

Sujets de stage Master 2 Informatique - Intelligence artificielle

Apprentissage et comparaison de réseaux bayésiens pour l'évaluation du risque de chute d'une personne âgée

Ce sujet est une contribution à un système de prévention des chutes pour les personnes âgées. L'objectif est de fournir un outil permettant d'évaluer le risque de chute d'une personne âgée à partir d'observations partielles.

Le travail repose sur des données réelles issues du CHRU de Lille, concernant plus d'un millier de personnes, et regroupant diverses observations (417 caractéristiques dont certaines sont redondantes, obsolètes ou écartées). La prise en main du sujet sera facilitée par une ontologie des facteurs de risque de chute qui sert de référence en termes de sélection des variables.

Les modèles utilisés sont des modèles graphiques probabilistes (réseaux bayésiens¹) bien adaptés pour encapsuler la distribution naturelle des paramètres dans une population et pour raisonner à partir de données partielles.

Les trois phases décrites ci-dessous seront développées en boucle de façon incrémentale : le prétraitement des données, l'apprentissage des classifieurs, les tests et les comparaisons.

1^{ère} phase : prétraitement des données et définition de variables de synthèse

A partir des données initiales, il s'agit de générer des fichiers de données utilisables pour l'apprentissage de réseau bayésien, de façon à ce que l'on puisse :

- sélectionner les caractéristiques et la classe (nombre de chutes dans les 6 mois précédents, nombre de chutes dans les 6 mois suivant la 1^{ère} consultation, niveau de risque de chute)
- appliquer des filtres pour ne conserver qu'une partie des cas
- gérer les données manquantes (divers choix)
- gérer le domaine des variables, en particulier les variables continues (discrétisation)
- générer le descriptif des traitements appliqués au fichier de données en sortie et les informations sur le fichier généré (nb lignes, nb colonnes, ...)

Un objectif supplémentaire vise à définir de nouvelles variables qui condensent l'information de plusieurs caractéristiques existantes. Pour cela, des techniques de réduction de dimension telle que l'analyse en composante principale (ACP) (ou autre²) seront utilisées.

2^{ème} phase : définition et apprentissage des réseaux bayésiens

Il s'agit d'appliquer des algorithmes d'apprentissage à partir des fichiers de données pour définir différents réseaux bayésiens permettant d'évaluer le risque de chute.

Voici les principaux classifieurs à étudier :

- réseau bayésien naïf (différents choix de classe et de nombres d'observation en entrée)
- réseau bayésien de type TAN (Tree Augmented Network)

¹ https://en.wikipedia.org/wiki/Bayesian_network

Réseaux bayésiens, 2007, Eyrolles, de Patrick Naïm, Pierre-Henri Wuillemin, Philippe Leray, Olivier Pourret, Anna Becker

² vidéo Apprentissage statistique et analyse prédictive en python

<https://www.youtube.com/watch?v=ZD5JJGq1rvQ>

- réseau bayésien causal, avec ajout de nœuds intermédiaires pour réduire la taille du nœud central, définis par exemple par analyse en composante principale (ACP) (ou autre³)

La méthode de la validation croisée sera utilisée pour exploiter les données pour l'apprentissage et les tests de chaque réseau bayésien, et générer la matrice de confusion.

3^{ème} phase : Comparaison et analyse des résultats

L'objectif est d'identifier quelle solution permet la meilleure évaluation du risque de chute même dans le cas où la description de la personne considérée en entrée est très partielle.

Outil de développement :

L'application sera développée en python et à l'aide des notebook (Anaconda, Jupyter lab)
Les réseaux bayésiens sont définis à l'aide des fonctions de la bibliothèque pyAgrum⁴.

Cadre : projet ELSAT2020 (Ecomobilité Logistique Sécurité et Adaptabilité dans les Transports à l'Horizon 2020), dans l'objectif stratégique n°1, L'humain dans les transports et sa mobilité, projet 2 : Mobilité et sécurité des Personnes à Mobilité Réduite.

Lieu : LAMIH - UMR CNRS 8201,
Université polytechnique des Hauts de France – UPHF
Le mont Houy, 59 313 Valenciennes

Durée : 5-6 mois

Gratification : taux légal en vigueur

Compétence requises :

bon niveau de programmation,
notions sur la classification appréciée
autonomie et rigueur

Candidature : envoyer un CV, un relevé de notes et une lettre de motivation à
veronique.delcroix@uphf.fr,
tel : 03 27 51 14 65 ou 06 03 62 61 53

³ vidéo Apprentissage statistique et analyse prédictive en python
<https://www.youtube.com/watch?v=ZD5lGq1rvQ>

⁴ <http://agrum.gitlab.io/>
<http://www-desir.lip6.fr/~phw/aGrUM/docs/last/pyAgrum/>