

Towards Anthropomorphic Machine Learning

Plamen P. Angelov, School of Computing and Communications, Lancaster University, LA1 4WA, UK; also Honorary Professor at Technical University – Sofia, Bulgaria; e-mail:

p.angelov@lancaster.ac.uk

Xiaowei Gu, School of Computing and Communications, Lancaster University, LA1 4WA, UK; e-mail: x.gu3@lancaster.ac.uk

Abstract

In this paper, we introduce and discuss the concept of *anthropomorphic machine learning* as an emerging direction for the future development in the area of artificial intelligence (AI) and data science. We start with outlining research challenges and opportunities, which the contemporary landscape offers. We focus on machine learning, statistical learning, deep learning and computational intelligence as theoretical and methodological areas of greater promise for breakthrough results and underpinning the future revolutionary changes in technology development as well as in our everyday life and societies. Our critical analysis brings us to the open problems and we formulate the paradigm shift in the understanding of machine learning. In a nutshell, our vision for the next generational machine learning methods and algorithms is *anthropomorphic*, which resembles the way people/humans learn from data. This concept brings machine learning from the statistics to the area of computational intelligence and AI.

Keywords- *anthropomorphic*, machine learning, computational intelligence

Data-rich environment, “digital obesity”, the new interpretation of Moor’s law in regards to the data

We are witnessing technological developments with revolutionary potential – they are re-shaping industries, the Society, our everyday life. These are based primarily on the exponentially growing amount, complexity and sources of data (streams). They strongly influence our ability to deal with such data-rich environments, in which there exists an abundance of data of various modalities. The data increasingly come in a form of streams, are non-linear, non-stationary and increasingly more multi-modal/heterogeneous: combining various physical variables, signals with images/videos as well as text. Now, since the famous Moors law is not so relevant for the hardware advance (because it reached the physical limits of miniaturization to a nanoscale), the breakthrough is required in the way we deal with data from algorithmic point of view. We are again witnessing an exponential growth similar to Moore’s law nowadays, but now it is in terms of the amount of data we produce and process rather than the hardware capacity and circuit integration. From personal aspect, human beings have to cope with this “digital obesity” from an earlier age with all positive and negative consequences. Nonetheless, in this paper we will focus on the technical and theoretical aspect of this revolutionary process.

This reality is drastically different from the reality when the fundamental results of the probability theory, statistics and statistical learning were developed few decades and centuries ago. It calls for new approaches bringing some specific new challenges, but also offering new opportunities.

Are traditional machine learning, statistical learning and even the most recent deep learning fit to this environment?

Traditional machine learning models [1] have internal structures, which are largely pre-determined by the assumptions made at the start of the process. For example, they strongly depend on the assumed *a priori* data generation model [2]. The interpretability of the traditional statistical machine learning models is typically lower than that of the so-called “first principles” models, which themselves are often based on differential equations representing mass, energy balance or approximations of the laws of physics. However, the level of their interpretability is typically higher than that of some computational intelligence models such as artificial neural networks (ANN). The level of interpretability and transparency of “first principles” models is closer to the level of transparency and interpretability of another form of computational intelligence - the fuzzy rule-based (FRB) systems. In addition, there are some parallels and dualism between FRB system and mixture Gaussian probabilistic or Bayesian models [1].

The most successful classifiers, for example, support vector machines (SVM) [3] and deep learning neural networks (DLNN) [4], [5], e.g. convolutional neural networks (CNN) and recurrent neural networks (RNN), have very low level of interpretability [4]. Their internal structure is not informative or clear to the human users and does not relate to the problem at hand (for example, they cannot explain clearly the reasons for certain decisions, and do not allow alternatives to be explored). This is an especially acute problem for the quickly popularized DLNNs, which are able to extract high-level abstractions from images and have already demonstrated excellent results in image processing. However, the contemporary DLNNs lack transparency, and the hundreds of millions of weights they use have values that cannot be linked to the problem in an analytical manner understandable for humans [4]. They also do not have a clear semantic of the inner model structure, the number of hidden layers and many other parameters are decided *ad hoc*. Opacity and incapability of working under uncertainty scenarios are two of the most challenging problems when using deep learning, as recently stated by Google and Drive.ai [6], [7]. Wrong decisions can lead to life-threatening situations. This is not a trivial problem and requires analytical insights into why DLNNs make particular decisions. The development of a transparent system that is designed to handle uncertainty by being able to recognize unknown scenarios can directly help solving several open problems that currently can only be addressed by deep learning.

What are the limits, challenges?

The mainstream DLNNs, despite their success (reported results comparable or surpassing the human ability [4]) and publicity (including the commercial one) and the increased media interest, still have a number of unanswered questions and deficiencies [6]–[8]:

- The internal architecture of DLNNs is opaque and lacks transparency (there are many *ad hoc* decisions and parameters such as number of layers, neurons, parameter values);
- The training process of DLNNs requires a large amount of data, time and computational resources, which preclude training and adaptation in real-time, therefore, DLNNs cannot cope with the evolving nature of the data (they have

- fixed structure and settings, e.g. number of classes), cannot be used and learn “from scratch”;
- DLNNs are prone to overfitting;
 - DLNNs cannot handle uncertainty. Not only they poorly perform on inference data that is significantly different from the training data, but they are also **unaware** of it (it is practically impossible to analyze the reasons for errors and failures); they can be easily fooled at outputting high confidence scores even when facing unrecognizable images [8].

What are the opportunities, alternatives?

Both machine learning and statistical learning as well as computational intelligence and deep learning as a form of the most advanced and complex artificial neural networks are quite different from the way people learn from data (for example, images). Statistical learning is based on the classical probability theory and iterative optimization. Due to the solid mathematical foundation, the performance of these learning algorithms are, in general, guaranteed when the amount of data tends to infinity and all the data come from known distributions. Nonetheless, the presumed random nature and known distribution(s) imposed on the data generation model are too strong and impractical to be holding true in real situations [2]. In addition, the predefined parameters of machine learning algorithms usually require a certain amount of *prior* knowledge of the problem, which, in reality is often unavailable.

If we compare the way people/humans learn from data (including, but not limited to, images) with the currently existing machine learning approaches, we can notice some **fundamental differences**, for example:

- People can easily **learn from a single or a handful of examples**;
- People do **learn lifelong** rather than “train and fix our recognition rules”;
- Our recognition rules (**classifiers**) **do evolve** (new rules and/or classes can emerge or some rules and/or classes are being merged as we grow up; the number of rules and/or classes is not fixed once and forever) [9];
- People **can add new classes, rules, etc., “on the fly”, in real-time** (we do not stop for re-training) [9];
- People **can explain why** they recognize a particular image (assign to a certain class), for example, colour of the hair, ears, nose if we consider face recognition or details if we talk about cars or houses, etc.;
- People can **learn collaboratively** (they can exchange aggregated data to each other and learn from it) [9];

We see great opportunities in developing the next generation Machine Learning in an anthropomorphic manner. In our view, the future intelligent machines and machine intelligence will be more human-friendly, human-like while offering a much higher throughput, automation and thus, augmenting our (human) capabilities.

The research challenge is, thus, to develop and implement accurate, transparent, adaptable, self-learning and computationally efficient machine learning methods (i.e. classifiers for images, predictors, controllers, anomaly detectors, recommender systems, etc.) capable of detecting unknown scenarios. These next generation machine learning

methods should not involve handcrafting, restrictive prior assumptions, user- or problem-specific parameters. That is, they have to be with high degree of autonomy. However, they have to be human-friendly, transparent and human-like (anthropomorphic).

Anthropomorphic machine Learning

Anthropomorphic characteristics of a machine learning method enable computers “to learn like humans do”. Nowadays, the vast majority of machine learning methods, including the mainstream DLNNs, require a huge amount of training data to be able to work. However, people can recognize an object that they have seen only once.

Just imagine the Opera House in Sydney, Australia or Tower Bridge in London, England— how many training images do we need to be able to associate new images with it next time. Humans are able to associate and make the respective reasoning, classification or anomaly detection based on a single or few examples. Therefore, **the ability of building a model from a single or very few training data samples, namely, to “start from scratch”, is a human-like (anthropomorphic) characteristic.**

Another human-like feature is **the ability to explain the decisions made**. When we recognize a bird, we never say “this is a parrot, but I do not know why”. Instead, we would say “this bird looks very much like a parrot because of its head, its beak, its claws, its wings, color of the feathers, sounds it makes, etc.” However, DLNNs are not able to do that. Identifying and **using prototypes** to which the new observations are associated makes it possible and easy to explain the reason for a certain decision. This is an anthropomorphic (human-like) feature that the current DLNNs, SVM as well as other state-of-the-art approaches cannot offer.

Further, **the ability to learn continuously** is also an anthropomorphic characteristic, which the vast majority of the existing methods (with the exception of the reinforcement learning method) cannot offer. However, reinforcement learning cannot guarantee convergence. In comparison, humans dynamically evolve their internal perception of the real world (the model of the environment) all the time. We all are able to learn **new information** and update the previously learned one. This ability is another anthropomorphic characteristic, which most mainstreaming machine learning methods lack, but is badly needed because of the raising challenges in today's world [9].

Last, but not least, human brain is remarkably efficient from the energy consumption point of view. It coordinates a large number of extremely complex tasks all day long. It consumes around 12 watts—a fifth of the power required by a standard 60 watt lightbulb, but makes miracles. The key to have anthropomorphic algorithms and methods for machine learning is to **learn in a lean, computationally efficient manner**. However, the state-of-the-art approaches, especially, DLNNs, are computationally expensive and cumbersome. Currently, they require HPC, accelerators (i.e. GPUs), etc. A lot of attempts were made in optimizing the hardware (e.g., neuromorphic chips Loihi by Intel [10]). However, we believe that **the huge untapped potential resource is the fundamentally different algorithmic processing using recursive, non-iterative, prototype-based methods.**

In short, the reality we live in requires **a shift towards anthropomorphic machine learning approaches.**

Bottleneck issues in mainstream approaches

Ideally, an anthropomorphic system will be able to:

- learn from a single or very few examples;
- learn lifelong from newly observed examples;
- explain what it knows **as well as know what it does not know and why it made specific errors**;
- detect data samples that are significantly different from all the previously observed ones and form new rules and/or classes when necessary (self-learning and self-organizing);
- perform learning in a lean, computationally efficient manner;
- collaborate with other anthropomorphic machine learning systems (if any).

However, the key bottleneck issue in the mainstream approaches is the lack of situational awareness. In addition, **the ability of presenting to the users the already learned knowledge in a transparent, human-interpretable manner** requires much more attentions as well. This also contributes to the **reliability** of the system.

Indeed, DLNNs can learn from the data under the most diverse scenarios but remain black boxes. They are able to perform highly accurate classification on data with familiar patterns. However, they completely fail when confronted with unfamiliar data patterns that deviate from the previously observed ones. Furthermore, DLNNs are unable to recognize when this happens. This deficiency is not trivial, but can cause serious problems, which further prohibits the use of DLNNs in real world applications. One well-known example is the self-driving vehicles which use DLNN-based computer vision technologies [6], [7]. They have been experimentally studied and even tested on real roads for a few years. However, although self-driving vehicles are able to perform well around **familiar** scenarios (highway, residential area, day, night, rain, snow, low traffic, high traffic, etc.), they are **unable to recognize** whether the current scenario is **known or not**. This can lead to a high risk and create a lot damages and even fatalities. As a result, **identifying the unknown situations** and autonomously learning from them are of paramount importance for the safety and wider use of this technology [9].

Ultimately, the goal has to be building a system that is able not only to recognize previously known patterns, but also to recognize the “unexpected” ones. In a way, the problem can be seen as **a system that is aware of its limitations** and is able to initiate a safety procedure when facing an unknown impossible to predict situation and learns from this autonomously. Another very important characteristic for an algorithm is whether it can work in the so-called “*one-pass*” mode. The term of “one-pass” means that a data sample (e.g. an image) once taken and processed by the algorithm is then thrown away, not stored in the memory nor used anymore. Learning models designed in offline mode may have a very good performance for validation data if the validation data share similar characteristics and patterns with the training data. However, their performance dramatically worsens facing the so-called data *shift* and *drift* [9]. In such cases, models designed offline require a recalibration or a complete re-design, which may be very expensive in terms of not only model design costs, but also the possible related industry stoppage costs. In real problems, this is often the case as significant changes may happen due to, for example, raw material quality variations (in oil refining), illuminations conditions (i.e. day, night) in image processing, etc.[9]

Being “one-pass” is a very demanding form of data processing. It requires the algorithm to rely only on the current data sample and possibly some statistical aggregations, model

parameters (such as cluster centers, prototypes, etc.), but not to use a (possibly, sliding) window of data samples or the so-called “data chunks”. “One-pass” is a very efficient mode of operation in terms of using memory of the machine, computational resources (search, storage) and can work lifelong. The difference is very significant. For example, in video processing, a “one-pass” algorithm only processes the current image frame and does not store past frames in the memory. Instead, it stores in the memory a small amount of statistical aggregations and model parameters only [9].

The ability of an algorithm to be “one-pass” is strongly depends on it being non-iterative and recursive [9]. The reason is simple, if an algorithm involves an iterative search procedure, there is no guarantee that it will get the result before the next data sample arrives. Having a recursive algorithm guarantees the ability of being “one-pass”, because recursive algorithms update their parameters at the current time step using only the values of these parameters at the previous time step and some simple arithmetic operations such as summation, multiplication by numerical coefficients.

First step towards anthropomorphic machine learning

As one of the recent advances in the field of machine learning, the recently introduced Deep Rule-Based (DRB) systems [11], [12] provide a possible solution to the problem. The DRB system is a principally new generic approach that combines the advantages of both, the self-organizing non-parametric FRB system and the multi-layer structure of DLNNs [11]. The DRB system is able to offer a self-organizing, self-adaptive, transparent, highly parallelizable rule-based architecture and learning algorithm with theoretically proven convergence. It has to be stressed that this new method is a general machine learning approach and is applicable to various classification and prediction problems with simple modifications, but in this paper, we will present the general concepts and principles focusing on image classification problems [11]–[13].

The general architecture of a DRB system for image classification is given in Fig. 1. As one can see from the figure, the classifier is composed of the following components [11]–[13].

1) **Pre-processing block**, which involves the widely used pre-processing techniques in the field of computer vision including:

- i)* normalization,
- ii)* scaling,
- iii)* rotation, and
- iv)* segmentation, etc.

Thus, it is, in fact, composed of a number of sublayers serving various purposes.

2) **Feature descriptor**, which projects the original images to a feature space that makes images from different classes separable, namely $\mathbf{I} \Rightarrow \mathbf{x}$.

3) **Massively parallel fuzzy rule-base**, which is a complex non-linear predictive model serving as the “learning engine” of the system. The predictive model itself is an ensemble of massively parallel AnYa type 0-order fuzzy rules [9], and each massively parallel fuzzy rule is composed of a (large, but not pre-determined) number of prototypes identified from samples of a particular class within the training set. Thus, for a training set containing data samples (e.g. images) of C different classes, there would be C fuzzy *IF ... OR...OR...OR... THEN...* rules identified in parallel (one rule per class).

number of prototypes

4) **Decision-maker**, which decides the winning class label based on the partial suggestions (degree of confidence) of the massively parallel local/sub-decision makers per IF...THEN...

rule/per class. Generally, the global decision-maker as well as the local ones within the DRB system uses the popular “winner takes all” principle. But, other decision-making strategies, i.e. “few winners take all,” “fuzzily weighted,” “average” can also be considered depending on specific problems.

Alternatively, a DRB system can also be viewed as a self-evolving ANN [11].

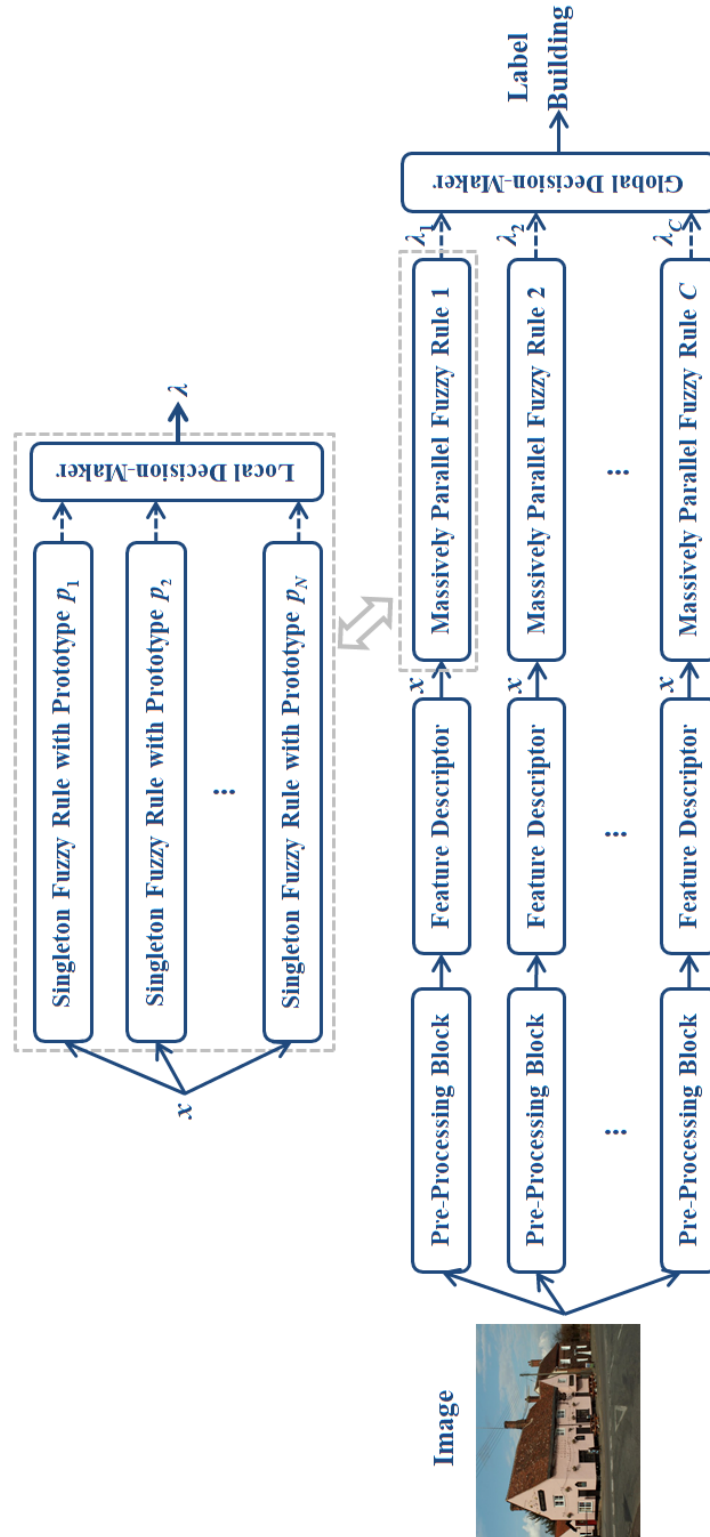


Fig. 1. General architecture of a DRB system for image classification

The DRB approach, which has anthropomorphic characteristics, works with prototypes of the data. In practice, these prototypes are the local maxima of the *multi-modal typicality/data density* automatically extracted from the observed data as described in [2]. Prototypes are instrumental to the DRB approach, they set this approach apart from the mainstream approaches including DLNNs [2], [11], [12]. These prototypes are identified based on the disclosed ensemble properties and mutual distribution of the data using parameter-free operators through a fully autonomous, online, non-parametric, non-iterative and “one-pass” learning process [2]. They are the most representative samples within the dataset, and directly reflect the knowledge that the DRB approach gained from the data [2].

Based on the identified prototypes, the DRB approach self-organizes and self-evolves from data a fully transparent and human-interpretable IF...THEN... massively parallel FRB system [11]. Each massively parallel fuzzy rule is formulated around a (large) number of prototypes identified from the training data samples of a particular class connected by a local decision-maker using “winner takes all” principle. An example of such fuzzy rule is given in expression (1) [11]–[13]:

$$\mathbf{R}_i : \begin{array}{l} \text{IF}(\mathbf{x} \sim \mathbf{p}_{i,1}) \text{OR}(\mathbf{x} \sim \mathbf{p}_{i,2}) \text{OR} \dots \text{OR}(\mathbf{x} \sim \mathbf{p}_{i,N_i}) \\ \text{THEN}(\text{Class } i) \end{array} \quad (1)$$

where “ \sim ” denotes similarity, which can also be seen as a fuzzy degree of membership/satisfaction; \mathbf{x} is a particular sample; $\mathbf{p}_{i,j}$ is the j^{th} prototype in the fuzzy rule corresponding to i^{th} class; N_i is the number of prototypes of the i^{th} class. Equivalently, \mathbf{R}_i can also be interpreted as a (large) number of simpler fuzzy rules sharing the same singleton consequent part as given by expression (2):

$$\begin{array}{l} \mathbf{R}_{i,1} : \text{IF}(\mathbf{x} \sim \mathbf{p}_{i,1}) \text{ THEN}(\text{Class } i) \\ \mathbf{R}_{i,2} : \text{IF}(\mathbf{x} \sim \mathbf{p}_{i,2}) \text{ THEN}(\text{Class } i) \\ \vdots \\ \mathbf{R}_{i,N_i} : \text{IF}(\mathbf{x} \sim \mathbf{p}_{i,N_i}) \text{ THEN}(\text{Class } i) \end{array} \quad (2)$$

The degree of membership/satisfaction of each fuzzy rule is calculated based on the identified prototype closest to the validation data sample [11]:

$$\lambda_i = \max_{j=1,2,\dots,N_i} \left(e^{-\|\mathbf{p}_{i,j}-\mathbf{x}\|^2} \right) \quad (3)$$

The singleton consequent parts of the massively parallel fuzzy rules further allow the DRB system to be free from user- and problem-specific parameters as well as from the need to handcraft the membership functions.

Each massively parallel fuzzy rule (one per class) can be viewed as a number of simpler fuzzy rules with a single prototype and the same singleton consequent part connected by an “OR” logic operator. Since each fuzzy rule is fully independent from each other, they are trained in parallel. Moreover, each rule can be updated or removed without influencing other rules in the rule base. This allows a very high level of parallelization.

Critically, the DRB approach is not “*black box*” (unlike the mainstream approaches [3], [5]). One of its most appealing aspects is the prototype-based nature, which offers human interpretability and transparency of the internal structure of the system. Most existing machine learning methods require a huge amount of training data, while the DRB system can learn even from a single example, that is, to start “from scratch”. The non-iterative

online autonomous learning algorithm further enables the DRB system to learn lifelong with newly observed samples continuously even after training/deployment, namely, to be “evolving” [9], [12].

Example of a DRB image classifier

For example, let us consider three fuzzy rules trained based on single images of a typical house taken from three countries (UK, Greece and Japan) as presented in Fig. 2(a). By giving more training samples, the DRB system identifies more new prototypes, the structure of the each fuzzy rule becomes massively parallel, the scale of the DRB system grows larger correspondingly and, most importantly, the DRB system gains a better understanding of the problem. The rules are given in Fig. 2(b), from which one can get a better understanding about the differences in the architectural styles extracted autonomously from the data.

The transparency and human-interpretability of the DRB approach brought by its prototype-based nature are of great importance for the real world applications [11]. One can check the aggregated information after the training process and can easily understand and analyze what has been learned by simply looking at the identified prototypes. This also enables users to identify possible mistakes when the deployed system goes wrong, and solve them easily by, for example, deleting the wrongly assigned prototypes, adding new prototypes, etc.

To give an idea of this, let us continue the example presented in Fig. 2(b). One may notice that, the second prototype $p_{3,2}$ of the rule R_3 is actually a typical house in China, which is wrongly labelled as a Japanese type house. To modify the DRB system, one can simply remove $p_{3,2}$ from R_3 and create a new rule R_4 with $p_{3,2}$ as its new prototype. This process is presented in Fig. 2(c) as an illustration.

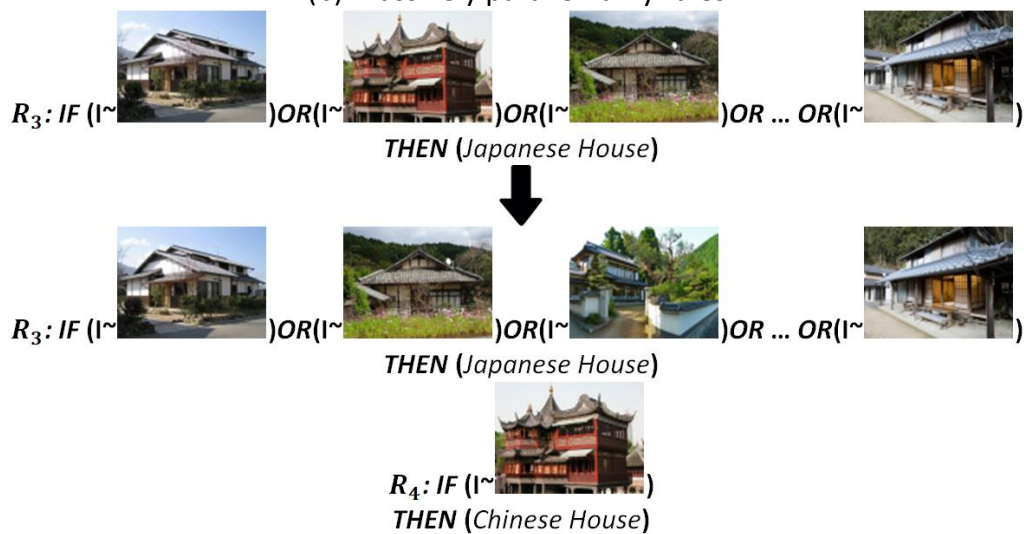
Another anthropomorphic characteristic of the DRB system is its strong sense of situation awareness, which is closely linked with its prototype-based nature [11]–[13] as well. For the newly observed samples, the DRB system will make a comparison with the previously identified prototypes. The system can easily distinguish new samples with unfamiliar patterns based on the lower degrees of membership/satisfaction. This further enables the system to learn continuously without or with very little human intervention or supervision by exploiting prototype-based semi-supervised and active learning algorithms [12]. Then, the DRB system can learn new prototypes as well as new classes actively in an autonomous manner “on the fly”, and users can be (optionally) involved in double-checking the newly gained information afterwards.



(a) Fuzzy rules with single prototype



(b) Massively parallel fuzzy rules



(c) Fuzzy rule modification by replacement of a prototype

Fig. 2. Illustrative examples of massively parallel fuzzy rules

Collaborative machine learning

The transparent, interpretable internal structure and the non-iterative learning process of the DRB systems further allow different machines to learn collaboratively on the same problem but from different aspects by exchanging very little information between each other, namely, the prototypes and the corresponding singleton consequent parts [9], [14]. Collaborative learning is an anthropomorphic characteristic since people usually exchange aggregated information through communication instead of the raw data. This is also a very appealing feature since it provides a feasible solution for handling very large scale data by parallelizing into a number of processing units, which not only speeds up the whole learning process but also lowers the requirement for computation and memory resources for each machine involved in the collaborative learning [9], [14]. In contrast, the training process of the mainstream approaches, i.e. DLNNs, requires iterative solutions, and thus, is very expensive and cumbersome for parallelization and scaling up.

Thanks to the prototype-based nature, the DRB systems can learn collaboratively in a flexible architecture. Each system continuously collects and exchanges with others the

partial information, and based on this aggregates the information. In this way, each DRB system has the key information about the problem learned by all the connected systems without the need to have all the observations.

Anthropomorphic decision support and recommender systems

Decision support systems (DSS) are traditionally based on FRB systems which are usually handcrafted in a subjective way involving human experts. Recommender systems, on the other hand, are often based on statistical analysis, but also require feature selection and many other *ad hoc* decisions. Both can often run into the so-called “curse of dimensionality” problem. For example, a medical DSS design would require a computing expert to collaborate with a medical expert and to determine:

- features of interest, e.g. blood systolic pressure, body temperature, haemoglobin level, etc.
- type of membership function if an FRB approach is used or type of data distribution model, e.g. Gaussian, etc. has to be determined;
- in the FRB approach the number of linguistic terms, e.g. *Low, Medium, High* or also *Very Low, Extremely Low*, etc. has to be determined
- parameters of the membership functions (or of the probability density function (pdf), if using statistical approach) have to be determined, etc.

It is obvious that the traditional approach (both, based on FRB systems or statistical one) are cumbersome, involve ad hoc decisions and human and may run into “curse of dimensionality” problems for realistic high scale applications. Moreover, once they are built, it is very difficult to update them later with new data (they are fixed, offline and not evolving dynamically).

The proposed recently alternative of empirical fuzzy sets and systems [15] possess many of the features of an anthropomorphic machine learning. The way they work is to only ask human experts (the subjective version) or automatically determine from the data (the objective version) to select few prototypes (minimum one per class). For example, for the medical DSS we described above the medical doctor (or based on objective data by applying so called Autonomous Data Partitioning approach [16] to identify prototypes of ill and healthy patients. Similarly, one can easily imagine a recommender system for selecting properties to buy or rent or cars or holidays, etc. Then the empirical fuzzy sets approach will automatically build the membership functions (or alternatively, one can design typicality which is an analogy to the pdf [2]). The system can then be used to automatically determine for future patients or recommend properties, cars, holidays, etc.

Such anthropomorphic DSS and recommender systems are:

- prototype-based;
- highly interpretable, allows the users to know how the decision/recommendation is made; can be used to extract human-intelligible and understandable, transparent linguistic information from data, formalize the knowledge and represent it in a mathematical form through membership function;

- personalized, not “one size fits all” type;
- computationally lean;
- dynamically self-evolving.

Furthermore, the anthropomorphic characteristics of the empirical fuzzy sets and systems [15] further facilitate the collaborations between experts from different areas and the computer scientists. For example, medical doctors and other domain experts are often reluctant to be involved in the design of traditional DSS mainly because of the cumbersome nature of the design process while providing few examples of ill and healthy patients (prototypes of the two classes, in this case) is not such a burden as asking a medical doctor to specify hundreds or thousands of Gaussian membership functions. Moreover, the consistency in determining the parameters and the smooth shape of the membership functions (or pdf) is usually not representing the real data correctly.

Conclusions and future directions

Currently existing machine learning methods were developed over past decades and the closely related statistical learning is based on centuries old probability theory. One particular technique, which borrows the key elements of its concept from the computational intelligence and artificial neural networks, called deep learning, has lately become the “golden standard” in computer vision and machine learning. However, among other problems, (e.g. it requires huge amounts of data and computational power), DLNNs are not adaptable to new scenarios and lack of transparency, being often referred as “black-boxes”.

In this paper, we introduce the concept of anthropomorphic machine learning as one which

- is transparent, not a black box;
- is prototype-based and is, thus, human-interpretable;
- is highly parallelizable, non-parametric, evolving;
- achieves human-level precision.

It further has human-like abilities, including:

- learning from a single example or very few examples;
- learning lifelong from new examples;
- learning efficiently “on the fly”;
- learning collaboratively;
- learning actively with little supervision;
- being aware of what has been learned and why mistakes take place;
- explaining clearly about what have been not learned;
- being aware of what is known, what is unknown and why.

We further demonstrated that anthropomorphic machine learning is possible by giving an example, namely the DRB approach, which was recently introduced. We also presented an approach to anthropomorphic decision support and recommender systems. Undoubtedly, there will be more types of anthropomorphic machine learning developed in the future.

References

- [1] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006.
- [2] P. P. Angelov, X. Gu, and J. Principe, "A generalized methodology for data analysis," *IEEE Trans. Cybern.*, p. DOI: 10.1109/TCYB.2017.2753880, 2017.
- [3] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press, 2000.
- [4] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nat. Methods*, vol. 13, no. 1, pp. 35–35, 2015.
- [6] <https://www.wired.com/2015/10/googles-lame-demo-shows-us-far-robo-car-come/>.
- [7] <http://spectrum.ieee.org/cars-that-think/transportation/self-driving/how-driveai-is-mastering-autonomous-driving-with-deep-learning>.
- [8] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 427–436.
- [9] P. Angelov, *Autonomous learning systems: from data streams to knowledge in real time*. John Wiley & Sons, Ltd., 2012.
- [10] M. Davies, N. Srinivasa, T. H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C. K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y. H. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: a neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.
- [11] P. P. Angelov and X. Gu, "Deep rule-based classifier with human-level performance and characteristics," *Inf. Sci. (Ny)*, vol. 463–464, pp. 196–213, 2018.
- [12] X. Gu and P. P. Angelov, "Semi-supervised deep rule-based approach for image classification," *Appl. Soft Comput.*, vol. 68, pp. 53–68, 2018.
- [13] X. Gu, P. Angelov, C. Zhang, and P. Atkinson, "A massively parallel deep rule-based ensemble classifier for remote sensing scenes," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 345–349, 2018.
- [14] P. Angelov, "Machine learning (collaborative systems)," US 8250004, 2012.
- [15] P. P. Angelov and X. Gu, "Empirical fuzzy sets," *Int. J. Intell. Syst.*, vol. 33, no. 2, pp. 362–395, 2017.
- [16] X. Gu, P. P. Angelov, and J. C. Principe, "A method for autonomous data partitioning," *Inf. Sci. (Ny)*, p. DOI: 10.1016/j.ins.2018.05.030, 2018.

Author Bios

Plamen P. Angelov (M'99–SM'04–F'16) is a Chair Professor of Intelligent Systems with the School of Computing and Communications, Lancaster University, UK. He obtained his Ph.D. (1993) and his D.Sc. (2015) from the Bulgarian Academy of Science. He is the Vice President of the International Neural Networks Society and a member of the Board of Governors of the Systems, Man and Cybernetics Society of the IEEE, a

Distinguished Lecturer of IEEE. He is Editor-in-Chief of the Evolving Systems journal (Springer) and Associate Editor of IEEE Transactions on Fuzzy Systems as well as of IEEE Transactions on Cybernetics and several other journals. He received various awards and is internationally recognized pioneering results into on-line and evolving methodologies and algorithms for knowledge extraction in the form of human-intelligible fuzzy rule-based systems and autonomous machine learning. He holds a wide portfolio of research projects and leads the Data Science group at Lancaster University.

Xiaowei Gu (S'16) received the B.E. and M.E. degrees from the Hangzhou Dianzi University, Hangzhou, China. He is currently a senior research associate and is pursuing the Ph.D. degree in computer science with Lancaster University, UK.