

# 研究レポート No.2 ～ROC 解析によるモデル精度評価～

2011年7月14日 株式会社アイズファクトリー <https://bodais.com/company/>

## 概要

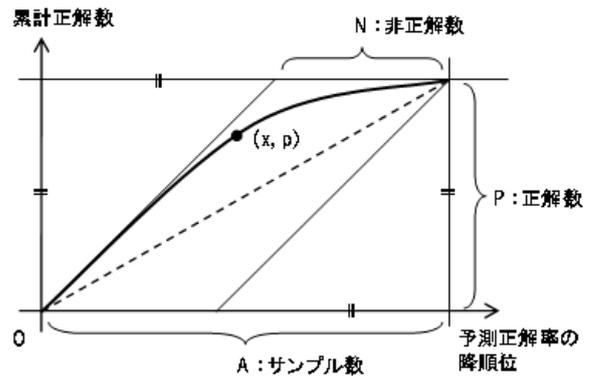
ダイレクトメールの効率化のために顧客データから予測入会率の高い顧客を抽出したり、医療診断において、X線画像から陽性部位を判別するなど、正解と不正解の混在するデータから効率よく正解を拾い出す技術が重要となってきている。その際、抽出技術の精度を評価する手法として重要なのが ROC 解析である。本レポートでは、ROC 解析による精度評価手法を説明し、その適用事例を紹介する。

### 1. はじめに

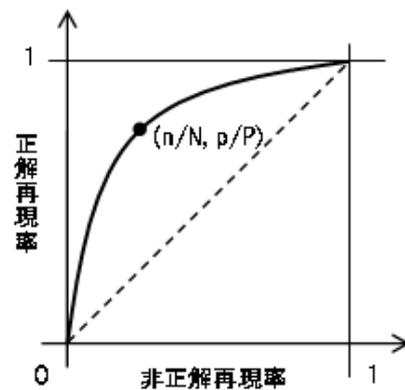
ダイレクトメールの効率化のために顧客データから予測入会率の高い顧客を抽出する場合、予め入会の有無が正解データとして紐づいている顧客リストをもとに、ロジスティック回帰などの手法で顧客ごとの予測入会率を算出し、その値の降順に顧客を抽出する。その場合、モデルの精度は、スコア上位に占める正解データの比率が高いほど良い。これを図式化したのがゲインチャートであるが、その精度評価指標となるゲインチャート面積比は、後述するように同じモデルでもデータ群により上限値、下限値が異なる。例えば、昨年の顧客リストによる予測モデルと今年の顧客リストによる予測モデルとを比較する場合 [1]、ゲインチャートをそのまま比較することはできない。異なるデータ間で精度比較を行うには、通常のゲインチャートを規格化した曲線を定義する必要がある。本レポートで取り上げる ROC が、正にそのような曲線である。ROC 解析では、ゲインチャートの縦横の両軸をデータ群の正解率に依存性のないように規格化して、異なるデータ間での精度比較を可能にする。さらに、ROC 解析は正解が定義されているデータならば、その背後にモデルが想定されてなくても使えるので、画像診断技術の精度評価などでも使用されている。

### 2. 正解予測モデルとゲインチャート

入会率予測のような正解予測モデルでは、各サンプルの予測正解率がスコアとして算出される。予測正解率（スコア）の降順にサンプルを並べたとき、スコア順位  $x$  までのサンプル集合をスコア上位群  $A(x)$ 、 $A(x)$  における正解数を  $p$ 、不正解数を  $n$  と定義しよう。 $A(x)$  のサンプル数は  $x$  であるから  $p+n=x$  が常に成り立つ。ゲインチャートとは、横軸にスコア順位  $x$ 、縦軸にスコア順位  $x$  までの累計正解数  $p$  を取って描かれる曲線のことであり、一般的に、ゲインチャートは、図 1 のような増加曲線となる。図に点線で示された対角線は、サンプルをランダム抽出して並べた場合に対応する。ゲインチャートが対角線より上に膨らむほどランダム抽出に比べて抽出効率がよく精度が高いことになる。従って、ゲインチャートの精度評価は、その曲線下面積と対角線下面積との比率で決まるが、この値の上限値と下限値はデータに含まれる正解データの比率に依存する。



【図 1】ゲインチャート



【図 2】ROC

具体的には、横軸の最大値をサンプル数  $A$ 、縦軸の最大値を正解数  $P$  としたとき、面積比の最小値は、正解率  $P/A$ 、最大値は  $2 - P/A$  となる。ゆえに正解率の異なる 2 つのデータ間でゲインチャート面積比による精度比較を行うと、間違った結論に達してしまう可能性がある。これを回避するためにゲインチャートを規格化する必要がある。

図 1 のように、ゲインチャートは底辺の長さが不正解数  $N (= A - P)$ 、高さが正解数  $P$  に等しい平行四辺形の内部に収まる。ゲインチャートの規格化は、平行四辺形領域を、一辺の長さが 1 の正方形領域に射影する線形写像として実現される（次式）。

$$\text{点}(x, p) \Rightarrow \text{点}((x - p)/N, p/P)$$

ここで、スコア上位群  $A(x)$  に対して、 $x-p$  はそれに含まれる非正解数  $n$  であるから、ゲインチャートの各点は、非正解再現率  $n/N$  と正解再現率  $p/P$  を座標値とする点に射影される (図2)。スコア順位  $x$  を変化させて得られるこれら射影点の集合が ROC である。ROC 下面積は 0 から 1 の連続値をとり上限値と下限値はデータによらない。

### 3. ROC 解析の検索理論による理解

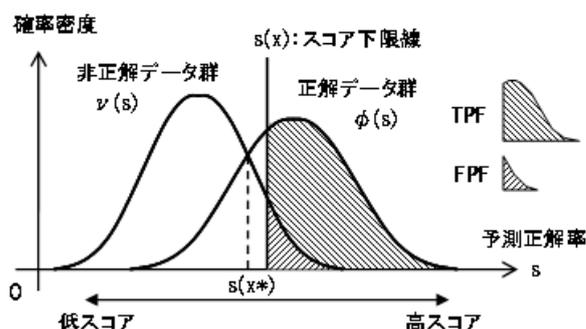
ROC とは、Receiver Operating Characteristic Curve の略であり受診者動作特性曲線と訳される。ROC を用いた識別精度評価法は ROC 解析と呼ばれている。もともとは、第 2 次大戦中に米国のレーダーの研究において、識別精度の評価法として利用された解析手法である。

データ全体は、正解データと不正解データとに分かれる。一方、正解予測モデルにより予測正解率がスコア (確信度) として算出されると、スコア上位群  $A(x)$  を以って予測正解群を定義できる。 $A(x)$  に含まれるサンプルは、予測により正解と判断されただけであり、実際には正解サンプルも非正解サンプルも含んでいる。医療診断にならって正解を陽性 (Positive)、非正解を陰性 (Negative) と呼ぶことにすれば、 $A(x)$  に含まれる正解群は、真陽性 (TP : True Positive) であり、 $A(x)$  中の非正解群は、偽陽性 (FP : False Positive) となる。予測正解群での正解データの再現率は真陽性率 (TPF : True Positive Fraction)、不正解データの再現率は偽陽性率 (FPF : False Positive Fraction) と解釈できる [2]。これとは逆に、非正解抽出群での正解データの再現率は偽陰性率 (FNF)、不正解データの再現率は真陰性率 (TNF) と解釈でき、定義上、 $FNF = 1 - TPF$ ,  $TNF = 1 - FPF$  が成り立っている [2]。ROC は、スコア降順にサンプルを並べたとき、当該順位における FPF を横軸、TPF を縦軸に取って描かれる曲線である。

スコア順位  $x$  を変化させたとき、スコア上位群  $A(x)$  が拡大するに従って ROC の座標点が動いていく様子は、正解データ、非正解データのスコア分布からも理解できる (図3)。一般に、予測正解率の高スコア側に正解データが分布し、低スコア側に非正解データが分布している。スコア上位群  $A(x)$  のスコア下限値を  $s(x)$  と定義すると (図3 の縦実線)、正解データの確率分布曲線の下側面積のうちスコア下限線の右側面積が正解再現率  $TPF(s(x))$  であり、ROC の縦軸座標となる。一方、非正解データの確率分布曲線の下側面積のうちスコア下限線の右側面積が非正解再現率  $FPF(s(x))$  であり、ROC の横軸座標となる。正解データの確率密度関数を  $\phi(s)$ 、非正解データの確率密度関数を  $\nu(s)$  とすると、スコア上位群の TPF、FPF はそれぞれ次式で定義できる。

$$TPF(s(x)) = \int_{s(x)}^{\infty} \phi(s) ds, \quad FPF(s(x)) = \int_{s(x)}^{\infty} \nu(s) ds$$

スコア順位  $x$  が下がるにつれて、図3のスコア下限線  $s(x)$



【図3】正解データ、非正解データのスコア分布

が右から左に動いて行く。図の場合、正解分布は非正解分布より高スコア側に位置するので、スコア下限線の位置に依らず  $TPF > FPF$  であり、ROC が対角線 ( $TPF = FPF$ ) より上に膨らんだ曲線になることが理解できる。

ここで、ROC の傾きを考えてみる。TPF、FPF の積分定義式を微分して次式が得られる。

$$dTPF(s)/ds = -\phi(s), \quad dFPF(s)/ds = -\nu(s)$$

正解データ、非正解データの確率密度関数は、ROC 上のスコア  $s$  に対応する点における接線ベクトルの各成分に対応している。従って、スコア  $s$  における ROC の傾きは、両密度関数の比として定義される (次式)。

$$dTPF(s)/dFPF(s) = \phi(s)/\nu(s)$$

スコア順位  $x_*$  において正解データ、非正解データの確率密度関数が互いに等しくなる場合 ( $\phi(s(x_*)) = \nu(s(x_*))$ )、ROC 上の対応する点における接線の傾きは 1 になる。スコア順位  $x$  が  $x_*$  より下位になるとスコア下限線  $s(x)$  が境界線  $s(x_*)$  より左側に位置し、ROC の傾きは定義により 1 より小さくなる。すなわちスコア順位  $x_*$  より下位では正解データに比して非正解データの増加率が大きくなるので、予測正解群を定義する切断点としてスコア順位  $x_*$  を採用することが考えられる。

### 4. まとめ

本レポートでは、統計モデルの精度評価の手法として ROC 解析を紹介した。本手法は、ダイレクトメールの入会率予測モデル等の正解予測モデルの精度評価に利用できる。その他にも、SVM や決定木など判別モデルの精度評価や複数モデル間での精度比較においても使われている。また、医療現場における X 線画像診断などの画像診断技術における精度評価法としても使われている [3]。

### 5. 参考文献

- [1] 研究レポート (2011) No.1
- [2] 桂川 (2004) : 日本放射線技術学会雑誌, 第 60 巻 3 号, 309-316
- [3] 池田 (2005) : 画像診断のための ROC 解析概論, 現代図書。