A FORMAL INDEXING MECHANISM FOR AN EMOTION-BASED AGENT

Rodrigo Ventura and Carlos Pinto-Ferreira
Institute for Systems and Robotics
Instituto Superior Técnico, Technical University of Lisbon
Av. Rovisco Pais, 1
1049-001 Lisboa, PORTUGAL
email: {yoda,cpf}@isr.ist.utl.pt

Abstract

The DARE architecture is an emotion-based agent model. This model is based on a double-representation of stimuli: a complex representation, oriented towards recognition, and simple one, oriented towards feature extraction. These two representations are associated and stored in the agent memory. This paper describes a formalization of an indexing mechanism. Given a new stimulus, the goal of this mechanism consists of finding in memory, efficiently, using the simple representation, the best matching item according to the complex representation. To assess the efficiency gain of this mechanism, an implementation was devised using the classic recognition problem of handwritten digits.

Key Words: Emotions, Agents, Representation.

1 Introduction

The fundamental idea of the DARE emotion-based architecture consists of representing stimuli under two different perspectives: a *cognitive* one, which is rich, complex, and oriented towards recognition, and a *perceptual* one, which is simple, basic, and oriented towards feature extraction and rapid response [1]. All further developments and implementations of DARE follow this basic idea. Some initial experimentation [2] followed the first ideas of DARE [1]. Márcia Maçãs improved the architecture and applied it to a labyrinth [3], and Pedro Vale has been exploring the learning capabilities of DARE [4]. The DARE architecture was also implemented in a real robot, with some interesting results [5]. A more sophisticated memory mechanism was proposed and applied to a control problem [6]. Sandra Gadanho has been exploring the combination of the architecture she previously developed, with DARE [7].

The double-representation paradigm of DARE is inspired on extensive work done by Antonio Damásio [8, 9]. He has hypothesized the idea of *somatic marker*, which consists of associating certain stimuli with desirable/undesirable body states. Failure to create and use these associations, namely in the case of lesions in the pre-frontal cortex, often leads to what Damásio calls “future myopia” [9]. This term refers to insensitivity to future consequences of certain decisions, mostly when the subject is faced with common-sense decision-making problems [8]. The idea of double-representation of stimuli, in the context of emotions, is not original. Joseph LeDoux has extensive work about double-pathways found in the brain [10].

On top of the double-representation idea, three mechanisms were hypothesized: the *marking* mechanism, which corresponds to associating, for a given stimulus, the cognitive with the perceptual representations, and storing them in memory; the *matching* mechanism, which searches the memory for a pair which cognitive and perceptual representations best match the ones extracted from a given stimulus; and the *indexing* mechanism which improves the efficiency of the matching mechanism by the means of exploiting the difference in terms of complexity of the two representations.

Research concerning emotions, in the context of AI, can be roughly divided in two domains: one concerning *external* manifestations of emotions (e.g., believable agents [11], affective computing [12]), and another one concerning its *internal* manifestations (e.g., alarm system [13], appraisal theory [14]). We position the present work in the context of this latter domain. Our research goal is to devise a formal model for a emotion-based agent, which is able to perform competently in complex and dynamic environments. There is some related research concerning the formalization of emotions. For instance, Zippora Arzi-Gonczarowski has been using mathematical category theory to model perceptions and emotions [15]. Piotr J. Gmytrasiewicz has also developed some research towards formalization of emotions [16], approaching it from a decision theoretic point of view.

The following section describes a formalization of the indexing mechanism. Some theoretical results are shown. In section 3 an implementation illustrates the gain in terms of efficiency of such a mechanism.

2 Proposed model

The agent model is based on a double representation of stimuli. Consider $\mathcal{S}$ to denote the set of all possible stimuli the agent can receive from the environment. At a time instant t, the agent extracts a cognitive representation, termed *cognitive image*, and a perceptual one, termed *perceptual image*. Considering the sets $\mathcal{I}_c$ and $\mathcal{I}_p$ to be the set of all possible cognitive and perceptual images, given a stimulus $s(t) \in \mathcal{S}$, the extraction of the cognitive and perceptual images $i_c(t) \in \mathcal{I}_c$ and $i_p(t) \in \mathcal{I}_p$ are formalized by the functions p_c and p_p:

$$\begin{cases} p_c : \mathcal{S} \longrightarrow \mathcal{I}_c \\ i_c(t) = p_c\left(s(t)\right) \end{cases} \tag{1}$$

$$\begin{cases} p_p : \mathcal{S} \longrightarrow \mathcal{I}_p \\ i_p(t) = p_p\left(s(t)\right) \end{cases} \tag{2}$$

Unless otherwise noted, the time dependency on t will be dropped, for clarity sake.

2.1 Marking

The marking mechanism simply associates, for a set of time instants for which the agent receives a stimulus from the environment, the cognitive and perceptual images extracted at those instants. Let $\mathcal{T}_M$ denote the set of time instants for which associations are formed and stored. The memory is then a set of pairs $\mathcal{M} \subseteq \mathcal{I}_c \times \mathcal{I}_p$

$$\mathcal{M} = \{(i_c(t), i_p(t)) \mid t \in \mathcal{T}_M\} \tag{3}$$

2.2 Matching

In order to be able to compare cognitive or perceptual images, a metric function will be used. The sets $\mathcal{I}_c$ and $\mathcal{I}_p$, each one equipped with a metric function, form metric spaces. We will designate these metric functions d_c and d_p, respectively:

$$d_c : \mathcal{I}_c \times \mathcal{I}_c \longrightarrow \mathbb{R}_0^+ \tag{4}$$

$$d_p : \mathcal{I}_p \times \mathcal{I}_p \longrightarrow \mathbb{R}_0^+ \tag{5}$$

both defined over the set $\mathbb{R}_0^+$ of non-negative real numbers. Both these metrics satisfy the usual metric axioms:

(i). $d(x,x) = 0$

(ii). $d(x,y) \geq 0$

(iii). $d(x,y) = d(y,x)$

(iv). $d(x,y) + d(y,z) \geq d(x,z)$

for d being either d_c or d_p, and for any x, y, and z in the respective space ($\mathcal{I}_c$ or $\mathcal{I}_p$).

The idea behind the cognitive and the perceptual images is for the former to be a complex representation of a stimulus, while the latter is a simple representation of the same stimulus. Therefore it is reasonable to view a perceptual image as a lower dimension projection of a higher dimensional cognitive image. In fact, we assume that there is a projection $p : \mathcal{I}_c \longrightarrow \mathcal{I}_p$ from the cognitive to the perceptual spaces such that p_p is the function composition of p_c followed by p:

$$p_p = p \circ p_c \tag{6}$$

With this idea in mind, we assume in this paper that the d_c and d_p metrics are such that:

$$d_c\left(p_c(s_1), p_c(s_2)\right) \geq d_p\left(p_p(s_1), p_p(s_2)\right) \tag{7}$$

for all $s_1, s_2 \in \mathcal{S}$. Intuitively this means that the projective nature of the perceptual space tends to keep closer two stimuli that are farther apart in the cognitive space, or in other words, the cognitive space has great resolution power than the perceptual space.

The goal of the matching mechanism is to find, in the memory $\mathcal{M}$, a pair which best matches a given stimulus received by the agent. Since we assumed that the cognitive image contains a richer representation of a stimulus, it makes sense to look for a match using the cognitive metric d_c. A cognitive match, for a given cognitive image i_c, is then defined as a minimization of the cognitive distance:

$$(i_c^*, i_p^*) = cm^*(i_c) = \underset{(i_c^M, i_p^M) \in \mathcal{M}}{\arg\min}\ d_c(i_c, i_c^M) \tag{8}$$

Using the perceptual representation (and metric) in this task would lead to poorer results, since it provides a coarse grained representation. However, due to its lower dimensionality, it would be computationally more efficient. The idea of the *indexing* mechanism, described in the next section, is to combine the efficiency of the perceptual representation with the fine grained accuracy of the cognitive one.

2.3 Indexing

The basic idea of the indexing mechanism consists of narrowing the search for a cognitive match function (8) to a subset of pairs in memory $\mathcal{M}$. We denote this subset as $S_p(i_p) \subseteq \mathcal{M}$, for a given perceptual image i_p, defined by

$$S_p(i_p) = \left\{(i_c^M, i_p^M) \in \mathcal{M} \mid d_p(i_p, i_p^M) \leq T_p\right\} \tag{9}$$

where T_p is some threshold. Having obtained $S_p(i_p)$, a cognitive match is then performed, restricted to that subset of pairs:

$$(i_c^+, i_p^+) = cm^+(i_c) = \underset{(i_c^M, i_p^M) \in S_p(i_p)}{\arg\min}\ d_c(i_c, i_c^M) \tag{10}$$

The efficiency gain of restricting the cognitive match to the $S_p(i_p)$ subset is as high as fewer pairs are contained in $S_p(i_p)$. The price to pay is the need to evaluate the perceptual distance $d_p(i_p, i_p^M)$ for all memory pairs in $\mathcal{M}$. That is why the perceptual representation is supposed to be simple, and the perceptual metric d_p fast to compute.

Using the assumption (7) made in the previous section, the following lemma can be trivially proved:

Lemma 1 *Given the d_c and d_p metrics satisfying the condition (7), and $S_p(i_p) \subset \mathcal{M}$ as defined in (9), whenever $d_c(i_c, i_c^M) \leq T_p$, we have $(i_c^M, i_p^M) \in S_p(i_p)$.*

Proof. Observing that $d_p(i_p, i_p^M) \leq d_c(i_c, i_c^M) \leq T_p$ we immediately have $(i_c^M, i_p^M) \in S_p(i_p)$ by definition of $S_p(i_p)$ in (9).

An interesting consequence of this lemma is that whenever the best cognitive match $(i_c^*, i_p^*) = cm^*(i_c)$ (from (8)) satisfies $d_c(i_c, i_c^*) \leq T_p$, we have necessarily that $(i_c^*, i_p^*) \in S_p(i_p)$ and therefore $(i_c^+, i_p^+) = cm^+(i_c)$ obtained from (10) is the same, *i.e.*, $(i_c^*, i_p^*) = (i_c^+, i_p^+)$. In other words, under the above conditions, we get an equally good cognitive match, using just the restricted set $S_p(i_p)$, thus preventing the calculation of the cognitive distance d_c for all memory pairs in $\mathcal{M}$, as in (8).

Predefining a value for the threshold T_p can be very problematic, since it depends on the metric properties of the domain, as well as on the cognitive and perceptual representations. On one hand, for a too high value of T_p, the $S_p(i_p)$ degenerates to $S_p(i_p) = \mathcal{M}$, if $d_p(i_p, i_p^M) \leq T_p$ for all $(i_c^M, i_p^M) \in \mathcal{M}$. On the other, for a too low value of T_p, it may happen that the desired i_c^* is such that $d_c(i_c, i_c^*) > T_p$, not only preventing the application of lemma 1, but also possibly leading to $(i_c^*, i_p^*) \notin S_p(i_p)$.

We call this strategy of predefining a threshold T_p, *thresholding*. To tackle the difficulty of pre-defining such a threshold value, an alternative strategy is proposed, which we call *N-best:* instead of obtaining $S_p(i_p)$ from T_p, the idea is to include in $S_p(i_p)$ the N_p memory pairs with the lowest perceptual distance $d_p(i_p, i_p^M)$. This means that

$$T_p \in \mathbb{R}_0^+ \quad \text{such that} \quad |S_p(i_p)| = N_p \tag{11}$$

assuming $|\mathcal{M}| \geq N_p$. This results on an upper bound to the number of cognitive distances d_c to be calculated[1].

We can then find the best cognitive match (i_c^+, i_p^+) in $S_p(i_p)$ using (10). Suppose now that there is one memory pair $(i_c^S, i_p^S) \in S_p(i_p)$ such that

$$d_p(i_p, i_p^S) \geq d_c(i_c, i_c^+) \tag{12}$$

Since $S_p(i_p)$ was defined in such a way that, for all $(i'_c, i'_p) \in \mathcal{M} \setminus S_p(i_p)$,

$$d_p(i_p, i'_p) \geq d_p(i_p, i_p^S) \tag{13}$$

it follows that, for all $(i'_c, i'_p) \in \mathcal{M} \setminus S_p(i_p)$

$$\begin{array}{rcll} d_c(i_c, i'_c) & \geq & d_p(i_p, i'_p) & \text{by (7)} \\ & \geq & d_p(i_p, i_p^S) & \text{by (9)} \\ & \geq & d_c(i_c, i_c^+) & \text{by (12)} \end{array} \tag{14}$$

which means that $d_c(i_c, i'_c) \geq d_c(i_c, i_c^+)$. These two relations obtained by transitivity of the inequalities have an interesting consequence: *the* (i_c^+, i_p^+) *pair minimizes the cognitive distance* $d_c(i_c, i_c^M)$ *over the whole memory* $\mathcal{M}$*, in other words,* $(i_c^+, i_p^+) = cm^+(i_c) = (i_c^*, i_p^*) = cm^*(i_c)$. This result proves the following lemma:

[1]This assertion fails if there are several memory pairs with the same perceptual distance value. Choosing $S_p(i_p)$ with the N_p best perceptual matches does not yield a unique solution in this case. And there is no T_p value such that expression (9) results in the same $S_p(i_p)$.

Lemma 2 *For a subset* $S_p(i_p) \subset \mathcal{M}$ *as defined in (9), and the minimizations in (8) and (10), whenever* $d_p(i_p, i_p^S) \geq d_c(i_c, i_c^+)$ *for some pair* $(i_c^S, i_p^S) \in S_p(i_p)$*, we have* $cm^+(i_c) = cm^*(i_c)$.

Both the thresholding and N-best strategies can be seen as stop criteria of the cognitive matching mechanism. Given a stimulus s, the agent calculates the perceptual distances of the extracted i_p to the ones stored in memory $\mathcal{M}$. The goal of the thresholding and N-best strategies is to select a subset $S_p(i_p)$, which will be used to perform the calculations of the cognitive distances and their minimization.

The lemma 2 can be used as a third stop criterion. The $S_p(i_p)$ subset can be constructed incrementally: first the memory pair with the smallest perceptual distance $d_p(i_p, i_p^M)$, then with the second best, and so on. Whenever the hypothesis of the lemma 2 is met, we have the guarantee that the cognitive match $cm^+(i_c)$, in the subset $S_p(i_p)$, is indeed the best one.

Note that the condition (7) can be replaced by

$$\lambda d_c(i_{c1}, i_{c2}) \geq d_p(i_{p1}, i_{p2}) \tag{15}$$

for some positive λ, which is the same as scaling the cognitive metric by a scalar λ. If condition (7) is not satisfied by some pair of metrics d_c and d_p, it may happen that for some sufficiently large value of λ, condition (15) is satisfied. However, if we re-write the hypothesis of lemma 2 using the λ scaling value:

$$d_p(i_p, i_p^S) \geq \lambda d_c(i_c, i_c^+) \tag{16}$$

we can observe that, as λ increases, this condition becomes more difficult to satisfy. By difficult we mean that more memory pairs need to be accumulated in $S_p(i_p)$, in order to satisfy (16).

However, this argument is reversible in the following sense: assuming that a pair of metrics already satisfy condition (7), we may be able to choose a value of λ, between 0 and 1, such that not only condition (15) is true, but also the inequality (16) is more easily satisfied. In other words, by scaling the cognitive metric, a stop criteria based on lemma 2 may improve the efficiency of the indexing mechanism. The drawback of scaling the cognitive distance is that it may break condition (7).

Let us summarize the three strategies of constructing $S_p(i_p)$ that we have proposed in this section:

- The *thresholding* strategy allows us to adjust up to which cognitive distance we want to obtain the best cognitive match. If we set T_p to that distance, the subset $S_p(i_p)$ is constructed by the means of perceptual distances, and then, by lemma 1, we have the guarantee that, if the best cognitive match distance does not exceed T_p, that match is in $S_p(i_p)$. However, the choice of a threshold value T_p is very sensitive to the domain, in terms of numerical values of the distances: too low or too high values of the threshold can lead to degenerated subsets $S_p(i_p)$ (empty or equal to $\mathcal{M}$);

- The *N-best* strategy solves the domain dependency problem of the *thresholding* by constructing $S_p(i_p)$ based on the N_p best perceptual matches, rather than on thresholding metric values. Moreover, depending on the computational resources and/or time available to process the stimulus, $S_p(i_p)$ can contain more or less pairs in $S_p(i_p)$ to perform the cognitive match. The drawback is that there is no guarantee of finding the best cognitive match (lemma 1 does not apply);

- The strategy based on lemma 2 can be used in conjunction with the previous one, in the sense that when the conditions of the lemma are met, there is no use on adding more pairs to $S_p(i_p)$ (assuming its incremental construction). However, depending on how larger the cognitive distances are, with respect to the perceptual ones, these conditions may never be met. Moreover, we can scale the cognitive metric in order to facilitate the satisfaction of the hypothesis of lemma 2. The drawback of this scaling is that its value is domain dependent.

In the following section, the ideas developed above are put into practice in an implementation. The results presented below show some interesting results which illustrate the gains, in terms of efficiency, that can be obtained with the indexing mechanism.

3 Illustrative Example

To experiment with the indexing mechanism formulated above, a simple example was devised. The problem consists of the classical hand-written digit recognition problem. Each digit consists of a binary image, and is classified with the respective digit symbol: 0 to 9. The task is to perform recognition using the emotion-based architecture, comparing the performance of the pure cognitive match (exhaustive search comparing cognitive images) with the guidance provided by the indexing mechanism. A stimulus is considered successfully recognized when its digit symbol is the same as the one of the best match.

3.1 Implementation

The cognitive image is the binary image itself ($i_c = s$, *i.e.*, p_c is the identity function, and $\mathcal{I}_c = \mathcal{S}$). Considering W to be the width and H the height (in pixels) of the images, the stimuli and the cognitive images have the form:

$$i_c = s = \begin{bmatrix} b_{11} & \cdots & b_{1W} \\ \vdots & \ddots & \vdots \\ b_{H1} & \cdots & b_{HW} \end{bmatrix} \tag{17}$$

where $b_{kl} \in \{0,1\}$ (for $k = 1,\ldots,H$ and $l = 1,\ldots,W$). The perceptual image is a vector of size W (same as the images width) constructed by counting the number of "1" pixels for each column, having the form:

$$i_p = \begin{bmatrix} n_1 & \cdots & n_W \end{bmatrix} \tag{18}$$

where each $n_k \in \mathbb{N}_0$ (non-negative integers) is calculated by the following expression:

$$n_k = \sum_{l=1}^{H} b_{lk}, \quad k = 1,\ldots,W \tag{19}$$

The perceptual metric d_p is a simple Euclidean distance between two vectors (the superscripts A and B distinguish each vector involved)

$$d_p(i_p^A, i_p^B) = \sqrt{\sum_{k=1}^{W} \left(n_k^A - n_k^B\right)^2} \tag{20}$$

while the cognitive metric corresponds to the Hamming distance between two binary images:

$$d_c(i_c^A, i_c^B) = \sum_{k=1}^{W}\sum_{l=1}^{H} \left|b_{lk}^A - b_{lk}^B\right| \tag{21}$$

These two metrics satisfy the metric axioms: the perceptual metric (20) is trivial, since it is an Euclidean norm; it is fairly easy to check that the cognitive one (21) also verifies them.

Considering two stimuli s^A and s^B, the cognitive and perceptual images extracted are denoted $i_c^A = p_c(s^A)$, $i_p^A = p_p(s^A)$, $i_c^B = p_c(s^B)$, and $i_p^B = p_p(s^B)$. Let us expand the following expression, which we call X,

$$X = \left[d_c(i_c^A, i_c^B)\right]^2 - \left[d_p(i_p^A, i_p^B)\right]^2 \tag{22}$$

Using the definitions for d_c, d_p, and n_k above, we obtain

$$X = \left[\sum_{k=1}^{W}\sum_{l=1}^{H} \left|b_{lk}^A - b_{lk}^B\right|\right]^2 - \sum_{k=1}^{W}\left[\sum_{l=1}^{H}\left(b_{lk}^A - b_{lk}^B\right)\right]^2 \tag{23}$$

Since $|a+b| \leq |a| + |b|$, and by induction $\left|\sum_i x_i\right| \leq \sum_i |x_i|$, we obtain the following inequality

$$X \geq \left[\sum_{k=1}^{W}\left|\sum_{l=1}^{H}\left(b_{lk}^A - b_{lk}^B\right)\right|\right]^2 - \sum_{k=1}^{W}\left[\sum_{l=1}^{H}\left(b_{lk}^A - b_{lk}^B\right)\right]^2 = Y \tag{24}$$

also defining Y as the right side of the inequality. Now considering that if $a, b \geq 0$, then $(a+b)^2 = a^2 + 2ab + b^2 \geq a^2 + b^2$, and by induction $\left(\sum_i x_i\right)^2 \geq \sum_i x_i^2$, we can write

$$Y \geq \sum_{k=1}^{W}\left[\sum_{l=1}^{H}\left(b_{lk}^A - b_{lk}^B\right)\right]^2 - \sum_{k=1}^{W}\left[\sum_{l=1}^{H}\left(b_{lk}^A - b_{lk}^B\right)\right]^2 = 0 \tag{25}$$

Therefore $X \geq Y \geq 0$, which is the same to say that

$$\left[d_c(i_c^A, i_c^B)\right]^2 \geq \left[d_p(i_p^A, i_p^B)\right]^2 \tag{26}$$

and since the metrics satisfy the metric axiom (ii), we can conclude that

$$d_c(i_c^A, i_c^B) \geq d_p(i_p^A, i_p^B) \tag{27}$$

which satisfies condition (7). This means that the theoretical results obtained in section 2 can be applied.

3.2 Results

In the following experiments a well-known test-set[2] was used. This test-set consists of 1934 samples of handwritten digits (0 to 9), scanned into binary images of 32 by 32 pixels ($W = H = 32$). From these samples, a training and a test set were randomly picked up, forming two disjoint sets. This corresponds to the usual cross-validation procedure.

The training process consists of running through all elements of the training set, and for each one of them, storing in memory the pair of the cognitive and perceptual images extracted. In order to evaluate recognition success ratio, the corresponding digit symbol was also attached to each pair.

Four matching mechanisms were tested. The results shown below were obtained averaging 10 trials, each one using disjoint training and test sets, containing 1500 and 200 digits respectively, randomly chosen from the pool of 1934 patterns.

1. Pure perceptual matching — the memory pair which the *perceptual image* is closer to the one extracted from the stimulus;

$$(i_c^{(1)}, i_p^{(1)}) = \arg\min_{(i_c^M, i_p^M) \in \mathcal{M}} d_p\left(i_p, i_p^M\right) \tag{28}$$

2. Pure cognitive matching — the memory pair which *cognitive image* is closer to the one extracted from the stimulus;

$$(i_c^{(2)}, i_p^{(2)}) = \arg\min_{(i_c^M, i_p^M) \in \mathcal{M}} d_c\left(i_c, i_c^M\right) \tag{29}$$

3. Guided cognitive matching *(indexing)*, using *thresholding* — like in the pure cognitive matching but where the set of memory pairs used is restricted by a pure perceptual matching. This restriction is based on thresholding the perceptual distances;

$$\begin{cases} (i_c^{(3)}, i_p^{(3)}) = \arg\min_{(i_c^M, i_p^M) \in S_p(i_p)} d_c\left(i_c, i_c^M\right) \\ S_p(i_p) = \left\{(i_c^M, i_p^M) \in \mathcal{M} \mid d_p(i_p, i_p^M) \leq T_p\right\} \end{cases} \tag{30}$$

4. Guided cognitive matching *(indexing)*, using *N-best* — like the one above, but the restriction corresponds to choosing the N_p best perceptual matches.

$$\begin{cases} (i_c^{(4)}, i_p^{(4)}) = \arg\min_{(i_c^M, i_p^M) \in S_p(i_p)} d_c\left(i_c, i_c^M\right) \\ T_p \in \mathbb{R}_0^+ \quad \text{such that} \quad |S_p(i_p)| = N_p \end{cases} \tag{31}$$

The implementation of the last two matching mechanisms was based on the following algorithm: *for a* (i_c, i_p) *pair extracted from a given stimulus,*

[2]Optical Recognition of Handwritten Digits, from E. Alpaydin, C. Kayna, URL: `ftp://ftp.ics.uci.edu/pub/machine-learning-databases/optdigits/` .

1. Calculate the perceptual distance $d_p(i_p, i_p^M)$ for all pairs in memory $(i_c^M, i_p^M) \in \mathcal{M}$;
2. Sort the pairs in memory $\mathcal{M}$ in increasing order of its perceptual distance $d_p(i_p, i_p^M)$;
3. Scan the obtained list until the corresponding stop criterion is met: threshold value, or number of best matching pairs.

To have an idea on the relationship between the cognitive and perceptual distances obtained after the sorting operation in step 2, figure 1 plots these two distances for a randomly chosen input stimulus, from the test set.

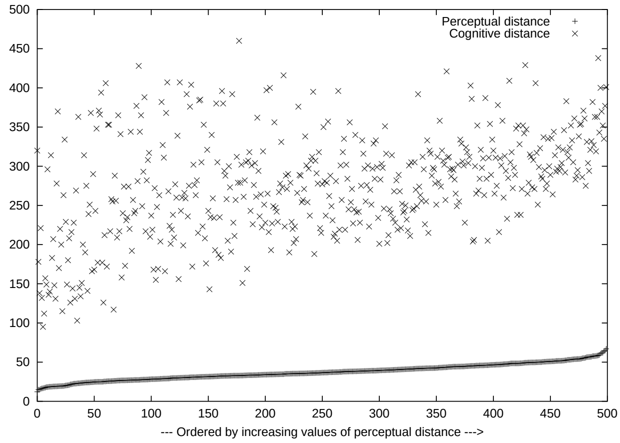


Figure 1. Cognitive and perceptual distances of a typical stimulus, with respect to the list of memory pairs, sorted by increasing perceptual distances $d_p(i_p, i_p^M)$. The lower trace corresponds to the perceptual distances.

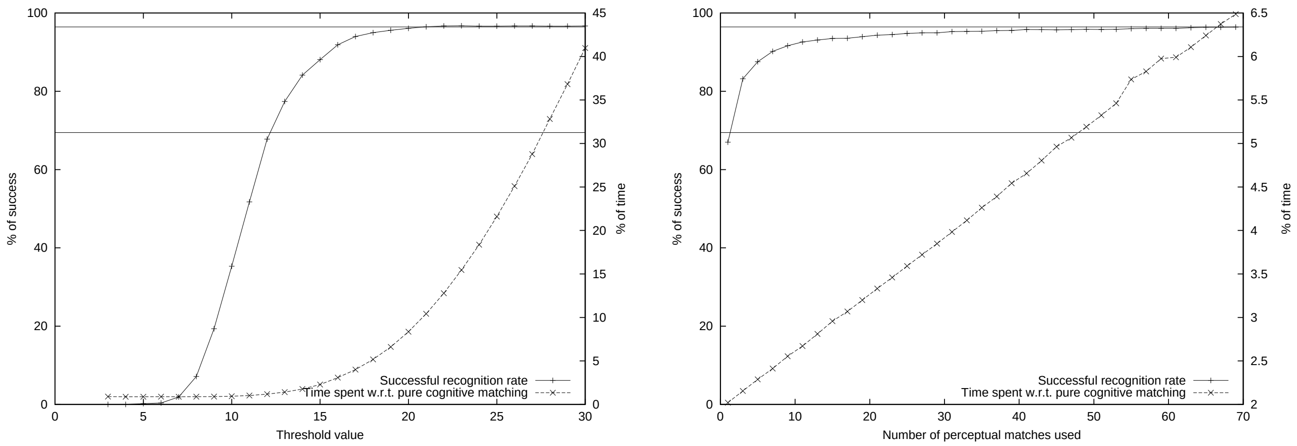
Table 1 shows the results for the first two tests. There is a clear trade-off between an extremely slow cognitive matching, with a high success rate, and the fast perceptual match leading to poor results.

mechanism	min (%)	avr (%)	max (%)	time
cognitive	94.0	96.45	100.0	130
perceptual	66.0	69.45	72.0	1

Table 1. Results for the pure cognitive and perceptual matching: the minimum, maximum, and average success rates for all trials, and the time ratio (perceptual=1).

With respect to the other two tests, which evaluate the indexing mechanism in this domain, the plots in figure 2 show how the success rates depends on a parameter. Using thresholding (figure 2a), the parameter is the threshold value (T_p), and using N-best (figure 2b), the parameter is the number of closest perceptual matches considered for indexing (N_p).

The two plots shown in figure 2 express basically the same outcome, since they both result from the indexing mechanism. What makes them different is the dependency on the parameter: in (a) the dependency on T_p



(a) (b)

Figure 2. Success rates obtained using the indexing mechanism: (a) *thresholding*, as function of the threshold value; (b) *N-best*, as function of the number of perceptual matches used. The values for execution time are expressed as a percentage of the time taken by the pure cognitive match (first line of table 1). In both plots, the two horizontal lines denote the average success rate for the pure cognitive (higher) and pure perceptual (lower) matching mechanisms. Note the difference scales in the rightmost axis (time).

is explicit, while in (b) the relationship is implicit. It is easy to realize that there is a non-linear monotonic relation between the horizontal axis, since each value of T_p leads to some number of pairs in $S_p(i_p)$. And this number of pairs increases monotonically with T_p.

In the (b) plot, the relative execution time increases linearly with the N_p parameter, because it determines how many cognitive distances have to be calculated. However, it is interesting to note that the success rate rises above 90% when just about 10 perceptual matches (in $S_p(i_p)$) are used. At this point, the cognitive match is using about 2.5% of the time taken by a pure cognitive match.

Up to now, the lemma 2 was not used. One possible use of it is as a stop criterion for an incremental construction of $S_p(i_p)$. An initial direct use in this implementation led to poor results: the subset $S_p(i_p)$ often degenerated to $\mathcal{M}$, because the hypothesis of the lemma was rarely met. Therefore, we tried scaling the cognitive metric as in (16). Of course, in doing so, the condition (27) may not be true. Figure 3 shows the results in function of λ, using this strategy.

This plot shows that the results are very sensitive to the λ parameter. For too low values of λ, (27) is easily satisfied, leading to the same success rate as a pure perceptual match. For too high values of λ, the stop criterion tends to be never used, degenerating in a slow pure cognitive match (the processing time raises at a significant rate, in direction of 100%). However, for a range of λ values it is possible to obtain very good results, keeping the processing time at low levels. Note that in these results, the stop criterion based on lemma 2 replaces entirely the thresholding and the N-best strategies.

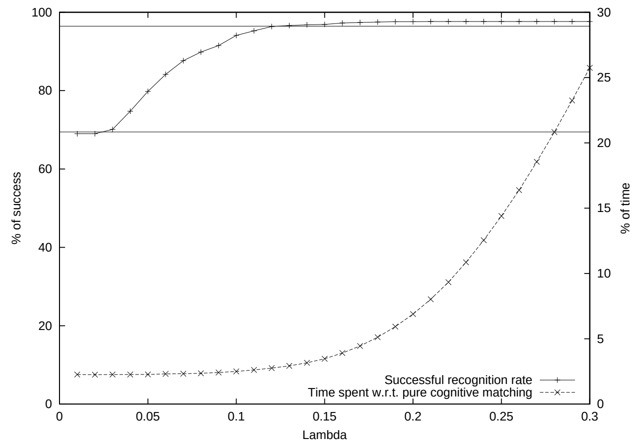


Figure 3. Results obtained using lemma 2 as a stop criteria (see figure 2 for further details).

4 Conclusions and future work

This paper presents the first steps towards the formalization of the DARE architecture. The underlying idea of DARE consists of manipulating two distinct representations of stimuli. A cognitive one, which is complex and rich, but slow to process, and a perceptual one, which is simple, but fast to process. This paper covers the formalization of the indexing mechanism, which consists of using the perceptual (simple) representation to guide the search for a cognitive (rich) match.

Section 2 has shown some theoretical results which can be drawn from a few assumptions on the (metric) structure of the cognitive and perceptual representations. Then, section 3 presented an implementation, along with experimental results, that illustrate some of the ideas raised by the theoretical discussion. Specifically, significant efficiency gains were obtained: for instance, identical recognition success rates (about 95%) were attained with just about 5% of the time taken by an exhaustive cognitive search. This corresponds to restricting the cognitive match to a few tenths of memory pairs, from a pool of 1500 pairs in memory.

Let us stress the fact that in this paper, the goal is not to obtain a good recognition rate. The recognition rates obtained are sole merit of the Hamming distance used in the cognitive metric[3]. The goal of the indexing mechanism is rather to approach the level of performance of the cognitive metric, *without the necessity of evaluating the cognitive metric for all memory pairs.* A good indexing mechanism should obtain the same results of a pure cognitive match, with much less calculations of the cognitive metric.

Future research, in the context of this paper, includes several topics. We are interested in extending the formalization to other mechanisms refered in section 2. An important research direction consists of closing the loop with the environment. With respect to the indexing mechanism, there are some possibilities open. For instance, it may be interesting to explore the consequences of modifying the structure of the perceptual image, namely by adding or removing components to it. Such modifications may be combined with an incremental construction of the agent memory $\mathcal{M}$. Another interesting topic corresponds to the possibility of categorizing the cognitive image, by the means of the perceptual representation.

Acknowledgments

The authors are grateful for the suggestions and comments made by Luis Custódio and by Sandra Gadanho on an early version of this paper.

References

[1] Rodrigo Ventura and Carlos Pinto-Ferreira. Emotion-based agents. In *Proceedings AAAI-98*, page 1204. AAAI, AAAI Press and The MIT Press, 1998.

[2] Rodrigo Ventura and Carlos Pinto-Ferreira. Emotion-based agents: Three approaches to implementation (preliminary report). In Juan Velásquez, editor, *Workshop on Emotion-Based Agent Architectures (EBAA'99)*, pages 121–129, May 1999.

[3] Márcia Maçãs, Rodrigo Ventura, Luis Custódio, and Carlos Pinto-Ferreira. Experiments with an emotion-based agent using the DARE architecture. In *Proceedings of the Symposium on Emotion, Cognition, and Affective Computing (AISB'01 Convention)*, UK, March 2001.

[3] In the test set used, the Hamming distance attains a good performance level. However, the Hamming distance is unable to deal with invariance to translation, rotation, and scaling of the digits.

[4] Pedro Vale and Luís Custódio. Learning individual basic skills using an emotion-based architecture. In *Proceedings of the Symposium on Emotion, Cognition, and Affective Computing, AISB'01 Convention*, 2001.

[5] Rui Sadio, Goncalo Tavares, Rodrigo Ventura, and Luis Custódio. An emotion-based agent architecture application with real robots. In *Emotional and Intelligent II: The Tangled Knot of Social Cognition*, 2001 AAAI Fall Symposium Series, pages 117–122. AAAI, 2001.

[6] Rodrigo Ventura, Luis Custódio, and Carlos Pinto-Ferreira. Learning courses of action using the "movie-in-the-brain" paradigm. In *Emotional and Intelligent II: The Tangled Knot of Social Cognition*, 2001 AAAI Fall Symposium, pages 147–152. AAAI, 2001.

[7] Sandra Clara Gadanho. Emotional and cognitive adaptation in real environments. In Robert Trappl, editor, *Cybernetics and Systems 2002*, pages 762–767. Austrian Society for Cybernetic Studies, 2002. Proceedings of EMCSR-2002, Vienna, Austria.

[8] Antonio R. Damásio. *Descartes' Error: Emotion, Reason and the Human Brain*. Picador, 1994.

[9] Antoine Bechara, Hanna Damásio, Daniel Tranel, and Antonio R. Damásio. Decising advantageously before knowing the advantageous strategy. *Science*, 275:1293–1295, February 1997.

[10] Joseph LeDoux. *The Emotional Brain*. Simon & Schuster, 1996.

[11] W. Scott Reilly. *Believable Social and Emotional Agents*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, May 1996. Technical Report CMU-CS-96-138.

[12] Rosalind W. Picard. Affective computing. Technical Report 321, M.I.T. Media Laboratory; Perceptual Computing Section, November 1995.

[13] Aaron Sloman. Beyond shallow models of emotion. In *i3 Spring Days Workshop on Behavior planning for life-like characters and avatars*, Sitges, Spain, March 1999.

[14] N. H. Frijda. *The Emotions*. Cambridge University Press, Editions de la Maison des Sciences de l'Homme, Paris, 1986.

[15] Zippora Arzi-Gonczarowski. Wisely non-rational — a categorical view of emotional cognitive artificial perceptions. In Dolores Cañamero, editor, *Emotional and Intelligent: The Tangled Knot of Cognition*, pages 7–12, 1998.

[16] Piotr J. Gmytrasiewicz and Christine L. Lisetti. Using decision theory to formalize emotions for multiagent systems. In *Second ICMAS-2000 Workshop on Game Theoretic and Decision Theoretic Agents*, Boston, 2000.