Studying Proton–Proton Interaction at Large Hadrons Collider Using Genetic Programming

Amr Radi

Department of Physics, College of Sciences, Sultan Qaboos University,
Al Khoudh, Muscat 123, Oman
Also at Department of Physics, Faculty of Sciences, Ain Shams University,
Abbassia, Cairo 11566, Egypt
E-mail: Amr.radi@cern.ch

Corresponding author: Amr Radi
E-mail: Amr.radi@cern.ch

Abstract. This paper describes how to use Genetic Programming (GP) as an evolutionary computational that is a family of algorithms for global optimization. GP, as a global optimization technique used by discovery of a new function for modeling physical phenomena. The p-p interactions are modeled at Large Hadron Collider (LHC) experiments, the number of charged particles multiplicity <n> and the total cross-section, σT, as functions of the total center of mass energy (from low to ultra-high energy), √s are discovered by using GP. In view of the discovered function for <n> (√s), the overall trend of the values predicted is consistent with LHC data [predicted values are 34.8638 and 35.3520 at √s = 13 TeV and √s = 14 TeV respectively]. The new function σT (√s), trained on experimental data of Particle Data Group (PDG) demonstrates a nice match to the other models. The predicted values of the total cross section at √s = 13 TeV, and 14 TeV are found to be 109.0381 mb and 111.8329 mb respectively. Furthermore, the values predicted are agreed with other models like Block

1. Introduction

Physics explores the basic particles and forces that form the world around us and explains how our universe functions at its most basic stage. The research of these interactions requires enormously large energy impacts, such as those in the LHC [1-8], up to 14 TeV.

Experimental findings offer great opportunity to uncover the lack of the Standard Model. Moreover, LHC may be progressing beyond the standard model towards our knowledge of particulate physics.
One of the basic interactions in high-energy physics is the proton-proton (PP) interaction. It's essential to have a full knowledge of the reaction mechanism to fully exploit the enormous physical potential. Some of those models are based on the mechanism of string fragmentation [9-11] and some on an exchange of Pomeron [12].

Recently, the implementation of Artificial Intelligence (AI) techniques included various modelling techniques based on soft computing systems. The evolution algorithms exist physically in high-energy physics [13-17]. As the interaction parameters are not linked to the output, the conduct of the interactions is often complex. Multipart data analysis is required and AI methods are critical to comprehend the interactions of basic particles. These techniques are becoming useful as alternatives to traditional techniques [18]. To that extent, AI techniques, such as Deep learning [19], Global Optimization Algorithm [20, 21 22], can be used to simulate these interactions as alternative methods [13-17, 21-23].

The GP is an algorithm used for global optimisation in particle physics [24]. The GP is based on Darwin's evolutionary theory of nature. GP utilizes individual populations, chooses them by fitness, and generates genetic variation by one or more genetic operators [25]. The motive of a GP strategy is its study algorithm which learns the interactions between variables in data collections and then creates models to explain those connections; (mathematically dependence)[26, 27].

In this article, I found by using GP method the functions that describe the whole intersection of interactions at the various law values with elevated energies. This article is divided into five parts. The basics of the GP method are examined in Section 2. Section 3 describes how the interaction is modelled by GP. Finally, sections 4 contain the outcomes and Sections 5 the conclusions.

2. An Overview of Genetic Programming algorithms

Solutions are expressed in GP as syntactic trees instead of as code lines. For example (Figure 1) displays a solution + (x+2, x*z) for tree description. The program variables and constants (x, z and 2) are tree leaves. They are called terminals in GP, whereas interne nodes called functions are arithmetic operations (+ and *). Sets of features and terminals are the basic set of a GP scheme. For example, *(x-y, x-2+z) becomes (* (- x y) (- x (+ 2 z))).

![Fig. 1. GP syntax tree representing *(x-y, x-2+z).](image)

A fitness value is assigned to each individual in the population, which quantify
how well the problem can be solved. The fitness value is calculated by a fitness feature that is issue dependent. The fitness feature that can be implemented for the performance evaluation of a produced solution is measured by the remaining square (SSR), the square is the proportion of the total of the regression coefficients (SSR) [28].

The following steps as shown in figure 2 are a typical implementation of the GP (i.e. the process of determining the optimal (or almost best) solution to a GP problem).

![Flowchart of GP algorithm](image)

**Fig. 2.** The flowchart of GP algorithm.

### 3. The proposed Genetic Programming for the p-p interaction

The strategy is to use genetic programming which is smart enough to find PP interaction features (complete cross-sections in relation to the entire core of mass energy). GPLAB, a MATLAB genetic toolbox was used in this article. In addition to most of the traditional features in genetic programming, see [29].

Two different GP models (total cross sections and multiplicity distribution of charged particles) are suggested. One by one with experimental data, two models are trained / predicted to simulate the PP interaction. The following are the two models:

1. The function was set to the number of charged particles multiplicity \( n \), was configured to have the total center of mass energy \( \sqrt{s} \), inputs at a given pseudo rapidity \( \eta \). The yield is the mean charge multiplicity \( \langle n \rangle(\sqrt{s}) \).
2. The input is the total center of mass energy \( \sqrt{s} \) in the second model that calculates the total cross sections \( \sigma_T(\sqrt{s}) \).
GP is able to find new models to explain that the data sets are split in three sets (training, testing and predication). By using the training / test set, GP can detect a fresh model while using the predefined set to examine its generalizational capacities.

I used the statistical measures for the fitness function to measure the difference between the experimental results and the simulated results. The statistical measures of sum squared error (SSE), the correlation coefficient $r$ and the standard error of the regression (root of square mean error RSME), see [28].

4. Results

The current models are evaluated with the guidance of training data considering both current experimental results in the LHC experiment [30—33] and the Particle Physics Data Group in the range of low and ultra-high energy interaction studies PP interaction [34-35].

4.1. The first interaction model:

The discovered functions with GP were tested by using the experimental data of the mean charge multiplicity, for $|\eta| < 2.4$, with the centre-of-mass energy. The training data is based on experimental observations at LHC experiments in the range [900 Gev-13 TeV] [1, 3, 6], data from lower energy experiments for $|\eta| < 2.5$, NA22 [31], UA1 [32] and UA5 [30] are presented.

| Table 1. | Shows GP optimum achieved parameters for $<n>$ ($\sqrt{s}$) |
|-----------------|-----------------|-----------------|
| Populations     | 1000-3000       |
| Generations     | 100-1000        |
| Terminal Set    | { $\sqrt{s}$, constant } |
| Function set    | log ,-,/*,+, and sqrt |
| Selection method| Elits and Tournament |
| Fitness function| SSE             |
| Crossover rate  | 0-0.90          |
| Mutation rate   | 0-0.01          |

Running GP is the minimum value of an experimental data error of 0.0001 for configurations in table 1, until the fitness function is reduced to an acceptable level. The feature is screened with the error function to match the input patterns with the destination output patterns. The final form of the charging features for the center of mass energy multiplicity described by $\eta$ $< 2.4$ is given by:

$$<n>(\sqrt{s}) = 2.83382 \ (\sqrt{\sqrt{s}}) - 6.92655$$
Fig. 3. Displays the $<n>(\sqrt{s})$ advancement for $|\eta| < 2.4$. Also, shows experimental data for $|\eta| < 2.5$, UA5 [32], UA1 [31] and NA22 [31].

The $<n>(\sqrt{s})$ feature found was used to estimate the average charge multiplicity of UA5 [30]. Figure 3 shows that the mean charging multiplicity used by our found features is well matched with the forecast (UA5) results of the experimental information. In addition, the predicted function gives more fit compared with Levin et al. [33] and Lichened et al. [34] models.

The results of the statistical parameters for $<n>(\sqrt{s})$ functions give that the value of RMSE is 0.00001 and correlation coefficient is 0.993. Moreover, $<n>(\sqrt{s})$ predicted at LHC are of values 34.8638 and 35.3520 at $\sqrt{s} = 13$ TeV and $\sqrt{s} = 14$ TeV respectively. The trends are useful for [1, 30, 31, 32].

4.2. The second interaction model:

The training data takes into account both latest LHC experiment experimental observations [35-38] and the Data Group on particle physics for P-P-interaction studies with low to high power ranges [39-40]. The discovered functions $\sigma_r(\sqrt{s})$ with GP were tested by using the experimental data.

Table 2. Shows GP optimal parameters that were obtained.

| Populations | 1000-4000         |
|-------------|-------------------|
| Generations | 100-1000          |
| Terminal Set| { $\sqrt{s}$, constant } |
| Function set| log, -, /, *, +, and sqrt |
| Selection method| Elits and Tournament |
| Fitness function | SSE |
| Crossover rate | 0-0.90 |
| Mutation rate | 0-0.01 |
Running GP is the minimum value of an experimental data error of 0.001 for configurations in table 1, until the fitness function is reduced to an acceptable level. The feature is screened with the error function to match the input patterns with the destination output patterns. The $\sigma, (\sqrt{s})$ is given by:

$$\sigma, (\sqrt{s}) = (((10*\sqrt{5}+6)/(((0.3)/((\sqrt{5})-(2)))+( ((\sqrt{5})-(3))) )+( (7)/((\sqrt{5}-(2)))+(35+((sqrt(5)*((5/7.7)))))))$$

Fig. 4. The progress of the $\sigma, (\sqrt{s})$ function of $\sqrt{s}$ between 10-50000 Gev was demonstrated.

In the following figure 3 and 4, the values of the $\sigma, (\sqrt{s})$ is matched with results at low to high energies from cosmic rays together with a comprehensive fit for the competition [35-39]. In addition, the model is well in accordance with low energy extrapolation.

The predicted values for $\sigma, (\sqrt{s})$ at LHC are 109.0381 mb and 111.8329 mb at $\sqrt{s} = 13.0$ TeV and $\sqrt{s} = 14.0$ TeV respectively. Those results match with Block [40].

5. Conclusions

The latest work introduces the GP as a fresh method for the construction of the features of the mean multiplicity distribution of charged particles, $<n>(\sqrt{s})$, and the total collision cross sections, $\sigma, (\sqrt{s})$, of p-p interaction. The functions that have been found match the experimental data. In addition, they are capable of anticipating experimental data not used during the training session for $<n>(\sqrt{s})$ and $\sigma, (\sqrt{s})$. 
The discovered function for \( <n> (\sqrt{s}) \) has an overall trend of forecast values is well agreed at LHC \([\text{predicted values are } 34.8638 \text{ and } 35.3520 \text{ at } \sqrt{s} = 13 \text{ TeV and } \sqrt{s} = 14 \text{ TeV respectively}]\). The new function \( \sigma_t (\sqrt{s}) \), trained on experimental data of PDG demonstrates a nice match to the other models. The predicted values of the total cross section at \( \sqrt{s} = 13 \text{ TeV}, \text{ and } 14 \text{ TeV} \) are found to be \( 109.0381 \text{ mb} \) and \( 111.8329 \text{ mb} \) respectively. Furthermore, the values predicted are agreed with other models like Block. To conclude, in high-energy physics GP has become one of the leading study fields.

6. REFERENCE

[1] ATLAS Collaboration, Eur. Phys. J. C 76:403, 2016.
[2] CMS Collaboration, J. High Energy Phys. 06 160, 2012.
[3] CMS Collaboration, J. High Energy Phys. 01 079, 2011.
[4] ATLAS Collaboration, Phys. Lett. B 707 330-348, 2012.
[5] ALICE Collaboration, Eur. Phys. J. C 68345-354, 2010.
[6] ALICE Collaboration, Phys. Lett. B 693 53-68, 2010.
[7] ATLAS Collaboration, Phys. Lett. B, 758, p. 67-88, 2016.
[8] TOTEM Collaboration, EPL 96 21002, 2011.
[9] M. Jacob, R. Slansky, Phys. Rev. D 5, 1847, 1972.
[10] R. Hwa, Phys. Rev. Lett. 26, 1143, 1971.
[11] R. Hwa, Phys. Rev. D 1, 1790, 1970.
[12] R. Engel, Z. Phys. C 66 (1995) 203; R. Engel, J. Ranft and S. Roesler, Phys. Rev. D 52, 1995.
[13] L. Teodorescu, IEEE T. Nucl. Sci. 53, 2221, 2006.
[14] L. Teodorescu, D. Sherwood, Comput. Phys. Commun. 178, 409, 2008.
[15] J.M. Link, Nucl. Instrum. Meth. A 551, 504, 2005.
[16] M. El-Bakry, A. El-Dahshan, A. Radi, M. Tantawy, M. Moussa, The Application of the genetic algorithm-back propagation neural network algorithm in the high energy physics, East Eur. J. Phys. Vol.3 No.2 4-14, 2016.
[17] S. El-Bakry, A. Radi, Int. J. Mod. Phys. C 18, 351, 2007.
[18] S. Whiteson, D. Whiteson, Eng. Appl. Artif. Intel. 22, 1203, 2009.
[19] S. Haykin, Neural networks a comprehensive foundation (3rd ed.), Prentice Hall 2009.
[20] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989.
[21] W. Banzhaf, P. Nordin, R. Keller and F. Francone, Genetic Programming: An Introduction, Morgan Kaufmann, San Mateo, CA, 1998a.
[22] R. Poli and W. Langdon and N. F. McPhee and J. Koza, A Field Guide to Genetic Programming. UK: Lulu Enterprises; 2008.
[23] A. Eiben, J. Smith, Introduction to Evolutionary Algorithms, Springer, Berlin, 2003.
[24] E. El-dahshan, A. Radi, M. El-Bakry, Int. J. Mod. Phys. C, 20, 1817, 2009.
[25] M. Mitchell, An Introduction to Genetic Algorithms, MIT Press, 1996.
[26] J. Koza, and J. Rice, Genetic Programming: The Movie, MIT Press, Cambridge, MA, 1992.
[27] W. Langdon and R. Polim, Foundations of Genetic Programming, Springer-Verlag, 2002.
[28] J. Cohen, P. Cohen, S. West, & L. Aiken, Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, NJ: Erlbaum, 2003.
[29] S. Silva and J. Almeida, “GPLAB - A Genetic Programming Toolbox for MATLAB”, Nordic MATLAB Conference, 2003
[30] UA1 Collaboration, Nucl. Phys. B335 261, 1990.
[31] EHS/NA22 Collaboration, Z. Phys. C37 215, 1988.
[32] UA5 Collaboration, Phys. Lett. B160 193, 1985.
[33] E. Levin and A Rezaeian, Rev. D82, 2010
[34] A. Likhoded, A.Luchinsky, and A.Novoselov, Phys. Rev. D 86, 034017, 2010
[35] CMS-Collaboration, Performance Analysis Note, CMS-PAS-FWD-11-001, 2011.
[36] G. Aad et al., arXiv: 1104. 0326 [hep-ex], 2011.
[37] M. G. Poghosyan et al., arXiv:1109.4510 [hep-ex], 2011.
[38] TOTEM Collaboration, EPL, 96 21002, 2011.
[39] Par C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001, 2016.
[40] M. M. Block and F. Halzen. Phys. Rev. D 70, 091901(R), 2004.