Research and implementation of no-tillage information monitoring strategy based on Hadoop

Hongwei Yang¹, Fucheng Xue¹, Sheng Gao¹, Hongmei Zhu²and Li Li¹*

¹School of computer science and technology, Changchun University of Science and Technology, Changchun City, Jilin Province, 130022, China
²School of electronics and information engineering, Changchun University of Science and Technology, Changchun City, Jilin Province, 130022, China
*Corresponding author’s e-mail: ll@cust.edu.cn

Abstract. Based on the Hadoop, this paper uses linear regression algorithm to study the intelligent monitoring strategy of no-tillage information. In this paper, the regional computers in this system use Flume and Kafka to transfer large amounts of farming information from local node databases to HDFS (Hadoop Distributed File System), then process this data with Spark, whatever, the system uses time series linear regression algorithm to predict the fault of no-tillage. Compared with the no-tillage designed before, the system described in this paper not only reduces time delay but also improves the fault tolerance of data storage and the system stability. Field experiments have proved that the research of this strategy has an accuracy rate of up to 90%, which reduces the huge economic losses caused by machine failure.

1. Introduction
In recent years, with the rapid development of smart agricultural machinery, the design and application of no-tillage machines have matured, which has greatly improved the quality of agricultural sowing and protected the soil structure[1-2]. In order to get the tillage data and failure information of the no-tillage machine in real time, scholars at home and abroad have made special studies on it, as detailed in the literature [3-6]: They added electronic systems to the no-tillage equipment to automatically control the tillage depth, or added side cabinet wheels to equalize the no-tillage cultivation depth. However, with the huge increase in tillage information and the urgent need for real-time processing of tillage information, there is still a lack of an effective no-tillage information monitoring method that allows users to understand the tillage status of no-tillage machines more quickly and intuitively, in order to avoid potential problems and huge economic losses caused by the failure of the tiller.

Based on above, this paper designs an intelligent monitoring strategy based on linear regression, which not only realizes the failure prediction of no-tillage machines, but also uses Hadoop to store and process the massive farming information of no-tillage machines, and finally realizes the display of large screen, App and web pages.

Multiple linear regression model for time series is a machine learning model that can be applied to numerical estimates sorted in time series. Linear regression can fit a linear equation according to the influencing factors. For example, when put the fertilization amount and seeding amount into the linear regression model, the fertilization amount of the no-till planter can be obtained at the next moment. Since the corresponding job information will change gradually with time when an error occurred while the no-till planter is working, people can set a threshold for the fault detection of the no-till planter, and
the threshold should contain the prediction error, so that it is possible to effectively predict the fault of the no-till machine based on the predicted value[7].

The amount of information on no-till planters is huge, especially in the northeast China. If the conventional structured database is used to directly store this information, it will cause system operation delay, increase time and hardware resource consumption [8] In order to avoid the above problems, Hadoop is used to store and process the collected no-till machine operation information in batches, reducing time delay and improving the stability of the system.

2. Research Status
With the gradual development of conservation tillage and the increasing precision requirements for agricultural machinery in agriculture, many papers have proposed methods for improving the structure of no-till seeders from different perspectives, or designing real-time monitoring system for no-till seeders.

The methods proposed in the literature or the use of hardware devices to design and control systems for different types of no-tillage machines can improve the quality of seeding. The real-time working status of no-tillage seeders cannot be obtained, and the working status is opaque to users. Some systems for real-time monitoring of no-till seeders only measure the tillage depth, but the working data layer of the no-till seeders is not comprehensive enough. Some use images for large-scale monitoring, and do not make full use of the working data of no-till seeders. Therefore, this paper proposes to design a set of no-till seeders operation quality monitoring system by combining software and hardware. The Hadoop is used to store massive no-till seeders operation data, predicted the working status of machines, and realize real-time control of no-till seeders operation information.

3. Material and Methods

3.1. Source of Thought
On the one hand, the application of big data can make the agricultural working be more efficient, According to the paper “A review on the practice of big data analysis in agriculture”: The application of big data can not only understand things more comprehensively, but also use this data to foresee the future. It helps people make accurate decisions and reduces losses [9]. Therefore, this paper proposes to establish a no-tillage intelligent seeding monitoring system based on linear regression and Hadoop. The aim is to reduce the economic loss of farmers while checking the operation status of no-tillage machines in real time, with the help of the intelligent no-tillage machine failure rate prediction function of the system.

3.2. Specific Implementation Methods

3.2.1. Data Collecting
During the data collection phase, the system uses LORA communication to transmit the working data information acquired by the collection terminal on the no-tillage machine. At the same time, local nodes are set. At each local node, a computer configured with MySQL. The data collected by the job information collection terminal will be stored in MySQL database. The tillage information of the no-tillage machine specifically includes: seed fall data, working position data, fertilization amount, no-tillage machine speed, plant spacing and line spacing, arable land area and heartbeat detection data of the collecting terminal.

By installing a job information collection terminal on each no-till machine including seed drop detection sensor, load cell, positioning module and Hall sensor to collect data information needed in this system. Among them, the falling seed detection sensor is used to detect the amount of falling seeds, the weighing sensor is used to measure the weight of the fertilizer applied, the positioning module is used to locate the no-tillage machine and obtain the plant spacing and line spacing information, and the Hall sensor is used to detect no-till Machine operating speed. In addition, the system requests the no-tillage
machine's operating status once at a frequency of 2 minutes, performs a heartbeat detection mechanism, and causes the acquisition terminal to return heartbeat detection data. After collecting these related information data, the system transmits the collected data information to local nodes in various places through LORA transmission, and then enables the local nodes to obtain data through the USB interface at the local nodes and stores them to the MySQL database in the local node computer.

3.2.2. Formatting author affiliations
In order to enable the mass operation information of no-tillage machines to be stored on the Hadoop component HDFS (distributed file storage system) from MySQL database, the computers at each local node are set as slave nodes of the Hadoop cluster, and Flume and Kafka are configured to be stored on the regional computers. Among them, the regional computer that collects the operation information of all no-till in a certain area is set as the slave node of the Hadoop cluster, and the server is set as the master node of the cluster. After storing the no-till job monitoring information data in HDFS (Distributed File Storage System).

After storing the no-till job information monitoring data on HDFS (Distributed File Storage System), this system uses Spark to process the data in real time. Specific job information data includes seed drop data, work position data, fertilization amount, no-till speed, plant spacing and line spacing, arable land area, faults, and heartbeat detection data from the collection terminal. After that, storing the processed results in a database and perform visual processing to achieve the purpose of real-time monitoring.

3.2.3. No-tillage machine failure prediction
As a large machine, no-tillage machines will continue to work for a period of time before they stop working and fail. During this time, the operating efficiency of no-tillage machines will change significantly from standard operations. Therefore, the difference can be calculated by predicting the relative amount of no-till seeding and fertilizing at the next moment and the actual value transmitted by the collection terminal at the next moment, and then setting the threshold according to the actual situation. If the difference is greater than the threshold, it proves that the no-tillage machine is about to fail, and the user needs to handle it in time. For the setting of the threshold, the difference between the predicted value and the actual value should be included to avoid the error interference of the prediction model. In addition, according to the field investigation, it was found that the no-tillage machine is dragged by the tractor to work. When the user artificially changes the speed of the tractor, the seeding amount and fertilizing amount will also change significantly. Therefore, the threshold will also be adjusted automatically according to the speed change. In response to this problem, the system considers setting parameters to control the threshold value, so that the effect of different speeds on the no-till seeder and fertilizer application will not affect prediction of failure.

As for the prediction of the sowing amount and fertilizer rate of the no-tillage machine at the next moment, what’s more, because another factor that affects the number of seeds is the speed change of the machine, as shown in Figure 1 and Figure 2, it is different from the situation when the normal machine is running. The amount of seed is greatly related to the speed. Speed is the second factor in this system. The value does not fluctuate greatly under the condition that the no-tillage machine works normally, the relationship model is very simple. Therefore, consider using a time series multiple linear regression algorithm to solve the problem of seeding and fertilization forecasting. Although the multiple linear regression algorithm is considered to be a low-precision prediction algorithm, for the no-tillage machine data with stable data, the linear regression model can build a model based on known good data with simple causality. And as long as the input parameters are accurate, the accuracy rate will be effectively improved, and the effect will be better than other artificial intelligence algorithms.
In this system, the amount that needs to be predicted is the seeding amount and the fertilizing amount, they are all in accordance with the same model. The predicted amount is \( \hat{y}_t \), and \( \hat{y}_t \) is (1):

\[
f(x) = \hat{y}_t = \bar{w}x + b
\]

Where \( \bar{w} \) represents the parameter. Assuming that the samples are independent of each other and follow the same distribution, the error \( \varepsilon \) between the actual values \( y_t \) and \( \hat{y}_t \). As shown in (2):

\[
\varepsilon = y_t - \hat{y}_t
\]

the error \( \varepsilon \) is also independent of each other, according to the central limit theorem, the variance \( \varepsilon \) is subject to a normal distribution with a mean of \( 0 \) and a variance of \( \sigma \). Get (3):

\[
p(\varepsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\varepsilon^{(i)})^2}{2\sigma^2}\right)
\]

Bring the error calculation formula into (4)

\[
p(y^{(i)} | x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)
\]

Use the maximum likelihood estimation method to estimate the value of \( \bar{w} \) to obtain the loss function (5):

\[
L(\theta) = \prod_{i=1}^{m} p(y^{(i)} | x^{(i)}; \theta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)
\]

Go logarithmic (6):

\[
L(\theta) = \log L(\theta) = \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)
\]

\[
= m \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)})^2
\]

Since in (13), \( m \log \frac{1}{\sqrt{2\pi}\sigma} \) and \( \frac{1}{\sigma^2} \) are known, here is an objective function for \( \theta \), \( J(\theta) \) (7):

\[
J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)})^2
\]
Solve for $J(\theta)$ a (8):

$$\theta = (X^T X)^{-1} X^T y$$

(8)

In order to prevent overfitting from adding perturbation values, and to prevent Runge caused by excessive perturbation values, adding the L2 regular term to $\theta$ regularization adds constraints to the loss function.

The L2 regular term is shown in (9):

$$Ridge P(W) = \|W\|^2$$

(9)

Using Ridge Regression to Predict the Working Situation of the No-tillage Machine at the Next Moment and Obtain the Estimate of $\theta$ (10):

$$\theta = (X^T X + \lambda I)^{-1} X^T y + \sum_j \theta_j^2$$

(10)

This system uses Python and Spark in the prediction of no-tillage machine faults, and specifically uses sklearn machine learning pants to build a linear regression model.

The model is now obtained. Extract some normal operating points and fit them to Figure 3. On the way, blue is the actual working point and red is the predicted data. As can be seen from the figure, the model fits well.

In order to ensure the real-time performance of the system, the system obtains the latest data from the database every 2 seconds according to the sending and receiving time interval of the sensor, and predicts the sowing and fertilization amount of the no-tillage machine at the next time, and the reception at the next time. The obtained data are compared to obtain the no-tillage machine failure prediction results.

The test results of the no-tillage machine failure prediction test results and the actual failure results are obtained from the test data, as shown in Figure 4:

![Figure 3. Model fitting diagram.](image)

![Figure 4. Prediction results in test data](image)

The red line in the figure is the working data of the no-tillage machine in the test, and the green line is the predicted data. The program output in the figure is consistent with the actual.

Compared with the model of univariate linear regression, using time series as the influence condition, the accuracy rate has been greatly improved. Using the test data to train the univariate linear regression model and the multiple linear regression model, respectively, the comparison is shown in Figure 5 and Figure 6.
Univariate linear regression that only uses time as a reference will mistake the rule change caused by the speed change as outlier error data, make incorrect predictions on the seeding amount, and have a great error impact on the no-till seeder's failure prediction.

3.3. Visualization

In order to evaluate the method, this paper applied this system to Dongfeng County, Jilin Province, and invested 211 no-trill planters with different tillage conditions to drive the simulated farming. Specifically divided into three regions, which also means that in this experimental scenario, there are three regional computers as slave nodes of the cluster. Figure 7 shows the results displayed on the large screen of the system. The data processing results show that the total area of no-tillage machines is 66 hectares, and the total number of connected agricultural machines is 211, of which the total number of no-tillage machines at work is 208, which is abnormal 2 machines, 1 no-tillage machine in standby, 1 no-tillage machine predicted to be abnormal soon. At the same time, the working area time distribution chart shows the tillage area of the no-tillage machine at different times, allowing users to intuitively understand the specific cultivation of the no-tillage machine.

Figure 8 shows the results of the system on the web page. The data displayed not only includes the work area and the number of seeds, but also the position information of the no-till seeder. Compared with large-screen display, the data displayed to users is more detailed. The working condition of the no-tillage machine is synchronized with the big screen display and displayed to the user: The tillage area of the no-tillage machine is 66 hectares and the number of seeds is 16,775,007.
4. Discussion

This section compares and analyzes the operation monitoring schemes of no-till seeders in Table 1, and compares different design schemes for monitoring the operation of no-till seeders. According to the analysis and research of the previous no-till seeder operation information monitoring system, the existing no-till seeder monitoring systems are all based on hardware design. In those designs, there is no connection between no-till seeders and the data is scattered. Therefore, this system has the advantages of combining hardware and software and applying big data methods to process data. In addition, in terms of obtaining job information, some methods use image processing methods to measure the cultivation effect. The system uses specific seed detection sensors to directly obtain the seeding amount of the no-tillage seeder, and the weighing sensor measures the amount of fertilizer applied is more specific. In this system, a load cell is added to measure the fertilization amount of the no-till seeder, which is the no-till seeder operation information not mentioned in the existing no-till seeder operation information monitoring system.

Table 1. Comparison of this paper scheme with the no-tillage seeding monitoring system in the research status described in this paper

|          | [1] | [2] | [3] | [4] | This paper |
|----------|-----|-----|-----|-----|-------------|
| seeds    | Yes | Yes | Yes | Yes | Yes         |
| location | No  | No  | No  | No  | Yes         |
| speed    | No  | No  | Yes | Yes | Yes         |
| fault    | Yes | Yes | Yes | Yes | Yes         |
| fat      | No  | No  | No  | No  | Yes         |
| spacing  | No  | No  | Yes | Yes | Yes         |
| area     | No  | No  | Yes | Yes | Yes         |

The monitoring method described in this article can more accurately and multi-dimensionally obtain the operation information of the no-till seeder. Each no-till seeder in the monitored area is equipped with a drop detection sensor, a three-in-one positioning module, a Hall sensor and a scale. The heavy sensor transmits measurable job monitoring information, uses HDFS (Distributed File Storage System) to store this information, and provides Spark to process the data. It uses the method of string division to obtain
the number of seeds dropped, the amount of fertilizer applied, fault information, and cultivated land. Speed, position information, plant spacing, and row spacing, and use the plant spacing and row spacing information to calculate the working area of the no-till seeder. Finally, the time series linear regression model was used to predict the failure condition of the no-tillage machine.

5. Conclusion
As a massive data processing solution for agricultural conservation tillage, this system uses Hadoop to store a large amount of no-till machine operation monitoring information data using appropriate sensors, and the number of fallen seeds and fertilization are really useful for fault prediction. According to the experimental data in Figure 2.5 and calling the accuracy_score function, it can be known that the accuracy rate of the model of this system is more than 90%. This shows that the system is more accurate in fault prediction. Through the fault prediction of the no-tillage machine, users can maintain the no-tillage planter that is about to fail in advance and reduce the loss caused by the no-tillage planter failure. In addition, the system's real-time monitoring of the no-tillage arable land speed, plant spacing, row spacing, and farming information are helpful for users to understand the specific operation status of the no-tillage planter and the farming status. The system displays the monitored information to users in the form of APP, large screen, and web client, which facilitates the user's query.

Acknowledgments
This research was funded by the Project of education department of Jilin province, the project name is: Remote Monitoring System of No-Tillage Planting Quality in Cloud Environment grant number JJKH20200801KJ. The authors express their sincere appreciation to the editors and the anonymous reviewers for their helpful comments.

References
[1] Xin Zhao, Shengli Liu, Chao Pu, Xiangqian Zhang, Jianfu Xue, Ran Zhang, Yuqiao Wang, Rattan Lal, Hailin Zhang, Fu Chen. (2016) Methane and nitrous oxide emissions under no-till farming in China: A meta-analysis. Global Change Biology, 22: 1372–1384.
[2] C.Germán Soracco, Rafael Villarreal, Luis Alberto Lozano, Santiago Vittori, Esteban M Melani, Damian Jose Gabriel Marino. (2018) Glyphosate dynamics in a soil under conventional and no-till systems during a soybean growing season. Geoderma, 52: 13-21.
[3] Pasi Suomi, Timo Oksanen. (2015) Automatic working depth control for seed drill using ISO 11783 remote control messages. Computers and Electronics in Agriculture. 31: 30-35.
[4] Rui Zhang, Tao Cui, Dandan Han, Dongxing Zhang, Kehong Li, Xiaowei Yin, Yunxia Wang, Xiantao He, Li Yang. (2016) Design of depth-control planting unit with single-side gauge wheel for no-till maize precision planter. International Journal of Agricultural and Biological Engineering, 9: 56-64.
[5] Honglei Jia, Mingzhuo Guo, Haibo Yu, Yang Li, Xianzhen Feng, Jiale Zhao, Jiangtao Qi.(2016) An adaptable tillage depth monitoring system for tillage machine. Biosystems Engineering, 151: 187-199
[6] Baojuan Zheng, James B. Campbell, Yang Shao, Randolph Wynne. (2013) Broad-Scale Monitoring of Tillage Practices Using Sequential Landsat Imagery. Soil Science Society of America Journal, 77: 1755-1764
[7] Sama Amid, Tarahom Mesri Gundoshmian. (2017) Prediction of Output Energies for Broiler Production Using Linear Regression, ANN(MLP, RBF), and ANFIS Models. Environmental Progress & Sustainable Energy, 36: 577-585
[8] Tong Zhou, Xingguang Lee, Lei Chen. (2018) Temperature Monitoring System Based on Hadoop and VCL. Procedia Computer Science, 131: 1346-1354
[9] Andreas Kamilaris, Andreas Kartakoullis, Francesco X Prenafetabol. (2017) A review on the practice of big data analysis in agriculture. Computers and Electronics in Agriculture, 143: 23-37