Classification par la théorie de l'évidence pour la gestion de tournée de véhicules

Clustering using evidence theory for vehicle routing problem

E. Lefevre J.P. Manata D. Jolly

Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A)
Université d'Artois, Faculté des Sciences Appliquées
Technoparc Futura, 62400 Béthune Cedex
Eric.Lefevre@iut-geii.univ-artois.fr

Résumé

Dans de ce papier, nous abordons le problème de routage de véhicules. Ce problème consiste à obtenir le meilleur chemin lors d'une tournée de véhicules. Toutefois, les techniques donnant le chemin optimal sont relativement gourmandes en temps de calcul lorsque le nombre de points à desservir est important. Afin de pallier à cette difficulté, des méthodes basées sur des heuristiques ont vu le jour. La technique envisagée dans ce papier consiste à décomposer le problème en sous problème afin d'en réduire la complexité. La décomposition du problème repose sur une technique de classification basée sur la théorie de l'évidence. Une fois les groupes définis, nous pouvons employer les méthodes de recherche de chemin optimal au sein de chaque groupe. Dans ce papier, nous présentons des tests sur des données synthétiques montrant le bon comportement de la méthode de classification employée. De plus, des tests et des comparaisons de l'approche proposée avec différentes heuristiques sont présentées sur des données utilisées couramment dans le cadre du problème de routage.

Mots Clef

Théorie de l'évidence, Classification, Routage de Véhicules.

Abstract

In this paper, we present the vehicle routing problem. This problem consists in obtaining the best way of a vehicle round. However, the techniques giving the optimal way are relatively greedy in computing times when the number of points to be served is significant. In order to reduce this problem, different methods based on the heuristic were used. The technique proposed in this paper consists in breaking down the problem into sub-problem to reduce the complexity. In order to break down the problem, we employ a classification technique based on the evidence theory. Once the groups defined, we can employ the research methods to find an optimal way within each group. In this paper, we present tests on synthetic data to illustrate the good behaviour of the classification method employed. Moreover, tests and comparisons between the approach proposed and different heuristic are presented on data usually used in the framework of vehicle routing problem.

Keywords

Evidence Theory, Clustering, Vehicle Routing Problem.

1 Introduction

Dans le domaine des transports (logistique), l'une des thématiques les plus abordées dans la littérature, concerne la recherche du trajet optimal dans le cadre de tournées de véhicules [1, 2, 16]. Les applications de cette thématique sont nombreuses : collecte/distribution de courrier postal auprès de clients, tournée de dépanneurs, ... Le trajet recherché doit alors satisfaire plusieurs contraintes telles que la minimisation du temps de route, des kilomètres parcourus, le temps d'attente des clients,... Ce problème couramment appelé VRP[1] est un problème NP-difficile. Il en résulte que les différents algorithmes proposés, afin d'obtenir le parcours optimal, sont relativement gourmands en temps de calcul dans le cas d'un problème de grande taille (c-à-d avec un nombre de clients important).

L'une des solutions qui est envisagée, dans le cadre de cet article, consiste à décomposer le problème en sous problèmes par un partitionnement géographique des clients. L'une des difficultés consiste alors à déterminer automatiquement le nombre de partitions. Pour cela, nous nous reposons sur des approches de classification automatique appelé aussi classification non supervisée ou catégorisation. Le problème de la classification automatique peut être formulé de la manière suivante : étant donné un ensemble fini

[1] Pour Vehicle Routing Problem

$\mathcal{X}$ d'objets non étiquetés (aussi appelé base de test), l'objectif est d'affecter à chaque objet, noté x, une étiquette afin de regrouper ces objets en différentes partition (ou classes) $\{\omega_1, \ldots, \omega_M\} = \Omega$ représentatives d'une certaine structure de l'ensemble $\mathcal{X}$. Dans notre cas, la valeur de M est inconnue. Une classe est constituée, par conséquent, des objets de $\mathcal{X}$ qui sont plus proches les uns des autres, au sens d'un certain critère de similarité, qu'ils ne le sont des objets des autres classes.

Plusieurs méthodes de classification automatique ont vu le jour depuis de nombreuses années : Nuées dynamique [13], Fuzzy C-means [4, 5, 15],... Une autre approche reposant sur la théorie de l'évidence, aussi appelée théorie de Demspter-Shafer, a été élaborée par Denoeux [12]. Cette approche, de part les fondements de la théorie de l'évidence, permet d'associer à chaque objet x :
– un degré de confiance vis-à-vis de son appartenance à l'une des classes,
– un degré d'ignorance quant à son affectation aux classes de Ω.

C'est cette dernière technique que nous avons choisi d'utiliser. L'article est organisé de la manière suivante. Tout d'abord, on rappelle, dans la section 2, quelques principes fondamentaux de la théorie de l'évidence. Ces rappels vont nous permettre d'introduire la démarche envisagée dans le cadre de cette théorie pour la classification automatique présentée dans la section 3. Quelques résultats sur des jeux de données synthétiques sont présentés afin de montrer le comportement de la méthode de classification. Enfin, nous illustrons et nous comparons l'approche proposée dans cet article à des heuristiques couramment utilisées dans le cadre de la gestion de tournée de véhicules (section 4).

2 La théorie de l'évidence

Nous rappelons dans cette section quelques éléments théoriques relatifs aux fonctions de croyance. La théorie de l'évidence fut initialement introduite par Dempster [8] dans le cadre ses travaux sur les bornes inférieure et supérieure d'une famille de distributions de probabilités. A partir de ce formalisme mathématique, Shafer [22] a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines. L'utilité des fonctions de croyance, comme alternative aux probabilités subjectives, a été démontrée plus tard de manière axiomatique par Smets [24, 25] au travers du *Modèle de Croyances Transférables*.

Dans le cadre de la théorie de l'évidence, l'ensemble Ω est appelé *cadre de discernement*. On suppose que les hypothèses dans Ω sont exclusives et que le cadre de discernement est exhaustif. On définit une masse de probabilité élémentaire, appelée *masse de croyance*, qui caractérise la véracité d'une proposition A pour une source d'information S. La masse m associée à cette source est alors définie par :

$$m : 2^{\Omega} \rightarrow [0, 1] \tag{1}$$

et vérifie les propriétés suivantes :

$$m(\emptyset) = 0 \tag{2}$$

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{3}$$

Cette fonction se différencie d'une probabilité par le fait que la totalité de la masse de croyance est répartie non seulement sur les hypothèses singletons ω_q, mais aussi sur les hypothèses combinées. La modélisation issue de la fonction m est appelée jeu de masses. A partir de la fonction m, on définit respectivement les fonctions de *crédibilité* Cr et de *plausibilité* Pl par :

$$Cr(A) = \sum_{B \subseteq A} m(B) \tag{4}$$

$$Pl(A) = \sum_{(A \cap B) \neq \emptyset} m(B) = 1 - Cr(\overline{A}) \tag{5}$$

où $\overline{A}$ représente l'événement contraire de la proposition A. La crédibilité $Cr(A)$ mesure la force avec laquelle on croit en la véracité de la proposition. La plausibilité $Pl(A)$, fonction duale de la crédibilité, mesure l'intensité avec laquelle on ne doute pas de A.

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 :

$$\forall\, A \subseteq \Omega \qquad m(A) = m^1(A) \oplus \ldots \oplus m^J(A) \tag{6}$$

où $\oplus$ représente l'opérateur de combinaison. Dans un cas à deux sources notées S^i et S^j, la combinaison peut se mettre sous la forme :

$$m(A) = \frac{1}{1-\kappa} \sum_{(B \cap C) = A} m^i(B).m^j(C) \tag{7}$$

où κ est défini par :

$$\kappa = \sum_{(B \cap C) = \emptyset} m^i(B).m^j(C). \tag{8}$$

Dans l'équation (7), le coefficient κ reflète le conflit existant entre les deux sources S^i et S^j et le quotient $\frac{1}{1-\kappa}$ est un terme de normalisation. Lorsque κ est égal à 1, les sources sont en conflit total et les informations ne peuvent être fusionnées. Au contraire, lorsque κ est nul, les sources sont en parfait accord. Cette règle de fusion, déduite de la règle de conditionnement [23], a été critiquée dans plusieurs travaux dont [27], en particulier dans le cas de sources en conflit total.

Pour la prise de décision, on suppose que nous avons une fonction de croyance m sur Ω qui résume l'ensemble des

informations apportées. La décision consiste à choisir une action a parmi un ensemble fini d'actions $\mathcal{A}$. Une fonction de perte $\lambda : \mathcal{A} \times \Omega \rightarrow \mathbb{R}$ est supposée définie de telle manière que $\lambda(a, \omega)$ représente la perte encourue si l'on choisit l'action a lorsque la vraie classe est ω. A partir d'arguments de rationalité, Smets [25] propose de transformer m en une fonction de probabilité P_{Bet} sur Ω (appelée fonction de probabilité pignistique) définie pour tout $\omega \in \Omega$ comme :

$$P_{Bet}(\omega) = \sum_{A \ni \omega} \frac{m(A)}{|A|}, \tag{9}$$

où $|A|$ est la cardinalité de $A \subseteq \Omega$. Dans cette transformation, la masse de croyance $m(A)$ est uniformément distribuée parmi les éléments de A. A partir de cette probabilité, on peut associer à chaque action $a \in \mathcal{A}$ un risque défini comme le coût espéré relatif à P_{Bet} si on choisit a :

$$R(a) = \sum_{\omega \in \Omega} \lambda(a, \omega) P_{Bet}(\omega). \tag{10}$$

L'idée consiste ensuite à choisir l'action qui minimise ce risque, généralement appelé risque pignistique. D'autres risques, basés sur les fonctions de crédibilité et de plausibilité, ont été développés et étudiés dans [6, 10].

3 Méthode de catégorisation par la théorie de l'évidence

Dans cette section, nous allons décrire le principe de la méthode de classification automatique fondée sur la théorie de l'évidence proposée dans [12].
Toutefois, avant d'aborder cette démarche et afin d'être plus clair, nous présentons, dans un premier temps, une méthode de classification supervisée, elle aussi basée sur les fonctions de croyance, qui est à l'origine de la méthode de classification automatique qui nous intéresse. Cette approche supervisée a été présentée dans [9, 11, 29].

3.1 Approche supervisée

L'objectif de ce type d'approche est d'associer à un objet x une classe parmi un ensemble Ω de M classes défini par :

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\} \tag{11}$$

Cette association est réalisée à partir d'un ensemble d'apprentissage $\mathcal{L}$ de N exemples de classe connue. Chacun des exemples d'apprentissage pourra alors être considéré comme une part de croyance quant à l'appartenance de x à l'une des classes de l'ensemble Ω. Ce degré de croyance peut-être assimilé à une fonction de croyance m^i qui possède 2 éléments focaux : la classe de x^i notée ω_q et Ω. Ainsi, si nous considérons que l'objet x^i est proche de x, alors une partie de la croyance sera affectée à ω_q et le reste à l'ensemble des hypothèses du cadre de discernement. La fonction de croyance peut-être alors obtenue à l'aide d'une fonction décroissante de la distance de la forme :

$$\begin{cases} m^i(\{\omega_q\}) & = \alpha \phi_q(d^i) \\ m^i(\Omega) & = 1 - \alpha \phi_q(d^i) \end{cases} \tag{12}$$

où $0 < \alpha < 1$ est une constante, $\phi_q(.)$ est une fonction décroissante monotone vérifiant :

$$\phi_q(0) = 1 \tag{13}$$

$$\lim_{d \rightarrow \infty} \phi_q(d) = 0 \tag{14}$$

avec d^i la distance euclidienne entre le vecteur x et le $i^{\text{ème}}$ vecteur de la base d'apprentissage $\mathcal{L}$. La fonction ϕ_q peut être une fonction exponentielle de la forme :

$$\phi_q(d^i) = \exp(-\gamma_q (d^i)^2) \tag{15}$$

où γ_q est un paramètre associé à la $q^{\text{ème}}$ classe. Le paramètre α empêche l'affectation de toute la masse de croyance à l'hypothèse ω_q lorsque x et le $i^{\text{ème}}$ prototype sont égaux. En outre, la contrainte $\alpha < 1$ garantit la possibilité de combiner m_i avec n'importe quelle autre fonction de croyance puisque quel que soit d^i, on aura toujours $m^i(\Omega) > 0$ (la certitude de ω_q pourrait entraîner un conflit total avec une autre source de croyance incompatible). Le paramètre γ_q, quant à lui, permet de spécifier la vitesse de décroissance de la fonction de masse. Ainsi, les paramètres γ_q peuvent permettre d'intégrer une information concernant la dispersion de chaque classe ω_q. Dans le cadre de l'approche supervisée, plusieurs méthodes ont été élaborées afin de définir ces paramètres [18, 28].
On obtient ainsi N fonctions de croyance $m^1, m^2, \ldots, m^N$ qu'il suffit de combiner à l'aide de l'équation (7).

3.2 Cas de la classification non supervisée

Nous allons montrer maintenant que l'approche présentée ci-dessus peut-être étendue à un apprentissage non supervisé. Nous considérons, dans un premier temps, l'ensemble $\mathcal{X}$ de N objets $x^1, x^2, \ldots, x^N$ qui ont été préalablement répartis en M classes $\omega_1, \omega_2, \ldots, \omega_M$. Cette partition de l'ensemble $\mathcal{X}$ est notée $\mathcal{P}$, et l'étiquette de l'objet $x^i \in \mathcal{X}$ est noté $L^i \in \{1, \ldots, M\}$. Soit $\mathcal{R}$, la règle de décision présentée précédemment. Supposons que le vecteur x^i soit associé à la classe $q = \mathcal{R}(x^i)$ en utilisant les $(N-1)$ objets de $\mathcal{X}$. Si $L^i = q$, alors on peut considérer que la règle $\mathcal{R}$ classe correctement l'objet x^i. Si, l'ensemble des éléments de $\mathcal{X}$ est correctement classé alors on peut considérer que la partition $\mathcal{P}$ est stable.
On peut alors considérer la procédure suivante pour la méthode de classification non-supervisée. Dans un premier temps, on réalise une partition initiale aléatoire. On associe à chaque vecteur de l'ensemble $\mathcal{X}$ une classe. On obtient alors autant de classes que de vecteur à la première itération. Ensuite, on classe chaque vecteurs x^i à l'aide de la règle $\mathcal{R}$. Si $\mathcal{R}(x^i) = L^i$ alors l'étiquette L^i reste inchangée sinon elle prend la valeur $\mathcal{R}(x^i)$. On répète alors ces opérations, jusqu'à ce que l'on obtienne une partition stable, c'est-à-dire lorsque les étiquettes des vecteurs x^i ne changent plus. Dans [12], il a été montré que cette démarche aboutit à une stabilisation de la partition en un nombre d'itérations fini. Dans la section suivante, nous reprenons cette démonstration.

3.3 Convergence de la méthode

Afin de calculer la fonction de masse m^i accordée à un objet x^i, nous utilisons l'ensemble des objets x^j de $\mathcal{X}$ tel que $j \neq i$. Ainsi chaque objet x^j, associé à une classe $L^j = q$ fournit une fonction de croyance $m^{i,j}$ concernant l'affectation de l'objet x^i à la classe q. Cette part de croyance sera obtenue à l'aide du système d'équations (12). Toutefois, au contraire de l'approche supervisée, il est impossible de déterminer les paramètres γ_q propre à chaque classe car dans notre cas nous n'avons aucune connaissance sur les classes (ni le nombre, ni la forme). Nous considérons, alors un paramètre unique à l'ensemble des classes, que nous notons : γ. Nous obtenons alors les équations suivantes :

$$\begin{cases} m^{i,j}(\{\omega_q\}) & = \alpha\phi(d^{i,j}) \\ m^{i,j}(\Omega) & = 1 - \alpha\phi(d^{i,j}) \end{cases} \tag{16}$$

où $d^{i,j}$ est la distance euclidienne entre l'objet x^i et l'objet x^j. A l'aide de cette équation, on peut définir $N-1$ fonctions de croyance où chacune de ces fonctions quantifie l'appartenance de l'objet x^i à la classe ω_q qui correspond à la classe de l'objet x^j. Ces fonctions sont combinées à l'aide de la règle de Dempster présentée à l'équation (7). Dans ce cas particulier, les fonctions de croyance ne possèdent que deux éléments focaux (L^j, Ω), la combinaison aboutit à une fonction de croyance m^i définie par :

$$m^i(\Omega) = \frac{1}{K} \prod_{j \neq i} T_{i,j} \tag{17}$$

$$m^i(\{\omega_q\}) = \frac{1}{K} \left(\prod_{\{j \neq i, L^j \neq q\}} T_{i,j} - m^i(\Omega) \right) \tag{18}$$

quel que soit $w_q \in \Omega$ et avec :

$$T_{i,j} = 1 - \alpha\phi(d^{i,j}) \tag{19}$$

Dans l'équation (18), K représente une constante de normalisation basée sur le conflit définit de la manière suivante :

$$K = \sum_{A \in \Omega} m^i(A). \tag{20}$$

La démonstration de convergence de la méthode présentée dans [12] s'inspire d'une analogie formelle avec le modèle connexionniste introduit par Hopfield [17]. Nous considérons alors la fonction de coût suivante :

$$E = \frac{1}{2} \sum_{q=1}^{M} \sum_{\{i \neq j \ | L^i = L^j\}} \omega_{i,j} \tag{21}$$

où $\omega_{i,j} = -\ln T_{i,j}$. Selon les notations employées par Shafer [22], $\omega_{i,j}$ représente le poids de l'évidence de l'appartenance de x^i à la classe de l'objet x^j. De plus, cette matrice de poids est symétrique c'est-à-dire que $w_{i,j} = w_{j,i}$. Nous allons, maintenant, montrer que chaque changement de classe d'un objet augmente strictement la valeur de E. Considérons que l'objet x^i change de classe, passant de la classe q à la classe p. Alors, la différence relevée sur la valeur de E s'exprime alors de la manière suivante :

$$\Delta E = \sum_{\{j \neq i | L^j = p\}} \omega_{i,j} - \sum_{\{j \neq i | L^j = q\}} \omega_{i,j} \tag{22}$$

Le changement de l'étiquette L^i de q en p, signifie qu'au niveau de la prise de décision la masse de croyance accordée à la classe ω_p est plus importante que celle attribuée à la classe ω_q :

$$m^i(\{\omega_q\}) < m^i(\{\omega_p\}) \tag{23}$$

Ce qui peut s'exprimer de la manière suivante :

$$\prod_{\{j \neq i, L^j \neq p\}} T_{i,j} > \prod_{\{j \neq i, L^j \neq q\}} T_{i,j} \tag{24}$$

$$\Leftrightarrow \sum_{\{j \neq i, L^j = p\}} w_{i,j} > \sum_{\{j \neq i, L^j = q\}} w_{i,j} \tag{25}$$

Par conséquent, la différence de la fonction de coût, définie par l'équation (22), consécutif à un changement de classe ne peut-être que positive $\Delta E > 0$. La fonction de coût E ne peut donc être que strictement croissante. Ceci démontre bien la convergence de la méthode de classification non-supervisée basée sur la théorie de l'évidence.

3.4 Implémentation de la méthode et estimation des paramètres

Cette méthode pose des problèmes d'implémentation. En effet, cette approche nécessite le calcul de $N-1$ fonctions de croyance à chaque itération. Si le nombre d'objets dans l'espace est très important, l'approche devient alors très coûteuse en temps de calcul. Toutefois, la complexité de cette approche peut être réduite en utilisant simplement les k plus proches voisins d'un objet x^i. En effet, on peut considérer que les fonctions de croyance obtenues par des objets trop éloignés apportent peu d'information (c-a-d $m(\Omega) = 1$) et peuvent-être négligées car elles correspondent à l'élément neutre de la règle de combinaison de Dempster. Malgré cette modification, l'approche proposée concerne ses propriétés de convergence [12].
Un autre aspect de la classification repose sur la détermination du nombre de classes. Nous avons vu, dans l'approche proposée, que le nombre de classes ne peut que diminuer par rapport au nombre initial. Ainsi la stratégie, qui a fait ses preuves dans la pratique, consiste à considérer une partition initiale pour laquelle il y a autant de classes que d'objets. Ainsi à l'étape initiale de cette méthode, il n'est pas nécessaire de connaître le nombre de classes.
Toutefois, cette approche nécessite la connaissance de plusieurs paramètres. Ainsi, si l'on considère la fonction :

$$\phi(d) = \exp(-\gamma d^2) \tag{26}$$

qui nous permet de construire les fonctions de masses, il nous reste à définir la valeur de α, de γ, et le nombre de voisins k.

La valeur de α n'a que très peu d'influence car elle est identique à toutes les fonctions de croyance.
Par contre, le paramètre γ, bien qu'identique pour l'ensemble des fonctions de croyance, influe sur la dispersion des classes. Afin de définir ce paramètre nous adaptons une heuristique proposée dans [28]. Pour notre approche, le paramètre γ sera déterminé à l'aide de la relation suivante :

$$\gamma = \frac{1}{d^2} \tag{27}$$

où d représente la distance moyenne entre les différents objets de $\mathcal{X}$. Ainsi défini, ce paramètre permet, de prendre en compte la dispersion des objets de $\mathcal{X}$.
Nous verrons par la suite, que le nombre de voisins k peut aussi être déterminé par une heuristique de manière assez robuste.

4 Résultats

Dans cette section, nous présentons, dans un premier temps, des tests sur des données synthétiques afin de valider la méthode de classification. Enfin, nous utilisons cette méthode dans le cadre du routage de véhicules. Cette approche est comparée à des approches classiques de routage de véhicule sur des données communément employées pour cette application.

4.1 Données synthétiques

De façon à illustrer l'approche de classification, nous présentons un test sur des données non-étiquetées générées artificiellement à partir de 4 distributions gausiennes dans un espace de dimension 3. Chaque distribution normale a pour moyenne μ_q et variance Σ_q :

$$\begin{aligned}
\mu_1 &= (0,0,0)^t \qquad \Sigma_1 = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0 & 0.7 \end{pmatrix} \\
\mu_2 &= (2,3,2)^t \qquad \Sigma_2 = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0 & 0.7 \end{pmatrix} \\
\mu_3 &= (-2,-2,-2)^t \quad \Sigma_3 = \begin{pmatrix} 0.9 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.7 \end{pmatrix} \\
\mu_4 &= (4,2,-2)^t \qquad \Sigma_4 = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned} \tag{28}$$

Chaque distribution gaussienne est constituée de 50 vecteurs. La figure 1 présente la répartition des objets dans l'espace.
Nous avons, dans un premier temps, étudié l'influence du nombre de voisins sur la partition obtenue. On peut constater sur la figure 2, que le nombre de classes dans la partition finale varie en fonction de k. En effet, si le nombre de voisins est trop faible, le nombre de classes obtenu est

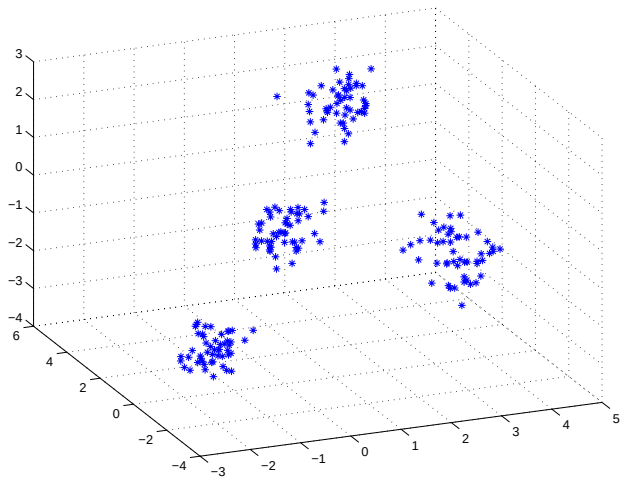

FIG. 1 – *Base de données synthétiques non étiquetées.*

plus important que celui souhaité sur notre exemple. Par contre, lorsque le nombre de voisins atteint un certain seuil le nombre de classes obtenu arrive à un minimum. Ceci s'explique par le fait que les voisins éloignés d'un vecteur i ont peu ou pas d'influence sur la fonction de croyance m^i finale. Dans l'exemple que nous avons pris, le minimum correspond au nombre de classes souhaitées (ici 4). L'objectif est alors de définir de manière suffisamment gé-

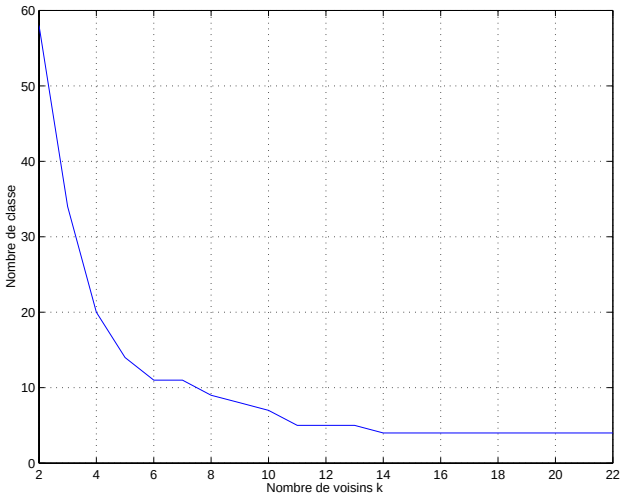

FIG. 2 – *Evolution du nombre de classes en fonction du nombre de voisins k.*

nérique la valeur de ce seuil. Pour cela, nous nous inspirons des travaux réalisés sur la méthode des *k* plus proches voisins (*k-ppv*). Pour cette méthode, différentes heuristiques ont été mises en oeuvre afin de définir la valeur de k. Nous avons choisi celle présentée dans [14] qui défini le nombre de voisins de la manière suivante :

$$k = \left[\sqrt{(N)}\right]_{ent} \tag{29}$$

où $[.]_{ent}$ représente la partie entière. La figure 3 représente la classification obtenue sur les données de la figure 1 en prenant 14 voisins. Nous pouvons ainsi constater que l'heuristique présentée permet de donner des résultats satisfaisants sur les données que nous avons généré. Par ailleurs, nous avons eu l'occasion de tester cette heuristique sur d'autres jeux de données et ainsi de valider sa robustesse. La figure 4 illustre l'évolution du coût E de l'équation (21).

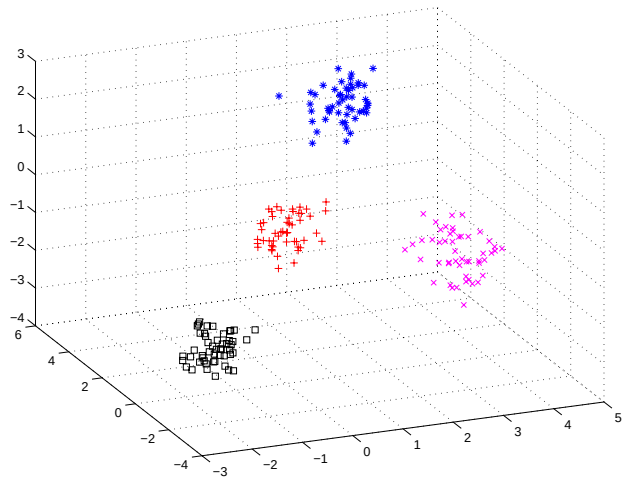

FIG. 3 – *Résultat de la classification évidentielle sur la base de données synthétiques avec* $k = 14$.

Cette figure montre ainsi que la fonction E est monotone croissante et qu'elle converge vers un extremum qui définit la stabilisation de la partition comme cela l'avait été démontré dans la section 3.3.

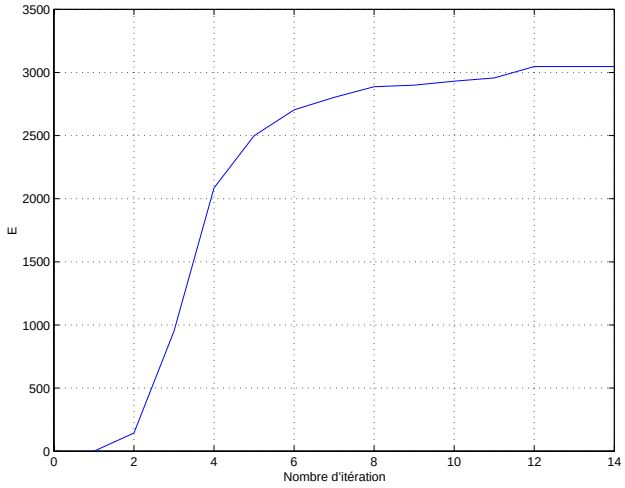

FIG. 4 – *Evolution du critère* E *pour* $k = 14$ *en fonction du nombre d'itérations.*

4.2 Application au routage de véhicule

A titre d'illustration nous appliquons notre démarche à un exemple classique de tournée de véhicules. Il s'agit dans ce cadre de faire desservir successivement par un véhicule un ensemble de nœuds géographique (ou villes) en minimisant la distance de parcours. Nous proposons de jumeler notre approche de classement non supervisé avec une recherche exacte du chemin minimum de type "branch and bound" [21].

Les méthodes "branch and bound" construisent un arbre de décision contenant uniquement la ou les solutions optimales. Dans l'application tournée de véhicule chaque déplacement possible d'un nœud à un autre devient une branche de l'arbre des parcours possibles. Chaque fois qu'une nouvelle branche est ajoutée en cours de construction, la distance déjà parcourue est évaluée. Si cette distance est trop importante alors la construction est interrompue (l'arbre est élagué) et le processus de construction reprend pour une autre branche. Dans ce type de construction, on conserve à tout instant une séquence de référence dans l'optique de la comparer à une seconde séquence en cours de construction. Malgré "l'élagage" les méthodes "branch and bound" ne peuvent pas en des temps raisonnables résoudre des problèmes qui dépassent la dizaine ou la vingtaine de nœuds à visiter. Nous contournons ce problème en proposant une heuristique en plusieurs étapes : Tout d'abord nous classons les points à desservir via l'algorithme de classement non supervisé. Nous appliquons alors la méthode "branch and bound" sur les centres d'inertie des classes plutôt que sur l'ensemble des points. Nous obtenons ainsi un parcours liant les classes entre elles du point de vue chronologique.

Il nous est possible alors d'appliquer la méthode "branch and bound" au sein de chaque classe. Pour la $i^{\text{ème}}$ classe nous recherchons un chemin dont le point d'entrée est le centre d'inertie de la $(i-1)^{\text{ème}}$ classe, dont le point de sortie est le centre d'inertie de la $(i+1)^{\text{ème}}$ classe et qui dessert tous les points de la classe i. La solution optimale du point de vue de la distance est retenue. Il est à noter qu'une fois la chronologie des classes connue, la recherche du chemin optimum au sein de chaque classe est indépendante et donc recherchée si besoin via un traitement parallèle.

Nous illustrons cette approche avec le jeu de données C101 issu de Solomon [26] constitué de 100 clients et d'un dépôt. La figure 5 représente la répartition des clients dans l'espace pour ce jeu de données. Sur cette figure le point situé aux coordonnés $(40, 50)$ représente le dépôt c'est-à-dire le point de départ du véhicule. La première étape à entreprendre, afin de mettre en œuvre la méthode présentée, consiste à réaliser un partitionnement. Celui-ci ne prend pas en compte le point correspondant au dépôt. Ainsi en référence à l'heuristique présentée à l'équation (29), nous prenons un nombre de voisins k égale à 10. Le résultat du partitionnement obtenu à l'aide de l'approche présentée est illustré sur la figure 6. La disposition particulière de ce jeu de données permet ici une expertise humaine simple qui valide le résultat.

Une fois le partitionnement obtenu, la seconde étape de la stratégie employée consiste à définir un parcours entre

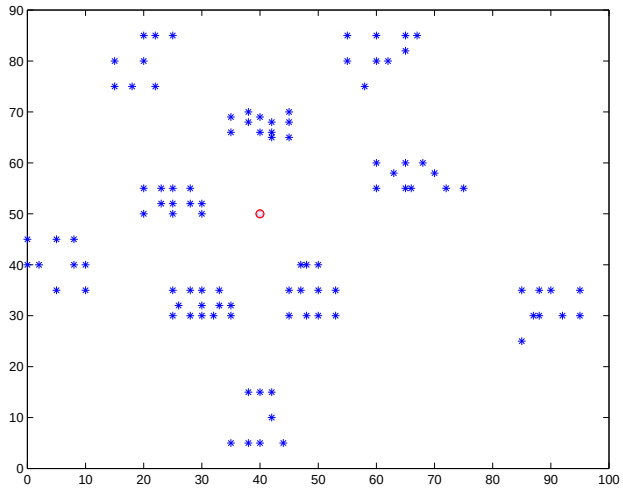

FIG. 5 – *Répartition dans l'espace des clients pour la base de données de Solomon.*

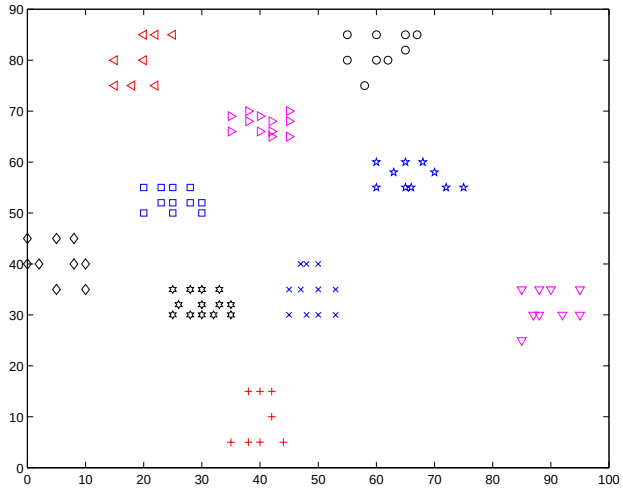

FIG. 6 – *Résultat du partitionnement sur la base de données de Solomon.*

les différentes classes. Pour cela, nous utilisons la méthode dite de "branch and bound" dans un espace réduit au dépôt et aux centres d'inertie des classes. Le chemin optimum dans ce cas est présenté sur la figure 7.

Reste alors à appliquer la dernière étape c'est-à-dire trouver le chemin optimum au sein de chaque classe comme décrit précédemment. Le parcours final résulte de la juxtaposition du chemin global (entre centres d'inertie de classes) et les chemins locaux (au sein de chaque classe). Ce parcours est représenté sur la figure 8.

Nous avons comparé l'approche proposée dans ce papier à différentes heuristiques communément employées dans la littérature. Celles-ci sont :

– La méthode du plus proche voisin (NN)[2] : l'algorithme déplace le véhicule systématiquement vers le client, non encore desservi, le plus proche géographiquement [3].

[2]Pour Nearest Neighbour.

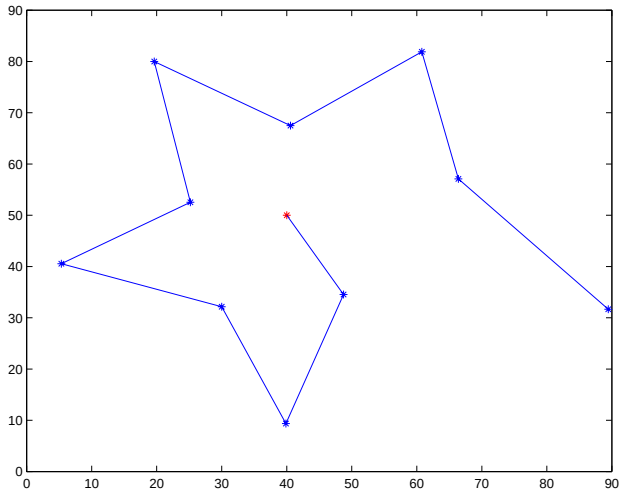

FIG. 7 – *Résultat de la stratégie "branch and bound" appliquées au centre d'inertie de chaque classes.*

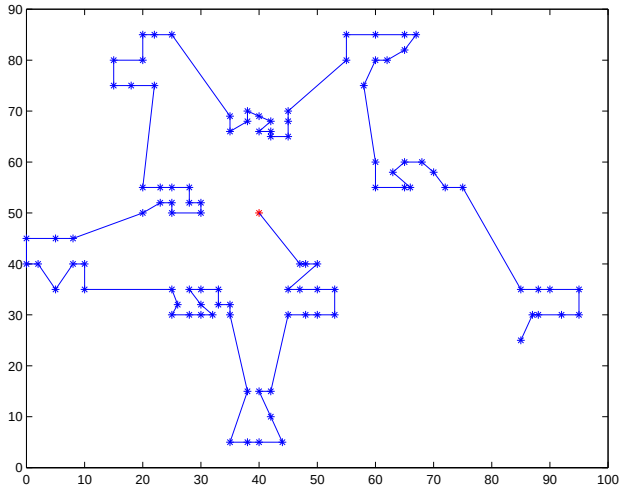

FIG. 8 – *Chemin obtenu avec la stratégie proposée dans ce papier.*

– Les méthodes de descente (ou k-opt) : ces méthodes utilisent une solution possible initiale (souvent obtenue par la méthode NN) qu'elles améliorent par itérations successives. Elles exploitent la notion de voisinage entre solutions : lorsqu'un nombre k de permutation dans la desserte des clients permet de passer d'une solution à une autre celles-ci sont dites voisines. A chaque pas ces méthodes vont alors explorer le voisinage de la solution courante et retenir la meilleure solution, s'il y a lieu, parmi ce voisinage. C'est la valeur du nombre de permutations k qui distingue les performances mais également la rapidité d'exécution de ces méthodes. Nous avons retenu 2-opt et 3-opt qui suffisent à obtenir de bons résultats pour le jeu de données utilisé [7, 19].

Les résultats de ces différentes techniques sont présentées

sur les figures 9, 10 et 11.

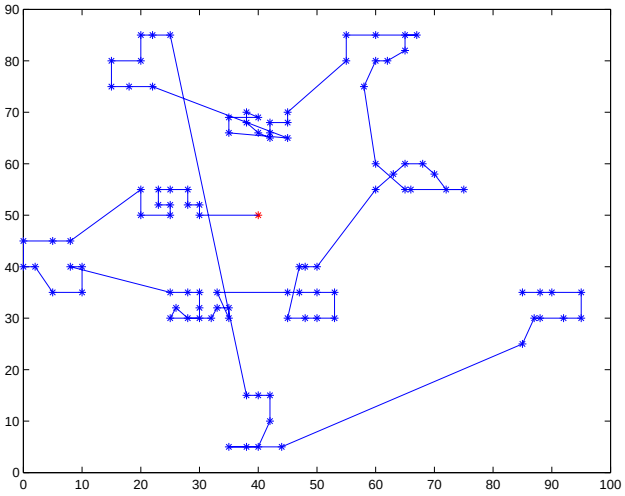

FIG. 9 – *Chemin obtenu avec la stratégie NN.*

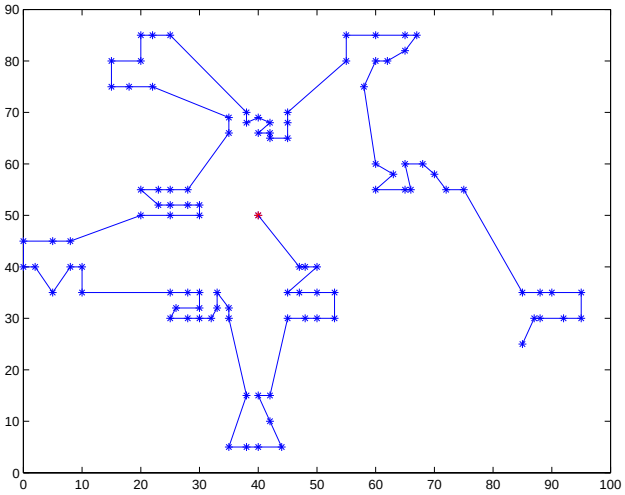

FIG. 10 – *Chemin obtenu avec la stratégie 2-opt.*

Enfin dans le tableau 1, un récapitulatif permet de comparer notre résultat à ceux obtenus avec ces heuristiques classiques.
Nous constatons que l'ordre de grandeur obtenu nous est favorable et voisin de celui obtenu avec des méthodes de descente qui sont réputées pour leurs performances dans ce type de problème.

5 Conclusion et perspectives

Les problématiques en rapport avec le routage de véhicules sont nombreuses. Nous nous sommes attachés à montrer dans cet article, que la théorie de l'évidence peut être un outil intéressant pour la résolution de celles-ci.
Nous utilisons une approche de classification basée sur cette théorie dans le cadre d'un problème de type VRP statique[3]. Nous avons obtenu des résultats intéressants sur un benchmark couramment employé dans ce type d'application.

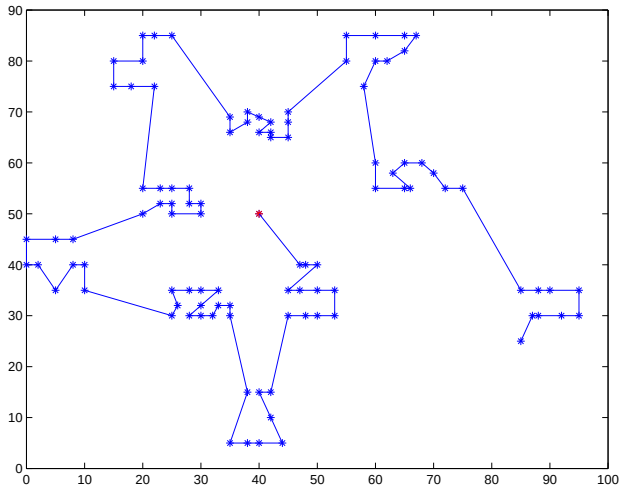

FIG. 11 – *Chemin obtenu avec la stratégie 3-opt.*

Méthode	*Distance parcourue*
NN	564.9
2-opt	471.5
3-opt	461.8
Classification + "Branch and Bound"	461.2

TAB. 1 – *Comparaison des distances parcourues.*

Il est évident que l'on ne peut généraliser ces bons résultats à toutes les applications. En effet, d'autres tests nous ont montré que l'intérêt de notre méthode se réduisait lorsque la taille des classes devenait par trop importante. Dans ce cas, la méthode exacte "Branch and Bound" que nous utilisons pour définir le chemin au sein de chaque classe devient alors trop gourmande en temps de calcul. Une perspective consisterai alors à remplacer la méthode "Branch and Bound" par une heuristique performante.
Une autre perspective, dans le cadre du problème VRP, consiste à prendre en compte l'aspect dynamique[4]. Lorsqu'une partie des clients est connue à l'instant initial le problème dynamique se résume à classer les nouveaux clients dans les groupes existants. L'arrivé d'un nouveau client dans une classe ne remet alors en cause que le chemin local de la classe modifiée. L'intérêt du partitionnement est d'ailleurs connu dans les problèmes de VRP dynamique [20], mais à notre connaissance ce partitionnement est pour l'heure réalisé avec un nombre de classes fixes et de manière aléatoire. Nous pensons donc que notre mé-

[3] c'est-à-dire que tous les clients sont connus lors de la recherche du chemin.

[4] Tout ou partie des clients apparaissent aléatoirement en fonction du temps.

thode peut apporter des perspectives intéressantes dans ce domaine.
Par ailleurs, d'autres applications de la classification dans des domaines connexes à celui du VRP sont possibles. En effet, être capable de classer des clients permet par exemple de définir le lieu et le nombre de dépôts et de prévoir les besoins en terme de véhicules. La résolution de ces problèmes peut donc être également obtenue par l'application de la classification par la théorie de l'évidence.

Références

[1] R. K. Ahuja, O. Ergun, J. B. Orlin, and A. P. Punnen. A survey of very large-scale neighborhood search techniques. *Discrete applied Mathematics*, 123(1-3) :75–102, 2002.

[2] D. Applegate, R. Bixby, V. Chvatal, and W. Cook. On the solution of traveling salesman problems. *Documenta Mathematica*, Extra volume ICM :645–656, 1998.

[3] D. Bertsimas and G. Van Ryzin. A stochastic and dynamic Vehicle Routing Problem in the Euclidean plane. *Operations Research*, 39 :601–615, 1991.

[4] J. C. Bezdek. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1 :57–71, 1974.

[5] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function*. Plenum, New York, 1981.

[6] W. F. Caselton and W. Luo. Decision making with imprecise probabilities : Dempster-Shafer theory and application. *Water Resources Research*, 28(12) :3071–3083, 1992.

[7] G. A. Croes. A method for solving traveling-salesman problems. *Operations Research*, 6 :791–812, 1958.

[8] A. Dempster. Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics*, AMS-38 :325–339, 1967.

[9] T. Denoeux. A k-nearest neighbour classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5) :804–813, 1995.

[10] T. Denoeux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7) :1095–1107, 1997.

[11] T. Denoeux. A neural network classifier based on Demspter-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics Part A*, 30(2) :131–150, 2000.

[12] T. Denoeux and G. Govaert. Combined supervised and unsupervised learning for system diagnosis using Dempster-Shafer theory. In *CESA'96 IMACS Multiconference, Computational Engineering Applications. Symposium on Control, Optimization and Supervision*, volume 1, pages 104–109, 1996.

[13] E. Diday. La méthode des nuées dynamiques. *Revue de Statistique appliquée*, 19(2) :19–34, 1971.

[14] B. Dubuisson. *Diagnostic et Reconnaissances de Formes*. Hermes, 1990.

[15] J. C. Dunn. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3) :32–57, 1973.

[16] F. Glover, G. M. Gutin, A. Yeo, and Zverovich. Construction heuristics ans domination analysis for the asymmetric TSP. Technical report, Brunel University, 1999.

[17] J. J. Hopfield. Neural network and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79 :2554–2558, 1982.

[18] E. Lefevre. *Fusion adaptée d'informations conflictuelles dans le cadre de la théorie de l'évidence : Application au diagnostic médical*. PhD thesis, Institut des Sciences Appliquées de Rouen, 2001.

[19] S. Lin. Computer solutions to the traveling salesman problem. *Bell System Tech. Journal*, 44 :2245–2269, 1965.

[20] X. Lu, A. C. Regan, and S. Irani. An asymptotically optimal algorithm for the dynamic traveling repair problem. Technical Report UCI-ITS-LI-WP-02-4, Institute of Transportation Studies, University of California Irvine, 2002.

[21] K. G. Murty. *Operations Research : Deterministic Optimization Models*, chapter 10, page 608. Prentice-Hall, 1995.

[22] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.

[23] P. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :447–458, 1990.

[24] P. Smets. What is Dempster-Shafer's model ? In R.R. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 5–34. Wiley, 1994.

[25] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2) :191–234, 1994.

[26] M. Solomon. Algorithms for the vehicle routing and scheduling problem with time window constraints. *Operations Research*, 32 :254–265, 1987.

[27] R. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41 :93–138, 1987.

[28] L. M. Zouhal. *Contribution à l'application de la théorie des fonctions de croyance en reconnaissance des formes*. PhD thesis, Université de Technologie de Compiègne, 1997.

[29] L. M. Zouhal and T. Denoeux. An evidence-theoretic k-NN rule with parameter optimization. *IEEE Transactions on Systems, Man and Cybernetics - Part C*, 28(2) :263–271, 1998.