Article submitted to " Multisensor, Mutlisource Information Fusion :
Architectures, Algorithms, And Applications " Symposium, Orlando, 2006.

Classifiers and Distance-Based Evidential Fusion For Road Travel Time Estimation

EL FAOUZI N.-E*. LEFEVRE E**.

(*) Laboratoire d'Ingénierie Circulation Transports (LICIT)
INRETS-ENTPE
25, Avenue François Mitterand Case 24
69675 Bron Cedex
nour-eddin.elfaouzi@inrets.fr

(**) Laboratoire d'Informatique et d'Automatique de l'Artois (LGI2A), EA 3926
Université d'Artois, Faculté des Sciences Appliquées
Technoparc Futura, 62400 Béthune
eric.lefevre@iut-geii.univ-artois.fr

Abstract

This paper addresses the road travel time estimation on an urban axis by classification method based on evidence theory. The travel time (TT) indicator can be used either for traffic management or for drivers' information. The information used to estimate the travel time (induction loop sensor, cameras, probe vehicle,...) is complementary and redundant. It is then necessary to implement strategies of multi-sensors data fusion. The selected framework is the evidence theory. This theory takes more into account the imprecision and uncertainty of multisource information. Two strategies were implemented. The first one is classifier fusion where each information source, was considered as a classifier. The second approach is a distance-based classification for belief functions modelling. Results of these approaches, on data collected on an urban axis in the South of France, show the outperformance of fusion strategies within this application.

Keywords: *Evidence Theory, Classifier combination, Travel Time Estimation*.

1. Introduction

Considering the huge development of automobile traffic, and more specifically, individual mobility, much research has been carried out in road traffic engineering to improve the operation of different road networks (urban and interurban) and the level of service provided. In the urban environment in particular, where congestion problems are frequent, road network operators are required to regulate and control traffic in order to improve the safety and comfort of users.

In this perspective, and with the appearance of new information technologies, current research is turning toward setting up dynamic information and guidance systems for road users. An example of this is the installation on highways and urban expressways of variable message signs (VMS) informing users about traffic conditions (fluid, bottlenecks, etc.) or the time needed to travel along the road they have taken. According to a large number of surveys, users appreciate the diffusion of this type of information. There is now a desire to equip major urban roads with similar information systems.

This raises the question of estimating travel time with acceptable accuracy, a very difficult task in the urban environment where a large number of theoretical, technical and methodological difficulties must be dealt with. In this respect, traditional detectors have proven to be inefficient in certain circumstances as a means of obtaining information on traffic conditions on urban thoroughfares. With the arrival of new measurement instruments (cameras, localization aided by GPS and cellular telephones, etc.), it is possible to rely increasingly on other sources of data. This data is intended to complete the information provided by traditional measurement devices with the result of improving the quality of travel time estimations. The problem of estimating travel time therefore becomes a typical data fusion problem.

Classically, the travel time estimation problem has been treated to two approaches:

- either the experimented travel times (ETT) by a certain number of vehicles (probe vehicles) are recorded, and the travel time for the itinerary is estimated on the basis of these measurements (e.g., by calculating a mean) ;

- or by using flow measurements from conventional traffic loop detectors (measuring flow and occupancy rate), which, by using a suitable method, are used to estimate an itinerary travel time (ITT).

These two techniques lead to very different types of result in terms of precision and representativeness. The first provides accurate estimation of the travel time of an itinerary, but this estimation suffers from poor statistical representativeness. Furthermore, the low availability of the RTT type data reduces the reliability of the estimation. Hence, the travel time information provided by this approach is precise but its representativeness varies according to the number of probes available, the measurement frequency and the distribution of travel times on the itinerary under consideration.

However, the second technique provides information on the entire flow on the itinerary continuously through time. These data are therefore considered as exhaustive and thus ensure that the measurements are representative. Nonetheless, an estimation of travel time on the basis of these data is less accurate than with those obtained by the former technique, for reasons specific to this approach (defective data, inaccurate travel time conversion methods, etc.). Whatever the case, it still remains the method used most due to its low cost.

In the framework of this work, we make use of both data from individual travel times and data characteristic of the flow. Therefore, we wish to improve the overall performance of travel time estimations resulting from the two approaches by combining them, i.e. by merging the estimations obtained from each of them. We also use travel time data collected by license plate matching which we consider as reference data and use as the basis for evaluating the performance of the fusion process.

Among the techniques of fusion of multi-sensors data known in the literature, we selected the evidence theory framework. Indeed, it seems that this theory allows taking into account in a more natural way inaccuracies as well as uncertainties associated to the information.

In this article, we present two different approaches for fusion process. First of all, an approach of fusion of classifiers in which each source of information is considered as classifiers. The second approach is based on a distance-based strategy for deriving mass functions associated to the evidence theory.

This article is decomposed as follows. We present in the section 2, the mathematical background of the evidence theory. Both methods of classification fusion used on the estimation of the travel time are presented in the section 3. Finally, in the section 4, results of these approaches on data collected on the site located in the south of France show the outperformance of the fusion process within the framework of this application.

2. Introduction

In this section, we recall briefly some basic concepts of the evidence theory. The point of view of the model of the transferable belief [16] is adopted in this article. This one distinguishes two levels of data processing: the credal level where the belief functions are modelled and revised and the pignistic level in which belief functions are transformed into probability functions for the decision-making step.

2.1. Credal level

Let Ω be a finite set, generally called frame of discernment. A belief function Bel is a non additive fuzzy measure mapping 2^{Ω} into $[0,1]$ defined as follows:

$$bel(A) \triangleq \sum_{\emptyset \neq B \subseteq A} m(B) \qquad \forall\, A \subseteq \Omega \qquad (1)$$

where m, called generally mass function, mapping 2^{Ω} into $[0,1]$ satisfying the following constraint:

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

Every subset $A \subseteq \Omega$ such as $m(A) > 0$ is called focal element of m. So, the mass function $m(A)$ represents the degree of belief attributed to the proposition A and which was not able, considering the state of the knowledge, to be allocated to a more specific subset than A.

A set of mass m such as $m(\emptyset) = 0$ is called normal. This condition was initially imposed by Shafer [11] but can be relaxed if we accept the opened world assumption which postulates for the non-exhaustiveness of the frame of discernment Ω.

From the seminal work of Shafer [11], the belief functions are nowadays recognized for the modelling of the uncertain information (from the total ignorance to the complete knowledge). So, a situation of complete ignorance corresponds to the belief function defined by m (Ω) = 1. The perfect knowledge will be represented by a sure belief function that is a function where the totality of the mass is assigned to a the only singleton of Ω. Another particular case can be met when the focal elements of m are singletons. In that case, the belief function is equivalent to probability measure and will be called Bayesian belief.

At the credal level thus intervenes belief functions modelling. Several methods were developed to define these functions. They can be grouped together into two approaches. Approaches of the first group are based on likelihood functions [1, 13, 19] while approaches of the second group are distance-based [3, 4, 10]. A comparison of these two families is realized in [17]. The findings of this study show that the performances of these two families, applied to classic problems, are relatively similar and that the choice thus remains a delicate subject. Moreover, it seems that information modelling belief function is application-dependant.

The second step at the credal level corresponds to the belief revision. Among tools defined in the initial works of Shafer, a rule of the combination of two belief functions was proposed. Given two mass functions m1 and m2, the conjunctive combination of these two functions (or sources) is given by:

$$m_{\cap}(A) \triangleq \sum_{A=B\cap C} m_1(B)m_2(C). \qquad (2)$$

This rule, called non-normalized Dempster’s rule, allows combining uncertain information extracted in the form of belief functions. If needed, the condition $m(\varnothing)=0$ can be obtained by dividing every mass by a normalization coefficient. The resulted rule, called orthogonal rule of Dempster, is defined as follows:

$$(m_1 \oplus m_2)(A) \triangleq \frac{(m_1 \cap m_2)(A)}{1-m(\emptyset)} \qquad (3)$$

Where the quantity $m(\varnothing)$ represents a degree of conflict between the functions m_1 and m_2 and can be calculated by using the following equation:

$$(m_1 \cap m_2)(\emptyset) = \sum_{B\cap C=\emptyset} m_1(B)m_2(C). \qquad (4)$$

The use of Dempster’s rule is possible if and only if m_1 and m_2 are not in total conflict, that is to say if there are two focal elements B and C of m_1 and m_2 which satisfy $B\cap C=\varnothing$. This rule possesses however interesting properties as the associativity, the commutativity and - idempotence but was very controversial [23, 22, 12]. That is why other schemes were developed [21, 7, 13, 5, 6]. A description of the conflict within management within the framework of evidence theory is addressed in [14].

2.2. Pignistic level

The aggregation step previously defined allows to obtain an exhaustive summary of the piece of information in the form of a belief function m used for the decision-making. Based on rationality concept developed in this transferable belief model (TBM), Smets [16] suggested transforming m into a probability function BetP defined on Ω (called pignistic probability function). For $\omega_k \in \Omega$

$$BetP(\omega_k) = \frac{1}{1 - m(\emptyset)} \sum_{A \ni \omega_k} \frac{m(A)}{|A|} \qquad (5)$$

Where $|A|$ represents the cardinal of A. In this transformation, the mass of belief m(A) is uniformly distributed among the elements of In. A justification of this transformation can be found in [15]. The derived distribution of probability allows classical tools of probability and statistical theory.

3. Implementation for Travel time estimation

To achieve the travel time estimation via the theory of evidence, we have implemented two different approaches. For the first approach, we define a space in 2 dimensions where every constituent corresponds to the estimation of a sensor (probe vehicle or traffic loop detector).

In this case, we calculate dissimilarities between a new couple of measurements and measurements in learning sample. These dissimilarities allow us to build belief functions and so to attribute a class of travel time to the new couple of measures.

The second approach corresponds to an approach of fusion of classifiers. In that case, the measures obtained by the available sensors are considered as classes. Output of each sensor can be interpreted as a classifier and the fusion problem is equivalent to the one of classifiers fusion.

In the following sections, we describe these both approaches in more detail with an emphasis on belief function building.

3.1. Approach based on distance method

Several work was completed on the modeling of the belief functions by dissimilarity measures [2,8,9,18]. In this section, we presented only the original work introduced by Denoeux [3,24]. This method can be described in the following way.

Let us consider the case where some pattern x has to be classified in one of the K classes $\omega_1, \ldots, \omega_n, \ldots, \omega_K$ using the training set χ of I pattern with known classification. Each training vector x^i sufficiently close to x according to some distance measure d_i can be regarded as a piece of evidence that influences our belief concerning the class of x. This item of evidence can be represented by a basic belief assignment (BBA) m_i over the frame of discernment $\Omega = \{\omega_1, \ldots, \omega_K\}$. If x^i belongs to class ω_n, then the unit mass should be distributed among two subsets of Ω: the singleton $\{\omega_n\}$ and Ω itself. If we consider as a reasonable assumption that the portion of belief committed to ω_n should be a decreasing function of the distance d_i between x and x^i, the m_i can be written in the following form:

$$\begin{cases} m_i(\{\omega_n\}) & = \alpha_i \phi_i(d_i) \\ m_i(\Omega) & = 1 - \alpha_i \phi_i(d_i) \end{cases} \qquad (6)$$

where $0 < \alpha_i < 1$ is a constant and ϕ_i is a monotonically decreasing function verifying $\phi_i(0) = 1$ et $\lim_{d \to \infty} \phi_i(d) = 0$. An exponential form for this function was postulated for ϕ_i:

$$\phi_i(d_i) = exp^{-\gamma_i (d_i)^2} \qquad (7)$$

γ_i being a positive constant associated to class ω_n. A method for optimizing parameters α_i and γ_i has been described in [24]. This method is based on the minimization of error quadratic between the pignistic probability and the membership vector. The above discussion concerned an arbitrary training pattern x^i. However, it is unlikely that all training patterns will be helpful in classifying x, so that we can focus our attention on the k nearest neighbors or on the prototype of each classes. This last solution was adopted in this paper.

3.2. Approach based on fusion of classifiers

In the last section, we presented a classifiers based on evidence theory. If several classifiers are available, the combination of these classifiers makes it possible to improve the results of classification.

Three types of classifiers are indexed:

- the combination is made based on the output information of the abstract level: a classifier only outputs a unique class)
- the combination is made based on the output information of the rank level: a classifier ranks all the labels in a queue with the label at the top being the first choice.
- the combination is made based on the output information of the measurement level: a classifiers attributes each label a measurement value to address the degree that ω_n has the label.

In the studied application, the outputs of the sensors (to similar at the classifier) are at the abstract level. The solution that we proposed rests on the errors of the classifiers taken individually. The error structure of each classifiers j (i.e. C^j) is usually described by a confusion matrix:

$$
M^j = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1j} & \cdots & n_{1N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ n_{i1} & n_{i2} & \cdots & n_{ij} & \cdots & n_{iN} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ n_{N1} & n_{N2} & \cdots & n_{Nj} & \cdots & n_{NN} \end{pmatrix}
$$

where N corresponds to the number of class in the problem. In this matrix, each row i corresponds to class ω_i and each column j corresponds to the class decided by the classifier C^j. With the knowledge of its confusion matrix M^j, an uncertainty on output of the classifier, could be described by the recognition rate ε^j_{rec} and the confusion rate ε^j_{con}. These rates can be defined by the following equation:

$$
\epsilon^j_{\text{rec}} = \frac{\sum_{i=1}^{N} n_{ii}}{\sum_{i,j}^{N} n_{ij}} \qquad \text{et} \qquad \epsilon^j_{\text{con}} = 1 - \epsilon^j_{\text{rec}}
$$

We supposed that we have the confusion matrix for each classifier. We present in the next section three methods to obtain the belief function with the confusion matrix.

1. METHOD N°1

The first approach, originally proposed by XU [20], is based on the evidence theory to combine classifiers. Within this approach, referred to as XU's method, the belief functions are defined as follows:

$$
\begin{cases} m_j(\{\omega_n\}) & = \epsilon^j_{\text{rec}} \\ m_j(\overline{\omega_n}) & = \epsilon^j_{\text{con}} \end{cases}
$$

in this equation the classifier C^j choices the class ω_n. The following method rises from this construction of belief functions.

2. METHOD N°2

The belief functions obtained previously are independent of the choice of the classifiers. The belief functions are similar whatever the class selected. In this approach, we defined a recognition rate for each class by classifier.

The associated belief functions are defined as follows:

$$\begin{cases} m_j(\{\omega_k\}) & = \frac{n_{kk}}{\sum\limits_{i=1}^{N} n_{ki}} \\ m_j(\overline{\omega_k}) & = 1 - \frac{n_{kk}}{\sum\limits_{i=1}^{N} n_{ki}} \end{cases}$$

when the classifier selected C^j the class ω_n.

3. METHOD N°3

Contrary to the preceding method, with this approach we place masses on the singleton, its opposite and on the frame of discernment Ω. The beliefs are distributed as follow:

$$\begin{cases} m'_j(\{\omega_i\}) & = \frac{n_{ki}}{\sum\limits_{j=1}^{N} n_{jk}} \qquad \forall i = 1, \ldots, N \\ m'_j(\Omega) & = \frac{\sum\limits_{j=1}^{N} n_{kj} - n_{kk}}{\sum\limits_{j=1}^{N} n_{kj}} \end{cases}$$

when the classifier selected C^j the class ω_k. The belief function m' is not normalized. We must pass by the step of following normalization:

$$m_j(A) = \frac{m'_j(A)}{\sum\limits_{B \subseteq \Omega} m'_j(B)} \qquad \forall A \subseteq \Omega$$

4. Results

4.1. Data

The data used in this study were collected during an experimentation carried out on an urban road axis in Toulouse (South of France). The urban axis on which was made the experiment is cut in 4 sections. Data collected consist in 3 different sources:

a. Data from conventional road sensors (magnetic loop detectors) every 6 minutes. This source delivers macroscopic traffic characteristics like flow, local density.
b. Probe vehicle data, mainly, travel time experienced by those vehicle.
c. And travel time derived from license-plate matching technique on the same urban axis.

The information delivered by the first 2 sources will be processed to achieve an estimation of the travel time while information from the last source, considered as reference travel time, will be used for evaluation purpose only.

As conventional road sensors (source 1) do not deliver travel time directly, it is necessary to convert the information supplied by magnetic loop detectors with a suitable estimation algorithm. For that purpose, we used the following conversion algorithm originally proposed by Bonvalet and Robin-Prévallée [7].

$$TP_i = \frac{(TO_i.Nmax)}{Q_i} + (1 - TO_i).TPL_i \qquad (14)$$

where TP_i is the travel time of section i, TO_i is the local density (or occupation) of the section, Qi is the flow, TPLi is the travel time prevailing at free flowing condition and Nmax is a maximum number of vehicle within the section.

Within this application, we have considered:

$$Nmax = d_i/5 \quad \text{et} \quad TPL_i = \frac{50000}{3600}.\frac{1}{d_i} \qquad (15)$$

Where d_i is the length (in meter) of section i. We so consider that an average length of vehicle is 5m and speed at free flowing condition is of 50 km/h.

This pre-processing step, allowed obtaining 230 valid observations from traffic data source whilst probe vehicles supplied only 156 observations. However, by cross-checking these two sources, it seems that only 143 observations are common. In addition, to apply the theory of evidence, travel time was discretized to form 6 non overlapping intervals.

In order to evaluate the performances of the various strategies of fusion implemented in this article, we present in the table 2 results of classification without fusion i.e. classification rate achieved individual sources.

	Traffic source	Probe source
Classification rate (%)	26.57	27.27

TAB.1 classification rate for individual sources

4.2. Behaviour of distance-based modeling

In this section, we evaluate the behaviour of the distance-based methods for designing belief functions according to their parameters: number of prototypes for the ProDS method and the number of neighbours for the Knn-ds method.

For this test, we used the cross-validation strategy: leave-one-out [13]. The ProDS method depends on the position of the prototypes we carried out 100 runs for a given number of prototypes. Thus, on the figures representing the evolution of the percentage of classification according to the number of prototypes, we indicated the minimal, maximum value as well as the average of the percentage of good classification. The various results obtained, in terms of percentage of good classification, are presented on the figure FIG. 1.

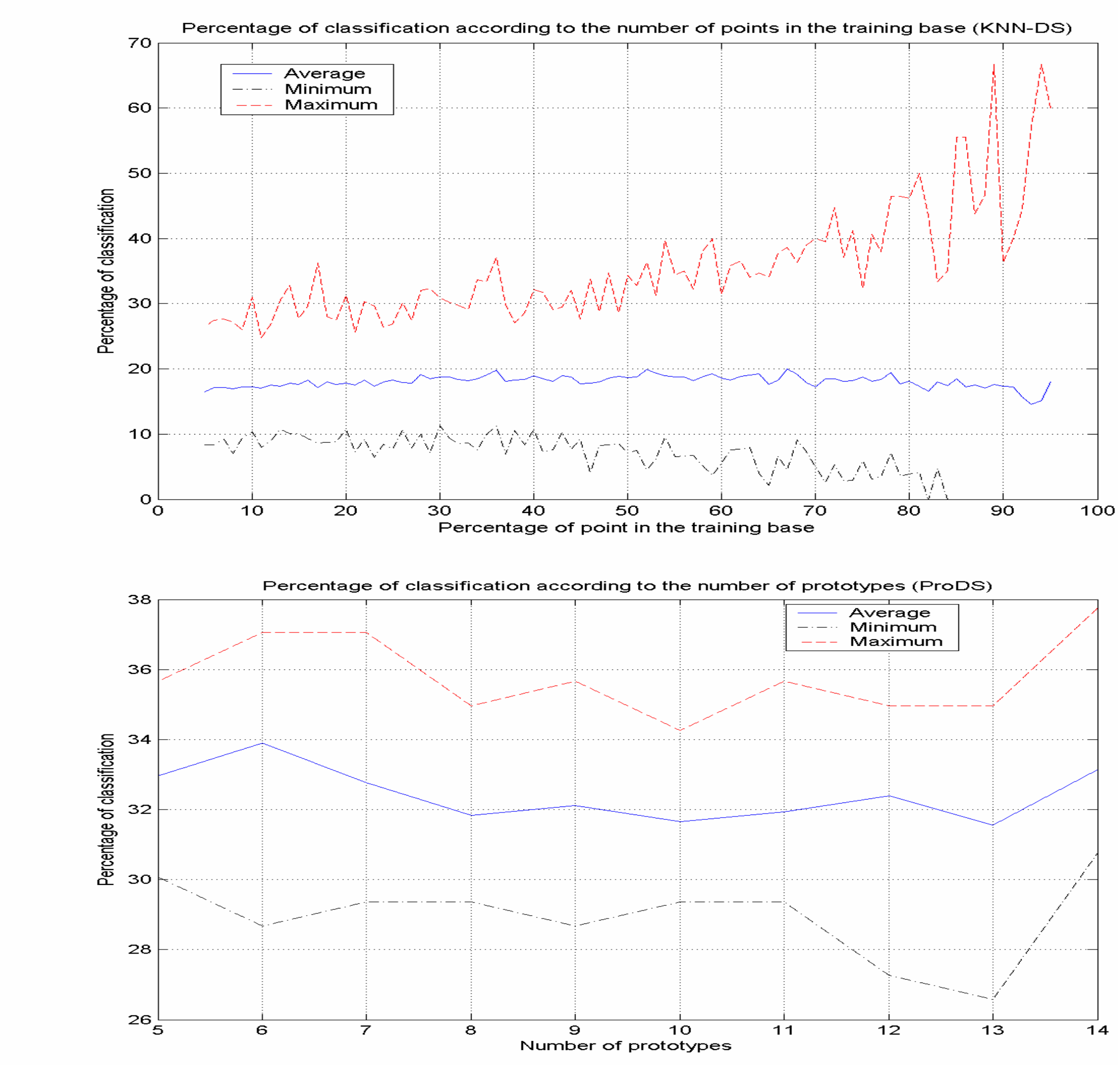

FIG.1 classification rate as a function of length of learning sample vs. number of prototype for KNN-DS ProDS

On this figure, one can observe that the ProDS method is much less sensitive to the variation of the number of prototypes contrary to the Knn-ds method of which the percentage of classification strongly evolves according to the number of neighbours. The percentage of good classification obtained by the ProDS method is around 3%. The percentage of classification obtained with the method of Knn-ds is definitely worse than those obtained by the ProDS method.

These results can be compared with those carried out by the classifiers fusion approach presented in section 3.2. By using the same strategy of training (leave-one-out), this approach gives results presented in the table 3. One can note that the results obtained by ProDS method are quite similar to those obtained by the 3rd strategy of fusion of classifiers.

	Xu's method	Method n°2	Method n°3
Classification rate (%)	27.97	27.27	32.87

TAB.2 Classification rate for the proposed methods and comparison with Xu's method.

4.3. Impact of training set on performance

To check the effect of training set, we varied the number of observations in the base of training from 5 to 95% of the initial population. This consists in taking part of the initial observations which are employed in the base of training and the remainder of the observations constituting the base of test then. 300 runs were carried out for a given proportion. Thus on the various curves, we represented the average, the minimal and maximum value as well as the standard deviation of the percentage of good classification. The figure FIG. 2 shows the evolution of the percentage of good classification obtained by the method of fusion of classifiers proposed by XU [29], according to the percentage of points in the training set.

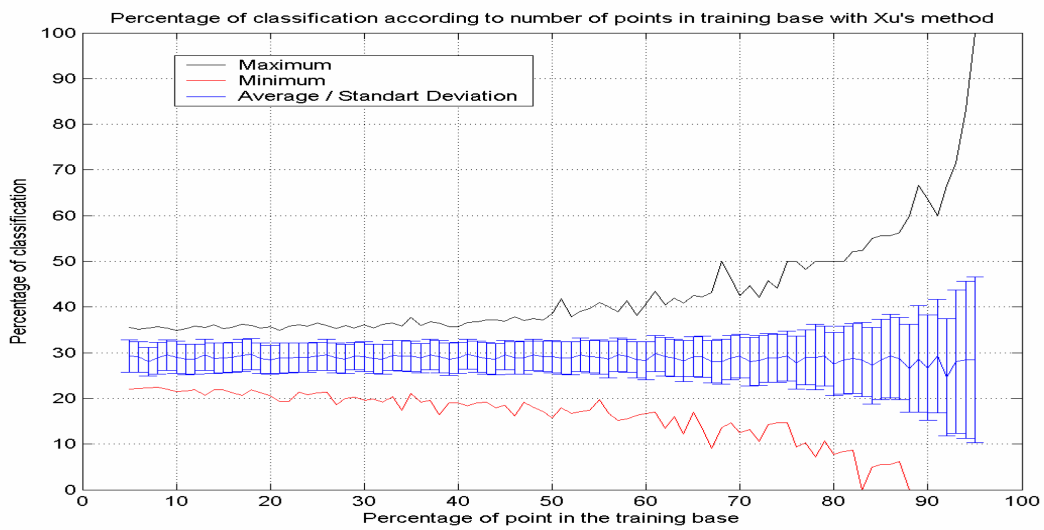

FIG.2 classification rate as a function of learning sample length for Xu's method.

The figure FIG. 3 shows the evolution of the percentage of classification rate according to the length of the training sample by using the improvement, suggested in [1], of the method of Xu and method n° 3.

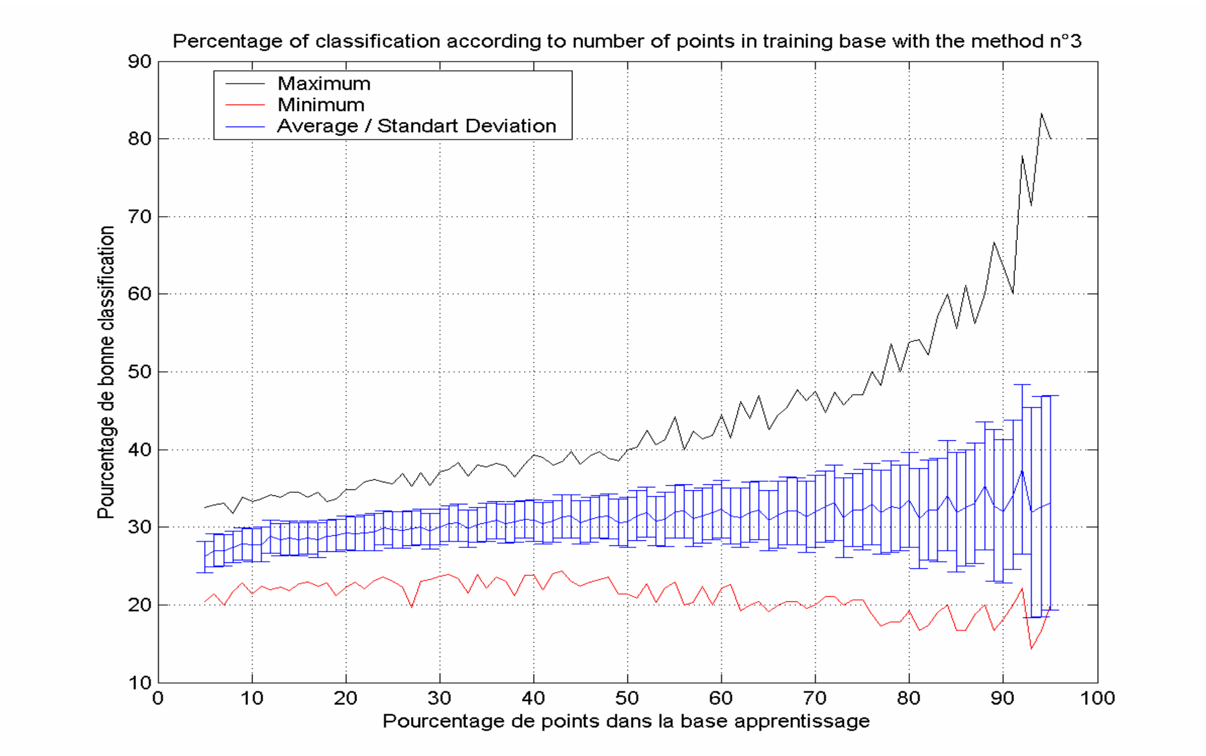

FIG.3 Classification rate according to the length of the training sample

One can note on these figures that the performances obtained by the method of Xu are less sensitive to the constituent numbers of observations which are in the training sample. However, these performances remain overall worse than those obtained by the methods n°2 and n°3 when 20% of the initial observations form the training set. In addition, the methods n°3 gives the best results, in terms of percentage of classification rate, once that the training set contains more than 20% of the initial observations.

The performance of Knn-DS method is lower than 20% no matter the length of training set. The ProDS method gives similar results to those obtained by the method n°3 of fusion of classifiers.

5. Conclusion

The objective of the work carried out in this article is to propose a methodological framework as well as solutions with the problem of estimate of travel time in the presence of data resulting from heterogeneous sources. Two sources were considered here: traditional sensors of traffic made up of an electromagnetic loop, which make it possible to measure the flow and the occupancy rate and to deduce an estimate of travel time, and a reduced sample of probe vehicles, which collect their experienced travel time.

In this article, we tackle the problem of the estimate of the travel time like a typical problem of fusion of classifiers by using the theory of evidence for its flexibility for modelling of knowledge. Within the framework of this theory, we presented two approaches as solutions to the estimation problem. The first technique uses the concept of classifiers fusion. In this case, each source of information is regarded as a classifier. For each one of these classifiers, we define a matrix of confusion which reflects capacity of discrimination of the sources. This matrix enables us thereafter to build the functions of belief. The second approach employs a traditional technique of classification based on the calculation of distance to build the functions of belief. The use of these two approaches within the framework of our real application proved to be effective compared to approaches mono-sensor. That is illustrated by one of our original techniques proposed here (method n°3)

It is obvious that one cannot generalize these results because they were obtained for a specific application. However, this first work enables us to think that the use of technique of fusion of data based on the theory of evidence will tend to improve the estimate of travel time.

The perspectives related to this work relate to the adaptive and dynamic fusion. For this prospect, it will be interesting to build up a fusion scheme which enables to account with temporal dimension of data sources.

Another direction relates to the use of the recent extensions of Evidence theory to take into account the continuous aspects of travel time. Those extensions would make it possible to estimate travel times either in a discrete way as that is currently the case (classes being intervals time) or in a way continuous way allowing a more precise estimate.

6. References

[1] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense,* 11:27-40, 1991.

[2] I. Bloch. Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters,* 17 :905-919,1996

[3] T. Denoeux. A k-nearest neigbour classification rule based on Dempster-Shafer theory. *IEEE transaction on System, Man and Cybernetics*, 25(5) : 804-813, 1995.

[4] T. Denoeux. A neural network classifier based on Dempster-shafer theory. *IEEE transaction on System, Man and Cybernetics* 30(2) : 131-150, 2000.

[5] E. Lefevre, O. Colot and P. Vannoorenberghe. Belief function combination and conflict namagement. *Information Fusion*, 3(2) : 149-162, 2002.

[6] E. Lefevre, O. Colot and P. Vannoorenberghe. Reply to the comments of R. Heanni on the paper " Belief functions combination and conflict management ". *Information fusion*, 4(1) :63-65, 2003.

[7] C.K. Murphy. Combining belief functions when evidence conflicts*. Decision Support Systems*, 29 :1-9, 2000.

[8] G. Ng and H. Singh. Data equalization with evidence combination for pattern recognition. *Pattern Recognition Letters*, 19 :227-235, 1998.

[9] G. Ng and H. Singh. Democraty in pattern classification : combination of votes from various patterns classifiers. *Artificial Intelligence in Engineering*, 12 :189-204, 1998.

[10] N.R. Pal and S. Ghosh Some classification algorithms integrating Dempster-Shafer theory of evidence with the rank nearest neighbor rules. *IEEE Transaction on Systems, Man and Cybernetics*, 31(1) :59-66, 2001.

[11] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University press, Princeton, New-Jersey, 1976.

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

[13] P. Smets, Belief functions : The disjonction rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9 :1-35, 1993.

[14] P. Smets, Analyzing the combination of conflicting belief functions. Submitted, 2004. http://iridia0.ulb.ac.be/psmets/.

[15] P. Smets. Decision making in the TBM : the necessity of the pignistic transformation. *International Journal of Approximate Reasonning*, 38 :133-147, 2005.

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

[17] P. Vannoorenberghe and T. Denoeux. Likelihood-based vs. Distance-based Evidential Classifiers. In *10th international Conference on Fuzzy Systems FUZZ-IEEE'2001*, pp. 320-323, 2001.

[18] A. Verikas, K. Malmqvist and M. Bacauskiene. Combining neural Networks, Fuzzy sets and evidence theory based approaches for analyzing colour images. In *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, pp 297-302, Como, Italy, July 200.

[19] P. Walley and S. Moral Upper probabilities based only on the likelihood function. *Journal of Royal Statistical Society*, Serie B, 61(Part 4) : 831-847, 1999.

[20] L. Xu, A. Krzyzak and C.. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transaction an System, Man and Cybernetics*, 22(3) :4187-435, 1992

[21] R.R. Yager. Hedging in the combination of evidence. *Journal of Information and Optimization Sciences*, 4(1) :73-81, 1983.

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

[23] L.A. Zadeh. *On the validity of Dempster's rule of combination of evidence*. University of California, Berkeley, 1979. ERL Memo M79/24.

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