RESEARCH

CrossMark

Research on precision management of farming season based on big data

Maoling Yan[1], Pingzeng Liu[1*], Fujiang Wen[2], Chao Zhang[1], Rui Zhao[1], Weijie Chen[1], Xuefei Liu[1] and Yuqi Liu[3]

Abstract

In order to strengthen the scientific management of saline-alkali land and the accurate management of agricultural production, a scientific data governance platform for saline-alkali land was developed. Based on the data accumulation of big data platform, the relationship between wheat growth and meteorology was taken as the research object. Thirteen variables including atmospheric pressure, temperature, light, and precipitation were extracted from surface meteorological data for correlation analysis, and temperature, precipitation, and sunshine were selected as the characteristic variables; through discretization processing, we have finally determined the three indicators that can be categorized: accumulative temperature, sunshine hours, and temperature. Finally, the three indicators are combined with months to build a model of farming season and weather based on Apriori. The results show that when judging the farming season with the month as the index, the accuracy of the model is between 78.81 and 100%. When the temperature or accumulated temperature is taken as the index to judge the winter wheat farming time, the accuracy of the model is above 90%. This shows that accurate analysis of farming season can be achieved through big data correlation analysis, which provides technical support for the timely adoption of agricultural production.

Keywords: Agricultural big data, Association rules, Mining analysis, Precise management

1 Introduction

In a natural environment, crop growth and cultivation are restricted by climatic conditions. Field crops not only need suitable climatic conditions but also need suitable farming time. The agricultural season means that each crop has a suitable farming season and suitable farming time. China has a long history of studying agrometeorological science. After a long period of production practice, a method of forming seasons based on solar terms and phenology has gradually emerged. Twenty-four solar terms reflect the annual apparent movement of the sun. The naming of the twenty-four solar terms reflects the seasons, phenology, and climate change [1]. Phenology is the science of studying the relationship between the seasonal phenomena of natural plants and the periodic changes of the environment. It takes full consideration of the growth laws of crops in the wild, taking the development signals of other plants in the environment as a reference, formulating the natural calendar, and exploring the periodic laws of the process of animal and plant development and activity and its dependence on environmental conditions.

With the improvement of the meteorological recording method, in 1735, Dullir first discovered that plants completed their life cycle requiring a certain accumulated temperature, that is, plants from sowing to maturity required a certain amount of daily average temperature accumulation. It can be said that solar terms, accumulated temperature, and phenology are the summary and reflection of the farmers' production experience, and they play a guiding role in agricultural production. However, this agricultural proverb has a long history and is highly targeted in its geographical scope. Accumulated temperature is one of the important indicators for judging the agricultural season, which lays the foundation for the automatic identification of farming season, but the cumulative temperature is only one of the many meteorological factors, and more factors need to be considered. Phenology is more reliable, but it is more dependent on human experience judgments; there are many limitations to

* Correspondence: lpz8565@126.com
[1]College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China
Full list of author information is available at the end of the article

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 2 of 10

achieve automatic identification. Overall, these three methods reveal a strong correlation between the crop growth cycle and agro-climatic resources and require the use of association rule algorithms for mining.

With the advent of the era of big data, the human ability to understand and process data has developed to a new stage. Agriculture is not only the source of big data but also a typical field of big data applications [2]. This is the application and practice of big data theory and technology in agriculture. The application of big data technology in agriculture is conducive to achieving a holographic three-dimensional reflection of the agricultural ecosystem. Big data technology helps to improve people's understanding and insight into the relationships between different things in an agro-ecosystem. With the aid of big data technology, it is also possible to predict the development trend of agriculture and help increase people's management capabilities.

2 The big data platform and data mining method
2.1 "Bohai granary" big data platform

The key factors affecting the growth of crops have the characteristics of complex regularity and space-time change, such as air temperature, light intensity, precipitation and so on, while traditional data acquisition and analysis techniques are difficult to accurately grasp the law of its change. In view of the large data capacity, diversity, and unstructured and redundant features of the "Bohai granary" construction project, a distributed, parallel, and efficient solution is proposed, and a service-oriented platform to support the whole life cycle of large data engineering is constructed [3]. The structure of the big data platform for "Bohai granary" is shown in Fig. 1. Data is the foundation and core of the large data platform. By integrating manual acquisition, automatic acquisition, remote sensing data acquisition, UAV data acquisition, and historical data acquisition methods, rich data resources have been
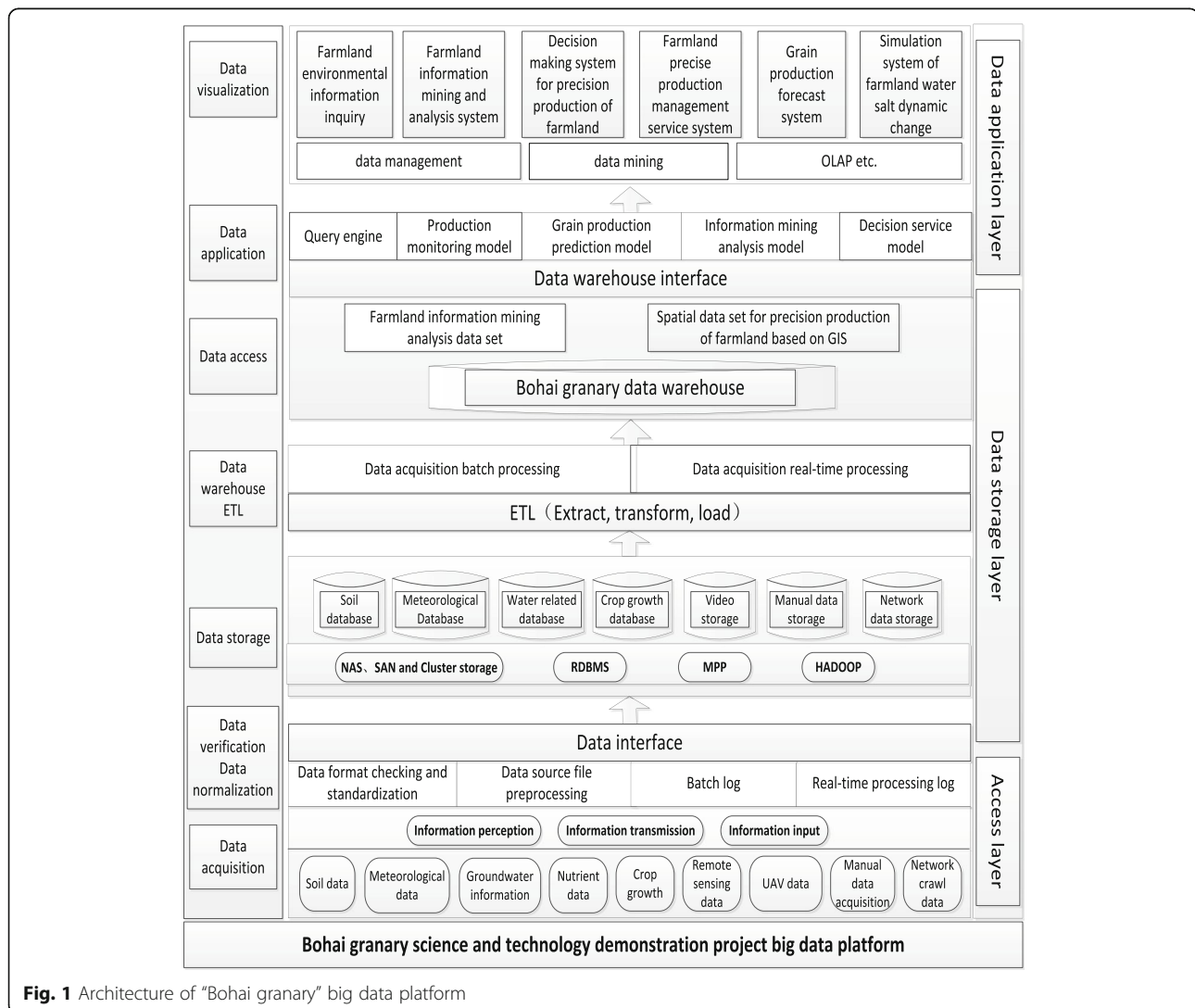

Fig. 1 Architecture of "Bohai granary" big data platform

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 3 of 10

accumulated. This laid the foundation for data mining and analysis services based on big data.

Among them, automatic collection refers to the agricultural microclimate information collected by wireless sensor equipment, including air temperature, air humidity, carbon dioxide concentration, light intensity, and so on. Artificial collection is aimed at recording crop growth stages such as winter wheat sowing time, seedling emergence period, and harvest time. Historical data refers to the meteorological and agricultural operations accumulated over the years in the region, in which meteorological data mainly come from the meteorological station, which usually reflects the average meteorological conditions of the city, and the statistics of agricultural data are carried out on the 10-day scale, which usually reflects the situation in the administrative region. Remote sensing data refer to the aerial map information obtained from satellite remote sensing technology, such as geomorphology, soil, and geology. UAV data refer to the hyperspectral and infrared data obtained from UAV aerial photography, which is used to analyze the pathological information of crop.

2.2 Big data analysis mining method

Data mining analysis is the core of big data's technical field, which directly produces value. It is the calculation process of discovering big data's centralized data model and aim to extract hidden value from a large number of incomplete, noisy, fuzzy, and random actual data that people do not know in advance, but also potentially useful information and knowledge processes that assist people in making more scientific and intelligent decisions. Data mining is a comprehensive application of statistics, database technology, and artificial intelligence technology [4]. It is a set of techniques that extract patterns from big data by using statistics and machine learning methods in database management system. The functions of correlation analysis, cluster analysis, prediction, classification, association analysis, sequence analysis, and so on are realized.

(1) Correlation analysis. Correlation is the law that exists between two or more variables in a certain sense. Its purpose is to explore the hidden correlation in the dataset. The commonly used correlation analysis methods include Pearson correlation coefficient method, canonical correlation analysis method, and principal component analysis method.

(2) Cluster analysis. Clustering analysis refers to the process of grouping sets of physical or abstract objects into groups of similar objects. The commonly used clustering methods are k-means clustering, k-center point clustering, Fisher clustering, Kohonen neural network, and fuzzy clustering.

(3) Predictive analysis. Predictive analysis is a statistical or data mining solution that includes algorithms and techniques that can be used in structured and unstructured data to determine future results. There are many analytical methods for prediction, such as prediction and simulation. The commonly used methods are mean value method, smoothing index method, linear programming, nonlinear regression, logistic regression, BP neural network, and support vector machine.

(4) Classification is to find out the common characteristics of a set of data objects in the database and divide them into different classes according to the classification pattern. The purpose of classification is to map the data items in the database to a given category through the classification model. The common classification methods are decision tree method, Bayes classification method, k-nearest neighbor, rough set, boosting and support vector machine, and so on.

(5) Association analysis. Association analysis is to find frequent patterns, correlation, or causality structure between sets of items or objects in transaction data, relational data, or other information carriers. The commonly used association analysis methods are Apriori algorithm and FP-Growth algorithm.

(6) Time series analysis is a statistical method of dynamic data processing. Based on stochastic process theory and mathematical statistical method, the statistical law of random data sequence is studied to solve practical problems. The commonly used time series analysis methods include AR model, MA model, ARIMA model, and time series analysis method based on prediction analysis.

(7) By analyzing the above data mining methods, we can see that the research focus of association analysis is to find out the correlation that exists in a large number of datasets and then to describe the rules and patterns of some attributes in the project at the same time. In the process of crop planting and production, climate attributes, soil environmental attributes, and other attributes that can affect crop growth can be mined by association analysis. Each crop has a certain farming season and time, which is, farming season, which plays an important guiding role in the production of food crops. We discuss the accurate judgment and management of farming season through the basic correlation analysis of the existing meteorological data and farming season data. Thus, the correlation model of farming season and meteorological data is established, and the intelligent judgment of farming season is realized.

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 4 of 10

2.3 Association rule mining

Association knowledge refers to the dependency or association that reflects an event and other events. Association analysis is also called association rule mining, association rule, in short, is like X → Y. The implication of X → Y means that Y can be deduced by X, where X and Y are called leaders of association rules, namely, antecedent or left-hand side (LHSs) and subsequent sequent or right-hand side, RHSN, respectively.

Machine learning, on the other hand, is dedicated to studying how computers simulate or implement human learning behavior, learning by analyzing large amounts of data, finding patterns in the data, and using these patterns to make predictions in order to acquire new knowledge or skills. Association rule mining is an algorithm for discovering association knowledge in machine learning.

In this study, the association knowledge mining based on the combination of machine learning and association analysis is used. The mining process of association rules mainly includes two stages: the first stage is to find out all the high-frequency item groups from the massive raw data; the second stage is to generate association rules from these high-frequency project groups. The commonly used algorithms are Apriori algorithm and FP-tree based on frequent pattern tree (FP-tree). Among them, Apriori association rule mining algorithm is one of the most influential algorithms for mining frequent itemsets of Boolean association rules, which has been widely used in user or consumer preference analysis [5, 6] and correlation analysis of economic indexes, medicine and pharmacy [7–9], and transportation [10]. In recent years, many applications have been made in agriculture, which has a broad application prospect.

The Apriori algorithm was proposed by R. Agrawal and R. Skrikant in 1994. The main idea of Apriori algorithm is based on prior knowledge, that is to say, the frequent items are generated according to the frequency items found in the previous time. The two concepts of "support degree" and "confidence degree" are used to quantify the association rules between things. They reflect the usefulness and certainty of the discovered rules respectively. Support degree support = P (AB) means the probability of event A and event B occurring simultaneously. The probability of occurrence of event B on the basis of event A is defined as the probability of occurrence of event B confidence = P (B|A) = P (AB)/P (A). The rule that satisfies both the minimum support threshold and the minimum confidence threshold is called the strong rule. First, the samples are divided into two independent training sets and test set. The correlation between crop farming time and weather is studied; the principle of which is shown in Fig. 2.

It iterates over the frequent itemsets in transaction databases by iterating method.

The main implementation steps of the Apriori algorithm are as follows:

(1) Given a minimum support threshold MINSUPPORT, the credibility threshold is MIN_CONF.
(2) Scan the transaction database D, generate candidate item C1, prune frequent 1 itemsets according to MINSUPPORT, and get frequent 1 itemsets L1.

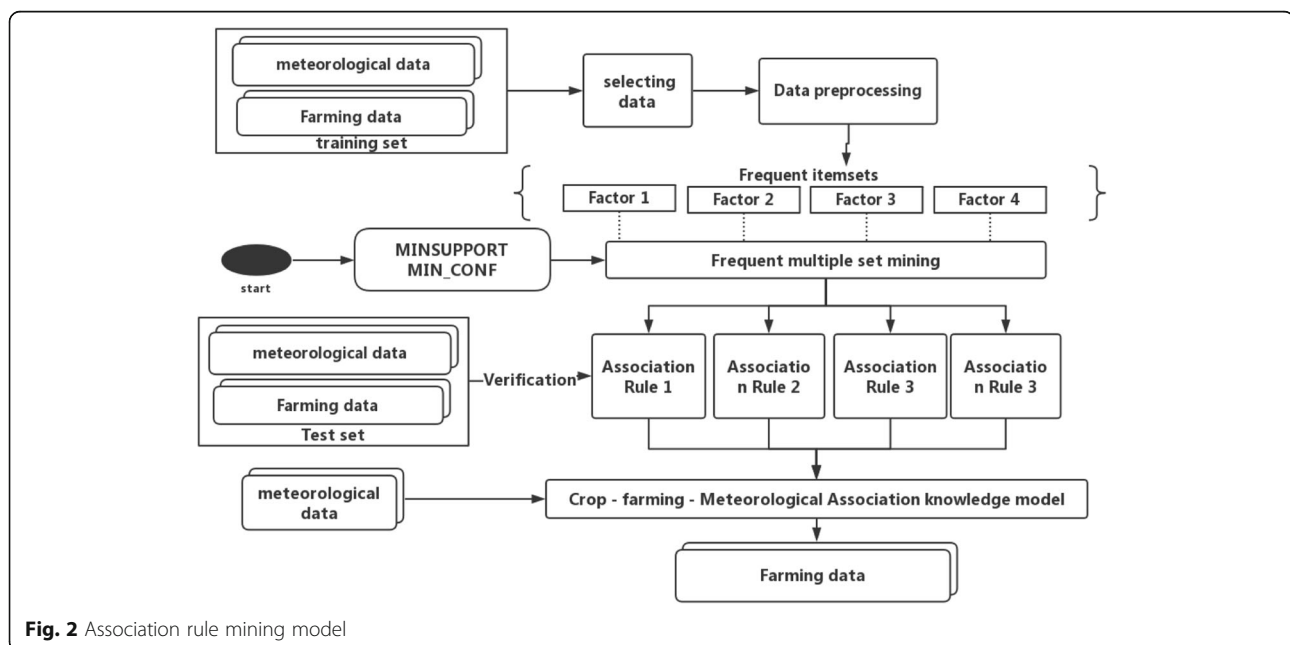**Fig. 2** Association rule mining model

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 5 of 10

(3) According to L1, we get the candidate 2 itemset C2, then prune the C2 according to MINSUPPORT, and generate frequent itemsets L2.

(4) Repeat iterations until the most frequent item Lk is found.

(5) Mining all strong rules that are greater than or equal to MIN_CONF from frequent itemsets L, that is, strong association rules.

The principle is used in the above steps to find a candidate 2—itemset C2 based on L1 is that if an itemset is frequent, then all its subsets are also bound to be frequent.

Apriori algorithm is the core algorithm of association rules. It is mainly used to mine association rules between items in the transaction database. However, in the big data environment, the computational efficiency of Apriori algorithm has obviously decreased, so scholars have done a lot of research to improve the efficiency of the algorithm. J.S. Park uses the hash technology to optimize the classical Apriori algorithm [11]; the main improvement is to simplify the generating process of the candidate 2 itemsets. J. Han and other proposed a new optimization method: FP-growth algorithm; the improved algorithm does not produce candidate itemsets when mining frequent itemsets. But after scanning the transaction library, the original transaction library is compressed and a FP-tree is constructed, which effectively improves the low efficiency of the classic Apriori algorithm [12]. A. Savasere improves the classical Apriori algorithm according to the partition principle and proposes a new algorithm partition [13]. The algorithm only needs to scan two transaction libraries to find all the frequent itemsets and improve the operation efficiency of the classical Apriori algorithm.

3 Data sources and data processing

3.1 Data preprocessing

The big data platform database of "Bohai granary" collected meteorological data from 2001 to 2016 in Yanzhou, Shandong province, and the corresponding statistical data of wheat growth stage. The whole growth period of wheat is divided into seeding, seedling emergence, tillering, wintering, returning to green, jointing, heading, booting, and flowering. Grain filling period of wheat, ripening, and the data in the database are 10-day granularity, covering the period from September to late

June (T1–T28), because the time between jointing to booting, booting to heading, heading to flowering, and so on is not more than 10 days, there is a great deal of overlap in the statistics of grain size of 10 days. For this, the whole production process is divided into nine stages (numbered F1–F9) to minimize the overlap of wheat growth stages, and the annual cycle distribution is different due to the different weather conditions, as shown in Table 1.

The daily meteorological data of Yanzhou mainly include the following attributes: area station number, year, month, day, average station pressure, daily maximum station pressure, daily lowest station pressure, average temperature, daily maximum temperature, daily minimum temperature, average water vapor pressure, the mean relative humidity, minimum relative humidity, precipitation, small evaporation, large evaporation and sunshine hours, 13 different characteristic variables, covering time, air pressure, temperature, vapor pressure, humidity, precipitation, evaporation, and sunshine hours. All the eight main aspects can be regarded as time series data with day granularity.

3.2 Correlation analysis

Firstly, the data and meteorological data of wheat growing stage were preprocessed, the missing data and abnormal value were eliminated, and the correlation analysis was made on the daily meteorological data (as shown in Table 2). The results show that the temperature is highly correlated with the air pressure and evaporation capacity, which can be replaced by more commonly used temperature indexes. The humidity and sunshine hours have a significant correlation. The sunshine and the precipitation have a micro-correlation, and the precipitation is related to the sunshine hours and the temperature. Thus, the final selected indexes are temperature, sunshine, and precipitation.

Research shows that during the growth period of crops, not only the temperature level is required, but also the sum of the total heat is needed. The sum of the total heat is usually expressed by the cumulative value of daily temperature in this period. This cumulative value is called accumulated temperature. Under the premise that other environmental conditions are basically satisfied, there is a positive correlation between temperature and the growth speed of biological organisms in a certain temperature range. The species, variety, and growth

Table 1 Division of winter wheat growth stages

F1	F2	F3	F4	F5
Sowing stage	Emergence stage	Stooling stage	Overwintering stage	Seedling establishment stage
F6	F7	F8	F9	
Jointing and booting stage	Heading flowering stage	Grain filling period	Mature stage	

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 6 of 10

Table 2 Correlation analysis of meteorological factors

Related items	Correlation coefficient	Related items	Correlation coefficient
Precipitation and sunshine duration	0.2381	Mean pressure and temperature	− 0.8851
Minimum relative humidity and sunshine duration	− 0.5755	Daily maximum pressure and temperature	− 0.9017
Minimum relative humidity and precipitation	0.3080	Daily minimum pressure and temperature	− 0.8641
Mean relative humidity and sunshine duration	− 0.4388	Daily maximum temperature and temperature	0.9781
Average relative humidity and precipitation	0.2655	Daily minimum temperature and temperature	0.9771
The average water vapor pressure and temperature	0.9100	Daily precipitation and temperature	0.0710
Small evaporation and temperature	0.8544	Sunshine duration and temperature	0.1710

[0.00, ± 0.30] related micro; [0.30, ± 0.50] real; [± 0.50, ± 0.80] significantly correlated; [± 0.80, ± 1] highly correlated

period of the species are different, and there are also differences in the starting temperature of the growth, that is, the lowest temperature at the beginning of growth and development. Only when the average temperature of the day is higher than the starting temperature of birth, the temperature factor can promote the growth and development of biological organisms. This reproductive initiation temperature is known as the biological lower limit temperature (also known as biological 0°). It is an index to study the relationship between temperature and the speed of organism development. It shows the effect of temperature on the growth and development of a biological organism from two aspects of intensity and action time. Therefore, as another form of temperature, it is necessary to add the accumulated temperature to the weather data. So, the temperature indicator includes mean daily temperature (°C) and the accumulated temperature (°C) of the two types of indicators, which accumulated from the beginning of September 20th; the average temperature is greater than the cumulative value of 0 °C. The precipitation index refers to the daily amount of precipitation (mm), and the sunshine duration index refers to the length of the daily sunshine duration (H hours).

4 Association mining and modeling
4.1 Discretization
The experimental data include temperature, sunshine duration, accumulated temperature, and farming season data from 2001 to 2016, with a total of 4234 data. First, the sample is divided into two separate parts: train set and test set according to the ratio of 8: 2; the data from 2001 to 2012 are taken as the training set, and the data from 2013 to 2016 are taken as the test set to mine the association rules between weather and winter wheat farming time.

Before the experiment, we should classify the selected indexes and discretize the existing meteorological data to minimize the occurrence of missed judgment. We use

(1) to calculate the rank of each factor; the statistical analysis is shown in Table 3 and Table 4.

$$d_i = \frac{\max(X_i) - \min(X_i)}{\text{count}(C)},$$

C is the count of development stages of winter wheat.

(1)

For the convenience of statistics, the contrast value is rounded, the temperature index is taken every 5 °C as a grade, the accumulated temperature is 300 °C as a grade, the precipitation is taken every 15 mm as a grade, and the sunshine is taken every 1 h as a grade; the probability distribution is shown in Table 5.

The original data include month, temperature, accumulated temperature, precipitation, and illumination, among which the items are divided into subsets according to the grade. The data format is temperature, accumulated temperature, precipitation, sunshine duration, month, and wheat growth stage.

If a subset of an itemset is not a frequent itemset, then the itemset is certainly not a frequent itemset. Therefore, the selected meteorological factors need to have better distribution effect after discretization classification to satisfy the calculation of support and confidence; the specific distribution is shown in Table 4. The precipitation grade 0 accounts for 98%, and the grade division is meaningless, indicating that the annual precipitation in Yanzhou area is less, and precipitation cannot be taken

Table 3 Meteorological factor statistics and grading

Index	Temperature	Accumulated temperature	Precipitation	Sunshine duration
Max	32.4	3120.1	144.8	13.5
Min	− 10.7	193.5	0	0
Range	43.1	2926.6	144.8	13.5
Mean	5	325.17	16.09	1.5

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 7 of 10

Table 4 Statistic of wheat growth stage

Growth stage	Count	Probability
F1	220	0.06
F2	326	0.08
F3	520	0.13
F4	1044	0.26
F5	609	0.15
F6	317	0.08
F7	284	0.07
F8	264	0.07
F9	398	0.10

as an index, which is abandoned here. Only three meteorological factors remain: temperature, sunshine duration, and accumulated temperature.

4.2 Model construction

The following training data is used to build a model by using related features.

1) Initialize the original dataset

$$X'_i = \left[\frac{X_i}{d_i}\right], [\ ] \text{for the integer operator} \qquad (2)$$

The event of temperature is set as event A, and the event of temperature grade i is A_i ($-2 \le i \le 6$).

The event of accumulated temperature is set as event B and the event of class j of accumulated temperature as B_j ($0 \le j \le 10$).

Table 5 Probability distribution of index grading

	Precipitation	Sunshine duration	Temperature	Accumulated temperature
− 2			2	
− 1			60	
0	4150	949	1321	235
1	50	89	651	304
2	21	113	602	959
3	5	154	757	1278
4	3	132	561	355
5	1	169	264	264
6	0	233	14	234
7	0	423		205
8	1	717		177
9	1	435		160
10		399		61
11		263		
12		140		

The event of sunshine duration is set as event C and the event of class k of accumulated temperature as C_k ($0 \le k \le 12$).

The event of precipitation is set as event D and the event of class n of accumulated temperature as D_n ($0 \le n \le 9$).

The event of month is set as event M and the event of class x of accumulated temperature as M_x ($9 \le x \le 12$ and $1 \le x \le 6$). Each month is divided into three stages, each of which is 10 days. For example, the first phase of September is expressed as M9.1.

The event of farming season is set as event F as shown in Table 1.

2) Input sample set S = {$(A_i, B_j, C_k, D_n, F_c, Y_n)$}
3) Computing frequent itemsets

In this step, the number of frequencies that appear at each level under the three sets of temperature, accumulated temperature, and sunshine duration is required to be counted respectively, and the subsets are formed.

With the time granularity of 10 days, the same wheat growth stage may cover multiple temperature or accumulated temperature levels. Therefore, the setting of support and confidence needs to take fully into account the overlap transition of wheat growth stage. Set the minimum support level for MINSUPPORT = 50; confidence MIN_CONF = 0.5. Support is the number of values in the candidate set, confidence is the ratio of the number of cycles contained in the candidate set to the total number of the cycles. Traversing discrete datasets and frequent itemsets, these itemsets appear at least as frequently as the predefined minimum support.

4) Connecting step

Frequent binomial sets can be generated by connecting frequently between one itemset and satisfying the minimum support degree and minimum confidence level. In order to avoid repeated calculation, only the link between meteorological factors and wheat growth stage is needed; the connection between meteorological factors can be saved so that the calculation time can be saved, and the calculation speed can be improved.

5) Pruning step

Because the time span of each growth stage is different, there is also the situation of crossing multiple levels on the same index. In order to avoid errors and omissions, it is necessary to improve the calculation of support degree in the traditional Apriori algorithm. By using the method of segmentation and integration, all frequent subsets are first calculated then classified; the

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 8 of 10

independent itemsets in the same category are accumulated, and then the itemsets are used as objects to judge the support degree, instead of subsets.

The pruning is based on the degree of support and confidence. The principle is that the frequency of each temperature or accumulated temperature level should be at least more than 50, and the continuous value of temperature and grade should account for at least 50% of the growth stage of the wheat.

$$A_i \geq \text{MIN\_SUPPORT} \tag{3}$$

$$P(A) = \sum_{i}^{j} p(A_i) \tag{4}$$

$$\frac{P(A)}{P(Y_n)} \geq \text{MIN\_CONF} \tag{5}$$

Because the length of each growing period is different, there are several temperature classes that correspond to a growing phenomenon. Therefore, it is not easy to prune each sub-itemset. Therefore, it is necessary to accumulate statistics for each sub-itemset and perform pruning processing on the itemset as the standard, as shown in Fig. 3, so that the integrity of the index can be better ensured.

5 Results and discussion
5.1 Results
Use the association rule model in 4.2 to match the test set to verify the accuracy of the model, as shown in Table 6. First of all, because the probability of
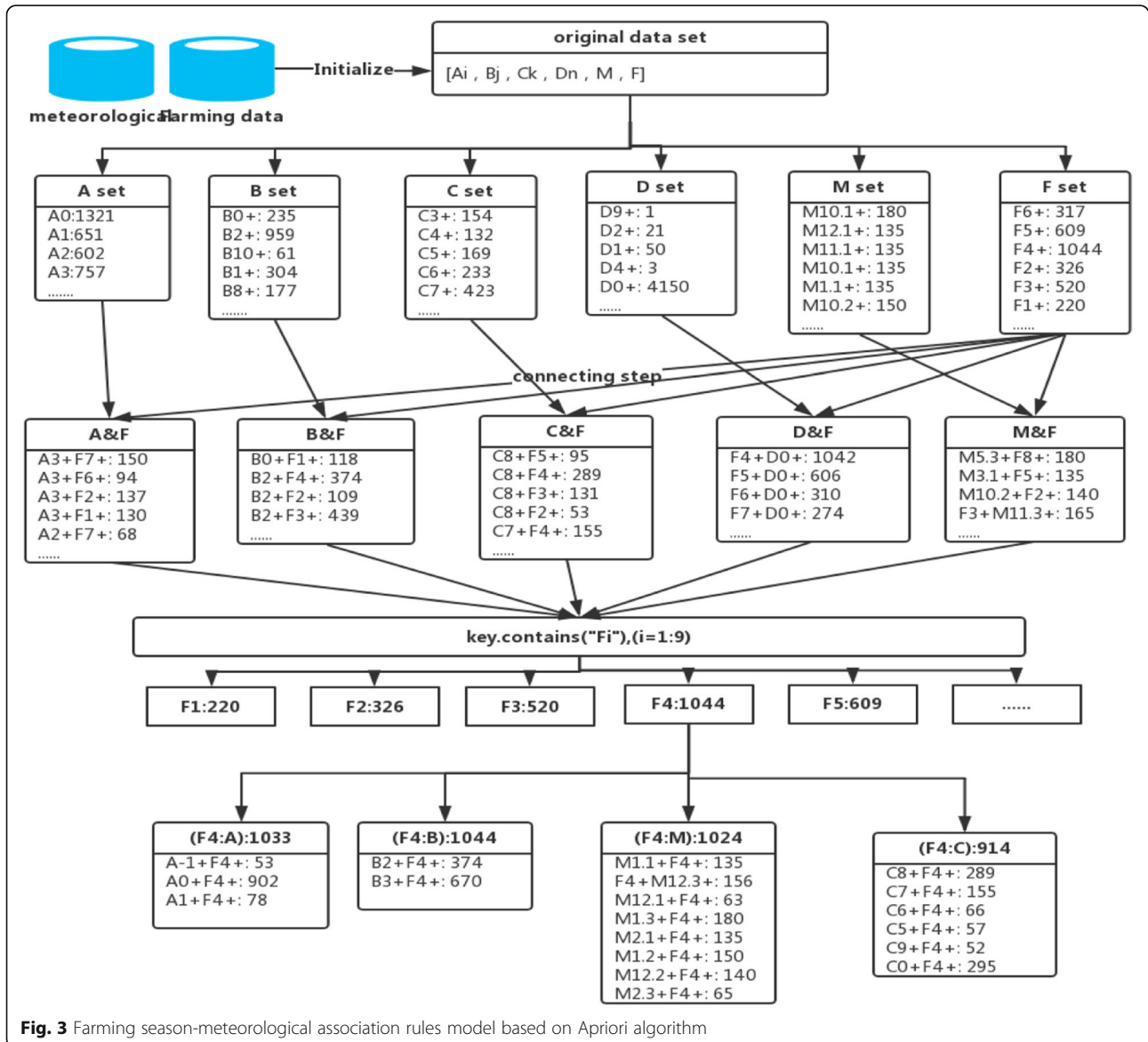

Fig. 3 Farming season-meteorological association rules model based on Apriori algorithm

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 9 of 10

Table 6 Analysis of model test results

	F1	F2	F3	F4	F5	F6	F7	F8	F9
M range	(M9.3, M10.1)	(M10.2, M10.3)	(M11.1, M12.1)	(M12.1, M2.3)	(M2.2, M3.3)	(M3.2, M4.3)	(M4.2, M5.1)	(M5.1, M5.3)	(M5.2, M6.3)
Conf. (%)	78.18	83.43	91.54	98.08	100	100	81.34	88.64	94.97
A range	(15, 20)	(10, 20)	(0, 10)	(− 10,5)	(0, 15)	(5, 20)	(10, 25)	(15, 25)	(20, 30)
Conf. (%)	59	84.04	95.96	98.95	93.59	88.96	94.36	87.88	95.48
B range	(0, 600)	(300, 900)	(600, 1200)	(600, 1200)	(900, 1500)	(1200, 1800)	(1200, 2100)	(1800, 2400)	(1800, 3000)
Conf. (%)	92.72	96.32	98.07	100	96.55	97.48	100	81.44	89.20
C range	–	(8, 10)	(7, 9)	(5, 10)	(7, 11)	(10, 12)	(10, 12)	–	(10, 13)
Conf. (%)		36.50	42.5	59.29	55.99	36.59	36.62		43.21

occurrence of sunshine itemsets meets the minimum confidence and minimum support at the same time, it is less likely to meet the requirements of strong association rule generation. It is no longer considered here. From Table 6, we can get that the index closely related to the growth stage of wheat is temperature and accumulated temperature.

There is the probability of 78.18% that the sowing period occurred between late September and early October, the probability of 59% covered the temperature range of 15°–20 °C, and the probability of 92.72% appeared at the accumulated temperature of 0°–600 °C.

The temperature of the seedling stage is about 10° to 20 °C, the time range is from the middle of October to late October, and the accumulated temperature distribution is 300°–900 °C. The confidence degree is above 80%. There is an overlap between temperature and accumulated temperature in seedling emergence and sowing.

The regularity of the tillering stage is that from early November to early December, the temperature is greater than 0 °C, the accumulated temperature is between 600° and 1200 °C, and the confidence of the three indexes exceeds 90%. There is an overlap of accumulated temperature between tiller and seedling emergence.

The wintering period mainly occurred from early December to late February, the temperature was below 5 °C, the accumulated temperature had no change compared with the tillering stage, and the confidence degree of the three indexes was above 98%. The temperature overlap between overwintering and tillering existed.

When the temperature is more than 0 °C, there is a transition from overwintering to turning green. The confidence degree of the three indexes of the period is above 90. There is an overlap between the time, temperature, and accumulated temperature between the green stage and the overwintering period. According to the principle of natural transition of winter wheat growth, the overlapping index region was taken out as the transition period of two adjacent cycles.

5.2 Model application

Big data platform makes full use of the Internet of things automatic sensing technology to obtain meteorological, soil, hydrological, and other comprehensive information. Up to now, Yanzhou project area has accumulated the data collected automatically by the Internet of things in 2014–2015 and 2015–2016. Here again, using the Apriori algorithm, the collected meteorological monitoring data are brought into the winter wheat farming season correlation model to determine the site of the site block of winter wheat farming time. Setting a minimum support of 50, it takes at least 5 days in 10 days to meet a certain cycle index before it can be considered a certain growth cycle. The temperature and accumulated temperature are obtained from the first step data processing in the growth cycle of winter wheat, a triple indicator of the date.

First of all, the time and temperature indexes involved in association rules are extracted from the data collected automatically from the multidimensional Internet of things; the time and temperature indexes are standardized, and the average temperature, accumulated temperature, and data element group are obtained. The association rule model is introduced to judge the growth cycle of winter wheat. Finally, the results are compared with the artificial statistical data of the Internet of things site, and the accuracy of the model is proved to be more than 95%.

6 Conclusions

In view of the low degree of automation in farming time judgment, this paper comprehensively compares the traditional farming season judgment method and studies how to use the data as the center to excavate the strong correlation between the winter wheat growth cycle (or the farming time) and the meteorological factors by the self-learning method.

Yan *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:143

Page 10 of 10

1) Based on the multi-source data accumulation of the agricultural big data platform, the key indicators for judging farming season are refined. A method and model for automatic farming season determination based on big data are proposed. This provides new ideas and methods for the precise management of agricultural production.

2) A self-learning, adaptive farming season, and meteorological correlation model was constructed. The growth stage of winter wheat could be automatically identified under the current farmland environment by using the real-time data collected from the field. Compared with artificial judgment, the automatic identification model can estimate the farming time with the accuracy of 90%.

3) The experiment proves that it is feasible to use the association rule model to predict the farm hour automatically. The experimental results, on the one hand, realized the dynamic judgment of the time and duration of the winter wheat farming season. On the other hand, it made up the shortage of agricultural record in agricultural production. This provides early warning and forecast information for local conditions and timely operation of farming.

With the popularization and application of the Internet of things technology in agriculture, more comprehensive and accurate information will be obtained in the future. On this basis, a more accurate model of association rules can be built to provide a more scientific and systematic technical support for precision agricultural production. Next, we plan to apply the automatic extraction of agricultural knowledge based on association rule mining to other regions and other crops.

Availability of data and materials
The simulation data can be downloaded at http://bhlc.sdau.edu.cn/.

Authors' contributions
MLY is the main writer of this paper. She proposed the main idea, designed the farming and meteorological association mining experiment, completed the simulation, and analyzed the result. PZL and FJW build this big data platform framework. CZ, RZ, and WJC helped to complete the analysis of the results of the experiment. XFL and YQL assisted in the collection and preprocessing of the data. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China. [2]Shandong Agricultural University, Tai'an 271018, China. [3]Wake Forest University School of Business, Winston-Salem, USA.

Reference
1. J Wang, Solar times, phenology, farming proverb and old farmer—the functioning mechanism of farming activity in Jiangnan of Modern Times. Ancient and Modern Agriculture (2005)
2. XF Li, SH Chen, LF Guo, Technological innovation of agricultural information service in the age of big data. J. Agric. Sci. Technol. **16**(4), 10–15 (2014)
3. S Wang, P Liu, Z Zhang, et al., in *Journal of Chinese Agricultural Mechanization*. Development of management methods for "Bohai Sea Granary" data (2016)
4. XU Shi-Wei, DJ Wang, LI Zhe-Min, Application research on big data promote agricultural modernization. Sci. Agric. Sin. **48**(17), 3429–3438 (2015)
5. Y Guo, M Wang, X Li, Application of an improved Apriori algorithm in a mobile e-commerce recommendation system. Industrial Management & Data Systems **117**(2) (2017)
6. GX Yang, The research of improved Apriori mining algorithm in Bank customer segmentation. Adv. Mater. Res. **2542**(760), 2244–2249 (2013)
7. W Zhang, D Ma, W Yao, Medical diagnosis data mining based on improved Apriori algorithm. Journal of Networks **9**(5) (2014)
8. Y Han Zhou, P Tian, An improvement of Apriori algorithm in medical data mining. Appl. Mech. Mater. **3458**(631) (2014)
9. MH Kuo, et al. Application of the Apriori Algorithm for adverse drug reaction detection. Studies in Health Technology and Informatics. **148**, 95 (2009)
10. S Prakash, An effective network traffic data control using improved Apriori rule mining. Circuits and Systems **07**(10) (2016)
11. JS Park, MS Chen, PS Yu, An effective hash-based algorithm for mining association rules. ACM SIGMOD Rec. **24**(2), 175–186 (2016)
12. J Han, J Pei, Y Yin, Mining frequent patterns without candidate generation//ACM SIGMOD International Conference on Management of Data. ACM, 1–12 (2000)
13. A Savasere, E Omiecinski, SB Navathe, An efficient algorithm for mining association rules in large databases//International Conference on Very Large Data Bases. Morgan Kaufmann Publishers Inc., 432–444 (1995)