Available online at www.sciencedirect.com

ScienceDirect

Pattern Recognition Letters 29 (2008) 1024–1032

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Learning to learn: From smart machines to intelligent machines

B. Raducanu [a,*], J. Vitrià [a,b]

[a] Computer Vision Center, Edifici "O" – Campus UAB, 08193 Bellaterra, Barcelona, Spain
[b] Computer Science Department, Autonomous University of Barcelona, 08193 Bellaterra, Barcelona, Spain

Available online 14 September 2007

Abstract

Since its birth, more than five decades ago, one of the biggest challenges of artificial intelligence remained the building of intelligent machines. Despite amazing advancements, we are still far from having machines that reach human intelligence level. The current paper tries to offer a possible explanation of this situation. For this purpose, we make a review of different learning strategies and context types that are involved in the learning process. We also present the results of a study on cognitive development applied to the problem of face recognition for social robotics.

Keywords: Intelligent systems; Cognitive development; Context; Social robotics; Face recognition

1. Introduction

The golden dream of artificial intelligence (AI) remains to design and build systems showing human-like intelligence. Nowadays, the machines can perform remarkable things: there are chess algorithms able to play at international masters complexity levels, applications to coordinate the deployment of troops on the battle field, computer aided tools which allow us to design from the most powerful microprocessors to the most sophisticated airplanes. But, on the other hand, despite of the high complexity of the previously mentioned systems, none of them is able to, for instance, interpret the objects that appear in an image, comment a story, answer a question, in general things that would not assume any difficulty for a normal person. At first sight, we cannot say we lack of necessary tools or we lack of high-skilled people (engineers and researchers) in order to develop systems which show a certain degree of intelligence (actually, there are a huge number of examples in the area of expert systems). On the other hand, the field of Artificial Intelligence is full of theories and learning algorithms. At this point, the natural question that arises is: what failed in AI's strategy?

One of the main reasons is that all the previously mentioned examples are task-oriented. In other words, the efforts of artificial intelligence were split in many different subdomains: computational linguistics, planning, computer vision, etc. All these areas focuses on separate tasks like inheritance, classification, control and ignore the richness and complexity of the human mind. One of the first who noticed this trend in AI was Allen Newell. In Newell (1973), he was criticizing the study of isolated components of human mind without considering their interaction. He also pointed out the fragmented character of AI research. As alternative, he proposed closer ties between AI and cognitive psychology.

For instance, we usually identify objects in several ways:

- by their appearance (color, luminance, shape, texture);
- by their physical properties (mass, elasticity);
- their presence depends on the spatial/temporal context;
- dynamic behavior which depends on how the action has been applied to them;
- their use.

* Corresponding author. Tel.: +34 93 581 2440; fax: +34 93 581 1670.
E-mail addresses: bogdan@cvc.uab.es (B. Raducanu), jordi@cvc.uab.es (J. Vitrià).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2007.09.003

In profound contrast with these complex aspects, we remind the story that took place at MIT's AI lab more than four decades ago. In the summer of 1966, Marvin Minsky asked one of his students, Gerald Sussman, to solve the following problem: connecting a TV camera to a computer, write a program in order to *categorize* the objects that appear in the scene.[1] Even if the outcome of this attempt proved to be finally unsuccessful, it served to show the general underestimation of the difficulties involved in simulating the cognitive processes. This failure can be explained by the fact that classical AI researchers adopted the hypothesis stated by Newell and Simon who claim that humans use symbolic systems to *think* (Newell and Simon, 1961).

Apart from the previous aspect of the relational representation of knowledge, another element to be taking into account refers to the temporal dimension, how the stored information changes over time and how new information is integrated with the existing one. Humans are endowed with an evolutionary mechanism, which allows them to learn since the birth time. This process is a continuous, incremental one which takes place all along our lifetime. Although some initial specifications of the learning scheme are codified in our genes, a crucial factor in our development is represented by our everyday experiences acquired both via our social interactions and via our interactions with the environment where we live. In Goren et al. (1975), Johnson et al. (1991), the authors claim that the newborn babies arrive to this world pre-wired with the ability to recognize face-like patterns. It looks like that they are attracted by moving stimuli which resemble human faces. Later on, and according with the evolution of our cognitive abilities, we learn to distinguish different subclasses within face class: males/females, young/mature/old, familiar/unfamiliar, etc. (de Gelder and Rouw, 2001). Opposite to this cognitive development, in classical AI the systems were pre-programmed for the detection of certain patterns both in time and frequency and correlations among them. This approach implied an explicit representation of the world, since the programmer was in charge to collect hand-picked samples of data. It was unlikely that with the human intervention meet the demands of many challenging cognition tasks that are critical for generating intelligent behavior. The architecture of such a system was a closed one, without any chance to evolve and adapt. In consequence, it was impossible to behave beyond what it was designed for. The paradox here is that the development of such a system was denied by the programmer himself.

These differences mentioned above between the classical AI and the cognitive perspective are summarized in Table 1.

In the current paper we propose a simple way to model the cognitive development for a social robot. The purpose of the study is to show how is possible to incrementally build a knowledge database (represented by human faces), starting with a limited set of data. The study was carried out using an AIBO robot (2000) which has the advantage of being biologically-inspired. The aim in the near future is to have the AIBO behaving in a personalized manner, depending on the frequency it sees a certain person. Thus, we expect the robot to develop a 'friendlier' attitude towards persons who are frequently seen, meanwhile to act more 'reserved' in front of a person who has been seen less frequent.

Table 1
Comparative between classical approach of intelligent systems and the cognitive perspective

Classical paradigm	Cognitive perspective
Centered on the programmer	Centered on the system
The programmer possess the knowledge	The system discovers and build the knowledge
The programmer provides the resources	The system finds the resources
Off-line learning	On-line learning
The programmer 'teaches' the system	The system learns by itself
Learning process is limited and isolated	Learning process is continuous and contextual
The system carries programmer's vision	The system develops its own'self'

The paper is structured as follows: in Section 2, we review the existing learning theories. Section 3 is dedicated to present the role of 'context' in the learning process. In Section 4 we present a study which is currently under way and is about the development of a cognitive model through incremental learning applied to the problem of face recognition by a social robot. Finally, in Section 5 we will draw some conclusions and present the guidelines for future work.

2. Learning theories

The term 'learning', from the Artificial Intelligence perspective, is referring to the ability of a machine to acquire some knowledge in order to improve its functionality/behavior over time. The questions we face at this moment are: Why we pretend the machines have the ability to learn? Why do we not design a machine which has the desired functionality from the very beginning? The answers to these questions come from several directions: some knowledge can only be acquired based on real-world data; it can be a direct relationship between the input and output data (or sometimes, a hidden correlation); the environment where the system is placed is changing over time or the structure of the represented data needs to be changed. The learning, in this case, comes from several domains: statistics, cognitive models, adaptive control theory, psychology, etc. In consequence, there is a very close relationship between data acquisition, knowledge representation and learning strategies.

[1] Of course, this still remains a big challenge nowadays.

2.1. Symbolic representation

In the traditional methodology of Artificial Intelligence, the case of problem solving was based on functional decomposition (Fodor, 1983) and data abstraction through symbolic representation (Newell and Simon, 1961). Based on it, the programmers built a 'world-model'. When the system was requested to give an answer, it was trying to find an one-to-one correspondence between the input pattern and the inference engine (set of rules), which represented the knowledge database. The description of the 'reasoning' process involved in this case could be summarized through the expression: 'sense-think-act'.

The main source of error was represented by data abstraction. For this purpose, were used some symbols which had nothing to do with the real world. These symbols existed only from the programmer's perspective. Through data abstraction, it was impossible for him/her to foresee all the possible situations which might appear during system functioning. The explanation resides in the fact that this way, the problem complexity was very much simplified. The architecture for this type of systems was a centralized one (that's why the expression 'brain-in-a-box') and it was quite often that the system got stuck. This was the case when it was presented with a request which was not foreseen at the programming stage and a result it had no representation in its knowledge database. A classical example is given by MYCIN system (Shortliffe, 1976). MYCIN was an expert system used in diagnostic of bacterial infections. This system had no idea about what and how a person is. If you told it, for instance, that the person had a cut and is bleeding, the system was trying to find a bacterial cause to this problem.

An explanation of the failure of these systems is given by the fact that the human mind does not represent the information only by its category, but also by the mean it was acquired. In other words, it depends on the sensorial mechanisms that contributed to its acquisition (Brooks and Stein, 1993).

2.2. Behavior-based systems

Facing the limitations of systems based on symbolic representation, it was obvious that a new paradigm was needed. In Brooks (1991), the author claimed that an intelligent system does not need a centralized representation of the 'world'. The idea behind his affirmation was that system's behavior could be represented only by relating directly the input data to its output, without the need of a centralized representation (knowledge database). This paradigm was referred with the expression: 'sensation-to-action'. In this case, the system had a distributed architecture, consisting of several modules each of them being responsible to perform a very simple behavior. The high level behaviors were obtained as a consequence of the combination of several low level behaviors. This paradigm aroused from Minsky's idea to explain human intelligence. In his vision, the mind was formed by a set of 'agents' that compete and collaborate between each other.

Without no explicit data representation, we can not talk about a possibility to measure directly the knowledge capacity of the system. In change, we can evaluate its learning (experience) level by analyzing its behavior. We can say that a system learned, when it manifests a change in its behavior. In consequence, a fundamental requirement for these systems is to place them directly in the environment where they will exist. That's because they develop their behavior-based on the direct interaction with their surroundings.[2]

2.3. The cognitive approach

The lack of explicit representation of knowledge in the case of behavior-based systems has a drawback: it is very difficult to say if the modification in behavior is exclusively due to the change of the environment or is due because of some internal changes of the system. For this reason, the researchers proposed the following solution: we have to focus also on the internal modifications of data representation (Omrod, 1999). With this new paradigm, the system can 'learn' without noticing any apparent changes. The cognitive approach consists of an integrated, recursive process that aims at building a model of the 'world' and at a continuous adaptation of this model. In consequence, it is the system itself which is responsible of how to analyze, interpret and represent the information. The system will learn new concepts (develop new competencies) based on previous data and the experience acquired over time. When a new piece of information becomes available, it is responsible to analyze it and in case it is relevant, should be added to the existing representation (at times, might be necessary a change of representation structure). This way, the system could present two classes of behaviors: one class, consisting of specifically learned behaviors and another one, corresponding to emergent behaviors. This cognitive learning strategy is depicted in Fig. 1.

In the case of cognitive learning, a very important characteristic is represented by information management. We have to distinguish between two types of memory: a short-term and a long-term memory (Atkinson et al., 1968; Palmer, 1999). The short-term memory is responsible for maintaining the information for a very brief period of time (acting like a buffer). On the other hand, the long-term memory represents the knowledge database built over time. If the information passed from short-term memory is relevant, than the knowledge content of long-term memory is updated. The long-term memory is responsible for guaranteeing the system viability over large period of times (weeks, maybe months). A phenomenon that can affect the long-term memory is the 'forgetting' or 'degradation'

[2] This idea is related with the well-known 'mind-body' problem, which refers to the fact that the human mind can not process beyond what is provided by the sensorial mechanisms attached to it.

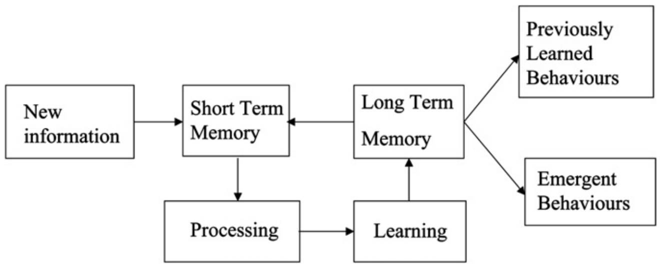

Fig. 1. Cognitive learning strategy.

(partial or total loss of some data). Sometimes, this process can be irreversible. These are also characteristics of human mind.

A key component of the cognitive learning process is represented by context: the set of factors that are determinant for the acquisition, representation and retrieval of information. The next section offers a discussion about this subject.

3. Context

After (Winograd, 2001), the term 'context' has its roots in linguistics. The word is composed of two parts: 'con' and 'text' and refers at the meaning extracted from a text. Nowadays, this term is used in a broader sense and has adopted in several domains (Brézillon, 1999).

In Artificial Intelligence, the notion of 'context' is associated with the 'frame problem': the challenge of representing the effects of action in logic without having to represent explicitly a large number of intuitively obvious non-effects. Is one of the most difficult problems with which Artificial Intelligence was confronted.

In case of written or verbal communication,'context' is referring to some properties of the interaction process that takes place between several agents. From this perspective, we can not separate 'context' from 'interaction'. 'Context' can be interpreted as the 'history' of what occurred during the period of time since the interaction has begun, the agents knowledge and the particularities derived from the interaction process. In this light, 'context' can be understood as a 'space of shared knowledge'. As a result of this observation, we can develop tools (assistants) which has the task to predict the intentions and actions of other agents. In the written communication, for instance, examples can be given related with text editors: as soon as we introduce a few letters of a word, the assistant displays a list of possible suggestions. Regarding the verbal communication, having an 'implicit' knowledge of what the conversation is about, can help us to eliminate a lot of uncertainty and ambiguity.

In the area of computer vision, we can refer to 'context' as the area surrounding an object as well as the intrinsic properties of the object. For artificial vision system, object recognition is still a partial-solved problem. Seen from the Artificial Intelligence perspective, the task of scene description for instance, (in terms of identifying the objects presented in the scene), remains one of the most challenging aspects. Among the factors that make difficult for an artificial vision system the task of visual recognition are: significant variations in illumination conditions, changes in object appearance (depending on the viewpoint), distortions, linear transformations, partial occlusions, etc. For this reason, the use of 'context' in computer vision is aimed to simplify the process of object recognition.

Experiments in scene analysis (Biederman et al., 1982) confirmed that the human visual system uses the notion of 'context' in an extensive manner to facilitate object detection and recognition. Furthermore, in the real world exists a very close relationship between the object and its surroundings. For this reason, the decision about the presence or the absence of an object in the scene is greatly influenced by it; on the other hand, the presence of different type of objects can be strongly related: for instance, if we can detect a table in the scene, we are also hoping to find a chair. It seems that initially, the human visual system performs a global analysis of the scene in order to estimate the object that might appear. The importance of using context to identify the existing objects is proved in two situations (Torralba and Sinha, 2001; Torralba et al., 2003): first, when the objects properties are partially observable or when they are affected by noise; and second, assuming that the identification was done successfully, the context can help eliminating the uncertainties (errors) in the object classification.

4. Aiboface: A case-study of cognitive development in robots

4.1. Overview

Since the beginning, the computational systems have been always focused towards machine and not people. Until now, human–computer interaction has assumed that the user must know the technical details of the machine, has to work on its terms, using its language and some specific devices to be able to communicate with it (keyboard, mouse, etc.). In case of Virtual Reality, things are even worst: we are immersed in a synthetic world, created by machine.

But the future is about to bring a fundamental change in the human–computer interaction, in the sense that the focus will be the user and not the machine. In other words, the user does not have to worry anymore how the machine is built or how it works. In change, the machine, through its perceptive capabilities, will have to identify the presence of a person in its neighborhood and to be aware when it becomes user's focus of attention, in order to respond to his/her demands (Weiser, 1991).

As part of this vision, the social robotics is a currently emerging field whose aim is the study and implementation of richer, natural forms of interaction between persons and robots. One area of application for this technology would be the assistance for elderly people living alone. Social and psychological experiments carried out in USA and Japan proved that these robots not only could be of help for improving the mood of these persons (Wada et al., 2005), but can also alert the competent services in case of an emergency. Another area of application for social robots is represented by their therapeutic use for autistic children. Psychological tests carried out in pediatric clinics revealed that children affected by autism have difficulties in relating with other people, but they show no fear when interacting with a robot (Yokoyama, 2002). An explanation of this behavior is that they hesitate to look in the eyes of another person, but have no problem looking in a robot's camera.

One of the most important tasks for a social robot is represented by the human presence detection. Faces represent by far the most distinguishable cue to assess a person presence. Psychological research revealed that for humans is much easier to recognize a face (if it is in normal position) than any other object.

4.2. Study description

In the current study, we analysis the cognitive development through incremental learning applied to the problem of face recognition for a social robot. The incremental learning is quite a new topic in the area of pattern recognition and it addresses the case when, for different reasons, not all the classes and not all the data from the existing classes are known from the beginning, but they become available over time. From a cognitive point of view, the short-term memory represents the information available over a very brief period of time[3], meanwhile, the long-term memory represents the database built from past experiences. When a new face image becomes available, it is added to the existing representation. This adaptive process will have a 'renewing' effect, in the sense that the most recent instances of a face will have greater weight in data representation, compared with previous ones. It will come a moment when the features corresponding to the earliest instances have been completely replaced.

Our study has been performed using an AIBO robot. AIBO is a biologically-inspired robot and is the flagship of its generation (social robotics). It comes with pre-built behaviors like for instance how to react to a variety of stimuli (visual, voice and tactile). At the same time, it is able to express a wide range of emotions and desires. A very important characteristic is that it possess a 'built-in' sense of curiosity. In consequence, it is in a continuous process of learning and discovering new things about itself and the environment where it 'lives'. For all these qualities, we consider it as the ideal candidate for our objectives.

The final goal of the current study is to build an application whose aim is to have AIBO behaving in a personalized manner, depending on the frequency it sees a certain person. Thus, we expect it to develop a 'friendlier' attitude towards a frequent seen person, meanwhile to act more 'reserved' in front of a person who got to 'know' from few instances.

[3] In our case, we will adopt the incremental learning in a sequential manner, so the short-term memory will consists of only one face instance.

4.3. Learning algorithm

To carry out the study described in the Section 4.2, we implemented the incremental LDA (linear discriminant analysis) algorithm (referred as IncLDA) from (Pang et al., 2005). In the next sections, we will briefly review, first, the classical LDA algorithm (referred as BatchLDA and its modification in order to allow a sequential updating of data representation.

4.3.1. Classical linear discriminant analysis (BatchLDA)

Let's assume that we have N data samples that belong to M classes $C_i, i = 1, 2, \ldots, M$. Each class C_i is formed by n_i samples $C_i = \{x_1^i, x_2^i, \ldots, x_{n_{C_i}}^i\}$. By $\bar{x}^{C_i}$ we will refer to the mean vector of class C_i and by $\bar{x}$ we will refer at the global mean vector. As introduced in Fukunaga (1990), the within-class scatter matrix S_w and between-class scatter matrix S_b are used as a measure of inter-class separability. They are defined as follows:

$$S_w = \sum_{i=1}^{C_M} \sum_{j \in C_i} (x_j - \bar{x}^{C_i})(x_j - \bar{x}^{C_i})^T \quad (1)$$

$$S_b = \sum_{i=1}^{C_M} n_{C_i} (\bar{x}^{C_i} - \bar{x})(\bar{x}^{C_i} - \bar{x})^T \quad (2)$$

In order to extract the optimal LDA features for knowledge representation several criteria can be used. One of the most common is the one that maximize the ratio of the between-class scatter matrix to that of the within-class scatter matrix, i.e.

$$\hat{W} = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} \quad (3)$$

This problem has an analytical solution and is mathematically equivalent to the eigenvectors of the matrix $S_w^{-1} S_b$.

4.3.2. Incremental linear discriminant analysis (IncLDA)

The shortcoming of the BatchLDA described in the previous section comes from the assumption that all the data are available at the classification. This is not the case for real applications, when the data is coming over time, at random time intervals, and the representation of the data must be updated. Computing from the beginning the scatter matrices, each time a new sample arrives, is not computationally feasible, especially when the number of classes is very high and the number of samples per class increases significantly. For this reason, we propose the IncLDA technique, that can process sequentially later-on added samples, without the need for recalculating entirely the scatter matrices. In order to describe the proposed algorithm, we assume that we have computed the S_w and S_b scatter matrix from at least 2 classes. Let's now consider that a new training pattern y is presented to the algorithm. The global mean is updated according to the following equation:

$$\bar{x}' = \frac{N\bar{x} + y}{N + 1} \tag{4}$$

From now on, in order to recursively update the two scatter matrices, we distinguish between two situations:

- The new pattern y belongs to one of the existing classes C_L, where $1 < L < M$).

In this case, the equation that updates S_b is given by

$$S'_b = \sum_{i=1}^{C_M} n'_{C_i}(\bar{x}'^{C_i} - \bar{x}')(\bar{x}'^{C_i} - \bar{x}')^{\mathrm{T}} \tag{5}$$

where $\bar{x}'^{C_i} = \frac{n_{C_i}\bar{x}^{C_i}+y}{n_{C_i}+1}$ and $n'_{C_i} = n_{C_i} + 1$, if $i = L$; otherwise, $\bar{x}'^{C_i} = \bar{x}^{C_i}$ and $n'_{C_i} = n_{C_i}$.

In the case of S'_w the update equation is the following:

$$S'_w = \sum_{j=1, j\neq L}^{C_N} S_w(C_j) + S_w(C_{L'}) \tag{6}$$

where

$$S_w(C_{L'}) = S_w(C_L) + \frac{n_{C_L}}{n_{C_L} + 1}(y - \bar{x}^{C_L})(y - \bar{x}^{C_L})^{\mathrm{T}} \tag{7}$$

and

$$S_w(C_i) = \sum_{j \in C_i}(x_j - \bar{x}^{C_i})(x_j - \bar{x}^{C_i})^{\mathrm{T}} \tag{8}$$

By $C_{L'}$ we refer to the class C_L after pattern y has been presented, i.e. $C_{L'} = C_L \cup \{y\}$

- The new pattern y belongs to a new class C_{M+1}

In this case, the equation that updates S_b is given by

$$\begin{aligned} S'_b &= \sum_{i=1}^{C_M} n_{C_i}(\bar{x}^{C_i} - \bar{x}')(\bar{x}^{C_i} - \bar{x}')^{\mathrm{T}} + (y - \bar{x}')(y - \bar{x}')^{\mathrm{T}} \\ &= \sum_{i=1}^{C_{M+1}} n_{C_i}(\bar{x}^{C_i} - \bar{x}')(\bar{x}^{C_i} - \bar{x}')^{\mathrm{T}} \end{aligned} \tag{9}$$

Regarding, the new S'_w matrix, this one remains unchanged, i.e

$$S'_w = S_w \tag{10}$$

4.4. Experimental results

For experiments, we used a custom face database built using the AIBO's camera. The image acquisition phase was extended over several weeks and performed in an automatic manner. For this purpose, we put the robot in an open space and snapshots were taken each time a person was passing in front of the camera. In Fig. 2 we extracted some frames from the face acquisition process.

The face was automatically extracted from the image using the face detector based on (Viola and Jones, 2004). We did not impose any restrictions regarding ambient conditions. Overall, our database consists of 6882 images of 51 people (both male and female).[4] Since no arrangements were previously made, some classes contain only a handful of images (as much as 20), meanwhile, the largest of them contains over 400. Face size is of 48×48 pixels. Because of the particularity of the acquisition process, face images reflect the changes in appearance suffered by subjects over time. Furthermore, since our application was thought to run in real-time (and to add it a more ad-hoc flavor), we did not perform any pre-processing step to face images before to pass them to the classifier. That's why the faces used in the experiment show a certain degree of variation in pose and size and are not constrained to be exactly frontal. That's why we allow the face images used to be a little wider than the face region itself. Some samples of these face images are presented in Fig. 3.

To test the IncLDA algorithm, we used 90% of the images (i.e. about 6000) as training set and the remaining ones as test set. From the training set, we used 15% of the images (belonging to 5 classes and representing 900 samples) to build the initial LDA eigenspace. In order to overcome the singularity problem, a PCA step was performed beforehand. This way, data dimensionality was downsized from 2304 to 60. The remaining samples (5100) from the training set were added later on in a sequential manner (the samples were drawn randomly) and this way the LDA eigenspace was updated. To test the efficiency of the IncLDA for classification, each data sample from the training set was encoded by projecting it to the updated LDA eigenspace. When a test sample is presented, its projection to the LDA eigenspace is computed and the classification is decided based on its 'nearest neighbor'.

In order to reflect the evolution of the learning process, we introduce the term of 'learning stage' to refer to the number of samples from the training set that have been added up to a certain moment. In Fig. 4 (left) we depicted the evolution of the learning process after each update (a new sample added) of the initial LDA eigenspace. In the early stages, there are a lot of new classes presented at very short intervals. It can be appreciated that, with almost 50% of the remaining training samples introduced, all classes have been represented. In Fig. 4 (right), we depicted the percentage of incremental training samples introduced so

[4] In the current study we put the accent in having a reasonable number of classes with a lot of instances rather having an excessive number of classes with very few instances.

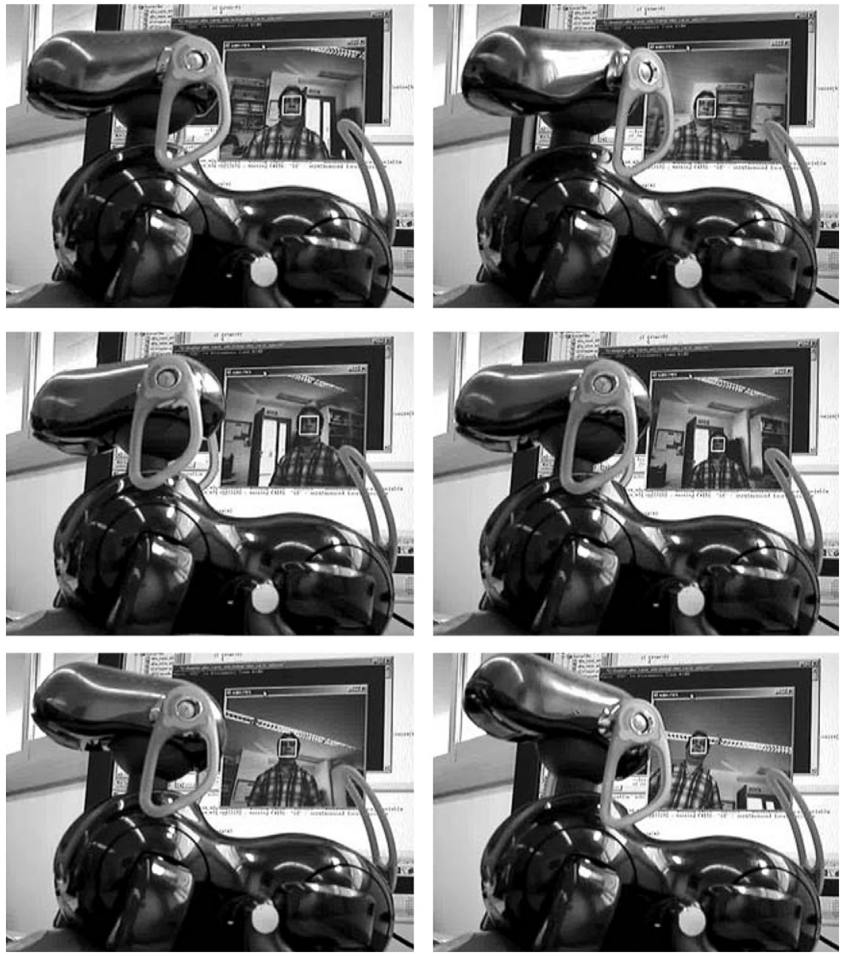

Fig. 2. Real-time face detection and tracking by an AIBO robot.

Fig. 3. Samples of face images from CVC custom database showing a certain degree of variation in illumination, pose and size.

far (the stars represent the moment when a new class has been added). This graphic should be read in concordance with the left one.

As a final proof of accuracy, we compared IncLDA with the classical LDA (referred as BatchLDA). In Fig. 5, we show that indeed the IncLDA is converging (at the end of the learning process) towards BatchLDA. The common recognition rate achieved is around 95%, which in our opinion is a very good result, taking into account the difficulty of the database. Both graphics were plotted after averaging the results obtained from a ten-fold cross-validation procedure (the training samples were chosen in a random manner in each run). The oscillation of the IncLDA in its early stages corresponds to the situation when a significant number of new classes have been added at very short intervals and only a very few samples of those classes were available. After some learning stages, when enough samples for each class became available, we can appreciate that the evolution curve regulates its tendency and becomes constantly ascending.

Some instances of misclassified faces are represented in Fig. 6. From the experiments performed, we arrive at the following conclusion. The misclassification occurs in three

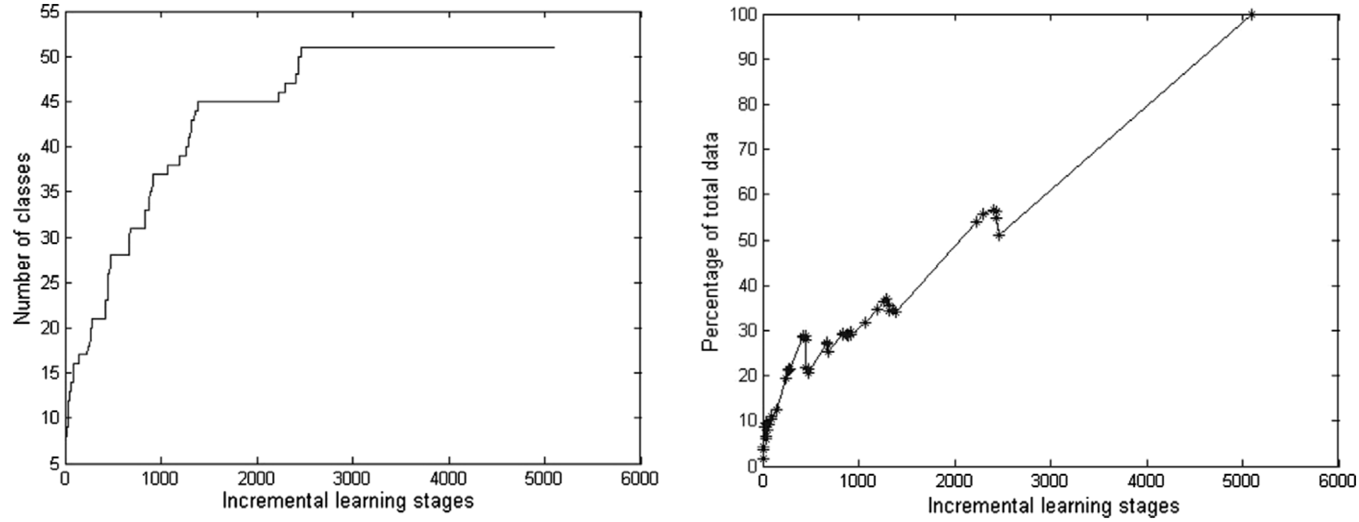

Fig. 4. Learning process: evolution of the number of classes function of learning stages (left) and the percentage of the training data function of learning stages (right).

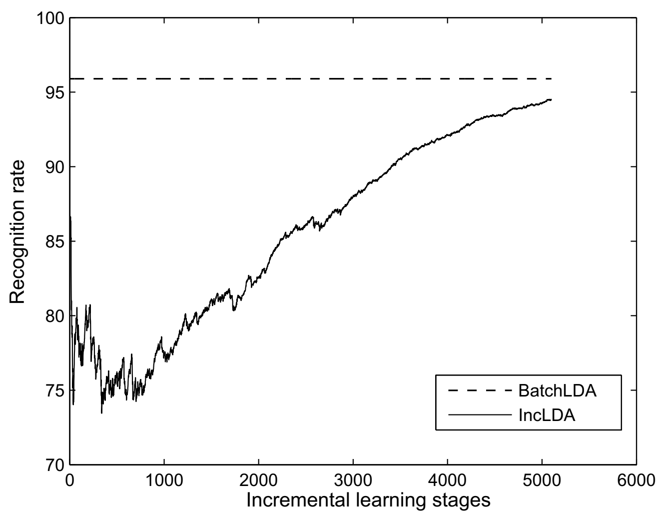

Fig. 5. IncLDA vs. BatchLDA curves. IncLDA converges towards BatchLDA at the end of the learning process.

Fig. 6. Some instances of misclassified faces.

situations: when there are too few face instances per class, when there are too few instances of a particular head pose/illumination conditions and when the image presents a high level of distortion (the 'blurring' effect due to person/camera movement).

5. Conclusions and future work

In this paper we presented our vision why the systems nowadays are smart, but not intelligent (in human terms). In our opinion, this failure is due to the fragmented study of AI and to the fact that the cognitive factors which are responsible for generating intelligent behavior have not received full consideration. After that, we made a brief review of the major learning strategies and the role that context can play in the learning process. Its use can help us disambiguate between uncertain situations or when we deal with partial or corrupted data. Finally, we presented a study for cognitive development in a social robot through an incremental learning algorithm applied to the problem of face recognition. In the future, we plan to endow our cognitive model also with 'forgetting' ability, so that if a person did not appear for a long time, his/her category will be eliminated from the knowledge database.

Acknowledgements

This work is supported by MCYT Grant TIN2006-15308-C02, Ministerio de Ciencia y Tecnología, Spain. Bogdan Raducanu is supported by the Ramon y Cajal research program, Ministerio de Educación y Ciencia, Spain.

References

AIBO robot. http://www.sony.net/Products/aibo/index.html.

Atkinson, R.C., Shiffrin, R.M., 1968. Human memory: A proposed system and its control processes. In: Spence, K.W. et al. (Eds.), The Psychology of Learning and Motivation, vol. 8. Academic Press, London.

Biederman, I., Mezzanotte, R.J., Rabinowitz, J.C., 1982. Scene perception: Detecting and judging objects undergoing relational violations. Cognit. Psychol. 14, 143–177.

Brézillon, P., 1999. Context in problem solving: A survey. Knowledge Eng. Rev. 14, 1–34.

Brooks, R.A., 1991. Intelligence without reason. In: Proc. Internat. Joint Conf. on Artificial Intelligence (IJCAI), Sydney, Australia, pp. 569–595.

Brooks, R.A, Stein, L.A., 1993. Building brains for bodies, AI Memo No. 1439, MIT.

de Gelder, B., Rouw, R., 2001. Beyond localisation: A dynamical dual route account of face recognition. Acta Psychol. 107, 183207.

Fodor, J.A., 1983. The Modularity of Mind. MIT Press, Cambridge, Massachusetts.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition, second ed. Academic Press, Boston, USA.

Goren, C.C., Sarty, M., Wu, R.W.K., 1975. Visual following and pattern discrimination of face-like stimuly by newborn infants. Pediatrics 54, 544–549.

Johnson, M., Dziurawiec, S., Ellis, H., Morton, J., 1991. Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition 40, 1–19.

Newell, A., 1973. You Cannot Play 20 Questions With Nature and Win. Visual Information Processing. In: Chase, W.G. (Ed.). Academic Press, New York.

Newell, A., Simon, H.A., 1961. Gps, A Program That Simulates Thought. Lernende Automaten. In: Billing, H. (Ed.). R. Oldenbourg, Munich, Germany, p. 109124.

Omrod, J.E., 1999. Human Learning, third ed. Merrill Prentice Hall, Upper Sadle River, NJ.

Palmer, S., 1999. Vision Science. MIT Press.

Pang, S., Ozawa, S., Kasabov, N., 2005. Chunk Incremental LDA Computing on Data Streams. In: Wang, J., Liao, X., Yi, Z. (Eds.), International Symposium on Neural Networks, LNCS 3497, Chongqing, China, pp. 51–56.

Shortliffe, E.H., 1976. MYCIN: Computer-based Medical Consultations. Elsevier, New York.

Torralba, A., Sinha, P., 2001. Statistical Context Priming for Object Detection. In: Proc. Internat. Conf. on Computer Vision, Vancouver, Canada, pp. 763–770.

Torralba, A., Murphy, K.P., Freeman, W.T., Rubin, M.A., 2003. Context-based Vision System for Place and Object Recognition. In: Proc. Internat. Conf. on Computer Vision Nice, France, pp. 273–280.

Viola, P., Jones, M.J., 2004. Robust real-time face detection. Int. J. Comput. Vis. 57, 137–154.

Wada, K., Shibata, T., Saito, T., Sakamoto, K., Tanie, K. 2005. Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In: Proc. Internat. Conf. on Robotics and Automation (ICRA), Barcelona, Spain, pp. 2796–2801.

Weiser, M., 1991. The computer for the twenty-first century. Scientific American 265, 94–104.

Winograd, T., 2001. Architectures for context. Hum. Comput. Interact. 16, 401–419.

Yokoyama, A., 2002. The Possibility of the psychiatric treatment with a robot as an intervention – from the viewpoint of animal therapy. In: Proc. 1st Internat. Joint Conf. on Soft. Comput. and Intellig. Syst and 3rd Internat. Symp. on Adv. Intellig. Syst. Paper number 23Q1–1.