Antenna Optimization Based on Co-Training Algorithm of Gaussian Process and Support Vector Machine

JING GAO, YUBO TIAN, AND XUEZHI CHEN
School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China
Corresponding author: Yubo Tian (tianyubo@just.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61771225, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province China under Grant KYCX20-3141, in part by the Natural Science Foundation of Jiangsu province of China under Grant BK20190956, and in part by the Qinglan Project of Jiangsu Higher Education.

ABSTRACT For the optimal design of electromagnetic components, surrogate model methods can usually be used, but obtaining labeled training samples from full-wave electromagnetic simulation software is most time-consuming. How to use relatively few labeled samples to obtain a relatively high-precision surrogate model is the current electromagnetic research hotspot. This article proposes a semi-supervised co-training algorithm based on Gaussian process (GP) and support vector machine (SVM). By using a small number of initial training samples, the initial GP model and initial SVM model can be trained by some basic parameter settings. Moreover, the accuracy of these two models can be improved by using the differences between these two models and combining with unlabeled samples for jointly training. In the co-training process, to ensure the performance of the proposed algorithm, a stop criterion set in advance to control the number of unlabeled samples introduced. Therefore, the accuracy of the model can be prevented from being reduced by introducing too much unlabeled samples, which can find the best solution in the limited time. The proposed co-training algorithm is evaluated by benchmark functions, optimal design of Yagi microstrip antenna (MSA) and GPS Beidou dual-mode MSA. The results show that the proposed algorithm fits the benchmark functions well. For the problem of resonant frequency modeling of the above two different MSAs, under the condition of using the same labeled samples, the predictive ability of the proposed algorithm is improved compared with the traditional supervised learning method. Moreover, for the groups of antenna sizes that meet the design requirements, the fitting effects of their return loss curve (S_{11}) are well. The effectiveness of the proposed co-training algorithm has been well verified, which can be used to replace the time-consuming electromagnetic simulation software for prediction.

INDEX TERMS Semi-supervised learning (SSL), Gaussian process (GP), support vector machine (SVM), co-training, antenna optimization.

I. BASIC INTRODUCTION

In the optimization fields of electromagnetic components, it is common to use numerical simulation calculation or full-wave electromagnetic simulation software such as high frequency structure simulator (HFSS), computer simulation technology (CST) combined with global optimization algorithms [1], [2]. As we know, inverse electromagnetic problem is a hotspot in the field of computational electromagnetics [3], [4]. From the perspective of applications, this problem can be divided into two categories: parameter identification problem and optimization design problem. The latter has been studied more in our research group, which the essence is to give the desired performance index of an electromagnetic system, achieving this goal by optimizing the parameters. It is also a comprehensive problem, according to the performance requirements of electromagnetic devices, so as to synthesize the structure and size of electromagnetic devices for antenna optimization.

According to the previous researches, HFSS simulation can be used to obtain high-precision results to acquire labeled samples for many times. However, if the structure of the

Received October 2, 2020, accepted November 15, 2020, date of publication November 19, 2020, date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039269
microwave device is complex, each call will require a lot of time. Therefore, the calculation cost is high, and the time needed is long. The usage of modeling methods instead of calling HFSS for analysis can effectively save time and has been a hot topic. Popular modeling methods such as artificial neural network (ANN) [5], [6], support vector machine (SVM) [7], [8], kernel extreme learning machine (KELM) [9], [10], Gaussian process (GP) [11], [12] et al., have been achieved.

Electromagnetic problems are generally small sample ones, both GP and SVM modeling methods are widely used in antenna optimization. GP is a machine learning (ML) method gradually developed in recent years. It has a strict statistical theoretical basis and is suitable for solving problems such as small samples, high dimensionality and nonlinearity. SVM is also a common ML method, which has unique advantages in solving problems such as small samples, nonlinear and high-dimensional modes. Resonant frequency is an important technical index in the optimal design of antennas. Obtaining resonant frequency quickly by known structural parameters of the antenna is often used in the research of modern antenna design [13], [14]. Both GP model and SVM model are widely used in antenna resonant frequency modeling [15], [16]. The trained model can establish a mapping relationship between the antenna-related parameters and the measured resonant frequency, so as to predict the resonant frequencies of other antenna parameters, reducing the number of calls to HFSS for accurate results.

The existing modeling of electromagnetic behavior is based on supervised learning, and the labeled training samples used are simulated by the simulation software such as HFSS [17], [18] et al. The frequency of full-wave electromagnetic simulation is the main factor affecting the training efficiency of the model such as GP and SVM mentioned above. Therefore, based on the existing researches, semi-supervised learning (SSL) [19], [20] method is proposed, so as to obtain the satisfied accuracy in a short time. Traditional ML techniques rely on a large number of labeled samples or unlabeled samples for training. In practical applications, it is difficult to obtain labeled samples, while unlabeled samples are easier to be obtained. Considering that unlabeled samples and labeled samples are usually distributed independently and identically, SSL method breaks through the limitations of just considering one type of samples, so as to mining hidden information of unlabeled samples to assist labeled samples for training. It is mainly divided into semi-supervised clustering [21], semi-supervised classification [22], semi-supervised dimensionality reduction [23], and semi-supervised regression [24], [25]. Co-training is a commonly used SSL method based on divergence. It has a solid theoretical foundation and a wide range of applications [26], [27]. The co-training algorithm was first proposed by A. Blum and T. Mitchell in 1998 [28], which has been continuously developed and gradually penetrated into many fields, such as natural language processing [29], image retrieval [30] et al. In the field of electromagnetism, to the best of our knowledge, there is no relevant researches, which is also the reason why this subject is worth researching.

Traditional co-training algorithm focuses on classification problems [31], lacking of researches on regression problems. Therefore, this study improves the traditional co-training method and proposes a co-training method based on GP and SVM, applying SSL algorithm to the field of electromagnetic optimization. Differences between GP model and SVM model have been utilized, and these two models use the same unlabeled samples to generate pseudo-labeled samples for updating. During the iteration process, the termination conditions have been set. If the test error of the next iteration is higher than that of the previous iteration and the test error of the previous iteration has reached the error threshold, the iteration will stop. Hence, the reduction of model accuracy can be prevented due to too many unlabeled samples introduced. In the cases study section, two different benchmark functions, and optimal design of Yagi microstrip antenna (MSA) and GPS Beidou dual-mode MSA are used to evaluate the effectiveness of the proposed co-training algorithm. The results show that the proposed co-training algorithm fits the benchmark functions well. For the experiments of two different antennas optimization, the proposed co-training algorithm has better predictive ability than that of traditional supervised learning method by using the same label samples.

II. CO-TRAINING ALGORITHM
Co-training algorithm uses the differences between two different models, improving the performances of the model by introducing unlabeled samples. This article improves the traditional co-training method to make it more suitable for antenna optimization.

A. GAUSSIAN PROCESS
GP describes the covariance of predicted data by the covariance of input data. The parameters in the kernel function are called hyper-parameters. The process of model training is the process of selecting the kernel function and determining the hyper-parameters. The network link for software package with relevant instructions for GP modeling is shown in reference [32].

GP is a set consisting of countless random variables and any subset is in accordance with Gaussian distribution, mean function \( m(x) = E[f(x)] \) and covariance function \( k(x, x') E[\{f(x) - m(x)\}\{f(x') - m(x')\}] \) determine the properties of GP, defined as

\[
f(x) \sim GP(m(x), k(x, x'))
\]

(1)

In the formula, \( x, x' \in X \) is any d-dimensional vector. The observed value is polluted by additive noise \( \epsilon \), which is a normally distributed random variable with a mean of 0 and a variance of \( \sigma^2_n \), that is

\[
\epsilon \sim N(0, \sigma^2_n)
\]

(2)
x is the input vector and the observation value polluted by noise. The obtained prior distribution is
\[
y \sim N \left( 0, K + \sigma^2_n I \right)
\]
(3)
n training sample outputs y and n* test sample outputs \( \tilde{y} \) form a joint Gaussian prior distribution, that is
\[
\begin{bmatrix} y \\ \tilde{y} \end{bmatrix} \sim N \left( 0, \begin{bmatrix} K(X, X) + \sigma^2_n I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right)
\]
(4)
where \( K(X, X^*) \) is covariance matrix of order \( n \times n^* \) between \( n^* \) test output samples and \( n \) training output samples, \( K(X^*, X^*) \) is the covariance matrix of order \( n^* \times n^* \) for the output sample itself.

The optimal hyper-parameters are obtained by maximum likelihood estimation. By establishing the log-likelihood function of the conditional probability, the derivate of hyper-parameters can be calculated. The conjugate gradient optimization method is used to search for the optimal hyper-parameters, and the negative log-likelihood function is expressed as
\[
\mathcal{L} = \log p(y|X) = -\frac{1}{2} y^T K^{-1} y - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi
\]
(5)

After the optimal hyper-parameters are obtained, predictions can be done. Given the new input \( x^* \), the input value \( X \) of the training sample set and the observation target value \( y \), the maximum possible predicted posterior distribution can be inferred and given by
\[
y^* | x^*, X, y \sim N(m, \sum) \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \ quad
The corresponding dual problem is expressed by

$$
\max = L_d (\alpha_i, \alpha_i^*) - \varepsilon \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*) \phi^T (x_i) \phi (x_j)
$$

where \( s.t. : \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i \leq C, i = 1, \ldots, N, \) and \( 0 \leq \alpha_i^* \leq C, i = 1, \ldots, N. \)

Therefore, \( w = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) x_i \) can be gotten. Introducing the kernel function, it is

$$
f(x) = w \cdot x = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K_{SVR} (x, x_i) + b
$$

where \( N \) is the number of support vectors and determines the complexity of the structure.

### C. CO-TRAINING ALGORITHM

The traditional co-training algorithms are actually iterative ones, which need significant differences for updating each other. If the number of iterations is too large, the number of unlabeled samples introduced increases, which leads to the accumulation of noise, and the performance of the model will be reduced. Therefore, in view of the above problems, this article proposes a semi-supervised co-training algorithm based on GP model and SVM model. Using less labeled samples for initial training, and then unlabeled samples are combined with labeled samples for updating each other. A termination condition is set in advance, which is if the test error of the next iteration is higher than that of the previous iteration and the test error of the previous iteration has reached the set error threshold, the iteration will stop. According to the above method, the number of unlabeled samples introduced can be effectively controlled, preventing reducing the accuracy of the model by introducing too many unlabeled samples. Moreover, a satisfied solution can be got in a short time.

The co-training algorithm proposed in this article mainly includes initial training and co-training. In the initial training process, a small number of labeled samples simulated by HFSS software are used as initial training samples, then initial GP model and initial SVM model are trained respectively by these samples. In the co-training process, the two initial models use the same unlabeled samples to generate pseudo-labeled samples, which are used to cross-train the GP model and SVM model and update these two models, denoted as GP\(_{time}\) and SVM\(_{time}\). The same test sample set simulated by HFSS is used to test the GP\(_{time}\) and SVM\(_{time}\). Their test errors are compared, then the pseudo-labeled samples used in the model with smaller error and the test sample corresponding to the number of iterations in the test sample set are all added to the initial training sample set to further train GP model and SVM model. Repeat the iterative process continuously until the above stop criterion is met. The test samples added to the training sample set are removed from the test sample set, and the remaining test samples form the test sample set for the next iteration. Fig.1 is a flowchart of the proposed algorithm, and Table 1 is the pseudo-code of the proposed algorithm.

![Flow chart of the proposed co-training algorithm.](image)

### III. CASES STUDY

Firstly, two benchmark functions have been used to verify the effectiveness of the proposed co-training algorithm, and then Yagi microstrip antenna (MSA) and GPS Beidou dual-mode MSA have been optimized by the proposed algorithm. The experimental results of resonant frequency modeling of the two MSAs are used as the basis for judging its generalizability. Moreover, the optimal trained co-training model is used to fit the return loss curve (S\(_{11}\)) of one group of antenna size that meets the design requirements. The S\(_{11}\) curve predicted by this model is compared with that of simulation by HFSS for verifying the performance of the proposed algorithm.

### A. BENCHMARK FUNCTIONS

In this study, the Griewank function and Quartic function are used firstly to test the performance of the proposed method, and their formulas are as (18)-(19). Both these two functions are set to 3 dimensions, the independent variable value interval is \([-5, 5]\). According to the complexity of functions and the initial errors, the error threshold is set to 1e-02, which is a well precision. Moreover, the maximum number of iterations is set to 100. At each iteration, the Relative Error (RE) is used for judging one test sample, and the Mean Relative
TABLE 1. Pseudo code of the proposed co-training algorithm.

| Algorithm: Co-training Algorithm of GP and SVM |
|-----------------------------------------------|
| 1 begin                                       |
| 2 \(e_1 \leftarrow 1, e_2 \leftarrow 1, t \leftarrow 0\) // assign initial value |
| 3 while min \((e_1, e_2)\) > error threshold do |
| 4 \(t \leftarrow t + 1\)                          |
| 5 train the GP model and SVM model by the initial training |
| samples simulated by HFSS, denoted as GP and SVM      |
| 6 select \(N_t\) samples \(X\) from the unlabeled sample set |
| 7 input \(X\) into GP                             |
| 8 output Co.GP                                     |
| 9 select the same \(N_t\) samples \(X\) from the unlabeled sample set |
| 10 input \(X\) into SVM                             |
| 11 output Co.SVM                                    |
| 12 use Co.GP to future train SVM, denoted as SVM_{true} |
| 13 use Co.SVM to future train GP, denoted as GP_{true} |
| 14 use test sample set \(\text{Test.} \, G\) by HFSS simulation to test the |
| \(\text{GP}_{\text{true}}\) and SVM_{true}               |
| 15 output \(e_1\) and \(e_2\)                      |
| 16 if \(e_1 > e_2\) then add Co.GP and the \(i\)th test sample \(\text{Test.} \, G_i\) into the |
| training sample set to future train GP, SVM         |
| 18 else                                           |
| 19 then add Co.SVM and the \(i\)th test sample \(\text{Test.} \, G_i\) into |
| training sample set to future train GP, SVM         |
| 20 end                                             |
| 21 end                                             |
| 22 end                                             |

Error (MRE) is used for judging the whole test sample set, formulas follow as (20)-(21).

\[
f(x) = \frac{d}{4000} \sum_{i=1}^{d} x_i^2 - \prod_{i=1}^{d} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \quad (18)
\]

\[
f(x) = \sum_{i=1}^{D} i x_i^4 + \text{random}(0, 1) \quad (19)
\]

\[
RE = \frac{|y_{\text{pred}} - y_{\text{test}}|}{y_{\text{test}}} \quad (20)
\]

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_{\text{pred}} - y_{\text{test}}|}{y_{\text{test}}} \quad (21)
\]

where \(y_{\text{pred}}\) is the label value predicted by the proposed semi-supervised co-training model, and \(y_{\text{test}}\) is the true label value of the test sample.

According to the independent variable value interval \([-5, 5]\), 200 training samples can be generated randomly to train the initial GP model and SVM model, and then conduct co-training process. During each iteration, 50 test samples are randomly generated for testing in the interval \([-5, 5]\).

After training, in order to further verify the effectiveness of the proposed algorithm, the optima co-training model is used to predict the output of other 50 randomly generated test points, and the predicted outputs are compared with the real outputs. Table 2 records the iteration numbers and the termination test error of these two benchmark functions. Fig.2 and Fig.3 are the iterative test results and the fitting effects of the test points. It can be seen from the error curves that the Griewank function reaches the error threshold at 45 times, and the Quartic function reaches the error threshold at 76 times, all of which have reached the threshold within 100 iterations. Moreover, it can be seen from the fitting curves that the fitting effect of the Griewank function is slightly better than that of the Quartic function, but the fitting effects of these two benchmark functions have all reached good levels. Therefore, according to these experiment results, the effectiveness of the proposed co-training algorithm has been initially verified.

B. YAGI MICROSTRIP ANTENNA

Yagi MSA has high gain, wide beam width, which has been widely used [35]. The structure diagram of the Yagi MSA is shown in Fig.4 (a), and the three-dimensional view in HFSS is shown in Fig.4 (b). The model is fed with gradual microstrip Barron. It has a lot of variables, such as the length of the reflective array, the length and width of the excitation array, the length and width of the guidance array and so on. The parameters that have a large impact on performance are selected, namely the length of the excitation array \(d_r\), the distance between the excitation array and the reflection array \(g_1\), the distance between the excitation array and the guide array \(g_2\), and the distance between the guide array \(g_3\) as input parameters.

1) RESONANT FREQUENCY MODELING

The design indexes of the Yagi MSA are working frequency 2.3GHz, and -10dB bandwidth covering 2.2GHz ~ 2.4GHz. The input parameter combination is \(y = [d_r, g_1, g_2, g_3]\), and \(d_r \in [40, 45], g_1 \in [15, 20], g_2, g_3 \in [8, 14], g_2 = g_3\), other fixed values of size parameters are shown in Table 3.

| TABLE 2. Iterative results of the benchmark functions. |
|-------------------------------------------------------|
| **Griewank function** | **Quartic function** |
| **Iteration number** | 45 | 76 |
| **Iteration termination error** | 0.0062 | 0.0074 |
The frequency sweep range of Yagi MSA is 1GHz $\sim$ 5GHz with a step size of 0.04GHz. Therefore, there are 101 frequency points in total for a group of parameters. Four variables are set in total with 6 levels for each variable, and the value ranges of parameters and sampling intervals of these 4 variables are shown in Table 4. For simulation,
the HFSS-MATLAB-API script [36] has been used. The HFSS software is controlled through the script interface to generate 3D model for analysis and solution. Finally, the simulation results are output. In other words, the scripts programmed in MATLAB are used to call HFSS software, taking the variable parameters of the antenna as input and the antenna return loss as output.

By using the partial orthogonal experiment, 16 groups of samples are simulated as initial training samples, another 10 groups of samples are simulated as test sample set, and 10 groups of samples are randomly set without HFSS simulation as unlabeled sample set. The computer processor used in this experiment is Intel(R) Core (TM) i5-10210U CPU @ 1.60GHz 2.11GHz, RAM is 8GB. It takes 104.79 seconds to obtain a set of labeled samples by HFSS. After simulation, the resonant frequency points corresponding to each group of antenna sizes can be got, and then the modeling experiments can be conducted.

Using the 16 initial training samples, taking \(d_r, g_1, g_2\), and \(g_3\) as input variables and \(f_{HFSS}\) as output, so as to establish the GP model and SVM model respectively. Test sample set is used to test these two models, and the initial errors of these two models are shown in Table 5. In each iteration, the GP and SVM take one unlabeled sample for cross-training. The test sample set is used to test the two models respectively, and the test error is evaluated by MRE. Following, the values of the two errors are compared. The smaller one is added as pseudo-label sample and the \(i^{th}\) test sample to the training sample set, further training the GP model and SVM model. At the same time, the test sample added to the training set will be deleted from the test sample set, and the remaining test samples are for the next iteration. Based on the magnitude of the initial error and the overall error, the error threshold of this experiment is 1e-03, which is a relatively small error. Table 6 records the test error of each iteration.

![FIGURE 4. The Yagi MSA.](image)

It can be seen from the results that the iteration ends at the 9th times, and the test error at the 8th iteration is the smallest, which is 0.0027. At this time, 7 labeled test samples are introduced to the training sample set. To ensure the fairness of the comparison, the updated training sample set including 23 samples are also, respectively, used to train the traditional supervised GP model and SVM model. These two models are tested by the same test sample set used in the \(8^{th}\) iteration of the co-training process. The test errors are 0.0050 and 0.0110, which are both greater than 0.0027. In this experiment, 23 sets of labeled samples are used, and the total computing time is 2410.17 seconds. Therefore, the prediction ability of the proposed co-training model is improved compared with the traditional supervised learning model by using the same label samples, namely the same computing time.

2) RETURN LOSS FITTING

The experimental results of resonant frequency modeling are used as the basis for judging the generalization ability of the trained model. In order to more intuitively reflect the fitting ability of the proposed model, a size conforming to the design standards is used for fitting the corresponding S\(_{11}\) curve. According to the optimization results, a set of antenna parameters whose performance satisfies the design index is [45, 18, 14, 14].

As shown in Fig.5, the X-axis is the frequency sweep range, which is from 1GHz to 5GHz. The Y-axis is the corresponding S\(_{11}\) parameter value of each frequency point. The S\(_{11}\) reaches \(-17.5dB\) at 2.3GHz and \(-10dB\) bandwidth covering 2.2GHz \(~\sim\) 2.4GHz, which meeting the design requirements. In order to further verify the effectiveness of the proposed co-training algorithm, the trained co-training model is used to predict the S\(_{11}\) of this group of antenna parameters, and compare them with these simulated by HFSS software. The solid blue line named ‘HFSS’ is the simulation results of HFSS, and the dotted red line named ‘Proposed’ is the
TABLE 6. Iterative results of Yagi MSA.

| Iteration number | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Test error       | 0.0069| 0.0068| 0.0069| 0.0067| 0.0057| 0.0055| 0.0047| 0.0027| 0.0037|

TABLE 7. Fixed size parameters of the GPS Beidou dual-mode MSA.

| Variable name                                      | Variable value/mm | Variable name                                      | Variable value/mm |
|----------------------------------------------------|-------------------|----------------------------------------------------|-------------------|
| Thickness of square dielectric substrate H          | 1.6               | Distance from feed point to y-axis X₀              | 6.75              |
| Side length of square dielectric substrate B        | 60                | Thickness of circular ground plate H₀              | 4                 |
| Width of radial branch width W₁                    | 4                 | Radius of circular ground plate R₀                 | 50                |
| Distance from feed point to x-axis Y₀              | 6.75              |                                                     |                   |

FIGURE 5. \( S_{11} \) fitting diagram of the Yagi MSA.

prediction results of the method proposed in this article. From the Fig. 5, the two curves are highly consistent, indicating that the proposed model is with high accuracy. This proposed model can be used to predict other parameters, replacing HFSS simulation effectively.

C. GPS BEIDOU DUAL-MODE MICROSTRIP ANTENNA

GPS Beidou dual-mode MSA can be used in GPS positioning systems and Beidou satellite navigation systems, which is widely used in various terminals [37]. In this experiment, a square MSA is used. The flat structure is shown in Fig.6 (a), and the three-dimensional model in HFSS is shown in Fig.6 (b). The four sides of the patch are branches with the same width and different lengths, corresponding to the two working modes of GPS L1 frequency band and Beidou B1 frequency band. Its performance is affected by the radiation patch side length \( W \), the low-frequency modal branch \( L₁ \) and the high-frequency modal branch length \( L₂ \), and these three antenna parameters are used as input variables to establish models.

1) RESONANT FREQUENCY MODELING

The design specification of the GPS Beidou dual-mode MSA is that the voltage standing wave ratio of the antenna at 1.58GHz (Beidou B1 operating frequency) and 1.61GHz (GPS L1 operating frequency) is less than or equal to 1.5. Several variables that have a significant impact on the performance are selected. The input parameter combination is \( \gamma = [W \, L₁ \, L₂] \), and \( W \in [42, 45] \), \( L₁ \in [5.1, 6.3] \), \( L₂ \in [3, 3.9] \). The values of other dimension parameters are fixed, which are shown in Table 7. For the simulation of the GPS Beidou dual-mode MSA, the frequency sweep range...
is 1.2GHz through 1.8GHz, and the step size is 0.005GHz. Therefore, there are 121 frequency points in total for each group of parameters. Each variable is set with 4 levels. The samples are also selected by using the partial orthogonal experiment. The sampling ranges and intervals of the three parameters are shown in Table 8.

By using the HFSS-MATLAB-API script, the input parameters of each set of antenna sizes correspond to a set of return loss values by electromagnetic simulation. In this experiment, 15 groups of samples are simulated by HFSS as the initial training samples, 10 groups of samples are simulated as the test samples, and other 10 groups of samples are randomly set without HFSS simulation as unlabeled samples. Through simulation, it takes 72.93 seconds for calling HFSS simulation to obtain one labeled sample. After the simulation, the resonant frequency points corresponding to each group of antenna parameters can be computed. Since two different operating modes have two different resonant frequency points, considering the accuracy of modeling, the two operating modes are modeled respectively by the proposed co-training algorithm.

Using 15 training samples and taking $W$, $L_1$, $L_2$ as input variables and $f_{HFSS}$ as output, the GP model and SVM model can be trained respectively. Table 9 records the initial errors of these two models. In each iteration, GP and SVM take one same unlabeled sample for cross-training, and the same test sample set is used for testing the two models respectively. The values of test errors are compared in each iteration, the pseudo-labeled sample with small error and the $i$th test sample are introduced into the training sample set to further train the GP model and SVM model. In this experiment, according to the magnitude of the initial errors of both two operating modes and the overall error trend, the error threshold is both set as 1e-04, which is a relatively small error. Table 10 records the iterative test errors of each iteration for the two models.

For the GPS mode, the iteration ends at the 9th times and the test error at the 8th iteration is the smallest, which is $8.3243 \times 10^{-4}$. At this time, 7 labeled test samples have been introduced. Considering the fairness of comparison, the updated training set including 22 samples is used to train traditional supervised GP model and SVM model. The test sample set used in the 7th iteration of the co-training algorithm is also used to test the above two models, and the test errors are $8.1689 \times 10^{-4}$ and $0.0065$ respectively, which are both greater than $6.1237 \times 10^{-4}$.

According to the above experimental results of both two operating modes, based on the same labeled samples, the predictive ability of the proposed co-training model is improved compared with that of the traditional supervised learning models. In order to ensure that the modeling accuracy of the models can all reach the high levels, the model in the 8th iteration is adopted as the best one in the experiment of GPS Beidou dual-mode MSA, which cost 1604.46 seconds totally.

2) RETURN LOSS FITTING

The results of the resonant frequency modeling experiment are used as the basis for judging the stability of the trained model. The trained model is used to describe the relationship between $W$, $L_1$, $L_2$ and $S_{11}$. A group of antenna size used for $S_{11}$ fitting is [44, 5.1, 3.6], conforming to the design requirements. Table 10 records the iterative test errors of each iteration for the two models.

For the GPS mode, the iteration ends at the 9th times and the test error at the 8th iteration is the smallest, which is $8.3243 \times 10^{-4}$. At this time, 7 labeled test samples are introduced. Considering the fairness of comparison, the updated training set including 22 samples is used to train traditional supervised GP model and SVM model respectively. The test errors are 0.0013 and 0.0124 respectively by the same test sample set used in the 8th iteration, both greater than $8.3243 \times 10^{-4}$.

For the Beidou mode, the iteration ends at the 8th times, and the test error at the 7th iteration is $6.1237 \times 10^{-4}$. At this time, 6 labeled test samples have been introduced. At the same time, the 21 labeled samples are used to train the traditional GP model and SVM model. The test sample set used in the 7th iteration of the co-training algorithm is also used to test the above two models, and the test errors are $8.1689 \times 10^{-4}$ and $0.0065$ respectively, which are both greater than $6.1237 \times 10^{-4}$.

According to the above experimental results of both two operating modes, based on the same labeled samples, the predictive ability of the proposed co-training model is improved compared with that of the traditional supervised learning models. In order to ensure that the modeling accuracy of the models can all reach the high levels, the model in the 8th iteration is adopted as the best one in the experiment of GPS Beidou dual-mode MSA, which cost 1604.46 seconds totally.

### TABLE 8. Experimental samples of the GPS Beidou dual-mode MSA.

| Geometry | Training samples (15 samples) | Test samples (10 samples) | Unlabeled samples (10 samples) |
|----------|-------------------------------|---------------------------|-------------------------------|
| variable | Min  | Max  | Step | Min  | Max  | Step | Min  | Max  | Step |
| $W$(mm)  | 42   | 45   | 1    | 42   | 45   | 1    | 42   | 45   | 1    |
| $L_1$(mm)| 5.1  | 6.3  | 0.4  | 5.1  | 6.3  | 0.4  | 5.1  | 6.3  | 0.4  |
| $L_2$(mm)| 3    | 3.9  | 0.3  | 3    | 3.9  | 0.3  | 3    | 3.9  | 0.3  |

### TABLE 9. Initial errors of the GPS Beidou dual-mode MSA.

| Model          | GP     | SVM    |
|----------------|--------|--------|
| Initial error of GPS | 0.0054 | 0.0153 |
| Initial error of Beidou | 0.0029 | 0.0071 |
and the dotted red line named ‘Proposed’ is the prediction results of the method proposed in this article. It can be seen that the two curves are basically consistent, indicating that the trained semi-supervised co-training model in this experiment has high accuracy, which can replace the HFSS simulation for predicting.

IV. CONCLUSION

In order to improve the optimization efficiency of electromagnetic components, reducing the number of times to call HFSS, and saving the time of obtaining labeled samples, this article proposes a semi-supervised co-training algorithm based on GP model and SVM model. GP model and SVM model are cross-trained with the same unlabeled samples. After comparing the test errors, the two models are future trained with pseudo-labeled samples with higher accuracy and corresponding test samples. Iteration termination conditions are set to control the number of unlabeled samples and the number of model updates to find the best solution in a relatively short time, which can effectively prevent the reduction of model accuracy. The benchmark functions are used to verify the effectiveness of the proposed algorithm. From the experimental results, it can be seen that the proposed algorithm has a good fitting effect for the Griewank function and the Quartic function. The Yagi MSA and GPS Beidou dual-mode MSA are optimized and the effectiveness of the proposed algorithm is verified by the resonant frequency modeling experiments. The experimental results show that, in the case of using the same labeled samples, the proposed co-training algorithm improves the prediction ability compared with the traditional supervised learning method. In order to further verify the effectiveness of the proposed algorithm, the trained resonant frequency model is used to fit the $S_{11}$ values. From the results, we can see the proposed model has a good fitting accuracy for $S_{11}$ of the above antennas. Therefore, the co-training algorithm proposed in this study are suitable for the problems that include many required label samples, high calculation cost and long time when training the model in antenna design. This proposed model can replace electromagnetic simulation effectively, which obviously save time in antenna optimization. It will further promote the applications of SSL method in the field of electromagnetic optimization.

REFERENCES

[1] A. F. Peterson, S. L. Ray, and R. Mittra, *Computational Methods for Electromagnetics*. New York, NY, USA: IEEE Press, 1998.
[2] X. H. Fan, Y. B. Tian, and Y. Zhao, “Optimal design of microwave devices by fitness-estimation-based particle swarm optimization algorithm,” *Appl. Comput. Electromagn. Soc. J.*, vol. 33, no. 11, pp. 1259–1267, 2018.
[3] X. Chen, Y. Tian, T. Zhang, and J. Gao, “Differential evolution based manifold Gaussian process machine learning for microwave Filter’s parameter extraction,” *IEEE Access*, vol. 8, pp. 146450–146462, 2020, doi: 10.1109/ACCESS.2020.3015043.
[4] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q.-J. Zhang, “Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of microwave filters,” *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4140–4155, Oct. 2019.
[5] T. Yu-Bo, Z. Su-Ling, and L. Jing-Yi, “Modeling resonant frequency of microstrip antenna based on neural network ensemble,” *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 24, no. 1, pp. 78–88, Jan. 2011.
[6] L.-Y. Xiao, W. Shao, X. Ding, Q. H. Liu, and W. T. Jinnes, “Multigrade artificial neural network for the design of finite periodic arrays,” *IEEE Trans. Antennas Propag.*, vol. 67, no. 5, pp. 3109–3116, May 2019.
[7] Zeng, Tan, Matsunaga, and Shirai, “Generalization of parameter selection of SVM and LS-SVM for regression,” *Mach. Learn. Knowl. Extraction*, vol. 1, no. 2, pp. 745–755, Jun. 2019.
[8] M. Ćmielia, J. Tabor, and P. Spurek, “SVM with a neutral class,” *Pattern Anal. Appl.*, vol. 22, no. 2, pp. 573–582, May 2019.
[9] L.-Y. Xiao, W. Shao, X. Ding, and B.-Z. Wang, “Dynamic adjustment kernel extreme learning machine for microwave component design,” *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 10, pp. 4452–4461, Oct. 2018.
[10] Y. Gao and Y. B. Li, “Misfire identification of automobile engines based on wavelet packet and extreme learning machine,” *J. Meas. Sci. Instrum.*, vol. 8, no. 4, pp. 384–395, 2017.
[11] E. Mark, “Gaussian processes for regression and classification: A quick introduction,” *Statist., Tech. Rep.*, 2015.
[12] J. P. Jacobs and S. Koziel, “Reduced-cost microwave filter modeling using a two-stage Gaussian process regression approach,” *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 25, no. 5, pp. 453–462, Jun. 2015.
[13] F. Chen and Y. B. Tian, “Modeling resonant frequency of rectangular microstrip antenna using CUDA-based artificial neural network trained by particle swarm optimization algorithm,” *The Appl. Comput. Electromagn. Soc. J.*, vol. 2, no. 12, pp. 1025–1034, 2014.
[14] D. Rittel, A. Dorogoy, G. Hiat, and K. Shemtov-Yona, “Resonant frequency analysis of dental implants,” *Med. Eng. Phys.*, vol. 66, pp. 65–74, Apr. 2019.
[15] J. Gao, Y. Tian, X. Zheng, and X. Chen, “Resonant frequency modeling of microwave antennas using Gaussian process based on semisupervised learning,” *Complexity*, vol. 2020, pp. 1–12, May 2020.

**TABLE 10. Iterative results of the GPS Beidou dual-mode MSA.**

| Iteration number | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      |
|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Test error of GPS | 0.0075 | 0.0064 | 0.0033 | 0.0025 | 0.0023 | 0.0015 | 0.0011 | 8.3243e-04 | 0.0018 |
| Test error of Beidou | 0.0037 | 0.0016 | 0.0015 | 0.0014 | 9.9543e-04 | 6.1237e-04 | 9.8790e-04 | 12.3243e-04 | 0.0018 |

**FIGURE 7.** $S_{11}$ fitting diagram of the GPS Beidou dual-mode MSA.
[16] S. Fei-Yan, T. Yu-Bo, and R. Zuo-Lin, “Modeling the resonant frequency of compact microstrip antenna by the PSO-based SVM with the hybrid kernel function,” Int. J. Numer. Model., Electron. Netw. Devices Fields, vol. 29, no. 6, pp. 1129–1139, Nov. 2016.

[17] Y. X. Xu, Y. B. Tian, and X. P. Hu, “Design of Dual-frequency Microstrip Antenna Based on Gaussian Process of a New Kernel Function,” Comput. Appl. Res., vol. 35, no. 8, pp. 2477–2479, 2483, 2018.

[18] J. C. Zhang, S. Y. Zeng, and Y. H. Jiang, “A Gaussian process based method for antenna design optimization,” Comput. Intell. Intellect. Syst., vol. 575, pp. 230–240, Dec. 2016.

[19] Z.-H. Zhou and M. Li, “Semi-supervised learning by disagreement,” Knowl. Inf. Syst., vol. 24, no. 3, pp. 415–439, Sep. 2010.

[20] O. Kostopoulos, S. Karlos, S. Kotsiantis, and O. Ragos, “Semi-supervised regression: A recent review,” J. Intell. Fuzzy Syst., vol. 35, no. 2, pp. 1483–1500, Aug. 2018.

[21] O. Chapelle, J. Weston, and B. Scholkopf, “Cluster kernels for semi-supervised learning,” in Advances in Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2002.

[22] D. Wu, M. Shang, X. Luo, J. Xu, H. Yan, W. Deng, and G. Wang, “Self-training semi-supervised classification based on density peaks of data,” Neurocomputing, vol. 275, pp. 180–191, Jan. 2018.

[23] Z. Long, H. Meng, and M. Sioutis, “Semi-supervised dimensionality reduction by linear compression and stretching,” IEEE Access, vol. 8, pp. 27308–27317, 2020.

[24] X. Guo and K. Uehara, “Graph-based semi-supervised regression and its extensions,” Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 6, pp. 260–269, 2015.

[25] E. Cheung and Y. Li, “Self-training with adaptive regularization for S3 VM,” in Proc. Int. Joint Conf. Neural Netw., May 2017, pp. 3633–3640.

[26] X. Li, H. Lu, J. Yang, and F. Chang, “Semi-supervised LIBS quantitative analysis method based on co-training regression model with selection of effective unlabeled samples,” Plasma Sci. Technol., vol. 21, no. 3, Mar. 2019, Art. no. 034015.

[27] W. Wang and Z. H. Zhou, “When does co-training work in real data?” IEEE Trans. Knowl. Data Eng., vol. 23, no. 5, pp. 675–680, Dec. 2010.

[28] A. Blum and T. Mitchell, “Combining labeled and unlabeled data with co-training,” in Proc. 11th Annu. Conf. Comput. Learn. Theory, Madison, WI, USA, 1998, pp. 92–100.

[29] L. Meng, X. Guo, L. Liu, and C. Cardie, “Limitations of co-training for natural language learning from large data sets,” in Proc. 6th Conf. Empirical Methods Natural Lang. Process., Pittsburgh, PA, USA, 2001, pp. 95–102.

[30] L. H. Zhou, K. J. Chen, and Y. Jiang, “Exploiting unlabeled data in content-based image retrieval,” in Proc. 15th Eur. Conf. Mach. Learn., Pisa, Italy, 2004, pp. 525–536.

[31] Z. H. Zhou and M. Li, “Semi-supervised Learning with Co-training style algorithm,” IEEE Trans. Knowl. Data Eng., vol. 19, no. 11, pp. 1479–1493, 2007.

[32] GP Package and SVM Package. [Online]. Available: https://www.mathworks.com/help/stats/fitrgp.html

[33] GP Package and SVM Package. [Online]. Available: https://www.mathworks.com/help/stats/fitrsvm.html

[34] Z. R. Wang and B. N. Tang, “Prediction and experimental investigation of air conditioning load based on support vector machine algorithm,” Chin. J. Refrigeration Technol., vol. 4, pp. 28–31, Dec. 2013.

[35] X.-H. Fan, Y.-B. Tian, and Y. Zhao, “Optimal design of yagi microstrip antenna based on particle swarm optimization with fitness estimation,” in Proc. Prog. Electromagn. Res. Symp., Toyama, Aug. 2018, pp. 1–5.

[36] W. C. Tian, D. W. Wu, and Q. Chao, “Application of genetic algorithm in M–N reconfigurable antenna array based on RF MEMS switches,” Modern Phys. Lett. R, vol. 32, no. 30, 2018, Art. no. 1850365.

[37] Z. Qiang, Y. Chen, and L. Xu, “Optimization of GPS antenna by PSO-based GP modeling,” J. Radio Sci., vol. 31, no. 5, pp. 927–932, 2016.