NESUS cost IC1305

Network for Sustainable Ultrascale Computing

Proceedings of the First PhD Symposium on Sustainable Ultrascale Computing Systems (NESUS PhD 2016)
Timisoara, Romania

Jesus Carretero, Javier Garcia Blas
Dana Petcu
(Editors)

February 8-11, 2016

Computational Intelligence Modeling of Pharmaceutical Properties

Hossam M. Zawbaa
Faculty of Mathematics and Computer Science, Babes-Bolyai University, Romania
Faculty of Computers and Information, Beni-Suef University, Egypt
hossam.zawbaa@gmail.com

Abstract

In the pharmaceutical industry, a good understanding of the casual relationship between product quality and attributes of formulations is very useful in developing new products, and optimizing manufacturing processes. Feature selection is mandatory due to the abundance of noisy, irrelevant, or misleading features. The selected features will improve the performance of the prediction model and will provide a faster and more cost effective prediction than using all the features. With the big data captured in the pharmaceutical product development practice, computational intelligence (CI) models and machine learning algorithms could potentially be used to identify the process parameters of formulations and manufacturing processes. That needs a deep investigation of roller compaction process parameters of pharmaceutical formulations that affect the ribbons production. In this work, we are using the bio-inspired optimization algorithms for feature selection such as (grey wolf, Bat, flower pollination, social spider, antlion, moth-flame, genetic algorithms, and particle swarm) to predict the different pharmaceutical properties.

Keywords Computational Intelligence, Pharmaceutical Roll Compaction, Bio-inspired Optimization, Feature Selection

I. Introduction

A feature is an measurable property of the problem under observation, over the past years the domain of features in machine learning and pattern recognition applications have expanded from tens to hundreds of variables or features used in such applications. Hence the use of reduction or selection techniques is essential to reduce the large number of feature in the problem. Feature selection is a process of selecting a subset of features from a larger set of features, which leads to the reduction of the dimensionality of features space for a successful classification task. Feature selection provides a way for identifying the important features and removing irrelevant or redundant features from a dataset [1]. Feature Selection helps in understanding data, reducing computation requirement, reducing the effect of curse of dimensionality and improving the predictor performance [2].

Formerly, an exhaustive search for the optimal or near to optimal solution in a enormous search space may be impracticable, many researches seek to model the feature selection as a optimization problem [3]. One of the most used methods to solve the feature selection problems are evolutionary and swarm intelligence methods. Swarm intelligence is a computational intelligence-based approach which is made up of a population of artificial agents and inspired by the social behavior of animals (fish, birds, fireflies, etc.) from the real world. Example of such methods are ant colony optimization [4], bat algorithm [5], and particle swarm optimization (PSO) [6].

Roller compaction is method of preparing drug granules for capsules or for tablet formulations used in the pharmaceutical industry with suitable densification. The most common filler binder excipient used in roller compaction are microcrystalline cellulose (MCC), dibasic calcium phosphate (DCP), and lactose. Roller compaction is a particle size enlargement technique that granulated the powder materials to obtain materials of intermediate sizes in tablets production. The use of latest technology facilitates to efficient production of high quality granules. The selection of the critical roll compaction parameters such as (constant compacting pressure and constant roller gap) is very important.

Being a part of the development of in-silico process models for roll compaction (IPROCOM) project, Marie Curie. IPROCOM project employs a multidisciplinary approach to understand the fundamental mechanisms of particulate manufacturing processes involving roll compaction, and to develop predictive in-silico tools that can be used by various industrial sectors in Europe. In addition, we in need to establish a computational intelligence framework that identifies the critical material and process parameters and defines the design spaces for robust formulations and efficient production.

The aggregate aim of this work is to propose the bio-inspired optimization algorithms for feature selection that maximize feature reduction and obtaining comparable or even better prediction results of roll compaction parameters from using full features and conventional feature selection techniques.

II. Related work

Evolutionary computational (EC) algorithms have been used in feature selection issues such as genetic algorithm (GA), genetic programming (GP), ant colony optimization (ACO), and particle swarm optimization (PSO). GA was the first evolutionary based algorithm introduced in the literature and developed based on the natural process of evolution through reproduction [7]. Particle swarm optimization (PSO) is one of the well-known swarm algorithms. In PSO, each solution is considered as a particle with specific characteristics (position, fitness, and speed vector) that defines the moving direction of each particle [8]. A hybrid methods can also be applied in which two evolutionary algorithms are used to solve the problem, for example [9] proposed a new feature selection approach that is based on the integration of GA and PSO. Artificial bee colony (ABC) is a numerical optimization algorithm based on foraging behavior of honeybees. In ABC, the employer bees try to find food source and advertise the other bees. The onlooker bees follow their interesting employer and the scout bee fly spontaneously to find the best food source [10]. Social spider optimization (SSO) algorithm is a population based algorithm and one of the comparatively recent swarm algorithms [11].

A virtual bee algorithm (VBA) is applied to optimize the numerical function in 2-D using a swarm of virtual bees, which move randomly in the search space and interact to find food sources. From the interactions between these bees results the possible solution for the optimization problem [12]. A proposed approach based on natural behavior of honeybees, which randomly generated worker bees are moved in the direction of the elite bee. The elite bee represents the optimal (near to optimal) solution [13]. Ant colony optimization (ACO) wrapper-based feature selection algorithm was applied in network intrusion detection with rough set theory [14]. Artificial fish swarm (AFS) algorithm mimics the stimulant reaction by controlling the tail and fin. AFS is a robust stochastic technique based on the fish movement and its intelligence during the food finding process [15].

III. Thesis idea

The main goal of this thesis study was to investigate the roller compaction and granulation characteristics of pharmaceutical formulations. During the roller compaction operation, uniformly mixed powder blends are passed continuously through the gap between a pair of counter rotating compression rolls to form solid ribbons or sheets which are then passed through a mill or granulator with a suitable sized screen to form dry granules. Compared to wet granulation processes, dry granulation by roller compaction has various advantages such as simpler manufacturing procedure, easier scale up and higher production throughput. Dry granulation is also energy efficient and suitable for processing pharmaceutical agents that are sensitive to moisture and heat. The complexity of formulation design is a highly specialised task, requiring specific knowledge and often years of experience. In this work, we have applied bio-inspired optimization algorithms such as (grey wolf optimization, Bat optimization, cuckoo search, flower pollination algorithm, social spider optimization, etc) for feature selection and prediction of different pharmaceutical properties. After that, we use machine learning techniques like (artificial neural network, k-nearest neighbour, extreme learning machine, etc) to predict the different pharmaceutical properties such as (true density, porosity, tensile strength, fines, etc).

Each optimization algorithm is run for 20 times to test the algorithm convergence capability. The used evaluation indicators to compare different optimization algorithms are:

1. **Average reduction** represents the average size of selected features to the total number of features.

2. **Mean square error (MSE)** measures the average of squared errors that means the difference between actual output and predicted ones.

The two evaluation criteria or objective function in the wrapper feature selection is commonly reflecting the regression performance as well as the feature reduction. A generic representation of the fitness function representing for both regression performance and feature reduction as described in equation (1):

$$f_\theta = \alpha * E + (1-\alpha)\frac{\sum_i \theta_i}{N}, \quad (1)$$

where f_θ is the fitness function given a vector θ sized N with 0/1 elements representing unselected / selected features, N is the total number of features in the dataset, E is the prediction error, and α is a constant controlling the importance of regression performance to the number of features selected.

A random controlling term (α) is used to balance the trade-off between exploration and exploitation and hence should be carefully adapted. Therefore, at the beginning of optimization (α) has its maximum value to allow for maximum exploration and at the end of optimization it has minimum value for more exploitation of search space. Each bio-inspired algorithm is initialized with n random agents, each agent (solution) representing a given selected feature combination. After that, each algorithm is iteratively applied for a number of iterations hoping to converge to a good solution. Individual solution is represented as a continuous valued vector with same dimension as number of attributes in the given dataset. The solution vector continuous values are limited to the range [0, 1]. At the solution fitness function evaluation the continuous valued solution is threshold to its binary representation using equation (2).

$$y_{ij} = \begin{matrix} 0 \text{ If}(x_{ij} < 0.5) \\ 1 \text{ Otherwise} \end{matrix} \quad (2)$$

where x_{ij} is the continuous value of the solution number i in dimension j, and y_{ij} is a discrete representation of solution vector x.

IV. Conclusion and future work

In this work, bio-inspired optimization algorithms were proposed and applied for feature selection in wrapper mode. The most recent bio-inspired optimization algorithms such as (GWO, ALO, BAT, SSO, and FPA) are hired in the feature selection domain for evaluation and results are compared against well-known feature selection methods particle swarm optimization (PSO) and genetic algorithm (GA). The evaluation is performed using a set of evaluation criteria to assess different aspects of the proposed system.

Acknowledgment

This work was supported by the IPROCOM Marie Curie initial training network, funded through the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement No. 316555. In addition, this work was partially supported by NESUS.

References

[1] Chizi, Barak and Rokach, Lior and Maimon, Oded and Wang, J, "A Survey of Feature Selection Techniques", 2009.

[2] Chandrashekar, Girish, Sahin, Ferat, "A survey on feature selection methods", Computers & Electrical Engineering, pp. 16-28, Vol. 40, No. 1, 2014.

[3] Duda, Richard O and Hart, Peter E and Stork, David G, "Pattern classification", John Wiley & Sons, 2012.

[4] Forsati, Rana and Moayedikia, Alireza and Jensen, Richard and Shamsfard, Mehrnoush and Meybodi, Mohammad Reza, "Enriched ant colony optimization and its application in feature selection", Neurocomputing, pp. 354-371, Vol. 142, 2014.

[5] Rodrigues, Douglas and Pereira, Luís AM and Nakamura, Rodrigo YM and Costa, Kelton AP and Yang, Xin-She and Souza, André N and Papa, João Paulo, "A wrapper approach for feature selection based on Bat Algorithm and Optimum-Path Forest", Expert Systems with Applications, pp. 2250-2258, Vol. 41, No. 2, 2014.

[6] Inbarani, H Hannah, Azar, Ahmad Taher, Jothi, G, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis", Computer methods and programs in biomedicine, pp. 175-185, Vol. 113, No. 1, 2014.

[7] Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, John Henry Holland, MIT press, 1992.

[8] R. C. Eberhart, and J. Kennedy, "A New Optimizer Using Particle Swarm Theory", Proceeding of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, pp. 39-43, 1995.

[9] Ghamisi, Pedram and Benediktsson, Jon Atli, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization", Geoscience and Remote Sensing Letters, IEEE, pp. 309-313, Vol. 12, No. 2, 2015.

[10] Dervis Karaboga, Bahriye Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm", Journal of Global Optimization, Vol. 39, No. 3, pp. 459-471, 2007.

[11] Cuevas, E., Cienfuegos, M., Zaldivar, D., Perez-Cisneros, M., "A swarm optimization algorithm inspired in the behavior of the social-spider", Expert Systems with Applications, Vol. 40, No. 16, pp. 6374-6384, 2013.

[12] Yang XS, "Engineering optimizations via nature-inspired virtual bee algorithms", In: Lecture notes in computer science, Springer (GmbH), pp. 317-323, 2005.

[13] Sundareswaran K, Sreedevi VT, "Development of novel optimization procedure based on honey bee foraging behavior", IEEE International conference on systems, man and cybernetics, pp. 1220-1225, 2008.

[14] H. Ming, "A rough set based hybrid method to feature selection", in Proc. Int. Symp. KAM, pp. 585-588, 2008.

[15] X. L. Li, Z. J. Shao, J. X. Qian, "An Optimizing Method Based on Autonomous Animates: Fish-swarm Algorithm", Methods and Practices of System Engineering, pp. 32-38, 2002.