

様々な評価方法に適応する機械学習

弱い教師データに基づく低分散なリスク推定方法の開発

包含 (つつみ ふくむ)

東京大学 情報理工学系研究科・博士課程

課題と研究の概要

Q. ユーザーの“目的”に応じた予測を行うには？

- 例
- 陽性の患者は漏れなく発見したい
 - クレジットカードの不正利用は必ず検知したい

成果: 与えられた評価指標の下で性能を最大化する学習方法

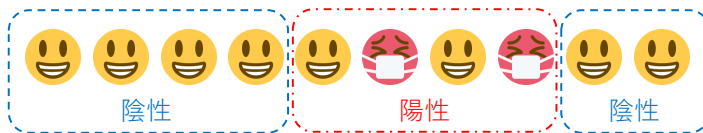
- F値など現実的な評価指標を考慮した分類器の学習方法を提案
- 勾配法を基にした効率的な最適化が可能
- 最大性能の達成を理論的に保証

ユーザーの意図を適応的に汲み取る学習・予測アルゴリズムが必要！

✗ 正答率80%だが、陽性患者が検出できず



○ 正答率80%だが、陽性患者を検出できている



研究成果の詳細

[Bao & Sugiyama AISTATS 2020]

Calibrated Surrogate Maximization of Linear-fractional Utility in Binary Classification

分数線形型の評価指標

$$U(f) = \frac{a_0 TP + b_0 FP + c_0}{a_1 TP + b_1 FP + c_1}$$

- 正答率よりも柔軟な真陽性(TP)と偽陽性(FP)のトレードオフ
- ユーザーが分類器に求める性質に応じて事前に指定

例: 正答率 $\frac{TP + TN}{2TP + FP + FN}$

F値 $\frac{2TP}{2TP + FP + FN}$

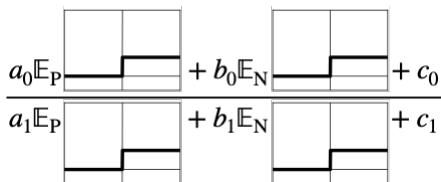
Jaccard指標 $\frac{TP}{TP + FP + FN}$

Q. どのように $U(f)$ を最大にする分類器 f を探す？

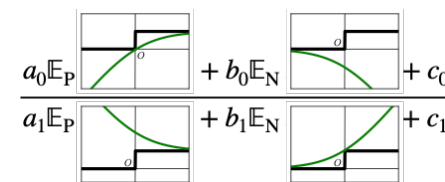
アプローチ: TP & FPを最適化しやすい関数に置き換える

元の評価指標

最適化しやすい「代理」関数



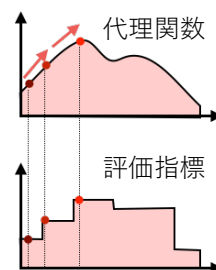
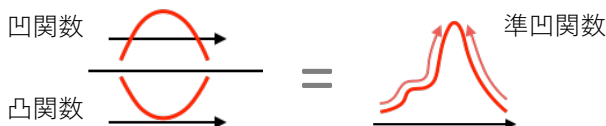
\geq



- 勾配が計算可能
- 「凹関数 / 凸関数」の形

性質① 効率的に最適化可能

性質② 最適解への収束保証



- 凹関数 / 凸関数 = 準凹関数 (山=最大点が1つだけ)
- (修正した) 勾配法で効率的に最適化可能

- 適合性: 最適解が元の関数の最適解と一致
- ある条件の下で適合的であることを証明

分数線形型の評価指標に対する初の適合性保証！