

ホームセンターにおける売上データの分析 ～アソシエーション分析を中心として～

南山大学情報理工学部 田中豊
南山大学数理情報研究科 橋本淳樹

研究体制

- ▶ オペレーションズ・リサーチ(OR):
 - ▶ 鈴木敦夫 教授 , ホームセンターからの研究員
RA(大学院生) , 学部学生(卒業研究)
- ▶ 統計学:
 - ▶ 田中豊 教授 , ホームセンターからの研究員
RA(大学院生) , 学部学生(卒業研究)

2008年度の研究テーマ: OR

1. 買い上げ点数増加のための最適店舗レイアウト*
2. 店舗回遊路を考慮した最適レイアウト
3. クロスMD商品の最適配置*
4. 折込広告の最適決定問題について*

*: アソシエーション分析と最適化

2008年度の研究テーマ: 統計学

1. クロスMD商品の選定*
2. 広告サイズの売上・客数への影響の分析
3. 店舗の売上予測とグループ化

*: アソシエーション分析, 多次元尺度法, クラスタ分析

アソシエーション分析(Visual Mining Studio)の応用事例

1. 部門レイアウトの統計的分析
2. 部門の最適レイアウトの設計
3. クロスMD商品の選定

アソシエーション分析

▶ 小売店などのPOSデータ・トランザクションデータから、頻出する商品の組み合わせの規則を漏れなく抽出し、その中から価値ある規則(相関ルール)を探し出すことが目的

▶ 「商品Aを買うと商品Bも買う」のような相関ルールを「A→B」の形で表す

アソシエーション分析の主な評価指標

- X,A,B: 商品の集合
- support(X) = (Xを含むレシート数) / (全レシート数)
- ▶ サポート
 $support(A \rightarrow B) = support(A \cup B) (=support(B \rightarrow A))$
- ▶ 信頼度
 $confidence(A \rightarrow B) = support(A \rightarrow B) / support(A)$
 $confidence(B \rightarrow A) = support(A \rightarrow B) / support(B)$
- ▶ リフト
 $lift(A \rightarrow B) = confidence(A \rightarrow B) / support(B) (=lift(B \rightarrow A))$

確率を用いた説明(1)

- ▶ A: 商品Aを購入するという事象
- ▶ B: 商品Bを購入するという事象
- ▶ $A \cap B$: 商品AとBを同時に購入するという事象
- ▶ $P()$: ()内の起こる確率

確率を用いた説明(2)

支持度(support) = $P(A \cap B)$

商品AとBを
同時購入する確率

確信度(confidence) $(A \rightarrow B) = \frac{P(A \cap B)}{P(A)}$

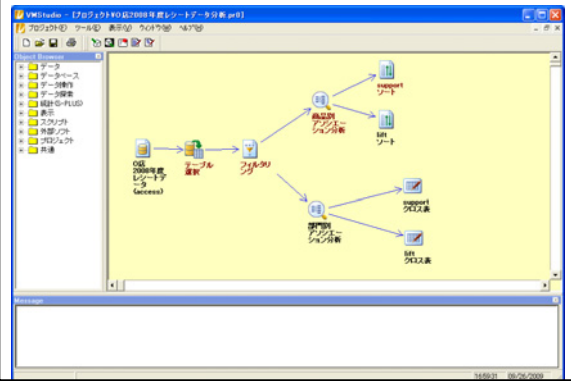
商品Aを購入したという
条件のもとで、商品Bを
購入する条件付き確率

確信度(confidence) $(B \rightarrow A) = \frac{P(A \cap B)}{P(B)}$

改善率(lift) = $\frac{P(A \cap B)}{P(A)P(B)} = \frac{P(A \cap B)}{P(A)} / P(B)$

・条件付き確率の無
条件確率に対する
改善率
・AとBが無関係なら
lift=1

VMStudioの分析フロー



アソシエーション分析結果

検出されたルール(support値上位抜粋)

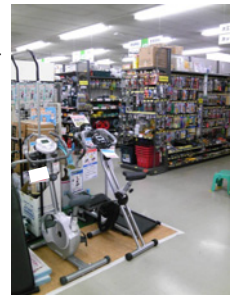
ルールのID	商品名	支持度	信頼度	リフト	商品数	レシート数	レシート数	レシート数	レシート数	レシート数
1	商品名	支持度	信頼度	リフト	商品数	レシート数	レシート数	レシート数	レシート数	レシート数
2	A市指定品	0.33	0.43	1.68	1675	4043	8577	509805		
3	A市指定品	0.33	0.43	1.23	1675	8577	4043	509805		
4	トレイロー	0.14	0.95	1.15	705	4876	7719	509805		
5	エルモアアシシュー	0.14	0.95	1.09	705	7719	4876	509805		
6	増化器HP フランラップ	0.13	0.46	1.14	685	5441	2903	509805		
7	増化器HP フランラップ	0.13	0.46	1.34	685	2903	5441	509805		
8	A市指定品	0.13	0.46	1.76	647	1499	4043	509805		
9	A市指定品	0.13	0.46	1.19	647	4043	1499	509805		
10	ラクスリリフレッシュCD	0.13	4.00	4.78	640	809	1008	509805		
11	ラクスリリフレッシュCD	0.13	4.00	2.73	640	1008	809	509805		
12	ビオラ レギュラー	0.11	0.32	1.34	581	2247	2454	509805		
13	ビオラ レギュラー	0.11	0.32	1.30	581	2454	2247	509805		
14	プロットリンス つめかえ用	0.10	343.7	1.96	523	1066	728	509805		
15	プロットリンス つめかえ用	0.10	343.7	3.54	523	728	1066	509805		
16	PRGバンタンEX	0.10	394.7	3.58	513	711	932	509805		
17	PRGバンタンEX	0.10	394.7	2.22	513	932	711	509805		
18	スピア キッチンタオル	0.10	8.40	1.08	508	6153	4943	509805		
19	スピア キッチンタオル	0.10	8.40	1.10	508	4943	6153	509805		
20	A市指定品	0.10	19.64	1.47	491	1499	8577	509805		
21	A市指定品	0.10	19.64	1.06	491	8577	1499	509805		
22	プリントロー	0.09	4.56	1.06	482	6977	7719	509805		
23	エルモアアシシュー	0.09	4.56	1.05	482	7719	6977	509805		
24	増化器HP フランラップ	0.09	5.74	1.08	479	5441	7719	509805		

2008年度1年間のレシート(509805件)のsupport上位

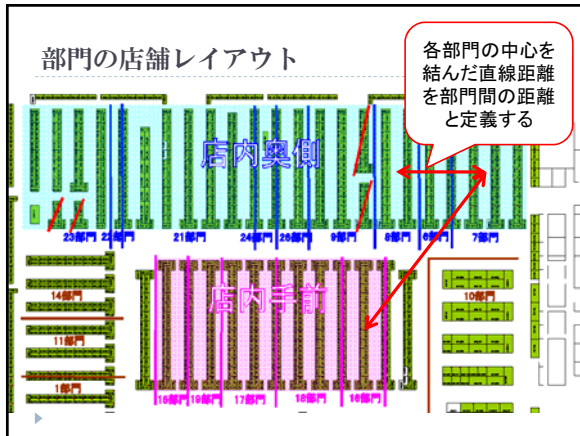
事例1：部門レイアウトの統計的分析

～多次元尺度法による部門配置の検討～

1. 部門間のアソシエーション分析
2. Support, Liftにもとづく
多次元尺度解析

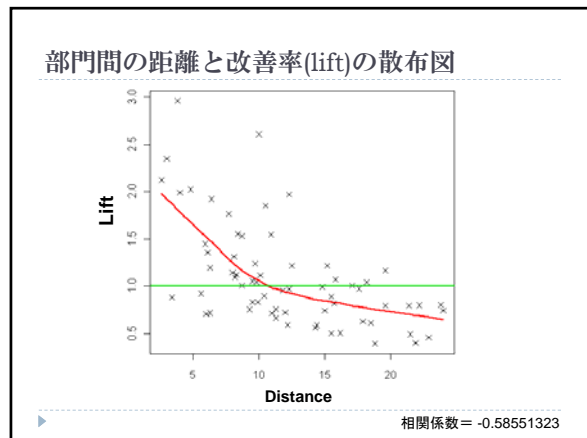
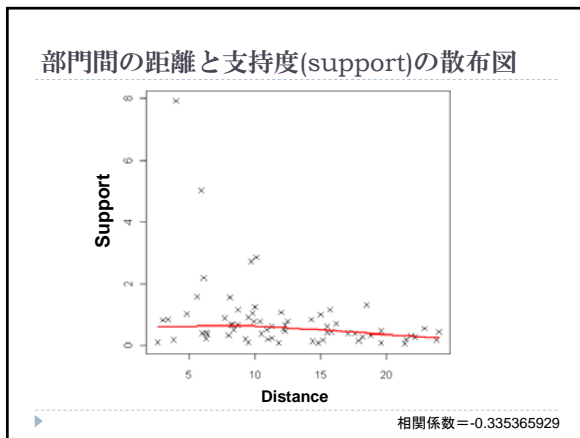


店内部門配置の様子



部門間のlift

	部門-7	部門-8	部門-9	部門-10	部門-11	部門-14	部門-15	部門-16	部門-17	部門-21	部門-22	部門-23
部門-7		2.024	1.763	2.607	0.745	0.46	0.613	0.896	0.742	1.074	0.797	0.799
部門-8	2.024		2.349	1.544	1.168	0.394	0.564	1.123	0.666	0.712	0.993	0.627
部門-9	1.763	2.349		1.967	1.004	0.501	0.72	1.093	0.833	1.143	0.957	0.587
部門-10	2.607	1.544	1.967		0.492	0.404	0.507	0.719	0.589	1.042	0.794	0.801
部門-11	0.745	1.168	1.004	0.492		0.883	0.923	0.887	0.832	0.758	1.054	0.755
部門-14	0.46	0.394	0.501	0.404	0.883		1.447	0.819	1.24	1.006	1.194	0.708
部門-15	0.613	0.564	0.72	0.507	0.923	1.447		1.115	1.99	1.315	1.553	1.044
部門-16	0.896	1.123	1.093	0.719	0.887	0.819	1.115		1.355	0.971	1.215	0.974
部門-17	0.742	0.666	0.833	0.589	0.832	1.24	1.99	1.355		1.534	1.852	1.22
部門-21	1.074	0.712	1.143	1.042	0.758	1.006	1.315	0.971	1.534		2.962	1.921
部門-22	0.797	0.993	0.957	0.794	1.054	1.194	1.553	1.215	1.852	2.962		2.121
部門-23	0.799	0.627	0.587	0.801	0.755	0.708	1.044	0.974	1.22	1.921	2.121	



多次元尺度法 (Multidimensional Scaling, MDS)

- ▶ n個の対象間の観測された非類似度 $\{s_{ij}\}$ の情報にもとづいて、対象を低い次元の空間内に配置する方法
- ▶ 計量的多次元尺度法(Metric MDS)
 - ▶ 観測された非類似度 $\{s_{ij}\}$ を距離とみなして、対象間の距離をできるだけ再現するように対象を空間内に配置する方法
- ▶ 非計量的多次元尺度法(Non-metric MDS)
 - ▶ 観測された非類似度 $\{s_{ij}\}$ の順序関係をできるだけ再現するように対象を空間内に配置する(Kruskal, 1964)

非計量多次元尺度法 (Non-metric Multidimensional Scaling)

- ▶ 再現できない度合を表す指標: ストレス

$$STRESS = \frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i \sum_j d_{ij}^2}$$

s_{ij} : 観測された非類似度
 d_{ij} : MDS空間上の (i, j) 間の距離
 \hat{d}_{ij} : d_{ij} の s_{ij} への単調回帰のあてはめ値

単調関係

$s_{jk} > s_{lm} \Rightarrow \hat{d}_{jk} \geq \hat{d}_{lm}$

$s_{jk} = s_{lm} \Rightarrow \hat{d}_{jk} = \hat{d}_{lm}$

非計量多次元尺度法 (部門間のsupport, liftの分析)

- ▶ 部門間の支持度 (support) と改善率 (lift) の逆数を非類似度としてKruskalの非計量MDSを適用し、2次元空間上での各部門の配置を求めた

(逆数に変換したLiftデータ一部抜粋)

	部門-7	部門-8	部門-9	部門-10	部門-11	部門-14
部門-7		0.494071	0.567215	0.383583	1.342282	2.173913
部門-8	0.494071		0.425713	0.647668	0.856164	2.538071
部門-9	0.567215	0.425713		0.508388	0.996016	1.996008
部門-10	0.383583	0.647668	0.508388		2.03252	2.475248
部門-11	1.342282	0.856164	0.996016	2.03252		1.132503
部門-14	2.173913	2.538071	1.996008	2.475248	1.132503	
部門-15	1.631321	1.77305	1.388889	1.972387	1.083424	0.691085

isoMDSの実行結果

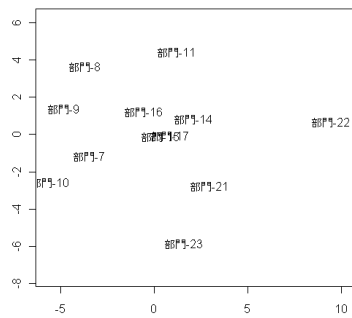
- ▶ supportの逆数を非類似度とした時のストレスの推移

	Stress	
initial value	27.612702	Final Stress 13.308887
iter 5 value	13.998077	
final value	13.308887	
converged		

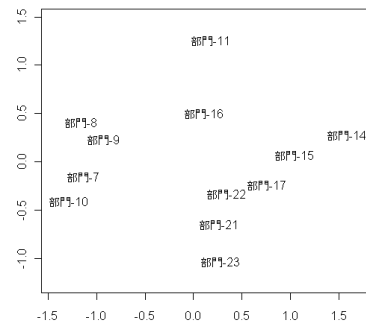
- ▶ liftの逆数を非類似度とした時のストレスの推移

	Stress	
initial value	14.901559	Final Stress 9.914947
iter 5 value	10.665580	
iter 10 value	10.029492	
final value	9.914947	
converged		

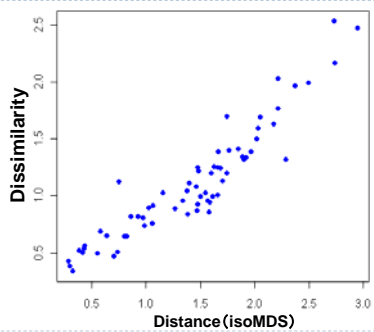
MDSによる2次元空間上の部門配置(support)



MDSによる2次元空間上の部門配置(lift)

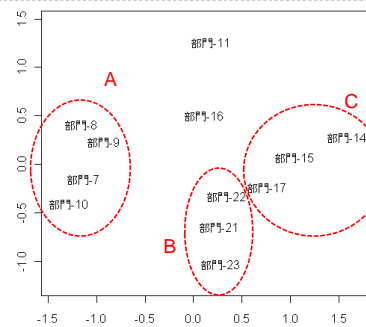


MDS空間上の距離と非類似度(liftの逆数)の散布図



相関係数 = 0.93035703

MDSによる2次元空間上の部門配置(lift)



部門間のliftクロス表 (並べ替えた結果)

	部門7	部門8	部門9	部門10	部門21	部門22	部門23	部門17	部門15	部門14	部門11	部門16
	A				B			C				
部門7		2.024	1.763	2.807	1.074	0.797	0.799	0.742	0.613	0.46	0.745	0.896
部門8	2.024		2.349	1.544	0.712	0.983	0.827	0.686	0.564	0.394	1.168	1.123
部門9	1.763	2.349		1.967	1.143	0.957	0.587	0.833	0.72	0.501	1.004	1.093
部門10	2.807	1.544	1.967		1.042	0.794	0.901	0.589	0.507	0.404	0.492	0.719
部門21	1.074	0.712	1.143	1.042		2.982	1.921	1.534	1.315	1.006	0.758	0.971
部門22	0.797	0.983	0.957	0.794	2.982		2.121	1.852	1.353	1.194	1.054	1.215
部門23	0.799	0.827	0.587	0.801	1.921	2.121		1.22	1.044	0.708	0.755	0.974
部門17	0.742	0.686	0.833	0.589	1.534	1.852	1.22		1.99	1.24	0.832	1.335
部門15	0.613	0.564	0.72	0.507	1.315	1.353	1.044	1.99		1.447	0.923	1.115
部門14	0.46	0.394	0.501	0.404	1.006	1.194	0.708	1.24	1.447		0.883	0.819
部門11	0.745	1.168	1.004	0.492	0.758	1.054	0.755	0.832	0.923	0.883		0.887
部門16	0.896	1.123	1.093	0.719	0.971	1.215	0.974	1.335	1.115	0.819	0.887	

▶ Lift値が1.2以上、2以上に色付け

事例2:部門の最適レイアウトの設計

～買い上げ点数増加のための最適店舗レイアウト～

1. レシートデータのアソシエーション分析
2. 得られた関連数を用いて、0-1整数計画問題を定式化
 - ▶ 部門間の関連数と距離の逆数の積の最大化
3. 最適な店舗レイアウトを作成

0-1整数計画問題の定式化

▶ 目的変数

$$\max \sum_{i \in N} \sum_{j \in N} D_{ij} M_{ij} x_{ij}$$

D_{ij} : 部門*i, j*の関連数 (同時購入頻度)

M_{ij} : 部門*i*と隣り合う部門*j*の距離の逆数

x_{ij} : 部門*i*と*j*が隣り合うときに1, そうでないときに0

i, j : 部門($i \in N, j \in N$) N : 部門の添字集合 n : 部門の個数

実行結果を利用した部門配置



▶ 部門配置変更の提案

▶ インテリアと寝具

▶ 作業用品と塗料・補修

事例3:クロスMD商品のレイアウト

▶ 商品のレイアウト

▶ 統計学:
クロスMD商品の選定

▶ OR:
クロスMD商品の最適配置



クロスMD商品陳列の様子

クロスMD商品の選定

▶ “ついで買い”されやすい商品をクロスMD商品として選定したらどうか?

▶ Aを購入したときBを購入する確率: 大 (確信度(A→B): 大)

▶ Bを購入したときAを購入する確率: 小 (確信度(B→A): 小)



“ついで買い”される商品とは

“確信度(A→B) > 2 × 確信度(B→A)”

という相関ルールを持つ商品Bは商品Aのついで買い?

クロスMD商品の選定

◆しかし、

$$\frac{P(A \cap B)}{P(A)} > 2 \times \frac{P(A \cap B)}{P(B)} \quad \text{より} \quad P(B) > 2 \times P(A)$$

となり、AとBの関連性に関係ない。この指標だけでは...

⇒ どういうものがクロスMD商品とされているか？

- ▶ 現在クロスMD商品とされている商品の支持度、確信度、改善率について検討中

結論と考察

- ▶ アソシエーション分析は有用な分析ツールである
 - ▶ どんな商品が同時購入されやすいかについては会社の人は経験に基づく知識と強い関心を持っている。
 - ▶ 得られた情報は最適化(売上げ、粗利、etc.)に結びつけやすい。
 - ▶ 大量のデータに対する第1段階分析法として有用
 - ▶ いきなり多次元尺度法やクラスター分析法を適用するより、まずアソシエーション分析を行った後、第2段階として適用する方がよい。
- ▶ VMStudioは使えるソフトである
 - ▶ Rのパッケージ"arules"と比べて操作が簡単でスピードが格段に速い。