

Support Vector Machine optimization with fractional gradient descent 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. SVM is a reliable linear classifier for linear or non-linear data, for large-scale data, there are computational time constraints. 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 investigate the current state of this new optimization method fractional derivatives that can be implemented in the classifier algorithm. The results of the SVM Classifier with fractional gradient descent optimization, it reaches a convergence point of approximately 50 iterations smaller than SVM-SGD. The process of updating or fixing the model is smaller in fractional because the multiplier value is less than 1 or in the form of fractions. The SVM-Fractional SGD algorithm is proven to be an effective method for rainfall forecast decisions.

1. Introduction

Large-scale data classification can be done using artificial intelligence, which can detect patterns from data then predict and provide conclusions automatically to users. Artificial intelligence can 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, the 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 the 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 minimizing 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 discusses 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 to increase performance in terms of accuracy and convergence speed. In both methods, parameters are updated iteratively 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 investigate the current state of the optimization method based on the fractional-order derivative conducted in the SVM classifier. The paper is organized as follows; Section II is devoted to the literature review of the fractional gradient method. Section III explained research and method for a classification task. Section IV result and discuss verify the proposed methods. At last, a conclusion is presented in Section V.

1.1. Literature Review

The optimization problem is generally formulated as a convex optimization problem because it is easy to solve the convex optimization problem than non-convex problems. After all, every local optimum is a global optimum in a convex optimization problem [11]. Generally, the categories of convex optimization problems are of two types: unconstrained convex optimization problem and constrained convex optimization problem. Constrained convex optimization problems 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. The Constrained convex optimization problem can be converted to the unconstrained convex optimization problem and then gradient-based methods can be applied to them [14] [15].

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

$${}^G L_a D_x^\nu f(x) = \lim_{h \rightarrow 0} h^{-\nu} \sum_{k=0}^{[(x-a)/h]} \binom{-\nu}{k} f(x - kh),$$

Definition 2. ${}^G L_a D_x^\nu$ Denotes the fractional differential operator based on G-L definition, (x) denotes a differentiable function, ν is the fractional-order, $[a,x]$ is the domain of $f(x)$, Γ is the gamma function, and is the rounding function.

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

$${}^{RL} D_x^\nu f(x) = \frac{1}{\Gamma(n - \nu)} \frac{d^n}{dx^n} \int_a^x \frac{f(y)}{(x - y)^{\nu-n+1}} dy$$

Where ${}^{RL} D_x^\nu$ denotes the fractional differential operator based on G-L definition; $n = [\nu+1]$.

Definition 3. The Caputo fractional derivative is given by;

$${}^c \mathcal{D}_x^v f(x) = \frac{1}{\Gamma(n-v)} \int_a^x (x-y)^{n-v-1} f^{(n)}(y) dy$$

Where ${}^c \mathcal{D}_x^v$ is the fractional differential operator based on Caputo definition, $n = [v+1]$.

2. Method

We can know that for the G-L and R-L definition, the fractional differential of a constant function is not equal to 0. Only with the Caputo definition, the fractional differential of a constant function equals 0, which is consistent with 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]. 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 modeling because it does not present a natural state or situation. Instead of 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 modeling.

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, the fractional derivative does not have a unified temporal definition expression up to now [14][18]. The commonly used definitions of the 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 the 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 take several values of the fractional parameters alpha which allows being more accurate than the integer-order version. Improvement in accuracy and reduction of a number of the parameter was expected since the fractional-order derivative has infinite memory and non-locality properties.

Bao in his research proposed a fractional-order deep backpropagation (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 an arbitrary real number bigger than 0. The numerical results support that the fractional-order BP neural networks with $L2$ regularization are deterministically convergent and can effectively avoid the overfitting 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 the classification tasks. 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).

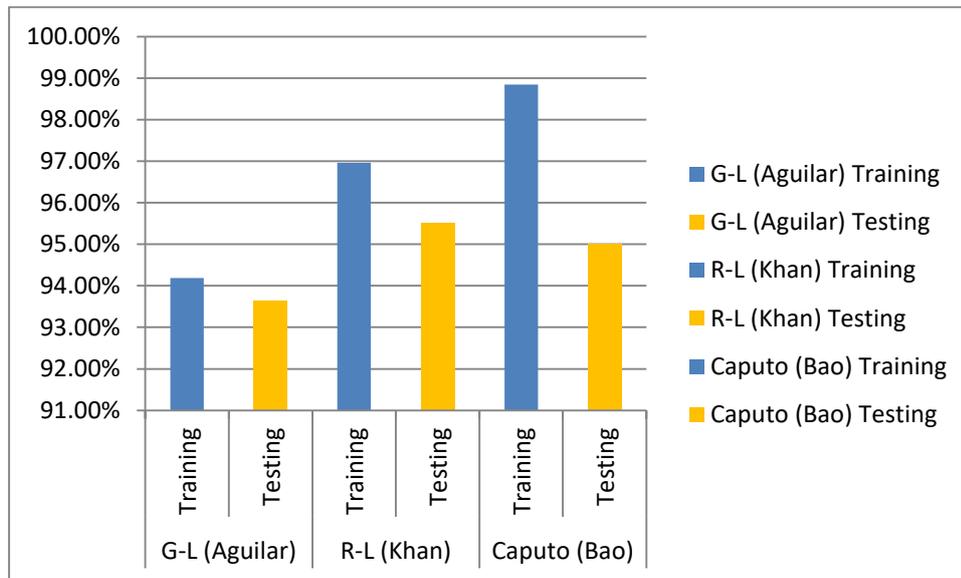


Figure 1. Fractional order gradient descent methods performance

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

3. Result and Discussion

To verify the convergence of the proposed Fractional-order Gradient Descent SVM algorithm, simulation has been done on the Iris dataset and the Rainfall dataset. For the simulation, we have a rainfall dataset with 1825 record data for 1095 training samples and 730 testing samples data. The Rainfall dataset is collected from ground-based meteorological data from Maritim Perak Station (ID WMO: 96937). For the assessment of the algorithms results, cross-validation and external testing were carried out. The datasets were divided into two subsets, training and test, comprising 80% and 20% of the original samples, respectively. The training set takes 80% of the samples randomly from the dataset as the training set. Test set we used other data remaining in the data set, which contained all the attributes except the rainfall data that the model is supposed to predict. The test set was never used for the training of any of the models.

Table 1. Error Rate for Data Iris and Data Rainfall

Data	Learning Rate	Metode		
		Stochastic	Caputo	
Iris	0,01	0	0	
	Linear	0,001	3,9624E-88	3,9943E-88
		0,0001	6,5161E-10	6,5032E-10
		0,01	0	0
	Non Linear	0,001	6,6462E-88	6,7129E-88
		0,0001	1,093E-09	1,0931E-09
0,01		0,2739597	0,27444902	
Rainfall	Linear	0,001	0,27484916	0,2750229
		0,0001	0,27519991	0,27518994
		0,01	0,27313254	0,27313254
	Non Linear	0,001	0,273066	0,273066
		0,0001	0,2730784	0,27308246
		0,01	0,2730784	0,27308246

In experiments conducted using rainfall data, the program runs a maximum of 1000 iterations so that it can be seen in the iteration of how many SVM-SGD and SVM-Fractional GD reach convergence. SVM classifier uses stochastic gradient descent optimization with 0.01, 0.001, 0.0001 a learning rate. Fractional gradient descent as SVM classifier optimization is also given the same learning rate value.

The results of the SVM Classifier with stochastic gradient descent optimization reach the convergence point at iteration 350. With a learning rate of 0.0001 with error rate = 0.273078, at learning rate 0.001 with error rate = 0.273066, and at learning rate 0.01 with error rate = 0.273132. For the results of the SVM Classifier with fractional gradient descent optimization, it reaches a convergence point of approximately 50 iterations smaller than SVM-SGD. With a learning rate of 0.0001 with an error rate = 0.273083, a learning rate of 0.001 with an error rate = 0.273070, lastly with a learning rate of 0.01 with an error rate = 0.273134.

4. Conclusions

The purpose of this study is to investigate the current state of the optimization method based on the Fractional stochastic gradient descent conducted on the Support Vector Machine classifier. Based on the findings of this study, it is possible to make an overview of the dominant optimization method based on gradient descent. We model the model parameters by using the Fractional SGD algorithm to optimize the model parameters. For the results of the SVM Classifier with fractional gradient descent optimization, it reaches a convergence point of approximately 50 iterations smaller than SVM-SGD. With a learning rate of 0.0001 with an error rate = 0.273083, a learning rate of 0.001 with an error rate = 0.273070, lastly with a learning rate of 0.01 with an error rate = 0.273134. The SVM-Fractional SGD algorithm is proven to be an effective method for rainfall forecast decisions. The SVM-Fractional SGD algorithm is proven to be an effective method for rainfall forecast decisions. The process of updating w or fixing the model is smaller in fractional because the multiplier value is less than 1 or in the form of fractions. For future studies, it is possible to implement an optimization method based on the fractional-order derivative for other classifiers such as SVM for a sparse dataset.

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

- [8] Z. A. Khan, N. I. Chaudhary, and S. Zubair, "Fractional stochastic gradient descent for recommender systems," *Electron. Mark.*, 2019.
- [9] D. Chen, H. Sheng, Y. Q. Chen, and D. Xue, "Fractional-order variational optical flow model for motion estimation," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, 2013.
- [10] C. J. Zuñiga Aguilar, J. F. Gómez-Aguilar, V. M. Alvarado-Martínez, and H. M. Romero-Ugalde, "Fractional order neural networks for system identification," *Chaos, Solitons and Fractals*, 2020.
- [11] H. Wang, Y. Yu, G. Wen, S. Zhang, and J. Yu, "Global stability analysis of fractional-order Hopfield neural networks with time delay," *Neurocomputing*, 2015.
- [12] M. U. Akhmet and M. Karacaören, "A Hopfield neural network with multi-compartmental activation," *Neural Comput. Appl.*, 2018.
- [13] H. P. Hu, J. K. Wang, and F. L. Xie, "Dynamics analysis of a new fractional-order hopfield neural network with delay and its generalized projective synchronization," *Entropy*, 2019.
- [14] M. Caputo and M. Fabrizio, "A new definition of fractional derivative without singular kernel," *Prog. Fract. Differ. Appl.*, 2015.
- [15] A. I. Journal, M. Caputo, and M. Fabrizio, "Progress in Fractional Differentiation and Applications A new Definition of Fractional Derivative without Singular Kernel," *Progr. Fract. Differ. Appl*, 2015.
- [16] Y. Chen, Y. Wei, Y. Wang, and Y. Q. Chen, "Fractional order gradient methods for a general class of convex functions," in *Proceedings of the American Control Conference*, 2018.
- [17] Y. Wei, Y. Chen, S. Cheng, and Y. Wang, "Discussion on fractional order derivatives," *IFAC-PapersOnLine*, 2017.
- [18] J. Losada and J. J. Nieto, "Properties of a new fractional derivative without singular kernel," *Prog. Fract. Differ. Appl.*, 2015.
- [19] C. Bao, Y. Pu, and Y. Zhang, "Fractional-Order Deep Backpropagation Neural Network," *Comput. Intell. Neurosci.*, 2018.
- [20] S. Khan, J. Ahmad, I. Naseem, and M. Moinuddin, "A Novel Fractional Gradient-Based Learning Algorithm for Recurrent Neural Networks," *Circuits, Syst. Signal Process.*, 2018.