Decision fusion for postal address recognition using belief functions

David Mercier[1,2], Geneviève Cron[2], Thierry Denœux[1] and Marie-Hélène Masson[1]

[1] UMR CNRS 6599 Heudiasyc
Université de Technologie de Compiègne
BP 20529 - F-60205 Compiègne cedex - France

[2] SOLYSTIC, 14 avenue Raspail,
F-94257 Gentilly Cedex, France

December 21, 2007

Abstract

Combining the outputs from several postal address readers (PARs) is a promising approach for improving the performances of mailing address recognition systems. In this paper, this problem is solved using the Transferable Belief Model, an uncertain reasoning framework based on Dempster-Shafer belief functions. Applying this framework to postal address recognition implies defining the frame of discernment (or set of possible answers to the problem under study), converting PAR outputs into belief functions (taking into account additional information such as confidence scores when available), combining the resulting belief functions, and making decisions. All these steps are detailed in this paper. Experimental results demonstrate the effectiveness of this approach as compared to simple combination rules.

Keywords: Mailing address recognition, Information fusion, Dempster-Shafer Theory, Evidence Theory, Transferable Belief Model.

1 Introduction

Postal address recognition is a complex task that involves different processes including image scanning, address block and writing fields location, character and word recognition, database querying and association of a group of words with a delivery address. Each of these stages requires different algorithms and complex decision procedures [15, 26]. Systems in charge of this task are referred to as *postal address readers (PARs)*.

A generic representation of a PAR is shown in Figure 1. A PAR is linked with a database containing the whole set of postal addresses of the concerned country. It assigns a handwritten or machine printed mail piece image to a delivery address. Additional pieces of information can also be output by a PAR such as alternative addresses, confidence scores, or other intermediate results.

In the past years, a lot of work has been devoted to improving the performances of PARs through refining various subsystems such as, e.g., block location [28] or handwriting recognition [16, 11]. A different way of improvement, which has comparatively received less attention, consists in combining different PARs. Indeed, a decision based on a great number of varied pieces of information is generally more robust than any decision made individually from a single piece of information [9, 3, 10].

As explained by Xu, Krzyzak and Suen in [27], the combination of multiple classifiers includes several problems: selecting the classifiers to combine, choosing an architecture for the combination, and combining the classifier outputs in order to achieve better performances than each classifier individually. In this article, we focus on the problem of combining the outputs supplied by different PARs, following the fusion scheme illustrated in Figure 2.

Various approaches to classifier fusion have been proposed in the literature (see, e.g., [3, 10]). These methods are based on various frameworks such as probabilistic theory, fuzzy sets, possibility theory, Dempster-Shafer theory, voting, etc. As remarked in [2], no general solution exists: the success of a fusion method strongly depends on the fusion problem itself, and the way a given formalism is applied.

In the postal domain, the fusion problem is very specific. Addresses for the purpose of physical mail delivery have a hierarchically structure: depending on the particular

application, they are characterized by a country, state, town, post office, street, street number or post office (PO) box, and secondary number such as apartment number. Moreover, PARs have the possibility to output a complete or a partial address. For example, a PAR can provide a solution for the town, but not for the street, if this latter has not been recognized. Moreover, as previously mentioned, some PARs can also supply additional information such as confidence scores.

Until now, approaches to PAR fusion have been very limited. They are essentially based on common sense combination rules [8, 4]. For example, the main idea in [8] is to choose the postal address provided by a majority of PARs, if it exists; otherwise, the postal address output by the most reliable PAR is chosen. Although such simple rules can be sufficient for some applications, higher recognition rates can be expected to result from finer modelling of the fusion problem.

In this paper, it is proposed to model this fusion problem using the Transferable Belief Model (TBM) [25, 22], a subjectivist and non-probabilistic interpretation of the Dempster-Shafer theory of belief function [18, 21]. This framework is known to offer great flexibility for the manipulation of partial knowledge and information represented at various granularity levels (see, e.g., [19, 12, 1]), thus providing a range of tools well suited to the application at hand.

This article is organized as follows. The postal problem is first presented in Section 2. Background material on belief functions is recalled in Section 3. A generic model for combining postal addresses is then introduced in Section 4, and experimental results are reported in Section 5. Finally, Section 6 concludes the paper.

2 Problem Description

2.1 Hierarchical Structure of Postal Addresses

A postal address is a string of characters indicating where mail sent to a person or an organization should be delivered. Each country has structured its addresses in a specific manner. However, all these arrangements have a hierarchical organization. From top to bottom, each field specifies the destination of the mail in a finer way.

In the simple model considered in this paper for illustrative purposes, a postal

2

address will be assumed to consist of only two parts:

- the *town part* indicating the city where the mail has to be delivered;

- the *distribution part* indicating the precise physical location within a particular city. This place can be a street or any country-specific notion like a PO box for example.

Denoting a town by T_i, a street by S_j and a PO box or any country specific distribution point by B_k, where i,j and k are indices, a complete address can take any of the two following forms:

- $T_i S_j$ for street S_j in town T_i;

- $T_i B_k$: for PO box B_k in town T_i.

Complete addresses form the *distribution level* (bottom level) of the hierarchy shown in Figure 3. When only the town is specified, one obtains a partial address of the form T_i. Such partial addresses for the second level of the hierarchy. Finally, a third level can be defined as the "empty" address noted "_", which corresponds to the case where the address is completely unspecified.

The output of a PAR for a given mail piece is usually a complete or partial address at any level in this hierarchy.

2.2 Performance Rates and Objectives

Using a test set of labelled letters (letters whose actual address is known), the performances of PARs can be assessed at the town and distribution levels, using the following three performance rates: the correct recognition rate, the error rate, and the rejection rate.

At the distribution level, an output is correct if and only if the town and the distribution are correct. At the town level, an output is correct if and only if the town is correct.

Example 1 When the truth is "$T_i \ S_j$":

- at the distribution level, only the output "$T_i\ S_j$" is considered to be correct; outputs "$T_iS_{j'}$" and "$T_{i'}$", with $j' \neq j$ and $i' \neq i$, are considered as errors, and outputs "T_i" and "$\_$" are considered as rejections;

- at the town level, outputs "$T_i\ S_j$", "$T_iS_{j'}$", and "T_i" are considered to be correct, while an output "$T_{i'}$" with $i' \neq i$ is an error, and an output "$\_$" is a rejection.

At each level and for each PAR, the sum of these three performance rates is equal to one.

As different PARs are available, our main objective is to improve the postal address recognition process by combining the outputs from different PARs. More precisely, one seeks a fusion scheme leading to the greatest possible correct recognition rate, while having an acceptable error rate at both town and distribution levels. In this work, the maximal tolerated error rate is set to the best error rate among individual PARs, as shown in Figure 4.

A fusion scheme designed to meet these objectives, based on the Transferable Belief Model, is presented in this paper.

3   The Transferable Belief Model

The Transferable Belief Model (TBM) is a model of uncertain reasoning and decision-making based on two levels [25, 7]:

- the *credal* level, where available pieces of information are represented by belief functions;

- the *pignistic* or *decision* level, where belief functions are transformed into probability measures, and the expected utility is maximized.

3.1   Basic Concepts

Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, called the *frame of discernment*, be a finite set comprising all possible answers to a given question Q. The beliefs held by a rational agent Ag regarding the answer to question Q can be quantified by a *basic belief assignment*

4

(bba) m_{Ag}^Ω, defined as a function from 2^Ω to $[0, 1]$ verifying:

$$\sum_{A \subseteq \Omega} m_{Ag}^\Omega(A) = 1 .$$

The quantity $m_{Ag}^\Omega(A)$ represents the part of the unit mass allocated to the hypothesis that the answer to question Q is in the subset A of Ω. When there is no ambiguity on the agent or the frame of discernment, the notation m_{Ag}^Ω will be simplified to m^Ω or m.

A subset A of Ω such that $m(A) > 0$ is called a *focal set of m*. A bba m with only one focal set A is called a *categorical bba* and is denoted m_A; we then have $m_A(A) = 1$. Total ignorance is represented by the bba m_Ω such that $m_\Omega(\Omega) = 1$, called the *vacuous bba*. A *normal* bba m satisfies the condition $m(\emptyset) = 0$. A bba whose focal sets are nested is said to be *consonant*.

The belief and plausibility functions associated with a bba m are defined, respectively, as:

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B),$$

and

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B),$$

for all $A \subseteq \Omega$. Functions m, bel and pl are in one-to-one correspondence, and thus constitute different forms of the same information.

The basic operation for combining bbas induced by distinct sources of information is the *conjunctive rule of combination* (CRC), also referred to as the *unnormalized Dempster's rule of combination*, defined as

$$m_1 \medcirc m_2(A) = \sum_{B \cap C = A} m_1(B)m_2(C), \qquad \forall A \subseteq \Omega. \tag{1}$$

3.2   Refinements and Coarsenings

When applying the TBM to a real-world application, the determination of the frame of discernment Ω, which defines the set of states on which beliefs will be expressed, is a crucial step. As noticed by Shafer [18, chapter 6], the degree of *granularity* of Ω is always, to some extent, a matter of convention, as any element of Ω representing

a given state can always be split into several alternatives. Hence, it is fundamental to examine how a belief function defined on a frame may be expressed in a finer or, conversely, in a coarser frame. The concepts of *refinement* and *coarsening* can be defined as follows.

Let Θ and Ω denote two frames of discernment. A mapping $\rho : 2^\Theta \to 2^\Omega$ is called a *refining* of Θ (see Figure 5) if it verifies the following properties [18]:

1. The set $\{\rho(\{\theta\}), \theta \in \Theta\} \subseteq 2^\Omega$ is a partition of Ω.

2. For all $A \subseteq \Theta$:

$$\rho(A) = \bigcup_{\theta \in A} \rho(\{\theta\}). \tag{2}$$

Θ is then called a *coarsening* of Ω, and Ω is called a *refinement* of Θ.

A bba m^Θ on Θ may be transformed without any loss of information into a bba m^Ω on a refinement Ω of Θ by transferring each mass $m^\Theta(A)$ for $A \subseteq \Theta$ to $B = \rho(A)$: $m^\Theta(A) = m^\Omega(\rho(A))$. This operation is called the *vacuous extension of m^Θ to Ω*.

Conversely, a bba m^Ω on Ω may be transformed into a bba m^Θ on a coarsening Θ of Ω by transferring each mass $m^\Omega(A)$ for $A \subseteq \Omega$ to the smallest subset $B \subseteq \Theta$ such that $A \subseteq \rho(B)$. Note that, in this case, some information may be lost in the process.

Thanks to these transformations, a piece of information, initially defined on a frame Ω, may be expressed on any frame obtained by refining and/or coarsening Ω.

Example 2 Let us consider a bba m^Θ defined on $\Theta = \{\theta_1, \theta_2, \theta_3\}$ by $m^\Theta(\{\theta_1\}) = 0.7$ and $m^\Theta(\{\theta_2, \theta_3\}) = 0.3$. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_7\}$ be a refinement of Θ as shown in Figure 5. The piece of evidence m^Θ can be expressed on the finer frame Ω by $m^\Omega(\{\omega_3, \omega_4, \omega_5\}) = 0.7$ and $m^\Omega(\{\omega_1, \omega_2, \omega_6, \omega_7\}) = 0.3$.

Conversely, let us consider a bba m^Ω defined by $m^\Omega(\{\omega_3\}) = 0.6$ and $m^\Omega(\{\omega_1, \omega_6\}) = 0.4$. This piece of information can be expressed on the coarser frame Θ by $m^\Theta(\{\theta_1\}) = 0.6$ and $m^\Theta(\{\theta_2, \theta_3\}) = 0.4$.

3.3 BBA Correction Mechanisms

When receiving a piece of information represented by a BBA m, agent Ag may have some doubt regarding the reliability of the source that has provided this BBA. Such

6

metaknowledge can be taken into account using the *discounting* operation introduced by Shafer [18, page 252], and defined by:

$$^{\alpha}m = (1 - \alpha)m + \alpha\, m_{\Omega}\,, \tag{3}$$

where $\alpha \in [0, 1]$ is a coefficient referred to as the *discount rate*.

A discount rate α equal to 1 means that the source is not reliable and the piece of information it provides cannot be taken into account, so Ag's knowledge remains vacuous: $m^{\Omega}_{Ag} = {}^1m = m_{\Omega}$. On the contrary, a null discount rate indicates that the source is fully reliable and the piece of information it provides is entirely accepted: $m^{\Omega}_{Ag} = {}^0m = m$. In practice, an agent rarely knows for sure whether the source is reliable or not, but it has some degree of belief in the thee source reliability, equal to $1 - \alpha$ [20].

Using metaknowledge on the reliability of the source, the BBA m can thus be corrected into $^{\alpha}m$ using the discounting operation. However, this discounting operation only allows agent Ag to weaken a source of information, whereas it could sometimes be useful to strengthen it, e.g., when this source is too cautious. In [14], the following mechanism for weakening or reinforcing a BBA m was introduced and justified:

$$^{\nu}m = \nu_1\, m_{\Omega} + \nu_2\, m + \nu_3\, {}^{tr}m\,, \tag{4}$$

with $\nu_i \in [0, 1]\ \forall i \in \{1, 2, 3\}$ and $\nu_1 + \nu_2 + \nu_3 = 1$. The BBA ^{tr}m corresponds to the BBA m *totally reinforced*, and is defined by:

$$^{tr}m(A) = \begin{cases} \frac{m(A)}{1 - m(\Omega)} & \text{if } A \subset \Omega, \\ 0 & \text{if } A = \Omega, \end{cases} \tag{5}$$

if m is not vacuous, and $^{tr}m = m$ otherwise. If m is not vacuous, the total reinforcement ^{tr}m of m consists in transferring uniformly the whole mass allocated to Ω to the focal sets of m different from Ω. As shown in [6], this process is the dual of the normalization process that consists in uniformly redistributing to non-empty subsets the mass initially allocated to the empty set.

Example 3 Let us consider a bba m^{Ω} defined on $\Omega = \{\omega_1, \omega_2, \omega_3\}$ by $m^{\Omega}(\{\omega_1\}) = 0.4$, $m^{\Omega}(\{\omega_2\}) = 0.1$, and $m^{\Omega}(\Omega) = 0.5$. The beliefs held by an agent Ag knowing that

m^Ω should be totally reinforced, is given by ${}^{tr}m^\Omega(\{\omega_1\}) = 0.4/(1-0.5) = 0.8$, and ${}^{tr}m^\Omega(\{\omega_2\}) = 0.1/(1-.8) = 0.2$.

In the correction mechanism (4), rates ν_1, ν_2 and ν_3 are equal, respectively, to the degree of belief that the source is unreliable (having to be discarded), fully reliable (having to be accepted without modifying it), and too cautious (having to be reinforced). An application of this correction mechanism will be presented in Section 4.2.2.

3.4 Decision-making

In Bayesian decision theory, modelling a decision process implies defining:

1. A set $\mathcal{D}$ of decisions that can be made;

2. A set Γ of considered states of nature;

3. A cost function $c : \mathcal{D} \times \Gamma \to \mathbb{R}$, such that $c(d, \gamma)$ represents the cost of making decision $d \in \mathcal{D}$ when $\gamma \in \Gamma$ is the true state of nature.

Rationality principles [17, 5] then justify the choice of the decision $d^* \in \mathcal{D}$ corresponding to the minimum expected cost or *risk* according to some probability measure P^Γ on Γ:

$$d^* = \arg\min_d \rho(d),$$

with

$$\rho(d) = \sum_{\gamma \in \Gamma} c(d, \gamma) P^\Gamma(\{\gamma\}) , \quad \forall d \in \mathcal{D}. \tag{6}$$

In the TBM, this decision-theoretic framework is accepted [23, 24]. The set Γ, called the *betting frame*, is often taken equal to Ω. However, it may be any refinement or coarsening of Ω, or it may obtained from Ω by a succession of refinings and coarsenings. The probability measure P^Γ is obtained by the *pignistic transformation* [23, 24], which consists in firstly expressing the piece of evidence m^Ω, initially defined on the frame of discernment, on the betting frame Γ, and then computing the pignistic probability $BetP^\Gamma = P^\Gamma$ as:

$$BetP^\Gamma(\{\gamma\}) = \sum_{\{A \subseteq \Gamma, \gamma \in A\}} \frac{m^\Gamma(A)}{|A|(1 - m^\Gamma(\emptyset))}, \tag{7}$$

8

where $|A|$ denotes the cardinality of A.

4 Application to Postal Address Recognition

The different stages of the fusion process based on the TBM are illustrated in Figure 6. First, the frame of discernment Ω on which beliefs are expressed has to be chosen (step 1). Outputs o_i, which are partial or complete postal addresses provided by each available PAR i are then converted into bbas on Ω (step 2), and combined into one bba m^Ω synthesizing the whole available information (step 3). The pignistic transformation is then applied to reach a final decision (step 4). These four steps are described in more detail below.

4.1 Frame of discernment

The question Q of interest is to determine the destination address indicated on a given mail piece. The set of possible answers to this question should naturally contain all delivery addresses, which are usually stored in a database. However, there are additional possibilities, as the address written on the envelope may correspond to no valid delivery address. For instance, a part of an address may be non-existent or unreadable, as shown in Figure 7. Such addresses are said to be *invalid*. An address in which the town cannot be identified is said to be *totally invalid*. Addresses depicted in Figures 7(a), 7(b) and 7(c) are totally invalid. An address for which the town can be identified but not the distribution part, as illustrated in Figure 7(d), is said to be *partially invalid*.

The frame of discernment Ω we associate to question Q is thus composed of all valid addresses contained in the database, the partially invalid addresses associated with the towns included in the database, and an additional element representing totally invalid addresses. Such of frame typically contains several millions elements.

Example 4 (Small database) Let us consider a toy problem in which the database contains only two towns T_1 and T_2. Town T_1 contains two streets S_1 and S_2. Town T_2 contains one street S_1 (S_1 in T_1, and S_1 in T_2 are physically different streets, but they may have the same name) and two P.O. Boxes B_1 and B_2.

The frame Ω is shown in Figure 8. It is composed of:

- valid addresses: T_1S_1, T_1S_2, T_2S_1, T_2B_1, T_2B_2;

- partially invalid addresses associated with towns T_1 and T_2, denoted by T_1 inv and T_2 inv, respectively;

- an additional element, denoted $\{inv\}$, representing totally invalid addresses.

The subset $\{T_1S_1, T_1S_2, T_1\ inv\}$ contains all addresses in town T_1 and is denoted T_1. Likewise, $T_2 = \{T_2S_1, T_2B_1, T_2B_2, T_2\ inv\}$ contains all addresses associated with town T_2. Finally, Ω is equal to $T_1 \cup T_2 \cup \{inv\}$. The corresponding hierarchy is shown in Figure 9.

Each output of a PAR is thus associated with a unique subset of Ω. Outputs corresponding to a complete address are singletons of Ω; for example, an address "$T_i\ S_j$" is associated with the singleton $\{T_i\ S_j\}$. Outputs corresponding to a partial address are subsets of Ω:

- an output "T_i" is associated with the subset T_i of Ω;

- the output "_" representing total ignorance is associated with Ω.

4.2   Construction of Basic Belief Assignments

Once the frame of discernment has been defined, the next step is to define a method for converting PAR outputs into belief functions. Here, it is proposed to base this conversion on the observed performances of the PAR on a learning set of manually labelled mail pieces. More precisely, we will use a *confusion matrix* counting the occurrences of different types of errors. This method is in the spirit of the work described in [27] in the context of character recognition. However, the problem considered here is much more complex because of the hierarchical structure of the address space, and the existence of different address categories (streets, P.O. boxes, ...). The method described here generalizes that introduced by the authors in [13].

10

4.2.1 Basic Approach

In this section, only postal addresses output by PARs are considered as input to the fusion process. The use of confidence scores will be addressed in the Section 4.2.2.

Each PAR provides an output o composed of a complete or partial address, i.e., a subset of Ω. Depending on the true value of the address, an output o, if not correct, may be partially correct. For instance, in Example 4, an output $o = \{T_1S_1\}$, when the truth is $\{T_1S_2\}$, is not correct. However, the town is correct: $\{T_1S_1\}$ and $\{T_1S_2\}$ are in the same set T_1 of level 2. We then say that the output o of level 1 is correct at level 2.

Formally, an output o of level p is said to be *correct at level* $q \geq p$, where q is the level of the first element in the hierarchy containing o and the truth. In the worst case, q is equal to P, where P is the number of levels in the hierarchy. An output of level p correct at level p is said to be *correct*. Let us introduce the following notations:

- $n_c^{p,q}$ denotes the number of mail pieces in the learning set classified in category c of level p that are correct at level q, with $q \in [p, P]$;

- H is the address hierarchy;

- u is the function that maps an element of H different from Ω to its parent element in H. For instance, in example 4, $u(\{T_1S_1\}) = T_1$ and $u(T_1) = \Omega$.

The total number of mail pieces classified at level p and in category c is then equal to $n_c^p = \sum_{q=p}^{P} n_c^{p,q}$.

When a PAR provides an output o of level p and category c, a simple approach could be to define the corresponding bba m as follows:

$$m(u^{q-p}(o)) = \frac{n_c^{p,q}}{n_c^p}, \quad \forall q \in [p, P] , \tag{8}$$

where

$$u^{q-p} = \underbrace{u \circ \ldots \circ u \circ u}_{q-p \ times},$$

and u^0 is the identity function. In other words, each parent of o at level $q \in [p, P]$ receives a mass equal to the number of mail pieces of the same level p and category c as o that were correctly classified at level q, divided by the total number of mail pieces

11

classified by the PAR at level p and category c. We note that the bba m computed using (8) is consonant.

Although Equation (8) seems reasonable, it can be refined to allow the allocation of belief masses to invalid addresses, based on past decisions made in such cases. An extension of the above scheme that addresses this issue may be defined as follows.

Let us call each element T_i inv an invalid address of level 1 and inv the invalid address of level 2. Let us denote by $n_c^{p,inv(q)}$ the number of time the truth is an invalid address of level q when the output is provided at level p and is of category c. When a PAR provides an output o of level p and category c, a bba m may be defined as:

$$m\left(u^{q-p}(o)\right) = \frac{n_c^{p,q} - n_c^{p,inv(q-1)}}{n_c^p}, \quad \forall q \in [p, P] , \tag{9}$$

$$m\left(\left[u^{q-p}(o)\right] inv\right) = \frac{n_c^{p,inv(q-1)}}{n_c^p}, \quad \forall q \in [p, P] , \tag{10}$$

where, by convention, $n_c^{p,inv(0)} = 0$, Ω $inv = inv$, and $\left[u^{q-p}(o)\right] inv$ denotes the invalid address associated to $u^{q-p}(o)$.

Applying assignment method (9)-(10) allows us to distinguish between two different hypotheses in case of rejection: either the true complete address exists but the PAR has not recognized it, or the true address is invalid.

Example 4 (continued) Let us consider a PAR with the confusion matrix shown in Table 1. This matrix describes the PAR performances on a learning set. For instance, this PAR proposed the address $\{T_1 S_1\}$ at level 1 for 95 mail pieces whose actual address was $\{T_1 S_1\}$, and for 2 mail pieces whose actual address was $\{T_1 S_2\}$. Likewise, it proposed the address T_1 at level 2, 24 times for mail pieces whose actual address was $\{T_1 S_1\}$, and 22 times for mail pieces whose actual address was $\{T_1 S_2\}$.

Let us suppose that this PAR provides an output $o = \{T_1 S_1\}$ for a new mail piece. This output belongs to level 1 and category $street$, denoted s. The number of correct outputs of level 1 and category s is equal to the sum of the numbers in bold in Table 1:

$$n_s^{1,1} = 95 + 88 + 98 = 281 . \tag{11}$$

Similarly, the number of outputs of level 1 and category s that are correct at level 2 (respectively, at level 3) is equal to the sum of the numbers in italics (respectively,

underlined) in Table 1:

$$n_s^{1,2} = 2 + 1 + 2 + 2 = 7 \tag{12}$$

$$n_s^{1,3} = 1 + 1 = 2 . \tag{13}$$

Finally, the number of outputs of level 1 and category s which were invalid at level 1 is $n_s^{1,inv(1)} = 2$. The bba m representing the piece of information provided by the PAR is thus given by:

$$m(\{T_1 S_1\}) = \frac{n_s^{1,1}}{n_s^1} = \frac{95 + 88 + 98}{290} = 0.969 ,$$

$$m(\{T_1 inv\}) = \frac{n_s^{1,inv(1)}}{n_s^1} = \frac{2}{290} = 0.007 ,$$

$$m(T_1) = \frac{n_s^{1,2} - n_s^{1,inv(1)}}{n_s^1} = \frac{2 + 1 + 2 + 2 - 2}{290} = 0.017 ,$$

$$m(\{inv\}) = \frac{n_s^{1,inv(2)}}{n_s^1} = \frac{0}{290} = 0.000 ,$$

$$m(\Omega) = \frac{n_s^{1,3} - n_s^{1,inv(2)}}{n_s^1} = \frac{1 + 1 - 0}{290} = 0.007 .$$

Likewise, assume that the same PAR provides an output $o = T_1$, which belongs to level 2 and category $town$ denoted t. We have

$$n_t^{2,2} = 24 + 22 + 3 + 20 + 13 + 16 = 98$$

$$n_t^{2,3} = 1 + 1 = 2$$

$$n_t^{2,inv(1)} = 3$$

$$n_t^{2,inv(2)} = 0.$$

Hence,

$$m(T_1) = \frac{n_t^{2,2} - n^{2,inv(1)}}{n_t^2} = \frac{98 - 3}{100} = 0.95 ,$$

$$m(T_1 inv) = \frac{n_t^{2,inv(1)}}{n_t^2} = \frac{3}{100} = 0.03 ,$$

$$m(\Omega) = \frac{n_t^{2,3} - n_t^{2,inv(2)}}{n_t^2} = \frac{2}{100} = 0.02 ,$$

$$m(\{inv\}) = \frac{n_t^{2,inv(2)}}{n_t^2} = \frac{0}{100} = 0 .$$

Finally, if $o = \Omega$, we have:

$$m(\{inv\}) = \frac{n^{3,inv(2)}}{n^3} = \frac{4}{16} = 0.25 \; ,$$

$$m(\Omega) = \frac{n^{3,3} - n^{3,inv(2)}}{n^3} = \frac{4+2+3+2+2+0+1+2-4}{16} = 0.75 \; .$$

4.2.2 Using confidence scores

In some cases, a PAR provides not only a partial or complete address, but also a *score* indicating the degree of confidence in the output. Such a score may typically be provided by character recognition algorithms. When available, this additional information may easily be incorporated in our framework, using the bba correction mechanisms introduced in Section 3.3.

A simple approach to use confidence scores is to reinforce the piece of information provided by a PAR when the score is high, and, conversely, to discount it when the score is low. For that purpose, we may define four thresholds T_i, $i = 1, 2, 3, 4$ such that information provided by the PAR is:

- totally discounted if the score is lower than T_1;

- discounted proportionally to the score if the score belongs to $[T_1, T_2]$;

- left unchanged if the score belongs to $[T_2, T_3]$;

- reinforced proportionally to the score if the score belongs to $[T_3, T_4]$;

- totally reinforced if the score is greater than T_4.

This may be achieved by defining three coefficient ν_1, ν_2 and ν_3 as functions of the score as shown in Figure 10, and use them to correct the output bba m using (4)-(5) as explained in Section 3.3.

Figure 11 shows a typical fusion scheme involving three PARs, two of which provide confidence scores that are used to correct their corresponding output bbas.

4.3 Combination

Once each PAR output has been converted into a bba, the different bbas can be combined using the conjunctive rule of combination (1).

14

Example 5 (Combination) Let us consider three PARs associated with outputs $o_1 = T_2$, $o_2 = T_1$, and $o_3 = \{T_2S_1\}$, respectively. Let us suppose that assignment method (9)-(10) has led to:

$$m_1(T_2) = 0.95, \quad m_1(\{T_2\ inv\}) = 0.03, \quad m_1(\Omega) = 0.02,$$

$$m_2(T_1) = 0.9, \quad m_2(\Omega) = 0.1,$$

$$m_3(\{T_2S_1\}) = 0.93, \quad m_3(T_2) = 0.05, \quad m_3(\Omega) = 0.02 \ .$$

Then, $m = m_1 \text{\char"24E9} m_2 \text{\char"24E9} m_3$ is given by:

$$
\begin{aligned}
m(\{T_2S_1\}) \quad &= \quad 0.0902, & m(T_1) \quad &= \quad 0.0004, \\
m(\{T_2inv\}) \quad &= \quad 0.0002, & m(\Omega) \quad &= \quad 0.0000, \\
m(T_2) \quad &= \quad 0.0068, & m(\emptyset) \quad &= \quad 0.9024 \ .
\end{aligned}
\tag{14}
$$

The large mass allocated to the empty set reflects the high conflict between the outputs from the three PARs.

4.4 Decision Making

As explained in Section 3.4, decision making in the TBM requires the definition of a set $\mathcal{D}$ of decisions, a betting frame Γ, and a cost function.

In this application, the definition of $\mathcal{D}$ is based on classifier outputs. Recall that each classifier output is an element for the hierarchy H of addresses, i.e., a subset of Ω. The set $\mathcal{D}$ of decisions is simply defined as the subset of 2^Ω composed of PAR outputs, their parents in the hierarchy, and Ω. For instance, in Example 5, the PARs outputs are $o_1 = T_2$, $o_2 = T_1$, and $o_3 = \{T_2S_1\}$. The corresponding possible decisions are $d_1 = \Omega$, $d_2 = T_1$, $d_3 = T_2$, and $d_4 = \{T_2S_1\}$.

The other ingredients of the decision model presented in Section 3.4, namely, the betting frame and the cost function, will be described in the following subsections.

4.4.1 Betting frame

The result of the combination is a bba m^Ω on Ω. As explained in Section 3.4, decision making in the TBM is based on the pignistic transformation applied after a betting frame has been defined. In this application, Ω cannot be chosen as the betting frame

because it is much too big. Additionally, the cardinalities of various subsets of Ω are unknown: we do not know exactly how many addresses are contained in a given town, for instance. Computing this kind of information from the database would be too time-consuming. For this reason, it was decided to define the betting frame Γ as a coarsening (i.e., a partition) of Ω. In order to avoid loosing information, Γ was chosen as the coarsest partition of Ω such that all non-empty focal sets of the combined bba m^Ω as well as their associated invalid addresses can be obtained as elements or unions of elements of Γ. The formal construction of Γ from Ω is detailed in Algorithm 1.

Algorithm 1 Construction of the Betting Frame Γ.

Require: m^Ω a bba; H a hierarchy;

    $\mathcal{F}(m) \leftarrow$ all the focal sets of m;

    $\mathcal{F}(m) \leftarrow \mathcal{F}(m) \cup \{\Omega\}$;

    for each element F of $\mathcal{F}(m)$ **do**

        add to $\mathcal{F}(m)$ the invalid address associated with F, if it is not already contained in $\mathcal{F}(m)$;

    end for

    $\Gamma \leftarrow \emptyset$;

    repeat

        $F \leftarrow$ a highest level element of $\mathcal{F}(m)$;

        $\gamma \leftarrow F \setminus \bigcup\{A \in \mathcal{F}(m)$ s.t. F is the smallest element of $\mathcal{F}(m)$ containing $A\}$;

        add γ to Γ;

        remove F from $\mathcal{F}(m)$;

    until $\mathcal{F}(m) = \emptyset$

Example 5 (continued) The non-empty focal sets of m given by (14) are $\{T_2 S_1\}$, T_1, $\{T_2 inv\}$, Ω and T_2. The invalid addresses associated to T_1 and Ω are $T_1 inv$ and inv, respectively. The corresponding betting frame Γ is represented in Figure 12. It

is composed of seven elements:

$$\begin{aligned}
\gamma_1 &= \Omega\backslash(T_2 \cup T_1 \cup \{inv\}) , & \gamma_2 &= \{inv\} , \\
\gamma_3 &= T_1\backslash\{T_1\ inv\} , & \gamma_4 &= T_2\backslash(\{T_2 S_1\} \cup \{T_2\ inv\}) , \\
\gamma_5 &= \{T_1\ inv\} , & \gamma_6 &= \{T_2 S_1\} , \\
\gamma_7 &= \{T_2\ inv\} .
\end{aligned} \tag{15}$$

It is easy to see that the focal sets of m may be recovered from the γ_i: we have

$$\{T_2 S_1\} = \gamma_6, \quad T_1 = \gamma_3 \cup \gamma_5, \quad \{T_2\ inv\} = \gamma_7,$$

$$T_2 = \gamma_4 \cup \gamma_6 \cup \gamma_7, \quad \Omega = \bigcup_{i=1}^{7} \gamma_i.$$

As explained in Section 3.2, the bba m on Ω given by (14) can be expressed on the coarsening Γ as:

$$\begin{aligned}
m^\Gamma(\{\gamma_6\}) &= 0.0902, & m^\Gamma(\{\gamma_3, \gamma_5\}) &= 0.0004, \\
m^\Gamma(\{\gamma_7\}) &= 0.0002, & m^\Gamma(\Gamma) &= 0.0000, \\
m^\Gamma(\{\gamma_4, \gamma_6, \gamma_7\}) &= 0.0068, & m^\Gamma(\emptyset) &= 0.9024 .
\end{aligned} \tag{16}$$

The pignistic probability computed from m^Γ using (7) is:

$$\begin{aligned}
BetP(\{\gamma_1\}) &= 0.000, & BetP(\{\gamma_2\}) &= 0.000 \\
BetP(\{\gamma_3\}) &= 0.002, & BetP(\{\gamma_4\}) &= 0.023 \\
BetP(\{\gamma_5\}) &= 0.002, & BetP(\{\gamma_6\}) &= 0.948 \\
BetP(\{\gamma_7\}) &= 0.025 .
\end{aligned}$$

4.4.2 Cost function

As explained in Section 3.4, the final decision d is the one that minimizes the risk defined as the expected cost with respect to the pignistic probability distribution. To compute this risk, we have to define a cost function $c : \mathcal{D} \times \Gamma \rightarrow \mathbb{R}$.

Let ω_0 denote the true address, γ_0 the element of Γ containing ω_0, and d_0 the smallest element of $\mathcal{D}$ containing γ_0. The costs are defined as follows. If $d = d_0$, then the decision is correct and the cost is equal to zero. If $d \neq d_0$, four situations are considered:

1. if $d = \Omega$ (the mail piece was rejected), then $c(d, \gamma) = C_R^{town}$; parameter C_R^{town} is called the *rejection cost at town level*;

2. if $d = T_i$ for some town T_i (only the town but no street or PO box was given) and $d_0 \not\subseteq d$ (the true address is not in town T_i), then $c(d, \gamma) = C_E^{town}$; parameter C_E^{town} is called the *error cost at town level*;

3. if $d = T_i$ for some town T_i and $d_0 \subset d$ (the true address is in town T_i), then $c(d, \gamma) = C_R^{dist}$; parameter C_R^{dist} is called the *rejection cost at distribution level*;

4. if $d = \{\omega\}$ for some $\omega \in \Omega$, $\omega \neq \omega_0$, then $c(d, \gamma) = C_E^{dist}$; parameter C_E^{dist} is called the *error cost at distribution level*;

Note that cases 2 and 4 correspond to errors (on the town or on the address within a town), whereas cases 1 and 3 correspond to situations where the decision is too imprecise, but not erroneous. Table 2 shows the costs associated with the decisions and betting frame elements of Example 5.

The cost function is thus defined using four parameters C_R^{town}, C_E^{town}, C_R^{dist} and C_E^{dist}. It is natural to assume the following ordering relation between these parameters:

$$0 \leq C_R^{dist} \leq C_R^{town} \leq C_E^{dist} \leq C_E^{town} . \tag{17}$$

For instance, an error in the distribution part of the address is less prejudicial than an error in the town part. Likewise, it is preferable to make a rejection than an error. Let us note, however, that this order may depend on postal agencies. For example, the error on the distribution when the town is correct may incur a smaller cost than the work implied by outright rejection. The rejection cost at the town level may be thus higher than the error cost at the distribution level.

Example 5 (concluded) With $I = \{1, \ldots, 7\}$, the expected risks of each decision are the following:

$$\begin{aligned}
\rho_{bet}(d_1) &= \sum_{\gamma \in \Gamma} c(d_1, \gamma) BetP(\gamma) = \sum_{i \in I \setminus \{1,2\}} C_R^{town} BetP(\{\gamma_i\}) \\
&\approx C_R^{town} , \\
\rho_{bet}(d_2) &= \sum_{\gamma \in \Gamma} c(d_2, \gamma) BetP(\gamma) = \sum_{i \in I \setminus \{3,5\}} C_E^{town} BetP(\{\gamma_i\}) \\
&\approx 0.996 \, C_E^{town} , \\
\rho_{bet}(d_3) &= \sum_{i \in \{1,2,3,5\}} C_E^{town} BetP(\{\gamma_i\}) + C_R^{dist} BetP(\{\gamma_6\}) \\
&\approx 0.004 \, C_E^{town} + 0.948 \, C_R^{dist} , \\
\rho_{bet}(d_4) &= \sum_{i \in \{1,2,3,5\}} C_E^{town} BetP(\{\gamma_i\}) + \sum_{i \in \{4,7\}} C_E^{dist} BetP(\{\gamma_i\}) \\
&\approx 0.004 \, C_E^{town} + 0.048 \, C_E^{dist} .
\end{aligned}$$

Then, according to the cost vector $\boldsymbol{c} = (C_R^{dist}, C_R^{town}, C_E^{dist}, C_E^{town})$, decision d_1, d_3 or d_4 can be made. As $C_E^{town} \geq C_R^{dist}$ (17), $\rho_{bet}(d_2) \geq \rho_{bet}(d_3)$, $\forall \boldsymbol{c}$.

For instance, if $\boldsymbol{c} = (1, 2, 3, 4)$ (resp. $(1, 2, 20, 40)$, $(1, 2, 100, 300)$), the final decision is $d_4 = \{T_2 S_1\}$ (resp. $d_3 = T_2$, $d_1 = \Omega$).

Ideally, the cost vector, reflecting real financial costs, should be provided by postal agencies. Unfortunately, such input is rarely available. However, as explained in Section 2.2, the goal of the postal fusion is to achieve the highest possible correct recognition rate associated with a satisfactory error rate. The costs can thus be determined using a set of labelled letters in order to obtain acceptable performances.

5 Experimental results

The fusion scheme described in this paper was applied to the combination of three PARs denoted $PAR\,1$, $PAR\,2$ and $PAR\,3$. A set of 56,000 mixed handwritten and machine printed mail pieces was divided into a learning set and a test set in equal proportions. The confusion matrix of each PAR was computed using the learning set, which was also used to tune the costs so as to achieve acceptable performances on the learning set. For different cost settings, error and correct recognition rates were then computed using the test set. Note that, in this real application concerning France, the address hierarchy H is more complex than the simple one considered above to explain the principles of the approach. This real hierarchy is depicted in Figure 13.

In addition to proposed addresses, PARs 1 and 2 also provide confidence scores in the case of handwritten mail, as considered in Section 4.2.2. Figure 14, which displays the confidence scores provided by PAR 1 for all handwritten mail pieces in the database, shows that errors at the town level tend to be associated with lower scores, which confirms that the score carries useful information. Figure 14 also shows the definition of the four thresholds T_i used for the correction mechanism detailed in Section 4.2.2.

Figures 15 and 16 show the performances of each of the individual PARs as well as various fusions schemes at the town and distribution levels. To preserve confidentiality on the performance level achieved, the origins of these two plots are not disclosed. Recognition rates, represented on the x-axis, are expressed relatively to a reference recognition rate, denoted R at the distribution level, and R' at the town level. Error rates, represented on the y-axis, are expressed relatively to a reference error rate, denoted E at both distribution and town levels. Rates R and R' are, respectively, greater than 50% and 80%, while E is smaller than 0.1%.

In addition to the three individual PARs, the performances of two reference fusion rules, noted Maj and C_{PAR1}, are plotted in Figures 15 and 16 for comparison. The former correspond to a simple majority voting scheme, whereas the latter is a variant given more importance to PAR 1 in the decision. The special role of PAR 1 in this combination rule was motivated by its good overall performances at the distribution and town levels.

The four combined PARs C_1 to C_4 were obtained using different cost settings, without using confidence scores. At the distribution level (Figure 16), each selected combination point has an acceptable error rate (i.e., an error rate lower than the minimum over individual PARs). At the town level (Figure 15), however, only combination points C_1 and C_2 remain under the maximal tolerated error rate. Overall, the cost setting corresponding to point C_2 meets the requirements mentioned in Section 2: it provides a combination scheme yielding more correct recognitions and fewer errors than each of the individual PARs, both at the town and distribution levels.

Finally, combined PARs with different cost settings and corrections based on confidence scores are also represented as crosses in Figures 15 and 16. We can see that

the use of scores further improves the performances of the combination. This fusion scheme allows us to obtain a combination point noted C_+, which is associated with an acceptable error rate and a higher correct recognition rate than previous combination point C_2, both at the town and distribution levels. This confirms the interest of using confidence scores in the fusion process.

6   Conclusion

A fusion scheme for postal address recognition based on the Transferable Belief Model has been presented. One of the key aspects of this scheme concerns the conversion of PAR outputs into belief functions, based on confusion matrices and confidence scores when available. Belief functions associated to the different PARS are then combined using Dempster's rule, and a final decision is made by minimizing the expected cost relative to a pignistic probability distribution defined on a suitable betting frame. This method has been shown experimentally to meet the requirements set by postal agencies.

In this work, PARs have been considered as black boxes, which allows great flexibility and modularity in the maintenance of the fusion system. For instance, PARs can be added or removed without modification of the whole system: only new confusion matrices having to be estimated. The fusion module is built on top of existing PARs and constitutes an independent component. Further improvements might be gained by using additional information provided by specific recognition algorithms inside each PAR, at the expense of some loss of generality. Situations where PARs provide more complex outputs such as ordered lists of addresses could also be studied. This is left for future research.

References

[1] T. Benouhiba and J.-M. Nigro. An evidential cooperative multi-agent system. *Expert Systems with Applications*, 30(2):255–264, 2006.

[2] I. Bloch. Fusion of information under imprecision and uncertainty, numerical methods, and image information fusion. In A.K. Hyder and al, editors, *Multisensor Data Fusion*, pages 267–294. Kluwer, 2002.

[3] I. Bloch. *Fusion d'informations en traitement du signal et des images*. Hermès, Paris, France, 2003.

[4] G. De Leo, M. Vicenzi, and C. Franzone. ELSAG SPA. European patent: Mail recognition method. Number: EP 1 594 077 A2, November 2005.

[5] M.H. DeGroot. *Optimal Statistical decisions*. McGraw-Hill, New York, 1970.

[6] T. Denœux and Ph. Smets. Classification using belief functions: the relationship between the case-based and model-based approaches. *IEEE transactions on Systems, Man and Cybernetics B*, 36(6):1395–1406, 2006.

[7] D. Dubois, H. Prade, and Ph. Smets. Representing partial ignorance. *IEEE Transactions on on Systems, Man and Cybernetics*, 26(3):361–377, 1996.

[8] M. Fisher. Siemens AG. International patent: System and method for smart polling. Number: WO 2005/050545 A1, June 2005.

[9] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.

[10] L.I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

[11] C. K. Lee and C. G. Leedham. A new hybrid approach to handwritten address verification. *International Journal of Computer Vision*, 57(2):107–120, 2004.

[12] J. A. Malpica, M. C. Alonso, and M. A. Sanz. Dempster-Shafer theory in geographic information systems: A survey. *Expert Systems with Applications*, 32(1):47–55, 2007.

[13] D. Mercier, G. Cron, T. Denœux, and M. Masson. Fusion of multi-level decision systems using the transferable belief model. In *Proceedings of the 8th International Conference on Information Fusion, FUSION'2005*, Philadelphia, USA, paper C8-2, July 25-29, 2005.

[14] D. Mercier, T. Denœux, and M. Masson. General correction mechanisms for weakening or reinforcing belief functions. In *Proceedings of the 9th International Conference on Information Fusion, FUSION'2006*, Florence, Italy, paper 146, July 10-14 2006.

[15] P. W. Palumbo and S. N. Srihari. Postal address reading in real time. *International Journal of Imaging Systems and Technology*, 7(4):370–378, 1996.

[16] R. Plamondon and S. N. Srihari. On-line and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):63–84, 2000.

[17] L.J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

[18] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

[19] I. U. Sikder and A. Gangopadhyay. Managing uncertainty in location services using rough set and evidence theory. *Expert Systems with Applications*, 32(2):386–396, 2007.

[20] Ph. Smets. Belief functions: the disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.

[21] Ph. Smets. What is Dempster-Shafer's model ? In R. R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer theory of evidence*, pages 5–34. Wiley, New-York, 1994.

[22] Ph. Smets. The Transferable Belief Model for quantified belief representation. In D. M. Gabbay and Ph. Smets, editors, *Handbook of Defeasible Reasoning and*

Uncertainty Management Systems, volume 1, pages 267–301. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[23] Ph. Smets. Decision making in a context where uncertainty is represented by belief functions. In R. P. Srivastava and T. J. Mock, editors, *Belief functions in business decisions*, pages 17–61. Physica-Verlag, Heidelberg, 2002.

[24] Ph. Smets. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005.

[25] Ph. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[26] S. N. Srihari. Handwritten address interpretation: A task of many pattern recognition problems. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(5):663–674, 2000.

[27] L. Xu, A. Krzyzak, and C.Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435, 1992.

[28] B. Yu, A. K. Jain, and M. Mohiuddin. Address block location on complex mail pieces. In *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, pages 897–901, 1997.

Figures
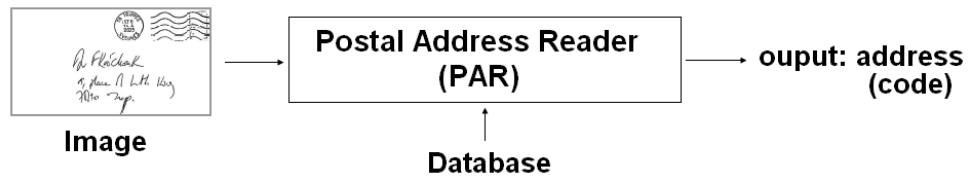
Figure 1: Generic representation of a postal address reader.
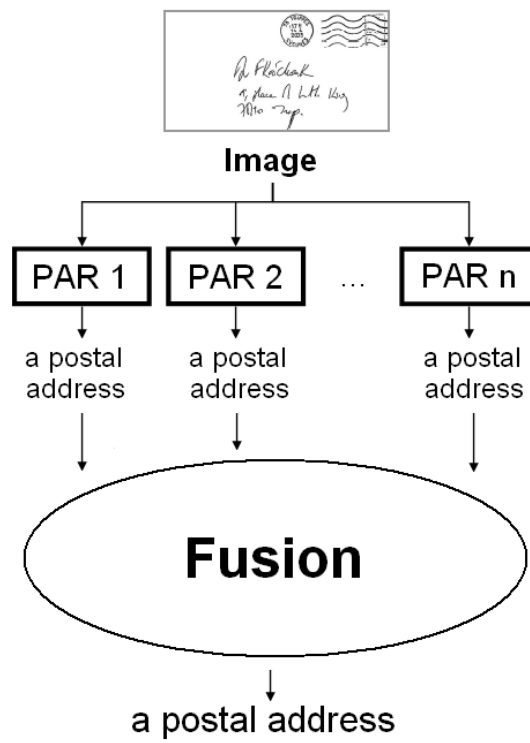
Figure 2: The PAR fusion problem.

(a) Postal addresses hierarchy (b) Overall hierarchy
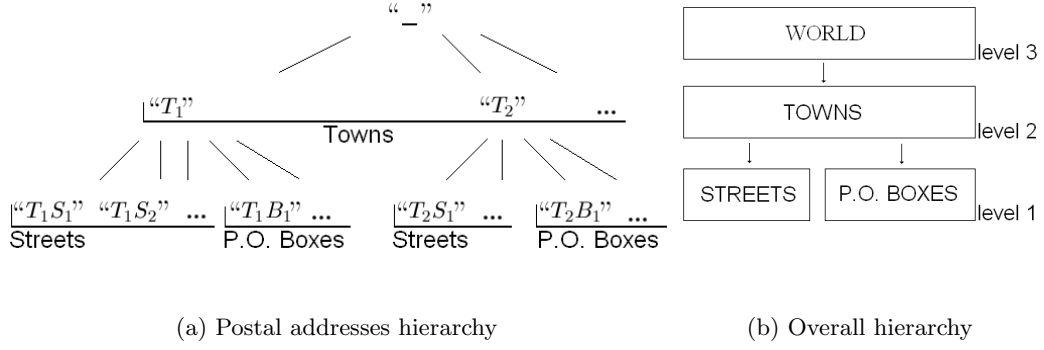
Figure 3: Hierarchy of postal addresses.
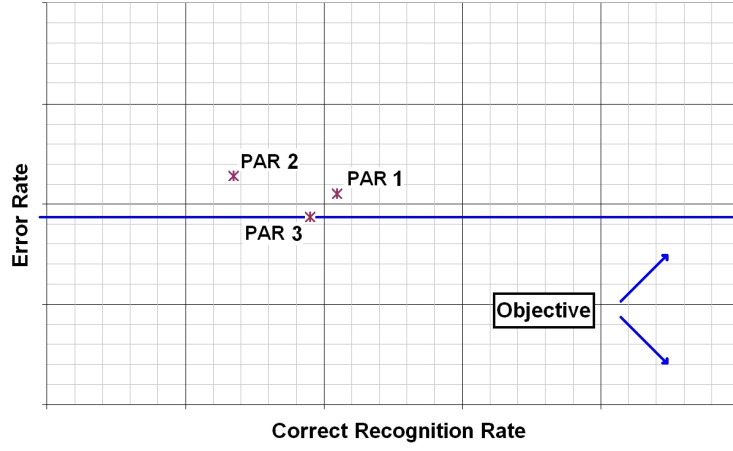
Figure 4: The objective of the fusion process: achieving the greatest possible correct recognition rate while keeping the error rate below that of the best individual PAR.
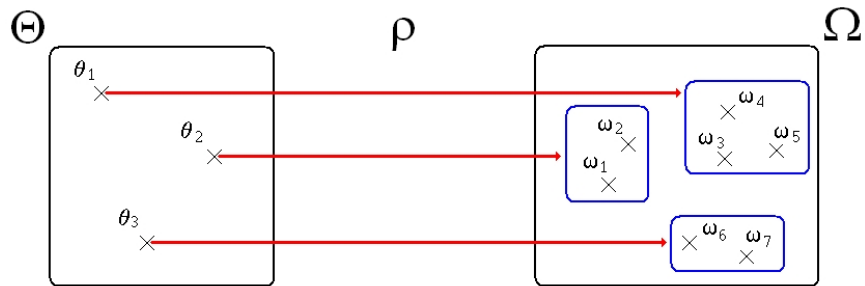

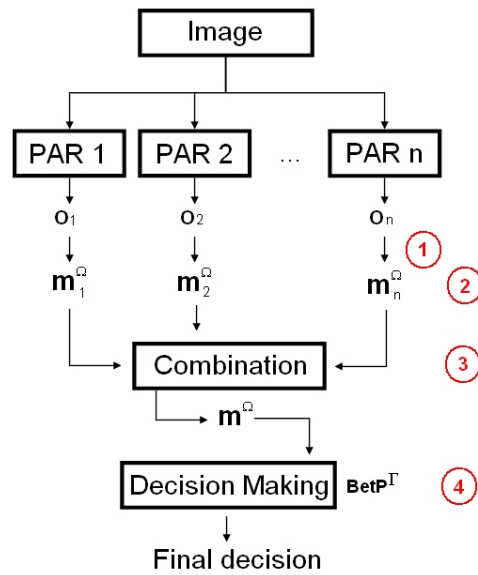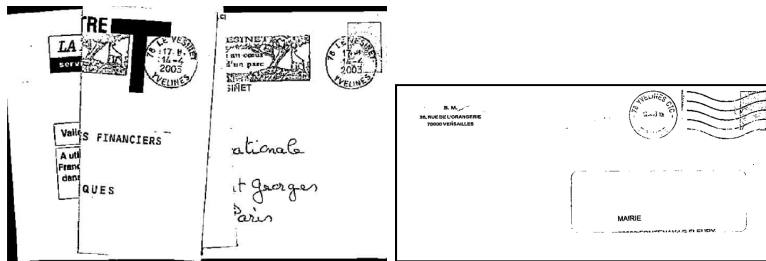Figure 5: A refining ρ from a frame Θ to a refinement Ω (Example 2).

26

Figure 6: Fusion scheme.

(a) A folded envelope.

(b) A letter slid into a window envelope.
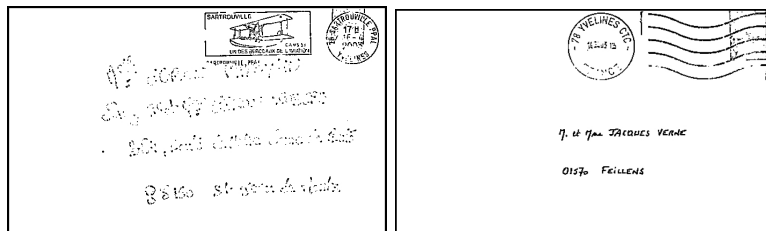
(c) Erased ink.

(d) A partial invalid address: the street is missing.

Figure 7: Some examples of invalid addresses.
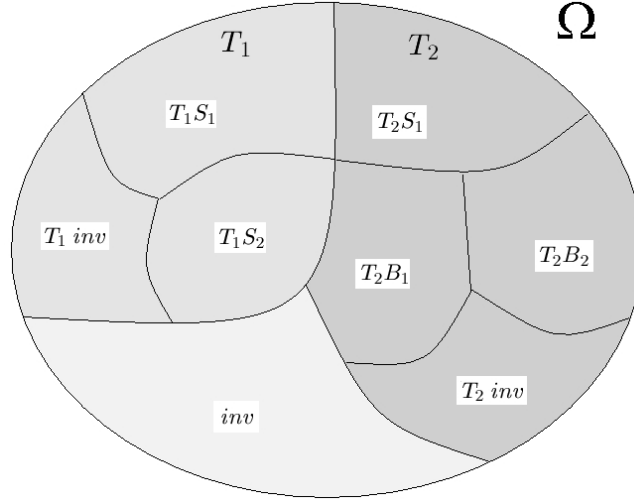
Figure 8: Frame of discernment associated with the small database of Example 4.
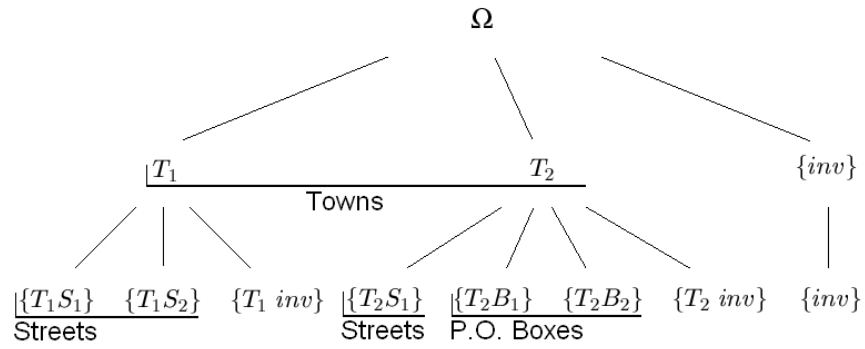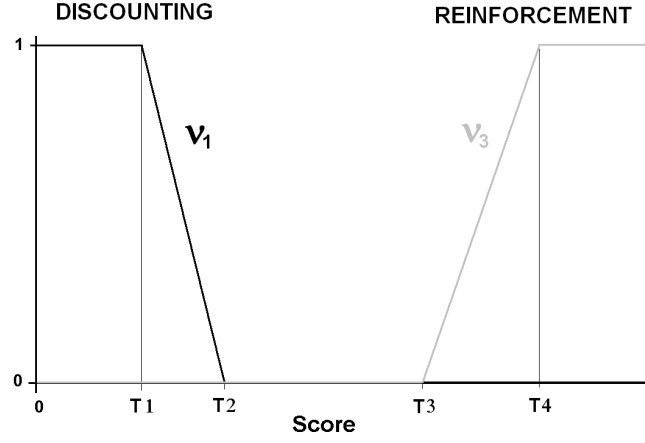

Figure 9: Address hierarchy of Example 4.

Figure 10: Coefficients ν_1 and ν_3 as functions of the score. The three coefficients ν_1, ν_2 and ν_3 are linked by the equation $\nu_1 + \nu_2 + \nu_3 = 1$.
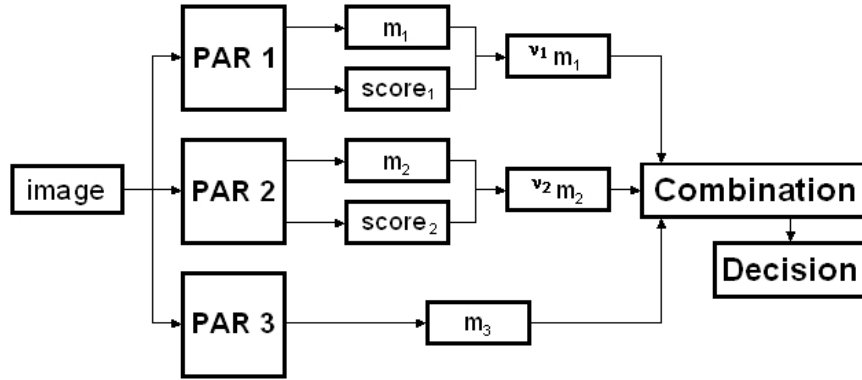

Figure 11: An extended model using a discounting/reinforcement correction mechanism based on the scores provided by $PAR\,1$ and $PAR\,2$.
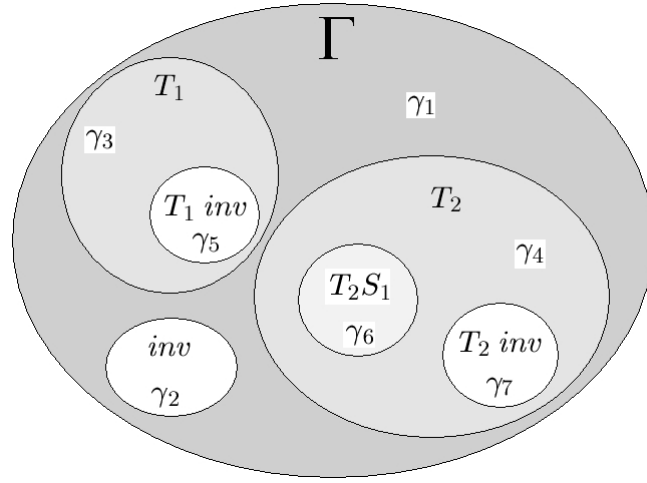
29

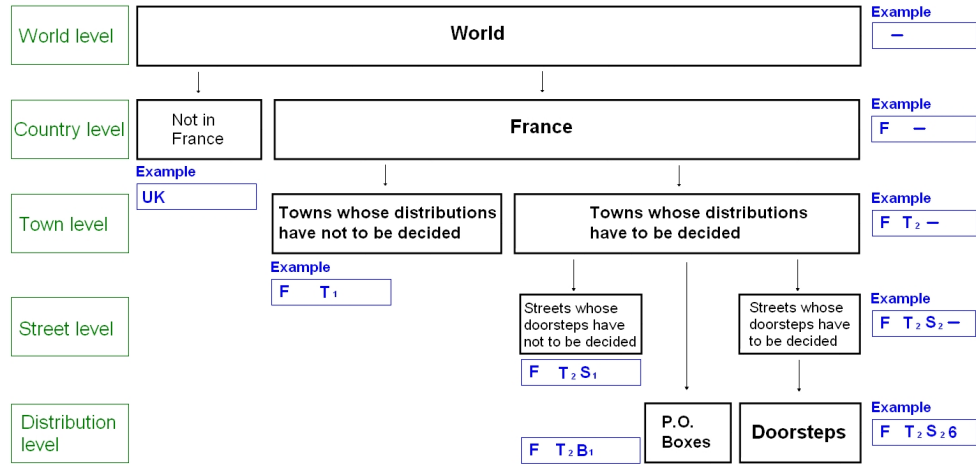Figure 12: Betting frame Γ for the bba m of Example 5.

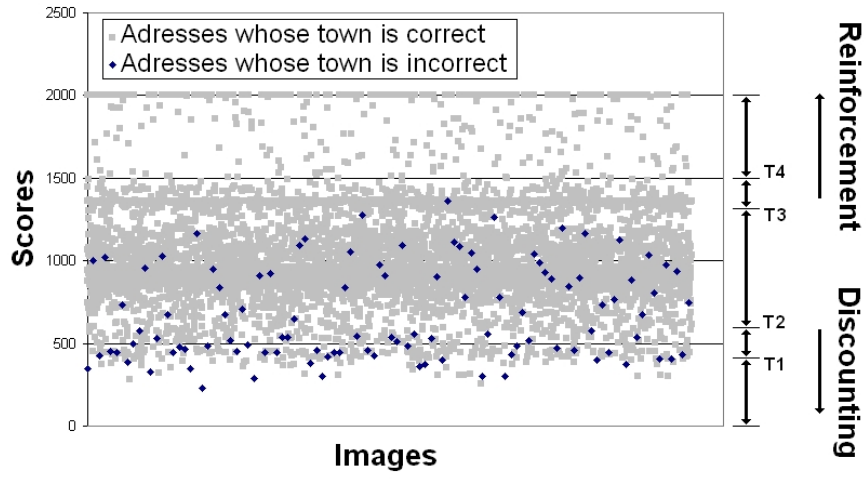Figure 13: Hierarchy of addresses in the real application.

Figure 14: Confidence scores and addresses provided by PAR 1 regarding the images of the learning set. A black dot corresponds to an address whose town is incorrect. A grey dot is associated with an address whose town is correct.
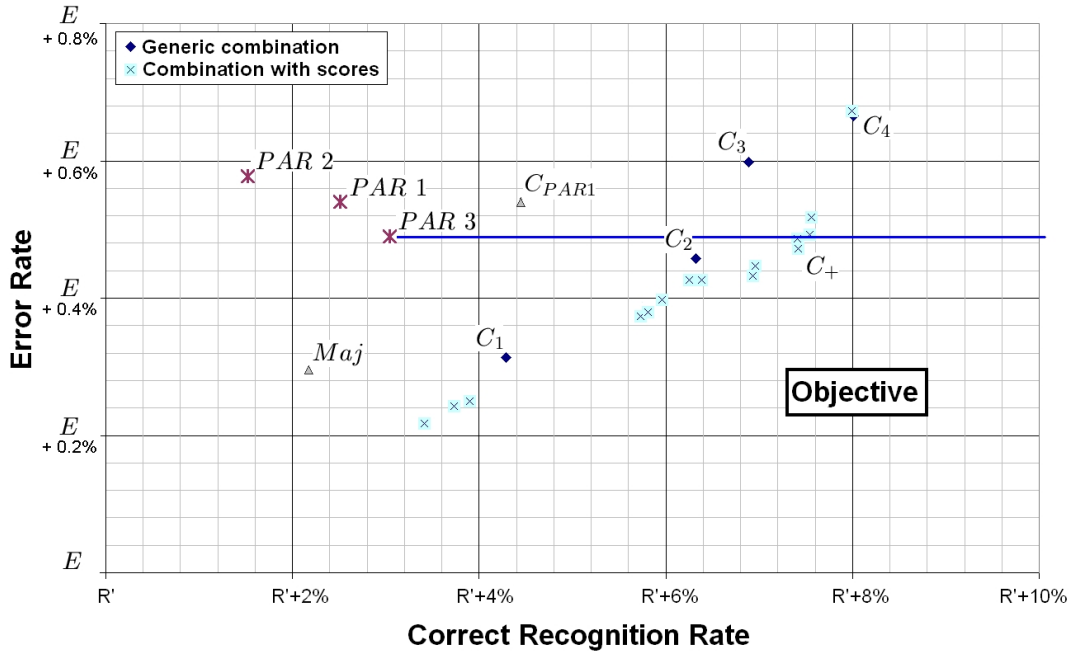

Figure 15: Performances of individual PARs and various fusion schemes on the test set at town level.
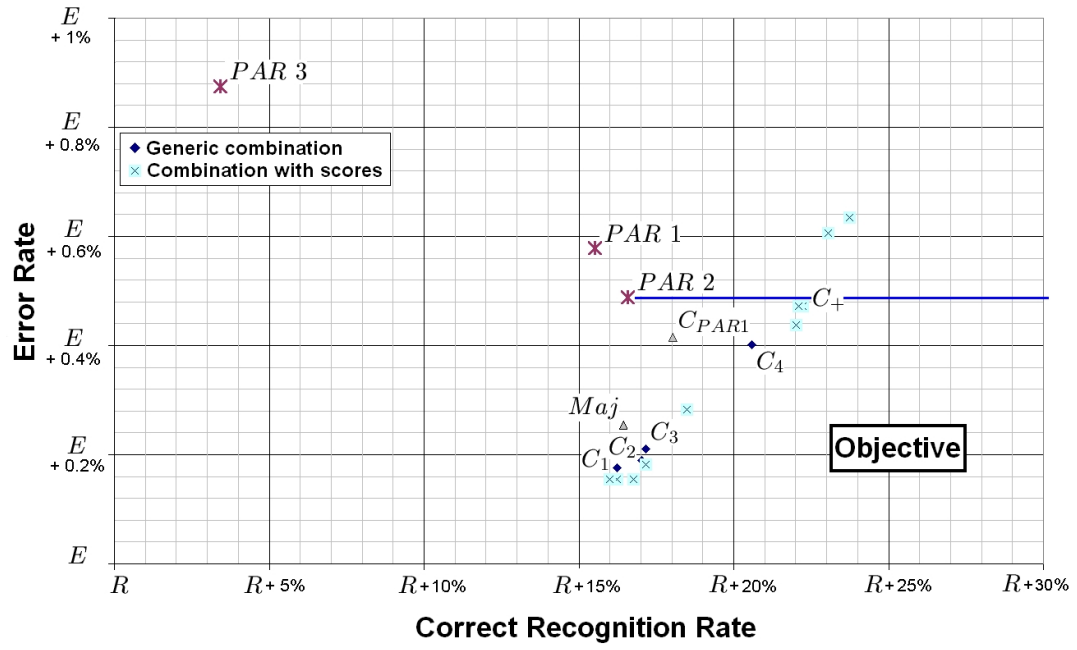
Figure 16: Performances of individual PARs and various fusion schemes on the test set at the distribution level.

Tables

Table 1: Confusion matrix associated with a PAR on Example 4. Notations: cat = category, pob = P.O. Boxes, s = streets, t = town.

Level	Cat	Truth / Output	inv	T_1S_1	T_1S_2	T_1inv	T_2S_1	T_2B_1	T_2B_2	T_2inv
1	s	T_1S_1	0	**95**	*2*	*0*	1	0	0	0
		T_1S_2	0	*1*	**88**	*2*	0	0	0	0
		T_2S_1	0	0	*1*	0	**98**	*2*	*0*	*0*
	pob	T_2B_1	0	0	0	0	1	55	1	0
		T_2B_2	0	0	0	0	0	0	46	0
2	t	T_1	0	24	22	3	1	0	0	0
		T_2	0	0	0	1	20	13	16	0
3		Ω	4	2	3	2	2	0	1	2

Table 2: Costs of making a decision d_i when the truth is an element γ_j of the betting frame in Example 5.

Truth / Decisions	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
$d_1 = \Omega$	0	0	C_R^{town}	C_R^{town}	C_R^{town}	C_R^{town}	C_R^{town}
$d_2 = T_1$	C_E^{town}	C_E^{town}	0	C_E^{town}	0	C_E^{town}	C_E^{town}
$d_3 = T_2$	C_E^{town}	C_E^{town}	C_E^{town}	0	C_E^{town}	C_R^{dist}	0
$d_4 = \{T_2S_1\}$	C_E^{town}	C_E^{town}	C_E^{town}	C_E^{dist}	C_E^{town}	0	C_E^{dist}