Unconstrained optimization based fractional order derivative for data classification

Dian Puspita Hapsari¹, Imam Utoyo², Santi Wulan Purnami³
¹Institute of Technology Adhi Tama, Arief Rahman Hakim street, Surabaya, East Java, Indonesia
²Airlangga University, H.Ir. Soekarno street, Surabaya, East Java, Indonesia
³Institute of Technology Sepuluh Nopember, Keputih street, Surabaya, East Java, Indonesia

E-mail: m.i.utoyo@fst.unair.ac.id

Abstract. Data classification has several problems one of which is a large amount of data that will reduce computing time. The Fractional gradient descent method is an unconstrained optimization algorithm to train classifiers with support vector machines that have convex problems. Compared to the classic integer-order model, a model built with fractional calculus has a significant advantage to accelerate computing time. In this research it is to conduct a qualitative literature review in order to investigate the current state of these new optimization method fractional derivatives can be implemented in the classifier algorithm.

1. Introduction

Large-scale data classification can be done using artificial intelligence, which is able to detect patterns from data then predict and provide conclusions automatically to users. Artificial intelligence is able to carry out predictive analyses that produce knowledge insights and are used as a basis for decision making. Predictive analysis is part of Knowledge Discovery in Databases (KDD) or Data Mining (DM) activities that use large-scale data sets to identify interesting patterns in past data sets to predict future conditions. In data mining tasks, predictive analysis will involve machine learning methods that conduct supervised learning for large scale data. In some fields of research such as Pattern Recognition [1], Sentiment Analysis [2], Bioinformatics [3], Image Processing [4].

The most interesting topics in the field of machine learning are optimization. Numerical optimization algorithms extensively were finding local optima of any given function. Finding best available values of some objective function given a defined domain (or input), including a variety of different types of objective functions and different types of domains, which works best with convex functions. Types of optimization algorithms for minimize a loss function we used First Order Optimization Algorithms - Gradient descent method is commonly used to train classifiers by minimizing the error function.

[5] Were introduce the classification optimization with the gradient method for large-scale data training and identify how optimization problems arise in machine learning and what makes them challenging. Through case studies on text classification and the training of deep neural networks, he discussed how optimization problems arise in machine learning, what makes them challenging including an investigation of two main streams of research on techniques that diminish noise in the stochastic directions and methods that make use of second-order derivative approximations.
To deal with large scale datasets and to factorize large scale matrices, different algorithms have been proposed. Stochastic gradient descent (SGD) algorithms much simple and efficient and are extensively used for matrix factorization [6] [7]. In the literature, different variants of gradient descent (GD) and stochastic gradient descent (SGD) have been suggested with the aim of increasing performance in terms of accuracy and convergence speed. In both methods, parameters are updated in an iterative manner to minimize the objective function.

In GD, a specific iteration involves running through all the samples in the training set for a single update of a parameter, while in SGD, a single or a subset of samples from the training set is taken for parameter update. This makes GD highly computationally complex for large numbers of training samples. Thus a suitable choice for many classification applications involving large numbers of training samples is the SGD. Update rules for SGD based techniques involve integer order gradient descent. The integer order gradient descent based SGD can be further improved by applying fractional order gradient descent using the concepts of fractional calculus as has been observed in different areas of research [8] recommender system, [9] motion analysis and [10] system identification.

The purpose of this study is to conduct a qualitative literature review to investigate the current state of the optimization method based on fractional order derivative conducted in Neural Network classifier. The paper is organized as follows; Section II is devoted to math preparation for the convergence problem of fractional gradient method. Section III explained dataset for classification task. Section IV shows some numerical examples to verify the proposed methods. At last, conclusion is presented in Section V.

2. Research methodology

Literature review is considered a rigorous scientific inquiry, limiting random biases and errors through planned methods and strategies. These strategies imply the search for all important papers and the adoption of selected criteria, which can be used by other researchers. Literature review is a suitable methodology for conducting research in the field of operations management, which produces a reliable stock of knowledge that can support practical applications. Our research questions are as follows:

Research Question: What are the practical applications of unconstrained optimization based fractional order derivative for data classification?

The optimization problem is generally formulated as a convex optimization problem because is easy to solve the convex optimization problem than non-convex problems because every local optimum is a global optimum in a convex optimization problem [11]. Generally, the categories of convex optimization problem are of two types: unconstrained convex optimization problem and constrained convex optimization problem. Constrained convex optimization problem are further of three types: equality constrained (EC), inequality constrained (IC) and hybrid constrained (HC) [12][13]. Gradient based methods are generally used for solving the unconstrained convex optimization problems. Constrained convex optimization problem can be converted to unconstrained convex optimization problem and then gradient based methods can be applied to them [14] [15].

Definition 1. Let $x: \mathbb{R} \to \mathbb{R} \times [t_0, t_f], \alpha \in (0,1], x \in C^1$then, the Grunwald-Letnikov (G-L) fractional derivative is given by;

$$\frac{\alpha}{\delta} \mathcal{D}^\alpha_x f(x) = \lim_{h \to 0} h^{-\alpha} \sum_{k=0}^{[\frac{x-\alpha}{h}]} \binom{-\alpha}{k} f(x - kh).$$

Definition 2. $\frac{\alpha}{\delta} \mathcal{D}^\alpha_x f(x)$ Denotes the fractional differential operator based on G-L definition, $(x)$ denotes a differintegrable function, $V$ is the fractional order,$[a,x]$ is the domain of $f(x)$, $\Gamma$ is the gamma function, and is the rounding function.

The Riemann-Liouville (R-L) fractional derivative is given by;

$$\frac{\alpha}{\delta} \mathcal{D}^\alpha_x f(x) = \frac{1}{\Gamma(n - \alpha)} \frac{d^n}{dx^n} \int_a^x \frac{f(y)}{(x - y)^{\alpha-n+1}} dy.$$
where \( \mathcal{D}_x^y \) denotes the fractional differential operator based on G-L definition; \( n = [v+1] \).

Definition 3. The Caputo fractional derivative is given by:

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y f(x) = \frac{1}{\Gamma(n-v)} \int_a^x (x-y)^{n-v-1} f^{(n)}(y) \, dy
\]

where \( \mathcal{C} \mathcal{D}_x^y \) is the fractional differential operator based on Caputo definition, \( n = [v+1] \).

We can know that for the G-L and R-L definition, the fractional differential of constant function is not equal to 0. Only with the Caputo definition, the fractional differential of constant function equals to 0, which is consistent to the integer-order calculus. Therefore, the Caputo definition is widely used in solving engineering problems and it was employed to calculate the fractional order derivative [16][17]. Fractional Order Derivative Caputo type with \( f(x) = C \), for \( n = 1 \),

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y f(x) = \frac{1}{\Gamma(n-v)} \int_a^x (x-y)^{n-v-1} f^{(n)}(y) \, dy
\]

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y C = \frac{1}{\Gamma(n-v)} \int_0^x (x-y)^{n-v-1} f'(y) \, dy
\]

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y C = \frac{1}{\Gamma(n-v)} \int_0^x (x-y)^{-v} f'(y) \, dy
\]

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y C = \frac{1}{\Gamma(n-v)} \int_0^x (x-y)^{-v} C' \, dy
\]

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y C = \frac{1}{\Gamma(n-v)} \int_0^x (x-y)^{-v} 0 \, dy
\]

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y C = \frac{1}{\Gamma(n-v)} \int_0^x 0 \, dy
\]

\[
\frac{\mathcal{C}}{d} \mathcal{D}_x^y C = \frac{1}{\Gamma(n-v)} (C - C) = \frac{1}{\Gamma(n-v)} \cdot 0 = 0
\]

Fractional Order Derivative Riemann-Liouville (R-L) type with \( f(x) = C \), for \( n = 1 \)

\[
\mathcal{R}_a^L \mathcal{D}_x^y f(x) = \frac{1}{\Gamma(n-v)} \frac{d^n}{dx^n} \int_a^x \frac{f(y)}{(x-y)^{v-n+1}} \, dy
\]

\[
\mathcal{R}_a^L \mathcal{D}_x^y C = \frac{C}{\Gamma(1-v)} \frac{d^n}{dx^n} \lim_{a+\rightarrow x} \int_0^x (x-y)^{-v} \, dy
\]

\[
\mathcal{R}_a^L \mathcal{D}_x^y C = \frac{C}{\Gamma(1-v)} \frac{d^n}{dx^n} \left( \frac{1}{v-1} (x-x)^{1-v} - \frac{1}{v-1} x^{1-v} \right)
\]

\[
\mathcal{R}_a^L \mathcal{D}_x^y C = \frac{C}{\Gamma(1-v)} \frac{d^n}{dx^n} \left( - \frac{1}{v-1} x^{1-v} \right)
\]

\[
\mathcal{R}_a^L \mathcal{D}_x^y C = \frac{C}{\Gamma(1-v)} \left( - \frac{1}{v-1} x^{1-v} \right)
\]
The fractional order derivative type (R-L) for \( f(x) = C \) is not equal to zero so it is not recommended to be used in system modelling because it does not present a natural state or situation. Instead the Caputo type for \( f(x) = C \) is equal to zero so it is considered to be able to present a natural state for system modelling.

3. Literature review

In this section, the basic knowledge of fractional calculus is introduced. Fractional calculus is a branch of mathematical analysis, which studies the several different possibilities of defining real number or complex number powers of the derivative of a function. Different from integer calculus, fractional derivative does not have a unified temporal definition expression up to now \[14\][18]. The commonly used definitions of fractional derivative are Grünwald-Letnikov (G-L), Riemann-Liouville (R-L), and Caputo derivatives.

Aguilar in his research proposed a fractional order neural network (FONN) model with Grünwald-Letnikov fractional derivative \[10\]. Gradient Descent algorithm with a fractional order derivative of Grünwald-Letnikov (G-L) type, have some advantage of the non-locality that the fractional calculus offers. The fractional order learning algorithm can be take several value of the fractional parameters alpha which allow to be more accurate than integer-order version. Improvement in accuracy and reduction of number of parameter was expected, since fractional order derivative have infinite memory and non-locality properties.

Bao in his research proposed a fractional-order deep back propagation (BP) neural network model with Caputo derivative \[19\]. The proposed model had no limitations on the number of layers and the fractional-order was extended to arbitrary real number bigger than 0. The numerical results support that the fractional-order BP neural networks with \( L_2 \) regularization are deterministically convergent and can effectively avoid the over fitting phenomenon.

Khan in his research proposed the radial basis function neural networks with Riemann–Liouville (R-L) derivative-based fractional gradient descent methods \[20\]. For the pattern classification problem, the proposed method demonstrated better accuracy in a fewer number of iterations. The proposed algorithm has shown better performance compared to the conventional RBF-NN in both, training and testing phases. For the problem of nonlinear system identification, the proposed framework achieved high convergence performance.

From the results of the performance of the three types of fractional order derivative methods for classifying optimization, it is clear that the G-L type has the lowest performance. With a value of 94.19\% for training data and 93.65\% for testing data on classification task. Caputo types have the best performance of the three. With a value of 98.84\% for training data and 95.00\% for testing data on large scale data classification activities (data size = 10,000-60,000).
Figure 1. Fractional order gradient descent methods performance

The training accuracy is slightly decreased but the testing accuracy significantly increased, which indicated that Caputo type suppress over fitting and improve the generalization of classifier. Then, the stability and convergence of the classifier converged fast and stably and were finally close to zero.

4. Conclusions and future developments
The purpose of this study was to lead a literature review in order to investigate the current state of the optimization method based on fractional order derivative conducted in Neural Network classifier. Based on the findings of this review, it is possible to establish an overview of the predominant optimization method based on fractional order derivative. The literature review highlights that optimization method based on fractional order derivative is very important to increase accuracy and increase computational time for online learning and dealing with large data sets. For the future study it is possible to implement optimization method based on fractional order derivative for other classifier such SVM.

Acknowledgments
The authors would like to express their gratitude to the editors and anonymous reviewers for their valuable comments and suggestions which improve the quality of this paper. Gratitude’s for the supervisor of the doctoral program, the faculty of science and technology, Airlangga University.

References
[1] Gallego A J, Calvo-Zaragoza J, Valero-Mas J J, and Juan R 2018 Clustering-based k-nearest neighbor classification for large-scale data with neural codes representation Pattern Recognit 74 pp 531-43.
[2] Al-Saqqa S, Al-Naymat G, and Awajan A 2018 A large-scale sentiment data classification for online reviews under apache spark Procedia Computer Science 141 pp 183-89.
[3] Gu X Q, Chung F L, and Wang S T 2018 Fast convex-hull vector machine for training on large-scale ncRNA data classification tasks Knowledge-Based Syst 151 pp 149-64.
[4] Liu H, Yang S, Gou S, Chen P, Wang Y, and Jiao L 2017 Fast Classification for Large Polarimetric SAR Data Based on Refined Spatial-Anchor Graph IEEE Geosci. Remote Sens. Lett., 14 pp 1589-93.
[5] Bottou L 2010 Large-scale machine learning with stochastic gradient descent Proceedings of COMPSTAT 2010 - 19th International Conference on Computational Statistics, Keynote, Invited and Contributed Papers.
[6] Pu Y F, Siarry P, Zhou J L, Liu Y G, Zhang N, Huang G and Liu Y Z 2014 Fractional partial differential equation denoising models for texture image Sci. China Inf. Sci 57 pp 1-19.
[7] Chen D, Chen Y Q, and Xue D 2015 Fractional-order total variation image denoising based on proximity algorithm Appl. Math. Comput 257.

[8] Khan Z A, Chaudhary N I, and Zubair S 2019 Fractional stochastic gradient descent for recommender systems Electron. Mark 29 pp 275-85.

[9] Chen D, Sheng H, Chen Y Q, and Xue D 2013 Fractional-order variational optical flowmodel for motion estimation Philos. Trans. R. Soc. A Math. Phys. Eng. Sci 371.

[10] Zuñiga Aguilar C J, Gómez-Aguilar J F, Alvarado-Martínez V M and Romero-Ugalde H M 2020 Fractional order neural networks for system identification Chaos, Solitons and Fractals 130 109444

[11] Wang H, Yu Y, Wen G, Zhang S, and Yu J 2015 Global stability analysis of fractional-order Hopfield neural networks with time delay Neurocomputing 154.

[12] Akhmet M U and Karacaören M 2018 A Hopfield neural network with multi-compartmental activation Neural Comput. Appl 29 pp 815-22.

[13] Hu H P, Wang J K, and Xie F L 2019 Dynamics analysis of a new fractional-order hopfield neural network with delay and its generalized projective synchronization Entropy 21.

[14] Caputo M and Fabrizio M 2015 A new definition of fractional derivative without singular kernel Prog. Fract. Differ. Appl. 1 pp 73-85.

[15] Journal A I, Caputo M, and Fabrizio M 2015 Progress in Fractional Differentiation and Applications A new Definition of Fractional Derivative without Singular Kernel Progr. Fract. Differ. Appl 1 pp 87-92.

[16] Chen Y, Wei Y, Wang Y, and Chen Y Q 2018 Fractional order gradient methods for a general class of convex functions Proceedings of the American Control Conference.

[17] Wei Y, Chen Y, Cheng S, and Wang Y 2017 “Discussion on fractional order derivatives,” IFAC-PapersOnLine 50 pp7002-06.

[18] Losada J and Nieto J J 2015 Progress in Fractional Differentiation and Applications A new Definition of Fractional Derivative without Singular Kernel Progr. Fract. Differ. Appl 1 pp 87-92.

[19] Bao C, Pu Y, and Zhang Y 2018 Fractional-Order Deep Backpropagation Neural Network Comput. Intell. Neurosci 2018.

[20] Khan S, Ahmad J, Naseem I, and Moinuddin M 2018 A Novel Fractional Gradient-Based Learning Algorithm for Recurrent Neural Networks Circuits, Syst. Signal Process 37.