最大广义信念熵模型的推导

2020 年 4 月 14 日

设A_i是辨识框架Ω的子集,A_i表示集合A_i的基数,$m(A_i) \in [0,1]$是幂集2^Ω的概率指派函数

1 取得最大广义信念熵的重要条件

刘帆提出的最大广义信念熵模型为:

$$E_{t,r}(m(A_i)) = \frac{1}{1-r}\left[\left[\sum_i (\frac{m(A_i)}{2^{|A_i|}-1})^t (2^{|A_i|}-1)\right]^{\frac{1-r}{1-t}} - 1\right]$$

又因为概率指派函数的性质:

$$\sum_i m(A_i) = 1$$

于是求解上面广义信念熵最大值模型的问题就转化为在由概率指派函数的性质所确定的条件下，求解广义信念熵模型的最大值。这个问题可以通过拉格朗日乘数法来求解，我们可以定义这个问题的拉格朗日函数如下:

$$F_{t,r}(m(A_i)) = \frac{1}{1-r}\left[\left[\sum_i (\frac{m(A_i)}{2^{|A_i|}-1})^t (2^{|A_i|}-1)\right]^{\frac{1-r}{1-t}} - 1\right] + \lambda(\sum_i m(A_i) - 1)$$

接着，我们通过这个拉格朗日函数对$m(A_i)$求偏导，因此我们可以得到:

$$\frac{\partial F_{t,r}(m(A_i))}{\partial m(A_i)} = \frac{1}{1-t}\left[\sum_i (\frac{m(A_i)}{2^{|A_i|}-1})^t (2^{|A_i|}-1)\right]^{\frac{1-r}{1-t}} t(\frac{m(A_i)}{2^{|A_i|}-1})^{t-1} + \lambda = 0$$

这里面，t, λ在这种情况下都是常数，因此我们对上面的式子进行如下处理:

$$\left[\sum_i (\frac{m(A_i)}{2^{|A_i|}-1})^t (2^{|A_i|}-1)\right]^{\frac{r-1}{t-1}} (\frac{m(A_i)}{2^{|A_i|}-1})^{t-1} = \frac{\lambda(t-1)}{t}$$

然后等式两边再同时乘$(t-1)$次方:

$$\left[\sum_i (\frac{m(A_i)}{2^{|A_i|}-1})^t (2^{|A_i|}-1)\right]^{r-1} (\frac{m(A_i)}{2^{|A_i|}-1})^{(t-1)^2} = \left(\frac{\lambda(t-1)}{t}\right)^{t-1}$$

到这里我们可以看得比较清楚了，中括号里面的式子对于一个给定的辨识框架来说是一定的。同时，上面提到过，t,λ 是常数，所以我们可以得到:

$$\left(\frac{m(A_1)}{2^{|A_1|}-1}\right)^{(t-1)^2} = \left(\frac{m(A_2)}{2^{|A_2|}-1}\right)^{(t-1)^2} = ... = \left(\frac{m(A_i)}{2^{|A_i|}-1}\right)^{(t-1)^2}$$

这也就是:

$$\frac{m(A_1)}{2^{|A_1|}-1} = \frac{m(A_2)}{2^{|A_2|}-1} = ... = \frac{m(A_i)}{2^{|A_i|}-1}$$

根据中学的学习的合比性质，我们可以得到:

$$\frac{m(A_1)+m(A_2)+...+m(A_n)}{(2^{|A_1|}-1)+(2^{|A_2|}-1)+...+(2^{|A_n|}-1)} = \frac{m(A_i)}{(2^{|A_i|}-1)}$$

这也就是:

$$\frac{1}{\sum_j 2^{|A_j|}-1} = \frac{m(A_i)}{2^{|A_i|}-1}$$

因此，我们可以得到:

$$m(A_i) = \frac{2^{|A_i|}-1}{\sum_j 2^{|A_j|}-1}$$

所以广义信念熵取得最大值的充要条件是:$m(A_i) = \frac{2^{|A_i|}-1}{\sum_j 2^{|A_j|}-1}$

其实这里只是求得了一个极值点，还有一个极值点就是$r=1$。$r=1$确切地说是函数的一个奇点，而刘帆的文章中已经证明了当$r\to 1$时，广义信念熵模型退化为 Reny-Deng 熵，所以在这种情况下问题就转化为求Reny-Deng熵的最值。关于Reny-Deng熵的最值，在下面的小节中会说明。(这段话没有被写在论文中)

2 广义信念熵最大值模型

我们将得出的充要条件:

$$m(A_i) = \frac{2^{|A_i|}-1}{\sum_j 2^{|A_j|}-1}$$

代入广义信念熵模型后我们可以得到:

$$F_{t,r}(m(A_i)) = \frac{1}{1-r}\left[\left[\sum_i \left(\frac{2^{|A_i|-1}}{\sum_j 2^{|A_j|}-1}\right)^t \times \left(\frac{1}{2^{|A_i|-1}}\right)^{t-1}\right]^{\frac{1-r}{1-t}} - 1\right]$$

接着我们对第二个中括号里面的进行化简后可以得到:

$$F_{t,r}(m(A_i)) = \frac{1}{1-r}\left[\left[\left(\sum_i 2^{|A_i|}-1\right)^{-t} \times \left(\sum_j 2^{|A_j|}-1\right)\right]^{\frac{1-r}{1-t}} - 1\right]$$

进一步进行化简我们可以得到:

$$E_r(A_i)_{max} = \frac{1}{1-r}\left[\left(\sum_j 2^{|A_j|}-1\right)^{1-r} - 1\right]$$

这就得到了最大广义信念熵模型

3 最大R-D熵和最大T-D熵

在上面提出的最大广义信念熵模型中，命题A_i是客观的，因此通过讨论参数r的取值来得到最大广义信念熵的最大值。对上面的最大广义信念熵模型求关于r的偏导数，即:

$$\frac{\partial E_r(A_i)_{max}}{\partial r} = \frac{\left(\sum_j 2^{|A_j|}-1\right)^{1-r}\left[-\ln\left(\sum_j 2^{|A_j|}-1\right)(1-r)-1\right]}{(1-r)^2} = 0$$

接着，可以得到两个r值:

$$r = 1 + \frac{1}{\ln\left(\sum_j 2^{|A_j|}-1\right)}, r = 1$$

将第一个r值代入最大广义信念熵模型中可以得到:

$$E_r(A_i)_{max} = \ln\left(\sum_j 2^{|A_j|}-1\right)\left[1 - \left(\sum_j 2^{|A_j|}-1\right)^{-\ln\left(\sum_j 2^{|A_j|}-1\right)^{-1}}\right]$$

因为 $u^v = e^{vlnu}$,将其应用于中括号内部的乘积式子，可以得到:

$$E_r(A_i)_{max} = \ln\left(\sum_j 2^{|A_j|}-1\right)(1-\frac{1}{e})$$

而上式是肯定小于: $\ln\left(\sum_j 2^{|A_j|}-1\right)$,即:

$$E_r(A_i)_{max}=\ln\left(\sum_j 2^{|A_j|}-1\right)(1-\frac{1}{e})<\ln\left(\sum_j 2^{|A_j|}-1\right)$$

而当$r\to 1$时，(这就是我们上面提到的(也是刘帆论文中证明了的)原来的广义信念熵模型退化为R-D模型的条件，这里因为在求得最大广义信念熵模型的过程中，始终没有对式子前面的 $\frac{1}{1-r}$进行本质变化，所以这个奇点的含义没有变化。这句话也没有写在论文中)，通过洛必达法则可以得到:

$$\begin{aligned}\lim_{r\to 1}\frac{1}{1-r}\left[\left(\sum_j 2^{|A_j|}-1\right)^{1-r}-1\right]&=\lim_{r\to 1}\frac{\left(\sum_j 2^{|A_j|}-1\right)^{1-r}\times\ln\left(\sum_j 2^{|A_j|}-1\right)}{1}\\&=\ln\left(\sum_j 2^{|A_j|}-1\right)\end{aligned}$$

所以，综上，最大广义信念熵模型的最大值:

$$E_r(A_i)_{max}=\ln\left(\sum_j 2^{|A_j|}-1\right)$$

3.1 最大R-D熵

从刘帆的论文得知，当$r\to 1$时，广义信念熵模型退化为R-D熵，同时，他的论文还指出R-D熵取得最大值的条件和这里所得出的条件相同即:

$$m(A_i)=\frac{2^{|A_i|}-1}{\sum_j 2^{|A_j|}-1}$$

但是我们的模型中却没有关于$m(A_i)$的式子，只有通过$r\to 1$的值来得到R-D熵的最大值。这里您可能会有这样的问题就是通过$r\to 1$来得到R-D熵的最大值到底靠不靠谱，就是说这里$m(A_i)$的式子代入R-D熵模型和通过$r\to 1$来得到R-D熵的最大值是否一样，我论文中算例2就是算的这个，结果是一样的，但是不知道的是$r\to 1$和这里的$m(A_i)$有什么关系。

这里将r的值推向1，其实上面已经讨论过了，最后得出的最值和上面的一样:

$$E_r(A_i)_{max}=\ln\left(\sum_j 2^{|A_j|}-1\right)$$

3.2 最大T-D熵

最大R-D熵的证明可能复杂一点，我不知道是不是我想复杂了。还有一点就是刘帆的论文中文字部分提到当$r \to t$时，广义信念熵模型退化为T-D熵，但是配图上标注的是当$r \to q$时，广义信念熵退化为T-D熵。共同点就是T-D熵取得最大值的条件都是一致的，这两种方式:$r \to t$ $r \to q$ 我都计算过了。其中前一种方式比较简单，算出来的最大T-D熵的结果和我们得出的最大广义熵模型是一致的。(当$r \to t$这种简单情况我没有在论文中提及)

比较复杂的是当$r \to q$时的最大T-D熵的求解:

当$r \to q(q > 0, q \neq 1)$时，我们可以得到:

$$
\begin{aligned}
&\lim_{r \to q} \frac{1}{1-r}\left[\left(\sum_j 2^{|A_j|}-1\right)^{1-r}-1\right] \\
&= \frac{1}{1-q}\left[\left(\sum_j 2^{|A_i|}-1\right)^{1-q}-1\right] \\
&= \frac{1}{1-q}\left[e^{(1-q)\ln\left(\sum_j 2^{|A_j|}-1\right)}-1\right]
\end{aligned}
$$

然后我们对上式进行进行幂级数展开:

$$
\begin{aligned}
originformula =& \frac{1}{1-q}\left[1+(1-q)\ln\left(\sum_j 2^{|A_j|}-1\right)\right. \\
&\left. + \sum_{i=2}^{n} \frac{(1-q)^n\left[\ln\left(\sum_j 2^{|A_j|}-1\right)\right]^n}{n!}-1\right] \\
=& \ln\left(\sum_j 2^{|A_j|}-1\right)+\frac{1}{1-q}\sum_{i=2}^{n}\frac{(1-q)^n\left[\ln\left(\sum_j 2^{|A_j|}-1\right)\right]^n}{n!}
\end{aligned}
$$

这里n是推向$+\infty$的。接着，我们对于q的范围分类:

当$0 \leq q<2$时，$1-q$在$[-1,1)$内，而对于一个确定的辨识框架来说,

$\left(\sum_j 2^{|A_j|}-1\right)$是确定的，因此我们可以得到:

$$\lim_{n\to\infty}\sum_{i=2}^{n}\frac{(1-q)^n\left[\ln\left(\sum_j 2^{|A_j|}-1\right)\right]^n}{n!}=0$$

故此时，T-D上的最大值是: $\ln\left(\sum_j 2^{|A_j|}-1\right)$

当$q>2$时，我们可以对最开始的式子(开始$r\to q$讨论的式子)进行变化，而得到以下式子:

$$\ln\left(\sum_j 2^{|A_j|}-1\right)+\frac{1}{1-q}\left[e^{(1-q)\ln\left(\sum_j 2^{|A_j|}-1\right)}-(1-q)\ln\left(\sum_j 2^{|A_j|}-1\right)-1\right] \tag{1}$$

我们现在考察式子:

$$e^{(1-q)\times\ln\left(\sum_j 2^{|A_j|}-1\right)}-(1-q)\ln\left(\sum_j 2^{|A_j|}-1\right)-1$$

令

$$x=(1-q)\ln\left(\sum_j 2^{|A_j|}-1\right)(x<0)$$

则考察的式子变为这样一个函数：

$$f(x)=e^x-x-1(x<0)$$

我们对它求导可以得到:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x}=e^x-1=0$$

得出$x=0$, 即函数$f(x)$在$x<0$时单调递减，故我们可以得出:

$$f(x)>f(0)=0$$

因此考察的式子是大于0的，但是$(1-q)$是小于0的，所以可以得到:

$$(1)<\ln\left(\sum_j 2^{|A_j|}-1\right)$$

故T-D熵的最大值也是: $\ln\left(\sum_j 2^{|A_j|}-1\right)$