The Research on the Early Warning System of Soybean Import Dependence Degree Based on Decision Tree*

Qiumei Liu[1] and Jun Meng*[1,2]
[1]Science College, Northeast Agricultural University, Harbin Heilongjiang, 150030, China
[2]National study center of soybean engineering, Northeast Agricultural University, Harbin Heilongjiang, 150030, China
Email: merd@mail.neau.edu.cn(*Corresponding Author)

Abstract *The early warning, which is usually built on the base of forecast, is the premise of remove-warning. The paper presents a new warning way-decision tree warning which can warn all kinds of conditions directly. First, the paper sets up a soybean early warning index system by analyzing soybean market of china in nearly the recent ten years. Then, some regulations are extracted according to the warning sign index's effects on the warning alert index in the way of decision tree. At last, the early warning system of the soybean import dependence degree is set up.*

Keywords *Decision tree Soybean import dependence degree Early warning system*

1. Introduction

Looking back the soybean export and import conditions, china is always the soybean net export country before 1995, but from then on, china has became a soybean net import country until to now. At present, soybean market of china is in turmoil, and seriously dependents on the foreign soybean. Foreign soybean has formed a great shock to the soybean production of china. Soybean import dependence degree is identified to the ratio of soybean net import to soybean demand amount. According to the constructed warning sign index system, the soybean import dependence degree is warned, and the degree is divided into non-warning area, gentle warning area, medium warning area and serious warning area. Decision tree, by which extracted regulations are accurate convenient and simple, is the basic content of data mining. So the soybean import dependence degree is researched in the way of decision tree, so that the serious condition in the soybean import can be found ahead, and scientific bases can be provided for macro regulations and control departments to make decision.

2. The Establishing of Warning Index System

The warning of soybean market in china mainly warns the factors which effect the market. The paper combines the soybean demand, soybean price and so on, warns the abnormal condition, and then provides the corresponding precautions. Warning is the basis process to ensure the safe of the soybean market, and only the serious conditions and degrees are timely warned, can we ensure the safety of soybean market in china.

2.1 The Choice of Warning Indexes

Whether the choice of warning indexes is reasonable affects directly the effect and accuracy of warning, so in the process of determining the warning index system of the soybean market, we must consider all kinds of factors which should reflect every side of the soybean market and have representative characters, especially consider the effect of international factors on soybean

* This research has been supported By Education Ministry Emphases Science and Technology Project in Heilongjiang (project number: 1153LZ10) and IRTNEAU.

market of china, so we can present the development condition of soybean market more accurately.

The paper considers overall the soybean production, demand, export, price, country policy and so on. The corresponding data can be seen in Table 1.

Table 1 The corresponding data of soybean market in China from 1995 to 2006

Year	Import dependent degree (%)	Soybean production (ten thousand tons)	Soybean import (ten thousand tons)	Soybean export (ten thousand tons)	Price ratio	Agriculture basis constructing expenditure (a hundred million)	Soybean production in USA (ten thousand tons)	Soybean sowing area (k hectare)
1995	0.02	1350.00	29.8	37.6	0.75	110.0	68444	8127
1996	0.08	1322.00	111.4	19.3	0.75	141.5	59174	7471
1997	0.17	1473.00	288.6	18.8	0.71	159.7	64780	8346
1998	0.18	1515.20	320.1	17.2	0.63	460.7	73176	8500
1999	0.24	1425.00	432.0	20.7	0.59	357.0	74598	7962
2000	0.41	1541.00	1041.9	21.5	0.64	414.4	72224	9307
2001	0.48	1540.70	1394.0	26.2	0.62	480.8	75055	9481
2002	0.41	1651.00	1131.5	30.5	0.69	423.8	78672	8720
2003	0.58	1539.32	2074.1	29.5	0.70	527.3	75010	9313
2004	0.54	1740.15	2023.0	34.9	0.65	542.3	66778	9589
2005	0.63	1634.78	2659.1	41.3	0.58	512.6	85013	9590
2006	0.64	1596.70	2827.0	39.5	0.57	504.3	83368	9279

The data come from web of Statistics of National Bureau of China directly or have been calculated.

Warning alert index is the ratio of soybean net import to soybean demand amount, named as Soybean import dependence degree, which reflects the dependence degree of soybean market in china on foreign soybean market, in which, the soybean demand amount is calculated as follows:

Soybean demand amount = soybean production + soybean import – soybean export

Warning sign indexes are soybean production, the ratio of soybean price in USA to that in china which is named as price ratio, soybean demand amount and agriculture basic construction expenditure, in which,

Price ratio = soybean price in USA × exchange rate of money ÷soybean price in china.

2.2 Decide the Foregoing Years of Warning Sign Indexes

The dividing of foregoing, synchronous and lag indexes is very important in the process of soybean warning, which is the key part of the whole warning system of soybean market. Usually, the approach of time difference relevant analytics is used. The foregoing warning sign indexes, which are the most important sign indexes of the warning system, have a role of predicting the future of soybean market in china. The paper decides the foregoing years by many ways and comprehensive analysis, the foregoing years can be seen in the Table 2.

Table 2 The foregoing years of warning sign indexes

Warning sign indexes	Foregoing years	Warning sign indexes	Foregoing years
Soybean production in china	3	Soybean demand amount	1
Price ratio	2	Agriculture basic construction expenditure	2
Soybean production in USA	3	Soybean sowing area	3

2.3 Dividing Alert Limits and Degrees of Warning Alert Index and Warning Sign Indexes

Dividing the alert limits and degrees is the key part in the design of warning system. Alert degrees reflect the level and intensify of warning contained in the actual values of warning alert index. In order to decide the warning degree of each index, first we ensure the alert limits of each index, and then ensure the prosperity station of each index each year according to the values of alert limits. Consulting Ge Huiling's paper 'The Forecast and Early Warning Research about Soybean Market in China', the station of soybean market import dependant degree and alert limits are described as Table 3.

Table 3 Status description of soybean market import dependent degree

Condition	Feature described	Alert limit
Non-warning area	Soybean market is stable and boom, soybean mainly comes from china, there are little import, doesn't depend on foreign soybean.	[0 0.2]
Gentle warning area	Soybean market is stable and boom, soybean mainly comes from china, there are a little of import, doesn't depend on foreign soybean.	[0.2 0.35]
Medium warning area	Soybean market is relatively stable, soybean mainly comes from china, there are a little of import, depends on foreign soybean in some degree.	[0.35 0.5]
Serious warning area	Soybean market is turbulent, soybean coming from china is a very small part, there were mass import soybean, largely depends on foreign soybean.	[0.5 1]

Making use of feedback theory, the warning sign indexes' limits are decided according to the warning alert index's limits. To represent in the boundary of two warning areas of warning alert index $x(i)$, and the one of warning sign index $y(i)$. So the boundary is between the non-warning and the gentle warning area, or between the gentle warning and medium warning area, or between the medium warning and serious warning area. To represent in the maximal of x and y of $\max_x$ and $\max_y$, the minimal of x and y of $\min_x$ and $\min_y$, then according to the radial principle, the warning area of x and y are decided by the following formula:

$$\frac{\max_x - x(i)}{\max_y - y(i)} = \frac{\max_x - \min_x}{\max_y - \min_y}$$

By calculating, the warning areas of warning sign indexes can be seen in Table 4.

Table 4 The warning areas of warning sign indexes

Warning sign index	Serious warning area	Medium warning area	Gentle warning area	Non-warning area
Soybean production	[1643 +∞)	[1537 1643]	[1441 1537]	(-∞ 1425]
Soybean demand amount	[3677 +∞)	[2944 3677]	[2211 2944]	(-∞ 2211]
Price ratio	[0 0.62]	[0.62 0.66]	[0.66 0.70]	[0.70 +1]
Agriculture basic construction expenditure	[441 +∞)	[337 441]	[233 337]	(-∞ 233]
USA soybean production	[8021 +∞)	[7364 8021]	[6706 7364]	(-∞ 6706]
Soybean sowing area	[9098 +∞)	[8587 9098]	[8076 8587]	(-∞ 8076]

Thought out the data and results, we set up the training table of soybean market import dependence degree in Table 5.

Table 5 Training table of soybean market import dependence degree

Year	Import dependent degree	Soybean production	Soybean demand amount	Price ratio	Agriculture basis constructing expenditure	USA soybean production	Soybean sowing area
1995	non-warning	non-warning	non-warning	non-warning	non-warning	gentle warning	gentle warning
1996	non-warning	non-warning	non-warning	non-warning	non-warning	non-warning	non-warning
1997	non-warning	gentle warning	non-warning	gentle warning	non-warning	non-warning	gentle warning
1998	non-warning	gentle warning	non-warning	medium warning	serious warning	gentle warning	gentle warning
1999	gentle warning	gentle warning	non-warning	serious warning	medium warning	medium warning	non-warning
2000	medium warning	gentle warning	gentle warning	medium warning	medium warning	gentle warning	serious warning
2001	medium warning	gentle warning	gentle warning	serious warning	serious warning	medium warning	serious warning
2002	medium warning	serious warning	gentle warning	gentle warning	medium warning	medium warning	medium warning
2003	serious warning	medium warning	medium warning	gentle warning	serious warning	medium warning	serious warning
2004	serious warning	serious warning	serious warning	medium warning	serious warning	non-warning	serious warning
2005	serious warning	medium warning	serious warning	serious warning	serious warning	serious warning	serious warning
2006	serious warning	medium warning	serious warning	serious warning	serious warning	serious warning	serious warning

3. Decision Tree Warning

3.1 Decision Tree Arithmetic

Decision tree that is a tree structure is similar to a flow chart. Each node designates an attribute, the branch designates the value of the attribute and the leaf designates a category. The highest level node is root. The way of setting up a tree is recursion from up to down.

3.1.1 The Creation of Decision Tree

(1) If all the examples in the trained sample set belong to the same type according to the category attribute, then the set can be called leaf node. The content of the leaf node is the sign of the type.

(2) Otherwise, according to some tactic (information gain ratio of each attribute), a non-category attribute is chosen whose information gain ratio is biggest. By the value of the attribute, the sample set is divided into several subsets.

(3) Then, deals with each subset by the way of recursion till all samples in the same subset have the same value of category attribute.

3.1.2 Choice of Determinant Attribute

About the recursion way of decision tree, usually the attribute which have the most information gain is the test attribute of present node, so the demand information is smallest for sorting the remained test sample set. That is, dividing the sample set by the attribute will make the mixture degrees of the subsets be slowest. However, this calculating way based on information theory is partial to the attribute which has more values, but the attribute is not always the best. So, information gain is represented by gain ratio.Definition of entropy：

$$Entropy(S)=\sum_{i=1}^{c}-p_i\log_2(p_i)$$

S is the training set, c is the number of categories of category attribute and pi is the ratio of the number of samples of category I to that of S. Definition of information gain

$$Gain(S,A)=Entropy(S)-\sum_{v\in Value(A)}\left(|S_v|/|S|\right)Entropy(S_v)$$

Sv is the subset whose value of attribute A is v and Entropy(S) is the entropy of the training set. The information gain of attribute is the reductive expected value after dividing.

Definition entropy of each non-category attribute.

$$SplitInfo(S,A)=\sum_{i=1}^{c}(|S_i|/|S|)\log_2(|S_i|/|S|)$$

S_i is the subset which has the ith value of attribute A in S. for the more average of the attribute divided by attribute value, the bigger of SplitInfo, information gain is represented by information gain ratio to avoid choosing that attribute. Definition of information gain ratio

$$GainRatio(S,A)=\frac{Gain(S,A)}{SplitInfo(S,A)}$$

Gain Ratio is information gain ratio，Gain is information gain，SplitInfo is the entropy of attribute.

3.2 Actual Application of Decision Tree

Putting the training data in Table 5 into the decision tree arithmetic, the information gain ratio of non-category attribute is obtained in Table 6.

Table 6 Information gain ratio of each attribute

Soybean production in china	Soybean demand amount	Price ratio
0.5068	0.4856	0.6652
Agriculture basis constructing expenditure	Soybean production in USA	Sowing area
0.5406	0.4192	0.6475

It can be seen from Table 6, the information gain ratio of price ratio is biggest, so it is the root node, then deal with other attribute by the way of recursion. We obtain the following tree:

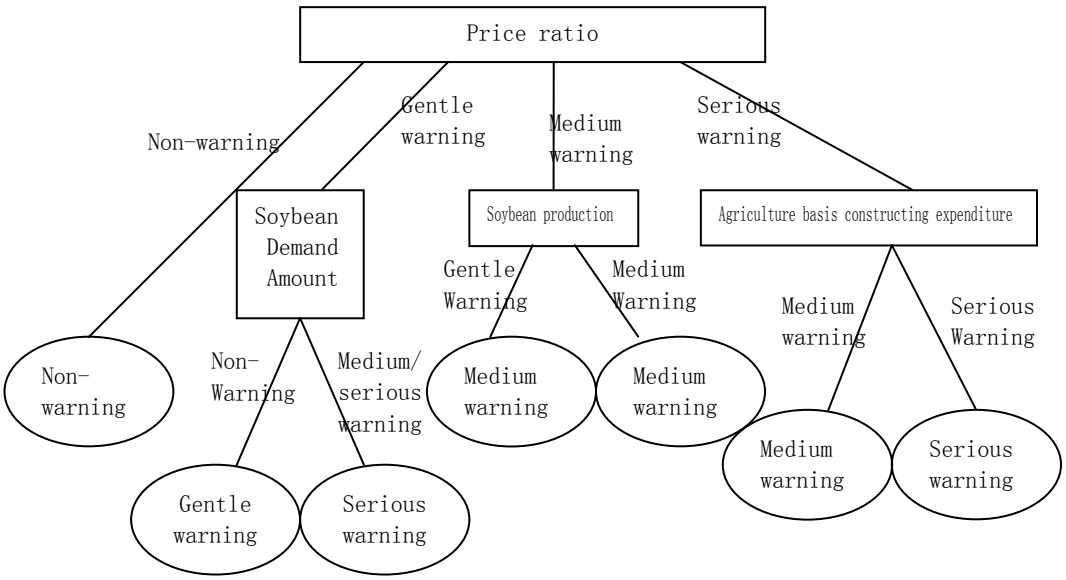

Figure 1 The decision tree

Obtained rules：

If price ratio is non-warning, soybean import dependence degree is non-warning

If price ratio is gentle warning and soybean demand is non-warning, soybean import dependence degree is gentle warning

If price ratio is gentle warning and soybean demand is gentle or medium warning, soybean import dependence degree is serious warning

If price ratio is medium warning and soybean production is gentle warning, soybean import

dependence degree is medium warning

If price ratio is medium warning and soybean production is medium warning, soybean import dependence degree is serious warning

If price ratio is serious warning and Agriculture basis constructing expenditure is medium warning, soybean import dependence degree is medium warning

If price ratio is serious warning and Agriculture basis constructing expenditure is serious warning, soybean import dependence degree is serious warning.

Making use of this tree, the import dependence degree of the soybean market in china in 2007 is tested, and that of 2008 is warned ahead, corresponding warn degrees of foregoing warn sign indexes are in Table 7.

Table 7 The Training table of the import dependence degree of the soybean market in china in 2007 and 2008

Year	Soybean production	Soybean Demand	Price ratio	Agriculture basis constructing expenditure
2007	Serious warn	Serious warn	Serious warn	Serious warn
2008	medium warn	Serious warn	Serious warn	Serious warn

Putting the training data in 2007 into decision tree, we obtain the result: the import dependence degree of soybean market in china in 2007 is serious warning. It has been known that the soybean production in china in 2007 is 1440 ten thousand tons, the soybean imports 3082 ten thousand tons and the soybean exports 39.2 ten thousand tons. By calculating, import dependence degree of soybean is 0.69 and in serious warning area. This accords with the result obtained by the way of decision tree. Furthermore, the paper warns the import dependence degree of soybean market ahead in 2008, and the result is that it is in the serious warning.

The tree comes true the warning by combining the soybean price, soybean production, soybean demand and country policy. USA supplies the most of soybean in the world, and china is one of the biggest import countries. The price ratio has a direct influence on china soybean market. The lower of the ratio, china is more partial to import foreign transgenic soybean in order to satisfy the demand, so resulting in warning affairs.

At the same time, along with the rapid development of economy in china, people's consumptions in albumen and oil are increasing gradually. Because of the limit of the sowing area in china, the soybean production is limited. In recent years, the gap between inland soybean production and market demand amount is being enlarged gradually, following which the amount of soybean import is increasing, so resulting in warning affairs.

Besides, agricultural basis construction expenditure has an obvious effect on soybean market of china, in which storage and construction of traffic have a direct effect on soybean purchase price. If the costs of storage and transportation are reduced, the amount of the soybean import will be less. However, the ability of transportation in some areas of china is very low, which results in the rise of the cost of transportation and indirectly stimulates the rise of the inland soybean price, so the Chinese soybean businessmen import the foreign soybean. Therefore, the investment in agriculture basis construction expenditure is the direct factor influencing soybean import degree. The investment in agriculture basis construction expenditure of china hasn't gone up for 4 years, which accords with the fact of serious warning of soybean import degree in recent years.

In order to improve the present condition of china, first, the change of foreign and native soybean prices should be paid more attentions, and the inspecting degree of import soybean should be intensified at any moment. Then, not only preventing dump of foreign country to soybean market in china, but also taking measures to retort the inferior position of domestic soybean price. At the same time, the policy is necessary and native soybean planting industry should also be stable. At last, the disadvantageous condition of depending foreign soybean should be changed overly by enhancing the completing ability.

4. Conclusion

The basic functions of decision tree are prediction and classifying. Both of the two functions are the basis of warning. First, the foregoing property of the warning sign indexes ensures the success of warning, and future import condition is estimated by the values of warning sign indexes, then the import condition is classified into 4 parts. The paper considers overall the factors affecting on the soybean import condition, so the non-category attributes are a little more, but there were only parts of the attributes which provides more information are used. Therefore, the advantage of decision tree can be seen, that is the way can auto-filter factors and needn't combining other ways.

Because of the limits of the foregoing years of warning sign indexes, the model can only warn the soybean import dependent degree of future in a year. Besides, the number of data in this paper is small and taking use of every province's data in china may be a good thing. The more of the data, the warning result is more accurate.

References

[1] Jun Meng. The Analysis of Soybean Potential Yield in Jiansanjiang Farm by Logistic Model. *Advances in Systems Sciences and Applications*, 2007, 7(1): 143-147.

[2] Jun Meng. The Research of Agricultural Producing Industry Structure's Evaluating and Demonstration Analysis. *Advances in Systems Sciences and Applications*, 2007, 7(2): 173-177.

[3] Fanliang Kong, Zhaowu Ni. A Martingale Method of Perpetual American Options Pricing With Ddefault-Risk. *Advances in Systems Sciences and Applications*, 2007, 7(1): 7-11.

[4] Minfen Shen, Bin Li, Qianhua Zhan, P.J.Beadle. Estimation of Visual ERP Signals with Nonextensive Entropy. *Advances in Systems Sciences and Applications*, 2007, 7(1): 26-31.

[5] Ming Zhu. Data Mining[M]. He Fei: China University of Science and Technology Publishing House, 2002: 67-74. (in Chinese)

[6] Jingmin Chen. Technology of Data Warehouse and Data Mining[M]. Publishing House of Electronics Industry, 2007, 6: 203-206 (in Chinese)

[7] DanYang Cao, JinHong Li, JinQiang Wei, YanFang Zhang. CET-4 Grade Performance Analysis Based on Decision Tree. *Journal of North China University of Technology Beijing China*, 2007, 1: 38-41. (in Chinese)

[8] Xue Bai, Fu Duan. The Study and Application of Decision Tree on Teaching Access. *The development and Application of Computer*, 2007, 20(2): 24-26. (in Chinese)