REGULAR PAPER

A combined approach for improving humanoid robots autonomous cognitive capabilities

Kurosh Madani[1] · Antonio M. Rinaldi[2] · Cristiano Russo[2] · Cristian Tommasino[2]

Abstract

Recent technologies advancements promise to change our lives dramatically in the near future. A new different living society is progressively emerging, witnessed from the conception of novel digital ecosystems, where humans are expected to share their own spaces and habits with machines. Humanoid robots are more and more being developed and provided with enriched functionalities; however, they are still lacking in many ways. One important goal in this sense is to enrich their cognitive capabilities, to make them more "intelligent" in order to better support humans in both daily and special activities. The goal of this research is to set a step in bridging the gap between symbolic AI and connectionist approaches in the context of knowledge acquisition and conceptualization. Hence, we present a combined approach based on semantics and machine learning techniques for improving robots cognitive capabilities. This is part of a wider framework that covers several aspects of knowledge management, from representation and conceptualization, to acquisition, sharing and interaction with humans. Our focus in this work is in particular on the development and implementation of techniques for knowledge acquisition. Such techniques are discussed and validated through experiments, carried out on a real robotic platform, showing the effectiveness of our approach. The results obtained confirmed that the combination of the approaches gives superior performance with respect to when they are considered individually.

C.Russo, K. Madani, A. M. Rinaldi, C.Tommasino have contributed equally to this work.

✉ Antonio M. Rinaldi
  amrinald@unina.it

  Kurosh Madani
  madani@u-pec.fr

  Cristiano Russo
  cristiano.russo@unina.it

  Cristian Tommasino
  cristian.tommasino@unina.it

[1] LISSI Laboratory, University of Paris-Est Creteil (UPEC), 36-37 rue Georges Charpak, 94000 Paris, France

[2] Department of Electrical Engineering and Information Technology, University of Napoli Federico II, Via Claudio, 21, 80125 Napoli, Italy

1 Introduction

Recently, autonomous robots have aroused deep interest in the scientific community for their groundbreaking enormous potential. Today, machines are amazingly sophisticated and cleverly engineered, however they still lack the high-level abilities required to advocate for and interact with humans in many tasks. Moreover, the "intelligence" exhibited rarely comes out from real autonomous decisions, that is the machine is not conscious of its own knowledge. Although the cognitive capabilities of existing robots are quite distant from those of humans, robots are expected to share their living space with humans (and vice versa) in this century. Moreover, they will no longer be run by expert technicians. Such a future perspective certainly promises to have a tremendous impact on the way we understand society compared to how we see it today. Besides implications related to cultural changes, we have to consider that this epochal transition is a very complex task to accomplish, requiring joint efforts from multiple scientific communities operating in very different fields. Robots need to be autonomous enough to accomplish tasks cooperatively with human users, who may not have a-priori technical skills. To meet this requirement, future work in robotics should aim at increasing the autonomy of robots. One of the best methods of contributing to this is to investigate in the research field of machine cognition and, in relation to this, on autonomous knowledge acquisition. This is also one of the aims of our proposed system, further discussed in the rest of the paper.

Our interest is on the relationship between autonomous robotics, knowledge management systems and artificial intelligence movements. In our vision, the dichotomy between symbolic AI and connectionist AI is of particular attraction. In fact, regardless of specific intents, a neat separation between these two "worlds" emerges. One of the objectives of this work is to establish a stronger link between them to better exploit the advantages of both in the tasks of knowledge acquisition and conceptualization. While often criticized in the past, the use of approaches based on symbolic artificial intelligence, could provide particular advantages to existing solutions in: i) the representation of knowledge by a robot to increase its autonomy in the conceptualization process, ii) providing prior knowledge (it could be done through Bayesian methods as well) to a robot contextualized to its operating environment, iii) human–robot and robot-robot interactions actively handled by machines and not guided by a human. In this view, the way we store and represent knowledge in an artificial system assumes significant importance. Features, such as flexibility and interpretability, become necessary to surmount these limitations and enhance human–machine and machine–machine interactions. Such features may be provided by means of artificial neural networks inspired by the human brain, i.e., by using networks designed to emulate some mental processes which are the biological roots of human behaviors. The main contributions of this paper refer to the design of individual modules for a general modular framework for knowledge management in autonomous robotics. In particular, we discuss here the refined development of the unsupervised low-level approach for constructing and conceptualizing the knowledge based on perception, whose improved results are compared with the ones from previous work, presented in [1] and briefly reported here; its combination with the supervised high-level approach for constructing and conceptualizing the knowledge based on semantic [2] into a unified approach through the use of merging operators and functions; the validation of the

models on a real autonomous robotic platform. The paper has the following structure: in Sect. 3, we introduce the framework for knowledge management along with a brief discussion of its components, while we provide a literature overview of knowledge acquisition in intelligent systems in Sect. 2. Section 4 discusses the approaches used for knowledge acquisition. In particular, for this work we focus on unsupervised and combined machine learning techniques; Sects. 5 and 6 provide implementation details for the proposed approaches, while Sect. 7 is left for conclusions and future works directions.

2 Related works

In the 80s and 90s, researchers laid the foundations to try to understand what mechanisms guided the human mind, resulting in intuition and intelligence. An interesting theory has been outlined by [3] whose objective is to merge the various areas of knowledge belonging to the many areas of research involved. From this we can see the need to consider knowledge from both biological and machine point of view, while it is at least hazardous to propose a universal definition of intelligence, since it can manifest in different ways depending on the context. Albus theory also delineates the minimum requirements intelligence should have: "intelligence requires the ability to sense the environment, to make decisions, and to control action. Higher levels of intelligence may include the ability to recognize objects and events, to represent knowledge in a world model, and to reason about and plan for the future. In advanced forms, intelligence provides the capacity to perceive and understand, to choose wisely, and to act successfully under a large variety of circumstances so as to survive, prosper, and reproduce in a complex and often hostile environment".

At the same time, Brooks [4–6] suggested a subsumption architecture for mobile robot control. It shows a hierarchy of layers in which higher level layers have more expertise and subsume the role of lower level layers. Following the considerations provided by Nonaka [7] on the role of interaction in knowledge creation, an approach to knowledge acquisition in artificial intelligent systems is the one investigated by the research field known as *Human–computer interaction* (HCI), and with more specificity to the field of robotics, the *human–robot interaction* (HRI).

The development of service robots with high cognitive skills, capable of friendly interactions with humans, is currently being investigated by a number of works: in [8, 9], the authors propose a method to create a spatial cognitive map. The goal is to establish friendly interactions between the robot and humans. Changes in human approaching behavior while having a conversation have been investigated in [10, 11] and validated on MiRob platform. Specific features of HRI, such as safety and dependability have been studied in [12]. Reinforcement learning is used in [13] for the development of a synergy-based framework for autonomous and intelligent robots. The use of robots as assistive technologies is one of the most investigated application domain in robotics. The neuro-robotics paradigm [14] aims to fuse competences and methods from neuroscience and robotics. Examples of intelligent assistive robots are [15, 16], while the employment of robots for people rehabilitation is witnessed in [17–19].

The field of human–robot collaboration is reviewed in detail in [20–22] Occupancy maps are used in [23] to record human's movement preference over time. This information is then used as input for an optimization-based motion planner to keep a safe distance between the human and the robot to improve the safety.

Another common approach to knowledge acquisition in intelligent systems relies on semantics-based techniques, such as production rules and ontologies. Ontologies have been widely used as a tool for knowledge management of robots and related cognitive architectures. Successful examples include the OpenRobots Ontology (ORO) [24, 25] and KnowRob [26, 27]. Semantic maps for task planning were used in [28]. A reasoner based on the geometry of the environment and the relations between humans and objects is presented in [29] to generate relevant symbolic information. Despite the innumerable benefits that the use of ontologies brings in the acquisition and sharing of knowledge, there are also serious disadvantages that have limited their spread over the years and therefore it is necessary to take them into account. One of the major concerns regarding ontologies is that they reflect a fundamentally monolithic view of knowledge [30], i.e., they are too rigid and cannot easily manage exceptions and shades of knowledge. In other words, the use of ontologies might be in contrast to the two major requirements of flexibility and interpretability used in the approaches of knowledge construction. The same author affirms that ontologies are also considered to be inadequate as a tool for knowledge representation, since the human language is used to manipulate concepts and to represent the world. Although such a claim can be considered reasonably true and acceptable, no alternative solution is proposed in this study. In contrast, ontologies are effectively used as a tool for knowledge integration in [31, 32]. Also, multimedia ontologies have been designed in [33, 34] to associate semantic concept with their visual representation. This approach is used to exploit both semantics and visual information in a combined manner.

3 Modular framework for knowledge management

This section introduces the theoretical scheme of the developed framework for cognitive systems briefly discussing its parts. The strategies advanced in this work are general enough to be considered valid for any machine or artificial intelligent system being it a device, like computers, smartphones, or a robotic platform. We restrict our application context to robots, as the development of the cognitive autonomy of such machines is an important step toward the next-generation of society. A logical view of the framework is provided in Fig. 1. It gives an overview of the aspects of knowledge management covered by organizing them in a unified framework, or logical architecture. The intended meaning is that of outlining its general structure.

Indeed, as multiple independent lines of research are involved, they are presented here as logical modules of a framework, in order to have a unified and more comprehensive view of areas that are explored and investigated as well as they interconnect each other. The picture aims to support the reader understanding the general workflow and to fix conceptual links among the different parts that cover the investigated research areas, which are depicted in the figure as squares with dashed borders. Before introducing all the framework components, it could be useful to discuss a concrete example which is helpful to understand how the different methodologies are integrated to make a machine (robot) able to acquire and conceptualize knowledge from the surrounding environment. First, we need to define the context within which to operate. One of the most interesting usages of humanoid robots for years to come is that of *personal robot*, i.e., a service robot whose main task is to assist elderly or sick people who are not able to move in autonomy or they need specific assistance in their homes. This conception of robot as an assistant to human beings is also definitely in line with the origin of the term *robot* itself. Therefore, a possible example would be to assist an
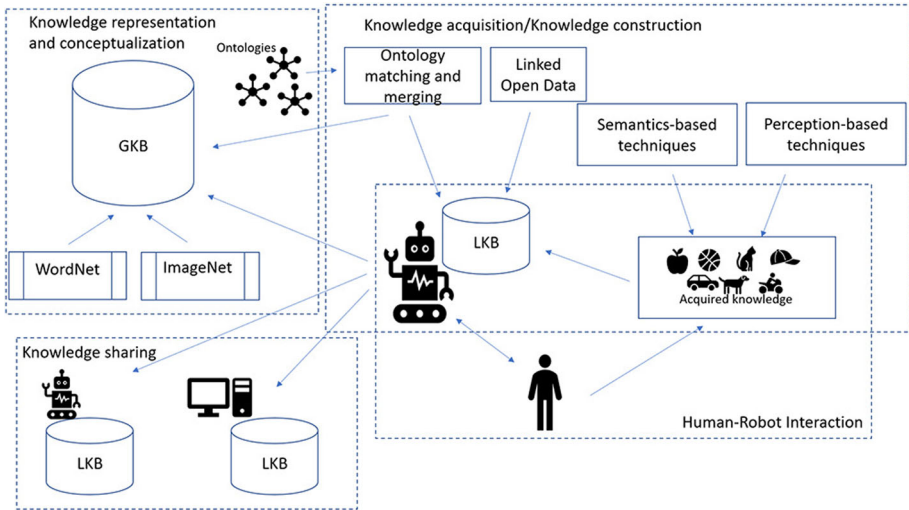
Fig. 1 High-level architecture of the framework. Dashed lines identify research areas, while arrows show the information flow and interconnections among the different fields

individual in his/her own home to carry out ordinary, everyday tasks, such as searching for objects, providing answers to general questions and so on. A common example is a robot searching for food. The first step is focused on knowledge design and representation in the framework. It consists in preparing the General Knowledge Base, described more in detail in Sect. 3.1, with general concepts and multimedia representations (in our case WordNet [35] and ImageNet [36]). According to the user specific needs, ontology matching and merging can be optionally performed to add more specific concepts related to the domain of interest. With regard to our example, this process would consist in looking for ontologies about food and anchoring them to the General Knowledge Base. The robot is initially provided with a basic amount of knowledge (few concepts) in its own Local Knowledge Base, since the goal is to test its own ability to acquire new knowledge. The subsequent phase is related to knowledge acquisition. Both semantics-based and perception-based techniques are used for this purpose. This phase can also involve the interaction with humans, who can trigger new requests and/or validate inferences made by the robot. Back to our example, the robot could be asked by the human to search for some food, let us say chocolate. First, the robot tries to pair the human utterance with a semantic concept and actions to be performed. Then, he looks for it in his Local Knowledge Base. In case the concept is absent, the robot retrieves it from the General Knowledge Base, he updates his own Local Knowledge Base and then searches for the item or most similar ones by using one of the above-mentioned techniques. Each element will be described in the following sections. We do not provide details about the component parts which are out of scope for this work. Our focus is on the unsupervised knowledge acquisition task which will be instead extensively described.

3.1 General knowledge base

One of the key-elements of the framework is the *General Knowledge Base*, alternatively named GKB. It plays a central role since the other modules are built upon it, or they make

use of it. A knowledge base provides a means for information to be collected, organized, shared, searched and utilized. In the context of this work, the knowledge base is called general since it has been implemented starting from a high-level ontological model proposed in [33] acting as overarching ontology, in turn defined following the structure of WordNet. This is the methodology used here for knowledge conceptualization. WordNet is a dictionary containing lexical-semantic connections between abstract concepts. The fundamental unit is a *synset*, i.e., a set of terms having the same meaning. Each synset represents a concept in WordNet. Synsets are interconnected each other by links that represent linguistic properties, in particular *lexical* and *semantic*; lexical relations hold between word forms while semantic relations hold between word meanings. Besides, visual features extracted from images contained in *ImageNet* have been imported in the General Knowledge Base. In fact, such a solution has advantages as well as drawbacks since its usage may not be ideal in certain situations and invalidate the development of applications requiring the acquisition of specific skills and not of general concepts. To address this issue and mitigate its effects, several countermeasures have been taken. First, the ontological model has been provided with the ability of being extensible through ontology matching and merging activities. This way, it is possible to integrate new knowledge coming from other ontologies. Second, each artificial system is provided with its own local knowledge base (LKB), which keeps the reference with the general knowledge base, but it is free to evolve on its own by adding new unknown concepts. Other techniques employable for the extension of knowledge acquisition process are based on the Semantic Web and Linked Open Data (LOD) principles. The use of LKBs brings important benefits. First, it allows to keep intact the structure of the general knowledge base, avoiding the inclusion of possible errors deriving from the observation of a real environment during the process of acquisition of new knowledge. Secondly, having a single centralized knowledge base would adversely affect its performance, when querying and retrieving information when the size of knowledge increases significantly. Third, with a view to sharing of knowledge, assuming that each entity provided with LKB acquires new skills, it will be possible to equip all the actors involved in the process with knowledge acquired individually. Moreover, since the LKBs always start their evolving process by keeping a reference with the GKB, the consistency of the whole system is guaranteed. This means that the knowledge that can be acquired by an agent/machine is framed to the knowledge contained in the GKB, so it cannot diverge possibly generating inconsistencies. In other words, it is very important to remark that the scope of knowledge acquisition is somehow limited to the knowledge contained in the GKB, i.e., when we talk about new knowledge we mostly refer to the acquisition from reality of physical entities as instances of concepts already included in the GKB. However, exceptions to violate such a constraint have been experimented for including knowledge from outside the GKB, i.e., knowledge acquired from linked open data.

4 Approaches to knowledge construction

This section focuses on the approaches defined for constructing and acquiring the knowledge based on how it has been conceptualized. By introducing different machine learning and human–computer interaction techniques, the goal is to enable an artificial intelligent system to autonomously or semi-autonomously learn and properly manage the knowledge contained in the general knowledge base. The process of ingesting further knowledge in an autonomous or quasi-autonomous way is challenging and may be achieved through different techniques and approaches. In our vision, the class of bio-inspired approaches has proved [37–40] to be

a good strategy in learning cognitive processes. Indeed, they are required to have satisfactory and friendly interactions as well as knowledge acquisition skills, either with humans or artificial systems. The rest of the section is arranged according to a classification of the approaches based on the nature of the techniques used that inspired them. In detail, we focus on approaches based on robots autonomous acquisition of knowledge. However, the foundation of our approach is still valid with application domains different than robotics.

The main goal is to allow a robot to derive, envision and conceptualize the knowledge with a dual process, combining perceptual and semantic knowledge. The perceptual knowledge originates from the enclosing environment. It commonly includes all the information retrieved from sensors, e.g., microphones for capturing audio signals such as human conversations, movements, temperatures and vision for extracting images. Visual sensors are employed to acquire images. The data collection process can be extended to gather information from other sensors, such as a microphone for audio and infrared for distances. The semantic knowledge is represented in our knowledge base. We validated the studied concepts on a humanoid robotic platform through real use cases. The dual-process above-cited looks in two opposed directions: the aim is to grasp the benefits of the two groups. The first one is more abstract, and we name it *top-down* (TD), it contributes with semantics, and it is closer to human comprehension. The Semantic Web technologies, ontologies, linked open data, production rules systems fall into this category. Further details about the TD approach are provided in [2].

The second one, discussed in this paper, is named *bottom-up* (BU). It is closer to machine-understanding, and it is related to perception of the environment, e.g., raw information extracted from sensors.

4.1 Bottom-up approach

The bottom-up approach aims to retrieve knowledge from the surrounding perceived elements. We say it is "guided by perception" in opposition with the top-down approach that is semantics-driven. The BU approach seeks knowledge through a low-level of abstraction.

It does not have any a-priori knowledge for the visual information hence, unsupervised learning represents a good strategy to pursue to find unknown patterns.

4.2 Self-organizing maps

Unsupervised machine learning techniques are particularly interesting, and they are widely used in a broad range of applications.

An interesting unsupervised neural network is the *self-organizing map* (SOM) [41]. Self-organizing maps are also called *Kohonen maps*, from the scientist who described their architectures in [41]. The basic idea follows the results of many neurobiological studies [42, 43].

SOMs follow a sort of principle of locality of information, which is also referred to as *the principle of topographic map formation* [44]: "The spatial location of an output neuron in a topographic map corresponds to a particular domain or feature drawn from the input space". The aim is to preserve the neighborhood of neurons which treat similar information during the processing.

The network is trained through competitive learning, where neurons are progressively tuned based on input stimuli. The formula used by SOM for updating the vector of weights

$W_v(i)$ for a neuron v is:

$$W_v(i + 1) = W_v(i) + \theta(u, v, i) \cdot \alpha(i) \cdot (D(t) - W_v(i)) \tag{1}$$

index i represents the iteration, t is the training sample, u is the index of the best matching unit, $\alpha(i)$ is a monotonically decreasing learning coefficient; $\theta(u, v, i)$ is the neighborhood function computing the distance between u and v at step i. The neighborhood function narrows over time and can be of different forms. In the context of this research, self-organizing maps have been employed for detecting semantic clusters starting from visual observation. A transfer-learning mechanism has been used for features employed for such an approach. An in-depth analysis of the whole method and how it has been implemented is provided in Sect. 5.

4.3 Growing hierarchical self-organizing maps

Classical self-organizing maps are a good strategy to pursue and they show good experimental results as it will be explained in Sect. 5. However, they have some drawbacks that limit their applicability. In fact, among the desirable features for an intelligent system, one of the most important, if not the most important, is the capability of evolving and enrich its own knowledge during time. Given their unsupervised nature, self-organizing maps provide in some ways such characteristic, but with strong limitations due to the intrinsic constraints related to the fixed size of the map and the flatness of the structure. As an example, the size parameter is often determined by experiments and it is based on the available data. Therefore, having a fixed size of the map forces the definition of clusters which could become at least not appropriate, or even meaningless, when new knowledge is added. Similarly, the flatness of the map does not allow to organize data hierarchically. This represents a strong limitation from a knowledge acquisition point of view, given that semantic hierarchies are a keystone of some cognitive processes as widely documented in Brooks' studies on subsumption architectures [4–6].

In order to overcome such limitations, a second unsupervised approach has been investigated. It makes use of *growing hierarchical self-organizing maps* (GHSOM) which have been first presented in [45, 46]. The GHSOM artificial neural-network model tackle the aforementioned issues with a hierarchical architecture of growing self-organizing maps. Moreover, contrary to similar solutions, such as *Growing grid* [47], the GHSOM is able to provide a global orientation to the independent maps facilitating the navigation across them.

An example of architecture for the GHSOM is shown in Fig. 2.

Stopping criteria are based on quantization error metrics (absolute error or mean error). In particular, two parameters named τ_1 and τ_2 are used during the training process to handle the growth of individual maps and the depth of the architecture.

The first parameter, τ_1, is used in the stopping criterion for the growth of a single map in the architecture. The formula for such criterion, assuming to use the absolute quantization error metric, is given in Eq. 2; q_{ej} is the quantization error of the father unit, that is the neuron of the map in the upper layer from where the map i has been generated.

$$MQE_{map_i} < \tau_1 \cdot q_{ej} \tag{2}$$

As an example, if $\tau_1 = 0.1$ it means that the mean quantization error of the map must be less than one-tenth of the quantization error of father neuron in the upper layer.

The second parameter, τ_2, is used to specify the minimum quality of data representation of each unit as a fraction of the quantization error of the zero unit, that is the initial layer of the map where all the data are mapped into a single neuron. It directly influences the
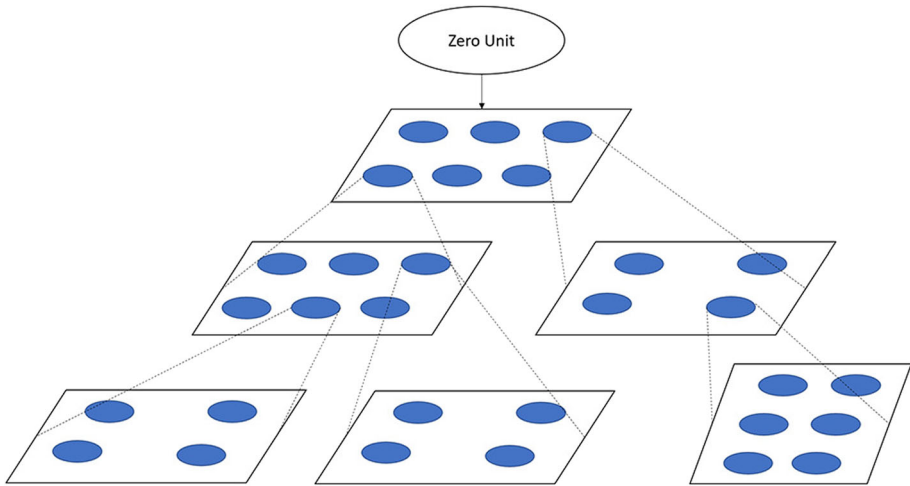
Fig. 2 GHSOM architecture

vertical growth of the GHSOM. The formula for such criterion, assuming to use the absolute quantization metric, is given in Eq. 3;

$$qe_i < \tau_2 \cdot q_{e0} \tag{3}$$

Same considerations made about self-organizing maps applies here, since the approach is basically the same but enhanced with more capabilities. Details are provided in Sect. 5.

4.4 Combined approach

All the methods explored until now present different facets, but they eventually belong to one of the two sides of the same coin. Given this premise, the further, most natural step is to attempt to unify the results obtained from the aforementioned approaches in order to evaluate if a novel combined approach gives optimal results exploiting the advantages of both. The idea is to have a more robust final classification by a voting-based decision which takes into account both supervised and unsupervised approaches results. This is a well-known practice in machine learning and artificial intelligence field, which proved to be successful over the years. A logical scheme for such a combination is given in Fig. 3.

To fuse the approaches together, it is necessary to define which operator to use, i.e., a mathematical function taking in input the output of individual algorithms and returning a novel output given from their combination. Both TD and BU approaches are already designed to output a list of concepts along with a probability, hence there is no need for a normalization step. The probabilities also take into account the hierarchies of classifications. For the combination function to take place, *Ordered-Weighted averaging* (OWA) operators have been used [48].

Formally, an OWA operator of order n is a function $F : R_n \rightarrow R$ having a collection of weights $W = [w_1, ..., w_n]$, whose elements fall into the unit range such that: $\sum_{i=1}^{n} w_i = 1$. The function is defined as in Eq. 4.
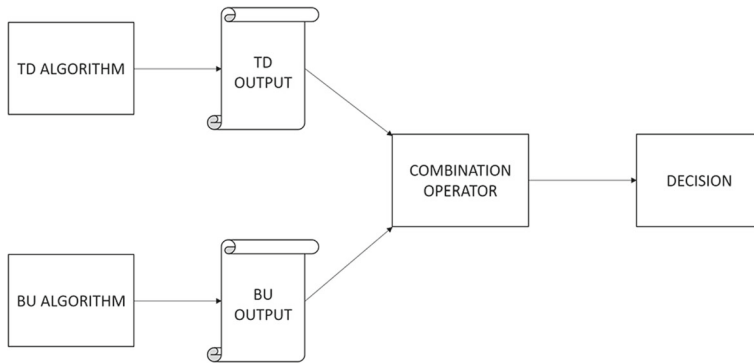
Fig. 3 Logical architecture of combined approach

$$F(a_1, ..., a_n) = \sum_{j=1}^{n} w_j b_j \tag{4}$$

where b_j is the $j - th$ largest of the a_i.

In the context of this research, their use is motivated by the fact that, since they provide a parameterized family of aggregation operators, it is possible to have more reliable computations when dealing with decision analysis under uncertainty.

5 Bottom-up approach implementation

As discussed in Sect. 4.1, bottom-up approaches are based on unsupervised learning techniques largely used in machine learning and artificial intelligence research field. In particular, the approaches tested for this research make use of classical SOMs and the more advanced GHSOMs. The implementation used for both algorithms is based on two libraries, written in Python language. The first one is already packaged with the Python installation, while the second one is a third-party[1] implementation, which has been adapted to our needs. We discuss several experiments we have performed to evaluate the BU approach. Our analysis is focused in particular on the knowledge acquisition of a machine and its evolving ability; hence, we discuss both qualitative and quantitative measures given the operating context.

5.1 Data preparation

At first, we manually selected 34 concepts (listed in Table 1) related to different semantic domains from our knowledge base.

In particular, we defined 5 semantic classes. The first four are well defined, hence strongly coherent from a semantic point of view, i.e., they share a strong semantic correlation. The last one instead includes two semantically uncorrelated concepts. The number of images per synset in the knowledge base is limited due to size constraints. Therefore, we collected further images from the service *google images* by using an automated Python script. We also manually checked inconsistencies in the retrieved images. By using this strategy, we

---

[1] https://github.com/enry12/growing_hierarchical_som.

Table 1 Concepts grouped by semantic domain

Semantic domain	List of synsets
Animals	Cat, giraffe, zebra, sheep, horse bird, dog, elephant, cow, bear
Food	Pizza, sandwich, orange, broccoli, watermelon hotdog, banana, apple, pear, donut, cake
Vehicles	Bus, truck, boat, motorcycle, bicycle, car, airplane, train
Dishes	Cup, knife, spoon
Other	Potted plant, person

Table 2 Example of dataset statistics

Class	No. of images
Banana	43
Horse	39
Broccoli	50
Bear	38
Boat	39
Truck	38
Hot dog	47
Spoon	44
Train	50
Bird	45
Cow	49
Apple	19
Cake	49
Person	37
Pear	38
Bicycle	49
Donut	46
Car	40
Cup	47
Sheep	47
Potted plant	45

eventually collected 1477 samples. Details about samples per concept are provided in Tables 2 and 3.

Figure 4 shows examples of images collected for some selected synsets.

Each image was processed to extract global features (JCD, CEDD, etc.) and activation features from the last pool layer of selected CNNs to be used as input in our experiments. This is the same approach used for the Top-down approach.

5.2 Bottom-up approach evaluation

Similarly to what we have done for the classical SOMs in our previous research [1], experiments have been carried out also for the approach based on GHSOMs. The same sets of data and deep features have been used. Moreover, a genetic algorithm has been developed for the

**Table 3** Example of dataset statistics (continuation)

Class	No. of images
Motorcycle	49
Giraffe	47
Zebra	47
Orange	40
Pizza	48
Elephant	50
Bus	49
Cat	38
Knife	45
Airplane	37
Sandwich	47
Dog	37
Watermelon	45

Fig. 4 Example images for synsets airplane, bear, pizza, spoon

fine tuning of parameters τ_1 and τ_2 in the GHSOM architecture. Details about this algorithm are provided in Sect. 5.3.

Figures 5, 6 and 7 depict three maps (with winning neurons) belonging to the resulting GHSOM architecture, got by running an experiment with the dataset *34-synsets* and deep features from *VGG16* model. For what concerns the grouping of data in semantic clusters a slight difference has been introduced. Since the cluster of dishes is made of just three classes, it has been included in the cluster, named *Others*, which is used as a noise element. Hence, the meaning for colors is now as follows: red color is for animals, blue color for vehicles, yellow color for food, cyan color for others. Numbers on the right represents the labels for classes in the dataset. More in detail, Fig. 5 shows the parent map, that is the map located at the first level of the hierarchy. The map shows a strong coherence between semantic domains and neurons. In other words, different classes with a strong semantic correlation are placed in the same neuron or in adjacent neurons, allowing for a first discrimination power of the system at domain-level. Such an ability is still more remarked if we look at the lower levels of the hierarchy. For example, Fig. 6 refers to the map at the second level of the hierarchy

Fig. 5 Graphical representation of winning neurons in parent GSOM. First level in the GHSOM architecture
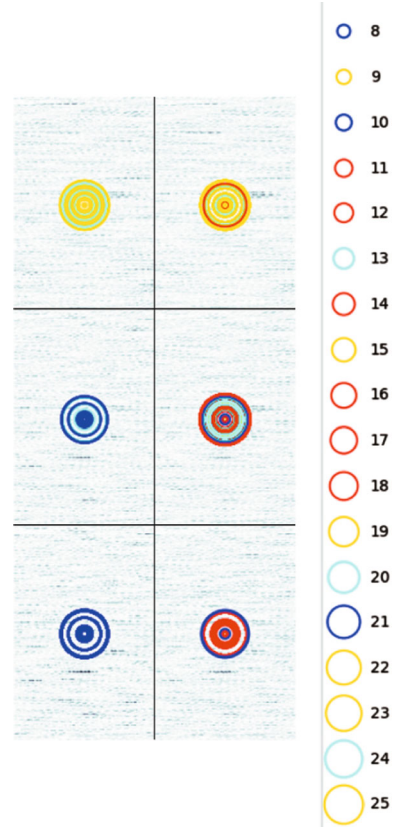
having its parent neuron in positions (0, 0). Most of the classes are depicted in yellow, which means they belong to the food domain. Similarly, Fig. 7 refers to the map at the second level of the hierarchy having its parent neuron in positions (1, 0). Most of the classes are depicted in blue, which means they belong to the vehicle domain.

It is worthy to note the presence of some instances belonging to the "disturbing" semantic domain, colored in cyan, for the illustrated maps in the second level. As it has been discussed, such a cluster has been intentionally included as a noise element in the analysis to have a more robust evaluation. Also, this behavior was expected, since classes belonging to this group do not really have a semantic correlation, since they are taken in a scattered manner. Moreover, with respect to the other semantic domains investigated, this one has a much lower instances number; hence, it is much smaller and less representative.

An experiment for computing centers of gravity for classes has been carried out for the GHSOM case. The technique is the same we used in [1] for the classical SOM architecture. There are, however, some differences that need to be highlighted and further clarified. Given that GHSOM has a hierarchical and dynamically growing structure, its final architecture is composed of many maps with different sizes at different levels of granularity, instead of just one with fixed size, as in SOM case. In particular, the whole GHSOM structure has a high number of maps (for the settings used in our experiments more than fifty), each exploring different portions of the dataset going down the hierarchy.

Fig. 6 Graphical representation of winning neurons in second level GSOM having its parent neuron in position (0,0)

The implication for this fact is that all the maps in the GHSOM, whose growth is controlled by parameters τ_1 and τ_2, have very different and smaller sizes compared to the SOM, where the size of the unique map is one of the parameters manually set for each experiment. To make an example, the GHSOM map in the first level (after the zero unit), which is the one receiving in input the whole dataset has a size after training of $3x2$ neurons, while for the SOM map analyzed with centers of gravity test the size is $30x30$ neurons (because they are all condensed in that unique map). The positions of centers of gravity for the map placed at the first level of the hierarchy, which is a map with $3x2$ neurons, are displayed on the grid in Fig. 8.

In Table 4, we report the detailed results obtained for centers of gravity, winning neurons and relative Euclidean distances related to the same map.

5.3 Optimizing parameters for controlling maps growth

An optimization algorithm has been implemented for a correct definition of proper values of the two parameters τ_1 and τ_2, responsible for the growth of the GHSOM architecture. The goal of this process is to shove the formation of visual clusters toward a semantics-based interpretation. The task has been accomplished by realizing a genetic algorithm used to fine-tune the parameters. Other optimization algorithms could be used as well; however, evolutionary
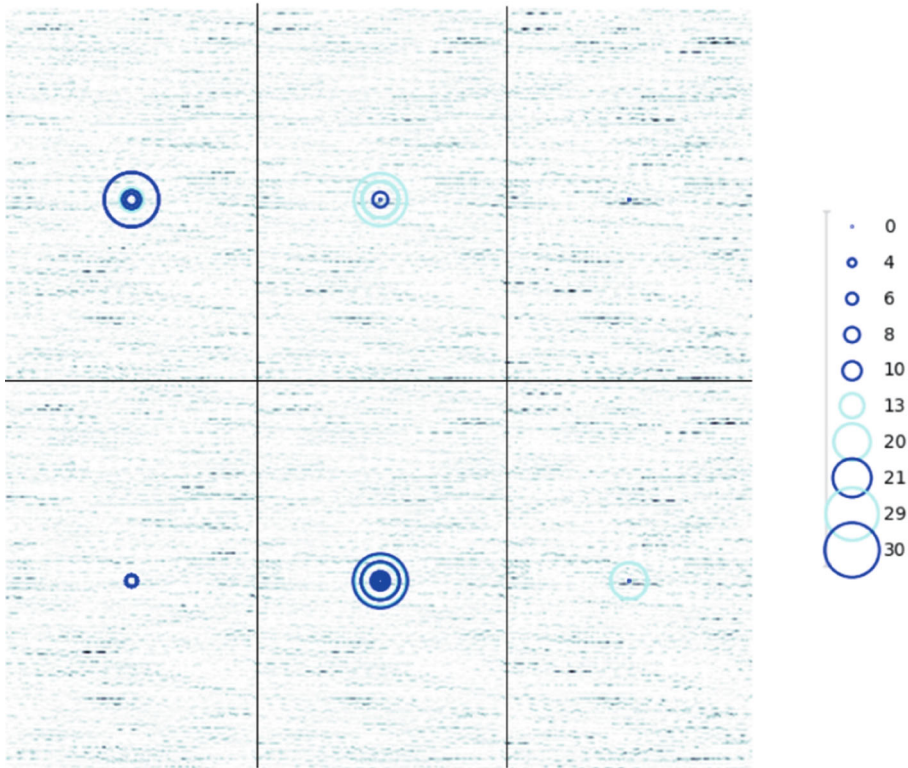
Fig. 7 Graphical representation of winning neurons in second level GSOM having its parent neuron in position (1,0)
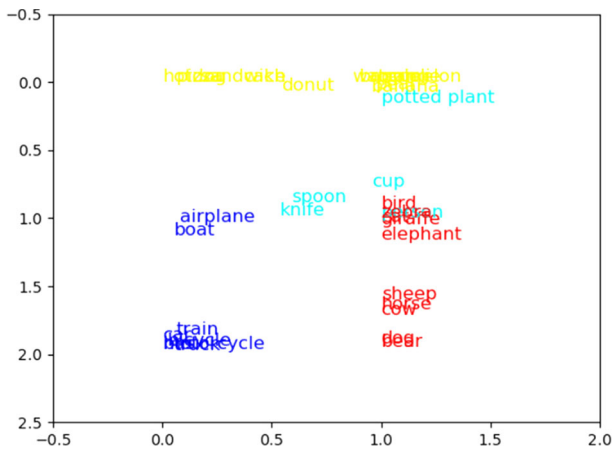

Fig. 8 Positions of centers of gravity with labels in the first-level map of GHSOM architecture

Table 4 Center of gravity positions, winning neuron position for center of gravity vector, Euclidean distances for first-level map in GHSOM architecture

Class	Center of gravity position	Winning neuron position	Eucl. distance
Banana	(0.07,0.95)	(0,1)	0.08
Horse	(1.66,1.0)	(2,1)	0.33
Broccoli	(0.0,0.9)	(0,1)	0.1
Bear	(1.95,1.0)	(2,1)	0.05
Boat	(1.13,0.05)	(1,0)	0.14
Truck	(1.97,0.05)	(2,0)	0.06
Hot dog	(0.0,0.0)	(0,0)	0.0
Spoon	(0.88,0.59)	(1,1)	0.42
Train	(1.86,0.06)	(2,0)	0.15
Bird	(0.93,1.0)	(1,1)	0.07
Cow	(1.71,1.0)	(2,1)	0.29
Apple	(0.0,0.95)	(0,1)	0.05
Cake	(0.0,0.37)	(0,0)	0.37
Person	(1.0,1.0)	(1,1)	0.0
Pear	(0.05,0.97)	(0,1)	0.06
Bicycle	(1.94,0.02)	(2,0)	0.06
Donut	(0.06,0.54)	(0,1)	0.46
Car	(1.89,0.0)	(2,0)	0.1
Cup	(0.77,0.96)	(1,1)	0.24
Sheep	(1.59,1.0)	(2,1)	0.4
Potted plant	(0.15,1.0)	(0,1)	0.15
Motorcycle	(1.96,0.0)	(2,0)	0.04
Giraffe	(1.04,1.0)	(1,1)	0.04
Zebra	(1.0,1.0)	(1,1)	0.0
Orange	(0.0,0.98)	(0,1)	0.02
Pizza	(0.0,0.06)	(0,0)	0.06
Elephant	(1.16,1.0)	(1,1)	0.16
Bus	(1.96,0.0)	(2,0)	0.04
Cat	(1.03,1.0)	(1,1)	0.03
Knife	(0.98,0.53)	(1,1)	0.47
Airplane	(1.03,0.08)	(1,0)	0.09
Sandwich	(0.0,0.17)	(0,0)	0.17
Dog	(1.92,1.0)	(2,1)	0.08
Watermelon	(0.0,0.87)	(0,1)	0.13

properties of genetic algorithms are the most suitable in this context. The sets of experiments on the dataset *34 synsets* have proven that a relation exists among visual features and semantics knowledge. Therefore, the idea is to exploit such information to minimize semantics diversity in visual clusters. To accomplish this, the details of the genetic algorithm are explained below. The parameters we want to optimize are the *genes*, while each combination of them is defined as a *chromosome*, that is a potential solution for the optimization problem.

FITNESS FUNCTION EVALUATION

Fig. 9 Evolution of the fitness function

A *population* instead is a set of solutions or chromosomes. The optimization problem has been set as a minimization problem, since our goal is to reduce semantics heterogeneity in clusters. Hence, a simple *fitness function* is used for the evolutionary process of the genetic algorithm. The fitness is defined as the ratio between the sum of numbers of different semantic labels encountered in each neuron and the total number of neurons. Such a function is given in Eq. 5 where K is the total number of neurons resulting from the growth process.

$$fitness = \frac{\sum_{i=1}^{K} labels_i}{K} \tag{5}$$

Figure 9 shows the evolution of the fitness function (lower values correspond to a better fitness). At each iteration, represented on the x axis of the chart, the algorithm evaluates the current population and chooses the best solution, which is kept, while other solutions are discarded and replaced by new chromosomes. The evolutionary process terminates according to the stopping criterion, which is triggered when the fitness does not improve for three consecutive iterations.

The best solution resulting after the algorithm execution has values 0.01 and 0.1 for parameters τ_1 and τ_2, respectively.

The fitness function provided in Eq. 5 just considers how many different classes are detected inside each neurons.

A more complex fitness function, shown in Eq. 6, has been realized to take into account the semantic information at our disposal.

$$fitness = \frac{\sum_{i=1}^{K} labels_i + \sum_{l=1,m=1}^{N}(1 - wup_{l,m})}{K} \tag{6}$$

Such a function adds to the previous fitness function another term in its numerator corresponding to the sum of all N^2 "semantic distances" between pair of samples (l, m) classified at each neurons. Each "semantic distance" is computed as the inverse of the *Wu-Palmer* [49] similarity ($wup_{l,m}$), which is a measure already normalized in the range [0, 1]. Hence, the inverse is calculated as $1 - Wu\text{-}Palmer$.
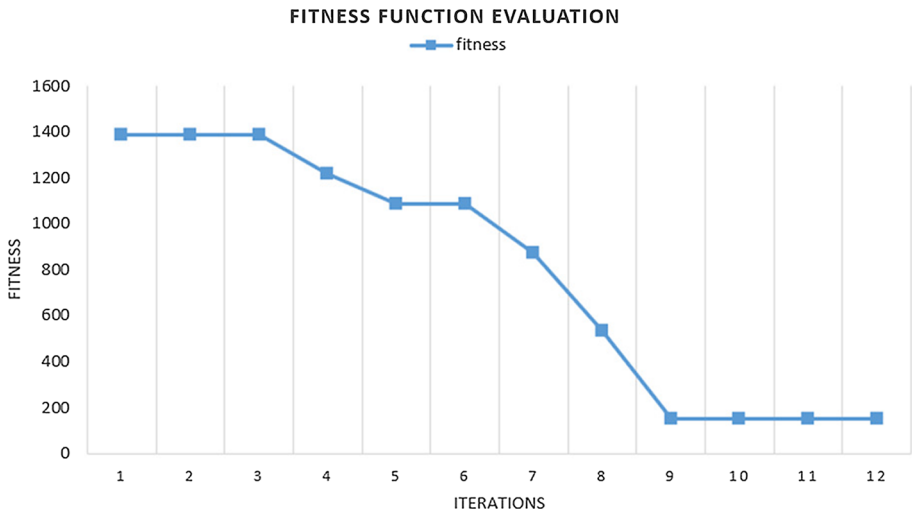
FITNESS FUNCTION EVALUATION

Fig. 10 Evolution of the fitness function with semantic similarity measure

Table 5 Combination schemes

Combination schemes		
Name	Weight TD	Weight BU
A	0,5	0,5
B	0,2	0,8
C	0,8	0,2
D	0,4	0,6
E	0,6	0,4
F	0,0	1,0
G	1,0	0,0

The experiment, reported in Fig. 10, shows an analogous behavior, in terms of best solutions, to the one obtained with the first fitness function (see Fig. 9). This means that, for the set of data used in the experiments, the first fitness function is as much effective as the second one. The second fitness function proposed is preferable since it is expected to behave better with larger data sets and above all, with more classes, given its grater discriminatory power in term of semantics.

6 Combined approach implementation and evaluation

As explained in Sect. 4.4, the goal of combination is to use a function or operator for merging the results of individual approaches together (see Fig. 3). The weights used in the experiments are reported in Table 5. Each combination scheme is identified by a letter. For example, the scheme named *A* gives equal importance to top-down and bottom-up approaches. Other schemes evaluate the performance for several different configurations of weights. Schemes *F* and *G* refers to individual performances of bottom-up and top-down approaches, respectively.

Table 6 Combination accuracy

Combination accuracy		
Name	Accuracy top-1 (%)	Accuracy top-5 (%)
A	97.2	99.6
B	89.19	99.6
C	86.49	99.6
D	97.2	99.6
E	94.6	99.6
F	83.7	99.6
G	86.49	86.49

The algorithm has been implemented in Python language and tested over a test-set extracted from the dataset, *34-synsets*. It receives in input two lists of predicted synsets with corresponding probabilities. One list is the output of Top Down approach, i.e., the combined chain of predictions from the "Semantic Multilevel classifier"; the second list is the result of Bottom Up approach, i.e., the chain of predictions from the "GHSOM" architecture. In particular, the algorithm considers the list of synsets from last level of predictions for both. The combination merges the two lists together and computes the final score for all the considered synsets, taking into account the weights given in Table 5. The final list is ranked according to the final scores; both top-one accuracy and top-five accuracy are computed.

Schemes *F* and *G*, show that individual algorithms have already good performances for top-1 accuracy. Scheme G, which corresponds to the usage of top-down approach only, does not show improvement in top-5 accuracy. In fact, the optimal results for top-5 accuracy are due to the usage of the bottom-up approach. This means that, also in cases of misclassifications, the real class is always present in the same cluster of predicted one. The results of other combination schemes show the effectiveness of this technique: schemes B and C, which use the two approaches in an 80–20 ratio, already show an improvement compared to the use of the approaches used individually. Still better performances are achieved when the weights of two approaches are more balanced; this is the case of schemes E, D and A. In particular, schemes D and A were able to reach a 97.2% top-1 accuracy, i.e., a delta improvement of 11–14% with respect to individual algorithms. This gives still more significance to the usage of combination since it is a sign that one approach is capable of correcting the other one when it fails. In conclusion, the experiments suggest that the best option is scheme D, which assigns weights of 0.6 and 0.4 to bottom-up approach and top-down approach, respectively. In fact, giving slightly more importance to bottom-up approach allows us to exploit goodness of both approaches and it could be helpful considering the advantage in top-5 accuracy of bottom-up approach, resulting into better final overall performance, which was our initial goal for the combination task. Same results, shaped as a chart, are reported in Fig. 11. Such considerations reflect the results obtained from the experiments we carried on the datasets at our disposal. Given their not very big size, further experimentation on wider datasets could help increase the confidence of these insights.

6.1 Validation on a real robotic platform

As previously stated, we also tested the proposed combined approach on a real robot. The platform used to test our work is the humanoid robot Pepper. It has 20 degrees of freedom.
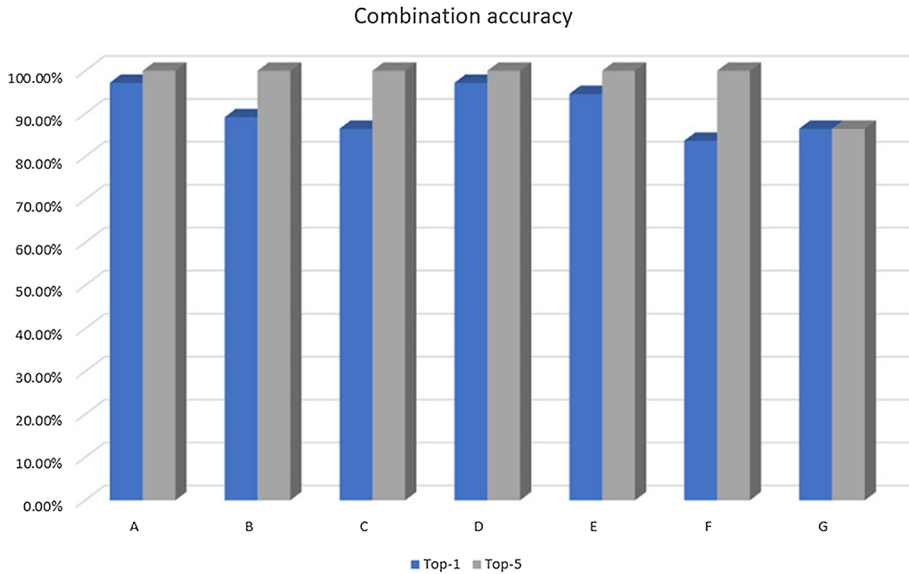
Combination accuracy

Fig. 11 Combination schemes accuracy

Fig. 12 Validation of combined approach on Pepper robot

It is designed to be fully interactive, as it can hold full conversations, and it is endowed with a number of sensors and modules, that allow him to speak, to detect obstacles, to take photos/videos, to move and so on. In the context of our research we have mainly used its 2D camera, and movement sensors. Figure 12 illustrates two frames of a demonstration which aims to show the final combined approach. In this scenario, we use multi-layer semantic classifier for the top-down approach, the GHSOM for the bottom-up approaches, combined according to the best values for the weights previously shown in combination experiments, that is 0.6 for the bottom-up approach and 0.4 for the top-down approach. The scenario is that of a simple interaction between the robot and a human that needs some help. The human asks for a pizza because he is hungry. In the environment, we have disseminated 3 objects (a plant, a spoon and an apple) along the path of Pepper. The robot starts the exploration of the environment and analyzes the objects he finds on his way. For each observation the robot uses the combined approach and then, since he could not find exactly what asked, he suggests the human for an apple which is the concept found "semantically" closest to the one asked.

The use of a robotic platform can give rise to multiple problems. Here, we discuss possible disadvantages related to the hardware with which the platform was equipped. Indeed, current humanoid robots come with very limited hardware compared to the software requirements of sophisticated machine learning algorithms. In addition, the sensors on board mainstream platforms are often inaccurate and/or not of high quality. On the Pepper robot we tried to overcome some of these limitations by communicating with the robot through its programming interface in Python and by replacing the Speech Recognition engine, with a more sophisticated one, by using a service for speech recognition, known as *IBM Watson*. In this way, we were able to reduce the computational load on the robot and avoid many issues with voice recognition when using the integrated microphone. In the real-world scenario tested, the robot accomplished the object recognition task with satisfactory response times, as we observed an average response time of about 4 s, resulting in a smooth interaction between the human and the machine. We did not carry out a systematic time performance analysis since response time strictly depends on the platform used for the communication with the robot. Our experiments show that our approach is potentially applicable also in real-world conditions, despite some the hindrances as previously discussed.

7 Conclusions and future works

In this paper, we have investigated several approaches as parts of a unified framework for knowledge management, conceptualization and construction in artificial intelligent systems. Such techniques shares the common goal of combining semantics-based and perception-based information to be successfully transformed into knowledge.

They have been implemented and evaluated as modules of the system. The methodologies explored are both based on supervised learning techniques and unsupervised learning techniques, where the knowledge construction process is semantics-driven for the first class and perception-driven for the second class. The paper was focused on the so called bottom-up approach, while unaware of semantics during the learning phase it has shown the ability to form meaningful semantic clusters from visual observations.

Given the same nature and the shared goal of the two kind of approaches, their results have been used to define a new combined approach by weighting the individual models. The approaches have been also validated via experimental setups using Aldebaran Pepper social humanoid robot. The simulations and real-world experiments on the mobile robotic platform (Aldebaran Pepper) show the ability of generalization as well as the emergence of self-consciousness capabilities.

An interesting perspective would be that of translating the implementation for faster compiler-based languages, such as C++, in order to speed up the whole performance in real-word applications which have real-time execution constraints. Another point of technical improvement, more in long-term way, concerns the use of more powerful techniques, such as instance-based segmentation deep neural network as an approach for knowledge construction. Since, both hardware and software are constantly evolving, the used approaches could be easily replaced with better analogues in the future, like the one cited, to be embedded in mobile CPUs of robots for better performances.

We are also investigating new techniques for improving the object recognition task by introducing the recognition of more complex objects to be stored in our knowledge base with the addition of spatiotemporal relationships for a more comprehensive representation of knowledge. Given the continuous drop in prices of the robot market and the consequent spread

of companion robots, it would likely be easier to pursue the idea of multi-agent distributed knowledge system, where each agent provides its knowledge and share it to other agents.

References

1. Russo C, Madani K, Rinaldi AM (2020) An unsupervised approach for knowledge construction applied to personal robots. IEEE Trans Cognit Dev Syst 13(1):6–15
2. Russo C, Madani K, Rinaldi AM (2019) Knowledge construction through semantic interpretation of visual information. In: International Work-Conference on Artificial Neural Networks, pp. 246–257. Springer
3. Albus JS (1991) Outline for a theory of intelligence. IEEE Trans Syst Man Cybern 21(3):473–509
4. Brooks R (1986) A robust layered control system for a mobile robot. IEEE J Robot Autom 2(1):14–23
5. Brooks RA (1990) Elephants don't play chess. Robot Auton Syst 6(1–2):3–15
6. Brooks RA (1991) Intelligence without representation. Artif Intell 47(1–3):139–159
7. Nonaka I (1994) A dynamic theory of organizational knowledge creation. Organ Sci 5(1):14–37
8. Bandara HRT, Muthugala MVJ, Jayasekara ABP, Chandima D (2018) Cognitive spatial representative map for interactive conversational model of service robot. In: 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 686–691. IEEE
9. Bandara HRT, Muthugala MVJ, Jayasekara ABP, Chandima D (2018) Grounding object attributes through interactive discussion for building cognitive maps in service robots. In: 2018 IEEE international conference on systems, man, and cybernetics (SMC), pp 3775–3780. IEEE
10. Samarakoon SBP, Muthugala MVJ, Jayasekara ABP (2018) Identifying approaching behavior of a person during a conversation: a human study for improving human-robot interaction. In: 2018 IEEE international conference on systems, man, and cybernetics (SMC), pp 1976–1982. IEEE
11. Samarakoon SBP, Muthugala MVJ, Jayasekara ABP (2018) Replicating natural approaching behavior of humans for improving robot's approach toward two persons during a conversation. In: 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 552–558. IEEE
12. Alami R, Albu-Schäffer A, Bicchi A, Bischoff R, Chatila R, De Luca A, De Santis A, Giralt G, Guiochet J, Hirzinger G et al (2006) Safe and dependable physical human-robot interaction in anthropic domains: state of the art and challenges. In: 2006 IEEE/RSJ international conference on intelligent robots and systems, pp 1–16. IEEE
13. Ficuciello F, Siciliano B (2016) Learning in robotic manipulation: the role of dimensionality reduction in policy search methods. Comment on" hand synergies: Integration of robotics and neuroscience for understanding the control of biological and artificial hands" by marco santello et al. Phys Life Rev 17:36–37
14. Vitiello N, Oddo CM, Lenzi T, Roccella S, Beccai L, Vecchi F, Carrozza MC, Dario P (2015) Neuro-robotics paradigm for intelligent assistive technologies. Springer Tracts Adv Robot. https://doi.org/10.1007/978-3-319-12922-8_1
15. Chibani A, Bikakis A, Patkos T, Amirat Y, Bouznad S, Ayari N, Sabri L (2015) Using cognitive ubiquitous robots for assisting dependent people in smart spaces, pp 297–316
16. Hortal E, Úbeda A, Iáñez E, Azorín JM (2015) Brain-machine interfaces for assistive robotics, pp 77–102
17. Neto AF, Elias A, Cifuentes C, Rodriguez C, Bastos T, Carelli R (2015) Smart walkers: advanced robotic human walking-aid systems, pp 103–131
18. Low KH (2015) Recent development and trends of clinical-based gait rehabilitation robots, pp 41–75
19. Raya R, Rocon E, Urendes E, Velasco MA, Clemotte A, Ceres R (2015) Assistive robots for physical and cognitive rehabilitation in cerebral palsy, pp 133–156

20. Bauer A, Wollherr D, Buss M (2008) Human-robot collaboration: a survey. Int J Humanoid Rob 5(01):47–66
21. Chandrasekaran B, Conrad JM (2015) Human-robot collaboration: a survey. In: SoutheastCon 2015, pp 1–8. IEEE
22. Ajoudani A, Zanchettin AM, Ivaldi S, Albu-Schäffer A, Kosuge K, Khatib O (2018) Progress and prospects of the human-robot collaboration. Auton Robots, 1–19
23. Zhao X, Pan J (2018) Considering human behavior in motion planning for smooth human-robot collaboration in close proximity. In: 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 985–990. IEEE
24. Lemaignan S, Alami R (2013) Explicit knowledge and the deliberative layer: lessons learned. In: 2013 IEEE/RSJ international conference on intelligent robots and systems, pp 5700–5707. IEEE
25. Lemaignan S, Warnier M, Sisbot EA, Clodic A, Alami R (2017) Artificial cognition for social human-robot interaction: an implementation. Artif Intell 247:45–69
26. Tenorth M, Beetz M (2009) Knowrob—knowledge processing for autonomous personal robots. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, pp 4261–4266. IEEE
27. Tenorth M (2011) Knowledge processing for autonomous robots. PhD thesis, Technische Universität München
28. Galindo C, Fernández-Madrigal J-A, González J, Saffiotti A (2008) Robot task planning using semantic maps. Robot Auton Syst 56(11):955–966
29. Milliez G, Warnier M, Clodic A, Alami R (2014) A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. In: The 23rd IEEE international symposium on robot and human interactive communication, pp 1103–1109. IEEE
30. Brewster C, O'Hara K (2007) Knowledge representation with ontologies: present challenges–future possibilities. Int J Hum Comput Stud 65(7):563–568
31. Caldarola EG, Rinaldi AM (2016) An approach to ontology integration for ontology reuse. In: Information reuse and integration (IRI), 2016 IEEE 17th international conference on, pp 384–393. IEEE
32. Rinaldi AM, Russo C (2018) A matching framework for multimedia data integration using semantics and ontologies. In: 2018 IEEE 12th international conference on semantic computing (ICSC), pp. 363–368. IEEE
33. Rinaldi AM (2014) A multimedia ontology model based on linguistic properties and audio-visual features. Inf Sci 277:234–246
34. Rinaldi AM (2015) A complete framework to manage multimedia ontologies in digital ecosystems. Int J Bus Process Integr Manag 7(4):274–288
35. Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41
36. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp 248–255. IEEE
37. Renk T, Kloeck C, Burgkhardt D, Jondral FK, Grandblaise D, Gault S, Dunat J-C (2008) Bio-inspired algorithms for dynamic resource allocation in cognitive wireless networks. Mobile Netw Appl 13(5):431–441
38. Ramík DM, Madani K, Sabourin C (2013) From visual patterns to semantic description: a cognitive approach using artificial curiosity as the foundation. Pattern Recogn Lett 34(14):1577–1588
39. Ramík DM, Sabourin C, Moreno R, Madani K (2014) A machine learning based intelligent vision system for autonomous object detection and recognition. Appl Intell 40(2):358–375
40. Erlhagen W, Mukovskiy A, Bicho E, Panin G, Kiss C, Knoll A, Van Schie H, Bekkering H (2006) Goal-directed imitation for robots: a bio-inspired approach to action understanding and skill learning. Robot Auton Syst 54(5):353–360
41. Kohonen T (1990) The self-organizing map. Proc IEEE 78(9):1464–1480
42. Thompson RF, Johnson RH, Hoopes JJ (1963) Organization of auditory, somatic sensory, and visual projection to association fields of cerebral cortex in the cat. J Neurophysiol 26(3):343–364
43. Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. Cereb Cortex 10(9):873–888
44. Dong Z, Yang D, Reindl T, Walsh WM (2015) A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance. Energy 82:570–577
45. Dittenbach M, Merkl D, Rauber A (2000) The growing hierarchical self-organizing map. In: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, vol. 6, pp. 15–19. IEEE
46. Rauber A, Merkl D, Dittenbach M (2002) The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. IEEE Trans Neural Netw 13(6):1331–1341
47. Fritzke B (1995) Growing grid–a self-organizing network with constant neighborhood range and adaptation strength. Neural Process Lett 2(5):9–13

48. Yager RR, Kacprzyk J (eds.) (1997) The ordered weighted averaging operators. https://doi.org/10.1007/978-1-4615-6123-1
49. Wu Z, Palmer M (1994) Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on association for computational linguistics. ACL '94, pp. 133–138. Association for Computational Linguistics, USA. https://doi.org/10.3115/981732.981751

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Kurosh Madani received the graduation degree in fundamental physics from Paris 7-Jussieu University, Paris, France, in June 1985, the M.Sc. degree in microelectronics and the PhD degree in electrical engineering and computer sciences from the University of Paris-Sud, Orsay, France, in September 1986 and February 1990, respectively, and the DHDR Doctor Hab. degree (senior research doctorate degree) from the University Paris 12 Val de Marne, Créteil, France, in 1995. He works as the Chair Professor of electrical engineering with the Senart-FB Institute of Technology, University Paris-Est Creteil, Créteil, where he was the Co-Creator with the Images, Signals, and Intelligent Systems Laboratory (LISSI/EA 3956) in 2005. He has authored and coauthored of more than 350 publications (in international journals, books, conferences', and symposiums' proceedings). Dr. Madani has been awarded by the Academic Palm of French Government (Chevalier de l'Ordre des Palmes Académiques) for his academic and scientific achievements in 2017.

Antonio Maria Rinaldi received the graduation degree in computer engineering and the PhD degree in computer science and systems from the University of Napoli Federico II, Napoli, Italy, in 2005. His fields of interests are knowledge representation and retrieval, big data, multimedia database, deep learning and GIS. He is currently an Associate Professor with the Department of Electrical Engineering and Information Technology, University of Napoli Federico II, Napoli, Italy. Dr. Rinaldi is the Head of the Intelligent KNOwledge Systems Lab (IKNOS LAB) LUPT-UNINA, member of the CINI-ITEM Laboratory Board, the ARhemLab-DIETI Technical Board and MISlab-DIETI. He is an Associate Researcher at CNR IEOS "G. Salvatore" - National Research Council of Italy Institute of Experimental Endocrinology and Oncology "G. Salvatore". He is member of several journal editorial boards, conference program and reviewer committees, and he is author and co-author of numerous research papers in international journals and conference proceedings. In the last years, he has been involved in several research and industrial projects related to information retrieval, multimedia database integration, and geographic information systems.

Cristiano Russo received the PhD degree from both Universitá degli Studi Di Napoli Federico II and Université Paris-Est in 2020. He received the MSc degree from the Department of Electrical Engineering and Information Technologies, Universitá degli Studi Di Napoli Federico II, in 2017. His research interests include bioinformatics, information retrieval, ontologies, NoSQL databases, cognitive robotics, machine learning and data mining. Russo is part of the Intelligent KNOwledge Systems Lab (IKNOS LAB) and COmputational and Systems Biology Lab (COSMO LAB) CNR IEOS-UNINA. He also collaborates with Pattern analysis and Intelligent Computation for Multimedia Systems Lab (Picus Lab) at UNINA. He is program committee member of international conferences, and he contributes as a reviewer for several international journals and conferences. He is currently studying the field related to gene regulatory networks at Consiglio Nazionale delle Ricerche(CNR) in Naples, Italy, and he has also didactical commitments as contract lecturer with different universities. He has several publications in international conference and journals. Russo has been supervisor of several MSc Management Engineering thesis and co-supervisor of BSc/MSc Computer Science Engineering thesis.

Cristian Tommasino received the MSc degree in Computer Science Engineering from the Department of Electrical Engineering and Information Technologies of the University of Napoli Federico II (UNINA) in 2019 with a thesis on web document retrieval using visual and semantic techniques. He is currently a PhD student in Information Communication Technologies for Health (ICTH) at the University of Napoli Federico II, where he is studying biomedical/biological data modelling through ontological representation and biological images processing and analyzing. His research interests include computational pathology, multimedia biodata analysis, information retrieval, ontologies, NoSQL databases, cognitive robotics, machine learning, Big Data, deep learning and data mining. Tommasino is part of the Intelligent KNOwledge Systems Lab (IKNOS LAB) and COmputational and Systems Biology Lab (COSMO LAB) CNR IEOS-UNINA. He also collaborates with Pattern analysis and Intelligent Computation for Multimedia Systems Lab (Picus Lab) at UNINA. Tommasino is a member of the program committee of the AIKE conference, and he contributes as a reviewer for several international journals and conferences. Tommasino has been supervisor of several MSc Management Engineering thesis and co-supervisor of BSc/MSc Computer Science Engineering thesis.