解説

正確な学習よりも得する学習 ―誤分類コストを考慮する分類学習―

(1) 評価編

鈴木 英之進
横浜国立大学大学院工学研究院
suzuki@ynu.ac.jp

正確でも損したらダメ

クレジットカードの不正利用に悩まされる信販会社が，怪しい利用に対して警告を発するシステムを見直すことにした．この種のシステムでは，過去の利用履歴，利用者の情報，および購入の条件などを総合して警告を発すかを決める．詐欺師は警告を発す基準に注意を払っており，その裏をかこうとしているのが常である．利口な敵に素早く対抗するために，この会社は人工知能技術を導入することにした．開発担当者は，「機械学習」や「データマイニング」が関係あると判断し，最も頻繁に使われている「決定木☆1」を使うことを決断した．不正利用に関連しそうな属性群を定義し，最近の利用をこれらの属性群で表し，不正利用か否かをクラスとし，学習データを用意した．データに学習手法を適用して得られた決定木を見ると，「すべての利用に警告を発するな」とだけ書いてあった．

これはもちろん笑い話であるが，知らない技術を用いる際には，その原理を十分理解すべきであるという教訓も示している．学習するソフトウェアの構築を目的とする機械学習は，種々の応用における実用的な利点が魅力であり，データマイニングの基盤ともなっている．機械学習技術を実応用において有効裏に用いるためには，データと学習法の両方の特性を考慮すべきであり，さもなければ落し穴にはまることになる．

冒頭の失敗は，全体のごくわずかの割合しか占めない不正利用が，この警告システムにおいて重要な意味を持つことからきている．後半部分をより正確にいうと，不正利用を通常利用と誤判断してしまうことは，その逆よりも損，すなわちコスト (cost) が大きい☆2．多くの決定木学習手法は正確な学習を目指しており，コストや利益 (benefit) を無視している．このため全体のごくわずかな割合しか占めない不正利用は，コスト面で重要なのに軽視され，大多数を占める通常利用を正確に予測する，非現実的な決定木が学習されてしまったのである．

本稿では，正確な学習よりも得する学習がよいという立場に立ち，この問題の重要な一面について考える．予測を間違えた際のコストは誤分類コスト (misclassification cost) と呼ばれる．誤分類コストを考慮する分類学習は，近年盛んに研究されており，重要かつ興味深い研究成果が出ている．誤分類コストを考慮する分類学習は，学習結果の評価法に関するものと，誤分類コストが低い学習法に関するものに分けられる．この解説においては，前者を評価編，後者を解決編で，2 回に分けて紹介する．なお，コストに関する研究は効用理論や人工知能において綿々と続けられてきたが，本連載で紹介する研究はそれらとは独立に機械学習分野で行われていると，筆者は認識している．

☆1 決定木は，次回の解決編で説明する．
☆2 本連載ではコストを損する度合いの尺度と抽象的に定義する．コストは金銭に関するものだけではなく，時間や重傷度など種々存在する．

誤分類コストを考慮する分類学習

たとえば，受診者 1, 2, …, n に関する健康診断結果 $d_1, d_2, \ldots, d_n$ から，受診者 i が精密検査を受けるべきか予測する分類モデル（関数）$f(\cdot)$ を学習する問題を考えてみよう．分類モデル $f(\cdot)$ に，i の健康診断結果 d_i を入力すると，$f(d_i)$=yes あるいは $f(d_i)$=no の出力を得る．分類モデルとしては，たとえば，d_i が検査Xに関する数値である場合，$f(d_i)$=yes ($d_i>5$ の場合)，no ($d_i \leq 5$ の場合）などが考えられる．この種の問題は，分類学習問題と呼ばれ，機械学習だけではなく，パターン認識や統計学でも研究されている．

分類学習においては，上記の場合の患者を例と呼ぶ．通常，例は複数個の属性に関する属性値群と，1 個のクラスによって表される．ただしクラスは，上記の場合における精密検査を受けるか受けないかなど，分類モデルが予測する対象であり，クラス間に順序はない．分類モデルは，訓練データから学習され，訓練データとは別のテストデータで評価される．訓練データとテストデータを構成する例をそれぞれ，訓練例とテスト例と呼ぶ．

通常，分類モデルの良さは，テスト例についての予測が正確である度合いで評価される．最も頻繁に用いられる評価指標は正答率と呼ばれ，m 例のうち m_f 例について予測が当たった場合，割合 m_f/m で与えられる．現在研究されているほとんどの分類学習は，正答率の向上が主要目的となっており，正確な学習を目指していると言えよう．

クラスが 2 種類の場合，予測対象として重要なクラスを正クラス，もう片方を負クラスと呼ぶ．予測の結果は，予測対象となる例が属すクラスと予測されたクラスに関して，**表-1** に示す 4 通りに分類される．

	実際は負クラス	実際は正クラス
負クラスと予測	true negative (TN)	false negative (FN)
正クラスと予測	false positive (FP)	true positive (TP)

表-1　予測結果の種類

正答率では，精密検査を受けるべき重病人に検査を受けさせずに見逃す間違いも，精密検査を受ける必要がない健康人に検査を受けさせて空振りする間違いも等しく見なす．もっとも現実においては，検査コストがかかっても重大な病気を早期に発見する方が重要であると考えられるため，空振りよりも見逃しの方が重罪である．逆に，不動産売買履歴や銀行高額送金履歴から犯罪者を特定する分類学習問題では，重要顧客を失う損失が大きいため，犯罪者を合法と予測し見逃すことよりも合法顧客を違法と予測し空振りする方が罪が重い場合がある．さらに不正検知などにおいては，クラス比が 99：1 など極端に偏っている分類学習問題も存在し，このような場合にはすべての例を多数派クラスと予測し正答率が 99% になることよりも，総合的な利益やコストを考慮することが求められる．実問題においては，正確な学習よりも得する学習を目指すべきである．

そもそも，分類学習では初期から，クラスが j である例をクラス i に属すと予測した際に被るコストを損失関数 $C(i|j)$ で表し，損失関数を考慮して学習結果を評価すると定義している．正答率では $C(i|j)=0$ ($i=j$ の場合)，$C(i|j)=1$ ($i \neq j$ の場合）と仮定しており，この場合の $C(i|j)$ は 0/1 損失関数と呼ばれる．近年，データマイニングの発展などに伴って機械学習の実応用が大幅に増え，0/1 損失関数に偏っていた研究が，やっと一般的な損失関数を対象とするようになってきたとの感がある．

誤分類コストを考慮する分類学習には，$C(i|j)$ が学習時に分かっていると仮定するアプローチと，分かっていないと仮定するアプローチがある[3)]．以下，両者を分けて解説する．

学習時にコストが既知の場合

■コスト行列と最適予測

以下，断らない限りクラスが 2 種類存在する 2 クラス分類学習問題を対象とする．クラス数が n の場合，予測結果の種類数は n^2 となるため，3 クラス以上の場合は 2 クラスの場合に比較して研究が遅れているとの感がある．例 x がクラス j に属す確率を $P(j|x)$ で表すと，x がクラス i に属すと予測する条件つきリスク $R(i|x)$ は，

$$R(i|x)=\sum_j P(j|x)C(i|j). \tag{1}$$

例 x のクラスに関する Bayes 最適予測は，$R(i|x)$ が最小となる i すなわち $\mathrm{argmin}_i\, R(i|x)$ である．誤分類コストを考慮すると，$P(j|x)$ が小さくてもクラスを j と予測する場合がある．たとえば信販会社は，クレジットカードでの支払いがほぼ合法的と見なせる場合でも，高額なために拒否する場合がある．

以下，$C(i|j)$ を要素とするコスト行列を**表-2** に示すように表し，コスト行列と最適予測に関して考察する[2)]．ただし負クラスを−，正クラスを＋で表す．表より，false positive のコストは $C(+|-)$，false negative のコストは $C(-|+)$ である．

コスト行列は，予測が当たった場合のコストは外れた場合のコストより低いという道理性条件 (reasonableness conditions) を満たす必要がある．すなわち $C(+|-) > C(-|-)$ かつ $C(-|+) > C(+|+)$．どちらか片方が成立し

	実際は負クラス	実際は正クラス
負クラスと予測	$C(-\|-)$	$C(-\|+)$
正クラスと予測	$C(+\|-)$	$C(+\|+)$

表-2　コスト行列．各要素は 1 列当たりのコストを表す

	実際は悪い	実際は良い
悪いと予測	0	1
良いと予測	5	0

表-3　クレジット決済に関する間違ったコスト行列

	不正	合法
拒否	\$20	−\$20
承諾	$-\gamma$	0.02γ

表-4　利益行列の例

	実際は負クラス	実際は正クラス
負クラスと予測	$M(-\|-)$	$M(-\|+)$
正クラスと予測	$M(+\|-)$	$M(+\|+)$

表-5　混同行列．各要素はテストデータにおける例数を表す

ないなら，すべての例を正あるいは負と予測すればよくなってしまう☆3．コスト行列の各要素に同じ定数を加えても乗じても，最適予測は無関係である．よって，コスト行列は$C(-|-)=0$, $C(-|+)=C(-|+)'$, $C(+|-)=1$, $C(+|+)=C(+|+)'$ と簡略化でき，その自由度は 2 であることが分かる．

十分注意しないでコスト行列を設定すると，**表-3** に示すような間違いを犯してしまう．このコスト行列は，信販会社がクレジットカードでの決済を良い (＝ローンを完済する) か悪い (＝ローンが焦げ付く) と予測する問題に関するものである．悪いと予測することは決済を拒否することを意味するから，本当は「悪いと予測」の行におけるコストは同じ値になるべきである．

間違いの原因は，コストと利益を混同しているためと予想できる．利益はコストに比較して損益分岐点 (baseline) が存在する上，通常は例に依存する．利益行列を設定する場合には，コスト行列の場合よりもさらに注意が必要である．上記の問題に関し，クレジットカード決済についての利益行列の例を**表-4** に示す．ただしγはクレジットカードでの決済額である．不正な決済を拒否すると将来の不正を防ぐため利益は正，合法な決済を拒否すると信頼を失うため利益は負であることが分かる．

例xについて正 (+) クラスの予測が最適とは，誤分類コストの期待値が負 (−) クラスと予測する場合以下であることと定義できるため，

$$P(j=-|x)C(+|-)+P(j=+|x)C(+|+)$$
$$\leq P(j=-|x)C(-|-)+P(j=+|x)C(-|+).$$

道理性条件が成立すると仮定し，正クラスが最適予測となる $p^*=P(j=+|x)$ を求めると，

$$p^*=\frac{C(+|-)-C(-|-)}{C(+|-)-C(-|-)+C(-|+)-C(+|+)}. \qquad (2)$$

結局，コスト行列は最適予測に関しては 1 自由度であることが分かる．誤分類コストを考慮する分類学習の観点からは，$P(j=+|x)\geq p^*$ の場合に限りxが正例 (クラス 1) と予測する分類モデルが得られればよい．

より正確には，分類モデルの良さは，期待誤分類コストの小ささで表される．分類モデル $f(\cdot)$ の期待誤分類コストは，すべての例xに対し，$\sum_x P(x)R(f(x)|x)$で与えられる．ただし期待誤分類コストは，正確に知ることはできないので，通常はテストデータをTとすると，次で代用する．

$$\sum_{x\in T}\hat{P}(x)\hat{R}(f(x)|x)=\sum_{x\in T}\hat{P}(x)\sum_{j}\hat{P}(j|x)C(i|j)$$

ただし，$\hat{P}(x)$, $\hat{R}(f(x)|x)$, $\hat{P}(j|x)$ は，それぞれ$P(x)$, $R(f(x)|x)$, $P(j|x)$の推定値を表す．この目標を達成しようとするために提案された種々の学習法は，次回紹介する．

■誤分類コストの信頼性評価法

テストデータにおける予測結果に関する例数$M(i|j)$は，**表-5** に示す混同行列 (confusion matrix) を構成する．正答率は，この表記では$\{M(-|-)+M(+|+)\}/\{M(-|-)+M(+|+)+M(+|-)+M(-|+)\}$と表される．期待誤分類コストは，誤分類コストを考慮する分類学習において，正答率よりも適した評価指標である．情報検索では，再現率$recall=M(+|+)/\{M(+|+)+M(-|+)\}$と適合率$precision=M(+|+)/\{M(+|+)+M(+|-)\}$の両方，あるいはF値$(1+\beta^2)recall\cdot precision/(\beta^2 recall+precision)$を評価指標として用い，誤分類コストを考慮する分類学習においてもこれに習う場合がある☆4．

ある分類モデルに関し，たとえば期待誤分類コストが

☆3　この条件はクラスが 3 個以上ある場合にも当てはまり，道理性条件は，各行は他を支配 (dominate) してはいけないと表せる．
☆4　通常$\beta=1$とする．

150となった場合，本当にこの値でこの分類モデルを評価してもよいのだろうか？　テストデータが，例数が同じでも例が異なれば，期待誤分類コストが145となる場合もありそうである．さらに別の分類モデルについて，期待誤分類コストが160となった場合，最初の分類モデルは2番目の分類モデルよりも，コストが低いと断言してよいのだろうか？　本節では，これらの問題に対し，統計的検定に基づく手法[3]を紹介する．これらの手法BCOSTとBCOSTDELTAは，期待誤分類コストが正規分布に従うと仮定する手法より，正確であることが実験で確認されている[☆5]．

以下，本節では，クラス数がsの場合を考える．BCOSTは，混同行列の各要素が起こる確率$P(i|j)$を推定し，期待誤分類コストの信頼区間を求める．まず，$P(i|j)$が多項分布に従うと仮定し，Laplace修正を用いて推定する．

$$\hat{P}(i|j)=\frac{M(i|j)+\lambda}{n+s^2\lambda} \tag{3}$$

ただし，λ, nはそれぞれLaplace修正のパラメータとテストデータの例数を表す[☆6]．次に推定された$\hat{P}(i|j)$を用いて，ブートストラップ法という一種のシミュレーションを行い，期待誤分類コストの信頼区間を求める．シミュレーション回数が1000の場合の，BCOSTの疑似コードを示す．

手続き：*BCOST*

入力：信頼水準ρ，テスト例数n，コスト行列C，
　　混同行列M

返り値：期待誤分類コストの信頼区間 $[c_{\mathrm{lb}}, c_{\mathrm{ub}}]$

```
式(3)を用いて混同行列の各要素のP(i|j)を推定
For(シミュレーション番号u) from 1 to 1000
Do
  混同行列M_uの各要素を0に初期化
  For (生成する例番号v) from 1 to n Do
    P̂(i|j)に基づき，例を生成
    生成した例に基づき，混同行列M_uを更新
  c_u=(M_uの期待誤分類コスト)
μ=⌊(1−ρ)/2＊1000⌋+1
c_lb=(μ番目に大きいc_u)
c_ub=(1001−μ番目に大きいc_u)
```

BCOSTDELTAは，2つの分類モデルの期待誤分類コストに差がないという帰無仮説を検定し，特定のデータ集合に関する期待誤分類コストの差を判定する．基本的な方法はBCOSTと同じであるが，実際のクラスがjである例について，分類モデル1がクラスi_1と予測し，分類モデル2がクラスi_2と予測する例数$M'(i_1, i_2|j)$についての3次元行列M'を用いる．$M'(i_1, i_2|j)$に対する期待誤分類コストの差$\Delta(i_1, i_2|j)$は，$\Delta(i_1, i_2|j)=C(i_1|j)-C(i_2|j)$．BCOSTDELTAは，BCOSTと同様の方法により2つの分類モデルについて，期待誤分類コストの差$\sum_{i_1}\sum_{i_2}\sum_j M'(i_1, i_2|j)\,\Delta(i_1, i_2|j)$に関する信頼区間を求め，これが0を含む場合に限り帰無仮説を棄却する．BCOSTとBCOSTDELTAは，誤分類コストを考慮する分類学習において，実用に耐え得る最初の検定法であると位置付けられている．

学習時にコストが不明の場合

■ROC曲線分析

前章では分類モデルが期待誤分類コストで評価できるとしてきたが，このためには誤分類コストの値（コスト行列の要素）$C(i|j)$とクラスの割合$P(k)$[☆7]が正確に分かっている必要がある．もっとも，具体的な不正検知の例を考えてみれば分かるように，これらは通常，正確には分からない上，時間とともに変化する．このような場合には，任意の$C(i|j)$と$P(k)$におけるコストが一目で分かる評価法を用いることが考えられる．ROC曲線(Receiver Operating Characteristics Curve)分析は，この考えに基づく，分類モデルの集合についての評価法であり，信号検出や医療の分野などで長い間用いられてきた．

ROCグラフは，x, y軸にそれぞれfalse positive率$FP\equiv P(+|-)/P(-)$，true positive率$TP\equiv P(+|+)/P(+)$[☆8]をプロットし，つなげることによって得られる[5]．1個の分類モデルは，ROCグラフ上において1個の点として表される．ある分類学習手法において，パラメータを変えたり，訓練データを複数通りに前処理すると，複数個の分類モデルが得られる．ROC曲線は，このような複数個の分類モデルに相当する点の集合をROCグラフ上に表してつなげた曲線である．**図-1**に，4個の分類

☆5　この結果により，BCOSTとBCOSTDELTAは，Chernoff Boundなどにより期待誤分類コストの上限値を返す手法よりも正確だと予想されるが，実験によるこれらの定量的な比較は興味深いと思われる．
☆6　直観的にいうと，λは観測された値を信じない度合を表す．
☆7　以降，各クラスの割合を総してクラスの分布と呼ぶ．
☆8　実際にはテストデータにおけるfalse positiveの数をテストデータの例数で割った値と，true positiveの数をテストデータの例数で割った値で代用することになるが，なめらかな曲線を得るために交差検定や曲線の当てはめを用いることも多い．

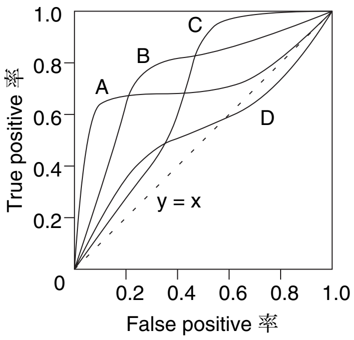


図-1 4個の学習法に関するROCグラフの例

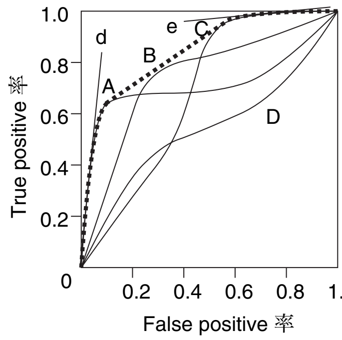


図-2 ROC凸包（太い点線）の例

学習手法に関するROCグラフの例を示す．各ROC曲線において，負クラスの間違いを減らそうとfalse positive率を下げると，つられて正クラスの正答率であるtrue positive率も下がることが分かる．

ROCグラフにおいて，隅の点は特別な場合を表す．左下の点 (0, 0) は常に負クラスを予測する場合，すなわち警告を出さない場合に相当し，右上の点 (1, 1) は逆に常に正クラスを予測する場合に相当する．左上の点 (0, 1) は常に予測が当たる場合，右下の点 (1, 0) は常に予測が外れる場合に相当する☆9．直線 y=x は，ランダムにクラスを予測する場合の期待値に相当する．直感的には，ROCグラフにおいて左上の点は，FPが小さくTP が大きいため右下の点より良い．もし分類学習手法Aが生成した分類モデルの集合に関するROC曲線が，分類学習手法Bが生成した分類モデルの集合に関するROC曲線よりも左上にあり，悪くとも等しい場合，前者は後者を支配する (dominate) という．この場合，任意の $C(i|j)$, $P(k)$ において，AはBよりも優れているか，悪くても同性能である．分類学習手法の良さを，ROC曲線$g(x)$下の面積AUC (Area Under the ROC Curve) $\int_0^1 g(x)dx$で評価する場合もある．

■ROC凸包に基づく評価法

ROC凸包法 (ROC convex hull method) は，多数のROC曲線を可視化・比較・管理するために考案された[4)]．以下，予測が当たった場合の誤分類コストを0と仮定する$C(-|-)=C(+|+)=0$．ROCグラフ上の点C：(FP, TP) で表される分類モデルの期待誤分類コストは，

$$E[C] \equiv P(+)\,(1-TP)\,C(-|+)+P(-)FP\cdot C(+|-). \quad (4)$$

よって，2 点(FP_1, TP_1)，(FP_2, TP_2) が同性能である場合，

$$\frac{TP_2-TP_1}{FP_2-FP_1}=\frac{P(-)C(+|-)}{P(+)C(-|+)} \quad (5)$$

期待誤分類コストが等しい点を結んだ直線を，等性能直線 (iso-performance line) と呼ぶ．式(5)は等性能直線の傾きを表し，等性能直線は左上にあるほど良い☆10．

図-2のように複数本のROC曲線が与えられたとき，最適である可能性がある分類モデルは，それらの左上方向☆11の凸包上の点で表されるものに限られる．これらの点集合をROC凸包と呼び，ROC凸包から得られる情報について考える．たとえば図において，分類学習手法BとDは決して最適にならないことが分かる．さらに考慮すべき分類モデルは，分類学習手法AとCによって学習された分類モデルのうち，ROC凸包上にあるものだけであることも分かる．ROC凸包は，誤分類コストの値とクラスの分布が分かった場合，最適な分類モデルを選択することにも使用できる．たとえば $P(-)=10P(+)$ のとき，$C(-|+)=C(+|-)$ あるいは$C(-|+)=100C(+|-)$であれば，等性能直線の傾きはそれぞれ10, 0.1 である．図のd,eに示すように，該当する直線とROC凸包の接点は，最適な分類モデルに該当する．

このようにして分類学習手法の性能を比較する評価法を，ROC凸包法と呼ぶ．n本の曲線の凸包は $O(n \log n)$ 時間で求められるため，ROC凸包は高速に構成できる．さらに新しい分類学習手法が加わった場合，凸包を拡張するかを調べればよいため，この方法は多数の分類学習手法を管理することに適している．誤分類コストの値やクラスの分布が変化しても，等性能直線を求め直せばよ

☆9 ICML-2003 国際会議で P. Flach は，左上の点と右下の点をそれぞれ ROC の天国，ROC の地獄と名付けた．地獄は予測を反転すれば天国に相当するので，それほど悪くないとのことである．

☆10 なお，(5) より，この評価法において，クラスの分布が偏っている問題と，誤分類コストがクラスに依存する問題は，本質的に同じであることが分かる．たとえば，P(+) を倍にすることは，$C(-|+)$ を倍にすることや $C(+|-)$ を半分にすることと同じ効果がある．

☆11 $y=x$ よりも左上にあること．

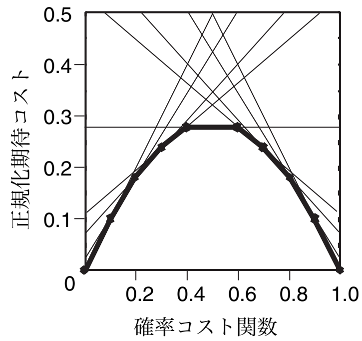


図-3　分類モデルの集合に対応する包絡線

いので，容易に対処可能である．誤分類コストの値やクラスの分布に幅がある不完全情報の場合においても，対応する等性能直線の傾きの範囲からROC凸包上の部分曲線を求め，最適な分類学習手法の候補を絞り込むことができる．

■コスト曲線

ROCグラフは広く用いられているが，分類モデルの性能を比較するには意外に不便な場合がある．たとえば，誤分類コストの値やクラスの分布が分かった場合に，分類学習手法Aと分類学習手法Bの期待誤分類コストを，それぞれが生成した分類モデルの集合に関するROC曲線を用いて比較する場合を考える．まず，等性能直線の傾きを求め，2本のROC曲線との接点(FP_1, TP_1)，(FP_2, TP_2) を求める手間が必要である．次に，これらで表される分類モデル2個の期待誤分類コストの差は，$(TP_2-TP_1)P(+)C(-|+)+(FP_1-FP_2)P(-)C(+|-)$ で与えられる☆12．これは重みつきマンハッタン距離に相当し，ROCグラフから読み取るのは難しい．分類学習手法の集合Cが分類学習手法の集合Dよりも期待誤分類コストが小さくなる条件を求める問題も，同様に手間がかかる．この問題を解くためには，CとDそれぞれのROC凸包に関する接線の傾きを比較する必要がある．ROC凸包の交点がこの問題にほぼ無関係であることは，ROCグラフの大きな欠点である．

対策として，クラス分布・誤分類コストと期待誤分類コストの関係がグラフから容易に分かる表現法が提案されている[1]．この表現法では，コスト平面上 (cost space) にコスト曲線 (cost curve) を描く．以下，クラスiに対し，次のように定義する．ただし，$\bar{i}$はクラスiとは異なるもう片方のクラスを表す．

$$w_i \equiv P(i)C(\bar{i}|i) \tag{6}$$

$$PCF(i) \equiv \frac{P(i)C(\bar{i}|i)}{P(+)C(-|+)+P(-)C(+|-)} \tag{7}$$

コスト曲線のx軸は確率コスト関数$PCF(+)$である．

$$PCF(+) = \frac{w_+}{w_+ + w_-} \tag{8}$$

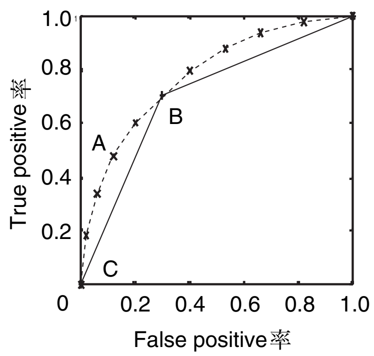
y軸は，期待誤分類コストをすべての例を誤予測した場合のコストで割った値，すなわち正規化期待コスト$NE[C]$を表す．

$$\begin{aligned} NE[C] &\equiv \frac{E[C]}{P(+)C(-|+)+P(-)C(+|-)} \\ &= (1-TP-FP)PCF(+)+FP \end{aligned} \tag{9}$$

ただし式の変形においては，$PCF(+)+PCF(-)=1$ を用いた．(9)より，ROC平面上の点 (FP, TP) は，コスト平面上において直線として表されることが分かる．逆にコスト平面上の点 ($PCF(+)$, $NE[C]$) は，ROC平面上において直線として表されることが示せる．

分類モデルの集合に対するコスト曲線を描くためには，**図-3**に示すように対応する各直線を描き，下方の包絡線を求めればよい．**図-4**にROCグラフの例と対応するコストグラフを示す．2つのグラフにおいて，実線と点線はそれぞれ対応するが，上記のようにROC曲線上の点はコスト曲線上の線分に対応する．すなわち，分類モデルは図左のROC平面では点だが，図右のコスト平面では線分として表されている．たとえば，実線で表されるROC曲線上の点Bは，台形のコスト曲線における一番上の辺に相当する．この辺の端点は，Bが有効となる範囲 ($0.3 \leq PCF(+) \leq 0.7$) を表す．y軸が正規化期待コストを表すため，2種類の分類モデル集合間の性能を比較することは容易である．たとえば図の左において，$PCF(+)=0.3, 0.7$ のときの差は，約20%であること

☆12 前節と同様に，予測が当たった場合の誤分類コストを0と仮定した $C(-|-)=C(+|+)=0$．



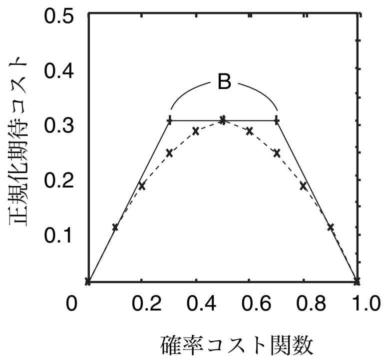


図-4　ROCグラフ(左)と対応するコストグラフ(右)

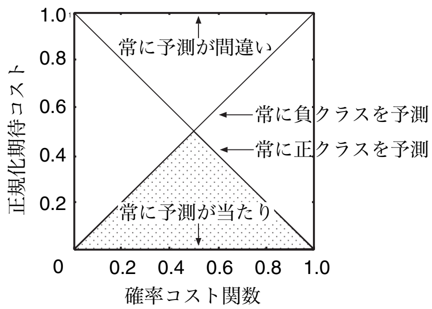


図-5　極端な分類モデル

が分かる．さらに分類学習手法の集合Eが分類学習手法の集合Fよりも正規化期待コストが小さくなる条件を求める問題も，EとFに対応するコスト曲線の交点が解となるため，分かりやすい．

図-5に，極端な分類モデルをコスト平面に4個示す．上辺は最も性能が悪い分類モデルであり，いつも予測を間違えるため正規化期待コストが常に1である．下辺は逆に最良の分類モデルであり，いつも予測が当たるために正規化期待コストが常に0である．常に正クラスを予測する分類モデルは，$PCF(+)=0, 1$のとき，正規化期待コストがそれぞれ1，0である．よって，実際の評価では，下の三角形で示される網かけ部分だけを考慮すればよいことが分かる．

1次元コストの評価の次は?

原子力プラントをfalse alarmで止めると経済的損失は多大，miss alarmで止めないと社会的影響と損失が多大である．このように業務運用や意思決定問題では，コストは多次元尺度であり，説明可能性と実用性が重要である．この辺を体系化しようとすると意思決定支援プロセスや業務プロセスとの統合まで踏み込む必要があり，対象問題に依存する地道な作業が多くなる．

もっとも，コストを誤分類コストに単純化した見通しがよい問題設定で，汎用的な議論をすることも必要である．今回は，機械学習手法を現実問題に適用する際には誤分類コストを考慮すべきことを述べ，その評価法を誤分類コストとクラス分布が既知の場合と未知の場合に分けて紹介した．誤分類コストを考慮する学習はまだ発展途上にあり，優れた提案が無視されている場合も多い．筆者が思うに，前者の場合においては信頼性評価，後者の場合においてはコスト曲線がこの状況に該当し，これらはより多くの関心を集めるべきである．

研究に限定しない一般論として，「問題が定義されれば半分は解決された」との格言がある．この意味において，今回の評価法は重要であるが，手持ちのデータから見えないデータのクラスを予測する分類学習は，一筋縄ではいかない．次回の解決編では，具体的な学習法を紹介する．

謝辞　本研究の一部は，文部科学省科学研究費特定領域研究「アクティブマイニング」の援助を受けている．大阪大学の鷲尾隆先生とIBMの鹿島久嗣氏たちから有益なコメントを得た．記して感謝する．

参考文献

1) Drummond, C. and Holte, R. C.: Explicitly Representing Expected Cost: An Alternative to ROC Representation, Proc. Sixth Intl. Conf. on Knowledge Discovery and Data Mining (KDD), pp.198-207 (2000).
2) Elkan, C.: The Foundation of Cost-sensitive Learning, Proc. Seventeenth Intl. Joint Conf. on Artificial Intelligence (IJCAI), pp.973-978 (2001).
3) Margineantu, D. D. and Dietterich, T. G.: Bootstrap Methods for the Cost-sensitive Evaluation of Classifiers, Proc. Seventeenth Intl. Conf. on Machine Learning (ICML), pp.583-590 (2000).
4) Provost, F. and Fawcett, T.: Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions, Proc. Third Intl. Conf. on Knowledge Discovery and Data Mining (KDD), pp.43-48 (1997).
5) Provost, F. Fawcett, T. and Kohavi, R.: The Case Against Accuracy Estimation for Comparing Induction Algorithms, Proc. Fifteenth Intl. Conf. on Machine Learning (ICML), pp.445-453 (1998).

(平成15年10月11日受付)