

Fonctions de croyance appliquées à la surveillance de la pollution atmosphérique

P. Vannoorenberghe¹, E. Lefevre² et O. Colot²

¹ HEUDIASYC, UMR 6599 CNRS

Université de Technologie de Compiègne

B.P. 20529, 60205 Compiègne Cedex

² PSI, UPRES EA 2120

Université/INSA de Rouen

Place Emile Blondel, BP 08, 76131 Mont-Saint-Aignan Cedex

Email : Patrick.Vannoorenberghe@hds.utc.fr

1 Introduction

Le problème envisagé concerne le diagnostic précoce de l'évolution de la pollution atmosphérique vers une situation du type "Jour Pollué". Nous nous sommes focalisés sur le réseau de surveillance de la ville de Lyon dont les données nous ont été fournies par COPARLY. Dans cet article, on se place dans le cadre d'un problème de discrimination où les données sont préalablement étiquetées (2 classes) en proposant une méthode basée sur l'apprentissage d'un historique de mesures. L'événement à diagnostiquer est produit sous forme binaire comme la prédiction du jour J de l'absence ou de la présence d'une pollution au jour J+1.

2 Analyse des données

Les informations délivrées par les capteurs de pollution sont de nature imparfaite c'est-à-dire imprécises, incertaines et incomplètes. En effet, une première imperfection réside dans l'incomplétude des bases de données. A titre d'exemple, sur les mesures de concentration en ozone, 24% des données sont manquantes. Certaines techniques d'analyse de données permettent de prendre en compte ce type d'imperfection soit en remplaçant les données manquantes ou tout simplement en les ignorant. De surcroît, la pollution étant un phénomène difficilement modélisable, il s'avère que le contexte du problème posé évolue en fonction du temps. A titre d'exemple, on a représenté sur la figure Fig. 1 l'évolution de la concentration moyenne journalière en ozone sur un capteur durant les cinq années d'étude. Les moyennes annuelles sur ce capteur sont également représentées sur ce graphique. On s'aperçoit que

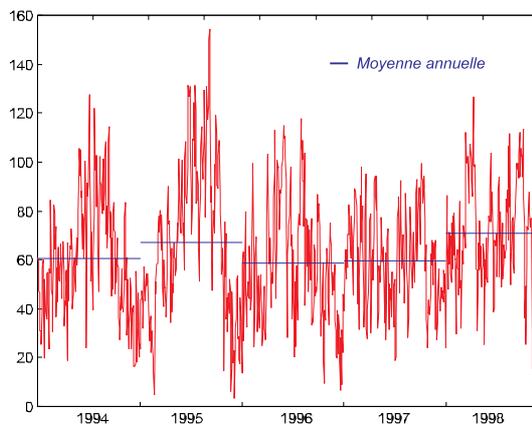


FIG. 1 – Evolution de la concentration moyenne journalière en ozone sur le capteur St-Just

ces moyennes varient d'une année sur l'autre rendant le problème de l'apprentissage d'autant plus délicat à appréhender. De plus, le problème étudié s'oriente autour de la détection d'événements rares. En effet sur les 897 jours de données observés, seulement 58 jours sont considérés comme des situations anormales représentant uniquement 6.3% des données. Ces constatations tendent en faveur d'une méthodologie appliquée aux traitements de données imparfaites.

3 Méthodologie proposée

De manière à assurer une gestion des incertitudes liées aux données environnementales, nous proposons une approche basée sur l'utilisation des fonctions de croyance et de la théorie de l'évidence. A partir d'un formalisme mathématique initialement introduit par Dempster [3] lors de ses travaux sur les bornes inférieure et supérieure d'une famille de distributions de probabilités, Shafer [7] a mis en évidence l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines. Cette alternative aux probabilités subjectives a été démontrée de manière axiomatique par Smets [8] au travers du *Modèle de Croyances Transférables*.

3.1 Aspects théoriques

Soit Θ l'ensemble des N hypothèses solutions du problème. L'ensemble Θ , appelé *cadre de discernement*, est défini de la manière suivante :

$$\Theta = \{H_n\} \quad \forall n \in [1, N]. \quad (1)$$

On définit une masse de probabilité élémentaire, appelée *masse de croyance*, qui caractérise la véracité d'une proposition \mathcal{H} pour une source d'information S_j donnée pour tout $j \in [1, Q]$. La masse associée à cette source S_j que nous noterons m_j est alors définie par :

$$m_j : 2^\Theta \rightarrow [0, 1] \quad (2)$$

et vérifie les propriétés suivantes : $m_j(\emptyset) = 0$ et $\sum_{\mathcal{H} \subseteq \Theta} m_j(\mathcal{H}) = 1$. Cette probabilité se différencie d'une probabilité au sens classique du terme par le fait que la totalité de la masse de croyance est répartie non seulement sur les hypothèses singletons H_n mais aussi sur les hypothèses combinées \mathcal{H} . La modélisation issue de la fonction m_j est appelée jeu de masses. La principale difficulté consiste à modéliser les connaissances sur le problème en initialisant de manière adéquate [5] les fonctions de croyance m_j . Cette modélisation dépend généralement de l'application envisagée. Cependant, il est intéressant de remarquer que certaines modélisations [1] permettent de s'octroyer trois propriétés fondamentales, qui sont : la cohérence avec l'approche bayésienne dans le cas où les probabilités *a priori* sont connues, la séparabilité des hypothèses H_n et la cohérence avec l'association probabiliste des sources. Nous avons retenu un des modèles proposés par A. Appriou. Il est défini par :

$$m_{nj}(H_n) = \frac{q_{nj} \cdot R_j \cdot p(x_j/H_n)}{1 + R_j \cdot p(x_j/H_n)} \quad (3)$$

$$m_{nj}(\overline{H}_n) = \frac{q_{nj}}{1 + R_j \cdot p(x_j/H_n)} \quad (4)$$

$$m_{nj}(\Theta) = 1 - q_{nj}. \quad (5)$$

Dans ces équations, R_j est un facteur de normalisation contraint par $R_j \in [0, \max_{n \in [1, N]} p(x_j/H_n)^{-1}]$ où $p(x_j/H_n)$ représente la densité de probabilité de la mesure x_j issue de la source S_j sous les différentes hypothèses H_n . Ces densités pourront être déterminées à partir de la connaissance *a priori* de la distribution de probabilité ou à partir d'estimations calculées sur un ensemble d'apprentissage. Celles-ci sont donc plus ou moins représentatives des densités réellement rencontrées. Les coefficients q_{nj} caractérisent ici le degré de représentativité. Lorsque les densités sont parfaitement représentatives de l'apprentissage alors les coefficients q_{nj} sont égaux à 1 et dans ce cas les structures de croyances ne sont pas affaiblies. A l'inverse, lorsque la distribution de probabilité est totalement méconnue, ce qui est caractérisé par un coefficient q_{nj} égal à 0, les structures de croyance deviennent élément neutre de l'opérateur de combinaison de Dempster. Dans le cadre de la théorie de l'évidence de Dempster-Shafer, la fusion des informations issues de sources distinctes est réalisée en utilisant la *loi de combinaison de Dempster*. Celle-ci, qui s'avère commutative et associative, est définie par :

$$\forall \mathcal{H} \in 2^\Theta \quad m(\mathcal{H}) = m_1(\mathcal{H}) \oplus \dots \oplus m_Q(\mathcal{H}) \quad (6)$$

où \oplus représente l'opérateur de combinaison. Dans un cas à deux sources notées S_i et S_j , une combinaison conjonctive peut se mettre sous la forme :

$$(m_i \cap m_j)(\mathcal{H}) = \sum_{\mathcal{H}' \cap \mathcal{H}'' = \mathcal{H}} m_i(\mathcal{H}') \cdot m_j(\mathcal{H}''). \quad (7)$$

Une fois la masse résultante $m(\cdot)$ ainsi obtenue, la décision peut alors être prise. Différentes règles de décision ont été définies, les plus courantes étant la règle du maximum de plausibilité et la règle du maximum de crédibilité ou de probabilité pignistique [4]. Celles-ci consistent généralement à retenir l'hypothèse la plus vraisemblable parmi l'ensemble des hypothèses en compétition.

3.2 Méthodologie proposée

La méthodologie mise en oeuvre s'articule autour de trois principales étapes. Dans un premier temps, une modélisation des connaissances sur le problème est réalisée par l'intermédiaire de l'apprentissage des densités de

probabilité $p(x/H_n)$. Ensuite, de manière à qualifier la bonne représentativité de cette modélisation, on réalise un apprentissage des coefficients q_{nj} permettant ainsi d'obtenir les jeux de masses m_{nj} . Enfin, une partie finale de fusion des informations permet de construire la fonction de croyance utilisée pour la prise de décision.

3.2.1 Jeu de masses

La construction des jeux de masse m_{nj} (cf. équations (3) à (5)) passent par une première estimation des densités $p(x_j/H_n)$. De manière à estimer ces densités suffisamment précisément, on utilise une méthode d'estimation par noyau [6] sur les données de la base d'apprentissage. A partir de ces estimations et des équations (3), (4) et (5), il nous reste à estimer les coefficients q_{nj} de manière à obtenir les différents jeux de masses.

3.2.2 Paramètres d'apprentissage

De manière à essayer de qualifier la bonne représentativité de l'apprentissage, nous proposons d'utiliser une mesure d'information qui est calculée pour chaque couple (S_j, H_n) . Le but de la démarche est de permettre d'évaluer la fiabilité de la source S_j relativement à l'hypothèse H_n lorsque les conditions d'apprentissage ont évolué. Pour satisfaire ce principe, on s'appuie sur l'examen des données d'une base de validation et la comparaison entre ces données et les données de la base d'apprentissage. Plus précisément, nous proposons l'évaluation des q_{nj} par le calcul de la distance de Hellinger entre les données relatives à (S_j, H_n) sur la base d'apprentissage et les données relatives à (S_j, H_n) sur la base de validation. La comparaison entre ces deux jeux de données est réalisée en comparant l'histogramme optimal $\hat{\lambda}_{C_{opt}/H_n}^a$ calculé sur la base d'apprentissage et l'histogramme $\hat{\lambda}_{C_{opt}/H_n}^v$ calculé sur la base de validation. Ces histogrammes sont considérés optimaux au sens du principe de conservation de l'information [2]. Ainsi, $\hat{\lambda}_{C_{opt}/H_n}^a$ et $\hat{\lambda}_{C_{opt}/H_n}^v$ représentent respectivement les histogrammes optimaux pour un découpage en C_{opt} classes réalisés sur les données relatives à l'hypothèse H_n dans la base d'apprentissage a et la base de validation v . Ce découpage en C_{opt} classes, obtenu par le calcul d'un critère d'information du type Akaïke, permet d'obtenir les estimations des densités respectives quant à ces données. Enfin, une mesure de dissemblance entre ces deux estimations de lois peut être calculée en introduisant la distance de Hellinger [2]. On choisit de faire évoluer le coefficient q_{nj} en suivant l'équation suivante :

$$q_{nj} = 1 - Hell(\hat{\lambda}_{C_{opt}^j/H_n}^a, \hat{\lambda}_{C_{opt}^j/H_n}^v) \quad (8)$$

où $Hell$ représente la distance de Hellinger entre les deux approximations. L'obtention de ces différents coefficients permet la construction des jeux de masses relatifs à chaque hypothèse H_n et à chaque source d'information S_j .

3.2.3 Fusion et décision

A partir des différents jeux de masse $m_{nj}(\cdot)$, on réalise une première étape de fusion de manière à obtenir un jeu de masses unique $m_j(\cdot)$ pour chaque source d'informations S_j . Cette étape conduit à utiliser l'opérateur précédemment défini (équation (7)) qui peut s'écrire :

$$m_j(\mathcal{H}) = \bigcap_{n \in [1, N]} m_{nj}(\mathcal{H}) \quad \forall \mathcal{H} \in 2^\Theta. \quad (9)$$

Une seconde étape de fusion consiste à agréger les informations de chacune des sources S_j de manière à obtenir le jeu de masse final $m(\cdot)$. Celui-ci est calculé par :

$$m(\mathcal{H}) = \bigcap_{j \in [1, Q]} m_j(\mathcal{H}) \quad \forall \mathcal{H} \in 2^\Theta. \quad (10)$$

Enfin, la prise de décision se fait en utilisant le critère de maximum de plausibilité. Ce critère conduit à retenir l'hypothèse singleton H_n qui maximise la fonction de plausibilité définie par $Pl(\mathcal{H}) = \sum_{(\mathcal{H} \cap \mathcal{H}') \neq \emptyset} m(\mathcal{H}')$ parmi l'ensemble des hypothèses en compétition.

4 Résultats

Dans cette section, nous présentons les résultats obtenus par la méthode sur les données du réseau de surveillance de la ville de Lyon. L'événement à prédire est de type binaire conduisant ainsi à un cadre de discernement Θ à deux hypothèses. La constitution des bases de données a été faite de manière chronologique. Les années 1994 à 1997 (488 jours) ont été choisies comme base d'apprentissage. Les bases de validation et de test correspondent respectivement aux années 1997 et 1998 représentant 245 (229 non pollués et 16 pollués) et 164 jours (152 non pollués et 12 pollués). En ce qui concerne les informations, nous avons choisi de travailler avec 3 sources qui sont respectivement le géopotential (S_1), la température (S_2) et la concentration en ozone (S_3) maximales calculées le

TAB. 1 – Résultats sur la base d’apprentissage

Observés	Prédits	
	Non Pollué	Pollué
Non Pollué	687	81
Pollué	8	38

TAB. 2 – Résultats sur la base de test

Observés	Prédits	
	Non Pollué	Pollué
Non Pollué	132	20
Pollué	1	11

jour précédent. Les tableaux Tab. 1 et Tab. 2 représentent les pourcentages de classification sur chacune de ces bases de données. On peut noter le taux de non-détection (8 et 1 non-détections respectivement pour la base d’apprentissage et la base de test) reste faible. En ce qui concerne le taux de fausse alarme, il reste élevé (81 et 20 fausses alarmes respectivement). Ces résultats permettent néanmoins d’augmenter sensiblement le taux de bonne classification dans le cas où les coefficients q_{nj} sont tous égaux à 1.

5 Discussion

Le tableau Tab. 3 présente les valeurs respectives des différents coefficients de fiabilité pour les trois sources et les deux hypothèses. Il permet de remarquer que le coefficient relatif à la source S_3 dans le cas de l’hypothèse

TAB. 3 – Coefficients q_{nj} pour $n \in [1, N]$ et $j \in [1, Q]$; N=2 hypothèses, Q=3 sources

Hypothèse H_n	Source S_j		
	S_1	S_2	S_3
Non Pollué	0.999	0.997	0.970
Pollué	0.974	0.978	0.918

de jour pollué est plus faible, ceci étant dû à la mauvaise représentativité de l’apprentissage et imputable au fait que le nombre d’exemples relatifs à cette hypothèse est faible comparativement à l’hypothèse de jour non pollué. Ce fait est particulièrement sensible sur les données relatives à l’ozone (source S_3) où les mesures de concentration sont très imprécises. Ce résultat permet de montrer l’utilité d’un apprentissage suffisamment précis des coefficients d’affaiblissement des fonctions de croyance et d’obtenir ainsi une amélioration des performances du système de surveillance. Il est à noter que la décision peut être étoffer d’un indice de confiance directement accessible par l’intermédiaire de la fonction de croyance. Cette confiance dans la décision constitue un des atouts majeurs de la méthodologie développée en vue de l’exploitation du résultat proposé.

Références

- [1] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, 11 :27–40, 1991.
- [2] O. Colot, C. Olivier, P. Courtellemont, A. El-Matouat, and D. de Brucq. Information criteria and abrupt changes in probability laws. In M. Holt, C. Cowan, P. Grant, and W. Sandham, editors, *Signal Processing VII : Theories and Applications*, pages 1855–1858. EUSIPCO’94, September 1994.
- [3] A. Dempster. Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics*, AMS-38 :325–339, 1967.
- [4] T. Denoeux. Analysis of evidence-theory decision rules for pattern classification. *Pattern Recognition*, 30(7) :1095–1107, 1997.
- [5] E. Lefevre, O. Colot, and P. Vannoorenberghe. Using information criteria in dempster-shafer’s basic belief assignment. In *International Conference on Fuzzy Systems, FUZZ-IEEE’99*, Seoul, KOREA, 1999.
- [6] G. Saporta. *Probabilités, Analyse des Données et Statistique*. Editions Technip, 1990.
- [7] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [8] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2) :191–234, 1994.