

Exploration visuelle d'images IRMf basée sur des Gaz Neuronaux Croissants

Jerzy Korczak*, Jean Hommet*, Nicolas Lachiche*, Christian Scheiber**

*LSIIT, CNRS, Illkirch

<jjk,hommet,lachiche>@lsiit.u-strasbg.fr

**CHU Lyon

christian.scheiber@chu-lyon.fr

Résumé. Les algorithmes actuels de fouille de données ne supportent que de façon très limitée les mécanismes de guidage et d'engagement d'expert dans le processus de découverte. Dans cet article, nous présentons une nouvelle approche interactive de fouille des images IRMf, guidée par les données, permettant l'observation du fonctionnement cérébral. La discrimination des voxels d'image du cerveau qui présentent une réelle activité est en général très difficile à cause d'un faible rapport signal sur bruit et de la présence d'artefacts. L'exploration de donnée visuelle se focalise sur l'intégration de l'utilisateur dans le processus de découverte de connaissance en utilisant des techniques de visualisation efficaces, d'interaction et de transfert de connaissances. Dans ce travail, nous montrons sur les données réelles, que l'exploration visuelle permet d'accélérer le processus d'exploration d'images IRMf et aboutit à de meilleurs résultats dotés d'une confiance accrue.

1 Introduction

Les techniques modernes d'imagerie cérébrale, comme l'Imagerie par Résonance Magnétique fonctionnelle (IRMf), offrent la possibilité d'enregistrer en même temps l'activité de l'ensemble du cerveau. C'est une force, mais cela génère une masse de données considérable (environ 300000 voxels, "pixels tridimensionnels", pour lesquels on recueille entre 100 et 1000 observations). Du point de vue de la fouille de données, le cerveau est un objet complexe par excellence. En général, la discrimination des voxels d'image du cerveau qui présentent une réelle activité est très difficile à cause d'un faible rapport signal sur bruit et de la présence d'artefacts. Les premiers tests des algorithmes actuels de fouille dans ce domaine ont montré que leurs performances et leurs qualités de reconnaissance sont faibles (Sommer and Wichert, 2003). En raison de la difficulté qu'il y a à manipuler de telles quantités d'informations, l'essentiel des études ne cherchent pas à les explorer, mais s'en servent pour tester un modèle par le biais de statistiques univariées effectuées en chacun des points. C'est le principe de logiciels de traitement tels que Statistical Parametric Mapping (SPM) (Friston et al., 1995), AFNI (Cox, 1996) ou BrainVoyager (Goebel, 1997) qui consiste à mettre en évidence les voxels plus actifs dans une condition par rapport à une autre.

Sur le plan international, un grand nombre de recherches méthodologiques sont en cours pour mettre en évidence les variations qui ont du sens. On peut regrouper celles-ci en deux grandes familles. La plus commune est l'approche par les statistiques

multivariées comme les MANCOVA, PCA, PLS, analyse canonique ou plus récemment l'ICA (Beckmann and Smith, 2003). La seconde famille d'approche regroupe toutes les méthodes de fouille de données, partant des techniques de clustering jusqu'aux algorithmes génétiques en passant par les réseaux neuronaux (Sommer and Wichert, 2003).

La fouille de donnée visuelle se focalise sur l'intégration de l'utilisateur, un expert-médecin, dans le processus de découverte de connaissance en utilisant des techniques de visualisation efficaces, d'interaction et de transfert de connaissances. Dans cet article, nous présentons une nouvelle approche interactive de fouille des images IRMf, guidée par les données. L'originalité de notre approche tient au fait que nous n'allons pas seulement mettre en oeuvre des techniques de fouille de données encore non appliquées au domaine de l'imagerie fonctionnelle, mais surtout que celles-ci vont être étendues par l'injection de connaissance *a priori* dans les mécanismes de fouille du système et l'interactivité, intégrant directement l'expert-médecin dans le processus de découverte et d'apprentissage de concepts pour mettre en évidence les zones fonctionnelles du cerveau et leur organisation.

Les algorithmes actuels de fouille de données ne supportent que de façon très limitée les mécanismes de guidage et d'engagement d'expert dans le processus de découverte. Les algorithmes de classification classiques requièrent des paramètres à définir à l'initialisation qui influencent de manière importante le résultat final. A contrario, l'approche visuelle autorise une gestion dynamique des paramètres de classification en intégrant l'utilisateur dans le processus de découverte de connaissance. Le principe de base consiste à visualiser dynamiquement l'apparition des classes, filtrer et détailler les plus prometteuses et affiner le processus de découverte. Les présentations visuelles augmentent la perception de l'utilisateur sur le déroulement de la classification et permettent d'arriver plus rapidement à des résultats concluants. Les régions intéressantes sont repérées immédiatement par l'expert qui peut agir sur les classes découvertes avec une précision désirée.

L'article est structuré de la façon suivante. Dans le chapitre 2 nous présentons les caractéristiques des données IRMf. Dans le chapitre 3 nous décrivons notre approche de fouille interactive d'images IRMf en particulier les techniques d'engagement d'expert dans le processus d'exploration d'images IRMf. Le chapitre 4 détaille les aspects d'interactivité de l'algorithme de classification neuronale : Gaz Neuronaux Croissants. Nous concluons dans le dernier chapitre en dégageant quelques perspectives.

2 IRMf comme objet complexe

L'activité cérébrale qui se caractérise par des phénomènes électriques rapides, peut être mesurée directement avec les techniques d'électro-encéphalographie (EEG) et de magnéto-encéphalographie (MEG). Si ces deux techniques ont une très bonne résolution temporelle, leur résolution spatiale par contre est quasi-nulle. Pour pouvoir savoir où et quand a lieu l'activité cérébrale, on a recours à l'Imagerie par Résonance Magnétique fonctionnelle (IRMf). Cette technique est une mesure indirecte car elle n'est pas sensible à l'activité même des neurones, mais à la consommation d'oxygène qui en dépend. L'effet BOLD (eng. *Blood Oxygenation Level Dependent*) est le phénomène qui permet

la mesure de cette consommation, il se traduit par une évolution du signal RMN induite à la fois par des variations de concentration d'oxygène et des variations de débit sanguin. L'IRMf consiste donc en une IRM optimisée sur l'effet BOLD. La mesure de l'activité d'un cerveau revient à enregistrer des images IRM à la suite les unes des autres pendant que ce cerveau travaille. Un tel enregistrement est donc composé d'une série d'images en trois dimensions d'un même cerveau immobile, et dont les variations d'intensité d'une image sur l'autre sont liées avec l'activité de ce dernier. Les zones actives sont ainsi localisées aussi bien spatialement que temporellement avec des résolutions typiques de 3 mm et 3 s. A ces échelles, le signal IRMf d'un voxel représente l'activité d'un agrégat de millions de neurones moyenné sur un intervalle de temps très grand devant le signal isolé d'un seul neurone.

L'IRMf permet de localiser précisément chez un sujet des zones liées à un fonctionnement ou à un dysfonctionnement donné. Le cerveau étant toujours en activité et le rapport signal sur bruit d'IRMf étant très faible, cette technique a recours systématiquement à un paradigme pour faire ressortir un effet escompté. Le paradigme est une représentation du fonctionnement de l'ensemble de l'acquisition. Intégrant la séquence de stimuli ainsi que les paramètres d'acquisition, le paradigme doit être optimisé de manière à maximiser les contrastes entre les différentes conditions du test afin de relever au mieux leurs effets.

Les données IRMf suivent principalement cinq grands axes de complexité. Elles sont à la fois des données en trois dimensions spatiales plus une dimension temporelle, elles peuvent correspondre à plusieurs sujets ou bien à l'historique d'un même patient, enfin elles peuvent être combinées avec des connaissances anatomiques ou plus généralement médicales. De plus, une acquisition IRMf génère une série de 100 à 1000 images IRM, formant une série IRMf de taille comprise entre 25 Mo et 1 Go.

Les bruits durant l'acquisition IRMf sont nombreux et très variés. On trouve tout d'abord les bruits physiques instrumentaux classiques comme par exemple des dérives ou des parasites électroniques qui sont fréquents dans les conditions instrumentales extrêmes de l'IRM où il s'agit de capter des signaux RF très faibles au sein de champs magnétiques très intenses. Autre source de bruit, les mouvements du sujet qui génèrent des signaux parasites, ce sont non seulement les déplacements de la tête et des yeux, mais aussi des déplacements internes induits par les pulsations cardiaques à proximité des veines. Enfin les activités cérébrales sans relation avec le paradigme comme des souvenirs ou des envies sont également générateurs de bruit.

Le protocole expérimental d'IRMf classique consiste à définir un paradigme du fonctionnement du cerveau pour une expérience donnée. Cela consiste à définir dans le même temps, le programme des tâches à soumettre au cerveau, aussi bien qu'un modèle de la réponse hémodynamique pour chacune de ces tâches. La réponse mesurée expérimentalement est ensuite comparée statistiquement avec ce modèle proposé *ab initio*. La méthode statistique est performante mais ne peut conclure en dehors du modèle préfixé. Avec cette méthode les résultats doivent être forcément anticipés, ce qui n'est pas toujours possible. Le concept de fouille de donnée, peut se révéler utile en complément ou en remplacement de la méthode classique lorsqu'il est délicat de prévoir ce qui va se passer durant l'acquisition. La démarche proposée ici est d'utiliser des méthodes d'exploration de données au sein d'un système interactif pour faire ressortir

les zones actives sans avoir recours à un modèle.

3 Engagement d'expert dans le processus de découverte

La fouille interactive d'images IRMf n'est pas basée sur un modèle mais au contraire est une approche guidée par les données qui de plus intègre directement un expert dans le processus de découverte. L'architecture de notre système de fouille d'images IRMf et le processus d'exploration sont détaillés dans (Korczak et al., 2005). Ce processus est composé de cinq phases que sont l'acquisition et la sélection des données, leurs prétraitements, la classification, l'extraction de règles et de concepts, enfin la validation. Les données IRMf sont issues de l'acquisition de 100 à 1000 images IRM séquentielles d'un cerveau, l'enregistrement ayant lieu alors que ce dernier effectue une série de tâches programmées. La préparation des données consiste à choisir des attributs mettant en forme des vecteurs de manière à établir une relation de distance sur les données. Une acquisition de m images de taille n voxels, génère un ensemble de données de n individus. Un individu est un point dans un espace à m dimensions dont chaque coordonnée est l'écart de l'intensité du voxel considéré par rapport à sa moyenne sur la série.

Le système permet de faire une fouille de données interactivement afin de pouvoir optimiser l'expérience et la renouveler aussitôt sur le sujet. Les images IRMf peuvent être classifiées à partir de séries complètes ou en cours d'acquisition avec l'engagement de l'expert médecin dans le processus de découverte. Dans ce dernier mode, en fonction des résultats obtenus, l'expert-médecin peut modifier l'acquisition d'un point de vue purement technique par des paramètres géométriques ou temporels (résolution, zoom, etc). Cette réaction sera influencée par un retour d'information uniquement lié à l'état des signaux. Les algorithmes de classification et d'explication symbolique des classes aident l'expert-médecin à comprendre les classes générées et à modifier l'expérience directement au niveau du paradigme. Ces modifications sont suscitées par les classifications générées par le système de fouille de donnée, et choisies par l'expert en fonction d'hypothèses sur le fonctionnement cérébral ou à partir de données connues comme des données anat-fonctionnelles préalablement saisies dans la base. Pour arriver à ce stade, le système doit être très performant, en retournant des informations à un niveau d'interprétation suffisamment élevé. Le système est capable d'apprendre et stocker des connaissances acquises pendant les expérimentations et de s'en resservir comme connaissances préalables dans de futurs examens. Cette imagerie fonctionnelle interactive demande une analyse très rapide de résultats préliminaires pendant l'acquisition, cette analyse n'est envisageable que par une méthode spécifique qui s'insère dans la démarche actuelle.

Le travail présenté dans cet article est centré essentiellement sur la phase d'exploration, dont la démarche se veut interactive et dynamique autour d'algorithmes de classification non supervisée. La classification consiste à décrire des voxels en les distribuant en un nombre limité de classes homogènes. Ces classes regroupent les voxels qui ont des caractéristiques et comportements similaires. La classification est dite "dirigée par les données" car le processus n'utilise pas de connaissance externe aux données et

ne dispose que de leur description pour extraire des informations sur la structure de l'ensemble.

Les méthodes de classification non supervisée (Bock and Diday, 2000) ont déjà été appliquées à l'IRMf à travers plusieurs méthodes dérivées de statistique, de logique floue ou encore de classification neuronale et hiérarchique (Dimitriadou et al., 2004), (Goutte et al., 1999), (Moller et al., 2001), l'Analyse en Composante Principale (ACP) (Andersen et al., 1999), (Lai and Fang, 1999), et l'Analyse en Composantes Indépendantes (ACI) (Esposito et al., 2002).

Dans nos travaux, nous nous sommes plus particulièrement focalisés sur les méthodes neuronales : l'algorithme de Kohonen (Kohonen, 1982), (Fischer and Hennig, 1999) (Ngan and Hu, 1999) et l'algorithme des Gaz Neuronaux Croissants (GNC) (Fritzke, 1995). Suite à nos expérimentations, nous favorisons l'algorithme GNC qui s'adapte bien à l'exploration visuelle interactive d'images IRMf. Dans cet article, nous ne considérerons donc que cet algorithme. L'originalité et l'intérêt des GNC résident dans le fait que le nombre de classes n'est pas fixé à l'avance contrairement à la plupart des autres méthodes. Le nombre de classes peut tout aussi bien augmenter que diminuer durant l'exécution, rendant l'algorithme particulièrement flexible. Nous avons rendu la procédure de classification interactive, de manière à engager l'expert dans le processus de découverte. L'utilisateur dispose de retours et de contrôles sur l'évolution de la classification, pouvant ainsi suivre et intervenir pour orienter le déroulement. Cette exploration visuelle en IRMf dispose des outils de vision 3D, des retours statistiques et un algorithme de classification suffisamment souple et dynamique pour supporter l'interaction [Fig. 1]. Nous avons étendu le système de visualisation SLICER (<http://www.slicer.org/>) en y intégrant des méthodes de classification ainsi que des outils d'exploration d'images IRMf. Les classes produites par l'algorithme sont évaluées par l'expert qui ne retient que les classes pertinentes. Chaque classe correspond à un ensemble de voxels, régions du cerveau, ayant la même réponse hémodynamique au cours du temps. Les caractéristiques des classes sont affichées par le système durant le processus d'exploration ; la figure [Fig. 1] montre un exemple de classe relevant d'une activation de type bloc associée aux statistiques sur le taux d'échanges de voxels entre classes informant sur la stabilité et la convergence de l'algorithme. Cette réponse peut être caractérisée explicitement par la construction de règles. Ces règles combinent des motifs temporels observés avec des informations spatiales telles que l'activité des voxels ou des régions voisines, en plus des connaissances du domaine, comme l'atlas des fonctions des régions du cerveau. Ces motifs peuvent être synchronisés avec le paradigme, par exemple pour découvrir l'interaction entre les régions du cerveau utilisées pour la mémoire visuelle. D'autres motifs temporels peuvent être indépendants de tout paradigme, par exemple pour mettre en évidence la succession d'activations de régions du cerveau typiques de l'hallucination.

4 Classification interactive avec GNC

Afin d'illustrer l'exploration interactive d'images, nous avons choisi GNC parmi les algorithmes implantés dans notre système (Korczak et al., 2005). Dans cet algorithme, des classes peuvent apparaître ou disparaître au cours du temps sous l'influence d'er-

Exploration visuelle d'images IRMf basée sur des Gaz Neuronaux Croissants

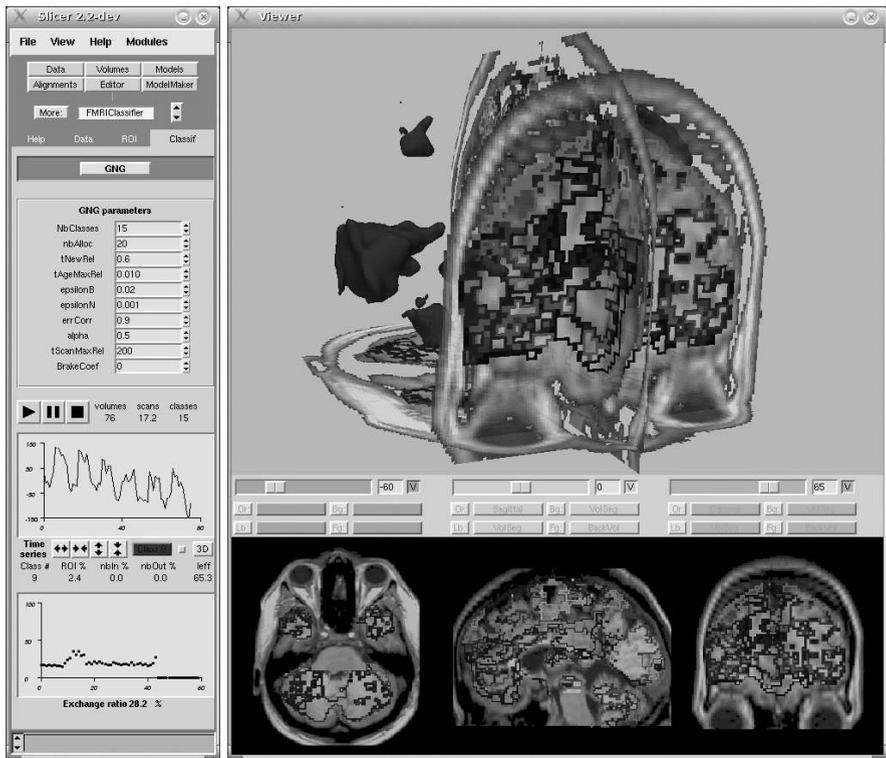


FIG. 1 – Exemple d'exploration interactive d'images IRMf

reurs évaluées. Ceci donne à GNC une souplesse qui lui permet de s'adapter à une démarche interactive, dans laquelle l'évolution de la classification est également influencée par un expert. Il s'agit d'un algorithme dynamique de réseau de neurones inter-connectés. Les connexions entre classes ont la propriété de vieillir et disparaissent lorsqu'elles atteignent un âge maximal prédéfini. Cette propriété est la cause de la disparition des classes qui sont éliminées lorsqu'elles ne sont plus connectées. L'apparition des classes se fait à intervalles de temps régulier en insérant une nouvelle classe auprès de la classe qui présente la plus grande erreur. L'erreur d'une classe est une évaluation basée sur la somme des distances de cette classe aux individus qui l'ont déclarée gagnante. Ainsi le réseau se renforce auprès des classes qui parcourent les plus grandes distances ou qui réunissent le plus d'individus. Nos contributions portent sur deux aspects de classification, notamment sur la réduction de l'espace des données et une aide au paramétrage de l'algorithme assisté par des outils statistiques et de fouille de données visuelle.

L'algorithme des Gaz Neuronaux Croissants adapté aux données IRMf a été spécifié

dans (Korczak et al., 2005). En bref, la modification des classes se fait itérativement à l'échelle de la donnée; on adapte la classe la plus proche ainsi que les classes voisines en fonction de la dernière donnée tirée. En modifiant de la sorte plusieurs classes simultanément, on accélère le processus et l'on donne une cohésion à l'ensemble des classes. De plus, les adaptations sont réglables par des coefficients qui prennent en compte le degré de voisinage. La contre partie de cette organisation en réseau de classes est qu'elle peut générer des classes vides.

Nous introduisons une pause dans l'algorithme durant laquelle le GNC est dans un état cohérent. Ceci permet d'effectuer des mesures sur cet état, de le modifier, de le sauver ou encore de le restaurer. Cette pause est ainsi invoquée systématiquement à intervalles de temps réguliers pour effectuer des mesures absolues sur la classification et alimenter les outils de visualisation. Elle est aussi utilisée à la demande lorsque l'expert intervient sur le cours de la classification. Ce dernier a donc accès et peut modifier tout point de la classification durant l'exécution. Tout peut être modifié sans avoir besoin d'arrêter et de relancer le processus. Les paramètres de l'algorithme ainsi que l'espace des données peuvent donc être optimisés et varier durant la classification. Quatre niveaux d'interaction peuvent être discernés. Le niveau minimum correspond à un retour de mesures sur l'état courant et permet à l'utilisateur de pouvoir décider quand interrompre le processus. Le deuxième niveau est la modification du volume des données, ce niveau est rendu possible lorsque l'adaptation des classes est indépendante de l'historique de la classification. Le troisième niveau est accessible aux algorithmes qui disposent de paramètres non nécessairement fixés. Dans ces conditions l'utilisateur peut les faire varier pour les optimiser en cours d'exécution. Au delà de ces trois niveaux d'interaction, une gestion de sauvegarde et de restauration des états de la classification constitue un dernier aspect d'interaction avec la classification.

L'expert-médecin peut suivre l'évolution de la classification au cours du temps à partir d'outils visuels alimentés à intervalles de temps réguliers par l'algorithme. Les informations fournies sont de trois ordres : statistiques, temporelles et spatiales. Les statistiques portent sur la classification en tant que telle, quant aux informations temporelles et spatiales, elles sont liées à la nature de l'IRMf. Les informations statistiques décrivent l'évolution de la classification par des mesures sur les classes générées. L'utilisateur peut suivre à partir de graphiques, l'évolution des erreurs, des inerties intra-classe et inter-classes, du nombre de voxels réunis au sein de chaque classe et du nombre de voxels qui changent de classe par unité de temps. L'ensemble de ces informations renseigne notamment sur la dispersion des classes, la convergence et la stabilité de la classification [Fig. 1, partie gauche].

Les informations temporelles et spatiales véhiculées dans chaque classe, représentent l'évolution des signaux IRMf au cours du temps et leurs localisations dans le cerveau. Les signaux sont affichés sous forme de courbes dans des graphiques où l'on peut également faire figurer le paradigme, soit par la séquence de stimuli, soit par un modèle de la réponse lorsqu'il existe. La visualisation spatiale est effectuée simultanément de deux manières, à partir de trois coupes perpendiculaires et directement en trois dimensions [Fig. 1, partie droite].

La classification en cours alimente un volume où les couleurs des voxels sont associées aux classes. Cette vision 3D de la classification peut être superposée suivant les

cas, à un des volumes de la série d'acquisition, à une IRM structurelle du sujet ou à un volume normalisé faisant office d'atlas anatomique.

Toutes ces informations sont rafraîchies régulièrement durant la classification permettant à l'utilisateur de suivre et de guider en direct l'évolution du processus. Il opère ainsi une visualisation active, en affichant les aspects qu'il juge opportuns et en intervenant sur le cours de la classification. Les leviers dont il dispose pour influencer le processus sont les paramètres de l'algorithme et le volume des données qu'il peut modifier quand il le désire.

Dans le mode d'exploration, l'expert-médecin peut facilement optimiser l'algorithme, par l'observation en direct des mesures statistiques et de leurs variations induites par des modifications de paramétrage. Ainsi, le réglage de l'algorithme peut être convenablement effectué sans avoir besoin d'arrêter et de recommencer sans cesse. D'autre part, les paramètres peuvent être adaptés en fin de classification pour stabiliser ou affiner un résultat.

Dans la découverte de régions actives du cerveau, les moyens d'intervention de l'expert-médecin se situent au niveau du réglage du nombre de classes nécessaires et de la sélection des régions explorées. Ainsi, l'augmentation du nombre de classes se fait naturellement, et sa réduction peut être réalisée directement par l'utilisateur qui a la possibilité d'éliminer des classes. Notons au passage que la suppression d'une classe est accompagnée d'une gestion appropriée des connexions afin de maintenir la cohérence du réseau. En définitive pour adapter au mieux le nombre de classes aux données IRMf, l'expert-médecin fixe la valeur du nombre maximal de classes en observant l'apparition des nouvelles classes durant l'exploration des images. Il peut par exemple décider de stopper cette croissance après l'apparition d'une classe escomptée ou bien revenir en arrière par l'élimination d'une classe de son choix.

Pour augmenter la vitesse d'exécution et diminuer la complexité des résultats, l'utilisateur a toujours intérêt à limiter au maximum le volume des données. L'espace d'exploration peut être ajusté au moyen de plusieurs techniques. Un outil inclus dans la visualisation 3D, permet de sélectionner des volumes d'intérêt ou de désintérêt. L'espace peut aussi être restreint à des structures anatomiques. Dans tous les cas un seuillage permet d'éliminer l'espace qui entoure le crâne. Et dans le cas d'images normalisées, il est possible de limiter les recherches à la matière grise. Les résultats intermédiaires de classification sont aussi utilisés pour la sélection des données. Ainsi, l'expert-médecin peut poursuivre une classification en focalisant son intérêt sur certaines classes après avoir éliminé l'ensemble des voxels des autres classes. L'augmentation de l'espace de recherche, également possible, est parfois désiré. Le cas se produit lorsque l'on veut étendre une classification à d'autres zones, par exemple après avoir optimisé certains paramètres de l'algorithme sur une région volontairement réduite pour gagner du temps.

Des expérimentation sont été effectuées sur un jeu de données caractéristique : une série IRMf de tests auditifs. Ces données proviennent du site de l'institut de recherche londonien "The Wellcome Department of Imaging Neuroscience" ¹. Les résultats obtenus (Korczak et al., 2005) sont encourageants. Le système identifie les régions pertinentes et facilite le guidage interactif de processus de fouille.

¹<http://www.fil.ion.ucl.ac.uk/spm/>

5 Conclusion

Dans cet article, nous avons présenté une approche d'exploration visuelle d'images médicales avec un engagement d'expert dans le processus de découverte. Les séries d'images 3D IRMf ont été considérées comme objets complexes en prenant en compte le volume, la dimension temporelle, les relations spatiales et le bruit. Nous avons détaillé la partie visualisation de notre système de fouille de séquences d'images 3D IRMf. Dans notre approche, nous avons mis en avant l'implication d'un expert-médecin dans le processus de classification d'une part, et d'autre part, l'intégration d'un algorithme de classification capable de s'adapter aux spécificités de l'IRMf avec de faibles connaissances préalables. La méthode d'exploration visuelle est indépendante de l'algorithme de classification utilisé. Plusieurs algorithmes ont été étudiés : K-means, LBG, SOM et GNC ; le modèle des Gaz Neuraux Croissants s'est avéré plus robuste et reproductible dans les mêmes conditions de fonctionnement et surtout plus adapté à l'exploration visuelle interactive. Tous ses paramètres peuvent être modifiés en cours d'exécution y compris le paramètre du nombre de classes. Cet aspect dynamique est déterminant pour l'exploration interactive. L'outil de fouille de données visuelle qui en découle permet notamment d'optimiser les paramètres de l'algorithme, de réduire l'espace de recherche, de présenter le processus de découverte sous plusieurs angles et participe ainsi à la diminution de la complexité des objets concernés pour fournir à l'expert-médecin des classifications de régions cervicales plus compréhensibles. Les premiers résultats sur des données de type bloc sont encourageants et nous permettent d'envisager la poursuite de ces travaux vers des expériences de type événementiel. Il s'avère que l'exploration visuelle permet d'accélérer le processus d'exploration d'images IRMf et aboutit à de meilleurs résultats dotés d'une confiance accrue.

Remerciements. Les auteurs remercient K. Friston et G. Rees pour les données de test SPM, ainsi que les étudiants de l'Université Louis Pasteur de Strasbourg, France, H. Hager, P. Hahn, V. Meyer, J. Schaeffer et O. Zitvogel, pour leur participation dans la phase initiale de réalisation du projet.

Références

- Andersen, A. H., Gash, D. M., and Avison, M. J. (1999), Principal component analysis of the dynamic response measured by fMRI : a generalized linear systems framework. *Magnetic Resonance Imaging*, 17 :795–815, 1999.
- Beckmann, C. and Smith, S. M. (2003), Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. on Medical Imaging*, 2003.
- Bock, H. and Diday, E. (2000), *Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data*, Springer Verlag, 2000.
- Cox, R. W. (1996), AFNI : Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29 :162–173, 1996.
- Dimitriadou, E., Barth, M., Windischberger, C., Hornick, K., and Moser, E. (2004), A quantitative comparison of fMRI cluster analysis. *AI in Medicine*, 31 :57–71, 2004.
- Esposito, F., Formisano, E., Seifritz, E., Goebel, R., Morrone, R., Tedeschi, G., and

- DiSalle, F. (2002), Spatial independent component analysis of fMRI time-series : To what extent do results depend on the algorithm used? *Human Brain Mapping*, 16 :146–157, 2002.
- Fischer, H. and Hennig, J. (1999), Neural network-based analysis of MR time series. *Magnetic Resonance in Medicine*, 41 :124–131, 1999.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. (1995), Statistical parametric maps in functional imaging : A general linear approach. *Human Brain Mapping*, 2 :189–210, 1995.
- Fritzke, B. (1995), A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural Information Processing Systems 7*, pages 625–632. MIT Press, Cambridge MA, 1995.
- Goebel, R. (1997), *BrainVoyager : Ein Programm zur Analyse und Visualisierung von Magnetresonanztomographiedaten*. T. Plesser and P. Wittenburg, Forschung und Wissenschaftliches Rechnen, 1997.
- Goutte, C., Toft, P., Rostrup, E., Nielsen, F. A., and Hansen, A. K. (1999), On clustering fMRI time series. *NeuroImage*, 9 :298–310, 1999.
- Kohonen, T. (1982), Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43 :59–69, 1982.
- Korczak, J., Scheiber, C., Hommet, J., and Lachiche, N. (2005), Fouille interactive de séquences d'images 3D d'IRMf. *RNTI*, 2005.
- Lai, S. H. and Fang, M. (1999), A novel local PCA-based method for detecting activation signals in fMRI. *Magnetic Resonance Imaging*, 17 :827–836, 1999.
- Moller, U., Ligges, M., Grunling, C., Georgiewa, P., Kaiser, W. A., Witte, H., and Blanz, B. (2001), Pitfalls in the clustering of neuroimage data and improvements by global optimization strategies. *NeuroImage*, 14 :206–218, 2001.
- Ngan, S. C. and Hu, X. (1999), Analysis of functional magnetic resonance imaging data using self-organizing mapping with spatial connectivity. *Magnetic Resonance in Medicine*, 41 :939–946, 1999.
- Sommer, F. T. and Wichert, A. (2003), *Exploratory Analysis and Data Modeling in Functional Neuroimaging*. The MIT Press, 2003.

Summary

Current data mining algorithms do not contain sophisticated guiding and engagement of expert facilities. In this paper a new interactive data-driven approach of brain fMRI images is presented. Discrimination of the image voxels of the brain that represent real activity is, in general, very difficult because of a weak signal-to-noise ratio and of the presence of artifacts. Visual exploration combines the expert involvement in brain mining process, interactive visualisation techniques, and the use of domain knowledge. As an example of unsupervised classification an algorithm of Growing Neural Gas has been developed and tested on sequences of fMRI images. The results of the tests have shown a strong influence on the classifier performances by the number of classes, signal-to-noise ratio, and volumes of activated and explored zones. The interface allows not only to accelerate the classification process but also to increase the user confidence in generated hypotheses.