原 著

系列カテゴリー法による観察者の確信レベル（平均値）の測定

島 村 正 道・船 間 芳 憲*・吉 田 日出輝
内 田 武 博**・上 田 新 也***・村 上 誠 一****

熊本大学医療技術短期大学部診療放射線技術学科

*熊本大学医学部附属病院中央放射線部

**医療法人天野整形外科

***熊本市民病院中央放射線部

****産業医科大学病院放射線部

（論文受理 1994年 1月28日）

（最終論文受理 1995年 2月 4日）

（Code No. 520）

Key words : Method of successive categories, Normal distribution, Confidence-rating level, Interval scale

MEASUREMENT OF THE AVERAGE VALUES OVER OBSERVER'S CONFIDENCE-RATING LEVELS BY THE METHOD OF SUCCESSIVE CATEGORIES

MASAMICHI SHIMAMURA, YOSHINORI FUNAMA*, HIDETERU YOSHIDA,
TAKEHIRO UCHIDA**, SHINYA UEDA***, SEIICHI MURAKAMI****

Department of Radiological Technology, College of Medical Science, Kumamoto University

*Department of Radiology, Kumamoto University Hospital

**Department of Radiology, Amano Plastic Surgery

***Department of Radiology, Kumamoto City Hospital

****Department of Radiology, University of Occupational and Environmental Health Hospital

Summary

ROC analysis and rating by cumulated points have been widely employed for visual evaluations regarding medical images, which have no objective interval scales of the observer's confidence levels to categories in which their judgements regarding the presence or absence of subtle signals are ranked based on different types of point confidence-rating scales.

By the method of successive categories, measurements of mean values over their confidence levels to categories and distances among them were carried out under the following assumptions.

Observers ranked their judgements regarding the presence or absence of subtle signals on a 5-point confidence-rating scale. Their interpretation of the rating scale should be kept constant throughout this

experiment.

These experimental results suggest that the interval scales of confidence level of the rating to categories depend on the observers.

1. はじめに

医用画像の視覚的評価法の中で、観察者に観察試料の判定を数個のカテゴリに求め、得られたデータを基に解析を進める手法は、ROC 曲線 (Receiver Operating Characteristic Curve) や累積点法などがあり多くのすぐれた報告¹⁻⁷⁾がなされている。これらの評価法では、観察者のカテゴリ分類に対する判定の尺度 (以下、判定尺度) を、カテゴリ順位にしたがった順序尺度⁹⁾でとらえたり、実験者による主観的尺度を利用した解析が行われる場合が多い。一方、観察者は試料のカテゴリ分類に際しては固有の判定尺度の下で分類を行っているものであり、一様な判定尺度の下で分類を行っているとは考えられない。したがって、個々の観察者のカテゴリ分類に対する等間隔性^{8,9)}を問題にしたり、正確な判定尺度を測定して、累積点法等での評価を行う場合には、客観性のある測定法が必要となる。

われわれは、観察者が示した試料のカテゴリ分類に対する判定尺度を利用するために、観察者の試料分類の分布が正規分布をなすことを利用した系列カテゴリ法¹⁰⁾を適用して、判定尺度の平均値を測定した。さらに、これらの平均値間を距離尺度で測定して、観察者間の判定尺度の相違について検討した。

2. 材料および実験方法

観察試料は、厚さ16 cm のアクリル板に厚さ0.9 mm に重ねたフィルム片 (5 mmφ), アクリルビーズ (2 mmφ) を別々に密着して撮影した信号を含むものおよび信号を含まないもので各々30枚のX線フィルムを使用した。画像システムはスクリーン/フィルム系で Hi-screen/RX (以下 HiS/RX), HR-4/SHR-S の 2 種類を、Computed Radiography (以下 CR) による HiS/RX と等線量 (以下 CR (1)) および 1/2 線量 (以下 CR (1/2)) の画像処理有り と画像処理無し の 4 種類 (以下 CR (1)R, CR (1)L, CR (1/2)R, CR (1/2)L) の計 6 種類である。撮影条件は、管電圧70 kV でグリッド (10:1) を使用し、フィルム濃度が 0.7 ± 0.05 になるように調整した。CR の画像処理は Semi Auto Mode でラチチュード2.0に固定して行い、処理条件は次の通りである。

L 画像: GA1.1, GT D, GC0.6, GS-0.45, RN3, RT Q, RE1.0

R 画像: GA0.9, GT A, GC0.7, GS-0.74, RN3, RT Q, RE5.0

観察者は経験年数が3~17年の5名の放射線技師であ

る。

観察者に試料の分類を求めたカテゴリは、信号は確かにない (カテゴリ 1), たぶんない (カテゴリ 2), どちらともいえない (カテゴリ 3), たぶんある (カテゴリ 4), 確かにある (カテゴリ 5) の 5 つである。また、これらのカテゴリは観察者の判定軸上で連続体として分けられているものとし、カテゴリの境界は常に一定に保たれているものとした¹⁰⁾。観察者に示された試料は 6 画像システム別にアクリルビーズもしくはフィルム片の信号を含むもの30枚、信号を含まないもの30枚を無作為に並べた合計60枚を 1 セットにした計12セット (6 画像システム×2 信号) の720枚である。判定は上に示した5つのカテゴリに分類されるものとし、判定の作業は連続的に行った。得られたデータは経験的に認められているように観察者の判定軸上で正規分布をなすものと仮定した。

Fig. 1 にカテゴリ分類に対する観察者の判定尺度の測定方法を具体例で示した。示した資料は各観察者あたり、信号を含んでいる場合の 6 画像システム×2 信号の合計の試料数360枚より算出したものである。算出手法は、観察者により各カテゴリへ分類された試料枚数を求めて、確率に変換する (T1)。カテゴリ 1 からカテゴリ 5 の右方向に累積確率を求める (T2)。規準正規分布の曲線下の面積は確率を表しているの、左から右方向へカテゴリ 1 からカテゴリ 5 までの確率を順次とる。5つのカテゴリ分類には4個の境界が存在し、規準正規分布関数より境界の縦座標 (Y 座標) を求める (T3)。隣接するカテゴリの境界の縦座標の差を求める (T4)。どのカテゴリにも2つの境界が存在し、カテゴリ J の上限は、また1つ上のカテゴリ J+1 の下限でもある。したがって、次式より、標準偏差を単位として、あるカテゴリに対する判定尺度の平均値 Z_{ij} が求まる (T5)。

$$(y_1 - y_2) / p = Z_{ij} \quad (j: 2 \leq j \leq 5)$$

但 y_1 : カテゴリ j の下限の縦座標

y_2 : カテゴリ j の上限の縦座標

i: 観察者, p: カテゴリ j の確率

今、観察者 i においてカテゴリ 1 の判定尺度の平均値を基準値 0 とすれば、カテゴリ 2, 3, 4, 5 の平均値への距離 D_{ij} (i: 観察者, j: $j \geq 2$) と隣接するカテゴリの平均値間の距離 $d_{j-1,j}$ (j: カテゴリ番号, $2 \leq j \leq 5$) を算出できる (T6)。

同様にして信号を含まない場合の処理は、カテゴリ 5 を基準値 0 とすればカテゴリ 4, 3, 2, 1 の平均値の距離が測定できる。また、カテゴリ 3 を基準値 0 とればカテゴリ 4, 5, カテゴリ 2, 1 の平均値への距離を

OBSERVER	C1	C2	S+ C3	C4	C5	TOTAL
A	13	52	28	135	132	360
B	26	43	70	75	146	360
C	49	58	29	57	167	360
D	51	38	26	55	190	360
E	21	42	41	114	142	360
TOTAL	160	233	194	436	777	1800

T 1

OBSERVER	C1	C2	C3	C4	C5	TOTAL
A	0.0361111	0.1444444	0.0777777	0.375	0.3666666	1
B	0.0722222	0.1194444	0.1944444	0.2083333	0.4055555	1
C	0.1361111	0.1611111	0.0805555	0.1583333	0.4638888	1
D	0.1416666	0.1055555	0.0722222	0.1527777	0.5277777	1
E	0.0583333	0.1166666	0.1138888	0.3166666	0.3944444	1

T 2

OBSERVER	C1	C2	C3	C4	C5
A	0.0361111	0.1805555	0.2583333	0.6333333	1
B	0.0722222	0.1916666	0.3861111	0.5944444	1
C	0.1361111	0.2972222	0.3777777	0.5361111	1
D	0.1416666	0.2472222	0.3194444	0.4722222	1
E	0.0583333	0.175	0.2888888	0.6055555	1

T 3

OBSERVER	C1	C2	C3	C4	C5
A	0.0791	0.2633	0.3231	0.3766	--
B	0.1372	0.2731	0.3825	0.3878	--
C	0.2182	0.3461	0.3801	0.3973	--
D	0.2247	0.3157	0.3571	0.398	--
E	0.116	0.2578	0.3417	0.3848	--

T 4

OBSERVER	C1	C2	C3	C4	C5
A	-0.0791	-0.1842	-0.0598	-0.0535	0.3766
B	-0.1372	-0.1359	-0.1094	-0.0053	0.3878
C	-0.2182	-0.1279	-0.034	-0.0172	0.3973
D	-0.2247	-0.091	-0.0414	-0.0409	0.398
E	-0.116	-0.1418	-0.0839	-0.0431	0.3848

T 5

OBSERVER	C1	C2	C3	C4	C5
A	-2.190461	-1.275230	-0.768857	-0.142666	1.0270909
B	-1.899692	-1.137767	-0.562628	-0.02544	0.9562191
C	-1.603102	-0.793862	-0.422068	-0.108631	0.8564550
D	-1.586117	-0.862105	-0.573230	-0.267709	0.7541052
E	-1.988571	-1.215428	-0.736682	-0.136105	0.9755492

T 6

OBSERVER	C1	C2	C3	C4	C5	C5-C1
A	0	0.9152307	0.5063736	0.6261904	1.1697575	3.2175524
B	0	0.7619248	0.5751388	0.5371885	0.9816591	2.8559114
C	0	0.8092399	0.3717931	0.3134373	0.9650866	2.4595571
D	0	0.7240123	0.2888744	0.3055216	1.0218143	2.3402229
E	0	0.7731428	0.4787456	0.6005776	1.1116545	2.9641207

Fig. 1 Protocol of calculation of the mean values over observer's confidence levels for detections of subtle signals by successive categories and standard normal distribution graphs.

1. Each observer ranked his judgement regarding the presence or absence of subtle signals on a 5 point confidence-rating scale.
2. Number of materials in each category is changed into probability (T1).
3. Cumulated probability is calculated rightward (T2).
4. Ordinates ($y_1, y_2, y_1 > y_2$) of two boundary lines on the j -category are taken from the normal distribution function (T3).
5. y_2 is subtracted from y_1 (T4).
6. Mean values (Z_{ij}) can be expressed as follows: $Z_{ij} = (y_1 - y_2) / p(i: \text{observer}, j: \text{category}, p: \text{probability on } j\text{-category})$ (T5).
7. When the value of Z_{i1} is replaced by 0, relative distance ($D_{ij}; i: \text{observer}, j: \text{category}$) to Z_{ij} ($j=2\sim5$) can be obtained by a single scale (T6).

両側尺度として測定する事もできる。

この手法を、信号の種類別、画像システム別に集めたデータに適用すれば、それぞれにおける観察者別の D_{ij} , D_{ij} , $d_{j-1,j}$ を求めることができる。

3. 結果と考察

Fig. 2 は 6 画像システムより得られた全試料を信号の有無に関してのみ分類して算出した 5 人の観察者の D_{ij} (+), D_{ij} (-) を示したものである。(+) は信号を含む場合を示し, (-) は信号を含まない場合である。Di5 を平均

Fig. 2 Variations in the relative distances (D_{ij}) among the five observers for detection of reading images contained subtle signals. (upper: fragments of film, lower: acrylic beads)

値間で算出した判定軸（以下、判定軸）の長さとする、全観察者とも信号を含む場合の判定軸は信号を含まない場合の判定軸に比べて短くなっている。また、各カテゴリの平均値間の距離 ($d_{j-1,j}$) では、信号を含む場合に $d_{2,3}$, $d_{3,4}$ が $d_{1,2}$, $d_{4,5}$ に比べて短くなっているが、信号を含まない場合はほぼ同じ長さになるかやや短くなり、観察者間の変動も小さかった。これらの結果を 5 人の業務経験年数の面からみてみると、経験年数の浅い観察者の判定軸が短くなり、信号の有無による判定軸の長さの差も小さくなっている。したがって、今回の 5 人の観察者でみる限り、業務経験豊かな人が慎重な判断を行っているのに対して、経験が浅い人は強い確信での判定を好んでいる傾向がみられる。Fig. 1 とこれらの結果を比較してみる。まず、規準正規分布の密度関数を $f(x)$ とすれば

$$\int_{-\infty}^0 f(x) dx = \int_0^{\infty} f(x) dx = 0.5$$

であり、各カテゴリの確率はカテゴリの下限から上限までの $f(x)$ の積分値である。次に、同じ確率であっても分布の中心付近では裾野に比べてカテゴリの幅は小さくなる。したがって、左側から右側へカテゴリの累積確率をとっていることを考慮すると、高い確信の判定であるカテゴリ 1, 5 の判定比率が高くなる場合がカテゴリ 2, 4 のより低い確信度の判定比率が高くなる場合より判定軸は短くなる。また、両端のカテゴリの距離 $d_{1,2}$, $d_{4,5}$ は $d_{2,3}$, $d_{3,4}$ に比べて長くなっている事がわかる。実際、Fig. 1 より判定軸の短い観察者 C, D はカテゴリ 3 までの累積確率は判定軸の長い観察者 A, E よりも大きい値を示している。しかし、カテゴリ 4 の累積確率では逆転している。したがって、観察者 C, D は観察者 A, E に比べ

Fig. 3 Comparison of relative distances (D_{i5}) among the five observers for detection of two different subtle signals.

て誤った判断を多く含んでいるが信号は“確かにある”といった高い確信での判断を多く行っている。

Fig. 3 は信号を含む試料を 2 種類の信号別に分類したデータから算出した判定軸の長さを示したものである。各観察者でフィルム信号を含む場合の判定軸が短くなり、特に観察者 A, C ではアクリル信号を含む場合よりかなり短くなっている。これはフィルム片の判定がアクリルビーズより強い確信での判定が行われたことを示している。

Fig. 4 に信号を含まない場合の試料から算出した判定

軸を示した。(+)、(-) は 2 種類の信号を含む試料との組合せで判定された場合を示している。観察者 C 以外ではフィルム片の信号を含んだ試料との組合せの場合がアクリルビーズを含んだ試料との組合せより判定軸は長くなり、Fig. 3 と比較すると逆の傾向を示している。また、信号を含む場合の判定軸に比べて長くなり、信号の存在を認識しない場合に、観察者が高い確信度の判定が行いにくい事を示している。

Fig. 5 は全観察者の信号を含む試料の判定を、画像システム別に合計して算出した観察者全体の画像システム

Fig. 4 Comparison of relative distance (Di5) among the five observers by reading images excluding signals.

Fig. 5 Comparison of relative distances (Di5) among the six imaging systems. Di5 is calculated from cumulated data of the five observers.

別の判定軸を示したものである。このような集積データによる尺度の測定を試みたのは線量等の特定のパラメータによる尺度への影響をみるためである。その結果、HiS/RX と CR(1)ではほぼ同じ長さとなるが、HR-4/SHR-S では0.2から0.3, CR(1/2)では0.5前後長くなり、撮影システムの相違や CR の線量低下による判定軸への影響が認められた。

観察試料をカテゴリに分類することで、連続体に位置づける手法は視覚的評価に多く利用されている¹⁰⁾。ROC 解析や累積点法などで使われる場合は、カテゴリ順位にしたがった順序尺度による解析や、いわゆるブラックボックスとして^{3,7)}主観的尺度で取り扱われる場合が多かった。これらの理由は、両端のカテゴリを除くと評定尺度を距離尺度としても実用的問題となることが少なかった事や主観的尺度が結果の解釈を容易にしている事が考えられる。しかし、観察は観察者個々のさまざまな条件の下で行われるものであり、カテゴリ分類に対する判定尺度は単一的に処理すべきものではないと思われる。また、累積点法等で用いられる序数尺度等による主観的数字は、尺度の性質や使用に際しての熟練が必要とされており、安易な利用は誤差や偏りが生じやすく、誤った結果を招きやすいといわれている。したがって、これらの欠点を解消したり、画像評価で利用されている物理量との関数関係を調べるために正確な距離尺度の測定値が必要となったり、観察者の分類したデータを積極的に生かし、観察者の意思をより反映させた評定尺度を求めるためには、客観的測定値が重要な意味を持つものと思われる⁹⁻¹¹⁾。

カテゴリ分類に対する判定尺度の平均値を正規分布の仮定の下で行った。今回利用したような微小信号の検出に対する判定のデータ分布は、仮定の妥当性が経験的に認められているからである。しかし、実際の臨床写真による評価を行う場合などは、観察者の医学的知識や経験年数、観察能力などの相違で、両端のカテゴリにデータが集中する場合があったり、逆に、極端な判定をさける傾向が表れてデータが中央に集中したりして正規分布の仮定が困難になる場合がある。このような傾向を避けるために、上述したように“自分の見る限り絶対に信号は存在しない”，というような極端なカテゴリを両端に追加したり、観察者の十分なトレーニングを行うことは、他の視覚的評価と同様に、判定尺度の正確な測定にも大事なことである^{1,2,5,7,9)}。

系列カテゴリ法による今回の測定値は、観察者のカテゴリ分類に対する判定尺度の平均値であるが、観察者固有のデータを基に測定している客観性のある測定値である。しかし、正規分布を仮定している事は目的とする属性は1次元であること、また、各カテゴリと平均値間の

距離の線形性は保たれないことや試料・観察条件等により、データの正規性が仮定できない場合^{12,13)}は利用できない欠点がある。

4. 結 論

観察者のカテゴリ分類に対する判定尺度を、系列カテゴリ法を用いて平均値で測定した結果、カテゴリ相互間の距離は観察者で相違することが認められた。また、同じ観察者であっても信号の種類や信号の有無の認識によりカテゴリ間の距離が異なることが認められた。今後、このような客観性のある測定値を、放射線画像評価に用いられる物理量との関係を調べたり、視覚的な医用画像の評価に利用することで、観察者の意思をより反映させた評価法の確立が可能になると思われる。

参 考 文 献

- 1) Mets CE: ROC 解析の基礎, 日放技学誌, **46**(6), 111-117, (1990).
- 2) Mets CE: ROC methodology in Radiographic Image Invest Radiol. **21**, 720-733, (1986).
- 3) 飯沼 武, 遠藤真広, 館野之男, 他: 読影過程を含むX線診断系の定量的評価(ROC 曲線と情報理論による試み), 医用電子と生体工学, **14**, 200-206, (1976).
- 4) 石見 浩, 大賀泰文, 田畑洋二, 他: 視覚評価を用いた現像システムの違いによる画質の評価, コニカ X-ray 写真研究, **40**(5), 148-154, (1989).
- 5) 久米祐司, 前田 要, 山下一也, 他: 画像の主観的評価法, 日放技学誌, **49**(4), 616-627, (1993).
- 6) 山下一也, 片桐敏男, 若松孝司: 系列範疇法によるX線画像の画像評価, 日放技学誌, **32**(6), 560-565, (1977).
- 7) 土井邦雄: X線画像の信号検出と視覚特性の重要性, 日放技学誌, **43**(6), 694-729, (1987).
- 8) 上田尚一: 統計データの見方・使い方, 103-141, 朝倉書店, (1991).
- 9) 田中良久: 心理学的測定法, 東京大学出版会, (1988).
- 10) 松原 望: 意志決定の基礎, 147-223, 朝倉書店, (1991).
- 11) 脇本和昌, 垂水共之, 田中 豊: パソコン統計解析ハンドブック I, 共立出版, (1987).
- 12) 林知己夫: 数量化の方法, 東洋経済新報社, (1986).
- 13) 西里静彦: 質的データの数量化, 朝倉書店, (1989).