In: N. K. Kasabov, *Time-Space, Spiking Neural Networks and Brain- Inspired Artificial Intelligence*, Springer-Verlag GmbH Germany part of Springer Nature, 2019, Springer Series on Bio- and Neurosystems 7, https://link.springer.com/chapter/10.1007/978-3-662-57715-8_5

PART I. Time-Space and AI

Chapter 1
Evolving processes in *Time-Space*. Deep learning and deep knowledge representation in time-space. Brain-inspired AI.

This chapter presents challenges to information sciences when dealing with complex evolving processes in time-space. The emphasis here is on processes/systems that evolve/develop/unfold/change in time-space and what characterises them. To model such processes, to extract *deep knowledge* that drives them and to trace how they change over time, are among the main objectives of the brain-like approach that we take in this book by using SNN. And before going to SNN in the next chapters, we introduce how evolving processes can be represented as data, information and knowledge in Artificial Intelligence (AI), and more specifically, what is *deep knowledge* that we will target to achieve through *deep learning* in SNN.

This chapter consists of the following sections:

1.1. Evolving processes in time-space
1.2. Characteristics of evolving processes: Frequency, entropy, energy and information.
1.3. Light and sound.
1.4. Evolving processes in Time-Space and Direction.
1.5. From data and information to knowledge
1.6. Deep learning and deep knowledge representation in time-space.
1.7. Statistical, computational modelling of evolving processes
1.8. Brain-inspired AI (BI-AI)
1.9. Chapter summary and further readings for deeper knowledge

1.1. Evolving processes in time-space

Time is defined in the Oxford Dictionary as “The indefinite continued progress of existence, events, etc., in past, present and future regarded as a whole....”. *Time* has been studied for many years by the most prolific scientists and cosmologies [33, 37].

Space is defined in the Oxford Dictionary as “A continuous, unlimited area of expanse which may or may not contain objects...”.

Science aims at understanding Nature and the humanity. Processes in Nature are evolving in both space and time (Fig.1). To understand them humans create models, initially only mental models, as at the time of Aristotle (4c BC) and now mathematical and computational models to extract information and knowledge, and more specifically *deep knowledge* as defined here.

1.1.1. What are evolving processes?

We call *evolving processes* or *evolving systems* those that change, develop, unfold in time. Most evolving processes evolve both in time and in space. Evolving spatio-temporal processes are characterised by sometimes complex interaction between space and time components in a continuous manner. This interaction may change over time. Such processes may also interact with other processes in the environment. It may not be possible to determine in advance the course of interaction, unless we discover the important features, the spatio-temporal patterns and rules that drive such processes and their evolution in time.

Evolving spatio-temporal processes are difficult to model because some of their evolving rules (laws) may not be known a priori, they may dynamically change due to unexpected perturbations, and therefore they may not be strictly predictable in a longer term. Thus, modelling of such processes is a challenging task with a lot of practical applications in life sciences and engineering.

When a real process is evolving, a modelling system needs to be able to trace the dynamics of the process and to adapt to changes in the process. For example, a speech recognition system has to be able to adapt to various new accents, and to learn new languages incrementally. A system that models cognitive tasks of the human brain, needs to be adaptive, as all cognitive processes are evolving by nature. (We never stop learning!) In bioinformatics, a gene expression modelling system has to be able to adapt to new information that would define how a gene could become inhibited by another gene, the latter being triggered by a third gene, etc. There is an enormous number of tasks from life sciences where the processes evolve and change over time.

It would not be an overstatement to say that everything in nature evolves in space and time. But what are the rules, the laws that drive these processes, and how these rules change over time, how do they evolve? If we knew these rules, we could create a computational model that can evolve in a similar manner as the real evolving process, and use this model to make predictions and to better understand the processes. But if we do not know these rules, we can still try to uncover them from the data collected from these processes using machine learning.

The term “evolving” is used here in a broader sense than the term “evolutionary”. The latter is related to a population of individual systems traced over generations [1-3], while the former, as it is used in this book, is mainly concerned with the development of the structure and functionality of an individual system in space and/or time during its lifetime [4]. An evolutionary (population/generation) optimisation of the parameters of this system can be applied as well.

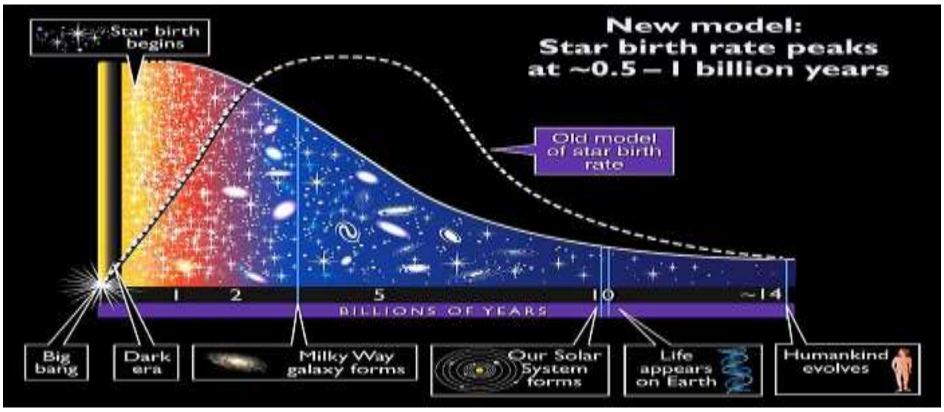


Figure 1. All processes in Nature are evolving in Time-Space, from the emergence of the universe to life and the human brain (after [23])

1.1.2. Evolving processes in living organisms

The most obvious example of an evolving process is life, defined in the Concise Oxford English Dictionary (1983) as “a state of functional activity and continual change peculiar to organized matter, and especially to the portion of it constituting an animal or plant before death, animate existence, being alive”. Continual change in space and time, along with certain stability, is what characterizes life. Modelling living systems requires that the continuous changes are represented in the model, i.e. the model adapts in a life-long mode and at the same time preserves features and principles that are characteristic to the process. The “stability–plasticity” dilemma is a well-known principle of life that is also widely used in connectionist computational models [5].

Perhaps, the most complex information system evolved so far is the human brain. Many interrelated evolving processes are observed at different “ levels” of brain functionality (Fig.2).

6. Evolutionary (population/generation) processes

5. Brain cognitive processes

4. System information processing (e.g. neural ensemble)

3. Information processing in a cell (neuron)

2 Molecular information processing (genes, proteins)

1. Quantum information processing

Figure 2. Many interrelated evolving processes are observed at different “levels” of brain functionality (after [13])

At the quantum level, particles are in a complex evolving state in space and time, being at several locations at the same time, which is defined by probabilities. General evolving rules are defined by several principles, such as entanglement, superposition, etc. [34, 36].

At a molecular level, RNA and protein molecules, for example, evolve and interact in a continuous way based on the DNA information and on the environment. The central dogma of molecular biology constitutes a general evolving rule, but there are specific rules for different species and individuals. Different spatio-temporal folding and unfolding of proteins in a 3D space define different functions cells in the same organism – Fig. 3 [35, 36] (for details see Chapter 15).

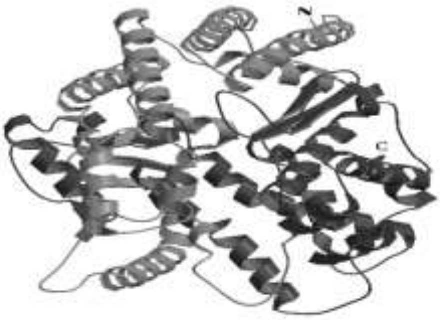


Figure 3. Evolving processes at a molecular level: Different spatio-temporal folding and unfolding of proteins in a 3D space define different functions of cells in the same organism (after [12])

At the cellular level (e.g. a neuronal cell) all the metabolic processes, the cell growing, cell division etc., are evolving processes in space and time. At the level of cell ensembles, or at a biological neural network level, an ensemble of cells (neuros) operates in a concert, defining the function of the ensemble or the network through learning, for instance - perception of sound, perception of an image or learning languages. An example of a general evolving rule is the Hebbian learning rule [6] where neurons create connections between them in space when they are activated in time [36].

In the human brain, complex dynamic interactions between groups of neurons can be observed when certain cognitive functions are performed, e.g. speech and language learning, visual pattern recognition, reasoning, and decision making [36]. When a person is performing a task brain activities are observed in different parts of the brain over time – Fig.4 (see Chapter 3 for details).

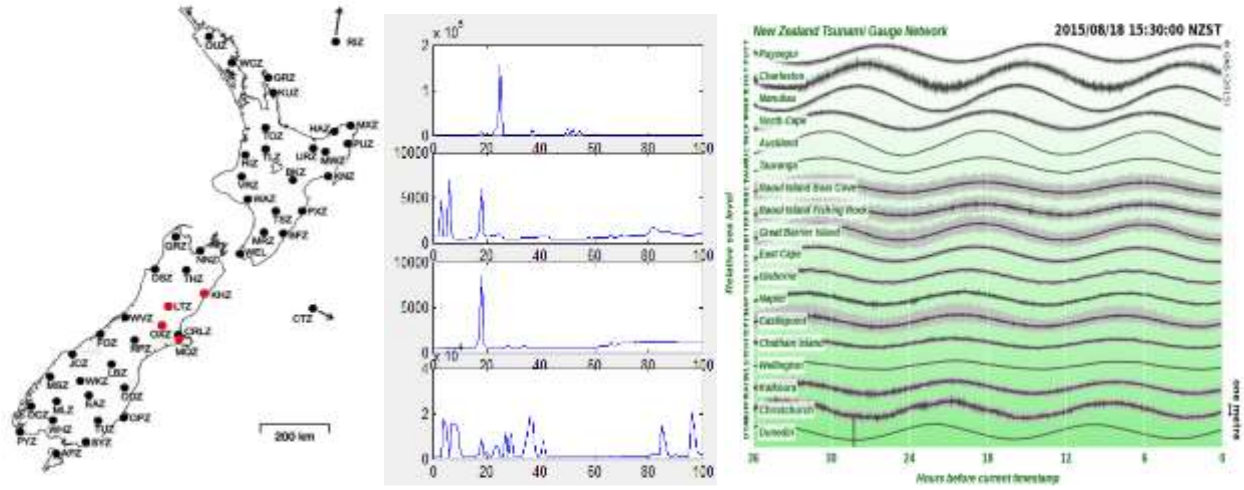
Figure 4. When a person is performing a task brain activities are observed in different spatially located parts of the brain at different times (after [23]).

At the level of population of individuals, species evolve through evolution A biological system evolves its structure and functionality through both lifelong learning of an individual and the evolution of populations of many such individuals [2, 3]. In other words, an individual is a result of the evolution of many generations of populations, as well as a result of its own developmental lifelong learning processes. The Mendelian and Darwinian rules of evolution have inspired the creation of computational modelling techniques called evolutionary computation (EC) [3, 7] (see Chapter 7 for details).

Interaction in Time-Space is what makes a living organism a complex one, and that is also a challenge for computational modelling. For example, there are complex interactions between genes in a genome, and between proteins and DNA. There are complex interactions between the genes and the functioning of each neuron, a neural network, and the whole brain. Abnormalities in some of these interactions are known to have caused brain diseases and many of them are unknown at present. An example of interactions between genes and neuronal functions is the observed dependence between long-term potentiation (learning) in the synapses and the expression of the immediate early genes and their corresponding proteins such as Zif/268 [8]. Genetic reasons for several brain diseases have been already discovered, where some genes are expressed at a later stage of live through interactions with other genes in the genome (see Chapters 16 and 18) .

1.1.3. Spatio-temporal and spectro-temporal evolving processes

The physical interaction between parts of the earth is measured as spatio-temporal seismic data (Fig.5) but what are these deep patterns of interaction that would trigger an earthquake? (see Chapter 19 for details).



(a) (b) (c)

Figure 5. Geophysical processes are both spatio-temporal and spectro-temporal: (a) Spatially located seismic sites in New Zealand; (b) Temporal seismic activities at four selected seismic sites

(spatially located) around Christchurch area manifest different frequency (spectral) characteristics; (c) Sea level at different harbours of New Zealand over time demonstrate both spatial and spectral characteristics (from: http://www.geonet.co.nz).

A sound signal represents a spectro-temporal evolving process in time, e.g. music as shown in Fig.6. as a wave form in time (see Chapters 12,13).

Several sources of signals located at different locations, represent a spatio/spectro-temporal process.

Figure. 6. A wave form of a segment from Mozart's music, represented as intensity of the sound over time, contains spectro-temporal information.

The processes of buying/selling shares on the stock market are spatio-temporal, sometimes presented as only spectro- temporal, i.e. the change of the stock prices in time.

To properly model and understand evolving processes, it is important to first understand their characteristics as discussed in the next section.

1.2. Characteristics of evolving processes: Frequency, energy, probability, entropy and information

Evolving processes are characterised by common characteristics, the most important ones being frequency, entropy, energy and information as explained below.

Frequency: Frequency, is defined as the number of a signal/event changes over a period of time (seconds, minutes, centuries, etc.). Some processes have stable frequencies (they are periodic), but other - change their frequencies over time. Different processes are characterised by different frequencies, defined by their physical parameters. Usually, a process is characterised by a spectrum of frequencies. For example, different frequency spectrums are observed as brain activities (e.g.

alpha waves), speech signals, image and video data, seismic processes, music, quantum processes, etc.

Frequency reflects on the changes in the signal (the data) in time. Evolving processes can manifest different behaviour, depending on the frequency of their changes:

- Random: there is no rule that governs changes of the process in time and the process is not predictable;
- Chaotic: the process is predictable but only in a short time ahead, as the changes of the process at a time moment depends on the process changes at previous time moments via a non-linear function.
- Quasy-periodic: the process is manifesting similarity of its changes over time, but slightly modified each time.
- Periodic: the process repeats same patterns of changes over time and is fully predictable (there are fixed rules that govern the process and the rules do not change over time).

Many complex processes in engineering, social sciences, physics, mathematics, economics and other sciences are evolving by nature and can be analysed using the above classification. Some dynamic time series in nature manifest chaotic behaviour, i.e. there are some vague patterns of repetition over time, and the time series are approximately predictable in the near future, but not in the long run [9 - 12]. Chaotic processes are usually described by mathematical equations that use some parameters to evaluate the next state of the process from its previous states. Simple formulae may describe a very complicated behaviour over time: e.g. a formula that describes fish population growth $F(t + 1)$ is based on the current fish population $F(t)$ and a parameter g [9]:

$$F(t + 1)= 4gF(t)(1 - F(t)) \qquad (1)$$

When $g > 0.89$, the function becomes chaotic.

A chaotic process is defined by evolving/changing rules, so that the process lies on the continuum of “orderness” somewhere between random processes (not predictable at all) and quasi-periodic processes (predictable in a longer time-frame, but only to a certain degree). Modelling a chaotic process in reality, especially if the process changes its rules over time, is a task for an adaptive system that captures the changes in the process in time, e.g. the value for the parameter g from the formula above.

All problems from engineering, economics and social sciences that are characterised by evolving processes require continuously adapting models to model them. A speech or sound recognition system (Chapters 12,13), an image recognition system (Chapters 12,13), a multimodal information processing system, a stock prediction system, an intelligent robot, a system that predicts the emergence of insects based on climate (Chapter 19), etc. should always adjust its structure and functionality for a better performance over time. This book offers one approach to achieving this using spiking neural networks (SNN).

Everything is evolving, living organisms for sure, but what are the evolving rules, the laws that govern these processes? Are there any common evolving rules for every material item and for every living organism, along with their specific evolving rules? And what are the specific rules? Do these rules change over time, i.e. do they evolve as well? These are questions that we will address in this book to certain degree, as the process of addressing these issues is also evolving, with our improved understanding of both the processes and the methods that we can use to deal with them.

An evolving process, characterised by its evolving, governing rules, manifests itself in a certain way and produces data that in many cases, can be measured. Through analysis of this data, one can extract patterns of relationship, rules that describe the processes at a certain time, but do they describe the evolving process in the future?

Here we will introduce some of the main characteristics of evolving processes, used to model and assess them in the other chapters of the book.

Energy.

Energy is a major characteristic of any object and organism. The Albert Einstein's most celebrated energy formula defines energy E as depending on the mass of the object *m* and the speed of light *c*:

$$E=m.\ c^2 \qquad (2)$$

The speed of light is used as a constant. It is appr. 300,000 km/sec. Some characteristics of light are important to note as they are used in some of the methods in this book and discussed in the next section.

Probability, entropy and information

Evolving processes generate data, that can be measured and then used to extract information and knowledge. Having data measuring an evolving process, the question is how do we measure the information contained in the data? There are several ways to define and to measure information depending on the processes.

One way is to use a measure of changes in a process called *entropy,* calculated with the use of measure of the uncertainties in these changes called *probability*, as explained below.

The formal theory of probability relies on the following three axioms, where p(E) is the probability of an event E to happen and p(¬E) is the probability of an event not to happen. E1, E2,…,Ek is a set of mutually exclusive events that form an universe U:

Axiom 1. $0 <= p(E) <= 1$

Axiom 2. $\Sigma p(Ei)=1$, $E1 \cup E2 \cup ... \cup Ek=U$, U- problem space

Corollary: $p(E) + p(\neg E) =1$

Axiom 3. $p(E1 \vee E2)=p(E1)+p(E2)$, where E1 and E2 are mutually exclusive events.

Probabilities are defined as:

- Theoretical – some rules are used to evaluate a probability of an event.
- Experimental – probabilities are learned from data and experiments, e.g. throw dice 1000 times and measure how many times the event “getting the number 6” has happened.
- Subjective – probabilities are based on common sense human knowledge, such as defining that the probability of getting number 6” after throwing dice is 1/6th , without really throwing it at all.

A random variable *x* is characterized at any moment of time by its *uncertainty* in terms of what value this variable will take in the next moment – its *entropy*. A measure of uncertainty $h(x_i)$ can be associated with each random value x_i of a random variable *x*, and the total uncertainty H(x), called *entropy*, measures our lack of knowledge, the seeming disorder in the space of the variable x:

$$H(X) = \Sigma_{i=1,\ldots,n}\ p_i.\ h(x_i), \qquad (3)$$

Where: p_i is the probability of the variable x taking the value of x_i; $h(x_i)= \log(1/ p_i)$.

The following axioms for the entropy H(x) apply:

- monotonicity: if n > n’ are number of events (values) that a variable x can take, then
Hn(x) > Hn’(x), so the more values x can take, the greater the entropy.

- additivity : if x and y are independent random variables, then the joint entropy H(x,y), meaning H(x AND y), is equal to the sum of H(x) and H(y).

The following log function satisfies the above two axioms:

$$h(x_i)= \log(1/ p_i) \qquad (4)$$

If the log has a basis of 2, the uncertainty is measured in [bits], and if it is the natural logarithm ln, then the uncertainty is measured in [nats].

$$H(X) = \Sigma_{i=1,\ldots,n} (p_i.\ h(x_i)) = -\ c.\ \Sigma_{i=1,\ldots,n} (p_i.\ \log p_i), \qquad (5)$$

where c is a constant.

Based on the *Claude Shannon’s* measure of uncertainty – *entropy,* we can calculate an overall probability for a successful prediction for all states of a random variable x, or the predictability of the variable as a whole:

$$P(x) =2^{-H(x)} \qquad (6)$$

The max entropy is calculated when all n values of a random variable x are equiprobable, i.e. they have the same probability 1/n – a uniform probability distribution:

$$H(X) = -\ \Sigma_{i=1,\ldots,n}\ p_i.\ \log p_i <= \log n \qquad (7)$$

Joint entropy between two random variables x and y (for example, an input and an output variable in a system) is defined by the formulas:

$$H(x,y) = -\ \Sigma_{i=1,\ldots,n}\ p(x_i \text{ AND } y_j).\ \log p(x_i \text{ AND } y_j) \qquad (8)$$

$$H(x,y) <= H(x) + H(y) \qquad (9)$$

Conditional entropy, i.e. measuring the uncertainty of a variable y (output variable) after observing the value of a variable x (input variable), is defined as follows:

$$H(y \mid x) = - \Sigma_{i=1,\dots,n} \; p(x_i , y_j). \log p(y_j \mid x_i) \qquad (10)$$

$$0 <= H(y \mid x) <= H(y) \qquad (11)$$

Entropy can be used as a measure of the i*nformation* associated with a random variable *x*, its uncertainty, and its predictability.

The *mutual* entropy between two random variables, also simply called *information*, can be measured as follows:

$$I(y; x) = H(y) – H(y \mid x) \qquad (12)$$

The process of on-line information entropy evaluation is important as in a time series of events, after each event has happened, the entropy changes and its value needs to be re-evaluated.

Bayesian conditional probability is calculated using the following formula, which represents the conditional probability between two events C and A (Tamas Bayes, 18 century):

$$p(A \mid C)= p(C \mid A). p(A) / p(C) \qquad (13)$$

It follows from the equations:

$$p(A \wedge C)= p(C \wedge A)= p(A \mid C) \; p(C) = p(C \mid A) \; p(A) \qquad (14)$$

Measuring information as correlation between variables

Correlation coefficients represent the relationship between variables. For every variable x_i (i = 1, 2,..., d_1) its correlation coefficients $Corr(x_i,y_j)$ with all other variables, including output variables y_j (j = 1, 2,..., d_2), are calculated. The following is the formula to calculate the Pearson correlation between two variables x and y based on n values for each of them:

$$Corr= SUM_i \, ((x_i- Mx)(y_i-My)) / [(n - 1) \; Stdx \; . \; Stdy,] \qquad (15)$$

where: Mx and My are the mean values of the two variables x and y, and Stdx and Stdy are their respective standard deviations.

Measuring the level (the value) of information carried in a variable (its importance).

The *t-test* and the *SNR* methods evaluate how important a variable is to discriminate samples belonging to different classes. For a case of two class problem, a SNR ranking coefficient for a variable x is calculated as an absolute difference between the mean value M1x of the variable for class 1 and the mean M2x of this variable for class 2, divided to the sum of the respective standard deviations:

$$SNR\_x = abs \; (M1x- M2x) / (Std1x + Std2x) \qquad (16)$$

A similar formula is used for the t-test:

$$\text{Ttest\_x} = \text{abs}\,(M1x - M2x) / (Std1x^2/N1 + Std2x^2/N2) \qquad (17)$$

where: N1 and N2 are the numbers of samples in class 1 and class 2 respectively.

Transformation of information spaces

A set of variables measured to carry information for an evolving process form the *problem, or the information space.* These variables can be used to create another set of variables in a new information space, that retains the main information from the original problem space but potentially reduces the dimensionality of the space into a smaller set of variables. Two of the most common transformations are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Principal Component Analysis (PCA)

PCA aims at finding a representation of a problem space X defined by its variables X= {x1, x2,..., xn} into another orthogonal space having a smaller number of dimensions defined by another set of variables Z= {z1, z2,..., zm}, such that every data vector **x** from the original space is projected into a vector **z** of the new space, so that the distance between different vectors in the original space X is maximally preserved after their projection into the new space Z.

Linear Discriminant Analysis (LDA).

LDA is a transformation of classification data from the original space into a new space of LDA coefficients that has an objective function to preserve the distance between the samples using also the class label to make them more distinguishable between the classes.

1.3. Light and sound

Two type of evolving processes known as Light and Sound are of a special importance as they, first, affect how we perceive the world, and second, the way we perceive them can be used as inspiration for brain-inspired AI that deal visual and audio information (Chapters 12,13)).

Light is important electromagnetic radiation that is characterised by certain frequencies and energy. Fig.7 shows a spectrum of electromagnetic radiations with *light* being part of it.

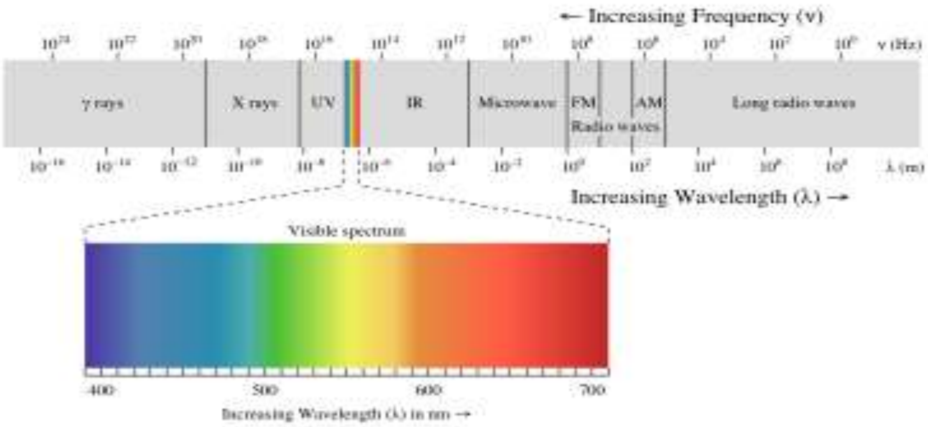


Fig. 7. The frequencies and the wavelengths of electromagnetic radiation, visible *light* being part of a small spectrum ([23])

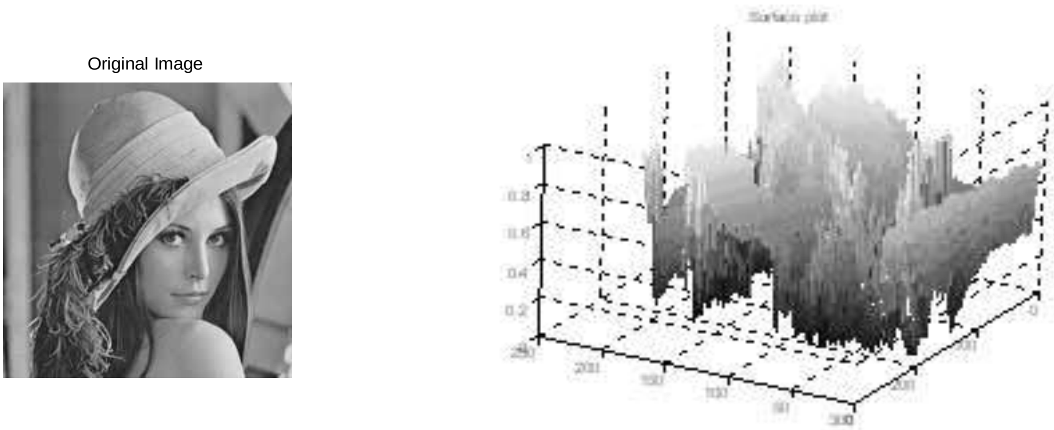
Visible light is having wavelengths in the range of 400–700 nanometres (nm), or 4.00×10^{-7} to 7.00×10^{-7} m, between the infrared (with longer wavelengths) and the ultraviolet (with shorter wavelengths). This wavelength means a frequency range of roughly 430–750 terahertz (THz). The speed of light is used as an universal constant. It is 299,792,458 meters per second.

The primary properties of visible light are: intensity, propagation direction, frequency or wavelength spectrum, polarization, energy.

Light has the properties of both:

- electromagnetic waves characterised by frequencies;
- quantum particles, called ‘photons” – that is the energy transferred from the light.

When white light illuminates an object or a face, the reflected light at different pixels may have different brightness as the reflected light has different frequencies (see Fig.8). Different brightness means different frequencies of the wave that reaches our retinas. The brighter spots activate earliest the corresponding cells and they send the first signals (spikes) to the brain. This principle is used in some SNN models described in Chapters 4, 5 and 6 as Rank Order Coding.



(a) (b)

Fig. 8. Original image (a) is represented as different intensities of brightness that represent the different frequencies of reflecting light at these pixels – (b). And this is how it activates the retina, first brighter pixels are perceived as shown on the z axis of the figure (b).

The human brain perceives visual information as a trajectory of activation of brain areas in time-space (Chapter 3). The creation of computational models for visual information processing is a subject of computer vision.

In Chapters 13 and 14 of the book SNN models are developed for both visual and audio information processing.

Sound is an oscillation under pressure, that is spread as waves in a medium. Sound waves are characterised by:

- Frequency,
- Amplitude
- Speed
- Direction

Sound that is perceptible by humans has frequencies from about 20 Hz to 20,000 Hz (see Fig. 9).

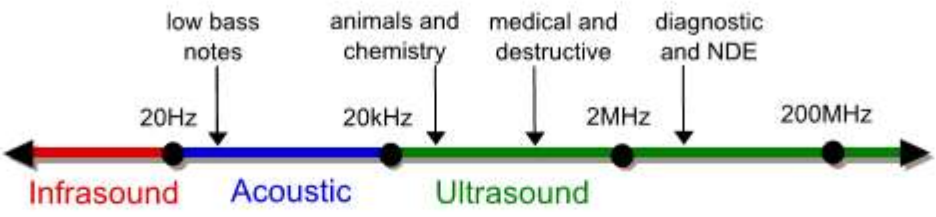


Fig. 9. Some sound frequencies with their approximate ranges for different uses (from [23])

In air, corresponding wavelengths of sound waves range from 17 m to 17 mm. Sometimes speed and direction are combined as a velocity vector, wave number and direction are combined as a wave vector and power of the signal at different frequencies over time is represented as spectrum (Fig. 10). A power spectrum represents frequencies on the signal as their power in time. Fig. 10 shows a spectrogram of the Mozart's music from fig.6, representing the power of frequencies (on the y axis) over time (on the x axis) as spectro-temporal data.

Figure 10. A spectrogram of the Mozart's music from fig.6, representing the power of frequencies (on the y axis) over time (on the x axis) as spectro-temporal data.

The way sound is perceived in the human brain is discussed in Chapter 3 and used in Chapter 14 for the creation of a BAI system for music recognition.

1.4. Evolving processes in Time-Space and Direction

Many evolving processes (in addition to light and sound as discussed above) are characterised by direction (or orientation) in which the signals or the waves spread. Examples are the spread of brain signals and the spread of seismic signals as illustrated below.

Deep learning trajectories of time-space directed connections are created during learning and recall in the brain as discussed in Chapter 3. Chapter 11 introduces a method for modelling time-space and direction on the case study of fMRI and DTI data (see also Fig.11).

Fig. 11 shows orientational information from a DTI image. Left image shows an axial slice of a single subject's DTI data, registered to structural and MNI standard space. The right image shows a close-up of the right posterior corpus callosum. Directions corresponding to each colour are as follows: Red - left to right or right to left; green - anterior to posterior or posterior to anterior and blue - superior to inferior or inferior to superior (see Chapter 11).

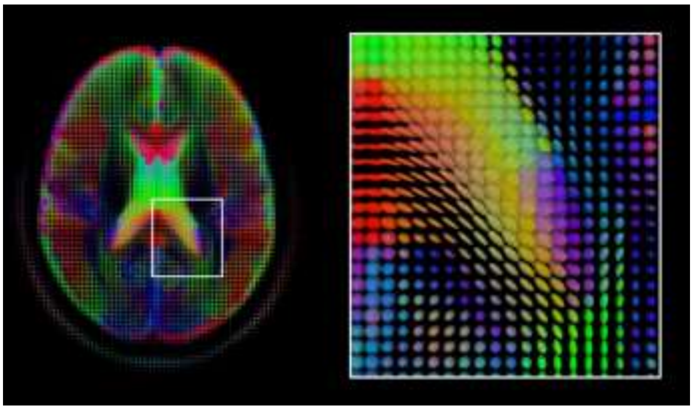


Figure 11. Orientation information from DTI image. Left image shows an axial slice of a single subject's DTI data, registered to structural and MNI standard space. The Right image shows a close-up of the right posterior corpus callosum. Directions corresponding to each colour are as follows: Red - left to right or right to left; green - anterior to posterior or posterior to anterior and blue - superior to inferior or inferior to superior [93].

Before an earthquake happens, tectonic pressure measured at one seismic centre, causes a pressure at another, etc. a chain of such reactions eventually manifested as an earthquake at a final place. Detecting time-space and direction of changes of seismic data may enable a better earthquake prediction. Fig.12 shows the map of New Zealand seismic centers and the created map of the direction of changes in these seismic data as edges of a graph developed in a SNN in Chapter 19 of the book. Time-spike learning in SNN allows for directions of changes in the data to be learned as directed connections between spiking neurons, showing which event happens first (a neuron Ni spikes) and which one follows (a neuron Nj spikes).

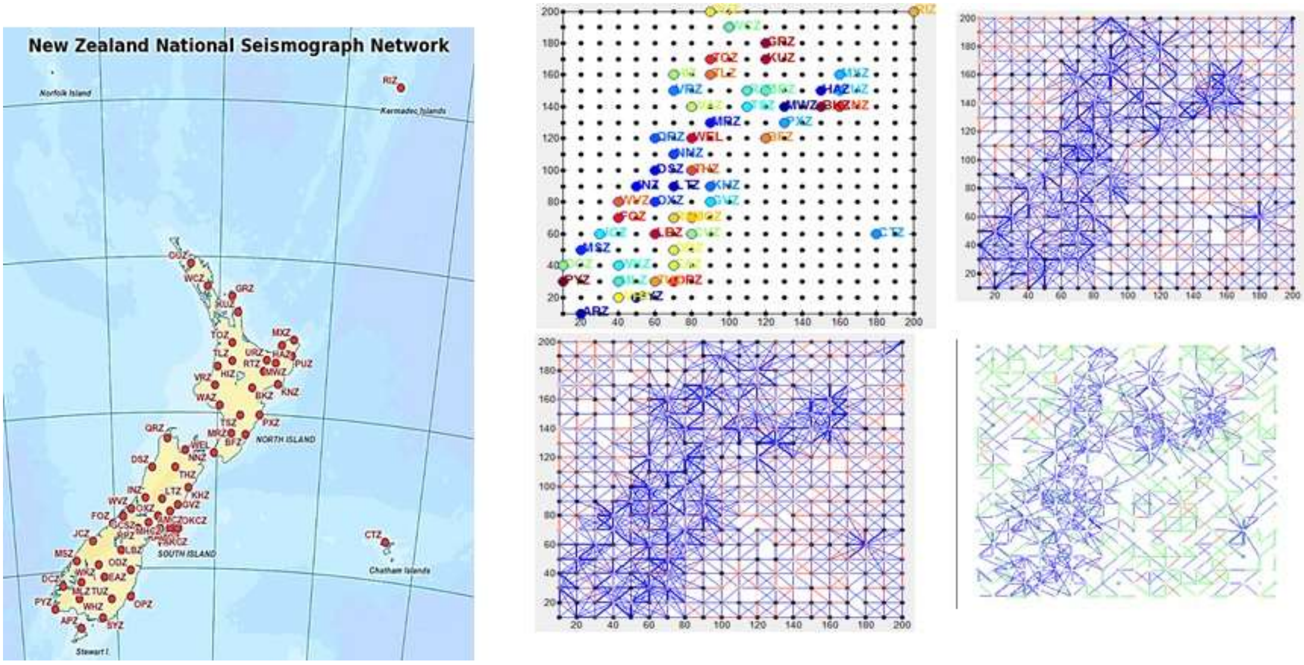


Fig.12. Before an earthquake happens, tectonic pressure measured at one seismic centre, causes a pressure at another, also measured there etc. as a chain reaction that eventually manifests as an earthquake at a final place. Detecting the direction of changes of the seismic data may enable a better earthquake prediction. Left figure shows the map of New Zealand seismic centers. Right figures show created maps of direction of seismic changes in the corresponding centres as edges of graphs representing deep knowledge as a result of deep learning in Brain-inspired SNN (Chapter 19)

Spike-time learning rules, such as STDP, to learn time-space and direction of events and changes in the data are discussed in Chapters 4,5 and 6 and applied across several applications in other chapters of the book.

Chapter 19 discusses also the detection of radio signals in space-time and direction from objects in the Universe called Pulsars. It also discusses recognition of fast moving objects in time-space and direction.

1.5. From data and information to knowledge

Generally speaking, *data* are raw entities: numbers, symbols etc., e.g., 36.

Information is labelled, understood, interpreted data, e.g., the temperature of the human body is 36 ◦C.

Knowledge is the understanding of a human, the way we do things, interpretable information in different situations, general information; e.g.:

IF the human temperature is between 36 ◦C and 37 ◦C degrees,

THEN it is most likely that human body is in a healthy state.

Some basic ways to represent data, information and knowledge of evolving processes are presented in this section, while next section discusses ways to represent deep knowledge, both acquired by humans and incorporated in a computer system**.**

The ultimate goal of information processing is the creation of knowledge. The process of knowledge acquisition from Nature is a continuous process that will never end. This knowledge is then used to understand Nature, to preserve it. From data and information, to knowledge discovery and back. This is what science is concerned with (fig.13). As shown in Fig.13, modelling evolving processes requires a sequence of procedures that involve dealing with data, information and knowledge, e.g.:

- **S**earching for *data*: Observe phenomena; collect data; store data;
- **A**nalyse data and extract *information* (e.g. pre-process data, filter, select features, visualise, label data);
- Create a model (learning, reasoning, validation)
- Extract *knowledge* (create/extract rules; reasoning with the knowledge - deductive, inductive)
- Adapt the model (accommodate new data and knowledge)

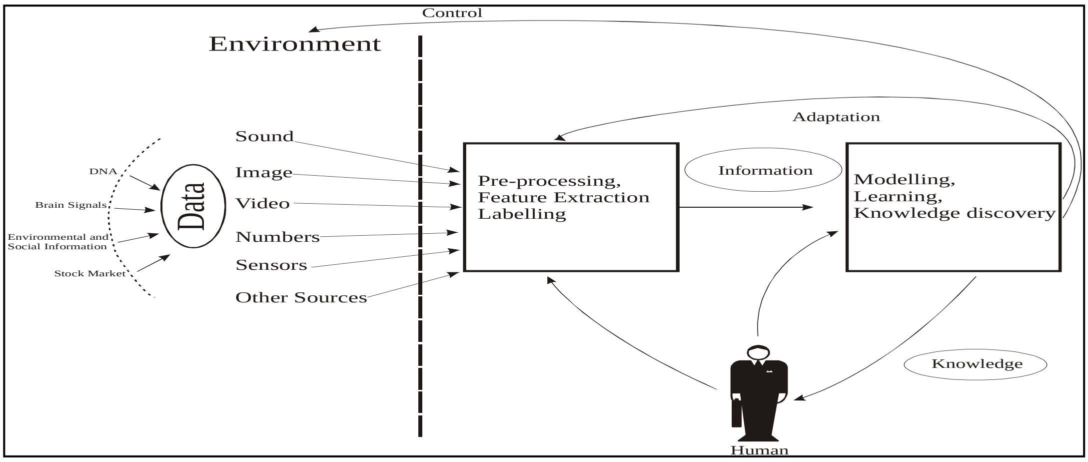


Figure 13. The flow from data and information to knowledge representation through computational modelling (after [13])

Extracting knowledge through observation of evolving processes has a long history. At the beginning, there was a school of learning that assumed that understanding of nature and its knowledge representation and articulation would not change with time. Aristotle was perhaps the most pronounced philosopher and encyclopaedist of this school.

Aristoteles (384-322 BC) was a pupil of Plato and teacher of Alexander the Great. He is credited with the earliest introduction of formal logic. Aristoteles introduced the theory of *deductive reasoning*.

Example:

All humans are mortal (i.e. IF human THEN mortal)

New fact: Socrates is a human

Deducted inference: Socrates is mortal

Aristoteles introduced *epistemology,* which is based on the study of particular phenomena which leads to the articulation of knowledge (rules, formulas) across sciences: botany, zoology, physics, astronomy, chemistry, meteorology, psychology, etc. [14, 15]. According to Aristotle this knowledge was not supposed to change. In places, Aristotle went too far in deriving 'general laws of the universe from simple observations and over-stretched the reasons and conclusions. Because he was perhaps the philosopher most respected by European thinkers during and after the Renaissance, these thinkers, along with institutions, often took Aristotle's erroneous positions, such inferior roles of women, which held back science and social progress for a long time.

Over many years after Aristotle, the logic he introduced was further developed into logic systems and rule based systems as a foundation of knowledge-based systems and AI. But this happened due to pioneers in programming analytical devices.

Perhaps the first one was the brilliant British mathematician **Ada Lovelace (1815-1852)** who is considered not only the first programmer, but the first person who demonstrated that an analytical device cannot only be used to crunch numbers, but to deal with symbols as well.

Based on symbolic representation several knowledge representation and reasoning models were developed, such as

- Relations and implications, e.g.: A-> (implies) B.
- Propositional (true/false) logic, e.g.: IF (A and B) or C THEN D.
- Boolean logic (George Boole).
- Predicate logic: PROLOG.
- Probabilistic logic: e.g. Bayes formula: p(A / C)) = p (C/A) . p(A) / p(C), where p(A/C denotes the conditional probability for an event A to happen if event C has already happened.
- Rule based systems, expert systems, e.g. MYCIN [4].

All above knowledge representations could not deal well with uncertainty of events. Human cognitive behaviour and reasoning is not always based on exact numbers and fixed rules. In 1965 **Lotfi Zadeh (1920-2018)** introduced fuzzy logic [16, 17] that represents information uncertainties and tolerance in a linguistically expressed rules. He introduced fuzzy rules, containing fuzzy propositions and fuzzy inference.

Fuzzy propositions can have truth values between true (1) and false (0), e.g. the proposition "washing time is short" is true to a degree of 0.8 if the time is 4.9 min, where *Short* is represented as a fuzzy set with its membership function – see fig.14.

Fuzzy rules can be used to represent human knowledge and reasoning, e.g.

IF washing load is small

THEN washing time is short.

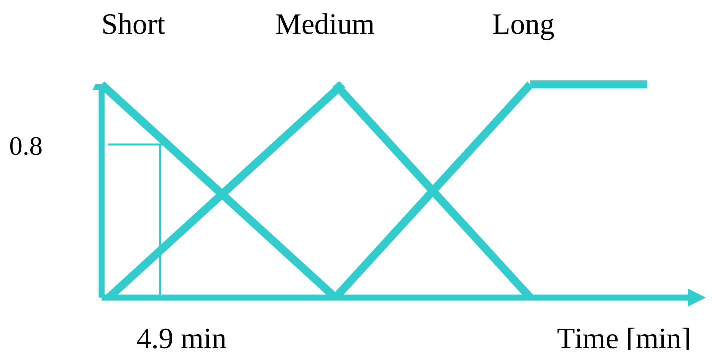


Figure 14. Fuzzy sets representing fuzzy terms of short, medium and long washing time, used to articulate and implement fuzzy rules, such as: IF Washing load is Small THEN Time of washing is Short.

Fuzzy inference systems calculate exact outputs based on input data and a set of fuzzy rules. However, fuzzy rules need to be articulated in the first instance, they need to change, adapt, evolve through learning, to reflect the way human knowledge evolves. And that is what artificial neural networks (ANN) can do as discussed in Chapter 2. In principle, logic systems and rules, while useful, could be too rigid in some cases to represent the uncertainty in the natural phenomena and some cognitive behaviour. They are often difficult to articulate, and in principle not adaptive to change.

We call the rules discussed above "flat rules", as they represent only single events represented as "flat" vectors of features and there is no time or space of series of events defined in their relationship.

1.6. Deep learning and deep knowledge representation in time-space

In contrast to the "flat rules" as discussed in the previous section, deep knowledge represents a series of events that happen in space and time in their continuous interaction.

Continuous learning of time-space data, to capture dynamically changing and informative patterns, 'hidden' deep in time and space, and to predict future events, has been a fundamental science challenge. We call this here *deep learning in time-space.* Inspired by the deep learning capabilities of the human brain, we introduce here the concept of deep knowledge in time-space. This is also related to concept formation from multimodal data.

Deep knowledge and understanding have been previously studied intensively from different aspects of the problem [38-40]. In [39] deep knowledge is defined as 'Knowledge that is concerned with underlying meanings and principles; integration of facts and feelings with previously acquired knowledge. Fundamental knowledge with general applicability, such as the laws of physics, which can be used in conjunction with other deep knowledge to link evidence and conclusions'.

Here we define deep in time-space knowledge in both brain-inspired and computational ways and that is how it is used in the rest of the book.

Let is consider a set of events E={E1, E2,..., En}. Each event Ei is defined by:

$$Ei=(Fi, Si, Ti, Pi), \quad (18)$$

defining: a function Fi; a space Si for function activity; time Ti of the activity; probability Pi.

An *event* could be a simple change in the value of a variable, or a complex cognitive process, or an earthquake, etc.

Time can be in the *past*, in the *present* or in the *future.*

Space is where an event happens.

Deep knowledge is defined here as time-space relationship between events, that can be represented in several ways.

One way to represent deep knowledge is through a relationship matrix **W**, such that the elements Wi,j of this matrix represent the relationship between events Ei and Ej in both time and space and the intensity of their interaction:

$$\mathbf{W}= \{Wi,j\}, i=1,...,n; j=1,..,n. \quad (19)$$

Another way to represent deep knowledge is through deep rules as explained below.

Events Ei and Ej for example are represented by corresponding functions Fi, Fj spatial locations Si, Sj, times Ti, Tj probabilities of the events to happen Pi, Pj and strength of the connection between the events Wi,j. All parameters of an event can be represented as *crisp* or as *fuzzy* values with corresponding membership functions (see Fig.14), e.g.:

- Location is around Si;
- Time is about Ti;
- Probability is about Pi (see about fuzzy probabilities in [38]);
- Strength is around Wi,j; or strength is High;

A hypothetical example of deep knowledge represented as a *deep fuzzy rule* is given below:

IF (event E1: function F1, location around S1, time about T1, probability about P1) (20a)
AND (strength W1,2,)
(event E2: function F2, location around S2, time about T2, probability about P2)
AND (strength W2,3,)
(event E3: function F3, location around S3, time about T3, probability about P3)
AND ...
....

(event En : function Fn, location around Sn, time about Tn, probability about Pn)
THEN (Task/event Q is executed)

The fuzzy rule above allows for the event/task Q to be recognised even if only partial match of new data is entered and the rule is applied. This is a brain-inspired principle. For example, we end up with crisp movements as a result of the activation of slightly different clusters of neurons at slightly different times in their sequence, as a reaction to crisp of fuzzy stimuli.

As a partial case, no fuzzy terms will be used, but crisp ones, e.g. the following *deep crisp rule*:

IF (event E1: function F1, location S1, time T1) *(20b)*
AND (strength W1,2,)
(event E2: function F2, location S2, time T2)
AND (strength W2,3,)
(event E3: function F3, location S3, time T3)
AND ...
....
(event En : function Fn, location Sn, time Tn)
THEN (Task/event Q is executed)

Crisp rules would be a case when activities of single neurons are measured in the brain at exact milliseconds time.

(20)

Deep knowledge is characterised by the following features:
(1) It represents informative patterns of multimodal data, deep in time (theoretically unconstrained) and in space (when dealing with spatio-temporal data);
(2) The knowledge is adaptable in an incremental, theoretically ‘life-long’ way;
(3) The knowledge is not restricted by fixed structures
(4) The knowledge is obtained in supervised-, unsupervised or semi-supervised modes
(5) The knowledge is interpretable for a better understating of the data and the processes that generated it;
(6) The knowledge can be used for early and accurate future event prediction.

Deep knowledge is what the human brain learns and manifests all the time, exemplified by:

- Listening or/and playing musical pieces;
- Playing a game;
- Visual perception;
- Predicting the movement of a predator;
- All sorts of cognition;
- Decision making;
- Consciousness;
- ...and everything else the brain does.

And the deep knowledge acquired in the human brain is manifested from hundreds of events in time and space activity of the brain, to hundreds of thousands, depending on the chosen scale to represent this knowledge, e.g. it can be represented at every 100msec or every single millisecond, at every large brain area or at every small neuronal cluster .

Some elements of deep knowledge are manifested in computational models and systems, some of them presented in the book, such as:

- Hidden Markov Models (next section);
- Deep brain EEG and fMRI patterns representing brain perception or cognitive activities (Chapters 8-11);
- Gene-regulatory networks in Bioinformatics and Neurogenetics (Chapter 17);
- Deep personalised patterns related to individual stroke prediction (Chapter 18)
- Deep climate patterns related to environmental events (Chapter 19);
- Deep geological patterns related to earthquake events (Chapter 19);
- And many other.

Illustration of deep rules as a result of deep learning are given in Chapter 3 (extracted from data measuring brain activities) and in Chapters 6,8,10, 8, along with other chapters of the book, where deep rules are extracted from a deep trained brain-inspired SNN using time-space data.

1.7. Statistical, computational modelling of evolving processes

Computational modelling of evolving processes aims at the development of mathematical and computational models that capture the essence of the dynamics of the processes and facilitate acquiring of knowledge.

1.7.1. Statistical methods for computational modelling

Here some of the most popular methods are presented, also used in other chapters of the book for a comparative analysis between their performance and the performance of new methods based on SNN.

Hidden Markov Models (HMM) are techniques for modelling the temporal structure of a time series signal, or of a sequence of events [18]. It is a probabilistic pattern matching approach which models a sequence of patterns as the output of a random process. A HMM consists of an underlying Markov chain.

$$P(q(t+1)|q(t),q(t-1),q(t-2),\ldots,q(t-n)) \approx P(q(t+1)|q(t)), \quad (20)$$

where $q(t)$ is state q sampled at a time t.

Multiple linear regression methods (MLR)

The purpose of multiple linear regression is to establish a quantitative relationship between a group of p predictor variables (X) and a response, y. This relationship is useful for:

- Understanding which predictors have the greatest effect.
- Knowing the direction of the effect (i.e., increasing x increases/decreases y).
- Using the model to predict future values of the response when only the predictors are currently known.

A linear model takes its common form of:

$$y = X A + b \quad (21)$$

where: p is the number of the predictor variables; y is an n-by-1 vector of observations; X is an n-by-p matrix of regressors; A is a p-by-1 vector of parameters; b is an n-by-1 vector of random disturbances. The solution to the problem is a vector, A' which estimates the unknown vector of parameters.

Support vector machines

This is a statistical learning technique introduced by V.Vapnik [19,20] which first transforms the data from the original space to a higher dimensional space where data belonging to different classes (outputs) can be discriminated by a hyperplane defined by a set of bordering new data points called support vectors. This is illustrated in Fig.15.

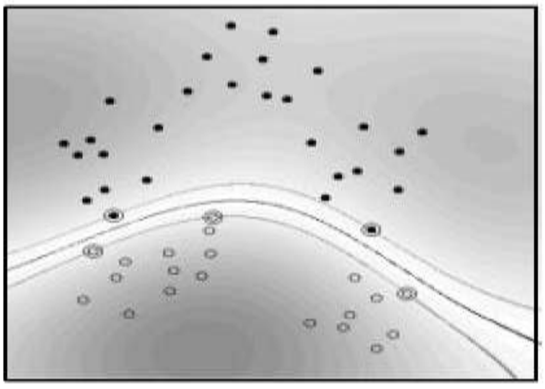


Fig.15. SVM hyperplane

Evaluating the error and accuracy of the computational models

The least squares solution is used, so that the linear regression formula approximates the data with the least root mean square error (RMSE) as follows:

$$RMSE= SQRT(SUMi=1,2,\ldots,n((y_i - y_i')\ 2\ )\ /\ n\ ) \qquad (22)$$

where: y_i is the desired value from the data set corresponding to an input vector x_i ; y_i' is the value obtained through the regression formula for the same input vector xi and n is the number of the samples (vectors) in the data set.

Another error measure is also used to evaluate the performance of a regression model – a non-dimensional error index (NDEI) – the RMSE divided to the standard deviation of the data set:.

$$NDEI = RMSE / Std \qquad (23)$$

A popular method to measure the accuracy of a computational model is the area under the curve (AUC, or also called ROC) - Fig.16, with a val7ue of 1.0 being the best and 0.5 being the worst.

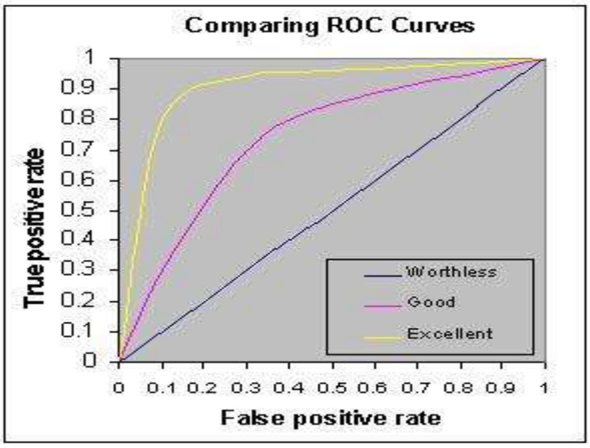


Fig.16. ROC curve is used to measure the accuracy of a computational model, with 1.0 being the best and 0.5 being the worst.

1.7.2. Global, local and transductive ("personalised") modelling [21]

Most of learning models and systems in artificial intelligence developed and implemented so far, are based on inductive inference methods, where a model (a function) is derived from data representing the problem space and this model is further applied on new data. The model is usually created without taking into account any information about a particular new data vector (test data). An error is measured to estimate how well the new data fits into the model.

The models are in most cases *global models,* covering the whole problem space. Such models are for example: regression functions; some ANN models, and also – some support-vector machine (SVM) models, depending on the kernel function they use. These models are difficult to update on new data without using old data, previously used to derive the models. Creating a global model (function) that would be valid for the whole problem space is a difficult task, and in most cases – it is not necessary to solve.

Some global models may consist of many local models, that collectively cover the whole space and can be adjusted incrementally on new data. The output for a new vector is calculated based on the activation of one or several neighbouring local models. Such systems are the evolving connectionist systems (ECOS), for example – EFuNN and DENFIS (Chapter 3).

Transductive modelling

In contrast to the inductive learning and inference methods, transductive inference methods estimate the value of a potential model (function) only in a single point of the space (the new data vector) utilizing additional information related to this point [19]. This approach seems to be more appropriate for clinical and medical applications of learning systems, where the focus is not on the model, but on the individual patient. Each individual data vector (e.g.: a patient in the medical area; a future time moment for predicting a time series; or a target day for predicting a stock index) may need an individual, local model that best fits the new data, rather then - a global model. In the latter case the new data is matched into a model without taking into account any specific information about this data.

Transductive inference is concerned with the estimation of a function in a single point of the space only. For every new input vector x_i that needs to be processed for a prognostic task, the Ni nearest neighbours, which form a sub-data set Di, are derived from an existing data set D and, if necessary, generated from an existing model M. A new model Mi is dynamically created from these samples to approximate the function in the point x_i. The system is then used to calculate the output value y_i for this input vector x_i.

A simple transductive inference method is the k-nearest neighbour method (K-NN). In the K-NN method, the output value y_i for a new vector x_i is calculated as the average of the output values of the k nearest samples from the data set Di. In the weighted K-NN method (WKNN) the output y_i is calculated based on the distance of the Ni nearest neighbour samples to xi:

$$y_i = \frac{\sum_{j=1}^{Ni} w_j y_j}{\sum_{j=1}^{Ni} w_j} \tag{24}$$

where: y_j is the output value for the sample x_j from Di and w_j are their weights measured as:

$$w_j = \frac{\max(\boldsymbol{d}) - [d_j - \min(\boldsymbol{d})]}{\max(\boldsymbol{d})} \tag{25}$$

The vector d = [d1, d2, … dNi] is defined as the distances between the new input vector x_i and Ni nearest neighbours x_j , for j = 1 to Ni; max(d) and min(d) are the maximum and minimum values in d respectively. The weights wj have the values between min(d)/max(d) and 1; the sample with the minimum distance to the new input vector has the weight value of 1, and it has the value min(d)/max(d) in case of maximum distance.

Distance is usually measured as Euclidean distance:

$$\|\mathbf{x}-\mathbf{y}\| = \left[\frac{1}{P}\sum_{j=1}^{P}|x_j - y_j|^2\right]^{\frac{1}{2}} \qquad (26)$$

Distance can be also measured as Pearson correlation distance, Hamming distance, cosine distance, etc. [20].

WWKNN: Weighted examples, weighted variables K-NN [21]

In the WKNN above the calculated output for a new input vector depends not only on the number of its neighboring vectors and their output values (class labels), as it is in the KNN method, but on the distance between these vectors and the new vector which is represented as a weight vector (W). It is assumed that all v input variables are used and the distance is measured in a v-dimensional Euclidean space with all variables having the same impact on the output variable.

But when the variables are ranked in terms of their discriminative power of class samples over the whole v-dimensional space, we can see that different variables have different importance to separate samples from different classes, therefore – a different impact on the performance of a classification model. If we measure the discriminative power of the same variables for a sub-space (local space) of the problem space, the variables may have a different ranking.

Using the ranking of the variables in terms of a discriminative power within the neighborhood of K vectors, when calculating the output for the new input vector, is the main idea behind the WWKNN algorithm [21], which includes one more weight vector to weigh the importance of the variables. The Euclidean distance dj between a new vector xi and a neighbouring one xj is calculated now as:

dj = sqr [sum l = 1 to v (ci,l (xi,l - xj,l))2] (27)

where: ci,l is the coefficient weighing variable xl for in neighbourhood of xi . It can be calculated using a Signal-to-Noise Ratio (SNR) procedure that ranks each variable across all vectors in the neighborhood set Di of Ni vectors Ci= (ci,1, ci,2,…..,ci,v)

ci,l = Sl /sum (Sl), for: l =1,2,…,v, (28)

where : Sl = abs (Ml (class 1) – Ml (class 2)) / (Stdl (class 1) + Stdl (class2)) (29)

Here Ml (class 1) and Stdl (class 1) are respectively the mean value and the standard deviation of variable xl for all vectors in Di that belong to class 1.

The new distance measure, that weighs all variables according to their importance as discriminating factors in the neighborhood area Di, is the new element in the WWKNN algorithm when compared to the WKNN.

Using the WWKNN algorithm, a “personalized” profile of the variable importance can be derived for any new input vector that represents a new piece of “personalised’ knowledge. Weighting variables in personalized models is used in the TWNFI models (Transductive Weighted Neuro-Fuzzy Inference) in [22].

There are several open problems related to transductive learning and reasoning, e.g. how to choose the optimal number of vectors in a neighbourhood and the optimal number of variables, which for different new vectors may be different [23].

1.7.3. Model Validation

When a machine learning model is built based on a data set S, it needs to be validated in terms of its generalisation ability to produce good results on new, unseen data samples. There are several ways to validate a model:

- Train-test split of data: Splitting the data set S into two sets: Str for training, and Sts for testing the model;
- N-fold cross validation (e.g. 3,5,10): in this case the data set S is split randomly into k sub-sets S1,S2, …,Sk and i=1,2,…k times a model Mi is created on a the data set S-Si and tested on the set Si; the mean accuracy across all k experiments is calculated.
- Leave-one-out cross validation (a partial case of the above method when the data set S is split N times, in each sub-set there is only one sample)

What concerns the whole task of feature selection, model creation and model validation, the above methods can be applied in two different ways:

- A “biased” way – features are selected from the whole set S using a filtering based method, and then a model is created and validated on the selected already features.
- An “un-biased” way – for every data subset Si in a cross validation procedure, first - features Fi are selected from the set S after set Si is removed from S (using some of the above discussed methods, e.g. SNR) and then – a model is created based on the feature set Fi; the model Mi is validated on Si using features Fi.

1.8. Brain-Inspired AI

Artificial Intelligence (AI) is part of the interdisciplinary information sciences area that develops and implements methods and systems that manifest cognitive behaviour [24-32].

Main features of AI are:

- learning,
- adaptation,

- generalisation,
- inductive and deductive reasoning,
- human-like communication.

Some more features are currently being developed:

- consciousness,
- self-assembly,
- self-reproduction,
- AI social networks.

Marvin Minsky (1961) articulated the term *Artificial Intelligence* as computer systems that are able to perform: search, pattern recognition, learning, planning, inductive reasoning [26].

In [41] AI is defined as computer systems that exhibit human like intelligence. It is a group of science fields and technologies concerned with creating machines that take intelligent actions based on inputs. And also in [41] AI is defined as "...advanced digital technologies that enable machines to reproduce or surpass abilities that would require intelligence if humans were to perform them. This includes technologies that enable machines to learn and adapt, to sense and interact, to reason and plan, to optimise procedures and parameters, to operate autonomously, to be creative and to extract knowledge from large amounts of data...."

There is a trend in AI called *Artificial General Intelligence (AGI)* that considers machines to become able to perform any intellectual task that humans can do.

Another trend in AI is called *Technological Singularity*. This trend argues that machines will become super intelligent that they take over from humans and develop on their own, beyond which point the human societies can collapse in their present forms, which may ultimately lead to the perish of humanity.

Stephen Hawking commented: "I believe there is no real difference between what can be achieved by a biological brain and what can be achieved by a computer. AI will be able to redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and could be superseded by AI. AI could be either the best or the worst thing ever to happen to humanity…"

A new trend in AI is the *Brain-Inspired AI (BI-AI)*, which is being developed and presented in this book. BI-AI systems use principles of deep learning in the human brain to reveal deep knowledge and to enable machines to manifest cognitive functions. BI-AI systems adopt structures and methods from the human brain to intelligently learn spatio-temporal data.

BI-AI systems have six distinctive features:

(1) They have their structures and functionality inspired by the human brain; they consist of spatially located neurons that create connections between them through deep learning in time-space by

exchanging information – spikes. They are built of spiking neural networks (SNN), as explained in Chapters 4-6 in the book.

(2) Being brain-inspired, BI-AI systems can achieve not only *deep learning*, but *deep knowledge* representation as well.

(3) They can manifest cognitive behaviour.

(4) They can be used for knowledge transfer between humans and machines as a foundation for the creation of symbiosis between humans and machines, ultimately leading to the integration of human intelligence and artificial intelligence (HI+AI) as discussed in the last chapter of the book.

(5) BI-AI systems are universal data learning machines, being superior than traditional machine learning techniques when dealing with time-space data.

(6) BI-AI systems can help us understand-, protect-, and cure the human brain.

Box 1 elaborates further on the main features above and lists 20 features of BI-AI as presented and demonstrated in various chapters of the book. Some of them are in a preliminary stage of development and more can be expected in the future.

Box 1. Twenty structural, functional and cognitive features of BI-AI systems

---

Structural Features:

1. The structure and organisation of a system follows the structure and organisation of the human brain through using a 3D brain template.
2. Input data and information is encoded and processed in the system as spikes over time.
3. A system is built of spiking neurons and connections, forming SNN.
4. A system is scalable, from hundreds to billions of neurons and trillions of connections.
5. Inputs are mapped spatially into the 3D system structure.
6. Output information is also presented as spike sequences.

Functional Features

7. A system operates in a highly parallel mode, potentially all neurons operating in parallel.
8. A system can be implemented on various computer platforms, but more efficiently on neuromorphic highly parallel platforms and on quantum computers (if available).
9. Self-organised unsupervised, supervised and semi-supervised *deep learning* is performed using brain-inspired spike-time learning rules.
10. The learned spatio-temporal patterns represent *deep knowledge*.
11. A system operates in a fast, incremental and predictive learning mode.

12. Different time scales of operation, e.g. nanoseconds, milliseconds, minutes, hours, days, millions of years (e.g. genetics), possibly in their integration.

13. A system can process multimodal data from all levels per Fig.1 (e.g. quantum; genetic; neuronal; ensembles of neurons; etc.), possibly in their integration.

Cognitive features

14. A system can communicate with humans in a natural language.

15. A system can make abstractions and discover new knowledge (e.g. rules) through self-observing its structure and functions.

16. A system can process all kinds of sensory information that is processed by the human brain, including: visual-, auditory-, sensory-, olfactory-, gustatory, if necessary in their integration.

17. A system can manifest both sub-conscious and conscious processing of stimuli.

18. A system can recognise and express emotions and consciousness.

19. Deep knowledge can be transferred between humans and machines using brain signals and other relevant information, e.g. visual, etc.

20. BI-AI systems can form societies and communicate between each other and with humans achieving a constructive symbiosis between humans and machines.

We will argue and will demonstrate in this book that BI-AI systems, if properly developed and used, can bring a tremendous technological progress across all areas of human activities and sciences and technologies, such as:

- Early disease diagnosis and disease prevention (Chapter 18);
- Affective robots for homes and for elderly (Chapters 8,14);
- Improved decision support and productivity (Chapter 20);
- Improved human intelligence and creativity (Chapters 12,13, 22);
- Improved lives and longevity (Chapter 17, 18);
- Predicting and preventing hazardous events (Chapter 19);
- And many more.

Some of the above applications are developed and illustrated in the book.

1.9. Chapter summary and further readings for deeper knowledge

This chapter discusses fundamentals of evolving processes in space and time and some of the challenges to model them and to acquire deep knowledge. All methods and concepts presented in this chapter are used in different chapters of the book as a fundamental information. More about this topic can be found in [12, 13].

With the large scale data collection across all science, technology and social areas, machine learning from data to create models and extract rules and knowledge became a necessity. This led to the establishment of artificial neural networks as major machine learning techniques that borrows some basic principle of information processing and learning from the brain (Chapter 2).

But, the human brain learns data in a *deep learning* mode and understands the evolving processes through the acquired *deep knowledge* (Chapter 3). How this could be used to create brain-inspired SNN systems is discussed in Chapters 4-7 and how SNN can be used to create BI-AI application systems is presented in Chapters 8-19. Chapters 20, 21 and 22 present some new directions of research in SNN and BI-AI.

Further recommended readings on specific topics can include:

- Aristoteles' epistemology [14, 15];
- Fuzzy logic [16, 17];
- Hidden Markov Models [18];
- Statistical Learning Theory [19,20];
- Neuro-fuzzy systems [4].

Acknowledgements

Some of the material in this chapter was first published by the author in [4, 12, 13].

References

[1] Charles Darwin (1859), On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life (1st ed.), London: John Murray, p. 502, retrieved 1 March 2011

[2] Charles Darwin (1860), On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life (2nd ed.), London: John Murray, retrieved 9 January 2009

[3] J. H. Holland (1975/1992), Adaptation in Natural and Artificial Systems. Cambridge, MA: MIT Press. Second edition (1992). (First edition, University of Michigan Press, 1975).

[4] N. K. Kasabov (1996), Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering, MIT Press, Cambridge, Massachusetts, USA, 1996.

[5] G. Carpenter, S. Grossberg, Adaptive Resonance Theory, MIT Press.

[6] D. Hebb, (1949), The Organization of Behavior. New York, John Wiley and Sons.

[7] D. E. Goldberg (1989), Genetic algorithms in search, optimization and machine learning. Addison Wesley, 1989.

[8] Abraham et al. (1993), Insect Biochem, Mol. Biol., Vol. 23, No. 8, pp. 905-912.

[9] J. Gleick (1987), Chaos Making A New Science, Viking Books, USA, 1987.

[10] O. Barndorff-Nielsen et. al. (1993), Chaos and Networks: Statistical and Probabilistic Aspects, London; New York, Chapman & Hall, 1993.

[11] F.C. Hoppensteadt (1989), Intermittent chaos, Self Organizing and Learning from Synchronous Synaptic Activity in Model Neuron Networks, Proc. Natl. Acad. Sci., USA, 1989, Vol. 86, pp. 2991-95.

[12] N. Kasabov, Springer Handbook of Bio-/Neuroinformatics, Springer, 2014.

[13] N. Kasabov (2007), Evolving Connectionist Systems: The Knowledge Engineering Approach (2nd ed.), Springer-Verlag, London, 2007.

[14] J. Ferguson (1972), Aristotle. New York: Twayne Publishers.

[15] J. De Groot (2014), Aristotle's Empiricism: Experience and Mechanics in the 4th Century BC, Parmenides Publishing, 2014, ISBN 978-1-930972-83-4.

[16] L. A. Zadeh (1965), Fuzzy Sets. Information and Control, **3** (vol.8), 1965, pp. 338-353.

[17] L. A. Zadeh (1988), Fuzzy Logic. IEEE Computer, **4** (vol.21), 1988, pp. 83-93.

[18] L. R. Rabiner (1989), A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE, vol. 77, no. 2, pp. 257 - 285, 1989.

[19] V. N. Vapnik (1998), Statistical Learning Theory. Wiley: New York, 1998.

[20] V. Cherkassky and F. Mulier (1998), Learning from data: Concepts, theory, and methods. New York: Wiley, 1998.

[21] N. Kasabov (2006), Adaptation and Interaction in Dynamical Systems: Modelling and Rule Discovery Through Evolving Connectionist Systems, Applied Soft Computing, 6, 3, 2006, 307-322.

[22] Q. Song and N. Kasabov (2005), NFI: A Neuro-Fuzzy Inference Method for Transductive Reasoning, IEEE Trans. on Fuzzy Systems, 13, 6, 2005, 799-808.

[23] Wikipedia, http://www.wikipedia.org.

[24] Daniel Dennett: *From Bacteria to Bach and Back. The Evolution of Minds.* W. W. Norton & Co. 2017.

[25] Stuart Russell and Peter Norvig *Artificial Intelligence: A Modern Approach* (3rd ed, Upper Saddle River, NJ, Prentice Hall, 2009).

[26] Nils J Nilsson *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge, UK, Cambridge University Press, 2010).

[27] Pamela McCorduck *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (2nd ed, Natick, MA, 2004).

[28] Nick Bostrom and Vincent C Millar “Future Progress in Artificial Intelligence: A survey of expert opinion” in Vincent C Muller (ed) *Fundamental Issues of Artificial Intelligence* (Synthese Library, Berlin, Springer, 2016) at 553.

[29] John H. Andreae (ed.) *Man-Machine Studies*, ISSN 0110 1188, nos. UC-DSE/4 and 5. 1974.

[30] John H. Andreae: *Thinking with the Teachable Machine*, Academic Press, 1977.

[31] Richard S. Sutton & Andrew G. Barto, *Reinforcement Learning*, MIT Press, 1998.

[32] Andy Clark, *Surfing Uncertainty*, Oxford University Press, 2016.

[33] Stephen Hawking, A brief history of time, Bantam Books, NY, 1990.

[34] Roger Penrose, The Emperors New Mind, Oxford University Press, 1989.

[35] Richard Dawkins, The selfish gene, Oxford University Press, 1989.

[36] Nikola Kasabov (ed) Spring Handbook of Bio-/Neuroinformatics, Springer, 2014.

[37] Albert Einstein, Ideas and Opinions, Wings Books, NY, 1954.

[38] B.Chandrasekaran, Sanjay Mittal, Deep versus compiled knowledge approaches to diagnostic problem-solving, International Journal of Man-Machine Studies, Vol.19, No. 5, Nov 1983, 425-436, https://doi.org/10.1016/S0020-7373(83)80064-9.

[39] Patricia L. Rogers, Gary A. Berg, Judith V. Boettcher, Caroline Howard, Lorraine Justice, and Karen D. Schenk, Encyclopedia of Distance Learning, (4 volumes), January, IGI Global, 2009,

[40] Alex Bennet and David Bennet, Knowledge-Based Development for Cities and Societies: Integrated Multi-Level Approaches, 2010, IGI Global, DOI: 10.4018/978-1-61520-721-3.ch009.

[41] The New Zealand AI Forum: http://aiforum.org.nz