Apprentissage d'un classifieur binaire par règles d'association

1 Introduction

Dans le présent travail, nous proposons une approche consistant à déterminer les profils, expression d'interactions entre les covariables, corrélés avec la variable réponse pour construire une fonction de classement. Cette approche est en étroite liaison avec la notion de règles d'association. Des approches similaires ont été proposées dans la littérature du domaine de l'intelligence artificielle ces dernières années [6–8]. L'idée principale consiste à rechercher un ensemble optimal de profils à partir d'un ensemble de profils fréquents. La stratégie consiste à élaguer les profils redondants et les profils de faible performance en se basant essentiellement sur les mesures statistiques suivantes : la sensibilité, la spécificité et les valeurs prédictives. Le présent travail vise à insérer cette approche dans le cadre de la statistique traditionnelle et à montrer la pertinence de son application dans un problème réel.

2 Profils et classement basé sur un profil

On considère un couple de variables aléatoires (Y, X), où Y est une variable de Bernoulli et $X = (X_j)_{j=1:p}$ est une suite finie de p variables aléatoires où chaque X_j est une variable non numérique à q_j modalités $m_{h(j)}, h(j) = 1 : q_j; j = 1 : p$.

2.1 Profil

Définition 1. On appelle profil toute suite finie d'événements $\left(X_j = m_{h(j)}\right)_{j\in J}$, où $J \subseteq 1 : p$ et $m_{h(j)}$ est une modalité de la variable X_j.

La longueur du profil $\left(X_j = m_{h(j)}\right)_{j\in J}$ est égale à la taille (cardinal) de l'ensemble $J \subset 1 : p$. Pour simplifier les notations dans la suite, on écrit $m_h^{X_j}$ pour désigner la modalité $m_{h(j)}$ de la variable X_j et on note $\left(m_h^{X_j}\right)_{j\in J}$ pour désigner le profils $\left(X_j = m_{h(j)}\right)_{j\in J}$.

Un profil peut être vu comme la réalisation conjointe de $|J|$ variables $(X_j)_{j\in J}$. Plus la taille du profil est grande, plus le nombre de variables conjointement réalisées augmente. Dans le domaine de l'intelligence artificielle et de l'apprentissage automatique, un profil est plus connu sous le nom d'item-set. Un profil de taille k est un k-itemset. Un profil $\left(m_h^{X_j}\right)_{j\in J}$ peut être compris comme l'expression d'une interaction entre les différentes variables non numériques $(X_j)_{j\in J}$ qui le définissent. La taille d'un profil est équivalente à la complexité d'une interaction dans un modèle paramétrique tel que la régression logistique. La gestion des interactions existant entre les covariables est l'un des avantages d'un profil par rapport aux modèles paramétriques. Un profil est pertinent lorsque sa probabilité d'occurrence est significative.

Définition 2. Soient $\left(m_h^{X_l}\right)_{l\in L}$ et $\left(m_h^{X_j}\right)_{j\in J}$ deux profils. On dit que $\left(m_h^{X_j}\right)_{j\in J}$ est emboîté dans $\left(m_h^{X_l}\right)_{l\in L}$ si les conditions suivantes sont vérifiées.

a) $L \subset J$

b) $\forall\, l \in L, \forall h \in \{1 : q_l\} \quad \exists !\, j \in J, \exists !\, k \in \{1 : q_j\}$ tel que $m_h^{X_l} = m_k^{X_j}$

Ils sont disjoints si $L \cap J = \emptyset$.

2.2 Classement associé à un profil et paramètres de performance

On peut associer à tout profil $U = \left(m_h^{X_j}\right)_{j\in J}$ une fonction indicatrice $\phi(\cdot, U)$ définie par :

$$\phi(X,U) = \prod_{j\in J} \mathbb{1}_{\left(X_j = m_h^{X_j}\right)}(X)$$

Par définition $\phi(\cdot, U)$ est un classifieur binaire. $\phi(X,U) = 1$ si tous les événements $[X_j = m_h^{X_j}]$ sont conjointement réalisés. Dans le domaine de l'intelligence artificielle, on appelle couverture du profil $U = \left(m_h^{X_j}\right)_{j\in J}$ la probabilité $\Pr\{\phi(X,U)=1\}$ et on appelle support du profil $U = \left(m_h^{X_j}\right)_{j\in J}$ la probabilité $\Pr\{\phi(X,U)=1, Y=1\}$.

Dans cette analyse, nous nous plaçons dans le cadre de la statistique pour aborder le problème. A chaque profil U, un seul classifieur $\phi(X,U)$ lui est associé. Par la suite, on peut remarquer que la pertinence d'un profil est étroitement liée avec la performance du classifieur qui lui est associé. Ainsi on peut donc utiliser les indicateurs de performance des classifieurs associés pour sélectionner un ensemble réduit de profils pertinents dont on se servira pour construire une règle de classement efficace. Cependant plusieurs indicateurs de performance ont été proposés dans la littérature pour évaluer les performances d'un classifieur donné. Parmi les plus utilisés figure l'erreur de classement. L'erreur de classement $Err(U,Y)$ d'un classifieur $\phi(X,U)$ engendré par un profil U est définie par :

$$Err(U,Y) = \Pr\{\phi(X,U) \neq Y\} = \Pr\{\phi(X,U)=1, Y=0\} + \Pr\{\phi(X,U)=0, Y=1\}$$

On peut en déduire alors l'expression suivante :

$$Err(U,Y) = \Pr\{Y=1\} + \Pr\{\phi(X,U)=1\} - 2\Pr\{Y=1,\phi(X,U)=1\}$$

On constate que l'erreur de classement est gouverné par le support $\Pr\{Y=1,\phi(X,U)=1\}$ du profil U. L'erreur de classement est une fonction décroissante du support du profil. Pour deux profils de même couverture, l'erreur de classement décroît avec le support des profils. Par conséquent, plus le support du profil est élevé meilleur est le profil. On s'intéressera alors aux profils pour lesquels les classifieurs associés réalisent des probabilités $\Pr(Y=1,\phi(X,U)=1)$ supérieurs à un seuil s_0.

Pour un classifieur binaire, on considère en particulier la sensibilité et la spécificité définie par

$$\text{Sensib}(U,Y) = \frac{\Pr(\phi(X,U)=1, Y=1)}{\Pr(Y=1)}$$

$$\text{Spécif}(U,Y) = \frac{\Pr(\phi(X,U)=0, Y=0)}{\Pr(Y=0)}$$

On observe que la sensibilité croît avec la probabilité $\Pr(\phi(X,U)=1,Y=1)$. Deux autres paramètres pourront aider à l'évaluation de la qualité du classifieur $\phi(X,U)$ donc à la sélection du classifieur dans un ensemble de classifieurs : la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN).

$$VPP(U,Y) = \frac{\Pr(\phi(X,U)=1, Y=1)}{\Pr(\phi(X,U)=1)}$$

$$VPN(U,Y) = \frac{\Pr(\phi(X,U)=0, Y=0)}{\Pr(\phi(X,U)=0)}$$

On peut établir les relations suivantes :

$$\text{Sensib}(U,Y) = VPP(U,Y)\frac{\Pr(\phi(X,U)=1)}{\Pr(Y=1)}$$

$$\text{Spécif}(U,Y) = 1 - [1 - VPP(U,Y)]\,\frac{\Pr(\phi(X,U)=1)}{1-\Pr(Y=1)}$$

$$VPN(U,Y) = [1 - VPP(U,Y)]\,\frac{\Pr(\phi(X,U)=1)}{\Pr(\phi(X,U)=0)}$$

Pour deux profils U_1 et U_2 de même probabilité d'occurrence (couverture), la spécificité croît avec la valeur prédictive positive du classifieur. Il en résulte que parmi les profils U de même couverture $\Pr(\phi(X,U)=1)$, on pourra s'intéresser à ceux pour lesquels les valeurs prédictives positives des classifieurs associés sont au dessus d'un seuil c_0 .

La valeur prédictive positive d'un profil est communément appelée confiance dans le domaine de

l'intelligence artificielle et de l'apprentissage automatique. En plus de la valeur prédictive positive (VPP) et de la valeur prédictive négative (VPN), on peut aussi baser la sélection des profils sur les paramètres suivants :
Le rapport de vraisemblance positif du profil U que nous notons par $RVP()$ est défini par :

$$RVP(U,Y) = \frac{\Pr(\phi(X,U)=1|Y=1)}{\Pr(\phi(X,U)=1|Y=0)}$$

on a alors

$$\begin{aligned} RVP(U,Y) &= \frac{\Pr\{Y=0\}}{\Pr\{Y=1\}} \frac{\Pr\{\phi(X,U)=1, Y=1\}}{\Pr\{\phi(X,U)=1, Y=0\}} \\ &= \frac{VPP(U,Y)}{1-VPP(U,Y)} \frac{\Pr\{Y=0\}}{\Pr\{Y=1\}} \end{aligned}$$

Le rapport de vraisemblance négatif du profil U que nous notons par $RVN()$ est défini par :

$$RVN = \frac{\Pr(\phi(X,U)=0|Y=1)}{\Pr(\phi(X,U)=0|Y=0)}$$

on a alors

$$\begin{aligned} RVN(U,Y) &= \frac{\Pr\{Y=0\}}{\Pr\{Y=1\}} \frac{\Pr\{\phi(X,U)=0\} - \Pr\{\phi(X,U)=0, Y=0\}}{\Pr\{\phi(X,U)=0, Y=0\}} \\ &= \frac{1-VPN(U,Y)}{VPN(U,Y)} \frac{\Pr\{Y=0\}}{\Pr\{Y=1\}} \end{aligned}$$

On a aussi $RVP(U,Y) = \dfrac{Sensibilite(U,Y)}{1-Specificite(U,Y)}$. Et donc $RVP(U,V) > 1$ entraîne que le classifieur $\phi(X,U)$ a de meilleurs indicateurs de performance que le classifieur de même sensibilité qui consiste à classer positive au hasard toute nouvelle observation. C'est à dire sur une courbe ROC [2], la courbe du classifieur $\phi(X,U)$ se situe au dessus de la première bissectrice.

Le risque relatif du profil U que nous notons $RR()$ est défini par :

$$RR(U,Y) = \frac{\Pr(Y=1|\phi(X,U)=1)}{\Pr(Y=1|\phi(X,U)=0)}$$

On peut établir que

$$RR(U,Y) = VPP(U,Y) \frac{1-\Pr\{\phi(X,U)=1\}}{\Pr\{Y=1\} - \Pr\{Y=1, \phi(X,U)=1\}}$$

Notons par G_1 le groupe d'objets vérifiant le profil U et G_0 le groupe d'objets ne vérifiant pas le profil

U. Le risque relatif est une mesure statistique qui permet de comparer la probabilité d'occurrence de l'événement $[Y = 1]$ dans G_1 par rapport à G_0. Le profil $U = \left(m_h^{X_l}\right)_{l \in L}$ est un profil à risque [4] pour Y si le risque relatif excède un seuil τ donné. La probabilité d'occurrence de l'événement $[Y = 1]$ dans G_1 est τ fois plus importante que la probabilité d'occurrence de l'événement $[Y = 1]$ dans G_0. Dans la suite, nous nous intéresserons alors aux profils U pour lesquels la probabilité d'occurrence de $[Y = 1]$ dans G_1 est τ fois plus importante que la probabilité d'occurrence de $[Y = 1]$ dans G_0. Par défaut le paramètre τ est supérieur à un $(\tau > 1)$. Le sous-ensemble de profils U pour lequel la probabilité conditionnelle $\Pr([Y = 1] | [\phi(X_i, U) = 1])$ est plus élevée que la probabilité conditionnelle $\Pr([Y = 1] | [\phi(X_i, U) = 0])$ constitue un ensemble potentiel pour construire un bon classifieur.

Les critères conventionnels d'évaluation utilisés, tels que la précision globale ou le taux d'erreur, ne fournit pas suffisamment d'informations dans le cas de l'apprentissage déséquilibré. En effet, des mesures d'évaluation plus performantes, telles que les courbes ROC (receiver operating characteristc), les courbes de précision-sensibilité et les courbes de coûts, sont nécessaires à l'évaluation concluante d'un classifieur en présence de données déséquilibrées. L'expression de l' aire en dessous de la courbe ROC d'un classifieur généré par un profil U est donnée par :

$$AUC(U, Y) = \begin{cases} \frac{1}{2}(\text{Sensib}(U, Y) + \text{Spécif}(U, Y)) & \text{si Sensib}(U, Y) + \text{Spécif}(U, Y) \geq 1 \\ \\ 1 - \frac{1}{2}(\text{Sensib}(U, Y) + \text{Spécif}(U, Y)) & \text{si Sensib}(U, Y) + \text{Spécif}(U, Y) < 1 \end{cases}$$

L'aire sous la courbe ROC (AUC) est une mesure utile pour évaluer la performance d'un profil. La comparaison des AUC de différents profils peut établir une relation de domination entre les profils. On peut l'utiliser alors pour la sélection d'un sous ensemble optimal de profils.

A partir de cette section, il apparaît clairement que les principaux paramètres d'apprentissage d'un classifieur basé sur un ensemble optimal de profils sont le support et la valeur prédictive positive. Ils permettent de gérer à la fois l'erreur de classement et la sensibilité du classifieur. Dans toute la suite, nous nous intéressons aux profils dont le support est supérieur à un seuil s_0 et la valeur prédictive positive (confiance) est supérieure à c_0.

Dans la littérature de la classification associative (classification supervisée basée sur les règles d'association), plusieurs mesures de performance ont été proposées pour l'extraction de règles d'association [3]. Une étude comparative exhaustive de plusieurs mesures de performance a été menée dans [9]. La plupart des mesures de performance sont destinées à découvrir les profils les plus fréquents. Raison pour laquelle la majeur partie d'entre elles ne sont pas appropriées lorsqu'il s'agit de traiter un problème de classification supervisée sur des données déséquilibrées. Le support et la confiance restent les mesures de performance les plus utilisées dans les algorithmes d'extraction des règles d'association basés sur la sélection des profils fréquents. Dans ces algorithmes, généralement le support est utilisé pour trouver les profils fréquents suivant sa propriété d'anti-monotonicité [1, 4]. Quant à la confiance elle est utilisée pour générer les règles à partir des profils fréquents et à les filtrer à l'aide d'un seuil de confiance minimum.

Selon ses propriétés, chaque mesure est utile pour certaines applications, mais pas pour d'autres [12]. Ces mesures peuvent produire des informations contradictoires sur l'intérêt et la pertinence d'un profil. Un exemple bien connu d'une telle mesure controversée est le support. D'une part, il est grandement utilisé à des fins de filtrage dans les algorithmes d'extraction [1, 10], puisque sa propriété d'anti-monotonicité simplifie le vaste ensemble de profils qui doit être exploré. D'autre part, il a presque tous les défauts qu'un utilisateur souhaite éviter par exemple la variabilité de la valeur sous l'hypothèse d'indépendance [11].

A notre connaissance, seuls la sensibilité connue sous le nom de support local, le risque relatif et l'odds ratio ont été utilisés pour la recherche d'un ensemble optimal de profil dans le cadre d'un problème de classification supervisée sur des données déséquilibrées par Li et al.[5]. En présence de données déséquilibrées, le support $\Pr\{\phi(X,U)=1, Y=1\}$ d'un profil U serait guère fréquent lorsque la classe d'intérêt $\{Y=1\}$ est rare. C'est pourquoi Li et al. ont définit le support local (sensibilité) $\Pr\{\phi(X,U)=1 \,|\, Y=1\}$ comme étant le support d'un profil dans le groupe d'observations vérifiant $\{Y=1\}$ puisque le support local vérifie la propriété d'anti-monotonicité du support. Ainsi un profil U est fréquent lorsque son support local est supérieure à un seuil minimum fixé. Leurs résultats ont montré que la sensibilité et le risque relatif sont des mesures statistiques pertinentes pour la sélection de profils optimaux lorsqu'on traite des données déséquilibrées.

Les algorithmes d'extraction de règles d'association basés sur les profils fréquents produisent en général un vaste ensemble de règles d'association dont la majeur partie sont triviales et sans intérêts. Pour construire un classifieur performant à partir du vaste ensemble de règles d'association explorées, nous allons donc établir une stratégie d'élagage des profils redondants et une stratégie de réduction de l'ensemble des profils fréquents et non redondants.

2.3 Profils redondants et sélection de profils

A l'instar des méthodes standards de classement, la procédure de sélection de profils que nous proposons s'intéressera en particulier aux profils qui sont corrélés avec la variable réponse.

Proposition 1. Soient $U = \left(m_h^{X_l}\right)_{l \in L}$ et $U' = \left(m_h^{X_j}\right)_{j \in J}$ deux profils. Si U' est emboîté dans U alors :

1. $\Pr\{\phi(X,U) = 1, Y = 1\} \geq \Pr\{\phi(X,U') = 1, Y = 1\}$
2. $\Pr\{\phi(X,U) = 0, Y = 0\} \leq \Pr\{\phi(X,U') = 0, Y = 0\}$

Preuve. Pour simplifier les expressions, on note $\phi(X,U)$ par ϕ_U et $\phi(X,U')$ par $\phi_{U'}$.
Par hypothèse U' est emboité dans U donc on a

$$\{\phi_U = 1\} \supset \{\phi_{U'} = 1\} \quad \text{et} \quad \{\phi_U = 0\} \subset \{\phi_{U'} = 0\}$$

On en déduit que :

$$\{\phi_U = 1\} \supset \{\phi_{U'} = 1\} \Rightarrow \{\phi_U = 1, Y = 1\} \supset \{\phi_{U'} = 1, Y = 1\} \Rightarrow \Pr\{\phi_U = 1, Y = 1\} \geq \Pr\{\phi_{U'} = 1, Y = 1\}$$

$$\{\phi_U = 0\} \subset \{\phi_{U'} = 0\} \Rightarrow \{\phi_U = 0, Y = 0\} \subset \{\phi_{U'} = 0, Y = 0\} \Rightarrow \Pr\{\phi_U = 0, Y = 0\} \leq \Pr\{\phi_{U'} = 0, Y = 0\}$$

□

Définition 3. Soient $U = \left(m_h^{X_l}\right)_{l \in L}$ et $U' = \left(m_h^{X_j}\right)_{j \in J}$ deux profils tels que U' soit emboîté dans U. On dit que le profil U' est redondant par rapport à U, si le(s) indicateur(s) de performance de la fonction de classement $\phi(\cdot, U)$ générée par le profil U est (sont) plus élevé(s) que le(s) indicateur(s) de performance de la fonction de classifieur $\phi(\cdot, U')$ générée par le profil U'.

Proposition 2. Soient $U = \left(m_h^{X_l}\right)_{l \in L}$ et $U' = \left(m_h^{X_j}\right)_{j \in J}$ deux profils. Si U' est emboîté dans U alors la valeur prédictive positive du classifieur généré par le profil U est comprise entre

$$Min\left\{VPP(U',Y)\,,\ \frac{\Pr(Y=1,\phi(X,U)=1) - \Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1) - \Pr(\phi(X,U')=1)}\right\}$$

et

$$Max\left\{VPP(U',Y)\,,\ \frac{\Pr(Y=1,\phi(X,U)=1) - \Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1) - \Pr(\phi(X,U')=1)}\right\}$$

Preuve. On a

$$\begin{aligned}
VPP(U,Y) &= \frac{\Pr(Y=1,\phi(X,U)=1)}{\Pr(\phi(X,U)=1)} \\
VPP(U,Y) &= \frac{\Pr(Y=1,\phi(X,U)=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)} + \frac{\Pr(Y=1,\phi(X,U)=1,\phi(X,U')=0)}{\Pr(\phi(X,U)=1)} \\
&= \frac{\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)} + \frac{\Pr(Y=1,\phi(X,U)=1) - \Pr(Y=1,\phi(X,U)=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)} \\
&= VPP(U',Y)\frac{\Pr(\phi(X,U')=1)}{\Pr(\phi(X,U)=1)} + \frac{\Pr(Y=1,\phi(X,U)=1) - \Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1) - \Pr(\phi(X,U')=1)}\left[1 - \frac{\Pr(\phi(X,U')=1)}{\Pr(\phi(X,U)=1)}\right]
\end{aligned}$$

On obtient une combinaison convexe de $VPP(U,Y)$ et $\frac{\Pr(Y=1,\phi(X,U)=1)-\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)-\Pr(\phi(X,U')=1)}$ par rapport à $\frac{\Pr(\phi(X,U')=1)}{\Pr(\phi(X,U)=1)}$. On en déduit que $VPP(U,Y$ est comprise entre

$$Min\left\{\frac{\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U')=1)},\frac{\Pr(Y=1,\phi(X,U)=1)-\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)-\Pr(\phi(X,U)=1)}\right\}$$

et

$$Max\left\{\frac{\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U')=1)},\frac{\Pr(Y=1,\phi(X,U)=1)-\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)-\Pr(\phi(X,U)=1)}\right\}$$

□

En résumé de la proposition 2, on a

1. Si $VPP(Y,U') < \dfrac{\Pr(Y=1,\phi(X,U)=1)-\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)-\Pr(\phi(X,U')=1)}$

   alors $VPP(Y,U') < VPP(Y,U)$. C'est à dire que U' est redondant par rapport à U. Par conséquent, on peut éliminer le plus long puisque sa sensibilité est plus faible et son erreur de classement est plus forte. Par contre

2. Si $VPP(Y,U') > \dfrac{\Pr(Y=1,\phi(X,U)=1)-\Pr(Y=1,\phi(X,U')=1)}{\Pr(\phi(X,U)=1)-\Pr(\phi(X,U')=1)}$

   alors $VPP(Y,U') > VPP(Y,U)$. Il est préférable de garder le profil U' au profit du profil U, puisque les indicateurs de performance du profil U sont meilleurs que ceux du profil U.

Proposition 3. Soients $U=\left(m_h^{X_l}\right)_{l\in L}$ et $U'=\left(m_h^{X_j}\right)_{j\in J}$ deux profils tels que U' soit emboîté dans U. Alors $\Pr\{\phi(X,U)=1\}=\Pr\{\phi(X,U')=1\}$ si et seulement si

1. $\Pr\{\phi(X,U)=1,Y=1\}=\Pr\{\phi(X,U')=1,Y=1\}$
2. $\Pr\{\phi(X,U)=0,Y=0\}=\Pr\{\phi(X,U')=0,Y=0\}$

Preuve. Supposons que $\Pr\{\phi_U=1\}=\Pr\{\phi_{U'}=1\}$. On a

$$\Pr\{\phi_U=1\}=\Pr\{\phi_U=1,Y=1\}+\Pr\{\phi_U=1,Y=0\}$$
$$\Pr\{\phi_{U'}=1\}=\Pr\{\phi_{U'}=1,Y=1\}+\Pr\{\phi_{U'}=1,Y=0\}$$

On obtient

$$\Pr\{\phi_U=1,Y=1\}+\Pr\{\phi_U=1,Y=0\}=\Pr\{\phi_{U'}=1,Y=1\}+\Pr\{\phi_{U'}=1,Y=0\}\qquad(a)$$

Puisque $[\phi_{U'}=1]\subset[\phi_U=1]$ alors

$$\Pr\{\phi_U=1,Y=1\}-\Pr\{\phi_{U'}=1,Y=1\}\geq 0\qquad(b)$$

$$\Pr\{\phi_U=1,Y=0\}-\Pr\{\phi_{U'}=1,Y=0\}\geq 0\qquad(c)$$

On peut déduire de (a), (b) et (c) les égalités suivantes :

$$\Pr\{\phi_U=1,Y=1\}=\Pr\{\phi_{U'}=1,Y=1\}\qquad(1)$$

$$\Pr\{\phi_{U'} = 1, Y = 0\} = \Pr\{\phi_U = 1, Y = 0\} \qquad (*)$$

Par ailleurs on a

$$\Pr\{\phi_{U'} = 1, Y = 0\} = \Pr\{Y = 0\} - \Pr\{\phi_{U'} = 0, Y = 0\} \qquad (**)$$

$$\Pr\{\phi_U = 1, Y = 0\} = \Pr\{Y = 0\} - \Pr\{\phi_U = 0, Y = 0\} \qquad (***)$$

En faisant la différence membre à membre des égalités $(**)$ et $(***)$ et en tenant compte de l'égalité $(*)$, on obtient

$$\Pr\{\phi_{U'} = 0, Y = 0\} = \Pr\{\phi_U = 0, Y = 0\} \qquad (2)$$

Supposons maintenant que les égalités suivantes soient vraies :

$$\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\} \qquad (1)$$

$$\Pr\{\phi_{U'} = 0, Y = 0\} = \Pr\{\phi_U = 0, Y = 0\} \qquad (2)$$

De l'égalité (2) on déduit

$$\Pr\{Y = 0\} - \Pr\{\phi_U = 1, Y = 0\} = \Pr\{Y = 0\} - \Pr\{\phi_{U'} = 1, Y = 0\}$$

On obtient alors les égalités suivantes

$$\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$$

$$\Pr\{\phi_U = 1, Y = 0\} = \Pr\{\phi_{U'} = 1, Y = 0\}$$

En faisant les sommes membres à membres des deux égalités on obtient :

$$\Pr\{\phi_U = 1, Y = 1\} + \Pr\{\phi_U = 1, Y = 0\} = \Pr\{\phi_U = 1\}$$

$$\Pr\{\phi_{U'} = 1, Y = 1\} + \Pr\{\phi_{U'} = 1, Y = 0\} = \Pr\{\phi_{U'} = 1\}$$

D'où

$$\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$$

□

Lorsqu'on divise par $\Pr(Y = 1)$ les deux termes de l'égalité 1 de la proposition 3, on obtient que le profil U' est redondant par rapport au profil U selon la définition 3. Le même résultat est obtenu en divisant les deux termes de l'égalité 2 par $\Pr(Y = 0)$.

Corollaire 1. Soients $U = \left(m_h^{X_l}\right)_{l \in L}$ et $U' = \left(m_h^{X_j}\right)_{j \in J}$ deux profils tels que U' soit emboîté dans U. Les propositions suivantes sont équivalentes :

1.

$$\Pr\{\phi(X, U) = 1\} = \Pr\{\phi(X, U') = 1\}$$

2.

$$\begin{cases} VPP(U,Y) = VPP(U',Y) \\ \\ \Pr\{\phi(X,U)=1, Y=1\} = \Pr\{\phi(X,U')=1, Y=1\} \end{cases}$$

3.

$$\begin{cases} VPN(U,Y) = VPN(U',Y) \\ \\ \Pr\{\phi(X,U)=0, Y=0\} = \Pr\{\phi(X,U')=0, Y=0\} \end{cases}$$

4.

$$\begin{cases} RVP(U,Y) = RVP(U',Y) \\ \\ \Pr\{\phi(X,U)=1, Y=1\} = \Pr\{\phi(X,U')=1, Y=1\} \end{cases}$$

5.

$$\begin{cases} RVN(U,Y) = RVN(U',Y) \\ \\ \Pr\{\phi(X,U)=1, Y=1\} = \Pr\{\phi(X,U')=1, Y=1\} \end{cases}$$

6.

$$\begin{cases} Err(U,Y) = Err(U',Y) \\ \\ \Pr\{\phi(X,U)=1, Y=1\} = \Pr\{\phi(X,U')=1, Y=1\} \end{cases}$$

7.

$$\begin{cases} RR(U,Y) = RR(U',Y) \\ \\ \Pr\{\phi(X,U)=1, Y=1\} = \Pr\{\phi(X,U')=1, Y=1\} \end{cases}$$

Preuve. Pour simplifier les expressions, on note par $\phi(X,U)$ par ϕ_U et $\phi(X,U')$ par $\phi_{U'}$.

1)Montrons que 1. est équivalent à 2.
Supposons que 1. est vrai
D'après la proposition 3, si $\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$ alors

$$\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$$

Si on divise les termes respectives de cette dernière égalité par $\Pr\{\phi_U = 1\}$ et $\Pr\{\phi_{U'} = 1\}$ respectivement, on obtient

$$\frac{\Pr\{\phi_U = 1, Y = 1\}}{\Pr\{\phi_U = 1\}} = \frac{\Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{\phi_{U'} = 1\}}$$

Réciproquement : supposons que 2. soit vrai
Si 2. est vrai alors on a $VPP(U, Y) - VPP(U', Y) = 0$
D'où

$$\frac{\Pr\{\phi_U = 1, Y = 1\}[\Pr\{\phi_{U'} = 1\} - \Pr\{\phi_U = 1\}]}{\Pr\{\phi_U = 1\}\Pr\{\phi_{U'} = 1\}} = 0$$

On en déduit que

$$\Pr\{\phi_{U'} = 1\} - \Pr\{\phi_U = 1\} = 0$$

2) Montrons que 1. est équivalent à 3.
Supposons que 1. soit vrai
On a $\Pr\{\phi_U = 1\} = 1 - \Pr\{\phi_U = 0\}$ et $\Pr\{\phi_{U'} = 1\} = 1 - \Pr\{\phi_{U'} = 0\}$ Donc

$$\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\} \quad \Rightarrow \quad \Pr\{\phi_U = 0\} = \Pr\{\phi_{U'} = 0\}$$

D'après la proposition 3, on a aussi $\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$ entraîne que

$$\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$$

En divisant les termes respectives de l'égalité ci-dessus par $\Pr\{\phi_U = 0\}$ et $\Pr\{\phi_{U'} = 0\}$ respectivement, on obtient

$$\frac{\Pr\{\phi_U = 0, Y = 0\}}{\Pr\{\phi_U = 0\}} = \frac{\Pr\{\phi_U = 0, Y = 0\}}{\Pr\{\phi_{U'} = 0\}}$$

Réciproquement : supposons que 2. soit vrai
Si 3.1 est vrai alors $VPN(U, Y) - VPN(U', Y) = 0$. On obtient donc

$$\frac{\Pr\{\phi_U = 0, Y = 0\}[\Pr\{\phi_{U'} = 0\} - \Pr\{\phi_U = 0\}]}{\Pr\{\phi_U = 0\}\Pr\{\phi_{U'} = 0\}} = 0$$

Il en résulte que

$$\begin{aligned} \Pr\{\phi_{U'} = 0\} &= \Pr\{\phi_U = 0\} \\ 1 - \Pr\{\phi_{U'} = 1\} &= 1 - \Pr\{\phi_U = 1\} \end{aligned}$$

d'où

$$\Pr\{\phi_{U'} = 1\} = \Pr\{\phi_U = 1\}$$

3) Montrons que 1. est équivalent à 4.
On suppose que 1. est vrai
Par définition on a

$$RVP(U, Y) = \frac{\Pr\{Y = 0\}}{\Pr\{Y = 1\}} \frac{\Pr\{\phi_U = 1, Y = 1\}}{\Pr\{Y = 0\} - \Pr\{\phi_U = 0, Y = 0\}}$$

Et d'après la proposition 3, on a $\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$ entraîne que

$$\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\} \qquad \text{et} \qquad \Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$$

Donc si on remplace $\Pr\{\phi_U = 1, Y = 1\}$ par $\Pr\{\phi_{U'} = 1, Y = 1\}$ et $\Pr\{\phi_U = 0, Y = 0\}$ par $\Pr\{\phi_{U'} = 0, Y = 0\}$ dans l'expression de $RVP(U,Y)$, on obtient

$$RVP(U,Y) = \frac{\Pr\{Y=0\}}{\Pr\{Y=1\}} \frac{\Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{Y=0\} - \Pr\{\phi_{U'} = 0, Y = 0\}}$$

D'où

$$RVP(U,Y) = RVP(U',Y)$$

Réciproquement : supposons que 4. soit vrai
Si 4. est vrai alors $RVP(U,Y) - RVP(U',Y) = 0$. Il en résulte que

$$\frac{\Pr\{\phi_U = 1, Y = 1\}}{\Pr\{Y=0\} - \Pr\{\phi_U = 0, Y = 0\}} - \frac{\Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{Y=0\} - \Pr\{\phi_{U'} = 0, Y = 0\}} = 0$$

Puisque $\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$ on en déduit donc que $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_U = 0, Y = 0\}$.
D'après la proposition 3, on a donc

$$\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$$

4) Montrons que 1. est équivalent à 5.
On suppose que 1. est vrai

Par définition on a

$$RVN(U,Y) = \frac{\Pr\{Y=0\}}{\Pr\{Y=1\}} \frac{\Pr\{Y=1\} - \Pr\{\phi_U = 1, Y = 1\}}{\Pr\{\phi_U = 0, Y = 0\}}$$

Et d'après la proposition 3, on a $\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$ entraîne que

$$\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\} \qquad \text{et} \qquad \Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$$

Donc si on remplace $\Pr\{\phi_U = 1, Y = 1\}$ par $\Pr\{\phi_{U'} = 1, Y = 1\}$ et $\Pr\{\phi_U = 0, Y = 0\}$ par $\Pr\{\phi_{U'} = 0, Y = 0\}$ dans l'expression de $RVN(U,Y)$, on obtient

$$RVN(U,Y) = \frac{\Pr\{Y=0\}}{\Pr\{Y=1\}} \frac{\Pr\{Y=1\} - \Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{\phi_{U'} = 0, Y = 0\}}$$

D'où

$$RVN(U,Y) = RVN(U',Y)$$

Réciproquement : supposons que 5. soit vrai
Si 5. est vrai alors $RVN(U,Y) - RVN(U',Y) = 0$. On peut en déduire que

$$\frac{\Pr\{Y=1\} - \Pr\{\phi_U = 1, Y = 1\}}{\Pr\{\phi_U = 0, Y = 0\}} - \frac{\Pr\{Y=1\} - \Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{\phi_{U'} = 0, Y = 0\}} = 0$$

Puisque $\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$, on obtient donc que $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$. D'où

$$\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$$

d'après la proposition 3

5) Montrons que 1. est équivalent à 6.
On suppose que 1. est vrai
On a

$$Err(U,Y) = \Pr\{Y=1\} + \Pr\{\phi_U = 1\} - 2\Pr\{\phi_U = 1, Y = 1\} \qquad (1)$$

$$Err(U',Y) = \Pr\{Y=1\} + \Pr\{\phi_{U'} = 1\} - 2\Pr\{\phi_{U'} = 1, Y = 1\} \qquad (2)$$

Si $\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$ alors $\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$ (proposition 3)

Il en résulte des égalités précédentes que

$$Err(U, Y) = Err(U', Y)$$

Réciproquement : supposons que 6. soit vrai
Si on les différences membre à membres des égalités (1) et (2) ci-dessus, on obtient

$$\Pr\{\phi_U = 1\} - \Pr\{\phi_{U'} = 1\} = Err(U, Y) - Err(U', Y) + 2(\Pr\{\phi_U = 1, Y = 1\} - \Pr\{\phi_{U'} = 1, Y = 1\})$$

donc si 6. est vrai alors $\Pr\{\phi_U = 1\} - \Pr\{\phi_{U'} = 1\} = 0$

6) Montrons que 1. est équivalent à 7.
On suppose que 1. est vrai
On a

$$\begin{aligned}\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\} \quad &\Leftrightarrow \quad 1 - \Pr\{\phi_U = 1\} = 1 - \Pr\{\phi_{U'} = 1\} \\ &\Leftrightarrow \quad \Pr\{\phi_U = 0\} = \Pr\{\phi_{U'} = 0\}\end{aligned}$$

alors

$$\frac{\Pr\{\phi_U = 0\}}{\Pr\{\phi_U = 1\}} = \frac{\Pr\{\phi_{U'} = 0\}}{\Pr\{\phi_{U'} = 1\}} \qquad (1)$$

D'autre part, on a $\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$ entraîne que

(a) $\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$

(b) $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$

d'aprés la proposition 3. Puisque $\Pr\{\phi_U = 0\} = \Pr\{\phi_{U'} = 0\}$ et $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$ alors

$$\Pr\{\phi_U = 0, Y = 1\} = \Pr\{\phi_{U'} = 0, Y = 1\}$$

On en déduit que

$$\frac{\Pr\{\phi_U = 1, Y = 1\}}{\Pr\{\phi_U = 0, Y = 1\}} = \frac{\Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{\phi_{U'} = 0, Y = 1\}} \qquad (2)$$

En faisant les produit membre à membre des égalités (1) et (2) on obtient

$$\frac{\Pr\{\phi_U = 0\}}{\Pr\{\phi_U = 1\}} \frac{\Pr\{\phi_U = 1, Y = 1\}}{\Pr\{\phi_U = 0, Y = 1\}} = \frac{\Pr\{\phi_{U'} = 0\}}{\Pr\{\phi_{U'} = 1\}} \frac{\Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{\phi_{U'} = 0, Y = 1\}}$$

Il en résulte que

$$RR(U, Y) = RR(U', Y)$$

Réciproquement : supposons que 7. soit vrai alors $RR(U, Y) - RR(U', Y) = 0$. Donc

$$\frac{VPP(U, Y)}{1 - VPN(U, Y)} - \frac{VPP(U', Y)}{1 - VPN(U', Y)} = 0$$

d'où $VPP(U, Y) = VPP(U', Y)$.
On a donc $VPP(U, Y) = VPP(U', Y)$ et $\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$. Il en résulte que

$$\Pr\{\phi_U = 1\} = \Pr\{\phi_{U'} = 1\}$$

□

Proposition 4. Soient $U = \left(m_h^{X_l}\right)_{l \in L}$ et $U' = \left(m_h^{X_j}\right)_{j \in J}$ deux profils tels que U' soit emboîté dans U. Si $\Pr\left\{\phi(X,U) = 1, Y = 1\right\} = \Pr\left\{\phi(X,U') = 1, Y = 1\right\}$ alors

1. $VPP(U,Y) \leq VPP(U',Y)$
2. $VPN(U,Y) \leq VPN(U',Y)$
3. $RVP(U,Y) \leq RVP(U',Y)$
4. $RVN(U,Y) \geq RVN(U',Y)$
5. $Err(U,Y) \geq Err(U',Y)$
6. $RR(U,Y) \leq RR(U',Y)$

Preuve. Pour simplifier les expressions, on note par $\phi(X,U)$ par ϕ_U et $\phi(X,U')$ par $\phi_{U'}$.
1) Montrons que $VPP(U,Y) \leq VPP(U',Y)$
On a U' emboîté dans U entraîne que $\{\phi_U = 1\} \supset \{\phi_{U'} = 1\}$. Donc

$$\frac{1}{\Pr\{\phi_U = 1\}} \leq \frac{1}{\Pr\{\phi_{U'} = 1\}}$$

Si l'égalité $\Pr\{\phi_U = 1{,}Y = 1\} = \Pr\{\phi_{U'} = 1{,}Y = 1\}$ est vérifiée alors

$$\begin{aligned}\frac{\Pr\{\phi_U = 1{,}Y = 1\}}{\Pr\{\phi_U = 1\}} &= \frac{\Pr\{\phi_{U'} = 1{,}Y = 1\}}{\Pr\{\phi_U = 1\}} \\ &\leq \frac{\Pr\{\phi_{U'} = 1{,}Y = 1\}}{\Pr\{\phi_{U'} = 1\}} \qquad (1)\end{aligned}$$

On obtient donc

$$VPP(U,Y) \leq VPP(U',Y)$$

2) Montrons que $VPN(U,Y) \leq VPN(U',Y)$
On a U' emboîté dans U entraîne que $\{\phi_U = 1\} \supset \{\phi_{U'} = 1\}$. Donc

$$\frac{1}{\Pr\{\phi_{U'} = 0\}} \leq \frac{1}{\Pr\{\phi_U = 0\}}$$

Par ailleurs si on a $\Pr\{\phi_U = 1{,}Y = 1\} = \Pr\{\phi_{U'} = 1{,}Y = 1\}$ alors

$$\Pr\{\phi_U = 0{,}Y = 1\} = \Pr\{\phi_{U'} = 0{,}Y = 1\}$$

On en déduit que

$$\Pr\{\phi_U = 0\} - \Pr\{\phi_U = 0{,}Y = 0\} = \Pr\{\phi_{U'} = 0\} - \Pr\{\phi_{U'} = 0{,}Y = 0\}$$

donc

$$\begin{aligned}1 - \frac{\Pr\{\phi_U = 0{,}Y = 0\}}{\Pr\{\phi_U = 0\}} &\geq 1 - \frac{\Pr\{\phi_{U'} = 0{,}Y = 0\}}{\Pr\{\phi_{U'} = 0\}} \\ \frac{\Pr\{\phi_U = 0{,}Y = 0\}}{\Pr\{\phi_U = 0\}} &\leq \frac{\Pr\{\phi_{U'} = 0{,}Y = 0\}}{\Pr\{\phi_{U'} = 0\}}\end{aligned}$$

On obtient donc

$$VPN(U,Y) \leq VPV(U',Y)$$

3) Montrons que $RVP(U,Y) = RVP(U',Y)$
Par définition on a

$$\begin{aligned} RVP(U,Y) &= \frac{1-\Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{\phi_U=1,Y=1\}}{\Pr\{Y=0\}-\Pr\{\phi_U=0,Y=0\}} \\ &= \frac{1-\Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{\phi_U=1,Y=1\}}{\Pr\{\phi_U=1,Y=0\}} \\ &= \frac{1-\Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{\phi_U=1,Y=1\}}{\Pr\{\phi_U=1\}-\Pr\{\phi_U=1,Y=1\}} \end{aligned}$$

donc si $\Pr\{\phi_U=1,Y=1\} = \Pr\{\phi_{U'}=1,Y=1\}$ et U' emboîté dans U alors

$$\Pr\{\phi_U=1\}-\Pr\{\phi_U=1,Y=1\} \geq \Pr\{\phi_{U'}=1\}-\Pr\{\phi_{U'}=1,Y=1\}$$

donc

$$RVP(U,Y) \leq \frac{1-\Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{\phi_{U'}=1,Y=1\}}{\Pr\{\phi_{U'}=1\}-\Pr\{\phi_{U'}=1,Y=1\}}$$

d'où

$$RVP(U,Y) \leq RVP(U',Y)$$

4) Montrons que $RVN(U,Y) \geq RVN(U',Y)$
Par définition on a

$$RVN(U,Y) = \frac{1-\Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{Y=1\}-\Pr\{\phi_U=1,Y=1\}}{\Pr\{\phi_U=0,Y=0\}}$$

par hypothèse $\Pr\{\phi_U=1,Y=1\} = \Pr\{\phi_{U'}=1,Y=1\}$ on a alors

$$RVN(U,Y) = \frac{1-\Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{Y=1\}-\Pr\{\phi_{U'}=1,Y=1\}}{\Pr\{\phi_U=0,Y=0\}}$$

Par ailleurs U' emboîté dans U entraîne que $\Pr\{\phi_U=0,Y=0\} \leq \Pr\{\phi_{U'}=0,Y=0\}$. On en déduit que

$$RVN(U,Y) \geq \frac{1-\Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{Y=1\}-\Pr\{\phi_{U'}=1,Y=1\}}{\Pr\{\phi_{U'}=0,Y=0\}}$$

d'où

$$RVN(U,Y) \geq RVN(U',Y)$$

5) Montrons que $Err(U,Y) \geq Err(U',Y)$
Par définition on a

$$Err(U,Y) = \Pr\{Y=1\} + \Pr\{\phi_U=1\} - 2\Pr\{\phi_U=1,Y=1\}$$

$$Err(U',Y) = \Pr\{Y=1\} + \Pr\{\phi_{U'}=1\} - 2\Pr\{\phi_{U'}=1,Y=1\}$$

Par hypothèse on a $\Pr\{\phi_U=1,Y=1\} = \Pr\{\phi_{U'}=1,Y=1\}$ donc

$$Err(U,Y) - Err(U',Y) = \Pr\{\phi_U=1\} - \Pr\{\phi_{U'}=1\}$$

et puisque U' est emboité dans U alors $\Pr\{\phi_U = 1\} \geq \Pr\{\phi_{U'} = 1\}$. On obtient donc

$$Err(U,Y) \geq Err(U',Y)$$

6) Montrons que $RR(U,Y) \leq RR(U',Y)$
On a

$$\Pr\{\phi_U = 0, Y = 1\} = \Pr\{Y = 1\} - \Pr\{\phi_U = 1, Y = 1\}$$

Puisqu'on a $\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_{U'} = 1, Y = 1\}$, on obtient alors implique aussi

$$\Pr\{\phi_U = 0, Y = 1\} = \Pr\{Y = 1\} - \Pr\{\phi_{U'} = 1, Y = 1\}$$

D'où

$$\Pr\{\phi_U = 0, Y = 1\} = \Pr\{\phi_{U'} = 0, Y = 1\}$$

Puisque que U' est emboîté dans U, on en déduit que

$$\frac{\Pr\{\phi_U = 0, Y = 1\}}{\Pr\{\phi_U = 0\}} \geq \frac{\Pr\{\phi_{U'} = 0, Y = 1\}}{\Pr\{\phi_{U'} = 0\}} \qquad (2)$$

Si on fait le rapport membre à membre des inégalités (1) et (2), il en résulte que

$$RR(U,Y) \leq RR(U',Y)$$

□

Il découle de la proposition 4 que lorsque les fonctions de classification générées par deux profils emboîtés ont la même sensibilité et des spécificités différentes alors la fonction de classification générée par le profil le plus long a une erreur de classement plus faible, une valeur prédictive positive (confiance) plus élevée, un rapport de vraisemblance positif plus élevé, un rapport de vraisemblance négatif plus faible et un risque relatif plus élevé que celui de la fonction de classification générée par le profil le plus court. De plus U' emboîté dans U implique que la fonction de classification générée par U' a une spécificité plus élevée que celle de la fonction de classification générée par U. Par conséquent on préférera le profil le plus long puisque ses indicateurs de performance (sensibilité, spécificité et erreur de classement) sont meilleurs.

Proposition 5. Soient $U = \left(m_h^{X_l}\right)_{l \in L}$ et $U' = \left(m_h^{X_j}\right)_{j \in J}$ deux profils tels que U' soit emboîté dans U. Si $\Pr\{\phi(X,U) = 0, Y = 0\} = \Pr\{\phi(X,U') = 0, Y = 0\}$ alors

1. $VPP(U,Y) \geq VPP(U',Y)$
2. $VPN(U,Y) \geq VPN(U',Y)$
3. $RVP(U,Y) \geq RVP(U',Y)$
4. $RVN(U,Y) \leq RVN(U',Y)$
5. $Err(U,Y) \leq Err(U',Y)$
6. $RR(U,Y) \geq RR(U',Y)$

Preuve. 1) Montrons que $VPP(U,Y) \geq VPP(U',Y)$

Par définition

$$\frac{\Pr\{\phi_U = 1,Y = 1\}}{\Pr\{\phi_U = 1\}} = \frac{\Pr\{\phi_U = 1,Y = 1\}}{\Pr\{\phi_U = 1,Y = 1\} + \Pr\{\phi_U = 1,Y = 0\}}$$

et

$$\frac{\Pr\{\phi_{U'} = 1,Y = 1\}}{\Pr\{\phi_{U'} = 1\}} = \frac{\Pr\{\phi_{U'} = 1,Y = 1\}}{\Pr\{\phi_{U'} = 1,Y = 1\} + \Pr\{\phi_{U'} = 1,Y = 0\}}$$

Comme $\Pr\{\phi_U = 0,Y = 0\} = \Pr\{\phi_{U'} = 0,Y = 0\}$, on sait que $\Pr\{\phi_U = 1,Y = 0\} = \Pr\{\phi_{U'} = 1,Y = 0\}$

et en plus si a, b, c sont des réels positifs et $a \geq c$ on a $\frac{a}{a+b} \geq \frac{c}{c+b}$. On peut déduire de ces deux conditions que

$$\frac{\Pr\{\phi_U = 1,Y = 1\}}{\Pr\{\phi_U = 1\}} \geq \frac{\Pr\{\phi_{U'} = 1,Y = 1\}}{\Pr\{\phi_{U'} = 1\}} \qquad (1)$$

On obtient donc

$$VPP(U,Y) \geq VPP(U',Y)$$

2) Montrons que $VPN(U,Y) \geq VPN(U',Y)$

Par définition

$$\frac{\Pr\{\phi_U = 0,Y = 0\}}{\Pr\{\phi_U = 0\}} = \frac{\Pr\{\phi_U = 0,Y = 0\}}{\Pr\{\phi_U = 0,Y = 1\} + \Pr\{\phi_U = 0,Y = 0\}}$$

et

$$\frac{\Pr\{\phi_{U'} = 0,Y = 0\}}{\Pr\{\phi_{U'} = 0\}} = \frac{\Pr\{\phi_{U'} = 0,Y = 0\}}{\Pr\{\phi_{U'} = 0,Y = 1\} + \Pr\{\phi_{U'} = 0,Y = 0\}}$$

Puisque U' est emboîté dans U alors $\Pr\{\phi_{U'} = 0,Y = 1\} \geq \Pr\{\phi_U = 0,Y = 1\}$.

d'où

$$\frac{\Pr\{\phi_U = 0,Y = 0\}}{\Pr\{\phi_U = 0\}} \geq \frac{\Pr\{\phi_{U'} = 0,Y = 0\}}{\Pr\{\phi_{U'} = 0\}}$$

puisque $\Pr\{\phi_{U'} = 0,Y = 0\} = \Pr\{\phi_U = 0,Y = 0\}$ donc

$$VPN(U,Y) \geq VPN(U',Y)$$

3) Montrons que $RVP(U,Y) \geq RVP(U',Y)$

Par définition on a

$$RVP(U,Y) = \frac{1 - \Pr\{Y = 1\}}{\Pr\{Y = 1\}} \frac{\Pr\{\phi_U = 1, Y = 1\}}{\Pr\{Y = 0\} - \Pr\{\phi_U = 0, Y = 0\}}$$

et

$$RVP(U',Y) = \frac{1 - \Pr\{Y = 1\}}{\Pr\{Y = 1\}} \frac{\Pr\{\phi_{U'} = 1, Y = 1\}}{\Pr\{Y = 0\} - \Pr\{\phi_{U'} = 0, Y = 0\}}$$

par hypothèse on $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$ donc le signe de $RVP(U,Y) - RVP(U',Y)$ dépend du signe $\Pr\{\phi_{U1} = 1, Y = 1\} - \Pr\{\phi_{U'} = 1, Y = 1\}$

or on a le profil U' emboîte dans le profil U. Ceci entraîne que

$$\Pr\{\phi_{U1} = 1, Y = 1\} - \Pr\{\phi_{U'} = 1, Y = 1\} \geq 0$$

d'où

$$RVP(U,Y) \geq RVP(U',Y)$$

4) Montrons que $RVN(U,Y) \leq RVN(U',Y)$

Par définition

$$RVN(U,Y) = \frac{1 - \Pr\{Y = 1\}}{\Pr\{Y = 1\}} \frac{\Pr\{Y = 1\} - \Pr\{\phi_U = 1, Y = 1\}}{\Pr\{\phi_U = 0, Y = 0\}}$$

et

$$RVN(U',Y) = \frac{1 - \Pr\{Y=1\}}{\Pr\{Y=1\}} \frac{\Pr\{Y=1\} - \Pr\{\phi_{U'}=1, Y=1\}}{\Pr\{\phi_{U'}=0, Y=0\}}$$

par hypothèse on $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$ donc le signe de $RVN(U,Y) - RVN(U',Y)$ dépend du signe $\Pr\{\phi_{U'} = 1, Y = 1\} - \Pr\{\phi_U = 1, Y = 1\}$
or on a le profil U' emboîte dans le profil U. Ceci entraîne que

$$\Pr\{\phi_{U'} = 1, Y = 1\} - \Pr\{\phi_U = 1, Y = 1\} \leq 0$$

d'où

$$RVN(U,Y) \leq RVP(U,Y)$$

5) Montrons que $Err(U,Y) \leq Err(U',Y)$
On a

$$\Pr\{\phi_U = 0, Y = 0\} = 1 - \Pr\{Y=1\} - \Pr\{\phi_U = 1\} + \Pr\{\phi_U = 1, Y = 1\}$$

$$\Pr\{\phi_U = 1, Y = 1\} = \Pr\{\phi_U = 0, Y = 0\} + \Pr\{\phi_U = 1\} + \Pr\{Y=1\} - 1$$

si on remplace $\Pr\{\phi_U = 1, Y = 1\}$ par son expression dans $Err(U,Y)$, on obtient

$$Err(U,Y) = -2\Pr\{\phi_U = 0, Y = 0\} - \Pr\{\phi_U = 1\} - \Pr\{Y=1\} - 2$$

de même on a

$$Err(U',Y) = -2\Pr\{\phi_{U'} = 0, Y = 0\} - \Pr\{\phi_{U'} = 1\} - \Pr\{Y=1\} - 2$$

et puisque on a par hypothèse que $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$ alors

$$Err(U,Y) - Err(U',Y) = -\Pr\{\phi_U = 1\} + \Pr\{\phi_{U'} = 1\}$$

par ailleurs $-\Pr\{\phi_U = 1\} + \Pr\{\phi_{U'} = 1\} \leq 0$ puisque U' est emboité dans U. d'où

$$Err(U,Y) - Err(U',Y) \leq 0$$

6) Montrons que $RR(U,Y) \geq RR(U',Y)$
On a

$$\frac{\Pr\{\phi_U = 0, Y = 1\}}{\Pr\{\phi_U = 0\}} = \frac{\Pr\{\phi_U = 0\} - \Pr\{\phi_U = 0, Y = 0\}}{\Pr\{\phi_U = 0\}}$$

et

$$\frac{\Pr\{\phi_{U'} = 0, Y = 1\}}{\Pr\{\phi_{U'} = 0\}} = \frac{\Pr\{\phi_{U'} = 0\} - \Pr\{\phi_{U'} = 0, Y = 0\}}{\Pr\{\phi_{U'} = 0\}}$$

en tenant compte que $\Pr\{\phi_U = 0, Y = 0\} = \Pr\{\phi_{U'} = 0, Y = 0\}$ et $\Pr\{\phi_U = 0\} \leq \Pr\{\phi_{U'} = 0\}$, on a

$$\frac{\Pr\{\phi_U = 0, Y = 0\}}{\Pr\{\phi_U = 0\}} \geq \frac{\Pr\{\phi_{U'} = 0, Y = 0\}}{\Pr\{\phi_{U'} = 0\}}$$

et il s'en suit que

$$\frac{\Pr\{\phi_U = 0, Y = 1\}}{\Pr\{\phi_U = 0\}} \leq \frac{\Pr\{\phi_{U'} = 0, Y = 1\}}{\Pr\{\phi_{U'} = 0\}}$$

d'où

$$\frac{\Pr\{\phi_U = 0\}}{\Pr\{\phi_U = 0, Y = 1\}} \geq \frac{\Pr\{\phi_{U'} = 0\}}{\Pr\{\phi_{U'} = 0, Y = 1\}} \qquad (2)$$

en faisant le produit membre à membre des inégalités (1) et (2) on obtient que $\mathrm{RR}(U',Y) \leq \mathrm{RR}(U,Y)$

□

Il résulte de la proposition 5 que si on a deux profils U et U' emboîtés tels que les fonctions de classification qui leurs sont associées ont des spécificités égales alors non seulement la sensibilité de la fonction de classification générée par U est plus élevée à cause de l'emboîtement mais aussi son erreur de classement est plus faible, sa valeur prédictive positive (confiance) est plus forte, son rapport de vraisemblance positif est plus élevé, son rapport de vraisemblance négatif est plus faible et son risque relatif est plus élevé que ceux de la fonction de classification générée par U'. On peut élaguer le profil U' qui est de plus grande taille. Cette proposition a été utilisée par Jiuyong Li et al [4] en premier en se basant sur la propriété anti-monotone du support.

3 Règles d'association binaires et classifieur associé à un profil

3.1 Règle d'association

Définition 4. Considérons $U = \left(m_h^{X_l}\right)_{l\in L}$ et $U' = \left(m_h^{X_j}\right)_{j\in J}$ deux profils disjoints. Une règle d'association est l'expression d'une implication de la forme $U \to U'$ signifiant que les probabilités $\Pr\left\{\left[\prod_{k\in L\cup J} \mathbb{1}\left(X_k = m_h^{X_k}\right) = 1\right]\right\}$ et $\Pr\left\{\left[\prod_{j\in J} \mathbb{1}\left(X_j = m_h^{X_j}\right) = 1\right] \mid \left[\prod_{l\in L} \mathbb{1}\left(X_l = m_h^{X_l}\right) = 1\right]\right\}$ sont significatives (supérieurs aux seuil s_0 et c_0 respectivement). On appelle U l'antécédent de la règle et U' la conséquence de la règle.

Une règle d'association $U \to U'$ exprime le fait que non seulement il y a une forte probabilité que les événements $\left[\prod_{j\in J} \mathbb{1}\left(X_j = m_h^{X_j}\right) = 1\right]$ et $\left[\prod_{l\in L} \mathbb{1}\left(X_l = m_h^{X_l}\right) = 1\right]$ aient lieu simultanément mais aussi que l'événement $\left[\prod_{j\in J} \mathbb{1}\left(X_j = m_h^{X_j}\right) = 1\right]$ ait une forte probabilité d'occurrence conditionnellement à l'événement $\left[\prod_{l\in L} \mathbb{1}\left(X_l = m_h^{X_l}\right) = 1\right]$.

Définition 5. Considérons une règle d'association $U \to U'$ où $U = \left(m_h^{X_l}\right)_{l\in L}$ et $U = \left(m_h^{X_j}\right)_{j\in J}$. La probabilité $\Pr\left\{\left[\prod_{k\in L\cup J} \mathbb{1}\left(X_k = m_h^{X_k}\right) = 1\right]\right\}$ est appelé le support de la règle d'association et la probabilité conditionnelle $\Pr\left\{\left[\prod_{j\in J} \mathbb{1}\left(X_j = m_h^{X_j}\right) = 1\right] \mid \left[\prod_{l\in L} \mathbb{1}\left(X_l = m_h^{X_l}\right) = 1\right]\right\}$ est sa confiance.

Il apparaît que le classifieur associé à un profil est une implication de la forme $[\phi(X,U) = 1] \to [Y = 1]$ dès lors qu'on exige que $\Pr(\phi(X,U) = 1, Y = 1) > s_0$ et $\Pr(Y = 1|\phi(X,U) = 1) > c_0$. Une telle règle d'association est dite binaire.

3.2 Classifieur basé sur un ensemble de profils

Dans un apprentissage statistique par règles d'association binaires, l'apprentissage automatique se résume en deux étapes. La première consiste à générer l'ensemble des profils $\mathcal{U}_\lambda$ défini par :

$$\mathcal{U}_\lambda = \left\{ U = \left(m_h^{X_j} \right)_{j \in J} ; \Pr(Y = 1, \phi(X, U) = 1) > s_0, \Pr(Y = 1 | \phi(X, U) = 1) > c_0 \right\}$$

où $\lambda = (s_0, c_0)$ est le paramètre qui spécifie l'ensemble $\mathcal{U}_\lambda$.
Le paramètre c_0 représente le seuil de confiance minimum et le paramètre s_0 représente le seuil de support minimum. Dans la pratique, on pourra étendre le paramètre λ en ajoutant le paramètre r_0 représentant le seuil de risque relatif minimum et le paramètre l_0 représentant la longueur ou taille maximale d'un profil.
La deuxième étape consiste à implémenter l'ensemble des fonctions indicatrices G_λ défini par :

$$G_\lambda = \{ \, \phi(X, U); \, U \in \mathcal{U}_\lambda \, \}$$

Lorsque la probabilité de la classe d'intérêt tend vers zero, la sensibilité du classifieur associé à un profil U (i.e., $\phi(X, U)$) peut être faible. En considérant un ensemble de profils, on peut espérer aboutir à un classifieur avec une meilleure sensibilité sans trop détériorer le niveau de spécificité. Etant donné un ensemble de profils G_λ pour un λ fixé, la fonction

$$\phi(X, \lambda, k) = \mathbb{1} \left(\sum_{U \in \mathcal{U}_\lambda} \phi(X, U) > k \right) \qquad k \in \{1, \cdots, |\mathcal{U}_\lambda|\}$$

définit également un classifieur.

4 Conclusion

L'objectif de cette analyse est de défendre une méthodologie permettant de mettre en place une fonction de classement binaire lorsqu'il s'agit d'une tâche de classification supervisée où la classe cible est un événement rare. Cet objectif est atteint par le recours à des règles d'association pour explorer les données afin d'identifier les profils qui sont corrélés avec la classe cible. Des profils pertinents sont sélectionnés sur la base de leurs sensibilités et spécificités, de leurs valeurs prédictives positives ou négatives, de leurs rapports de vraisemblance positifs ou négatifs et de leurs risques relatifs pour constituer un ensemble optimal de profils.
Dans la suite, nous allons mettre en place un algorithme d'apprentissage statistique pour établir une règle de classement (classifieur) basé sur un ensemble optimal de profils lorsque : (1) nous disposons d'un ensemble d'observations indépendantes et identiquement distribuées ; (2) les observations ne sont pas indépendantes et identiquement distribuées.

Bibliographie

[1] Agrawal, R., and Srikant, R. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases* (San Francisco, CA, USA, 1994), VLDB '94, Morgan Kaufmann Publishers Inc., pp. 487–499. 18

[2] Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett. 27*, 8 (2006), 861–874. 16

[3] Lenca, P., Meyer, P., Vaillant, B., and Lallich, S. On selecting interestingness measures for association rules : User oriented description and multiple criteria decision aid. *European Journal of Operational Research 184*, 2 (2008), 610–626. 18

[4] Li, J., Fu, A. W.-c., and Fahey, P. Efficient discovery of risk patterns in medical data. *Artificial intelligence in medicine 45*, 1 (2009), 77–89. 17, 18, 31

[5] Li, J., Fu, A. W.-c., He, H., Chen, J., Jin, H., McAullay, D., Williams, G., Sparks, R., and Kelman, C. Mining risk patterns in medical data. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining* (New York, NY, USA, 2005), KDD '05, ACM, pp. 770–775. 18

[6] Li, W., Han, J., and Pei, J. CMAR : accurate and efficient classification based on multiple class-association rules. In *ICDM 2001, Proceedings IEEE International Conference on Data Mining, 2001* (2001), pp. 369–376. 13

[7] Liu, B., Hsu, W., and Ma, Y. Integrating classification and association rule mining. pp. 80–86. 13

[8] Liu, B., Ma, Y., and Wong, C.-k. Classification using association rules : Weaknesses and enhancements. *In Grossman, R. L., et al (eds), Data Mining for Scientific and Engeneering Applications. Kluwer Academic Publishers* (2001), 591–601. 13

[9] Ohsaki, M., Kitaguchi, S., Okamoto, K., Yokoi, H., and Yamaguchi, T. Evaluation of rule interestingness measures with a clinical dataset on hepatitis. In *Knowledge Discovery in Databases : PKDD 2004*, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds., no. 3202 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 362–373. 18

[10] Pasquier, N., Bastide, Y., Taouil, R., and Lakhal, L. Discovering frequent closed itemsets for association rules. In *Proceedings of the 7th International Conference on Database Theory* (1999), Springer-Verlag, pp. 398–416. 18

[11] Piatetsky-Shapiro, G. Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*. pp. 229–248. 18

[12] Tan, P.-N., Kumar, V., and Srivastava, J. Selecting the right objective measure for association analysis. *Inf. Syst. 29*, 4 (2004), 293–313. 18