APPORT DES CLASSIFIEURS POUR L'ANALYSE DES MOUVEMENTS OCULAIRES EN SITUATION DYNAMIQUE

Sophie **LEMONNIER**[a,b,], Abderrahmane **BOUBEZOUL**[a], Roland **BREMOND**[a] & Thierry **BACCINO**[b]
[a] *IFSTTAR, LEPSIS, Champs sur Marne, France*
[b] *Université Paris VIII, CHART, LUTIN, Paris, France*

INTRODUCTION

La nature du lien entre le mouvement des yeux et la tâche est un problème ouvert (Yarbus, 1967). Les méthodes statistiques de classification sont utilisées depuis plus de vingt ans dans de nombreux domaines, mais elles n'ont été introduites pour l'analyse des mouvements oculaires par Greene *et al.* qu'en 2012. Lors d'une expérience, ils proposaient différentes tâches pour un même stimulus. Ils ont obtenu des scanpaths, puis ils ont utilisé un classifieur pour prédire la tâche à partir des mouvements oculaires (avec un résultat négatif). L'an dernier, Henderson *et al.* (2013) ont repris la même question avec un protocole et des tâches différentes. Leur classifieur a réussi à distinguer des tâches de recherche visuelle et de mémorisation. Suite à ces travaux, nous nous sommes demandé s'il était possible d'utiliser un classifieur pour prédire une tâche à partir du mouvement des yeux, mais dans une situation dynamique.

CLASSIFICATION SUPERVISEE

Pour alimenter un classifieur, un « label » (ici, la tâche, par exemple mémorisation ou recherche visuelle) est associé à chaque « donnée » (une donnée étant l'ensemble des variables dépendantes (VD) associées à un essai, par exemple, l'amplitude des saccades et la durée des fixations). La démarche comporte ensuite deux étapes :

- Le classifieur doit d'abord « apprendre ». A partir d'un ensemble de données labellisées, il va proposer un modèle, c'est-à-dire un opérateur capable de classer toute nouvelle donnée dans l'une des classes (chaque classe correspond à un label, tel que mémorisation ou recherche visuelle).
- La seconde étape consiste à tester ce modèle sur de nouvelles données (sans fournir le label), et à évaluer la qualité des prédictions. Avec deux classes par exemple, une classification au hasard donnera 50% de succès, et donc un bon classifieur doit prédire le label avec un taux supérieur à 50%. Une matrice de confusion permet de comparer les labels prédits et ceux attendus.

MATERIEL & RESULTATS

Notre situation expérimentale était l'approche d'un carrefour (sur simulateur de conduite) que les participants devaient traverser (Cf. Fig. 1). La question concernait la présence ou l'absence d'une tâche de prise de décision. Le panneau de signalisation pouvait être un stop, une priorité ou un cédez-le-passage. En stop et prioritaire, les conducteurs n'ont pas à prendre de décision puisque celle-ci découle du panneau (la décision elle-même peut être Go ou Stop). En cédez-le-passage, c'est au conducteur de décider s'il s'arrête ou non.

Les VD suivantes ont été recueillies : nombre et durée moyenne des fixations, nombre et amplitude moyenne des saccades, pentes (évolution temporelle) de ces données (fixations et saccades), durée et nombre de clignements, diamètre pupillaire. A chaque donnée était associé le label *avec* ou *sans* prise de décision, en fonction de la condition (cédez le passage, ou stop et prioritaire). Il y avait donc deux classes (Cf. Fig. 2).

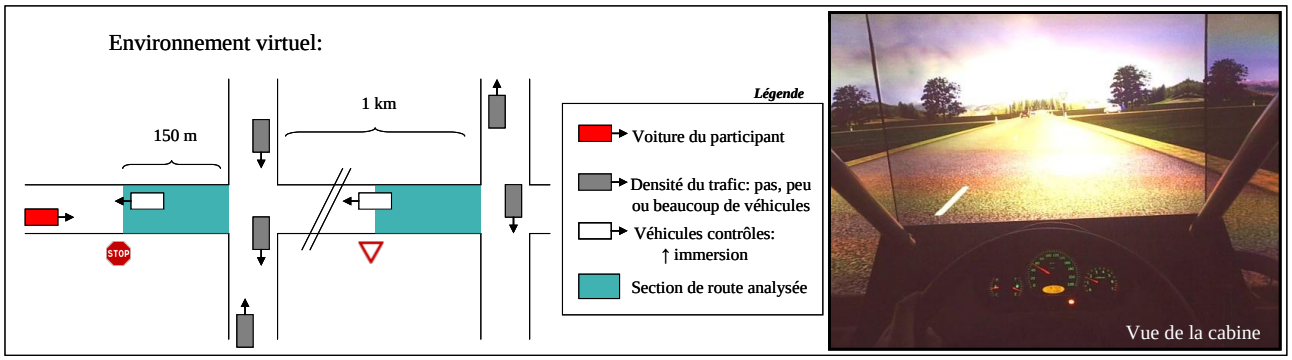

Figure 1 : Le design expérimental avec à gauche, une vue de dessus schématique de l'environnement virtuel, et à droite, une image issue de la caméra de scène de l'oculomètre lors d'une approche de carrefour.

Il existe différents types de classifieurs, basés sur des théories statistiques différentes. Il est possible que certains classifieurs soient plus appropriés que d'autres pour les données oculaires. Pour examiner cette hypothèse, nous en avons utilisé trois de manière à comparer leurs résultats : un *Support Vector Machines* (SVM ; Séparateurs à Vaste Marge ; Vapnik, 1995), un *Random Forest* (RF, forêt d'arbre décisionnel ; Breiman, 2001) et un *k-Nearest Neighbors* (k-NN, k plus proches voisins ; Cover & Hart, 1967).

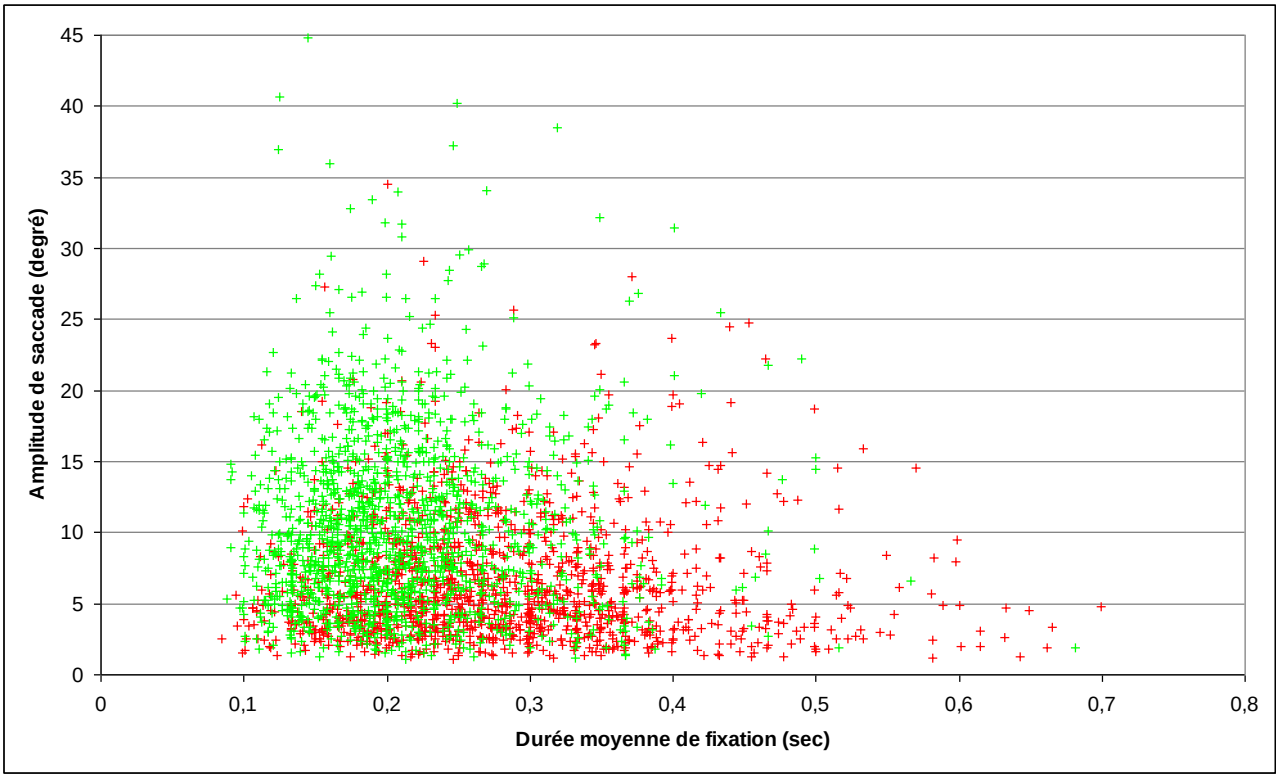

Figure 2 : Représentation des données dans un espace à deux dimensions correspondant à l'amplitude des saccades et la durée de fixation (ces deux variables faisant partie de celles qui ont le plus contribuées aux prédictions du classifieur RF). La couleur représente les deux classes, avec la tâche de prise de décision présente (vert) ou absente (rouge).

Pour le SVM, dans le cas où les données sont non-linéairement séparables comme ici, on utilise une fonction de décision non linéaire. L'idée sous-jacente est de tenter de trouver une projection (mapping) de l'espace d'entrée vers un autre espace de représentation (feature space) dans lequel les données sont linéairement séparables. Cela se nomme l'extension. Le SVM utilise alors un noyau (kernel) et divise l'espace des

données avec un hyperplan séparateur. Dans cette étude, nous avons utilisé un noyau Gaussien type RBF (Radial Basis Function). Après avoir normalisé les données, deux paramètres devaient ensuite être choisis afin de définir le noyau. Le premier - le gamma - définit la largeur de la gaussienne selon l'équation suivante : $\Gamma = 1/\sigma^2$. Le second - la constante C - est souvent appelée la constante de compromis (tradeoff). Elle permet d'indiquer l'importance que l'on accorde aux erreurs commises sur l'ensemble d'apprentissage par rapport au fait de maximiser la marge. Elle peut être interprétée comme une tolérance au bruit du classifieur. Pour ces analyses, les valeurs de ces deux paramètres étaient respectivement de 2^{-4} et 2^5.

Le RF utilise des arbres de décision et choisit le meilleur. Il fonctionne selon le même principe que le SVM en cherchant un hyperplan séparateur. Néanmoins, le SVM est conçu pour traiter des espaces de grande dimension et prend pour cela l'ensemble des mesures ; le RF n'a pas cette spécificité, et va attribuer un poids à chacune des VD. Ce poids permet par la suite de connaître la contribution de chacune de ces mesures dans le modèle final. Le paramètre à estimer est le nombre de forêts, ici égal à 50.

Le k-NN fonctionne au plus proche voisin. Pour chaque nouvelle donnée, il va calculer sa distance de similarité avec l'ensemble des données d'apprentissage. Il va ensuite ordonner ses valeurs, choisir les plus petites, et prendre le label majoritaire qu'il va ensuite attribuer à cette nouvelle donnée. Le nombre de valeurs choisi correspond au paramètre k, préalablement défini (plus il y a du bruit dans les données, et plus le k aura une valeur petite). Dans notre cas, nous avons testé les valeurs entre 1 et 30, et c'est k = 10 qui fonctionnait au mieux. La Figure 3 illustre une projection des données dans un espace 2D, avec la frontière entre les deux classes telle qu'elle est estimée par le modèle.

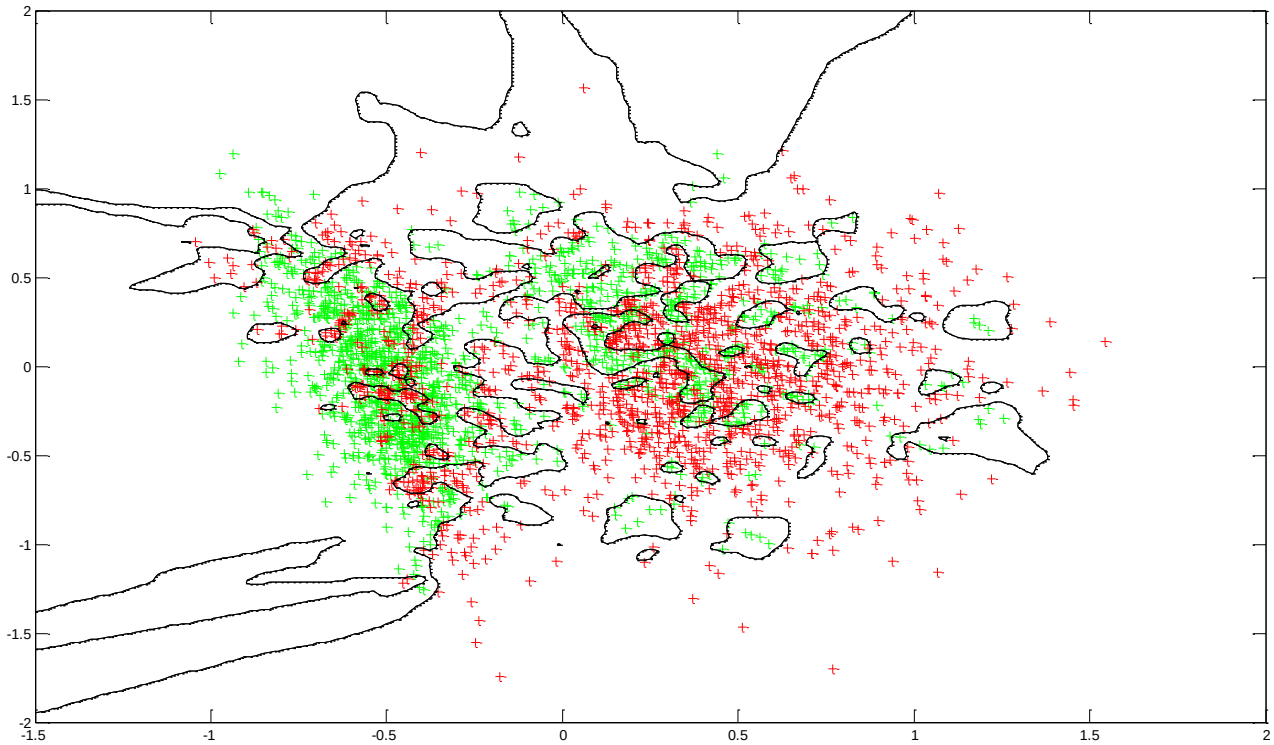

Figure 3 : Représentation schématique en deux dimensions des données de la Fig. 2 dans l'espace du classifieur k-NN. Les axes ont été choisis selon une analyse de type ACP ; dans le modèle il y a beaucoup plus de dimensions. Les lignes continues noires correspondent à la séparation faite entre les deux classes prédites par le classifieur, ici projetées dans un espace restreint à 2D (ce qui explique que certaines frontières semblent étranges). La couleur représente les deux classes, avec la tâche de prise de décision présente (vert) ou absente (rouge).

Sur nos données, la performance de prédiction du SVM était de 70% (hasard = 50%), celle du RF de 67.17% et celle du k-NN de 78.65%. De plus, il est possible avec certains classifieurs d'estimer les variables qui ont le plus contribué au résultat, de manière quantitative : le modèle fournit la contribution de chacune des variables. Pour le modèle RF utilisé, les VD qui ont le plus contribuées au résultat sont le nombre de saccades, la durée moyenne des fixations et les amplitudes de saccades.

DISCUSSION

Les résultats obtenus semblent confirmer qu'il est possible de prédire mieux que le hasard la présence d'une tâche (prise de décision en approche de carrefour) sur la base du mouvement des yeux en situation dynamique. Partant de données oculométriques qui semblaient difficilement séparables (Cf. Fig. 2), l'utilisation de classifieurs nous a permis de construire des modèles meilleurs que le hasard, que ce soit avec les méthodes SVM, RF ou k-NN. Il est cependant important de comprendre que chaque famille de classifieur a une logique spécifique, ce qui conduit à des modèles différents, et donc à des résultats différents : dans notre cas, le k-NN donne un taux de prédiction meilleur que le SVM et le RF.

BIBLIOGRAPHIE

Breiman, L. (2001). Random Forests. *Machine Learning 45* (1), 5–32.

Cover, T. M., Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory 13* (1), 21–27

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision research*, *62*, 1-8.

Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., & Olejarczyk, J. (2013). Predicting cognitive state from eye movements. *PloS one*, *8*(5), e64937.

Vapnik, V. (2000). *The nature of statistical learning theory*. springer.

Yarbus, A. L. (1967). *Eye movements and vision* (Vol. 2, No. 5.10). L. A. Rigss (Ed.). New York: Plenum press.