

麻酔とテクノロジー

徹底分析
シリーズ

臨床応用の 恩恵と落とし穴

機械学習モデル開発時の クラス不均衡への対応

尾形 宗士郎 15a 新M

近年、人工知能 (AI) 技術の進展により機械学習や深層学習を用いた予測モデルの医療応用が急速に進んでいる。特に、患者の予後や重篤な合併症の発生を予測するモデルは、麻酔科領域を含む多くの臨床分野で注目されている。しかし、こうしたAIモデルの開発には、医療データ特有の構造、特に「クラス不均衡」への理解と対応が欠かせない。

医療分野におけるAI・機械学習モデルの構築に際し、避けては通れない課題の一つが「クラス不均衡」である^{1,2)}。これは、予測対象となる臨床イベント、例えば術後合併症の発生、急性の低酸素状態、入院中の死亡などが、実際にはごく少数の症例にしか発生しないことに起因する。イベントの発生率は概して数~10%程度であり、学習に用いるデータの大部分は「イベントが発生しなかった」症例、すなわち負例 (negative) によって占められる。その結果、機械学習アルゴリズムは多数派である負例の特徴に過度に引き寄せられ、予測出力に大きな偏りを生じることとなる。

このようなクラス不均衡を放置したままモデルを構築すると、少なくとも三つの重大な弊害をまねく可能性がある。第一に、「偽の高性能」である。モデルがすべての症例を「イベントなし」と予測するだけでも、計算上の予測精度 (accuracy) は高くなる。例えば、イベントの発生率が5%であれば、すべての症例を「起きない」と予測するだけで95%の精度が得られることになる。しかし、このようなモデルは感度 (sensitivity: 実際にイベントが発生した症例のうち、それを正しく「発生あり」と予測できた割合) に乏しく、実臨床における有用性は極めて限定的である。

第二に、重要なイベントの見逃しが生じる。特に、発生頻度は低い为重篤な合併症の早期検出を目的とする予測モデルにおいて、感度の低下は患者安全に直結するリスクとなる。たとえ受信者動作特性曲線下面積 area under the receiver operating characteristic curve (AUROC) が高くとも、実際にイベントが発生した症例をとらえられなければ、臨床的な有効性は著しく損なわれる。

第三に、こうした非現実的なモデル出力は、共同研究者間の信頼関係を損なう。モデルの予測と臨床現場での判断がかけ離れている場合、「やはりAIは実地には使えない」「機械学習は現場の実態を理解していない」といった否定的な見解が共有されやすくなる。クラス不均衡に対する認識不足は、AI活用そのものに対する評価の低下へとつながりかねない。

このようにクラス不均衡は単なる統計的偏りではなく、モデル性能の評価、患者安全、そして臨床応用における信頼性を根本から揺るがす要因である。したがって、予測モデルの構築にあたっては、初期段階からクラス不均衡の存在を正しく認識し、戦略的かつ計画的に補正手法を導入することが不可欠である。

クラス不均衡への対策は大きく分けて、「モデルを工夫する」「学習データを工夫する」「アンサンブルを活用する」の三つの戦略があり、これらにもとづく多様な手法が提案されている。その中でも比較の実装が容易であり、効果・安定性・計算コストのバランスに優れた代表的な3手法を紹介し、それらの比較および選定の基準について述べる。その3手法とは、「クラス重み付けと閾値最適化」「標準モデル + synthetic minority over-sampling technique (SMOTE) 系のオーバーサンプリング」「アンダーサンプリングとアンサンブルの組み合わせ手法」である (表1)。

クラス重み付けと閾値最適化

表1 クラス不均衡への対策方法

観点 / 手法	(a) クラス重み付け+閾値	(b) SMOTE 系	(c) EasyEnsemble
手法の要点	損失+閾値	データ合成	アンダーサンプリング+アンサンブル
実装難易度	低	中	中~高
計算コスト	低	中	高
特徴空間要件	特になし	数値特徴が必要	多様な特徴に対応可

患者安全、そして臨床応用における信頼性を根本から揺るがす要因である。したがって、予測モデルの構築にあたっては、初期段階からクラス不均衡の存在を正しく認識し、戦略的かつ計画的に補正手法を導入することが不可欠である。

クラス不均衡への対処: まず試すべき3手法と その選び方

クラス不均衡への対策は大きく分けて、「モデルを工夫する」「学習データを工夫する」「アンサンブルを活用する」の三つの戦略があり、これらにもとづく多様な手法が提案されている。その中でも比較の実装が容易であり、効果・安定性・計算コストのバランスに優れた代表的な3手法を紹介し、それらの比較および選定の基準について述べる。その3手法とは、「クラス重み付けと閾値最適化」「標準モデル + synthetic minority over-sampling technique (SMOTE) 系のオーバーサンプリング」「アンダーサンプリングとアンサンブルの組み合わせ手法」である (表1)。

クラス重み付けと閾値最適化

クラス不均衡への対処法として最初に検討すべき手法の一つが「クラス重み付けと閾値最適化」である。この方法では、モデル学習時に少数クラスの損失を大きく評価することで分類器が多数クラスばかりを予測してしまうバイアスを抑制する³⁾。また、予測結果を発症/非発症の二値に変換する際には、機械的に0.5という固定閾値を用いるのではなく、感度や適合率 (precision) といった臨床上の目的に応じて、最適な閾値を再設定する⁴⁾。この調整により、例えば「0.3以上を陽性と判定する」といった柔軟な対応が可能となる。

この手法の利点は、実装の容易さにある。重みや閾値を調整するだけで、学習および推論の挙動を制御できる。例えば重み付けは、モデルアルゴリズムの引数オプション指定によりクラス出現頻度に応じた重みを自動的に設定できる。特に、不均衡比が中程度のクラス不均衡に対しては、軽量かつ高精度なベースライン構築手法として非常に有効である。ただし、重みの設定が過剰であると学習が不安定になったり、出力確率が歪む可能性があったりする。

SMOTEを代表とするオーバーサンプリング手法

SMOTEを代表とするオーバーサンプリング手法として挙げられるのが、SMOTEを代表とするオーバーサンプリング手法である。これは少数クラスのサンプルを単に複製するのではなく、近傍のサンプル間を補間することで新たなデータを人工的に生成し、データ空間に厚みをもたせる手法である。これにより、モデルはより明確なクラス境界を学習しやすくなる²⁾。

SMOTEには複数の拡張手法が存在する。例えば、分類が困難な境界領域に焦点を当てて合成を行う方法、数値とカテゴリが混在する特徴量に対応する方法、合成量をサンプルの難易度に応じて調整する方法などである。

SMOTEの利点は、もとのデータを保持したまま情報量を増やせる点にあるが、いくつか注意点も存在する²⁾。特に重要なのは、SMOTEは「学習用データにのみ適用すべき」であり、評価用のテストデータに適用してはならない点である。また、生成されたサンプルが実際の分布から逸脱している場合には、かえって学習の精度が損なわれるおそれがある。そのため、SMOTE適用前後で分布に大きな変化がないか、特徴量とアウトカムの散布図などを用いて確認することが望ましい。

アンダーサンプリングとアンサンブルの組み合わせ手法

最後は、EasyEnsembleを代表とするアンダーサンプリングとアンサンブルの組み合わせ手法である^{3,4)}。この手法では、少数クラスのデータは学習用データとしてすべて使用し、多数クラスはアンダーサンプリング、すなわち少数クラスのデータ数に合わせてランダムに抽出を行う。このアンダーサン

OGATA, Soshiro
国立循環器病研究センター
予防医学・疫学情報部

▼図1 アンダーサンプリングとアンサンブルの組み合わせ手法の概要

