Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

Research on Soft-sensing Method of Water Content of Fixed Tea Based on Improved Clustering

Jiangwen Tang[1], Zhuang Yang[1], Li Jiang[1], Liang Xue[1], Yongguang Hu[2], Huiliang Dai[3], Haiwei Xu[1]

[1]National Institute of Measurement and Testing Technology, Chengdu, Sichuan 610021, China
[2]School of Agricultural Engineering, Jiangsu University, Zhenjiang, Jiangsu 210213, China
[3]Zhejiang Shangyang Machinery Co., Ltd., Quzhou, Zhejiang 324000, China

Abstract:

The water content of tea leaves is an important process parameter of fixation of tea, existing measurement method for water content of tea is not suitable for on line measurement because of the limits of measurement speed, sample handling and cost. So, the water content of fixation leaves can only be evaluated by sense organ and be controlled by setting process parameters according the experience. In the process of removing tea activity,in order to predict the water content by the process parameter of fixation, the process of fixation in a cylinder is analyzed and the parameters having effect on the water content of tea is found in this paper.And 350 groups of production scene data including the water content of tea leaves and the process parameters of fixation were collected, while the 300 groups of them are used to cluster with K-means algorithm, linear model is built in each class with recursive least squares. Taking the remaining 50 groups of data as test data, to calculate the water content of tea leaves with the linear model of the class in which the clustering center is nearest to the test data in distance, the result is compared with the measured value. The tests indicate that the MSE of the soft-sensing Modeling is less than 0.02 and the maximum error is 0.0463.

Keywords: *Fixation leaf, Water content, Soft-sensing, K-means, Least square.*

I. INTRODUCTION

The purpose of fixation of tea is to inactivate polyphenol oxidase, retain TP (tea polyphenol) and make fresh leaves lose proper water for subsequent processing. Inactivation of polyphenol oxidase in fresh leaves can be achieved by rapidly heating the leaves to the temperature at which the enzyme activity is lost [1]. The water content of fixation leaves is often judged by senses and controlled by setting fixation process according to experience. Most of the existing fixation machines adopt drum-type structure as shown in Fig 1. The drum temperature, rotation speed, feeding speed and other parameters of the fixation machine can be controlled. Users should get these parameters according to their experience before fixation on a large scale. Although the automation of leaf fixation is realized, the control of water content of

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

fixation leaves is essentially open loop control, which is mainly limited by the difficulty of online measurement of water content.

Both constant temperature oven method and rapid moisture measurement method need to heat tea leaves at a certain drying temperature for a period of time and then complete the measurement of fixation leaves water content by weighing. Although the measurement has a high accuracy, it takes a long time. Near-infrared spectroscopy can measure the water content of fixation leaves quickly (0.2s) and non-destructively with high precision (within 1%), but the measurement is easily affected by the type of fixation leaves and the instrument is expensive [2]. The measuring range of water content by microwave method is only 0.3% to 30%, which is not suitable for water content measurement of fixation leaves. Electric measuring method is a kind of measuring method that converts the water content of fixation leaves into resistance and other electric quantities by sensors. This method has a large measuring range and fast response, but it is easily affected by the type and grade of fixation leaves. Besides the measurement reference value needs to be determined before measurement. In sum, this method is not suitable for online measurement.

Fig 1: Basic structure of drum-type fixation machine

To realize on-line low-cost measurement of water content of fixation leaves and further improve the quality of tea fixation process control, a multi-model soft-sensing algorithm of water content of fixation leaves based on K-means clustering is designed in this paper.

II SELECTION OF AUXILIARY VARIABLES

The drum-type fixation machine heats the outer wall of the drum by a heat source, and the inner wall of the drum directly heats the tea leaves. Because the water inside the tea leaves is heated, it continuously diffuses to the surface of the tea leaves and forms saturated water vapor on the surface. Because of the rotation of the drum, hot air flows to the discharge end of the drum inside the drum [3]. The water vapor on the surface of the tea leaves is taken out of the drum by the hot air, and finally the tea leaves water properly. The principal calculation formula of water content of fixation leaves is as shown in formula (1). Where P is the water content of fixation leaves; p0 is the water content of fresh leaves; m is the feeding amount depending on the feeding rate of the fixation machine; T is the fixation time, and K is the water drying rate.

$$p = \frac{mp_0 - kt}{m - kt} \quad (1)$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

Leaf moisture drying in drum can be regarded as a forced convection mass transfer process in a straight tube. According to the principle of heat transfer, the mass flux density of forced convection in the drum, that is, the drying rate of fixation leaves water, is related to the drum temperature, air velocity, air humidity and drum diameter.

The fixation time depends on the residence time of tea leaves in the drum. The residence time of fixation leaves is related to the length and inclination angle of the drum, the diameter of the drum, the rotation speed of the drum, the gas flow in the drum and the diameter of fixation leaves.

Based on this, we can get all the variables related to the water content of fixation leaves, as shown in TABLE 1. The diameter and length of the drum are invariable. The drum temperature is the key factor that affects the water loss rate. The drum rotation speed and inclination angle are the key factors that affect the fixation time, and the endowment and feeding speed of fresh leaf are the key factors that determine the water content of fixation leaves. Considering the above factors, drum temperature, drum inclination angle, drum rotation speed, water content of fresh leaves and feeding speed are selected as auxiliary variables for soft sensing.

TABLE I. Variables of water content in the fixation leaves

Grade	Variable	Unit	Measurable	Is it a critical variable
1	Tube temperature	℃	Measurable	Yes
2	Air velocity	m/s	Measurable	No
3	Airflow humidity	None	Measurable	No
4	Gas-flow rate	kg/s	Measurable	No
5	Roller diameter	m	Measurable	No
6	Roller length	m	Measurable	No
7	Roller inclination angle	Sinusoidal value	Measurable	Yes
8	Rotary speed of drum	rpm	Measurable	Yes
9	Rate of feed	kg/s	Measurable	Yes
10	Fresh leaf water content	无 None	Measurable	Yes
11	Tea leaf diameter	m	Measurable	No

III K-MEANS CLUSTERING

K-means clustering algorithm is widely used in data mining algorithms because of its fast convergence speed and simple structure [4]. Its clustering effect depends on the selection of initial clustering centers and the number of classifications. Production data are collected according to vector s= (t, n, a, v, y0, y), where: t is the drum temperature; n is the drum rotation speed; a is the sine value of the drum inclination angle; v

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

is the feeding speed; y0 is the water content of fresh leaves, and y is the water content of fixation leaves. Part of the data are used for modeling, and the rest is used to verify the model.

2.1 Selection of Initial Clustering Centers

The implementation steps of the algorithm of initial clustering center selection are as follows:

(1) Calculate the distance $d(i,j)=\sum_{r=1}^{6}(x_{ir}-x_{jr})^2$ between samples. x_{ir} is the r-th component of the i-th sample.

(2) Calculate the average spacing between samples:

$$\bar{d}=\frac{\sum_{i=1}^{300}\sum_{j=1}^{300}d(i,j)}{300\times 299} \tag{2}$$

(3) Calculate the sample density $den(i)$. Wherein, $den(i)$ is the number of distances satisfying $d(i,j)\le\frac{1}{3}\bar{d}$ in d_{ij} (j=1…300). The sample point with the highest density is taken as the first clustering center.

(4) Calculate the distance between the remaining samples and the existing clustering center. Wherein, the distance between the i-th sample and the existing clustering center is dis(i)=min{$d(i,j)$}. j is the number of the existing clustering center; in this way, the sample i with the maximum value in formula (3) is the next clustering center.

$$J=den(i)+dis(i)-(den(i)-dis(i))^2 \tag{3}$$

(5) Repeat step (4) until the number of clustering centers is the number of clusters.

The clustering center selected according to formula (3) ensures the high density of the new initial clustering center itself and the far distance from the existing initial clustering center, thus avoiding selecting the point with low density or at edge as the initial clustering center [5].

2.2 Determination of the Number of Clusters

Like most other clustering algorithms, k-means clustering algorithm also requires to give the number of clusters in advance [6]. The purpose of clustering is to transform the nonlinear model into the linear model of each sub-region, and the number of clusters is the final number of linear sub-regions. The number of clusters has an important impact on the similarity of samples in each class and the difference of samples between classes, and then on the modeling accuracy and speed. However, it is difficult to determine the number of clusters in advance without knowing the distribution boundary of samples. Comparison method, fusion method and trial-and-error method all need a large amount of calculation [7]. In the absence of prior

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

knowledge, this paper adopts the method based on satisfactory clustering to determine the number of clusters. The adopted cluster evaluation index is the root mean square error of modeling:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{nt}(y_i - \hat{y}_i)}{nt}} \tag{4}$$

y_i is the actual output value of the i-th sample, and $\hat{y}_i$ is the predicted output value of the model of the i-th sample. nt is the number of cluster samples. The initial value of N is set to 2. If RMSE<0.02, N is the number of clusters; otherwise, N is incremented and the initial clustering center cluster is determined according to the 2.1 algorithm. And the cluster is modeled until the cluster accuracy meets RMSE<0.02. Because the above clustering initial value algorithm will not recalculate the existing clustering centers when increasing the number of clusters, it can greatly improve the operation speed. Different from the comparison method to find the number of clusters, the number of clusters we determine here is the minimum number of clusters that can meet the modeling accuracy under the condition of low computation, rather than the number of clusters that can meet the highest modeling accuracy.

2.3 Clustering Implementation

The initial clustering centers determined according to the above algorithm are all points in the sample. Given that its vector set is $Z_1(j)(j=1,2,3,...N)$ and N is the number of clusters, the implementation process of k-means clustering algorithm is as follows:

(1) Calculate the Euclidean distance from each sample to the clustering center $d(v(i), Z_1(j))$, where $i=1,2,3,...,n$ and $j=1,2,3,...,N$. n is the sample size. If it is satisfied:

$$d(x(i), Z_I(k)) = \min\{d(x(i), Z_I(j))\} \tag{5}$$

Then, sample x (i) is classified into class K.

(2) Calculate the error criterion function. n_i is the number of class i samples, and $x()^i$ indicates the samples divided into class i.

$$S(I) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left\|x(j)^i - Z_I(i)\right\|^2 \tag{6}$$

(3) Judgment: if $|S(I)-S(I-1)|<\zeta$, the algorithm is finished, otherwise, $I=I+1$, and the new clustering center is recalculated according to the following formula.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

$$Z_{\cdot}(j) = \frac{1}{n_j}\sum_{i=1}^{n_j} x(i)^{(j)} \tag{7}$$

(4) Return to step (1).

IV MULTI-MODEL MODELING OF WATER CONTENT OF FIXATION LEAVES

The essence of multi-model modeling is the process of model decomposition and synthesis, which is an effective solution to nonlinear system modeling [8-9]. The principle of multi-model modeling is shown in Fig 2. According to the above algorithm, process data sets are classified, and then linear models Mi of water content and process parameters are established in each category. For the new process parameter X, find out the class to which X belongs according to the corresponding mode switching algorithm, and then calculate the water content of fixation leaves when the process parameter is X according to the linear model in the corresponding class. To facilitate the real-time update of various linear models, this paper adopts recursive least square method to establish various linear models, and selects the linear models of the class with the shortest distance to X for prediction. In this way, the global nonlinear model is transformed into the local linear model, and then the global model is implemented according to the local model identification and model switching.

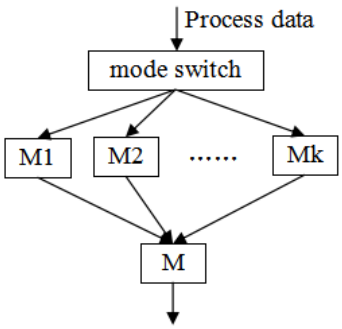

Fig 2: Structure diagram of multi-model modeling principle

3.1 Establishment of linear models in various classes

After dividing the tea production process data samples into N classes according to the above algorithm, recursive least square method is used to establish linear models of the sine value of water content of fixation leaves and drum temperature, drum speed, drum inclination angle as well as feeding speed and fresh leaf water content:

$$y_i = w_{i1}t + w_{i2}n + w_{i3}a + w_{i4}v + w_{i5}y_0 \tag{8}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

Namely $y_i = W_i^T x$, where:

$$x = [t, n, a, v, y_0] \quad W_i = [w_{i1}, w_{i2}, w_{i3}, w_{i4}, w_{i5}]^T$$

$i = 1, 2, 3, ..., N$ and $w_{ij} (j = 1, 2, 3, ..., 5)$ are the parameters that need to be identified when establishing a linear model in class i, and the recursive least square identification process is as follows:

$$W(k) = W(k-1) + K(k)[y(k) - \varphi^T(k-1)W(K-1)] \tag{9}$$

$$P(k) = [I - K(k)\varphi^T(k-1)]P(k-1) \tag{10}$$

$$K(k) = \frac{P(k-1)\varphi(k-1)}{1 + \varphi^T(k-1)P(k-1)\varphi(k-1)} \tag{11}$$

Wherein:

$$\varphi(k) = [t(k), n(k), a(k), v(k), y_0(k)]^T k = 1, 2, 3, ..., ni$$

ni is the sample number of class i. $K(k)$ is calculated first, and then $W(k)$ and $K(k)$ are calculated. The initial value of $K(k)$ is a small parameter, and the initial value of $P(k)$ is a diagonal matrix with a sufficiently large diagonal value.

3.2 Establishment of Water Content Model Based on Various Linear Models

For the tea-making process parameter vector $x(i) = [t(i), n(i), a(i), v(i), y_0(i)]$ set according to the user's requirements, the distance $d(x(i), Z_I(k))$ from $x(i)$ to each clustering center is calculated. Clustering centers $Z_I(r)$ and $Z_I(s)$ which are closest to $x(i)$ are selected. Let the distances from $x(i)$ to $Z_I(r)$ and to $Z_I(s)$ be d_1 and d_2 respectively and $d_1 \le d_2$.

If $d_1 \le 0.5 d_2$, then

$$y = W_r^T x \tag{12}$$

Conversely

$$y = \frac{d_1}{d_1 + d_2} W_r^T x + \frac{d_2}{d_1 + d_2} W_s^T x \tag{13}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

V MODEL VERIFICATION

The original data collection process was completed in a tea factory in Chongqing. The tea factory adopted the 6CST-55 numerical control infrared drum-type fixation machine produced by Sichuan Zhongce Measuring Instrument Company. This type of fixation machine can accurately control the drum temperature, the drum speed and the tilt angle of the drum feeding end. To control the feeding speed, a 6CST-10 automatic feeder was arranged in front of the 6CST-55 fixation machine during the experiment. The IR-3000 near infrared online moisture meter of MoistTech Company in the United States was used to measure the water content of fresh leaves and water content of fixation leaves. When measuring, the distance between the sensor and the sample was 20cm. The water contents of fresh leaves and fixation leaves were sampled and measured five times and the average value was taken as the final measurement value. The water content of fixation leaves was measured after they were cooled to room temperature.

The fresh tea leaves were autumn tea from the same tea garden. According to the picking progress of tea leaves, 350 experiments were arranged in 2014. Among them, 300 sets of data were used as modeling data, and the remaining 50 sets of data were used to verify the accuracy of the model.

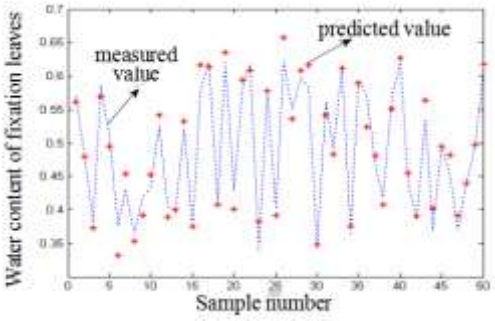

Fig 3: The model predicts the water content of fixation leaves and the actual water content of fixation leaves.

Based on the sample data, according to the method of determining the number of clusters described in 2.3, it is found that when the number of clusters is 5, better modeling accuracy can be obtained. At this time, the predicted value and the measured value of the water content model of the fixation leaves of the test sample are shown in Fig 3, with the root mean square error RMSE of 0.01945 and the maximum error MAXE of 0.0463. MAXE is defined as:

$$MAXE = \max_{i=1}^{nt}(|y_i - \hat{y}_i|) \tag{14}$$

The prediction error of the model is shown in Fig 4. It can be seen that the measurement model has good prediction accuracy in each sample interval. At the same time, Fig 3 shows that there is a big deviation between the predicted value and the measured value at individual sample points, which is mainly

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

due to the low precision of local linear model caused by the small number of modeling samples. Of course, the reason may also be that the test samples are distributed at the edge of each sub-region.

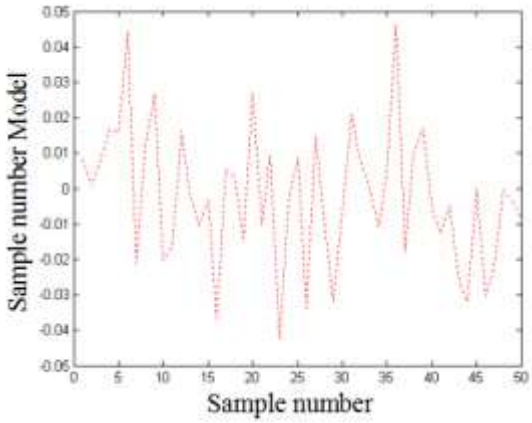

Fig 4: Model prediction error

VI CONCLUDING REMARKS

Aiming at the problem of on-line real-time measurement of fixation leaves of drum-type fixation machine, an initial clustering center algorithm and a method for determining the number of clusters are designed based on K-means clustering method. The K-means algorithm is used to cluster the process parameter samples of drum-type fixation machine, and a linear model of water content of fixation leaves is established in each class. Besides, a model switching algorithm is designed. Then, a soft-sensing model of water content of fixation leaves is established based on the linear model. The model experiment shows that the soft-sensing method proposed in this paper can effectively predict the water content of the fixation leaves of the drum-type fixation machine, and provide conditions for realizing the control of the water content of the fixation leaves.

Because the experimental data collected in this paper came from the same tea garden and the same fixation machine, the prediction effect of the established soft-sensing model on the water content of fresh leaves from different tea gardens at different picking times and after water removing by fixation machines with different specifications needs to be further tested. At the same time, K-means clustering is a hard clustering method, and the predicted values of sample models at various edges are affected by multiple linear models. Therefore, improving the accuracy of each linear model and optimizing the weight of the predicted values affected by each model are the keys to further improve the prediction accuracy.

ACKNOWLEDGEMENTS

This research was supported by National Science and Technology Support Program (2014BAD06B06-03).

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1810-1819
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

REFERENCES

[1] Yun Wang, Chunhua Li. Changes of Amino Acid Content in Famous Tea and Its Influence.
[2] Shousong Chen, Yang Zhan, Gongyu Zheng, Xinyi Jin. Common Determination Methods and Comparative Analysis of fixation Leaves Water Content.China Tea Leaf Processing, 2013, (3): 33-36.
[3] Xiaodong Liu, Zhoubin Tang. Tea Leaf fixation Machine and Tea-Making Quality Characteristics. Journal of Guangxi Agriculture, 2006, 23 (3): 21-23.
[4] Shengyun Yang, Dehui Yuan, Guoming Lai. A New Clustering Initialization Method. Computer Applications and Software. 2007 (08).
[5] Yanchi Liu. Determination Method of Optimal Cluster Number Based on Density. China Management Informationization. 2011 (09).
[6] Lifang Zhou, Luwen Zhou. A New Evaluation Function Of Multi-Model Modeling Based on Improved K-Means Clustering Algorithm. CIESC Journal, 2007, 44 (23): 2051-2054.
[7] Xiujuan Sun, Xiyu Liu. A New Genetic K-Means Clustering Algorithm Based on Initial Center Optimization. Computer engineering application, 2008, 44 (23): 166-168.
[8] Luwen Zhou, Lifang Zhou. Multi-Model Modeling Method Based On Improved K-Means Clustering Algorithm. Journal of University of Science and Technology, 2005, 23: 62-67.
[9] Ning Li. Research on Some Problems of Multi-Model Modeling and Control. Shanghai: Shanghai Jiaotong University, 2002.