日本オペレーションズ・リサーチ学会
2005年春季研究発表会

従来の整数計画法による判別モデル研究に対する批判

1202720　成蹊大学　新村秀一 SHINMURA Shuichi

1980 年代以降、整数計画法を用いた判別分析のモデルが数多く提案されているが、それらが統計分野で注目されたことはない。その理由を、Liittschwager, J.M. & Wang, C. (1978) (以下、Liittschwager モデルと略す) と新村が開発した IP-OLDF と比較することで例示する。

1 Liittschwager モデル

Liittschwager らは、線形判別関数を次の整数計画法モデルとして定式化している。ここで、P_i, Q_i, D_r, E_r, は 0/1 の決定変数。$c_1, c_2, \cdots, c_k$, b は自由変数。C は大きな整数。f_1 と f_2 はケース数に比例した事前確率。g_1 と g_2 は誤分類のリスク。M と N は 2 群のケース数。k は説明変数の数である。

$$\text{Min } f_1 g_1 M^{-1} \Sigma_{(i=1,\cdots,M)} P_i + f_2 g_2 N^{-1} \Sigma_{(j=1,\cdots,N)} Q_j$$

St

$$c_1x_{i1}+c_2x_{i2}+\cdot\cdot\cdot+c_kx_{ik} \leqq b+CP_i, \quad i=1,2,\cdots,M, \qquad (2.7)$$

$$c_1y_{j1}+c_2y_{j2}+\cdot\cdot\cdot+c_ky_{jk} \geqq b-CQ_i, \quad j=1,2,\cdots,N, \qquad (2.8)$$

$$-1+2D_r \leqq c_r \leqq 1-2E_r, \quad r=1,2,\cdots,k, \qquad (2.9)$$

$$\Sigma_{(r=1,\cdots k)} D_r + \Sigma_{(r=1,\cdots,k)} E_r = 1 \qquad (2.10)$$

これに対して、新村は標本誤分類数を最小化基準にして、データ空間の双対空間である判別係数の空間で次の IP-OLDF を定式化した。

$$\text{Min } \Sigma_{(i=1,\cdots,M)} e_i + \Sigma_{(j=1,\cdots,N)} e_j$$

St

$$c_1x_{i1}+c_2x_{i2}+\cdot\cdot\cdot+c_kx_{ik} +1 \leqq Ce_i, \quad i=1,2,\cdots,M,$$

$$c_1y_{j1}+c_2y_{j2}+\cdot\cdot\cdot+c_ky_{jk} +1 \geqq -Ce_i, \quad j=1,2,\cdots,N,$$

IP-OLDF は、次の点で Liittschwager モデルの亜流と誤解される。

1) 事前確率とリスクを 1 に固定し、
2) 誤分類確率最小化基準の変わりに標本誤分類数最小化基準を用い、
3) b の変わりに定数項を 1 に固定する事で、

形式的に Liittschwager モデルから導かれる。ただし、IP-OLDF では式(2.10)の制約を用いる必要はない。

しかし、Liittschwager モデルは式(2.10)の制約を除いて、単に判別関数を定式化したに過ぎない。

これに対して、IP-OLDF は統計的判別分析の理論を踏まえた上で、誤分類数最小化基準と定数項を 1 に固定する事で、判別係数の空間で定式化でき、それが Fisher の線形判別関数と異なった優位な情報を得られる事を示した点に新規性がある。しかも、Liittschwager モデルは実際のデータを分析せず、ヒューリスティック手法を提案している。この分野のヒューリスティック手法は、最適化手法の代替案として意味がああり、それ単独では存在意義は無いといえよう。

2. 総当り法による IP-OLDF と Liittschwager モデルの比較

表 1 は、「スイス銀行偽札データ」を用いて、IP-OLDF と Liittschwager モデルを比較したものである。

「OLDF」列は IP-OLDF の誤分類数を、「Fisher」列は Fisher の線形判別関数の誤分類数を、「差」は「=Fisher-OLDF」を、「Wang」は Liittschwager モデルを IP-OLDF に作り変えたモデルの誤分類数、「LDF=0」判別境界点上のケース数を表す。多くの統計研究家や数理計画法の研究家は、判別境界上のケースの扱いを正確に捉えていない。IP-OLDF はそれを考慮しているので、「Wang の差」は (=「Wang」+「LDF=0」－「OLDF」) で計算を行った。制約式(2.10)は単に解の空間を狭めているだけであり、この結果 OLDF と大きく結果が異なるモデルがあることが分かる。

表 1　IP-OLDF と Liittschwager モデルの比較　(3 変数モデルの結果を省く)

Mpdel	p	OLDF	Fisher	差	Wang	LDF=0	Wangの差
length,left,right,bottom,top,diagonal	6	0	1	1	0	6	6
left,right,bottom,top,diagonal	5	0	1	1	0	5	5
length,right,bottom,top,diagonal	5	0	1	1	0	4	4
length,left,bottom,top,diagonal	5	0	1	1	0	5	5

length,left,right,bottom,diagonal	5	0	1	1	0	5	5
length,left,right,bottom,top	5	2	7	5	3	5	6
length,left,right,top,diagonal	5	1	2	1	1	5	5
right,bottom,top,diagonal	4	0	1	1	0	4	4
left,bottom,top,diagonal	4	0	1	1	0	3	3
length,bottom,top,diagonal	4	0	1	1	0	4	4
left,right,bottom,diagonal	4	0	1	1	0	5	5
length,right,bottom,diagonal	4	0	1	1	0	4	4
length,left,bottom,diagonal	4	0	2	2	2	4	6
length,right,bottom,top	4	2	7	5	5	4	7
left,right,bottom,top	4	2	6	4	4	4	6
length,left,bottom,top	4	2	8	6	5	4	7
length,right,top,diagonal	4	1	2	1	1	4	4
left,right,top,diagonal	4	1	1	0	1	3	3
length,left,right,diagonal	4	1	2	1	1	4	4
length,left,top,diagonal	4	1	1	0	1	4	4
length,left,right,bottom	4	14	19	5	12	4	2
length,left,right,top	4	19	21	2	18	7	6
bottom,diagonal	2	0	3	3	0	0	0
bottom,top	2	3	8	5	58	7	62
right,diagonal	2	1	2	1		2	1
top,diagonal	2	1	1	0	1	4	4
left,diagonal	2	1	1	0	1	2	2
length,diagonal	2	1	2	1	1	3	3
right,bottom	2	14	19	5	27	7	20
left,bottom	2	14	20	6	22	7	15
length,bottom	2	14	18	4	20	8	14
right,top	2	25	29	4	–		
left,top	2	30	33	3	–		
length,right	2	34	36	2	–		
length,top	2	32	45	13	–		
left,right	2	34	44	10	–		
length,left	2	41	47	6	–		
diagonal	1	1	2	1	1	2	2
bottom	1	14	17	3	100	1	87
top	1	42	49	7			
right	1	34	43	9			
left	1	42	53	11			
length	1	67	100	33			

3．まとめ

判別データとして統計学で有名な「偽札データ」を用いて、IP-OLDF と Liittschwager モデルの比較を行なった。その結果、後者は真の誤分類数を最小化するモデルで無いことが分かった。

文献

1） Liittschwager, J.M. & wang, C. (1978). Integer programming Solution of a Classification Problem, Management Science, 24(14), 1515-1525.