2002年日本オペレーションズ・リサーチ学会
秋季研究発表会

データマイニング支援の事例

－　ＯＲリテラシーの普及事例（第５報）　－

０１１０２３４５　オーアールとく塾　権藤　元　　GONDO　Hajime

１．はじめに

Ａ社のデータ処理をしているＢ社の担当者に対して、Ａ社のデータマイニングに関連する定式化とでもいう作業を支援した事例の報告である。Ｂ社の担当者はデータマイニングのソフトの機能はよく理解しているが、本来の目的に対してどう対処するかという段階になるとお粗末といわざるを得ず、この局面でＯＲリテラシーが有効であることを体験した。その事例を２つ紹介する。１つは避けたい異常な事態例えば退会という現象への対応に役立つ情報を手に入れたいのであるが，その退会という現象のデータのみをデータベースから抽出して何か掴まらないか首をひねっているという状態で、もう１つは膨大なデータベースから関心の深い事項例えば代理店別・商品別・ユーザの属性・使用期間などの影響を評価したいのであるが．それらの周辺分布あるいは２次元分布からいろいろと評価をしようとしている状態である。第１の事例にはマハラノビスの距離を求め、ある閾値を越えるときに異常発生を予知する方式を導入し、第２の事例には周辺分布に拘らず多次元の各セルを対象に関心事項を評価するため、いくかの回帰モデルを描きながらモデルを選択する方式を導入した。ともにある程度の成果をあげた。

２．マハラノビスの距離の事例

退会という異常な事態のみ対象としないで、継続しいるメンバーからより正常なグループを把握して、それを対象にマハラノビスの基準空間を設け、退会者はマハラノビスの距離が大きくなるようなデータ項目を探した。その結果、ある程度の識別はできたが、データ項目に基本的なデータを欠落していることが判明し、以後そのデータ項目をデータベースに収録することによりその効果をあげることができた。なお、エクセルによるマハラノビスの距離算出については既発表[(1)]を参照されたい。

３．回帰分析の事例

代理店別・商品別・ユーザの属性・使用期間の影響を求めた１例を図表１に示す。データ項目はカテゴリ化し、その組み合わせにより生じるセルの内容はデータマイニングのソフトからエクセルに連結して得られる。特性を目的変数としカテゴリ化された説明変数の係数を最小２乗法で求める。これは、エクセルのソルバーを使用することにより，数ケースの回帰分析を一度に実行でき説明変数の選択・合成などに便利であった。

４．おわりに

今回のデータマイニング支援を始めてすぐに次のことに気がついた。それは、「データマイニングのツールとは適切なパラメーターを入力するとひとりでにねらいとする結果が出力されるもの。」と思い込んでおり、「自分で実態を表現するモデルをいくつか作り、データマイニングのツールを用いて抽出編集したデータによりどのモデルが適切かを自分で判断するもの」という考えを持たないことである。そこで、迂遠と思われたかも知れなかったが，ＯＲリテラシー（注参照）を身につけてもらうために、文献２をテキストとして促成教育を行った。その結果、Ｂ社の担当者は自らいろいろと仮設を立てモデルを作り、その検証用のデータをデータマイニングのツールを用い取得して、モデルを選択評価することができるようになった。ＯＲリテラシーを唱えたメンバーの一人としてまことに感慨深いものがある。ご意見をお待ちしている。　Ｅメール：hajime.gondo @nifty.com

注　ＯＲリテラシー（文献２のあとがきより抜粋）

何とかＯＲを世に広め使ってもらうにはどうしたらよいかということで、1990 年に学会の有志がＯＲ広報研究部会を作った。研究部会はその後ＯＲリテラシー研究部会さらにＯＲリテラシー研究グループと名称は変わっているが、その中から文献２のテキストが誕生した。これら一連の研究部会の初期の時点では、広めたいＯＲとは一体何かを確認することから始め、さらに情報リテラシーに対応してＯＲリテラシーというものを提唱した。… 途中省略 …、つまり、ＯＲリテラシーはＯＲ専門家が持つものというよりは、広くビジネスマンや各層の意思決定者に知っておいていただきたいＯＲの考え方や方法の常識とでもいうものである。

図表1 ソルバーによる回帰分析

	A	B	C	D	E	F	G	H	I	J	K	L	M
1				係数									
2	ケース0:ウエイト平均			21.86									
3	ケース1:ABCDすべて			23.76	-2.80	0	1.15	-0.79	0	0.50	0.82	-6.11	-3.0
4	ケース2:BCD(除くA)			23.06	0	0	0	-0.52	0	0.64	0.95	-6.14	-3.3
5	ケース3:ACD(除くB)			23.87	-2.77	0	1.13	0	0	0	0	-6.09	-3.0
6	ケース4:ABD(除くC)			23.48	-3.11	0	1.06	-0.60	0	0.70	0.60	0	0
7	ケース5:ABC(除くD)			23.21	-2.39	0	1.60	-1.53	0	-0.30	0.81	-6.17	-3.02
8		周辺分布			-2.68	0.00	1.48	-0.92	0.00	0.23	0.69	-6.16	-3.26
9		平均		21.86	19.63	22.31	23.79	20.92	21.84	22.07	22.53	16.56	19.46
10	セル名	ウエイト	特性	定数	A1	A2	A3	B1	B2	B3	B4	C1	C2
11	A1B1C1D1	20	7.0	1	1	0	0	1	0	0	0	1	0
12	A1B1C1D2	73	10.2	1	1	0	0	1	0	0	0	1	0
13	A1B1C1D3	3	15.8	1	1	0	0	1	0	0	0	1	0
14	A1B1C1D4	81	16.4	1	1	0	0	1	0	0	0	1	0
15	A1B1C2D1	52	10.9	1	1	0	0	1	0	0	0	0	1
16	A1B1C2D2	73	14.4	1	1	0	0	1	0	0	0	0	1
17	A1B1C2D3	42	16.2	1	1	0	0	1	0	0	0	0	1
18	A1B1C2D5	56	21.8	1	1	0	0	1	0	0	0	0	1
19	A1B1C3D2	19	18.3	1	1	0	0	1	0	0	0	0	0
	途中省略												
264	A3B4C5D5	19	32.8	1	0	0	1	0	0	0	1	0	0

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1									分散	標準偏差				
2									28.86	5.37				
3	0	2.06	4.02	-5.80	-2.85	0	2.62	3.40	0.77	0.88				
4	0	2.05	4.02	-5.55	-3.11	0	2.65	3.53	3.46	1.86				
5	0	2.04	4.00	-5.84	-2.77	0	2.71	3.43	1.15	1.07				
6	0	0	0	-5.73	-3.62	0	2.57	2.57	13.88	3.73				
7	0	1.88	3.64	0	0	0	0	0	12.95	3.6	61.07	<--分散の合計		
8	0.00	1.84	3.61	-5.48	-3.92	0.00	2.61	2.63	***** 残差 *****					
9	22.72	24.57	26.33	17.44	19.00	22.92	25.53	25.55	28.86	0.77	3.46	1.15	13.88	12.95
10	C3	C4	C5	D1	D2	D3	D4	D5	ケース0	ケース1	ケース2	ケース3	ケース4	ケース5
11	0	0	0	1	0	0	0	0	-14.9	-1.3	-3.9	-2.2	-7.0	-6.1
12	0	0	0	0	1	0	0	0	-11.7	-1.0	-3.1	-2.0	-5.9	-2.9
13	0	0	0	0	0	1	0	0	-6.1	1.7	-0.6	0.8	-4.0	2.7
14	0	0	0	0	0	0	1	0	-5.5	-0.3	-2.7	-1.3	-5.9	3.3
15	0	0	0	1	0	0	0	0	-11.0	-0.4	-2.8	-1.4	-3.1	-5.4
16	0	0	0	0	1	0	0	0	-7.5	0.1	-1.7	-1.0	-1.7	-1.9
17	0	0	0	0	0	1	0	0	-5.7	-0.9	-3.0	-1.9	-3.6	-0.1
18	0	0	0	0	0	0	0	1	-0.1	1.3	-0.9	0.2	-0.5	5.5
19	1	0	0	0	1	0	0	0	-3.6	1.0	-1.1	0.0	2.2	-1.0
	途中省略													
264	0	0	1	0	0	0	0	1	10.9	-0.4	1.2	0.4	5.1	3.6

埋め込み関数の例示　この例示をコピーすることで完成する

L7　=SUM(V9:AA9)

F9　=+SUMPRODUCT(D11:D264,$C$11:$C$264,$B$11:$B$264)/SUMPRODUCT(D11:D264

J9　=SUMPRODUCT(V11:V264,V11:V264,$B$11:$B$264)/SUM($B$11:$B$264)

E11　=IF(ISERROR(FIND(E$10,$A11,1)),0,1)

J11　=+C11-$D$2

K11　=$C11-SUMPRODUCT($D$3:$U$3,$D11:$U11)

参考文献

(1) 権藤、血栓症予測に用いたマハラノビス距離基準空間選定方法の確認シミュレーション－ＯＲリテラシーの普及事例（第３報）－、ＯＲ学会秋季研究発表会予稿集、2000.9

(2) 高井・真鍋編著、問題解決のためのオペレーションズ・リサーチ入門、日本評論社、2000.4、