Optimal Pattern Discovery to Reveal the High Yield Inhibition Factor of Soybeans

Namba, Midori ; Umejima, Kohei ; Nishide, Ryo ; Ohkawa, Takenao ; Ozawa, Seiichi ; Murakami, Noriyuki ; Tsuji, Hiroyuki

(URL)
https://hdl.handle.net/20.500.14094/90004850

Paper

Optimal pattern Discovery to Reveal the High Yield Inhibition Factor of Soybeans

Midori Namba*†   Member,   Kohei Umejima*   Non-member
Ryo Nishide*   Member,   Takenao Ohkawa*   Non-member
Seiichi Ozawa*   Non-member,   Noriyuki Murakami‡   Non-member
Hiroyuki Tsuji‡   Non-member

Abstract: Our research group is working on soybeans which the quantity of yielding is difficult to predict. We focus on the common characteristics observed at multiple cultivation points, in order to examine methods to acquire new knowledge in deciding the work based on the amount of yields. Our previous study has examined a method to discover optimal patterns using qualitative value of cultivation data. In this method, one qualitative value is assigned based on the boundary value. However, by typically generating qualitative values around the boundary value, different qualitative values are given, even if they are almost the same value, which tremendously affect the calculation of the optimal patterns. In this study, by considering ambiguity in boundary values when generating qualitative values, we discover new knowledge that was not found by the previous method. This method was applied to actual data. As a result, it was verified that the elements of the frequently appearing pattern have changed, and the pattern with the lower evaluation value has appeared in the higher rank. Moreover, by analyzing the optimal pattern extracted by the proposed method and the existing method, the trend of the well-known factors for high yield was discovered and factors that prevent high yield from the optimal pattern were eliminated.

Keywords: Optimal pattern, Soybean, Knowledge discovery

1. Introduction

There is a high demand in agricultural efficiency due to the aging society and a lack of successor in Japan. To cope with this problem, knowledge discovery using data mining is being applied for crops that can be managed precisely. For example, there is been studied to increase yielding for tomatoes and strawberries[1][2]. On the other hand, several attempts have appeared to discover efficient knowledge for crops grown under various environmental changes.

Soybeans, which is the target of our work, are known for high-yielding constraints (pestilence, weed, lodging, *etc.*)[3], and not much method has been established to improve the amount of yield and quality. We use the daily cultivation data such as environmental, growth and work data, and focus on the common characteristics observed in several cultivation points in order to analyze and examine the factors for high and low yielding.

Since a small numerical difference of the value of the climate factor does not affect much, we consider grasping with a rough value. Moreover, cultivation environments important for each growing stage are different for soybeans.

Therefore, environmental data are qualitatively valued for each growing stage[4]. At this time, since the interpretation of value for each climate factor differs for each cultivation point, qualitative value conversion is conducted based on the type of condition the cultivation spot is compared with the average year value. By using the qualitative cultivation data, we try to analyze factors affecting growth by exploring features common to many cultivation spots as frequent patterns. However, only redundant information can be obtained from features that appear in both high and low yield cultivation data. Therefore, we consider that an important factor that determines the quality of yield is included in the characteristic of either the high yield or low yield cultivation data, and a function to evaluate the frequency pattern.

In the previous study[5], when expressing the qualitative value, the top 33% was expressed as *High*, the bottom 33% as *Low*, and the rest as *Average*, as compared with the average value. However, directly converting the value near the boundary to the qualitative value may give different qualitative values even if they are almost the same value. When looking for features commonly found at many cultivation points, different qualitative values can be regarded as different features. Since the optimal pattern is extracted according to the frequency of the feature that has appeared, giving different qualitative values causes a problem as such feature to disappear frequently.

† Corresponding Author: namba@cs25.scitec.kobe-u.ac.jp

* Graduate School of System Informatics, Kobe University, Kobe 657-8501, JAPAN

‡ Naro Hokkaido Agricultural Research Center
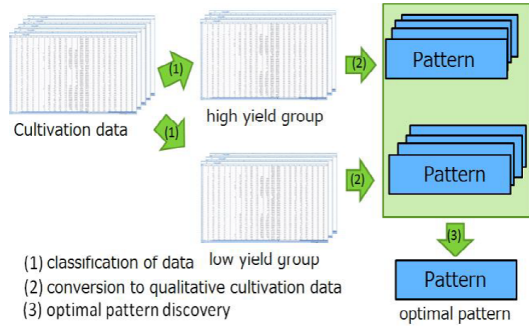 Sapporo, Hokkaido 062-8555, JAPAN

Figure 1: Framework of proposed method

In this paper, we aim to develop a method to discover new knowledge by extracting the optimal pattern from cultivation data. In the proposed method, the ambiguous boundary value is converted into two qualitative value in order to deal with the above problem.

2. Exploring Optimal Pattern with Cultivation Data

2.1 Cultivation Data We use three types of cultivation data, namely environmental, growth and work data, which is obtained for each cultivated spot for experiment. Environmental data includes weather data or other data related to environment for each cultivated spot. Growth data is a data that expresses the condition of crops, including image data shot with a camera from ground or bird's eye view, crop height and length measured by machine or by hand, and a number of flowers. Work data is a data related to motivation of farmers during the high-yielding constraints, or basic actions during cultivation period such as seeding date and amount of harvest.

2.2 Overview of optimal pattern discovery In the previous study, the cultivation data was classified depending on whether the crop yield at the cultivation spot is high or low[5]. Features appearing in either one of high or low yield cultivation data were considered important, and unique features were extracted. The method can be summarized into the following three steps Fig. 1.

1. Division of cultivation data by crop yield

2. Conversion to qualitative cultivation data with ambiguity for each growing stage

3. Optimal pattern discovery using evaluation function based on the difference between high and low yield group

2.3 Qualitative value considering ambiguity In the conventional method, one qualitative value is given for one data item. However, if the value near the qualitative value is simply converted to a qualitative value, there is a possibility that it cannot capture ambiguous information, for example, information existing between *High* and *Average* or between *Average* and *Low*. Therefore, by creating ambiguity in the boundary value when generating the qualitative value conversion, we aim to discover new knowledge that cannot be found before.

First, the number q of qualitative values is determined and expressed as a qualitative value $QV = \{QV_1, \cdots, QV_q\}$. At this time, the number of boundary values is $q - 1$ and is represented as $BV = \{BV_{1,2}, \cdots, BV_{q-1,q}\}$. Then, the value near the boundary value has an intermediate qualitative value. Represent the intermediate qualitative value as a set of $\{QV_{1,2}, \cdots, QV_{q-1,q}\}$. For example, values near the boundary value $b_{1,2}$ are converted into an intermediate qualitative value $QV_{1,2}$ having qualitative values QV_1 and QV_2. There are two types of qualitative valued data generation; majority generation and section generation.

In the case of majority generation, the value of the data item is initially compared with the average year value. Then, the qualitative value is generated, and the qualitative value of the data item is decided by majority when merging in the stage unit. If the intermediate qualitative value is selected as the representative value in the stage unit by the majority vote, the intermediate qualitative value is divided and decomposed into two data items having the original qualitative value. For example, assume that the observation period in one stage is 3 days. When the temperature values are $QV_{1,2}$, $QV_{1,2}$, and QV_1 in order from the first day, majority faction is $QV_{1,2}$, as the value of the temperature of the stage. At this time, since $QV_{1,2}$ is an intermediate qualitative value, it is decomposed into patterns having qualitative values of QV_1 and QV_2.

In the case of section generation, data items are divided into stage units first. Since one stage consists of several days and there is a value for each day, the value of the data item is merged into one value for each stage. As a method to merge the values, attributes such as total value and maximum value are prepared. Likewise, normal year data is divided for each stage, and attributes are prepared. Compare the attribute values summarized for each stage with normal year values, and determine qualitative values. As in the case of majority vote generation, when an intermediate qualitative value is chosen, the intermediate qualitative value is divided and decomposed into two data items with original qualitative values.

Generally, one qualitative-valued data is generated from the value of one data item, but since two qualitative-valued data are generated from the value classified as the intermediate qualitative value, simply calculating the frequency will cause the frequency to be doubly counted. Therefore, correction frequencies α and α' are assigned to each of the two qualitative valued data generated from the values classified as the intermediate qualitative value $QV_{k,k+1}$. The correction frequency is defined as follows so that $\alpha = \alpha' = 0.5$ when the original value is equal to the boundary value, and the value decreases as it becomes farther from the boundary value.

$$\alpha' = \max\left(\min\left(\frac{BV - x}{2D} + 0.5, 1\right), 0\right) \quad (1)$$

$$\alpha' = 1 - \alpha \quad (2)$$

Where α is a correction frequency of the qualitative-valued data corresponding to a portion smaller than the

boundary value BV, and α' indicates a correction frequency corresponding to a portion larger than BV. D is a constant that determines the distance from the boundary value BV the ambiguity is to be applied. x is the value before generating it as a qualitative value. It is possible to always set the sum of the frequencies for the doubled qualitative-valued data to 1, by the correction frequency.

2.4 Optimal Pattern Express cultivation data qualitatively valued using the method in Section 2.3 as a set of $p_i = \langle S, D, A, V \rangle$, that is, patterns $P = \{\cdots, p_i, \cdots\}$. In this quadruplet, S is a growth stage, D is a data item of environmental data, A is an attribute representing a method of qualitative value conversion, and V is a qualitative value. Moreover, in the qualitative value method of Section 2.3, frequency 1 is basically applied, and correction frequency less than 1 is assigned to p_i near the boundary value. The set of qualitative value data in which the pattern P appears in the target group DB_C is $H(P)$, and the set of qualitative value data that appears in the inverse group DB_R is $R(P)$. In this definition, the target group represents either a high-yield cultivation data set or a low-yield cultivation data set, and the inverse group is defined as a data set opposite to the target group.

$$H(P) = \{DB_i | P \subseteq DB_i, DB_i \in DB_C\} \qquad (3)$$

$$R(P) = \{DB_i | P \subseteq DB_i, DB_i \in DB_R\} \qquad (4)$$

$|H(P)|$ and $|R(P)|$ are the frequency of the pattern P in DB_C and DB_R, and $\frac{|H(P)|}{|DB_C|}$ and $\frac{|R(p_i)|}{|DB_R|}$ is the appearance rate of the pattern P. The evaluation function $f(P)$ of the pattern P is defined as follows.

$$f(P) = \sum_{p_i \in P} \left(\frac{|H(P)|}{|DB_C|} - \frac{|R(p_i)|}{|DB_R|} \right) \qquad (5)$$

In other words, the more the appearance rate of the pattern P in the object group is high and the appearance of each element p_i in the inverse group of P is small, the better the pattern becomes. The pattern P with the maximum value of $f(P)$ is defined as the optimal pattern.

2.5 Optimal pattern discovery using frequent closed pattern It is not efficient to enumerate all patterns and calculate the value of the evaluation function to find the optimal pattern. Therefore, we focus on the fact that the pattern with low appearance rate in a certain group is extremely unlikely to be the optimal pattern. Consider reducing the number of patterns to the ones that satisfy $\frac{|H(P)|}{|DB_C|} - \frac{|R(p_i)|}{|DB_R|} > 0$ for all elements $\forall_p \in P$ included in the pattern P, and excluding patterns that do not satisfy the appearance rate $\frac{|H(P)|}{|DB_C|} \geq minSup$ (minimum support). However, if the minimum support is set lower, then the number of patterns become extremely large. Therefore, we focus on a frequent closed pattern that is maximum and that exceeds the minimum support for pattern extraction. By using the closed

pattern, it is possible to omit processing for redundant patterns, and to reduce calculation time and number of outputted patterns. By using the algorithm LCM [6], frequent closed patterns that are candidates for the optimal pattern are enumerated.

In the LCM algorithm, disregarding the correction frequency α, any qualitative-valued data is counted as frequency 1 and frequent closed patterns are extracted by calculating support. On the other hand, it is necessary to count the frequency of qualitative valuation data including ambiguity as correction frequency of 1 or less. Therefore, if the support is recalculated from the correction frequency in the pattern extracted by LCM, there is a possibility that a pattern that does not satisfy the minimum support is included. Even though this frequent closed pattern cannot be an optimal pattern candidate, focusing on the frequent occurrence of a frequent closed pattern that its partial pattern is also frequent, local maximum pattern that satisfies the minimum support is extracted as an optimal pattern candidate.

First, if the support calculated by the correction frequency α of the frequent closed pattern satisfies the minimum support, it can be regarded as the optimal pattern candidate. If not, for each pattern obtained by removing each element constructing the frequent closed pattern, only the pattern that the support calculated by the correction frequency α satisfies the minimum support is enumerated. By recursively performing this process for the pattern that does not satisfy the minimum support, a local maximum pattern that has qualitative valuation data including ambiguity that satisfies the minimum support is enumerated as an optimal pattern candidate.

There is a possibility that the frequent closed pattern may include the element p_i that decreases the evaluation value of $\frac{|H(P)|}{|DB_C|} - \frac{|R(p_i)|}{|DB_R|} \leq 0$. By removing such p_i from the frequent closed pattern, frequent pattern that is a candidate for an optimal pattern with highest evaluation value from each p_i is obtained. Among them, the pattern with the highest evaluation value is the optimal pattern.

The calculation time of the proposed method depends on the calculation time of LCM.

3. Experiment and Analysis

3.1 Dataset For the actual cultivation data, we extract the optimal patterns that become the high yield factors. We compare the method that does not deploy ambiguity into the qualitative value (referred to as the existing method from here on) and the optimal pattern obtained by the proposed method. Then, we examine whether it is possible to acquire new knowledge useful for improving yield from discovered patterns while considering ambiguity.

The target of our experiment is the data of soybeans cultivated in Memuro and Hitsujigaoka at NARO Hokkaido Agricultural Research Center. Environmental data is acquired from Meteorological Observation System in Memuro and Hitsujigaoka[7] respectively. Meteorological Observation System is located at 300 meters from the field in Memuro, and at 750 meters in Hitsujigaoka. Rain sensors are shown in Fig. 2, humidity and wind sensors are
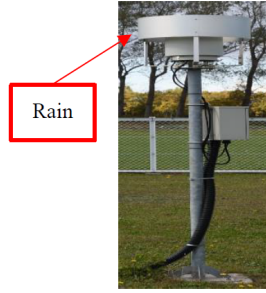
Figure 2: Rain sensors

Figure 3: Humidity and Wind sensors

Figure 4: Radiation sensors

Table 1: Fluctuation of rank for 100-grain weight

Proposed method	Method with no ambiguity in qualitative value			
1th	6th	72th	6th	18th
2nd	10th	40th	8th	286th
3rd	18th	80th	102th	22th
4th	28th	35th	152th	101th
5th	1th	74th	165th	166th
minSup	0.9	0.8	0.7	0.6

Table 2: Fluctuation of rank for seed weight

Proposed method	Method with no ambiguity in qualitative value			
2nd	4th	7th	20th	35th
3rd	6th	12th	18th	6th
4th	2th	4th	12th	32th
5th	11th	21th	2th	24th
minSup	0.9	0.8	0.7	0.6

shown in Fig. 3, and radiation sensors are shown in Fig. 4. Soil sensors are completely buried in the ground.

Growth data and work data are recorded each time the work is performed. There are several cultivation data that were acquired only in one of the two districts. When extracting the optimal pattern, data that appears only in one district has a few amount of information, so we use data acquired in both districts

The data used for the experiments are the 40 kinds of environmental element data recorded at the weather observation site, image data to examine the growth stage, work data with the record of seeding date and harvest date, and harvest data with the record of yield. Data period used in this experiment covers from 2013 to 2015, and from seeding date to harvest date in all cases of years and districts. From the image of the growth data, we manually determine the growth stage. Since the start time the images have been obtained depending on the cultivated place, there is a variation in the period when the growth stage is established. The optimal pattern extraction covers stages that appear commonly in all cultivation places, so the target stages in this experiment are V4 to R5. When generating qualitative values, divide number $q = 3$, $QV = \{High, Average, Low\}$. The correction frequency is computed with $D = 1$. There are five types of attributes: majority M, total value T, maximum value M_x, minimum value M_n, fluctuation range R. Two types of harvesting data used for dividing cultivation data are used: 100-grain weight (g) and seed weight (g/m^2), and the minimum support is $0.55 \leq minSup \leq 0.9$.

3.2 Experiment Results and Discussion

3.2.1 Changes of High-ranked Patterns and Rank Order Fluctuation While changing the minimum support, we compare the patterns of the highest evaluation value including the optimal pattern in the high yield group output by the existing method and the proposed method. By applying the proposed method, there is a possibility that a new frequent pattern appears that has not appeared as the candidate of the optimal pattern output for the existing method. On the other hand, there is a possibility that no frequent pattern appears as the candidate of the optimal pattern output that has appeared in the existing method. Therefore, it is meaningless to simply compare the number of

candidates of the optimal pattern output. Moreover, in order to examine how much the elements of the pattern have changed by qualitative value considering ambiguity,
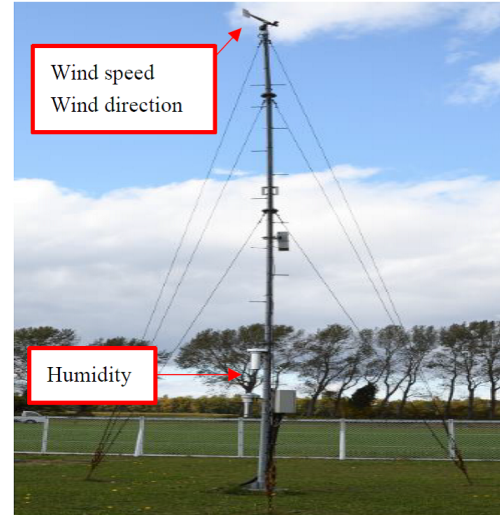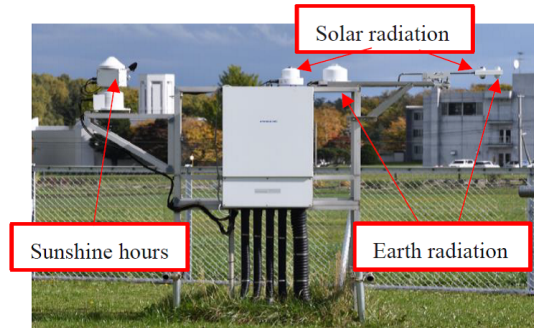
we compare patterns that appear at common observation points in the same group. Among the patterns with the highest evaluation value, Table 1 and Table 2 show ranking fluctuations of the top-five evaluation values of the existing method and the proposed method. Table 1 shows the rank-

Table 3: A subset pattern of 100-grain weight for proposed method

stage	Ground temperature		
	30 cm	10 cm	5 cm
V5	$T = High$	$R = Ave$	$R = Ave$
R4	$M_x = Ave$	$T = Ave$	$T = Ave$
R5	$R = High$	$T = Ave, R = High$	$T = Ave$

stage	Water volume		
	30 cm	10 cm	5 cm
V5	$M_x = Low$	$M_n = Low$	$M_x, M_n = Low$
V6	$T, M = Low$	$T = Low$	$T = Low$
R4			$M_x = Ave$

stage	Sunlight hour
R4	$M_x = High$
R5	$M_n = High$

Table 4: A subset pattern of 100-grain weight for existing method

stage	Long-wave radiation downward	Water volume	
		30 cm	10 cm
V4	$T, M_n = High$	$M_x = Low$	$R = Ave$
V5		$M = Low$	
V6		$M = Low$	
R3	$M_n = Ave$		
R4	$M_x = Ave$		
R5	$M = Ave$		

ing fluctuation of the evaluation value of the pattern in the 100-grain weight, and Table 2 shows the seed weight.

From Table 1, it can be observed that the pattern found ranked top with the proposed method is also ranked high with the existing method disregarding the minimum support. As a result, it is verified that the elements of the pattern with the top evaluation value do not have much ambiguity. On the other hand, for the pattern ranked 5, the order of the pattern of the existing method is higher only when $minSup = 0.9$. The new pattern generated by considering the ambiguity does not satisfy the minimum support because the minimum support is high, so it is considered that there is not much change in the pattern in the existing method and the proposed method. Similar to the pattern at the 100-grain weight in Table 2, the pattern ranked highest is also ranked high even with the existing method disregarding the decrease of the minimum support. As a result, it is verified that the elements of the pattern with the top evaluation value do not have much ambiguity, and there is not much difference between the patterns.

3.2.2   Analysis on comparison of detected optimal patterns   We compare the optimal patterns in the high yield group obtained by the existing method and the proposed method, then extract the unique partial pattern in each pattern, and analyze the result. We analyze the optimal pattern of the 100-grain weight and the rough seed weight obtained with $minSup = 0.65$. The patterns characteristic to the top ten among the extracted optimal patterns is selected.

Optimal pattern in 100-grain weight
Table 3 shows examples of unique partial patterns of the optimal pattern obtained by applying the proposed method, and Table 4 shows examples of partial patterns extracted by applying the existing method in the 100-grain weight. From the unique pattern by the proposed method, the following features are mainly observed.

- The ground temperature tends to be higher than

the standard before flowering and the latter of R period.

- Sunlight hours tend to be high in the latter of R period.

- Moisture content decreases directly before flowering.

From the unique pattern obtained by applying the existing method, the following features are mainly observed.

- Earth radiation is high before flowering and becomes standard after flowering.

- Moisture content tends to be low before flowering.

From these observations, it is newly discovered that by considering the ambiguity, the ground temperature tends to be high after flowering, but the characteristic that the high earth radiation before flowering has not been discovered. In addition, since the moisture content of each unique pattern was low before flowering, the same characteristics were observed regardless of ambiguity.

Optimal pattern in seed weight
For seed weight, examples of partial patterns unique to the optimal pattern obtained by the proposed method are shown in Table 5, and examples of partial patterns extracted by the existing method are shown in Table 6. From the unique pattern by the proposed method, the following features are mainly observed.

- The humidity is entirely high, and tends to be especially high in the latter of the R period.

- Sunlight hours tend to be high in the latter of the R period.

- The amount of solar radiation is high before flowering and becomes standard after flowering.

From the pattern unique obtained by the existing method, the following features are mainly observed.

- Humidity tends to be entirely high.

- Average pressure is low before flowering.

Table 5: A subset pattern of seed weight for proposed method

| stage | Humidity | | |
	mean	maximum	minimum
V4	$T, M_n = Ave$		
V5			$T = Low$
V6		$M_n = High$	
R2		$M_x = High$	
R3		$T = High$	
R4	$T = High$	$M, T, M_n = High$	
R5		$M_x, M_n = High$	

stage	Sunlight hours	Amount of solar radiation
V4		$T = High$
R3	$M_n = High$	
R4	$M_x = High$	

Table 6: A subset pattern of seed weight for existing method

| stage | Humidity | | Atmosphere |
	maximum	Minimum	
V5	$T, R = High$		
V6			$M, T = Low$
R1	$M = High$	$M = Low$	
R2	$R = High$		
R4		$M_x = High$	

From these observations, it is verified that solar radiation tends to be high before flowering by taking ambiguity into consideration, while low average pressure before flowering has not been observed. In addition, since the characteristics of high humidity are observed in common to each unique pattern, it can be analyzed that this characteristic has been obtained regardless of ambiguity.

Analysis for the whole optimal pattern

The paper by S. Nakamura et al. (1996) indicated that the high temperature in the flowering period (V4 to R3) and a low rainfall inhibited the high yield of soybeans. On the other hand, it is also clarified that desirable environment at maturity (R4 to R5) is large day temperature difference, long sunshine hours and low rainfall amount[8]. Referring to these findings as a conditions for high yield, we analyze and discuss our result.

As a characteristic of high-yield group, in both 100-grain weight and seed weight, the sunshine hours are *High* at maturity. This feature is found only in the proposed method. Moreover, the humidity is high throughout the entire growth period (V4 to R5). The amount of humidity might have correlation with the amount of precipitation, however, the desirable environment at maturity is low amount of rainfall. In the future, we plan to investigate humidity and precipitation individually.

Moreover, in the existing method, characteristics of atmospheric pressure is *High*, and global radiation is

Low, that are considered as the factor to raise the temperature, during flowering period were obtained. In the proposed method, it was shown that these patterns generally regarded as high yield inhibiting factors are removed from the optimal pattern.

Furthermore, the feature that the ground temperature is high at maturity was discovered only in the unique pattern in the optimal pattern obtained by the proposed method. Therefore, it can be considered that appropriate knowledge that was not discovered from the optimal pattern of the existing method can be extracted by considering ambiguity.

4. Conclusions

In this paper, focusing on the ambiguity in the vicinity of the boundary value when qualitative value is converted, we propose a method to generate a pattern with qualitative values and find the optimal pattern. As a result of applying the proposed method to the actually acquired time series cultivation data, it was confirmed that a pattern with a low evaluation value among frequent patterns obtained by the existing method is a top candidate of the optimal pattern. In addition, as a result of examining the unique pattern of the optimal pattern obtained by the existing method and the proposed method, while discovering new patterns that could be the knowledge of general high yield factors, it was possible to remove elements that could be the knowledge of high yield inhibition factors from optimal patterns.

The problems in this study are as follows. First, it is necessary to confirm this result in the actual cultivation environment. However, for that purpose, it is necessary to improve the experimental environment such as lowering the cost of setting sensors and cameras for information acquisition. Then, when dividing the stage, we manually judge the stage from the image of the crop, so the criteria for dividing the stage becomes ambiguous. Therefore, it is necessary to consider a method of incorporating information that is not visible by the stage when the stage is divided. Moreover, in this study, we analyze the environmental information by examining the stage independently. However, since the state of the crop is greatly affected by the former environment, environmental information itself is insufficient. It is necessary to consider methods incorporating growth information. To deal with this problem, it is possible to measure NDVI (Normalized Difference Vegetation Index) as growth information. Therefore, there is a possibility to develop a method that applies the value of NDVI and its amount of change as growth information.

References

[1] P. Makelaa, K. Jokinend, M. Kontturic, P. Peltonen-Sainioa, E. Pehua, and S. Somersaloa, "Foliar application of glycine-betaine –a novel product from sugar beet– as an approach to increase tomato yield", Industrial Crops and Products, Vol. 7, Issues 2-3, pp.139-148, 1998.

[2] B. Z. Yuan, J. Sun, and S. Nishiyama, "Effect of Drip Irrigation on Strawberry Growth and Yield inside a Plastic Greenhouse", Biosystems Engineering, Vol. 87, Issues 2, pp. 237-245, 2004.

[3] Japan Agricultural Development and Extension Association (JADEA) 2012, "Soybean Making for Improvement of Yield and Quality and Stable Production Q & A", http://www.jadea.org/houkokusho/daizu/documents/daizu-kaitei.pdf, pp.2-5, Accessed 15 June 2017 (in Japanese).

[4] W. R. Fehr and C. E. Caviness, "Stages of Soybean Development", Cooperative Extension Service, Iowa State University, Special Report 80, 1977.

[5] K. Umejima, F. Arimitsu, S. Ozawa, N. Murakami, H. Tsuji, and T. Ohkawa, "Optimal Pattern Mining form Time-Series Cultivation Data of Soybeans for Knowledge Discovery", Proceedings of the Workshop on Time Series Analytics and Applications, pp.19-24, 2016.

[6] T. Uno, T. Asao, Y. Uchida, and H. Arimura, "An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases", Lecture Notes in Artificial Intelligence, Vol. 3245, pp.16-31, 2004.

[7] R. Sameshima, T. Hirota, T. Hamasaki, K. Kato, and Y. Iwata, "Meteorological Observation System at the National Agricultural Research Center for Hokkaido Region since 1966", Meteorological Observation System at the National Agricultural Research Center for Hokkaido Region since 1966, pp.1-8, 2008.

[8] S. Nakamura, S. Yumoto, K. Takahashi, and Y. Takada, "Climatic Factors During Growth Phases on Soybean Seed Yield", Tohoku Agricultural Research, Vol. 49, pp.81-82, 1996 (in Japanese).

Midori Namba (Member) received her B.E. degree in Department of Computer Science and Systems Engineering, Kobe University, Japan, in 2016. She is currently a Student in the Department of Information Science, Graduate School of System Informatics, Kobe University, Japan. Her research interest is in smart agriculture.

Kohei Umejima (Non-member) received his B.E, M.E., degrees from Kwansei Gakuin University in 2015 and Kobe University in 2017, respectively. He is currently a System Engineer in financial industry.

Ryo Nishide (Member) received his B.E., M.E. and Ph.D. (Informatics) degrees from the Faculty of Informatics, Kansai University in 2002, 2004 and 2008, respectively. He is currently an Assistant Professor in the Graduate School of System Informatics, Kobe University, Japan. His research interests include mobile computing, networking, pedestrian flow analysis and data mining. He is a member of IEEE, ACM, IPSJ and DBSJ.

Takenao Ohkawa (Non-member) received his B.E, M.E., and Ph.D. degrees from Osaka University in 1986, 1988, and 1992, respectively. He is currently a Professor in the Department of Information Science, Graduate School of System Informatics, Kobe University, Japan. His research interests include intelligent data processing, bioinformatics, and smart agriculture. He is a member of the IEEE, the Information Processing Society of Japan, the Institute of Electronics, Information, and Communication Engineers, the Institute of Electrical Engineers in Japan, and the Japanese Society for Artificial Intelligence.

Seiichi Ozawa (Non-member) received the B.E. and M.E. degrees in instrumentation engineering from Kobe University in 1987 and 1989, respectively. In 1998, he received his Dr. Eng. in computer science from Kobe University. He is currently a full professor with the Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kobe University, Japan. His current research interests are neural networks, machine learning, online learning, pattern recognition, big data analytics especially in cybersecurity, SNS and smart agriculture. He is a member of Neural Networks TC, Data Mining and Big Data Analytics TC, and Smart World TC of IEEE CI Society.

Noriyuki Murakami (Non-member) enter Ministry of Agriculture, Forestry and Fisheries in 1991. After working as National Agricultural Research Center, moved to Hokkaido Agricultural Experiment Station in 1998 (reorganized as NARO Hokkaido Agricultural Research Center in 2001). He was a visiting researcher at the Robotics Institute of Carnegie Mellon University from 2000 to 2001. He is currently a director at division of farming system research , Hokkaido Agricultural Research Center, NARO, Japan. His research interests include on robots and information and communication technologies in the field of agriculture, engaged in development of proper diagnostic software for agricultural chemicals and research on agricultural chemical mixing apparatus *etc.*

Hiroyuki Tsuji (Non-member) received his B. Agr. degrees from Utsunomiya University in 1988, and Ph.D. in Life and Environmental Sciences from Osaka pref. University in 2015, respectively. He is currently a leader of Crop Rotation System Group at the Hokkaido Agricultural Research Center, NARO, Japan. His primary research interest is in developing low-labor cost system, high-yield crop rotation systems, which are wheat, sugar beet, potatoes, beans and vegetable crops. He is a member of the Crop Science Society of Japan, the Japanese Society of Farm Work Research, the Japanese Society for Root Research, and the Weed Science Society of Japan.