Physically Inspired Artificial Learning Models

Dymitr Ruta* and Bogdan Gabrys†

1 Introduction

Information is something that can only be encoded in the state of a physical system, thus the quest for better ways to acquire, store, transmit, and process information leads us to seek more powerful methods for understanding and controlling the physical world. Limitations inherent in the physical world (such as the size of atoms, the quantum effects or speed of the light) pose limitations in our explorability and perception of the world and likewise set the boundaries on the acquiring and processing of information. There are further indications that the links between the information and the physical world are much more fundamental. It is even claimed that the reality of material world arises at the very bottom from the elementary yes-no questions capturing single bits of information [1]. The comprehensive success of computing seems to appreciate this concept of it from bit and only now reveals the vastness of complexity of the physical world. Despite the natural familiarity with information, there is still no uniform definition covering all its aspects. Newly emerged disciplines of machine learning, artificial intelligence and data mining showed the strength of numeric data-type information processing and highlighted its benefits for knowledge discovery and predictive forecasting. Other forms of information including symbolic, textual and multi-medial are gaining a massive momentum on the exponentially expanding internet, yet they lack a mature unified information processing apparatus and hence apart from very targeted intelligent access methodologies this rich multi-form information source remains virtually unexploited for delivering explanatory, discovery and predictive information intelligence to the society.

2 The physics of information, complexity and learning

We believe there are 3 key concepts that require coordinated formulation and addressing in order to devise a unified information processing and learning framework and successfully harness it to the benefit of future smart information systems. These are information uncertainty, complexity and learning. Uncertainty is becoming to be seen as the main interchangeable facet of information often phrased as capacity to obtain information, by the direct analogy with physical energy as a capacity for doing work. This analogy was further enhanced with the recently revealed multidimensional nature of information uncertainty beyond traditionally perceived statistical ambiguity seen as an uncertainty over a choice from many alternatives. Vagueness, non-specificity or fuzziness perceived as inability to make sharp distinctions supplement the perception of uncertainty from the perspective of Fuzzy Sets Theory (FST) and General Theory of Evidence (GTE) [2] and provide further resemblance to the physical energy which can take many different forms like thermal, kinetic, gravity, electric etc. The newly introduced Generalized Theory of Uncertainty (GTU) [3] based on core perception of information as a constraint on the values that a variable is allowed to take appears to be the cornerstone of emerging unified information/uncertainty framework. The new generalised interpretation of information seen as a certain fixation or a constraint on the variable value space consistently follows the analogy of physical matter - de facto certain embodiment of energy - seen as a constraint (or as recently known as curvature) of the time-space. Further research into modelling and processing of information particularly imprecise and perceived is needed to complement a mature precise information analysis and further physical inspirations are anticipated to guide this process ultimately leading to the hypothetical law of the total preservation of information uncertainty in the isolated system built upon its physical counterpart of energy preservation law.

Information complexity is another concept that is defined in many different ways depending on the application. From the simplest definitions attributed simply to the amount of data and dimensionality through the amount of resources (time, operations etc) needed to complete the task up to the data structure itself, complexity is seen as a degree or a level of difficulty in applying different formalisms, methodologies, modelling and processing in order to achieve certain tasks [1]. Tasks in the information theory domain are nothing but obtaining specific information, hence generalized complexity can be seen as a certain measure of the cost of obtaining information. Kolmogorov proposed an algorithmic interpretation of information complexity seen as a shortest program code that can obtain the requested information [4]. Further, it has been shown that algorithmic complexity sets limits on the thermodynamic cost of computations, which in turn led to a definition of the generalised measure of information distance between 2 strings [5]. Distance measure is a key characteristic that on the one hand formally constructs the space of objective analysis and on the other hand defines spatial relationships among the objects of analysis, hence allowing to apply a number of well-defined physical models capturing spatial interaction of particles. Another physically inspired model of complexity is the thermodynamic depth of a system which tries to find

*Intelligence Systems Research Centre, British Telecom Group, Research & Venturing, Orion Building 1st floor, pp12, Adastral Park, Martlesham Heath, Ipswich IP5 3RE, UK, dymitr.ruta@bt.com

†Computational Intelligence Research Group, Bournemouth University, School of Design, Engineering & Computing, Poole House, Talbot Campus, Fern Barrow Poole BH12 5BB, United Kingdom, bgabrys@bournemouth.ac.uk

out how much information its final state contains about the trajectory leading to that state, and thereby is particularly suitable to directly describe and control the complexity of the learning process in the smart information systems.

In any intelligent information systems a core process defining its interaction with surrounding information space is the process of learning. From the information-theoretic point of view learning is a process of gaining knowledge about how to obtain required information given certain input evidence in an automated repetitive manner. More precisely it is a process of extracting the knowledge carried in the data samples and transfering it onto the model parameters to perform repeatable information transformation. In practise the process of learning is typically an optimisation problem in which the goal is to find the model parameters such that the error between model extracted information and the target information is minimum possible. There is a wide range of optimisation methods operating on continuous numerical data, yet there is virtually a lack of optimisation on non-numerical information space and on top of that it is difficult to evaluate and compare different methods due to multitude of optimisation criteria. We believe that the solution to these problems lies in further extension of recent advances on applying mutual information (MI) criterion of optimisation that tries to maximise the amount of information content by which outputs' uncertainty is reduced after observation of inputs [6]. New information thermodynamics models and further generalisation into categorical and possibly textual/string data is needed to allow for multiform information exploitation in the optimisation process that could be facilitated by massive global internet resources. On the other hand using generalised information representation as a string of symbols, the minimised difference of the conditional Kolmogorov complexities could become the prime criterion in future optimisation methods. For both MI maximisation and information distance minimisation there is a tremendous equivalence found in physical world. It has been shown that the nature of mutual information maximisation lies in allowing the data to move according to information forces generated by the information potentials built upon the interaction between inputs and outputs.

3 The new breed of artificial learning models

Designing the learning process on grounds of physical interactions among the particles is one of the most fruitful inspirations of physical phenomena for the machine learning domain. It is consciously or unconsciously being exploited in well known kernel machines methodology and very popular data density estimation methods. What is surprising however is the scarcity of physically inspired models in classification, clustering and regression methods which are becoming increasingly popular in the upcoming data mining age of business analytics. Authors' pioneering efforts into devising new classification, clustering and data condensation algorithms based on various potential fields found in nature confirm there is a huge potential for improving the performance of current predictive models [7]. Moreover, contrary to some beliefs, the simulation style optimisations run on large datasets according to simple fixed rules actually pose extremely small amount of information complexity according to the algorithmic Kolmogorov complexity measure. By analogy, our planetary system is a naively simple system once the Newton's and Keppler's laws are known, although for many it was and possibly still is an insolvably complex system otherwise. In physically inspired optimisation process the knowledge brought in by a physical equivalence transforms the space of the optimisation in such a way that the process of learning itself is naively simple. This is equivalent to the gravitational model, which for many is regarded as extremely simple model of free move in the time-space that is curved by the presence of mass. It is our well-justified belief that even further gain in predictive performance can be obtained by designing physically-inspired optimisation process run on large populations of data which can interact with each other during the learning process and thereby combine the supervised and unsupervised elements of learning. In such hybrid systems physically-inspired global models mainly defining and constraining the space of information, and rules of its interaction with the data will be combined with the biologically-inspired population or network of data-agents that are empowered to interact among each other during the optimisation process. Such systems should reflect the perfect composition of the knowledge that civilisation laboriously accumulated over thousands of years about both the alive and nonliving nature and are predicted to outperform current machine learning techniques in many application areas.

Based on the above discussion we would therefore like an opportunity to form a task force during the Brainstorming Meeting in Mallorca and further present our vision of various yet unrealised but emerging physically and biologically inspired artificial learning models that can be harnessed to the greater benefit of the businesses and society in the upcoming data mining age of the business analytics.

References

[1] W.H. Zurek: Complexity, Entropy and the Physics of Information. Proc. of the Workshop on Complexity, Entropy, and the Physics of Information. Santa Fe (1989).

[2] G.J. Klir, T.A. Folger: Fuzzy Sets, Uncertainty, and Information. Prentice-Hall International Edition (1988)

[3] L.A. Zadeh: Toward a Generalized Theory of Uncertainty (GTU)-An Outline, to appear in Information Sciences

[4] A.N. Kolmogorov: An Introduction to Kolmogorov Complexity and Its Applications. Springer-Verlag, New York (1997)

[5] C.H. Bennett, P. Gacs, M. Li, P.M.B. Vitanyi, W.H. Zurek: Information Distance, IEEE Trans. Information Theory 44:4 (1998), 1407-1423

[6] J. Principe, I. Fisher, D. Xu: Information Theoretic Learning. In S. Haykin (Ed.): Unsupervised Adaptive Filtering. New York NY (2000)

[7] D. Ruta, B. Gabrys: Nature Inspired Learning Models. Proc. European Symposium on Nature Inspired Smart Information Systems, Albufeira, Portugal (2005)