Discrete and Continuous Aspects of Nature Inspired Methods

Martin Macaš, Miroslav Burša, Lenka Lhotská

Gerstner Laboratory, Czech Technical University in Prague
Technicka 2, 166 27 Prague 6, Czech Republic
mmacas@seznam.cz, bursam@fel.cvut.cz, lhotska@fel.cvut.cz

Abstract. In nature, industry, medicine, social environment, simply everywhere we find a lot of data that bear certain information. A dictionary defines data as facts or figures from which conclusions may be drawn. Data can be classified as either numeric or nonnumeric. The structure and nature of data greatly affects the choice of analysis method. Under the term structure we understand the facts that the data might be not a single number but n-tuples of measurements. Structure is also very closely linked to the reason of data collection and method of measurement. The discussion is focused on the similarities and differences of nature inspired methods and their natural counterparts in light of continuous and discrete properties. Different examples of nature inspired methods will be inspected in terms of data, problem domains and inner structure and principles.

Data can be classified as either numeric or nonnumeric. With respect to this classification we can describe data in the following way. Qualitative data are nonnumeric and usually represent some descriptive features (e.g. colours, type of material, subjective view – poor, fair, good, better, best). Qualitative data are often termed categorical data. Some literature sources use the terms individual and variable to reference the objects and characteristics described by a set of data. They also stress the importance of exact definitions of these variables, including what units they are recorded in. The reason the data were collected is also important. Quantitative data are numeric. They are further classified as either discrete or continuous. Discrete data are numeric data that have a finite number of possible values. They are represented by integer or whole numbers. When data represent counts, they are discrete. Counts are usually considered exact and integer. Continuous data have infinite possibilities and are represented by real numbers.

The structure and nature of data greatly affects the choice of analysis method. Under the term structure we understand the facts that the data might be not a single number but n-tuples of measurements, e.g. pressure, temperature, humidity at a certain time point. Structure is also very closely linked to the reason of data collection and method of measurement.

Data are collected by mapping entities in the domain of interest to symbolic representation by means of some measurement procedure, which associates the value of a variable with a given property of an entity. The relationships between objects are represented by numerical relationships between variables. Obviously the measurement process is crucial. It underlies all subsequent data analytic and data mining activities. Let us briefly touch the issue of measurement. The experimental method depends on physically measuring things. The concept of measurement is related to the concepts of numbers and units of measurement. Statisticians categorize measurements according to levels, namely nominal, ordinal, interval, and ratio. There exist many taxonomies for measurement scales. Sometimes they are based not on the abstract mathematical properties of the scales but rather on the sorts of data analytic techniques used to manipulate them. Examples of such alternatives include counts versus measurements; nominal, ordinal, and numerical scales; qualitative versus quantitative measurements; metrical versus categorical measurements; and grades, ranks, counted fractions, counts, amounts, and balances. In most cases it is clear what is intended by these terms. Ranks, for example, correspond to an operational assignment of integers to the particular entities in a given collection on the basis of the relative „size“ of the property in question: the ranks are integers which preserve the order property.

Transforming Data. Sometimes raw data are not the most convenient form and it can be advantageous to modify them prior to analysis. There is a duality between the form of the model and the nature of the data. Certain transformations of the data may lead to the discovery of structures that were not at all obvious in the original scale.

Nature inspired methods. In the discussion we will briefly describe the basic properties of neural networks, binary and discrete PSO, genetic algorithms and ant systems with respect to different data types.

The main task of natural neural network is to process and transfer information included in continuous signal. The information is imparted between neurons by the help of continuous synaptic potentials. However, for example receptors can represent a source of binary information. The inner structure of both the natural and artificial neuron is continuous with some discrete features (thresholds and activation functions). On the other hand, there are both the discrete and continuous types of tasks solved by artificial neural networks. For example, while the function approximation is continuous, the classification task could be considered as discrete. The artificial neural networks can handle arbitrary data types.

The natural evolution is continual process from all points of view except the data representation by a ternary encoding. While the goal of Darwinian evolution is to evolve a well-adapted organism, the genetic algorithms can perform on combinatorial problems.

The main goal of real creatures (e.g. birds and fish) in the nature is often to move in physical space and find some source of food; the behaviour, which enables these creatures to follow this goal, results from complex mixture of knowledge and random elements. The PSO method working in real values is probably more suitable for problems from continuous domain; however using a special representation can enable the PSO to solve discrete or binary problems.

Of course, the process of making a decision is discrete. However the mechanism of choice of some decision is a complex problem, which is far from being purely discrete. The inner structure of binary PSO inspired by these processes has continuous probabilistic character with certain thresholds.

In ant system modelling, the goal is to find a robust and efficient method to solve a problem and not to model the natural process as a whole. In the cruel nature the main goal of an ant system is to survive. This is accomplished by positive feedback and stigmergic communication. The main tasks performed (and modelled) are: food search, food transport and item clustering (brood sorting, item clustering).