

scpQCA

A New and more Powerful QCA Algorithm

分享人:复旦大学复杂决策分析中心 傅曼青

指导老师:唐世平





01 QCA的因果逻辑

02 QCA解决问题的类别

03 scpQCA的贡献

04 scpQCA的有效性

QCA的因果逻辑 必要条件



必要条件:如果一个条件总在某个结果产生时出现,那么这个条件即该结果产生的必要条件。换句话说,没有该条件,该结果就无法产生。

$$consistency_N(C) = \frac{\sum_{i=1}^{n} c_i = 1 | o_i = 1}{\sum_{i=1}^{n} o_i = 1}$$

必要一致性阈值>0.9 (Ragin 2008)

QCA的因果逻辑 充分性分析



充分条件:如果一个结果总在某个条件出现时产生,那么这个 条件即该结果产生的充分条件,但这一结果同样可以产生于其 他条件下。

$$consistency_{S}(C) = \frac{\sum_{i=1}^{n} c_{i} = 1 | o_{i} = 1}{\sum_{i=1}^{n} C_{i} = 1}$$

充分一致性阈值>0.75 (Ragin 2008)

- **多重并发因果关系原则**:多个不同条件的组合产生结果;多个不同的条件组合产生相同的结果;组合中的条件取值可能不同。
- **复杂性&简约性**:QCA关注因果关系的多样性并且遵从奥卡姆 剃刀原则。

QCA的因果逻辑 QCA与机器学习



- 贝叶斯网络
- 决策树

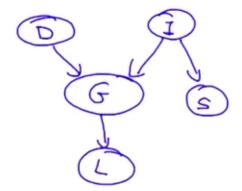
"……在排除虚无假设的过程中,这种方法可以缩小找到真实假设的范围。而且,即使这种方法不能够排除所有的无关条件,它也可以让我们在一定程度上接近并发现某个现象的'发生条件'……"(Cohen & Nagel, 1934, p.267)

QCA的因果逻辑 QCA与机器学习



- Grade
- · Course Difficulty
- Student Intelligence
- Student SAT
- Reference Letter

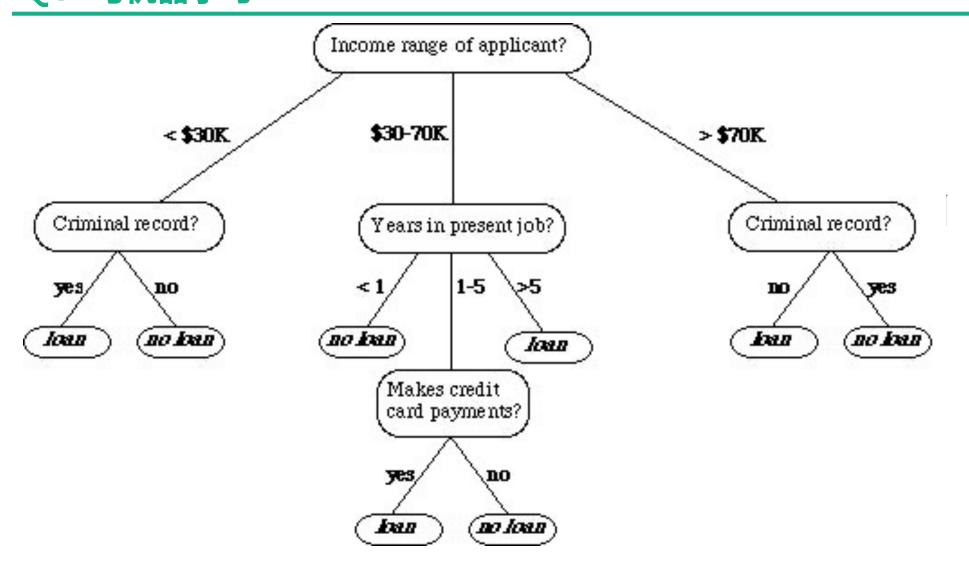
P(G,D,I,S,L)



▲ via Coursera——Probabilistic Graphical Models

QCA的因果逻辑 QCA与机器学习





QCA解决的问题类别 csQCA



- csQCA (清晰集) :
 - 构建二分数据表;
 - 构建"真值表" (truth table) ;
 - 解决矛盾组态;
 - 布尔最小化;
 - 纳入"逻辑余项";
 - 解释。(Ragin 2009)

QCA解决的问题类别 布尔最小化

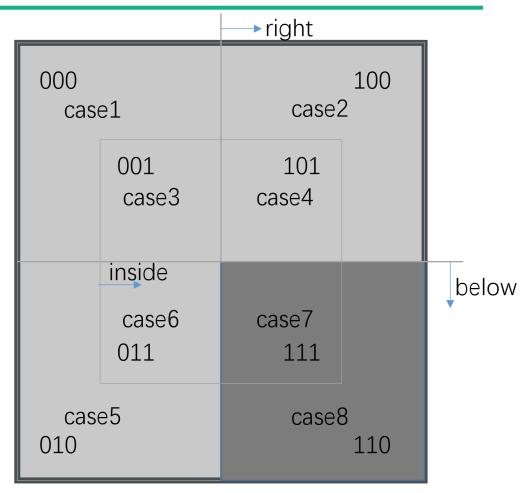


caseid	right	below	inside	out	come
case1		0	0	0	0
case2		1	0	0	0
case3		0	0	1	0
case4		1	0	1	0
case5		0	1	0	0
case6		0	1	1	0
case7		1	1	1	1
case8		1	1	0	1

 $RIGHT * BELOW * INSIDE + RIGHT * BELOW * inside \rightarrow OUTCOME$ (1)

 $RIGHT * BELOW \rightarrow OUTCOME$ (2)

如果, case6的outcome也为1, 纳入逻辑余项case5后结果变为为: $BELOW \rightarrow OUTCOME$ (3)



▲真值表对应的韦恩图

QCA解决的问题类别 mvQCA



• mvQCA(多值集):多值条件(可以从多分类定类、定序和定距的多个阈值中获得);最小化(条件的所有取值均满足才能进行化简)。

描述	年龄范围(岁)	mvQCA值
婴儿	0~1	0
幼童	2~5	1
童年	6~10	2
少年	>11	3

$$A_1B_0 + A_1B_1 + A_1B_2 \rightarrow O$$

$$A_1 \rightarrow O$$

QCA解决的问题类别 fsQCA



- fsQCA(模糊集):同时拥有定性和定量的属性,结合了集合隶属的类别(kind)和程度(degree)。
- Ragin的直接校准法和间接校准法。

清晰集	三值模糊集	四值模糊集	六值模糊集	"连续"模糊集
1=完全隶属	1=完全隶属	1=完全隶属	1=完全隶属	1=完全隶属
0=完全不隶属	0.5=既非完全隶属,也非完全不隶属	0.67=偏隶属	0.9=非常隶属	偏隶属: $0.5 < X_i < 1$
	0=完全不隶属	0.33=偏不隶属	0.6=有些隶属	0.5=交叉点,既非隶属也非 不隶属
		0=完全不隶属	0.4=有些不隶属	偏不隶属:
			0.1=非常不隶属	$0 < X_i < 0.5$
			0=完全不隶属	0=完全不隶属

QCA解决的问题类别 scpQCA与qmcQCA功能的对比



数据特征	scpQCA	qmcQCA						
清晰集(csQCA)	清晰集(csQCA)							
样本数量	不限	<=30						
变量数量	不限	<8						
多值集(mvQCA)								
样本数量	不限	<=30						
变量数量	<16	<8						
变量取值数量	<=6	<=2						
模糊集(fsQCA)								
样本数量	不限	不限						
变量数量	<16	<8						
校准	N值模糊集不需要(N<=6)	需要						

- scpQCA可以对所有非连续型变量进行计算;
- scpQCA可以免除校准过程;
- scpQCA可以打破样本和变量 数量上的限制;
- scpQCA可以无人工操作,保证结果和过程的透明度;
- scpQCA可以输出唯一解,不 再有简约解和复杂解的影响。

scpQCA的贡献 1.候选路径表



• 利用候选路径表代替真值表。

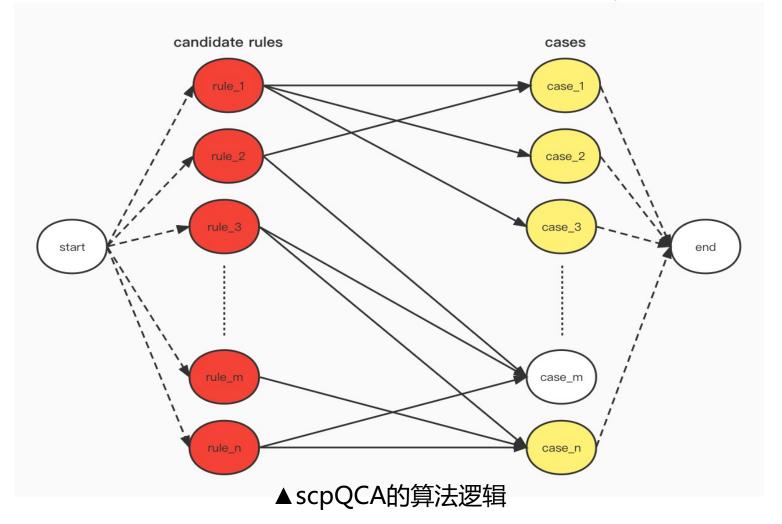
			conditions					
row	aknowledge	sknowledge	ecdiversity	meltingpot	selfexpress	number	consistency	coverage
1	1	-	-	-	1	57	0.8421	0.5393
2	1	-	-	1	-	53	0.8302	0.4944
3	1	1	-	-	-	51	0.8235	0.4719
4	-	-	1	1	-	45	0.8	0.4045
5	-	-	1	-	1	43	0.814	0.3933
	-	1	-	1	-	40	0.85	0.382
1	1	-	1	-	1	42	0.8095	0.382
3	1	1	-	1	-	39	0.8462	0.3708
)	1	-	-	1	1	40	0.8	0.3596
.0	-	-	-	1	1	40	0.8	0.3596
1	1	1	-	-	1	35	0.8	0.3146
2	-	1	-	-	1	35	0.8	0.3146
3	1	1	-	1	1	30	0.8	0.2697
4	-	1	-	1	1	30	0.8	0.2697

▲ (Rutten 2019)数据的候选路径表

scpQCA的贡献 2.集合覆盖算法化简



• 用贪心集合覆盖算法取代Quine-McCluskey。



scpQCA的贡献 Large-N dataset



• Rutten (2019)数据集,主要探究地域特征对产业创新的影响因素。共有108个事件,5个自变量。

	qmc	:QCA	scpQCA			
Configuration	[]	III	IV	1	2	3
Analytical knowledge	•					•
Synthetic knowledge	•	•			•	
Economic diversity	•					
Melting pot						
Self-expression						
Solution coverage	0.7	895		0.8315		
Solution consistency	0.9	441		0.8506		
Consistency threshold	0.	85		0.8		
Frequency threshold	ļ	5	5			
Truth table rows/Candidate rules	9 rd	OWS	14 rules			

▲(Rutten 2019)用不同QCA的结果比较

scpQCA的贡献 Multi factors dataset



• Schneider (2019)数据集,主要探究政体特征与投资维和基金的因果关系。共有22个事件,8个自变量。

	Schneider2019			На	esebrouck20	019	scpQCA			
Configuration	I	II	III	A	В	C	1	2	3	
MC		•		•			•			
MS					0	0		0	0	
PI			•			•				
GP										
LE	•	•	•	*	•	*		•		
LP	•			•		•	•			
ED	•	•	•	*	*	*	•	•		
PV		0	•					0		
Covered cases	BE, ES, FR, GR, PL, PT	BE, FR, GB, GR, IT, PL, PT	CZ, FI, IE, SK	BE, ES, FR, GR, PL, PT	BE, ES, FR, GR, IT, PL, PT	FI, FR, PL, PT	BE, ES, FR, GR, PL, PT	BE, FR, GR, IT, PL, PT	FI, FR, PL, PT	
Consistency	1	0.8571	0.5	1	1	1	1	1	1	
Unique coverage	1	2	4	0	1	1	1	1	1	
Solution coverage	1			0.8888			0.8888			
Solution consistency	0.8182			1			1			

▲ (Schneider 2018)数据用不同QCA的结果比较

scpQCA的有效性 内部有效性



• 内部有效性测试是检查 QCA 在不同的一致性和频率阈值下是 否仍产生稳定的结果。

Test		A	•	В				C			
configuration	1	2	3	1	2	3	4	1	2	3	4
Analytical knowledge		•	•		•	•	•		•	•	•
Synthetic knowledge		•							•		
Economic diversity	•			•				•			•
Melting pot	•			•				•			
Self-expression			•			•				•	
Solution coverage		0.8315		0.8652				0.8652			
Solution consistency		0.8506		0.8462			0.8462				
Consistency threshold	0.8		0.75			0.7					
Frequency threshold	5		5			5					
Num. of Candidate rules		14 rules			19 rules			25 rules			

▲scpQCA在(Rutten 2019)数据上的internal validity检验

scpQCA的有效性 内部有效性



 内部有效性测试是检查 QCA 在不同的一致性和频率阈值下是 否仍产生稳定的结果。

Test		A			В			С	•	-	D
Configuration	1	2	3	1	2	3	1	2	3	1	2
MC										•	
MS		0	0		0	0		0	0		0
PI											
GP											
LE							•				
LP							•		•	•	
ED							•	•			
PV		0			0			0			0
Solution coverage		0.8888		0.8888			0.8888			0.7778	
Solution consistency		1		1		1		1			
Consistency threshold		0.8		0.8		0.8		0.8			
Frequency threshold		2		3		4			5		
Num. of candidate rules		59 rules			37 rules		13 rules			9 rules	

scpQCA的有效性 外部有效性



- 外部有效性测试是检查 QCA 在抽取少部分案例后是否仍产生稳定的结果。
- *用python展示scpQCA的使用方法、有效性和预测能力

scpQCA判断类别	解释
Perfectly correct	只满足当前结果的条件特征
Confusion mistake	满足当前结果和其他结果的条件特征
Totally wrong	完全不满足当前结果的特征,但满足其他结果的条件特征
Not found	不符合任何结果的条件特征



