

Démonstration du diagnostic automatique de l'état dépressif

S. Cholet

H. Paugam-Moisy

Laboratoire de Mathématiques Informatique et Applications (LAMIA - EA 4540)

Université des Antilles, Campus de Fouillole - Guadeloupe

Stephane.Cholet@univ-antilles.fr

Résumé

Les troubles psychosociaux sont un problème de santé publique majeur, pouvant avoir des conséquences graves sur le court ou le long terme, tant sur le plan professionnel que personnel ou familial. Le diagnostic de ces troubles doit être établi par un professionnel. Toutefois, l'IA (l'Intelligence Artificielle) peut apporter une contribution en fournissant au praticien une aide au diagnostic, et au patient un suivi permanent rapide et peu coûteux. Nous proposons un outil d'aide au diagnostic automatique de l'état dépressif à partir d'observations du visage en temps réel, au moyen d'une simple webcam. A partir de vidéos du challenge AVEC'2014, nous avons entraîné un classifieur neuronal à extraire des prototypes de visages selon différentes valeurs du score de dépression de Beck (BDI-II).

1 Introduction

La démonstration est associée à l'article "Diagnostic automatique de l'état dépressif" présenté à la conférence CNIA'2018. Des visages dont les points d'intérêt sont extraits à partir de vidéos sont associés à un score mesurant l'état dépressif du sujet. Le système permettant cette association est un classifieur neuronal incrémental dont l'apprentissage a été réalisé sur les données du challenge AVEC'2014 [1]. Le présent article décrit comment ce classifieur peut être mis en œuvre pour estimer, en temps réel, l'état dépressif d'un sujet filmé par une webcam.

Le système d'estimation de l'état dépressif est composé de quatre phases qui s'enchaînent comme décrit Figure 1 : la capture de la vidéo est suivie de l'extraction des descripteurs, du traitement par le classifieur et de la présentation du diagnostic. On se focalise ici sur la description et le fonctionnement des procédés mis en œuvre pour la classification. L'attention du lecteur est attirée sur le fait que cette chaîne de traitement est encore à l'état de prototype, en phase active de développement et en attente de validation, en particulier par des experts psychiatres.

2 Chaîne des traitements

2.1 Capture vidéo

La capture vidéo est réalisée au moyen d'une webcam connectée à l'ordinateur utilisé pour la classification. Afin

d'optimiser le traitement du flux, les images sont capturées par un thread dédié et bufferisées dans une file d'attente. De plus, la taille des images est réduite à 300x300 pixels avant leur traitement. La bufferisation des images assure la continuité de la capture pendant leur traitement. La fréquence de capture est de 25 images par secondes.

2.2 Prétraitement des images

Dans la mesure où l'état dépressif s'évalue sur la durée, on ne traite que n images par seconde (malgré une fréquence de capture fixe). Ce paramètre est notamment utilisé pour optimiser la fluidité du traitement global. Chaque image fait l'objet d'une recherche de visage et d'estimation de la position des points d'intérêts via la méthode de Kazemi et Sullivan implémentée dans dLib.

L'alignement des visages consiste en une série de transformations géométriques ayant les objectifs suivants :

- que les visages aient la même taille ;
- que les visages soient centrés dans l'image ;
- que les yeux soient alignés horizontalement.

Cette phase, capitale, assure aux données l'homogénéité nécessaire pour la classification. Afin de réduire la durée de l'alignement, ce dernier est réalisé directement sur les points extraits et non sur les images

La sortie du prétraitement est, pour chaque image, un vecteur de 136 composantes comprenant les abscisses et ordonnées des 68 points d'intérêt extraits.

2.3 Classification

Les données sont classifiées au moyen d'un classifieur neuronal incrémental à base de prototypes. C'est un réseau de neurones à trois couches, la première recevant les entrées, la seconde étant constituée de "neurones-distance", i.e. des prototypes, qui sont totalement connectés aux neurones d'entrée. A chaque présentation d'un exemple, ce dernier est comparé à tous les prototypes en mémoire. Dans la troisième couche, plusieurs prototypes sont associés à un neurone de sortie, chacun représentant un score dépressif BDI-II.

La démonstration dont est l'objet ce papier n'utilise qu'en phase de généralisation un classifieur déjà entraîné, donc il n'y a pas création de nouveaux prototypes. Les règles de fonctionnement du classifieur permettent de sortir, ou non, une classe de score dépressif. La possibilité d'obtenir une

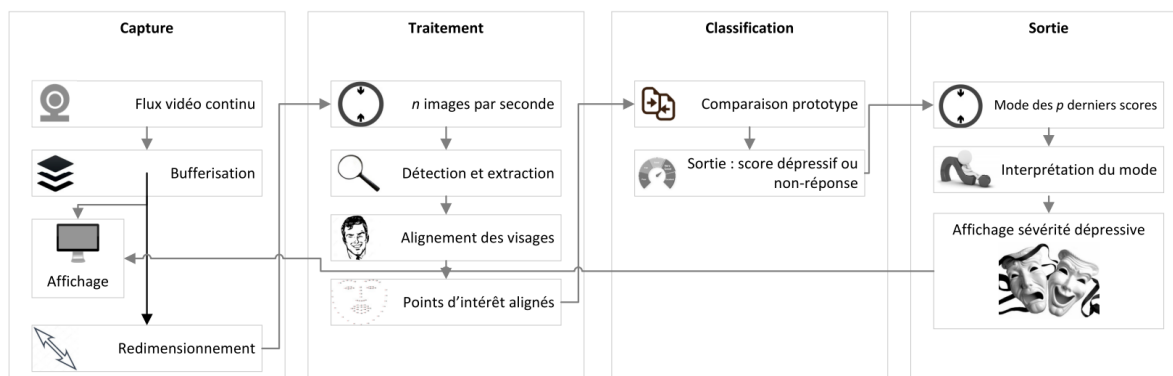


FIGURE 1 – Schéma d'ensemble du système d'estimation de l'état dépressif

TABLE 1 – Interprétation du score au test BDI-II

Score obtenu	Sévérité de la dépression
0-13	Minimale
14-19	Moyenne
20-28	Modérée
29-63	Sévère

"non-réponse" est une spécificité précieuse de ce classifieur qui le rend plus proche d'un diagnostic humain. Lors de la classification en temps réel, les non-réponses peuvent permettre de moduler la décision. Ainsi, le nombre de non-réponses rendues peut permettre le calcul d'une mesure de fiabilité de l'interprétation du score dépressif.

2.4 Diagnostic

La sortie du système est l'interprétation du score (voir Table 1) le plus représenté lors de la classification des images sur les p scores les plus récents. Le paramètre p est à ajuster en accord avec la durée de l'interaction. Les résultats en généralisation sur le corpus AVEC'2014 atteignent un taux de succès de 94% pour la classification de l'intervalle dépressif, contre 90% pour la classification du score BDI-II, d'où le choix de cette sortie qui se veut plus précise et interprétable en l'état.

3 Scénario

On présente ici un cas d'utilisation simple du système de classification de l'état dépressif, où l'utilisateur lit un passage affiché à l'écran pendant qu'il est filmé par une webcam.

Le nombre d'images traitées par seconde est fixé à $n = 10$, et la durée de l'enregistrement à 20 secondes. On fixe à $p = 20$ le nombre de scores pris en comptes dans l'affichage du résultat.

Quelques secondes après le démarrage de la procédure, l'interprétation de l'état dépressif est affichée. On présente également un indicateur de détection du visage, faisant savoir à l'utilisateur s'il doit modifier sa posture face à la

caméra. La Figure 2 présente un aperçu de la sortie du système.



FIGURE 2 – Exemple capturé depuis la sortie du système de classification

4 Conclusion

Le système proposé permet de classifier en temps réel l'état dépressif d'un sujet humain au moyen d'un classifieur neuronal incrémental. Une chaîne de traitement rapide du flux vidéo est mise en œuvre afin de produire une interprétation de l'état dépressif, au regard du test BDI-II. Les résultats obtenus sont encourageant pour la poursuite du développement de l'outil. Toutefois, il est important de considérer avec précaution le résultat. D'une part, le système n'a pas encore pu bénéficier de l'expertise d'un professionnel de la psychiatrie. D'autre part, dans sa définition, le score BDI-II est évalué sur des sujets exprimant des symptômes dépressifs depuis au moins deux semaines, d'où l'importance de fixer un cadre d'utilisation du système en amont.

Références

- [1] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "Avec 2014 : 3d dimensional affect and depression recognition challenge," *Proc. 4th Int. Work. Audio/Visual Emot. Chall. - AVEC '14*, pp. 3–10, 2014.