Relationship Study of Meltblowning Variables Based on Machine Learning Algorithm

Peng Cheng[1], Fuyou Mao[2,*], Haomin Zhao[2]

[1]Department of Water Conservancy Science and Engineering, Sichuan University, Sichuan, 610065, China
[2]College of Computer Science and Technology, Shenyang Jianzhu University, Shenyang, 110168, China
*Corresponding author: 2319723892@qq.com

Abstract: For the missing intercalation rate data, we first filled it in, and then split it according to the variable pairs to observe the changes of six indexes before and after intercalation. Then, we made grey correlation analysis between the changes and intercalation rate, and got the influence of intercalation rate on each index. In order to further explore the relationship between process parameters and structural variables, nine models, KNN, linear regression, ridge regression, lasso regression, decision tree, support vector machine, robust model, XGBoost and random forest, are used to train the data, and finally XGBoost regression model has the highest accuracy. Then, using the prediction results of structural variables obtained by XGBoost, we firstly make factor analysis on three indexes of structural variables and product performance, and the index with the largest factor load represents structural variables and product performance, and then make Pearson correlation analysis on these two indexes to get the relationship between structural variables and product performance. Through Pearson analysis of the three internal variables of structural variables and product performance, the internal correlation is obtained. For three indexes in the structural variables, respectively, they are linearly fitted with two variables of process parameters to obtain three fitting equations, and then they are fitted with filtration efficiency to obtain the fitting equation between filtration efficiency and structural variables. Finally, the linear regression equation between filtration efficiency and process parameters is sorted out. However, the effect of changing the linear regression model is not good in model testing, and we think it has a complex nonlinear relationship. Therefore, we use machine learning to carry out regression training on variables. The results show that when the result variables are used to regress the product performance, the effect of using random forest is better, but the filtering efficiency can't be achieved by using many kinds of machine learning. After that, considering the internal influence relationship of product performance, we used structural variables and all other indicators of product performance for regression training, and found that XGBoost algorithm had good effect, so we established a multiple regression model based on machine learning. By controlling the process parameters and observing the predicted structure, it is found that the filtration efficiency is the highest when the receiving distance is 10cm and the hot air speed is 1400r/min. Finally, we set up a multi-objective planning model, and globally optimize the planning model through the sand dune cat population optimization algorithm, and finally

get the approximate optimal matching scheme of process parameters.

1. Introduction

Meltblowing is a unique one-step process, in which the polymer is converted into fibers with a diameter of 1-5 μ m. In this process, the polymer melt is extruded into high-speed hot air flow through the capillary of the die, and the polymer melt bundle is accelerated, cooled, stretched and formed into small fibers at the same time. The method can produce fine fibers [1-2] with high surface area with high productivity and low cost. These characteristics of meltblowing technology make it an ideal choice technology for producing high-quality air filter media. Nowadays, melt-blown technology is widely used to produce different types of filter media, such as cartridge filter, clean room filter, mask, cabin air filter and HVAC filter. Melt-blown nonwoven is a good air filtration material with small fiber diameter, large specific surface area and high porosity. Its characteristics of good filtration effect, low penetration resistance, light weight, large capacity, environmental protection and moderate price represent the main development direction of future filtration materials. Meltblown nonwoven is a good air filter material with small fiber diameter, large specific surface area and high porosity. Its characteristics of good filtration effect, low resistance, light weight, large capacity, environmental protection and moderate price represent the main development direction of future filter materials [1]. Therefore, meltblown nonwovens have become important raw materials for mask production due to their excellent filtration performance, simple production process, low cost and light weight, and have attracted wide attention from domestic and foreign enterprises. However, because the fibers of melt-blown nonwovens are very fine, the performance of melt-blown nonwovens is often not guaranteed because of its poor compression resilience in use. Therefore, scientists created the intercalation meltblowing method, that is, by inserting polyester (PET) staple fiber and other fiber dimensions into the melt-blown fiber flow during the melt-blown preparation of polypropylene (PP), the "Z-shaped" structure intercalation meltblown nonwovens were prepared. However, there are many process parameters for the preparation of intercalated meltblown nonwovens, and the interaction among them is even more complicated after the intercalation air flow. Therefore, the research on determining the structural variables (thickness, porosity, compression resilience) by the process parameters (receiving distance and hot air speed) and the final product performance (filtration resistance, filtration efficiency, air permeability) by the structural variables becomes more complicated. So this paper explores the relationship between structural variables and product performance. The contributions of this paper are summarized as follows:

(1) Using grey correlation analysis, the influence of intercalation rate on each index is obtained [2, 3].

(2) 9 machine learning algorithms are used to explore the relationship between process parameters and structural variables.

(3) Factor analysis of structural variables and three indexes of product performance (thickness, porosity, compression resilience) shows that the index with the largest factor load represents structural variables and product performance, and Pearson correlation analysis of these two indexes shows the relationship between structural variables and product performance.

(4) Build a multiple regression model based on XGBoost [4, 5]. By controlling the variables of process parameters and observing the prediction structure, the process parameters with the highest filtration efficiency are obtained.

(5) The minimum filtering resistance and the maximum filtering efficiency are taken as objective functions, the existing constraints are digitized, and a multi-objective planning model is constructed by combining the trained planning models (2), (3) and (4). Then, the sand dune cat population optimization algorithm is used to globally optimize the planning model, and finally the approximate optimal matching scheme of process parameters is obtained.

2. Data set

First of all, for the record of melt-blown samples, that is, no intercalation, we fill in the missing intercalation rate as 0. Secondly, the data before and after intercalation are split and reorganized, and the statistical charts of six secondary indicators before and after intercalation are obtained, as shown in Figure 1 and Figure 2.
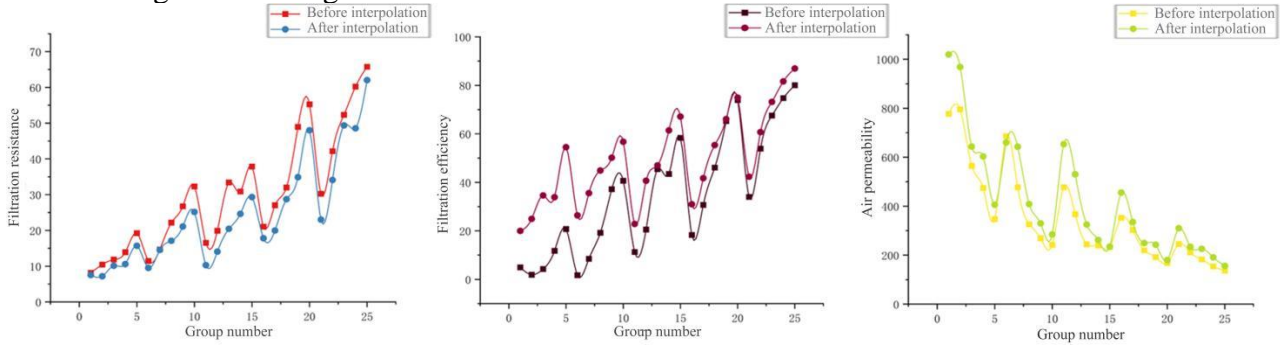

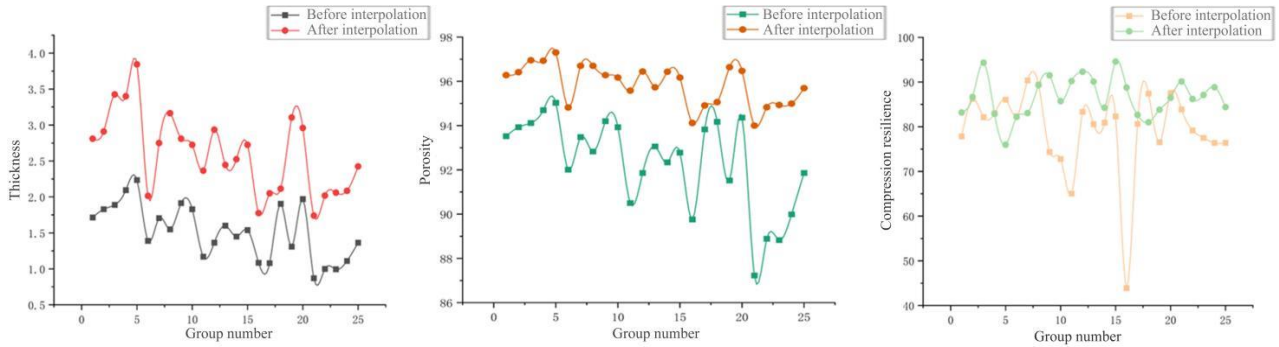Figure 1: Changes of structural variables before and after intercalation

Figure 2: Changes of product performance before and after intercalation

And calculate that mean and variance of the data before and after the change accord to the results:

Table 1: Changes of indexes before and after intercalation

	thickness	1.5188	0.3748	2.6074	0.5378
structure variable	void ratio	92.3492	2.0486	95.8612	0.8994
	Compressive resilience	79.4416	9.1376	86.6228	4.3437
	filtration resistance	29.7844	16.1298	24.1573	14.4848
product performance	filtration efficiency	34.99204	24.6227	49.3866	18.493
	gas permeability	347.604	185.7449	422.1344	232.032

The data is normalized by Equation 1. The results are shown in Figure 3.

$$x_{ni} = \frac{x_{ni}}{max\{x_{ni}\}}, n=0,1,2,\ldots,6, i=1,2,\ldots,50 \tag{1}$$
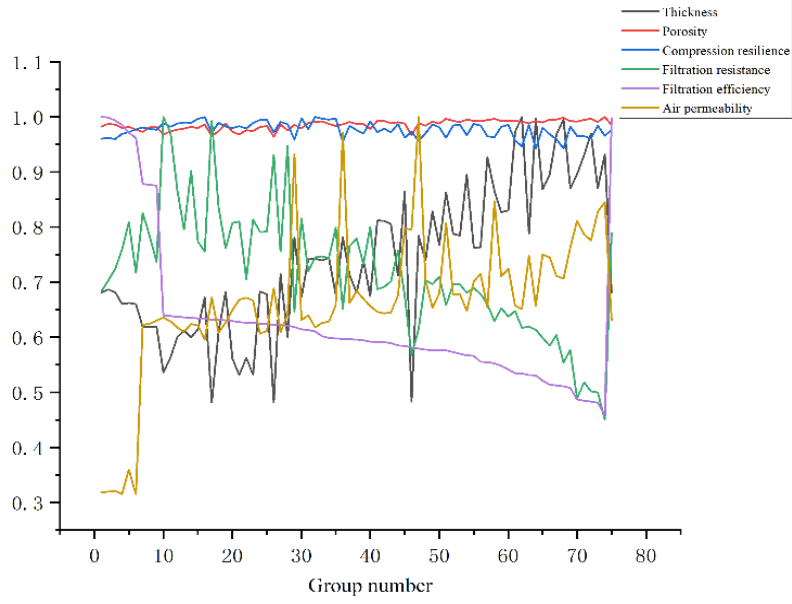
Figure 3: Overall change trend

3. Grey Correlation Analysis

In order to explore the influence of intercalation rate on structural variables and product performance, we take intercalation rate as the parent sequence, and all the secondary indexes of structural variables and product performance as the subsequence. There are 6 subsequences, and each sequence has 50 samples, so the data is expressed as Equation 2.

$$x_n = \left(x_n(1), x_n(2), ..., x_n(m) \right)^T, n = 0, 1, 2, ..., 6, m = 50 \tag{2}$$

Then, the correlation coefficient between the subsequence and the mother sequence is calculated by Equations 3.

$$y\left(x_0(k), x_i(k) \right) = \frac{a + \rho b}{\left| x_0(k), x_i(k) \right| + \rho b} \tag{3}$$

Where,

$$a = min(i) min(k) \left| x_0(k) - x_i(k) \right|$$
$$b = max(i) max(k) \left| x_0(k) - x_i(k) \right| \tag{4}$$

Finally, the grey correlation degree of each subsequence is calculated by Equation 5. The results of gray correlation analysis are shown in Figure 4 and Table 1.

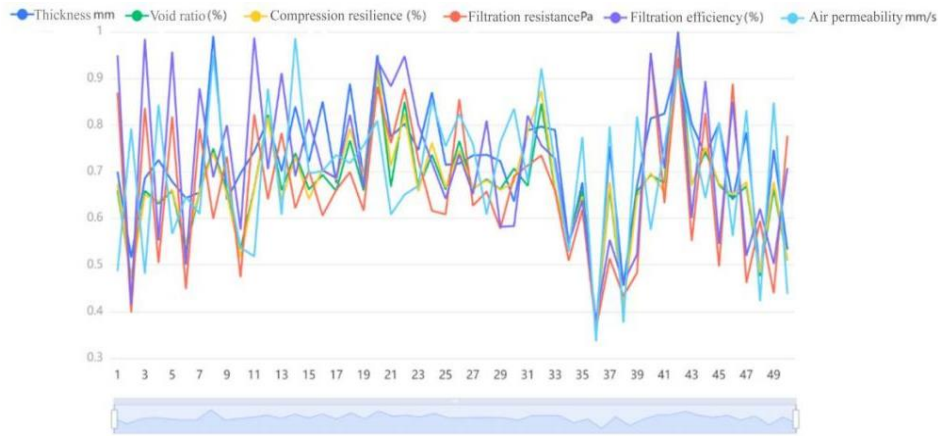$$y\left(x_0, x_i \right) = \frac{1}{n} \sum_{k=1}^{n} y\left(x_0(k), x_i(k) \right) \tag{5}$$

20

Figure 4: Gray Correlation Analysis

Table 2: Grey correlation results

Evaluation item	degree of association	ranking
Thickness mm	0.724	1
Filtering efficiency (%)	0.718	2
Air permeability mm/s	0.697	3
Compression resilience (%)	0.671	4
Porosity (%)	0.668	5
Filtration resistance Pa	0.662	6

Combined with the above correlation coefficient results, the correlation degree value is finally obtained, and the correlation degree value is used to evaluate and sort the six evaluation objects. The correlation value is between 0 and 1, and the larger the value, the stronger the correlation between it and the parent sequence, which means the higher its influence. As can be seen from the above table 2, among the six evaluation items, thickness mm is the highest (correlation degree: 0.724), followed by filtration efficiency (correlation degree: 0.718).

4. Prediction of product structure based on machine learning algorithm

In order to further explore the relationship between process parameters and structural variables, this paper uses KNN, linear regression, ridge regression, lasso regression, decision tree, support vector machine, robust model, XGBoost and random forest to train the data. The training framework is shown in Figure 5.
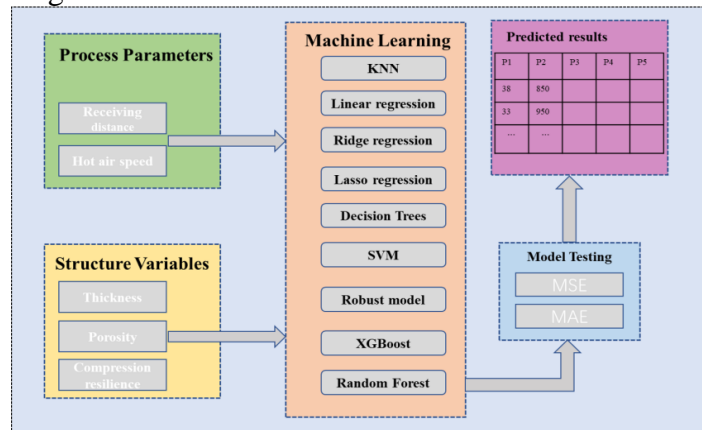

Figure 5: The training framework

We first divided the data into 0.7 training set and 0.3 testing set by Python, and tested the data by calling each model. The regression effects of thickness, porosity and compression rebound rate in structural variables are shown in Figure 6. The thickness training accuracy of the model is shown in Table 3.
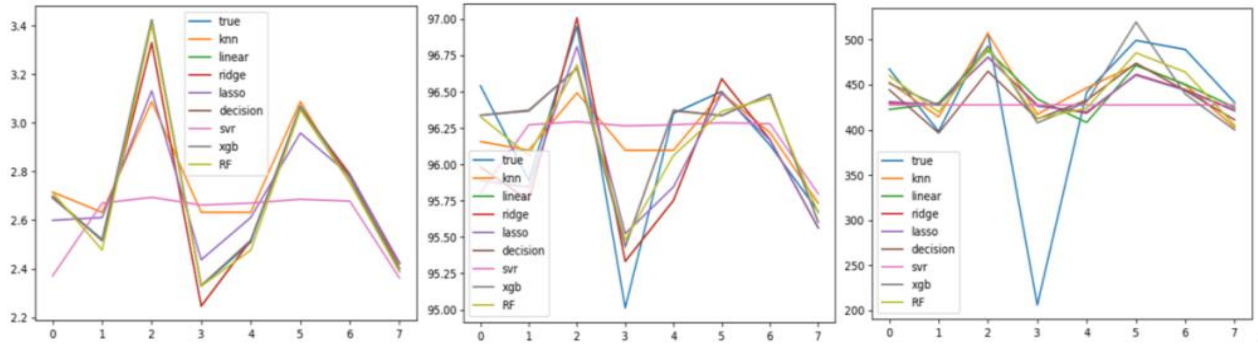

Figure 6: Regression effect of structural variables

Table 3: The thickness training accuracy of the model

index algorithm	thickness		poriness		Compression rebound rate	
	MSE	MAE	MSE	MAE	MSE	MAE
KNN,	0.0292	0.1209	0.2062	0.3171	0.4964	0.6293
linear regression,	0.0020	0.0298	0.1022	0.2390	1.0617	0.9449
ridge regression,	0.0020	0.0298	0.1022	0.2389	1.0617	0.9449
lasso regression	0.0170	0.1043	0.1220	0.2512	1.0817	0.9448
decision tree	0.0000	0.0051	0.0859	0.2501	0.1058	0.2424
support vector machine	0.1186	0.2752	0.3466	0.4483	1.0326	0.9168
robust model	0.0022	0.0309	0.0998	0.2369	1.0310	0.9337
XGBoost	0.0000	0.0046	0.0850	0.2484	0.1056	0.2423
Random forest	0.0192	0.0192	0.2343	0.2343	0.2335	0.2335

5. Factorial analysis

In order to make the complicated relationship between variables easier to explain, we reduce the dimension of structural variables and product performance by factor analysis [6]. The factor analysis method is to analyze the correlation coefficient matrix among various index variables, and summarize their relationships into a few comprehensive factors. Because the number of summarized factors is less than that of the original variables, but because they contain the information of the original variables, it can achieve the function of dimensionality reduction of the index [7].

The results of factor analysis are shown in Table 4.

Table 4: Factor molecular results

Name	Factor load	Name	Factor load
Thickness	0.952	Thickness	0.922
Poriness	0.926	Poriness	0950
Compression rebound rate	0.709	Compression rebound rate	-0.906

It can be seen from the above table that thickness and filtration resistance are selected as the descriptive variables of the first-class index among the structural variables and product performance. Then we make Pearson analysis of structural variables and descriptive variables of product performance through Python, and the results are shown in Figure 7:

	Thickness	Filtration efficiency
Thickness	1	0.138
Filtration efficiency	0.138	1

Figure 7: Correlation heat diagram

The relationship between thickness and filtration efficiency can be known from the thermal diagram, and thus the relationship between structural variables and product performance is:

$$F_{product\ performance} = 0.138 F_{structure\ variable}.$$

Pearson analysis of three variables in structural variables and product performance is shown in Figure 8.
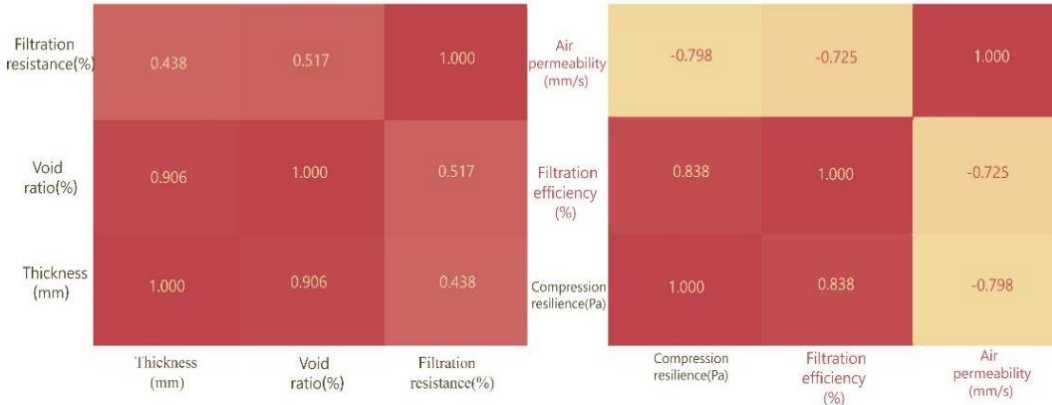

Figure 8: Correlation Analysis Chart

From the thermal diagram analysis of the internal correlation between structural variables and product performance, it can be seen that there is a positive correlation of 0.906 between thickness and porosity in structural variables, and a negative correlation of 0.798 between filtration resistance and permeability in product performance.

6. Multi-objective programming model

It can be seen from the above that the two optimization objectives are low filtration resistance and high filtration efficiency, and the solution is the collocation of process parameters [8]. The regression model uses symbols as shown in Table 5:

Table 5: Function description in multi-objective programming model

Symbol	Explain
$f_{XG\_h}(x_1, x_2)$	Process parameters use XGBoost to predict thickness.
$f_{XG\_k}(x_1, x_2)$	Process parameters XGBoost is used to predict porosity.
$f_{XG\_y}(x_1, x_2)$	Process parameters XGBoost is used to predict compression rebound rate.
$f_{RF\_z}(x_1, x_2, x_3)$	Structural variables use random forest to predict filtration resistance.
$f_{RF\_t}(x_1, x_2, x_3)$	Structural variables use random forest to predict permeability.
$f_{RF\_x}(x_1, x_2, x_3)$	Structural variables use random forest to predict filtering efficiency.
$f_{RF\_x}(x_1, x_2, x_3, x_4, x_5)$	Five variables use random forest to predict filtering efficiency.

Based on this, we establish the objective function as follows:

$$max\ y_1 = f_{RF\_x}\left(f_{XG\_h}(x_1,x_2), f_{XG\_k}(x_1,x_2), f_{XG\_y}(x_1,x_2), f_{RF\_z}(x_1,x_2,x_3), f_{RF\_t}(x_1,x_2,x_3)\right) \tag{6}$$

$$min\ y_2 = f_{RF\_z}\left(f_{XG\_h}(x_1,x_2), f_{XG\_k}(x_1,x_2), f_{XG\_y}(x_1,x_2)\right) \tag{7}$$

Among them, x_1 and x_2 are the receiving distance and hot air speed of process parameters, and y_1 and y_2 are the filtration resistance and filtration efficiency of product performance, respectively.

Then we set the constraint conditions. The constraint conditions of the receiving distance, hot air speed, thickness and compression resilience given in the title are as follows:

$$x_1 \le 100, x_2 \le 2000 \tag{8}$$

$$f_{XG\_h}(x_1,x_2) \le 3 \tag{9}$$

$$f_{XG\_y}(x_1,x_2) \ge 0.85 \tag{10}$$

In this paper, the improved swarm optimization algorithm simulating sand dune cat swarm is used to solve the dynamic programming problem [9]:

The intelligent algorithm of sand dune swarm divides cat behavior into two forms, one is search form; The other is the form of tracking.

We use Java language to solve the dynamic programming problem by using the improved swarm optimization algorithm which simulates the sand dune cat swarm. First of all, we assume that the individual of the population is a cat with memory, and each iteration, that is, the selection of schemes, needs to be recorded in the memory module of the cat group.

Set the number of cats to 500 and the number of iterations to 500, search, capture and move inside the cat swarm, and finally get a feasible solution set. In the set, select the top 10% with high filtering efficiency, and then select the solution with the lowest filtering resistance as the optimal solution of this iteration as the planning model, and add it into the memory of the cat swarm.

For the above-mentioned model constraints, every time a cat swarm is generated, constraints are carried out, and according to the memory of the cat swarm, feasible solutions meeting the constraint conditions are generated.

Algorithm 1. Sand cat swarm optimization algorithm pseudocode
Initialize the population
Calculate the fitness function based on the objective function
Initialize the r, r_G, R
While (t<= maximum iteration)
For each search agent
Get a random angle based on the Roulette Wheel Selection ($0° \le \theta \le 360°$)
If(abs(R)<=1)
Update the search agent position based on the Eq.5   ; $\overrightarrow{Pos_b}(t) - \overrightarrow{Pos_{rnd}} \cdot cos(\theta) . \vec{r}$
Else
Update the search agent position based on the Eq.4   ; $\vec{r}.(\overrightarrow{Pos_{bc}}(t) - rand(0,1) \cdot \overrightarrow{Pos_c})$
End
End
t=t++
End

At last, after iteration, we get the best approximate solution with the highest filtration efficiency and lowest filtration resistance within the specified parameter limit when the receiving distance is 17cm and the hot air speed is 1850r/min. The local iteration of population is shown in Figure 9:
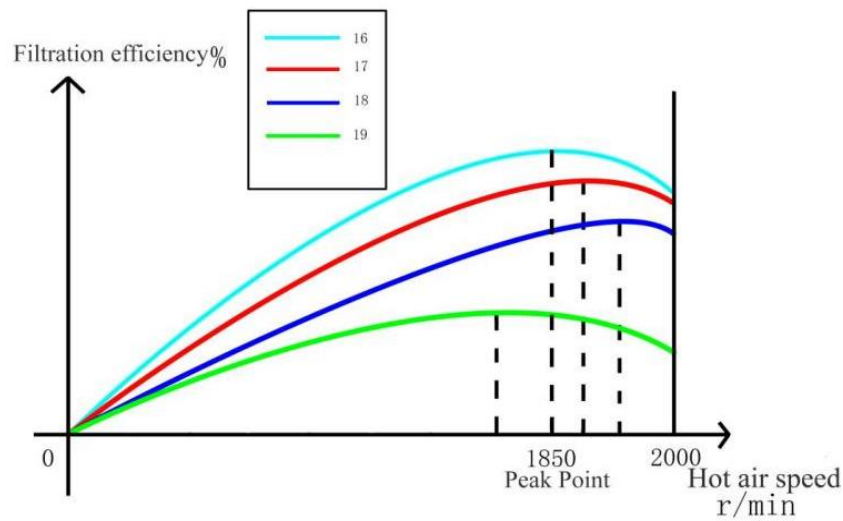
Figure 9: Optimal solution and its neighborhood results

As shown in the figure above, according to the conclusion of the third question, when the hot air speed is constant, the smaller the receiving distance, the greater the filtration efficiency. At a certain receiving distance, the higher the hot air speed, the higher the filtration efficiency. We broke through the range of the given data for optimization, and found that when the distance is 17cm, the peak value of the influence of hot air speed and filtration efficiency is the highest, and the peak point is close to 1850r/min. Therefore, we took the corresponding process parameters as the optimal solution of this problem.

7. Conclusion

Based on the process parameters and structural variables of the meltblowing data and the product performance dataset, this paper analyzes them through different operational research methods and machine learning methods to obtain the intrinsic complex nonlinear relationships between them, and finally finds the optimal process parameter solution by setting certain constraints and using the Sand Cat heuristic algorithm.

References

[1] Jafari Mehran and Shim Eunkyoung and Joijode Abhay. (2021). Fabrication of Poly (lactic acid) filter media via the meltblowing process and their filtration performances: A comparative study with polypropylene meltblown. Separation and Purification Technology, 260.

[2] Hasolli Naim, Park Young Ok & Kim Kwang Deuk. (2021). Multi-layered nonwoven filter media for capture of nanoparticles in HVAC systems. Korean Journal of Chemical Engineering (5). doi:10.1007/S11814-021-0773-9.

[3] Tianqi Chen & Carlos Guestrin. (2016). XGBoost: A Scalable Tree Boosting System.. CoRR.

[4] Ye Yuguang et al. (2021). Management of Medical and Health Big Data Based on Integrated Learning-based Health Care System: A Review and Comparative Analysis. Computer Methods and Programs in Biomedicine, 209 (prepublish), pp. 106293.

[5] Abedi Rahebeh, Costache Romulus, Shafizadeh Moghadam Hossein & Pham Quoc Bao. (2022). Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto International (19). doi:10.1080/10106049.2021.1920636.

[6] Feng Tengfei, Shen Yunzhong, Chen Qiujie, Wang Fengwei & Zhang Xingfu. (2022). Groundwater storage change and driving factor analysis in north china using independent component decomposition. Journal of Hydrology.

[7] Zhu Weidong et al. (2022). Research on optimization of an enterprise financial risk early warning method based on the DS-RF model. International Review of Financial Analysis, 81.

[8] Wang Yongli, Huang Feifei, Tao Siyi, Ma Yang, Ma Yuze, Liu Lin & Dong Fugui. (2022). Multi-objective planning

of regional integrated energy system aiming at exergy efficiency and economy. Applied Energy (PB). doi: 10.1016/J. APENERGY.2021.118120.

[9] Seyyedabbasi A, Kiani F. Sand Cat swarm optimization: a nature-inspired algorithm to solve global optimization problems [J]. Engineering with Computers, 2022: 1-25.