

Annexe1 :

Eléments de mathématiques pour la fusion de données **Christophe Baehr (Météo-France/CNRS - CNRM/GMEI)**

Ce court exposé vise à présenter les objets mathématiques généralement décrits par la fusion de données et survolera trois techniques classiques que l'on rencontre lors de la résolution des problèmes de fusion de données. Il ne s'agira pas d'être exhaustif mais de rappeler les idées et le vocabulaire. Selon la demande, suite à ces ateliers dans le cadre des groupes de travail ATIS du CNRM, trois présentations plus longues pourront être données pour détailler ces objets et techniques issues des mathématiques appliquées.

La fusion de données cherche à déterminer l'état d'un système réputé méconnu à partir d'informations multiples et incertaines ou ne permettant pas de décider chacune de l'état du système considéré. En ingénierie météorologique, une technique de fusion de données est bien connue, l'assimilation de données fournissant quotidiennement l'état de l'atmosphère analysée à partir d'observations multiples et de nature différente.

Il existe plusieurs manières de réaliser la fusion de données. On peut considérer trois niveaux de fusion :

- la fusion au niveau de la mesure ou des images, fusion de bas niveau ;
- la fusion d'attributs, fusion de niveau intermédiaire. Elle porte sur la combinaison d'informations après traitement des mesures ;
- la fusion de décisions, fusion de haut niveau. C'est la combinaison des décisions obtenues à partir de chaque source.

Mathématiquement la fusion de données renvoie à la théorie des ensembles et de la mesure. Placé dans un cadre particulier, on retrouvera les objets probabilistes. Les lois conditionnelles sont alors les quantités à calculer et la règle de Bayes est l'instrument principal. Nous parlerons alors d'algorithmes stochastiques et notamment de la méthode de Monte-Carlo par chaînes de Markov. L'assimilation opérationnelle se rattache à cette famille de techniques.

En changeant les règles de mesure des ensembles, nous définirons la théorie des possibilités. Notamment on ajoutera une règle de maxitivité qui rompt le caractère additif des mesures mais amène des outils spécifiques. On terminera par un pont jeté entre probabilité et possibilité avec une transformation entre distribution de probabilité et distribution des possibilités.

En apprentissage automatique (machine learning), un classifieur est une famille d'algorithmes de classement statistique. En classification automatique les algorithmes sont nombreux, citons les réseaux de neurones artificiels, les arbres de décision, la régression logistique, les algorithmes génétiques..., pour ne pas citer les méthodes de filtrage non-linéaires.

On peut alors définir trois niveaux de fusion de classifieurs:

- fusion au niveau de la représentation. On quantifie les degrés de ressemblance entre une forme entrante et les autres formes de la classe ;
- fusion au niveau des rangs. On quantifie les rangs de toutes les classes, le rang un qualifiant le choix préféré ;
- fusion au niveau de la décision. Un classifieur indique sa décision d'une classe ou de sous-ensemble de classes.

Si le temps nous le permet nous esquisserons rapidement le principe des réseaux de neurones artificiels.