New Trends of Soft Computing Methods for Industrial and Biological Processes

Bernardo Penna Resende de Carvalho and Leonardo Carneiro de Araújo

Abstract This work presents real world examples of using different Soft Computing methods in both industrial and biological processes from the years 2005 to 2009. Multi-Objective Algorithm, Least Squares Support Vector Machine and Fuzzy Inference were applied in steel industry processes, while Decision Tree, Recursive Feature Elimination and Genetic Programming were evaluated in biological processes. Soft Computing methods were capable to predict quantities, recognize patterns and select relevant attributes in order to improve each process. This paper shows the growing development on Soft Computing and the integration of process knowledge points to a direction of increasing possibilities to achieve better performances in industrial and biological processes.

1 Introduction

It's widely known that industries are an essential component of sustainable development of the economy. On the other hand, industries often deal with a degradative process which can result in pollution and exhaust of natural resources. Biological processes play an important role in the society nowadays for the development of innovative treatments and medicines, but the development of drugs is still a slow process which demands thousands of biological tests on different organisms, which leads to the death of many animals in order to understand the action mechanism of proteins and other molecules.

At those points of opposing interests rises the necessity to optimize the processes in order to achieve the best usage of the available resources, which can be done using state-of-art technologies available. In the last two decades, advances have been made

Bernardo Penna Resende de Carvalho
Optimization Department, ABG Information Systems, Brazil, e-mail: bpenna@gmail.com

Leonardo Carneiro de Araújo
CEFALA, Federal University of Minas Gerais, Brazil e-mail: leolca@gmail.com

in the technologies of Soft Computing. Beyond classical methods [8, 9, 10, 11], many techniques were recently developed [7, 12, 13, 14] and are proven to give good results in the fields of pattern recognition [1, 4], synchronization [5, 6] and prediction [2, 3] of process variables.

Steel industries need not only to produce their products with the required quality, but also to ensure that their production process is performed using the best practices with ecological sustainability. This includes the higher final product quality, the higher production speed and also the lower and efficient energy and raw-materials consumption.

Soft Computing methods are used in biological industries in order to help the development of new drugs and treatments, as well as reducing the market entrance time of a new medicine or even making possible some processes previously impracticable. This can be possible because *in silico* tests are much faster to be carried through than *in vivo* ones.

This work aims at giving real world examples of using different Soft Computing methods in both industrial and biological processes developed during the years 2005 to 2009. Multi-Objective, Least Squares Support Vector Machine and Fuzzy Inference were applied in steel industry processes (Sect. 2), while Decision Tree, Recursive Feature Elimination and Genetic Programming were evaluated in biological processes (Sect. 3). It can be observed in Sect. 4 that the usage of these technologies on real industrial and biological cases made possible to predict quantities, recognize patterns and select relevant attributes in order to improve each process.

2 Applications of Soft Computing on Steel Industrial Processes

Soft Computing methods provide the capability to generate knowledge directly from raw industrial data, which improved the know-how and the productivity of industrial processes. In this section we present steel industry cases where Multi-Objective Algorithm, Least Squares Support Vector Machine and Fuzzy Inference were successful implemented in optimization systems from 2005 to 2007.

2.1 Multi-Objective Algorithm

Multi-Objective Algorithm (MOBJ) [13] is a learning scheme for improving generalization of MLPs (Multilayer Perceptrons) [8], which are based on a connectionist approach to computation composed by building blocks, the neurons, interconnected as a network. Each neuron of MLP is a computing unit that weights its input data and provides an output according to its transfer function. MOBJ minimizes both sum of squared error and norm of network weight vectors to obtain the Pareto-optimal solutions, which are not unique. After the usage of a validation set, the final solution is expected to balance network variance and bias. Since the training parameters pro-

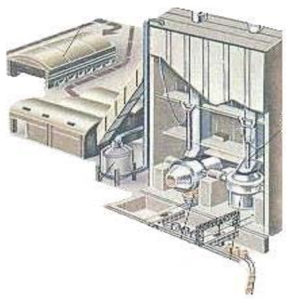

Fig. 1 Blast furnace and converter

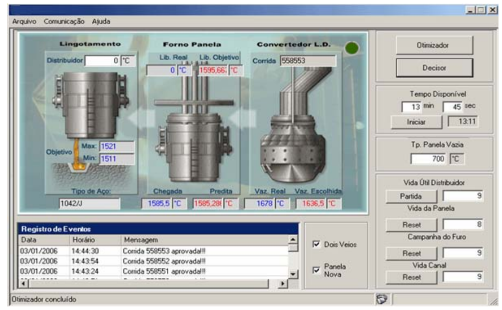

Fig. 2 MOBJ based system

duce minor effects on the final solution, tuning the best set of training parameters is an easy task. MOBJ's result is the best MLP generated with the Pareto-optimal solutions, where the criterion is the validation accuracy.

The steel processing depends on the synchronization of temperature in both the converter and the blast furnace, presented on Fig. 1, in order to achieve a desired temperature at the continuous casting. The system presented in Fig. 2 was developed to control temperature in the steel process of the blast furnace. It helps the operator to take decisions, being able to predict and track the thermal behavior of the steel during its processing inside the converter, blast furnace and continuous casting. This system, which is based on 14 variables of the process, is able to predict the thermal losses from the converter to the blast furnace. It got an accuracy of 78%, when considering an error as a difference greater than 150 degrees between the real and the prediction temperature [6]. The consequences of the system were an increase in the productivity of the steelworker and a lower raw-material consumption of the hole process.

2.2 Least Squares Support Vector Machine

Least Squares Support Vector Machine (LS-SVM) [12] is a learning machine that corresponds to a modified version of Support Vector Machine (SVM) [14]. This new SVM formulation consists of using equality constraints at the primal cost function to be minimized, instead of inequality ones. Moreover, the 2-norm is used for the slack variables of the primal problem, instead of the 1-norm. As a result of these modifications, the problem generated by LS-SVM can be solved with a system of linear equations, which is less complex than the quadratic programming used in SVM. LS-SVM is a way to deal efficiently with outliers. It uses the principle of structural risk minimization, which results in a high capacity of generalization, even if the training set is not very representative. A LS-SVM drawback is the parameter tuning difficulty during the training phase.

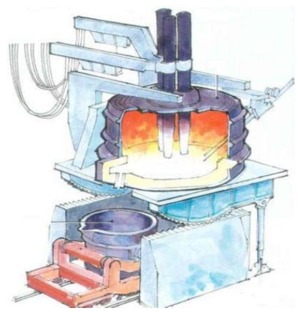

Fig. 3 Electric arc furnace

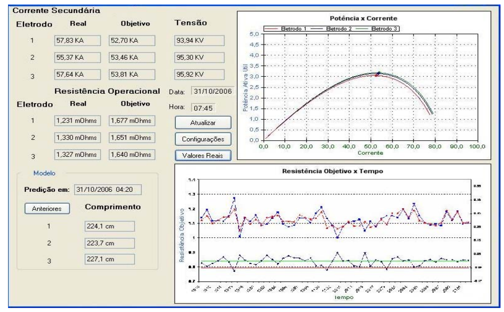

Fig. 4 LS-SVM based system

Electric arc furnace is an equipment often used for steel making with three graphite electrodes (Fig. 3). One of the most laborious and also very dangerous action in steel industry is the measurement of the electrode length. A system developed to estimate the three electrode lengths of an electric arc furnace is presented in Fig. 4. This system was capable to estimate each electrode length during the time, decreasing the frequency that the furnace has to be opened, avoiding damage on it and also increasing its productivity. It got an accuracy higher than 70% in the prediction of three electrode lengths, with a hit been considered when there is a difference lower than 15cm between the prediction and the real value [2].

2.3 *Fuzzy Inference*

Fuzzy Inference [9] is a method which formulates a mapping from a given input to an output using fuzzy logic instead of using sharp switching between values, dealing with the concept of approximate rather than precise. In the traditional view of sets, there are crispy limits for each set, making it possible to determine precisely either an element belongs to the set or not. In fuzzy sets, the concept of possession is rather diffuse, which implies in membership values that states the degree a certain element belongs to a set or another. Just like the fuzzy sets, in fuzzy logic a statement has a degree of truth that ranges from 0 to 1, not being constrained to only two values as in classic propositional logic. Fuzzy Inference uses the provided input information to derive inferred conclusions based on fuzzy logic and than use this result to guide the decision making using a set of if-then fuzzy rules.

A continuous casting process in steel industry is presented in Fig. 5. It's very important to control the synchronization of this process in order to allow the correct usage of the next equipments. A fuzzy system (Fig. 6) was implemented in this process phase, where fuzzy rules were used in order to estimate the production status of a casting process: on time, delayed or advanced. These rules were used to synchronize the casting phase and resulted in an advance of two runs a day and an economy of electricity consumption [5].

Fig. 5 Continuous casting process

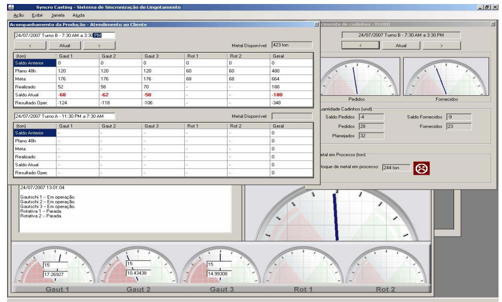

Fig. 6 Fuzzy based system

3 Applications of Soft Computing on Biological Processes

Biological and pharmaceutical industries have a straight relationship with computing and molecular biology, which aim at analyzing large amounts of biological data, predicting gene functions or demonstrating the relationships between genes and proteins in order to improve medicine development. In this section it's presented the results of Decision Tree, Recursive Feature Elimination and Genetic Programming applied into biological processes from 2006 to 2009.

3.1 Decision Tree

Decision Tree is a method [11] which uses symbolic relationships, in the form of a hierarchical tree structure, in order to classify samples based on a series of rules about their attributes. It classifies samples from the root to the leaves, which corresponds to a specific class. Each node of Decision Tree performs an attribute test for the sample, in a manner that each alternative way corresponds to a possible value for that variable. The process of walking through the tree is repeated until it reaches a leaf, where the pattern is stated for the correspondent class. It does not require any knowledge or parameter setting. The attributes of the classes can be any type of variables from binary, nominal, ordinal, and quantitative values, while the classes must be qualitative type (categorical or binary, or ordinal).

Glands play a very important role in the overall control function of organism bodies (Fig. 7). It was developed a system based on Decision Tree in order to classify samples of glands from various species based on their levels of genomic expression existent in a database of glands and non-glands. The system based on Decision Tree presented in Fig. 8 was capable to evaluate a classification between glands and non-glands with 100% accuracy using information of only 2% of the complete database of more than 4500 expression levels [1].

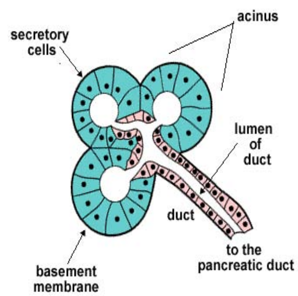

Fig. 7 A gland structure

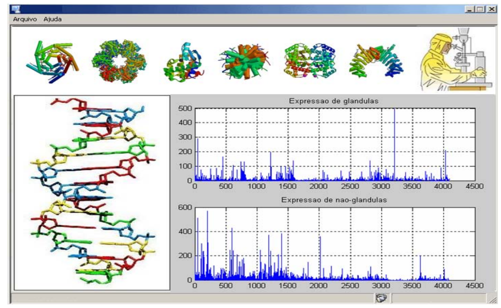

Fig. 8 Decision tree based system

3.2 Recursive Feature Elimination

Recursive Feature Elimination (RFE) [7] is a feature selection technique that uses SVM [14] to detect the most relevant features of a given database. RFE selects a subset of relevant features by using a ranking criterion that computes the variation in an objective function upon removing each feature in a SVM classifying task. In order to improve the efficiency of SVM training, this objective function is represented by a cost function for each feature and it's computed by using the training set only. When a feature is removed or its correspondent weight is reduced to zero that attribute is considered as irrelevant, because it has not any influence in building the separation hyperplane. The great advantage of RFE is the minimization of the original search space of the problem into a significantly simpler space, what makes easier the classification task.

Synapses are spaces filled with a fluid, between the neurons, which separate the emitting and receiving cells. The nervous signal is transmitted from the pre-synaptic to the post-synaptic neuron, through the synapse (Fig. 9). Some types of proteins can be found in the post-synaptic neurons, being important to the transmission of information. More interesting than only determining if a given protein possess post-synaptic activity or not, it is important to discover which are the most relevant fea-

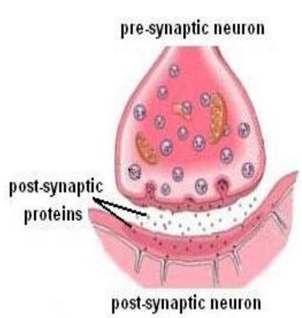

Fig. 9 A synapse structure

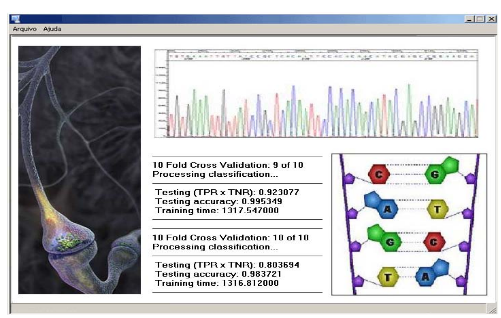

Fig. 10 RFE based system

tures related to this task, which is done by the developed system (Fig. 10). It uses RFE in order to select only the relevant features from the complete protein database. Once the reduced subset was found, it was used to predict its classes. The results show that the system led to a reduced representation to the database, using only 6% of the original information, and yielded a classification accuracy of 96% [3].

3.3 Genetic Programming

Genetic Programming (GP) [10] is an evolutionary algorithm inspired by biological evolution. It aims at finding computer programs, represented by trees, that perform a given task. The evolution process is performed according to a fitness function, with new individuals being created in each generation. Every tree node (individual) has an operator function and every terminal node has an operand, making mathematical expressions easy to evolve and evaluate. There are two operators that simulate the evolutionary process: mutation and crossover. Mutation affects only one individual in the population, by replacing a whole node in the selected individual or just one node's information. Crossover is applied by simply switching one of individual nodes with another node from another individual in the population. With a tree-based representation, replacing a node means replacing the whole branch.

Some toxins are able to increase the permeation of ions through the cellular membranes, by binding to ion channels (Fig. 11). Ion channels are proteins normally involved with the functioning of the nervous system of many organisms. Fig. 12 presents a system that can store raw data from protein purification and characterization experiments and uses GP to discover novel patterns within the primary and secondary structures of a set of toxins. The system discovered patterns made possible to differentiate these toxins by their function: binding to specific channels for sodium, calcium or potassium ions. The experiments were performed using 802 toxin primary sequences labeled as channel functions. The system got an average classification accuracy of 80% for scorpion and spider toxins, with correctness of 97%, 67% and 55% respectively to sodium, potassium and calcium channels [4].

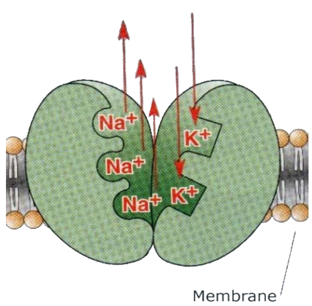

Fig. 11 An ion channel structure

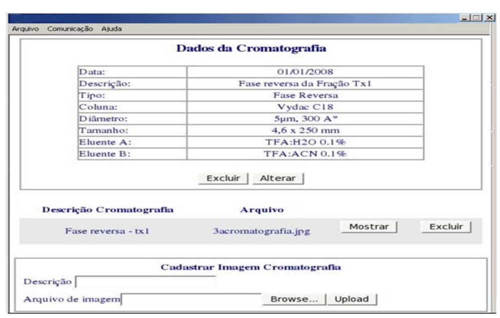

Fig. 12 GP based system

4 Conclusion

Several applications of Soft Computing in both industrial and biological processes were presented in this work. All systems presented were developed, implemented and tested by the authors and colleagues. The study on Soft Computing methods and optimization techniques is a wide research area still under development. The main point of this paper is to show the growing possibility and stimulate the usage of Soft Computing to solve real world problems.

Every process has its own peculiarities and there is no rule of thumb on how to optimize it. The development of a new solution requires great knowledge of the process, what is better achieved with a collaborative work of the process specialists and the specialists in Soft Computing. This work shows successfull cases where this kind of collaboration occurred.

References

1. Carvalho B P R (2006) Classification of Gland Cells Using Decision Tree into a KOG Database. ABG Systems Technical Report. Belo Horizonte, Brazil
2. Carvalho B P R, Carvalho D H D, Yovanovic A P (2006) Intelligent System for Prediction of Electrode Length in an Electric Arc Furnace. In: XVI Automatic Brazilian Congress, CBA 2006. Salvador, Brazil
3. Carvalho B P R, Medeiros T H, Ribeiro R S (2007) Prediction of Post-Synaptic Activity in Proteins Using Recursive Feature Elimination. In: XV European Symposium on Artificial Neural Networks, ESANN 2007. Bruges, Belgium
4. Carvalho B P R, Mendes T M, Ribeiro R S, Fortuna R, Veneroso J M, Mudado M A (2009) A System for Recognition of Biological Patterns in Toxins Using Computational Intelligence. In: IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2009. Nashville, United States
5. Carvalho D H D, Jesus F R, Martins G C A G (2007) A Fuzzy System for Casting Operation Synchronization. ABG Systems Technical Report. Belo Horizonte, Brazil
6. Carvalho D H D, Maia B T, Moreira A P, Fonseca A, Fortuna R, Ramos, K C F (2005) Blast Furnace System for Synchronization of Operations using Computational Intelligence. In: XXXV Seminar of Melting, Refining and Solidification of Metals, ABM 2005. Brazil
7. Guyon I, Weston J, Barnhill S, Vapnik V N (2002) Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning 46(3):389–422
8. Haykin S (1994) Neural Networks, A Comprehensive Foundation. Macmillan, New York
9. Jang J S R, Sun C T, Mizutani E (1997) Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Prentice Hall, United States
10. Koza J R (1992) Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, Cambridge
11. Quinlan J R (1985) Decision Trees and Multi-Valued Attributes. Machine Intelligence 1(1):81–106
12. Suykens J A K, Vandewalle J (1999) Least Squares Support Vector Machine Classifiers. Neural Processing Letters 9(3):293–300
13. Teixeira R A, Braga A P, Takahashi R H C, Saldanha R R (2001) Recent advances in the MOBJ algorithm for training artificial neural networks. International Journal of Neural Systems 11(3):265–269
14. Vapnik V N (1995) South J, Blass B (2001) The Nature of Statistical Learning Theory. Springer, London