Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 122 (2017) 130–137

Procedia Computer Science

www.elsevier.com/locate/procedia

Information Technology and Quantitative Management (ITQM 2017)

Analysis of Grain Storage Loss Based on Decision Tree Algorithm

Xueli Liu [a], Bingchan Li [b], Dongqin Shen [c], Jie Cao [a,c], Bo Mao [a,c,*]

[a] College of Information Engineering, Collaborative Innovation Center for Modern Grain Circulation and Safety, Jiangsu Key Laboratory of Grain Big-data Mining, Nanjing University of Finance & Economics, Nanjing 210023, China
[b] School of Electrical and Automation Engineering, Jiangsu Maritime Institute, Nanjing 211170, China
[c] Computer Science and Technology, Computer Science and Engineering College, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract

Different grain storage factors will cause different degrees of grain loss. In this paper, the data mining method is used to study the loss of grain storage, and the grain loss analysis and forecasting model based on decision tree algorithm is proposed. The paper analyzes and predicts the grain loss caused by different grain storage factors. And the influence of model parameters on model fitting and accuracy is verified by the verification curve. Then the decision tree model is optimized by the method of grid search and cross validation, which improves the prediction accuracy of the decision tree model to analyze the grain loss.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Key words: Decision tree; Grid search; Cross validation; Validation curve; Grain storage loss

1. Introduction

For a long time, farmers' grain storage has been an important supplement to the national grain reserves and played an important role in safeguarding national food security and stabilizing food prices. Due to the impact of various factors for grain in the storage process, it will occur to reduce the number and quality of decline, resulting in grain loss. In [1] the size of the loss is closely related to the maturity of the grain harvest, the harvest period, the moisture content, the storage variety structure and so on after the grain entering the storage link. There are significant differences in the loss of food due to different storage conditions. Therefore, it is important to analyze the loss caused by grain under different conditions.

* Corresponding author. Tel.: +86-25-83493900; fax: +86-25-83493900.
E-mail address: maoboo@gmail.com.

1877-0509
Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.
10.1016/j.procs.2017.11.351

In the face of massive storage of data, how to effectively find valuable information or knowledge, it is a very difficult task. Data mining is to respond to this requirement and developed rapidly. In [2] data mining is to extract people interested in the knowledge from the large database of data, the knowledge is implicit, previously unknown potential useful information, extracted knowledge expressed as the concept, rules, patterns and other forms.

This paper presents a predictive analysis model of grain loss based on decision tree algorithm. With the decision tree algorithm in data mining technology, the paper analyzes and predicts the loss factors of grain storage, and optimizes our decision tree model by using grid search and cross validation method to improve the accuracy of grain loss analysis and forecasting model.

2. Related work

Research has been carried out through the questionnaire survey and empirical analysis of the assessment of the situation of food losses in China. In [3] statistics, China's grain production by the farmers caused by the loss of about 8%. In [4] based on empirical analysis of the post-natal aspects of the loss of the assessment was made that China's grain post-natal losses of about 9% to 16%. In [5] found that especially the storage of farmers, China's grain post-natal losses are higher than the United Nations Food and Agriculture Organization 5% of the grain loss standard. In [6] evaluated the impact of natural disasters on food losses in China, believing that the rate of food loss caused by natural disasters was about 3.3 percent. In [7] food storage losses were investigated based on 5000 farmers of Henan Province, the results show that storage losses of more than 8%, depending on the loss of different varieties of food varieties. Lack of proper storage infrastructure at farm level causes high wastage and loss in value of food grains. In [8] the survey data of 357 households in Jiangxi Province are analyzed, the harvest during the grain loss rate of 3.7% of which the field threshing losses accounted for 3.01%, the whole sun loss accounted for 0.56% of total output; farmers accounted for the average stock of storage of 4.37%, of which the loss of insects accounted for 1.67%, mouse damage accounted for 2.56%, mold powder loss accounted for 0.05%, other losses accounted for 0.08%.In [9] shows that China's annual post-natal losses caused by mildew up to 21 million , accounting for 4.2% of the total grain output. In [10] had studied the post-natal characteristics and impairments of major grain crops in China. In [11] through the wheat grain storage test, compared the performance of different grain storage equipment and found that traditional farmers grain storage equipment can’t effectively use the storage technology, the average annual storage loss of more than 6%, while the new grain storage equipment can be so that grain loss fell below 1%.

The above research has carried on the investigation and the appraisal to the grain loss situation in our country from many aspects. For the grain loss analysis model, in [12] using the least squares support vector machine (SVM) method, a complex nonlinear relationship model between pest occurrence and influencing factors was established to study the distribution of pests in granaries. In [13] use the comprehensive clustering indicators to predict the year of scab caused by wheat yield loss. On this basis, this paper compares the application of classification forecasting algorithm in data mining to the analysis of grain loss, and puts forward a grain loss analysis and forecasting model based on decision tree model to analyze the grain loss caused by different storage conditions.

3. Methodology

3.1. Data preprocessing

The experimental data comes from the 2015 food public welfare industry research project survey data. According to the survey data, it can measure the farmers' grain loss. The data of this analysis is for the loss of grain storage links in the paper, the main selection of the loss, including grain worm and mildew loss. In [14] the