2K－6

抽象的内容の判断評価に関する研究

江原信郎　明治大学
江原義郎　順天堂大学

１　緒　言

抽象的なことについて論ぜよといわれた場合、人は程度の差は別として何らかの論点からある論述を試みることができる。このような論述がなされた場合、これを評価してあるグレイドをつけなければならぬことがある。多くの場合、直観的にともかく採点することができる。それは、自分が出題者ではない場合であっても、また、採点の基準が示されないような場合であっても、比較的容易に行なうことができる。このような判断評価を求められた場合、人はいったいどのように行なうのか、そのメカニズムを検討していこうとするものが本研究の目的である。

２　直観的な評価

ここでは、直観的な採点評価、３項目・３段階の項目別評価、および７項目１０段階の項目別評価を考える

評価の尺度が”よい・ふつう・おそまつ”というような非常に漠然としたものであるから、ファジイ論理を用いた表現を考えることもできよう。また、専門的な知識を持つ立場のものが採点するのであるから、その知識をエキスパートシステムとして取り扱うといった立場も考えることができる。しかし判断の基本的な要因を探ろうとする第一歩としては、自分はどこにどれほどのウエイトをかけて考えているのかといった情報の方が役にたつことが多い。

具体例とした問題は、大学院において提示した次のようなレポート課題である。

”新しい概念（理論でも方式でも哲学、アイディアでもよい）にもとずく新しい制御（理論でもシステムでも実現技術でもよい）とその効果について４００字で論ぜよ。”

提示されたレポートに対して、普通は直観的にある点数を割り振る。このような採点を行なう場合、一般に点数の境界条件がある。たとえば、大学院では何点以上をつけようとか、全体的にみた分布はどの程度が好ましいかというような要因である。このようないわゆる総合的判断にもとずいて直観的評価を行った一例が、表１に示す直観採点の項である。

３　評価式による評価アルゴリズム

２で述べた採点評価を例えば３項目の評価要因で表わすことは、かなり乱暴な近似では有るが、それにもかかわらず評価の特徴を把握することができる。

いま、評価すべき項目(I_1,I_2,I_3)の評価点を(x_1,x_2,x_3)とし、その係数を(a_1,a_2,a_3)、ある定数値をｄとする。評価点数Ｙを、

$$Y=a_1x_1+a_2x_2+a_3x_3+d \quad (1)$$

と表わす。

(1)式に示した評価式をつくる基本的な手順は次の通りである。

１　最低のグレード$(x_1,x_2,x_3)=(1,1,1)$において何点（P_0）を割り当てるかを決める。

２　３項目(I_1,I_2,I_3)についての重み付け(α,β,γ)を行なう。このばあい1:2:4とか1:2:3とかその程度の重み付けでよい。係数とこの重みとの関係を、kをある定数として

$$(a_1,a_2,a_3)=(k\alpha,k\beta,k\gamma) \quad (2)$$

とする。

３　各項目についてｒ段階の評価を行なう場合には、最高点(r,r,r)の得点に対して全体が100点となるように定数ｋを決め、このｋを用いて係数$(a_1,a_2,a_3)=(k\alpha,k\beta,k\gamma)$を決める。$(x_1,x_2,x_3)=(1,1,1)$に対して(1)式の値が$P_0$、$(x_1,x_2,x_3)=(r,r,r)$に対して(1)式の値が100となるのであるから(2)式を用いて、

$$k=(100-P_0)/(r-1)(\alpha+\beta+\gamma) \quad (3)$$

$$d=P_0-(\alpha+\beta+\gamma)k=(P_0r-100)/(r-1) \quad (4)$$

与えられたP_0、(α,β,γ)に対して、(1)式は、

$$Y=\frac{(100-P_0)}{(r-1)(\alpha+\beta+\gamma)}(\alpha x_1+\beta x_2+\gamma x_3)+(P_0r-100)/(r-1) \quad (5)$$

４　(5)式のような評価式を用いた評価の得点と直観的な評価の得点との比較を行ない、両者の差が最小となるように重み(α,β,γ)を修正する。この場合、ふたつの得点の差の大小は、考える全データに対する得点の差の２乗和で比較する。もちろん、この場合線形回帰アルゴリズムを用いて、係数を決めることもできるが、演算量は増える。

４　システムによる評価検討

表１は20件の対象レポートについて、同一採点者が直観的評価、および項目別の評価点付与を行ったものである。

ここで、３項目・３段階の評価項目としては、

１　新しさ・奇抜さ・おもしろさ
　　あり・どちらともいえない・なし
２　論旨の展開の良さ・完結度
　　あり・どちらともいえない・なし
３　取り組みの真剣さ・美しさ
　　よい・ふつう・おそまつ

を用い、７項目の場合には、１：概念の新しさ・奇抜さ　２：制御イメージの新しさ・奇抜さ　３：効果の説得性　４：論旨展開の良さ　５：可能性・妥当性の有無　６：取り組みの真剣さ　７：論文文章としての質の良さを用いた。

この項目別の評価に対して、$(\alpha,\beta,\gamma)=(3,2,1)$、(2,1,1)、(1,1,1)の重みをかけた場合の採点計算を示したものが表２である。この計算結果による採点結果を表１の直観的な採点と比較すると、両者の差の二乗和は、

Basic study of human judgement.

Noburo EHARA　Meiji Univ.
Yoshiro EHARA　Juntendo Univ.

(3,2,1)で527.9、(2,1,1)で506.8、(1,1,1)で657.3となり、(2,1,1)の場合が誤差が一番小さくなる。このとき、1件当りの点数の平均差は±5.0である。

与えたP_0のもとで、評点(x_1,x_2,x_3)に対して、(α,β,γ)を(10,10,10)から(1,1,1)までかえて、(3)(4)(5)をもとめ、計算による採点Yと直観的な採点との二乗誤差の和を最小とする(α,β,γ)を求めることができる。表1では、(α,β,γ)=(2,1,1)が最小となる。

表3は、別の20件のレポートに対して、出題者（採点者1）と専門以外のエキスパート（採点者2）との評価の結果を示したものである。なお、表3に示した項目別の評点は採点者1のものである。ここで、B*=90　B=85　C=75　とし、P_0=55として、採点者1について(5)式を導く。採点者1の項目別の評点は、(2,3,3)(2,3,1)(2,1,1)と非常に類型的なものであり、この結果は直観的な採点結果と概ね良くあっている。この場合の重み(α,β,γ)を、二乗誤差の和最小の条件から求めると、(4,1,1)となる。このとき、ひとつのデータ毎の平均の推定誤差は±4.7である。この場合採点者1は、”新しさ”の項目を非常に重視して採点したことになる。表1・表2の場合と比べると、(2,1,1)が(4,1,1)となったわけであるから、”新しさ”へのウエイトがかなり大きくなったことを示している。

このような試みを何回か（n回）くりかえすと最大n本の係数の異なる評価式が得られる。このn本の評価式の係数が大幅に異ならない場合にはこの評価はおおむね一定の状態で評価されたと判断することができる。

表3の採点者2の採点結果は採点者1の場合とはかなり異なっている。これだけのデータから、採点者2がどのような考えで採点しているのかを類推することは困難であるが、試みに各レポートに対する評点(x_1,x_2,x_3)が、採点者1の場合と同じであると仮定すれば、重み(α,β,γ)は(4,1,1)もしくは(3,1,1)となり、このときのひとつのデータ毎の平均誤差は±7.9もしくは±8.6となる。すなわち、同じ項目評点であるとすれば、重みを同じ(4,1,1)と考えた場合に採点者1に比べてこの推定は約2倍の誤差を含んでいることになる。

ここでは3項目・3段階の例を考えたが、同じ評価問題に対して7項目・10段階評価を試みた結果に本質的な差はみられない。

5　結　言

本報告ではまず、抽象的な課題のレポートについて、直観的な採点、ある程度の項目に分けた採点評価を試み、その結果を用いて一つの評価式をつくり出した。次にこのような評価をもとに、直観的な採点評価の分析を試み、直観的な採点評価がどの様な立場から行われたものであるのかの推定を行なうことを検討した。また、例えば200-400件というようなたくさんのレポートを、時間をかけて採点する際の採点者の内部状態の変化を時々刻々トレースする方法を考えた。

今後さらに多くの事例を検討し、特に表3に示した採点者2の採点根拠を類推する方法などについて考えていきたい。

表1　直観的採点と項目別評点

対象	直観採点	(x_1,x_2,x_3)
R1	90	3 3 2
R2	90	3 3 2
R3	90	3 2 2
R4	85	3 2 2
R5	85	3 2 2
R6	85	3 2 1
R7	85	2 3 2
R8	85	2 2 2
R9	85	2 2 2
R10	85	2 2 2
R11	85	2 2 1
R12	85	2 1 2
R13	75	3 1 1
R14	75	2 1 2
R15	75	2 1 1
R16	75	1 2 2
R17	65	1 1 2
R18	65	1 1 2
R19	65	1 1 2
R20	55	1 1 1

表2　重みと採点計算結果

	(321)	(211)	(111) 重み
R1	96	94	92
R2	96	94	92
R3	88	88	85
R4	88	88	85
R5	88	88	85
R6	85	83	77
R7	85	83	85
R8	77	77	77
R9	77	77	77
R10	77	77	77
R11	73	71	70
R12	70	71	70
R13	77	77	70
R14	70	71	70
R15	66	66	62
R16	66	66	70
R17	58	60	62
R18	58	60	62
R19	58	60	62
R20	55	55	55

表3　採点評価データ

	直観的採点 採点者1	直観的採点 採点者2	採1評点 (x_1,x_2,x_3)
D1	C	B	2 1 1
D2	B	B	2 3 1
D3	C	B	2 1 1
D4	B	C	2 3 1
D5	C	C	2 1 1
D6	C	B	2 1 1
D7	B	B	2 3 1
D8	B	C	2 3 1
D9	B	C	2 3 1
D10	C	C	2 3 1
D11	B*	C	2 3 3
D12	B	C	2 3 1
D13	C	C	2 1 1
D14	B	C	2 3 1
D15	C	C	2 1 1
D16	B	B	2 3 3
D17	B	B*	2 3 3
D18	B*	B*	2 3 3
D19	B*	B*	2 3 3
D20	C	C	2 3 1