



Proceeding Paper

Machine Learning Explicit and Implicit Model for Factor Classification Based on Factor Space Theory [†]

Kaijie Zhang ^{1,*}, Fanhui Zeng ², Xiaotong Liu ¹, Kaile Lin ¹ and Ying Wang ¹

¹ College of Science, Liaoning Technical University, Fuxin 123000, China; liu_xiaotong2299@163.com (X.L.); 15525501190@163.com (K.L.); 15309877632@163.com (Y.W.)

² Institute of Intelligent Engineering and Mathematics, Liaoning Technical University, Fuxin 123000, China; zfh3351@sina.com

* Correspondence: 19824855218@163.com

[†] Presented at the 2023 Summit of the International Society for the Study of Information (IS4SI 2023), Beijing, China, 14–16 August 2023.

Abstract: Factor implicit thought in factor space theory can be used to solve the classification problem of machine learning. On the basis of factor implicitness, a serial sweeping class classification algorithm is proposed, and a factor implicit model is constructed with this algorithm, and then tested and classified. On the basis of a serial scanning algorithm, this paper proposes a fine-tuning Sweeping Learning Algorithm, an ascending Side-By-Side Serial Scanning Algorithm, and a combination algorithm to solve the two problems in the running of the serial scanning algorithm. In order to solve the problem that the training speed of a traditional serial scanning algorithm is too slow due to the large amount of data, a new method which can select mixed domains in advance—the partial-side serial scanning algorithm.

Keywords: factor space; factor implicit; serial scanning algorithm; support vector machine; side-by-side serial scanning algorithm; support vector machine based on sweeping class chain algorithm; factor support vector machine



Citation: Zhang, K.; Zeng, F.; Liu, X.; Lin, K.; Wang, Y. Machine Learning Explicit and Implicit Model for Factor Classification Based on Factor Space Theory. *Comput. Sci. Math. Forum* **2023**, *8*, 88. <https://doi.org/10.3390/cmsf2023008088>

Academic Editors: Zhongzhi Shi and Wolfgang Hofkirchner

Published: 4 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the core of artificial intelligence, machine learning plays a very important role in the realization of computer intelligence. With the continuous innovation of science and technology and the rapid development of computer network technology, machine learning has become an increasingly important part of the field of artificial intelligence, and the classification problem in machine learning is also one of the main tasks of machine learning, which is widely used in various fields of real life, and the accuracy rate of the goals it can achieve is getting higher and higher. Among them, the binary classification problem is also an important part of machine learning; whether it is medical, agriculture or daily production and life, the problem of binary classification is everywhere. Looking for a more accurate algorithm to solve the binary classification problem is an important research direction in the field of artificial intelligence.

The advent of the information revolution and the era of big data promoted the development of artificial intelligence, followed by the need to find how to save, extract, and process the required factor data for huge data. Causal analysis between factors provides an important tool for artificial intelligence, data mining, etc. However, the main difficulty in the field of artificial intelligence is that the key factors to solve practical problems have not been revealed, and how to find the key factors has become an important research direction. Factor explicitness is a new theory under the factor space theory proposed by Wang Peizhuang [1]. As a bottleneck problem in the field of artificial intelligence, factor explicitness has great significance in helping artificial intelligence problems find the key factors. As long as the key factors are found, the corresponding problems will naturally be

solved. Sun Hui et al. [2] proposed a serial sweep algorithm. Aiming at the classification problem of machine learning, the algorithm defines the sweeping direction and the explicit and implicit factors by using the factor space theory. In order to reduce algorithm's complexity, the ordered set of swept class vectors is defined, and the factor implicit model is constructed. The results of numerical experiments show that the algorithm is feasible and effective. Zeng Fanhui et al. [3] proposed the application of the serial scanning algorithm in multi-classification. On the basis of this algorithm, this paper proposes a fine-tuning sweeping learning algorithm, a dimension-raising side-by-side serial scanning algorithm, and a combination algorithm [4] to solve the two problems in the operation of the serial sweep classification algorithm.

2. Sweeping Learning Algorithm

(1) Algorithm steps:

Algorithm steps:

- 1: input $S^- = \{x_1^-, \dots, x_l^-\}; S^+ = \{x_1^+, \dots, x_j^+\};$
 - 2: $w = o^+ - o^-;$
 - 3: $l = \max\{x_l^-, w\}; u = \min\{x_1^+, w\}; r = (u - l)/2; o = (u + l)/2;$
 - 4: if $l < u$, then go to Step 3; otherwise:
 - $S^- = S^- - \{x_l^- \in S^- \mid (x_l^-, w) < u\};$
 - $S^+ = S^+ - \{x_j^+ \in S^+ \mid (x_j^+, w) > l\};$
 - go back 2.
 - 5: for each deleted x_l^- from S^- , if $l < (x_l^-, w) < u$, then $S^- = S^- + \{x_l^-\};$
 for each deleted x_j^+ from S^+ , if $l < (x_j^+, w) < u$, then $S^+ = S^+ + \{x_j^+\}.$
 go back 2.
 - 6: if x is all deleted from $S^- + S^+$, (x, w) never enter into (l, u) ; then, output w , which can divide two classes' points.
-

(2) Fine-tuning algorithm step

Fine-tuning algorithm step

- 1: calculate the sum of two types of centers u_i^- and u_i^+ and obtain $w_i;$
 - 2: take any integer $t = [0, T)$; it is best to select the serial classification algorithm before sweeping into infinite iteration; then, adjust the displacement of the remaining positive and negative mixed points class X_{t+1}^+, X_{t+1}^- to fine-tune the positive and negative class data set.
 - 3: use linear discriminant analysis (LDA) to solve $w^*, w_{t+1} = w^*.$
 - 4: starting from $t = t + 1$, calculate solve, stop, and output the explicit and implicit factors by using the scanning serial classification algorithm.
-

(3) Steps of dimension-raising algorithm

Steps of dimension-raising algorithm

- 1: calculate the sum of two types of centers u_i^- and u_i^+ and obtain $w_i;$
 - 2: Take any integer $t = [0, T)$; it is best to select the serial classification algorithm before sweeping into infinite iteration; then, adjust the displacement of the remaining positive and negative mixed points class X_{t+1}^+, X_{t+1}^- to fine-tune the positive and negative class data set.
 - 3: use the kernel function to raise the dimension of the fine-tuned positive and negative data set, and use the corresponding formula to perform projection calculation.
 - 4: Starting from $t = t + 1$, calculate, solve, stop, and output the explicit and implicit factors using the sequential classification algorithm of sweeping classes.
-

(4) Example application

The data of Sweeping Learning Algorithm are derived from the Cryotherapy data set, Immunotherapy data set and Somerville Happiness in the UCI database based on the survey data set; Matlab software is used to show that the serial scanning algorithm has a better classification effect compared with a support vector machine, and the experimental results also show the feasibility and practicability of the serial scanning algorithm. The

experimental results, such as related data set physical properties and experimental results, are shown in Tables 1–4.

Table 1. Attributes for the two data sets.

| Data Set | Sample Number | Number of Categories | Characteristic Number |
|---------------------------------|---------------|----------------------|-----------------------|
| Cryotherapy | 90 | 2 | 6 |
| Immunotherapy | 90 | 2 | 7 |
| The Somerville Happiness Survey | 143 | 2 | 6 |

Table 2. Performance comparison of Cryotherapy data sample set.

| Contrast Term | Number of Training Steps | Number of Samples | Accuracy % |
|---------------|--------------------------|-------------------|------------|
| SVM | 74 | 9 | 100% |
| sweep | 74 | 9 | 100% |

Table 3. Performance comparison of Immunotherapy data sample set.

| Contrast Term | Number of Training Steps | Number of Samples | Accuracy % |
|---------------|--------------------------|-------------------|------------|
| SVM | 80 | 10 | 90% |
| sweep | 80 | 10 | 90% |

Table 4. Performance comparison of the Somerville Happiness Survey data sample set.

| Contrast Term | Number of Training Steps | Number of Samples | Accuracy % |
|---------------|--------------------------|-------------------|------------|
| SVM | 133 | 10 | 70% |
| Sweep | 133 | 10 | 100% |

3. Side-By-Side Serial Scanning Algorithm

(1) Algorithm steps:

Algorithm steps:

- 1: input $S^- = \{x_1^-, \dots, x_l^-\}; S^+ = \{x_1^+, \dots, x_j^+\};$
- 2: $w = o^+ - o^-; l = \max \{(x_l^-, w)\}; u = \min \{(x_1^+, w)\}; r = (u - l)/2; o = (u + l)/2;$
- 3: if $l < o < u$, then goto 4; Else
 $S^- = S^- - \{x_l^- \in S^- \mid (x_l^-, w) < u\};$
 $S^+ = S^+ - \{x_j^+ \in S^+ \mid (x_j^+, w) > l\};$
 go back 2.
- 4: for each deleted x_l^- from S^- , if $l < (x_l^-, w) < u$, then $S^- = S^- + \{x_l^-\};$
 for each deleted x_j^+ from S^+ , if $l < (x_j^+, w) < u$, then $S^+ = S^+ + \{x_j^+\};$
 go back 2.
- 5: if x is all deleted from $S^- + S^+$, (x, w) never enter into (l, u) ; then, output w , which can divide two classes' points.

Attachment: In the process of program calculation, the size of the sample centers of the positive and negative class data will change in actual situations, so it is difficult to maintain the positive class center above the negative class center; thus, it is necessary to consider the position transformation of the two types of centers. When it appears below, the calculation method of class vector scanning is: $w_t = -o_t^+ + o_t^-$.

(2) Example application:

Using the data sets of Iris and Haberman’s Survival UCI, SVM, sweep, s-sweep, and matlab2018a coding were used for multiple experimental comparisons. In the experimental process, 80% of the data sets were randomly selected for testing and the remaining 20% were verified. The corresponding attributes of the six data sets shown in Tables 5 and 6 are the average values of multiple experimental results.

Table 5. Attributes for the two data sets.

| Data Set | Sample Number | Number of Categories | Characteristic Number |
|---------------------|---------------|----------------------|-----------------------|
| Iris | 150 | 3 | 4 |
| Haberman’s Survival | 306 | 2 | 3 |

Table 6. Performance comparison on Iris data sample set.

| Contrast Term | Number of Training Steps | Number of Samples | Training Time/ms | Accuracy % |
|---------------|--------------------------|-------------------|------------------|------------|
| SVM | 100 | 50 | 259.156 | 96.5% |
| sweep | 100 | 50 | 7.369 | 100% |
| s-sweep | 100 | 50 | 1.987 | 100% |

4. Summary and Prospects

This paper solves two problems in the operation of a serial class scanning algorithm and puts forward corresponding processing methods. By contacting the vertical bisector of the class scanning vector, the sample points are retained according to the distance between the projection distance of the sample points in the class scanning vector and the vertical bisector, and a new training sample is obtained. New training samples are used to replace original training samples for traditional serial scanning training, as shown in Table 6. This algorithm can reduce the training samples without affecting the classification ability of the class scanning vector and speed up the data-processing process. Meanwhile, when the positive class center is smaller than the negative class center in the process of class scanning, the influence of the obtained class scanning vector on the data prediction and classification is also a problem. At the same time, the simulation results show that the algorithm is effective and feasible, and it has faster data-processing ability and accuracy than a traditional SVM algorithm and serial scanning algorithm.

Author Contributions: There are five authors in this paper. K.Z. provided the algorithm and software coding for verification analysis and writing of the paper; F.Z. provided the guidance for writing and preparing the first draft; X.L., K.L. and Y.W. reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted by Fuxin Social Science Project: Mathematical Basic Research of Factor Space Theory under the background of Fuxin Business Environment Optimization, project number: 2022Fslx111; Basic Scientific Research Project of colleges and universities of Liaoning Provincial Department of Education, key research project: Theory and Application Research of Factor Space-based intelligent incubation under the Digital Background, project number: funded by LJKZZ20220047.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The experimental data in this paper are derived from the data set of UCI machine learning repository.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, P.Z.; Zeng, F.H. *Factor Space Theory, Mathematical Basis of Unified Intelligence Theory*; Science Press: Beijing, China, 2023.
2. Sun, H. Sweeping Chain Learning Algorithm. Master's Thesis, Liaoning Technical University, Fuxin, China, 2022.
3. Zeng, F.H.; Wang, Y.; Wang, P.Z.; Sun, H. Multi-classes sweep learning based on factor space theory. *J. Liaoning Tech. Univ. Nat. Sci.* 2022, *accepted*.
4. Zhang, K.; Zeng, F. Sweeping class linkage algorithm in factor space. *J. Intell. Syst.* 2023, *submitted*.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.