

Sensibilité des indices de qualité d'un classifieur probabiliste

Ndèye Awa Dieye*, Ndèye Niang *, Giorgio Russolillo*

* Laboratoire Cedric-MSDMA, Paris, France
{ndeye-awa.dieye, ndeye.niangkeita, iorgio.rugssolillo}@lecnam.net

1 Contexte et objectifs

En classification supervisée binaire, nous disposons d'un échantillon de n observations décrites par un ensemble de variables explicatives et une variable de réponse $y \in \{0, 1\}$. Un classifieur probabiliste estime la probabilité π de l'évènement pour chaque observation. L'évaluation d'un classifieur se fait en examinant son pouvoir discriminant, sa calibration ou les deux à la fois. Le pouvoir discriminant est la capacité à bien classer les individus et sa calibration mesure la fiabilité de l'estimation de la probabilité d'appartenance aux classes. La capacité discriminante est communément évaluée avec l'aire sous la courbe ROC (AUC). Pour l'évaluation de la calibration (Huang et al. (2020)), on distingue la calibration *in the small*, souvent mesurée par l'Expected Calibration Error (ECE) de la calibration *in the large*, mesurée en comparant le taux de l'évènement et la probabilité estimée moyenne. Le score de Brier est par contre utilisé pour mesurer la qualité globale du classifieur. Dans ce travail nous montrons que la sensibilité de ces indices à différents niveaux d'écart à l'ajustement parfait varie selon la forme de distribution des probabilités π et de la distribution des écarts à ces probabilités.

2 Méthodologie

Quatre vecteurs x de taille $n = 2000$ sont générés à partir de différentes lois (beta, normale et logistique). A partir de chacun de ces vecteurs, un vecteur de probabilités π est calculé par le biais de la fonction logistique $\pi = \frac{e^{\beta x}}{e^{\beta x} + 1}$ avec $\beta = 1$. Ces vecteurs de probabilités présentent des distributions variées (Fig.1). 1000 vecteurs y , issus de lois de Bernoulli de paramètre π sont générés à partir de chaque vecteur de probabilités. Les quatre vecteurs π sont déviés selon trois méthodes :

- Méthode 1 : la valeur du coefficient β est remplacée par les valeurs 0.75, 0.5 et 0.25. Cette méthode donne des déviations monotones, symétriques, plus accentuées sur les valeurs extrêmes.
- Méthode 2 : des bruits gaussiens de moyenne nulle et d'écart-type 0.2, 0.5, 1 et 2 sont ajoutés aux vecteurs π . Cette méthode donne des déviations non monotones, symétriques, plus accentuées sur les valeurs centrales.
- Méthode 3 : le premier tercile de la distribution des π est multiplié par 1.075, 1.15, 1.3, 1.4, 1.5 et le dernier tercile par 0.925, 0.85, 0.7, 0.6, 0.5. Cette méthode donne des déviations non monotones, asymétriques, plus accentuées sur les valeurs proches de 1.

Sensibilité des indices de qualité d'un classifieur

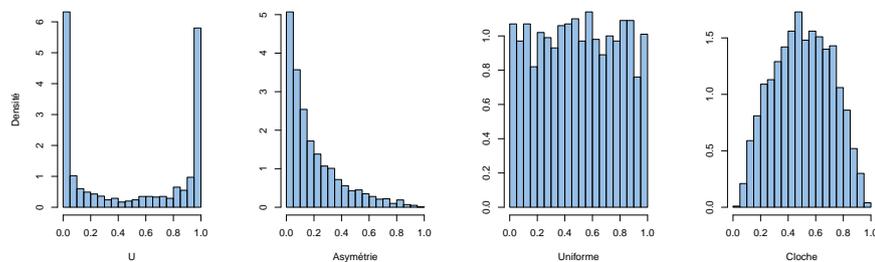


FIG. 1 – Distributions des probabilités π . De droite à gauche : forme en U (symétrique avec 2 modes aux valeurs extrêmes), asymétrie positive, uniforme et en cloche

Les 12 vecteurs de probabilités déviées π_{dev} obtenus avec ces méthodes sont considérés comme des estimations des 4 vecteurs π . Les indices de calibration, de discrimination et de qualité globale issus de la comparaison de chaque vecteur y et π_{dev} sont calculés.

3 Résultats et conclusion

Les valeurs optimales de l'AUC et du score de Brier, obtenues en comparant les vecteurs de réponse y et les probabilités π , dépendent de la distribution de ces probabilités. Lorsque la distribution présente plus de valeurs extrêmes (forme en U), la valeur optimale tend à être meilleure. En revanche, lorsque la distribution est plus concentrée autour des valeurs centrales (forme en cloche), la valeur optimale tend à être moins bonne.

Lorsque la distribution des vraies probabilités est caractérisée par beaucoup de valeurs extrêmes (grande variance), l'AUC montre une faible sensibilité. La mesure de la calibration *in the large* est plus sensible lorsque on retrouve une asymétrie concernant soit la distribution des probabilités, soit l'ampleur des écarts. L'ECE se révèle plus sensible lorsque les plus larges écarts se situent au niveau des probabilités avec une densité plus élevée.

Même si en théorie l'AUC et le score de Brier sont compris respectivement dans les intervalles $[0.5, 1]$ et $[0, 1]$, dans la pratique les valeurs optimales de ces indices dépendent de la distribution des probabilités π qui ont généré la variable de réponse binaire, qui est inconnue. De plus, la sensibilité de tous les indices considérés dans l'étude dépend de la distribution de π . Il convient donc d'être prudent dans l'interprétation de la qualité des classifieurs probabilistes avec ces indices.

Références

Huang, Y., W. Li, F. Macheret, R. A. Gabriel, et L. Ohno-Machado (2020). A tutorial on calibration measurements and calibration models for clinical prediction models. *Journal of the American Medical Informatics Association* 27, 621–633.