参考にした書籍

- 河本薫：「会社を変える分析の力」、講談社現代新書
小山昇：「数字は人格」、ダイヤモンド社
永野裕之：「ビジネス×数学＝最強」、すばる舎
竹内薫：「数学×思考＝ざっくりといかにして問題をとくか」、丸善出版
中西達夫：「統計データをすぐに分析できる本」、アニモ出版
中西達夫：「すぐれた判断が統計データ分析から生まれる」、実務教育出版
豊田裕貴：「マンガでわかる ビジネスを成功に導くデータ分析」、ナツメ社
向後千春、富永敦子：「統計学がわかる」、技術評論社
石井俊全：「意味がわかる統計学」、ベレ出版
涌井良幸、涌井貞美：「中学数学でわかる統計の授業」、日本実業出版
涌井良幸、涌井貞美：「統計学の図鑑」、技術評論社
西内啓：「統計学が最強の学問である」、ダイヤモンド社
西内啓：「統計学が最強の学問である(実践編)」、ダイヤモンド社
西内啓：「統計学が最強の学問である(ビジネス編)」、ダイヤモンド社
森岡毅、今西聖貴：「確率思考の戦略論」、角川書店
デビッド・マクアダムス：「世界一流企業はゲーム理論で決めている」、ダイヤモンド社
河村真一ほか：「本物のデータ分析力が身に付く本」、日経BPムック
末吉正成、末吉美貴：「Excel ビジネス統計分析 この分析できますか?」、翔泳社
谷岡一郎：「社会調査のウソ リサーチ・リテラシーのすすめ」、文藝春秋
林知己夫：「調査の科学」、ちくま学芸文庫
八谷大岳：「データ解析」シリーズ 全15回、(例)[データ解析 第1回 ベクトルの復習 - YouTube](#)