Breeding Science 47 : 253～257 (1997)

Fuzzy Logic Evaluation of Soybean Plant Shape

Jack Ambuel[1], Seishi Ninomiya[2] and Nobuo Takahashi[3]

[1]*Department of Agricultural and Biosystems Engineering, Iowa State University,* Ames, Iowa 50011, USA
[2]*National Institute of Agro-Environmental Sciences,* Tsukuba, Ibaraki 305, Japan
Present Address: *National Agriculture Research Center,* Tsukuba, Ibaraki 305, Japan
[3]*Nagano Chushin Agriculture Experimental Station,* Shiojiri, Nagano 391-64, Japan

Summary

Soybean plant shape evaluation is an important part of the soybean plant breeding process in Japan. This selection process is currently performed by visual inspection by the soybean plant breeder. This paper describes a method to evaluate soybean plant shape quality automatically. The method developed was an expert system using fuzzy logic rule sets to evaluate soybean shape quality. The evaluator operated on 4 shape indicators extracted from digitized images of each soybean plant. The evaluator placed the shape of each soybean plant into one of three categories: good (3), fair (2), and poor (1). Only those rated as good were selected by the soybean breeders. The goal was to develop an evaluator that would give the same ratings as those given by the soybean plant breeders. The shape quality evaluation results based on the fuzzy logic rule sets developed in this study were slightly better than those obtained using statistical discriminant analysis. The efficiency of the correct evaluation was about 76 % for both the plants to be selected (good) and the plants to be removed (fair or poor). Fuzzy logic evaluation has two advantages in contrast to statistical discriminant analysis. One is that it does not require any assumption on statistical distribution of the shape features and the other is that its structure is easy to understand.

Key Words : decision making, fuzzy logic, *Glycine max* L. Merr., plant shape evaluation.

Introduction

In soybean breeding, one important factor used in plant selection is plant shape. Plant shape is believed to be related to lodging resistance, light interception, yield and suitability for machine harvesting (Ninomiya and Shigemori 1991). Currently, soybean plant shape quality is evaluated visually by expert soybean breeders. Recently, attempts have been made to automate the process using image analysis and statistical discriminant analysis (Ninomiya and Shigemori 1991). The purpose of that approach was to develop quantitative measures of soybean plant shape based on the heuristic rules developed by breeders, and to provide an automatic method of evaluating soybean plant shape to assist the breeders in the selection process.

There were several reasons why it was decided to investigate the use of a fuzzy logic expert system for evaluation of soybean plant shape. One reason was that fuzzy logic evaluation is based on a heuristic instead of mathematical approach and is therefore inherently easier to understand. The rules are described linguistically in common language. This makes working with the evaluator much easier for researchers when tuning the system and for plant breeders when applying the system. In contrast, the techniques based on discriminant analysis and neural networks do not have the direct connection with the breeder's rules for evaluating plant shape. The second reason for using fuzzy logic was that because it is an expert system, there are no constraints on the type of processes it is used to describe or simulate. This is in contrast to discriminant analysis where there are a number of constraints. For example, when discriminant analysis is used on the 18 shape variables to characterize soybean plant shape, the shape variables are assumed to have a normal distribution. Because the fuzzy expert system is not subject to constraints, it has the potential to provide a better match with the breeders' evaluations.

This paper describes the development of the expert system and compares the results of the program when used to evaluate the shape of 875 soybean plants with the evaluations of a team of soybean plant breeders and with the results of the evaluation based on discriminant analysis.

Materials and Methods

Fuzzy Logic Expert Systems

Fuzzy logic, invented by Zadeh (1965), has in the last 10 to 15 years been applied in a large number of consumer goods and in industrial and transportation control (Williams 1991, Kosko 1993). Originally envisioned as a technique for analyzing complex biological systems that are not subject to mathematical analysis, fuzzy logic has found its greatest use to date in process and device control. However, recently more attempts have been made to apply the techniques of fuzzy logic to problems of the life sciences. The soybean plant shape evaluator described in this paper was a modified version of a fuzzy logic expert system yield model (Ambuel *et al.* 1994). The fuzzy logic soybean plant shape evaluator

Received February 24, 1997. Accepted June 9, 1997.

was developed using a spreadsheet (Lotus 123 Ver3.0, Lotus Develop. Corp., Cambridge, USA). Contained in the spreadsheet were the fuzzy logic rule sets, the soybean plant shape variable membership functions, the soybean plant shape variable raw data, and the programs (macros) used to link the raw data with the rule sets and extract the results. The evaluator used simple overlapping triangle functions for the membership functions. The evaluator used two variables per rule set and organized the rule sets hierarchically, with the output of the top level rule set being the plant shape quality. The evaluator was tested on the images from the 875 soybean plants of the above mentioned study (Ninomiya and Shigemori 1991).

Selection of Image Variables

The variables used in the fuzzy logic plant shape evaluator were selected from the variables employed in the study of plant shape classification based on discriminant analysis (Ninomiya and Shigemori 1991). In that study, 18 plant shape indicators were extracted using image analysis techniques from the digitized images of each of the 875 soybean plants. The plant shape quality was then obtained by performing discriminant analysis on the 18 variables for each plant. For the initial implementation of the fuzzy logic program, it was decided to use a reduced set of variables in order to simplify the development process. Therefore, only the variables that were most important in determining shape quality (as defined by expert breeders) were used. The most important characteristics for good soybean plant shape quality are that the plant be symmetrical in shape (not skewed), that the stem be straight, that there be a moderate amount of leaves (not too many and not too few) and that the plant be relatively slim (height greater than width). From these considerations the following four shape indicators were selected for development of the fuzzy rule sets:

1. The degree of occupancy shape indicator (a good indicator of plant leaf area) denoted as D in Ninomiya and Shigemori (1991).
2. The normalized width of the plant (an indication of the plant slimness) denoted as WDT* in Ninomiya and Shigemori (1991).
3. The skew of the plant with respect to the horizontal (the third moment of the horizontal distribution and an indication of horizontal symmetry) denoted as XSK in Ninomiya and Shigemori (1991).
4. The lag in the horizontal direction (a measure of the degree of bending of the plant) denoted as XD1 in Ninomiya and Shigemori (1991).

Rule Descriptions

Having selected the variables for use in the shape evaluator, the next step was to develop a general statement of the rules for what constitutes a good plant shape. As seen above, this step was actually a part of the variable selection procedure. The main requirements for good plant shape on which the variable selection was based were:

1. A plant is symmetrical.
2. The stem is straight not bent.
3. A medium amount of leaves. Not too many and not too few.
4. Height is larger than its width.

It is these four heuristic rules combined with the ranges of the selected variables that were used to develop the fuzzy rules described below.

Determination of Ranges

The third step in the evaluator development was the assignment of fuzzy ranges to each of the selected variables. In that evaluator, as described above, simple overlapping triangle functions (Fig. 1) were used to represent membership functions for each variable. Rule sets were limited to two variables each, and rule sets were combined hierarchically to generate the final output (plant shape quality). The number of fuzzy ranges for each variable was set equal to seven as VS=Very Small, S=Small, MS=Medium Small, M=Medium, ML=Medium Large, L=Large and VL=Very Large.

The procedure used to assign specific values to each of the fuzzy range centerpoints (peaks of the range triangle membership functions) was identical for each variable. First, the range (max and min) of each variable and its mean were determined. The mean value of the variable was then assigned to the Medium range centerpoint. Values close to the maximum and minimum values of the variable were then assigned to the Very Large and Very Small range centerpoints respectively. The values of the remaining range centerpoints were set so that the separation between ranges was a constant.

The results of the range assignments for the four variables are summarized in Table 1. In addition, the maximum, minimum and mean values for the variables are listed.

Development of fuzzy rule sets

The final step in the evaluator development was the transformation of the general rules for good soybean plant shape into corresponding fuzzy rules. Since four variables were used in the initial implementation, the number of rule sets required was three. Two on the first level to process the four variables and one on the second level to combine the outputs of the two first

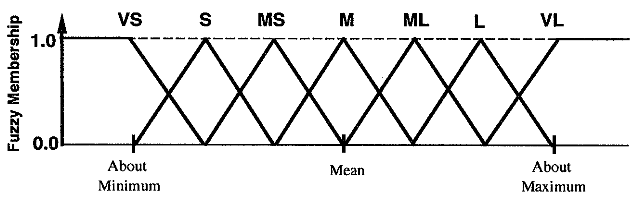

Fig. 1. Overlapping triangle fuzzy membership function. VS=Very Small, S=Small, MS=Medium Small, Medium, ML=Medium Large, L=Large and VL=Very Large.

level rule sets and generate the final output (plant shape rating). The development of the initial rule set was a somewhat uncertain process. The strategy was to use the general rules obtained from expert soybean plant breeders as a guide. The plan was to make an initial estimate of the fuzzy rule sets and then refine the estimate after analyzing the results. From the breeders' rules on soybean shape the following guidelines for use in the fuzzy rule sets were obtained:

1. The degree of occupancy (D) should be Medium.
2. The slimness (WDT^*) should be Medium Small or lower.
3. The skewness (XSK) should be Small or Very Small.
4. The lag (XD1) should be low, Small or Very Small.

Table 1. Shape indicator variables and the fuzzy ranges for the shape evaluator

VALUE	D	WDT^*	XSK	XD1
Maximum	0.54362	1091	0.826	322.0
Minimum	0.14684	332	0.00033	0.690
Mean	0.3452	712	0.41313	161
VS	0.20	400	0.070	10
S	0.25	500	0.185	60
MS	0.30	600	0.300	110
M	0.35	700	0.415	160
ML	0.40	800	0.530	210
L	0.45	900	0.645	260
VL	0.50	1000	0.700	310

D = Degree of Occupancy. As D increases, leaf area increases.
WDT^* = Normalized Width. As WDT^* increases, slimness increases.
XSK = Horizontal Skew. As XSK increases, horizontal symmetry decreases.
XD1 = Horizontal Lag. As XD1 increases, stem bending increases.
VS = Very Small. S = Small. MS = Medium Small. M = Medium.
MS = Medium Large. L = Large. VL = Very Large.

Plant Shape Quality Ruleset

	LI-VS	LI-S	LI-MS	LI-M	LI-ML	LI-L	LI-VL
WS-VS	100	100	100	100	100	100	100
WS-S	100	200	200	200	200	200	300
WS-MS	100	200	300	300	300	300	300
WS-M	100	200	300	400	400	500	500
WS-ML	100	200	300	400	500	500	600
WS-L	100	200	300	500	500	600	700
WS-VL	100	300	300	500	600	700	700

Light Interception Ruleset

	D-VS	D-S	D-MS	D-M	D-ML	D-L	D-VL
WDT-VS	100	200	500	700	700	700	500
WDT-S	100	200	500	700	700	700	500
WDT-MS	100	200	500	600	600	600	400
WDT-M	100	200	400	500	500	500	300
WDT-ML	100	100	200	400	400	400	200
WDT-L	100	100	200	300	300	300	100
WDT-VL	100	100	100	100	100	100	100

Width Symmetry Ruleset

	XSK-VS	XSK-S	XSK-MS	XSK-M	XSK-ML	XSK-L	XSK-VL
XD1-VS	700	700	550	300	200	100	100
XD1-S	700	700	500	300	200	100	100
XD1-MS	600	550	450	200	200	100	100
XD1-M	500	450	400	200	100	100	100
XD1-ML	200	200	200	200	100	100	100
XD1-L	100	100	100	100	100	100	100
XD1-VL	100	100	100	100	100	100	100

Fig. 2. Fuzzy rule sets for the soybean shape evaluator. See text for the detail.

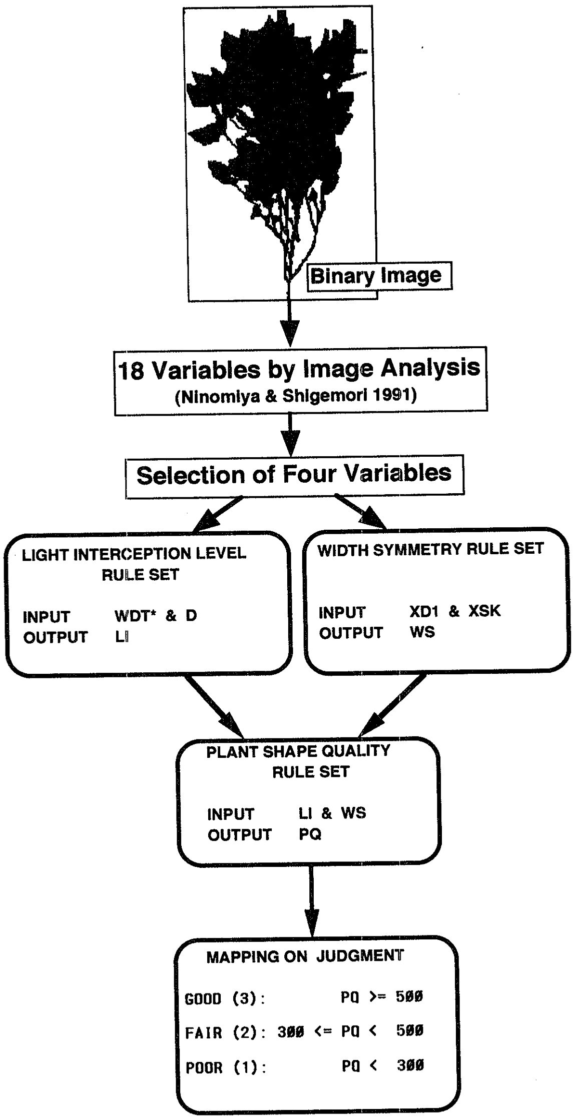

Fig. 3. Structure of the fuzzy rule sets for the soybean shape evaluator.

Results and Discussion

Using the above guidelines, the rule sets shown in Fig. 2 and Fig. 3 were developed. The approach taken to modifying the rule sets was to examine the variables for those cases where there were mismatches and to attempt to determine what changes in the rule sets would result in a match. This was a brute force approach to adjust each element of the rule sets shown in Fig. 1 manually, and after four iterations, a distinct improvement was obtained. The first rule set combined the leaf area (D) and the slimness (WDT^*) variables. The output of this rule set was named light interception (LI), because the slimness and leaf area affect the amount of light intercepted. The second rule set combined the horizontal skew (XSK) and the horizontal lag

(XD1) variables. The output of this rule set was called the width symmetry (WS).

The basic concept of the fuzzy logic procedure taken in this study was the same as that of Ambuel *et al.* (1994). For example, if WDT* = 575, the value intersects two fuzzy sets; the MS set with a membership value of 0.25 and the M set with a value of 0.75. Similarly, if D = 0.375, the value intersects the M and ML sets with membership values of 0.5 and 0.5, respectively. A total of four light interception rules (Fig. 2) are partially satisfied because each variable has nonzero membership in two sets. The four rules are: (1) if WDT* is MS and D is M then LI is 600; (2) if WDT* is M and D is M then LI is 500; (3) if WDT* is MS and D is ML then LI is 600; (4) if WDT* is M and D is ML then LI is 500. Then, the final output is determined by some combination of outputs for each individual rule. The method to determine the final output is a two-step process. First, the degree of fulfillment equal to the minimum value of the degrees of membership of all input values to the rule is established for each rule. For example, the degree of fulfillment of 0.25 is given to the rule (1). When the degree of fulfillment is established for each rule, the outputs of all partially satisfied rules are combined in a weighted average using the degree of fulfillment as the weight. For the above four rules, the final LI output is 533 $\left(=\frac{600\times0.25+500\times0.5+600\times0.25+500\times0.5}{0.25+0.5+0.25+0.5}\right)$.

The third rule set on the second level uses the outputs of the first two rule sets as its input variables. The output of this rule set for the initial shape evaluation program was the plant shape quality (PQ). Since seven levels were used for both inputs to and outputs from the fuzzy rule sets, the output of the plant shape quality rule set had to be transformed into the three levels assigned by the breeders: good, fair and poor (the assignment was done to each binary image of the soybean plants). This was done with the mapping shown in Fig. 2.

The transformed output was then compared with the breeder evaluations and with the evaluation performed using discriminant analysis with cross-validation (a SAS procedure, DISCRIM, SAS Institute Inc. 1985) based on just the same four variables used in the first level fuzzy rule sets. The results are summarized in Table 2. The discriminant analysis evaluation was better than the fuzzy logic shape evaluation in terms of the number of matches for the poor shape plants. In contrast, the fuzzy logic evaluation was slightly better than the discriminant analysis for both the fair and good shape plants. In the practical breeding process, only the good shape plants are selected and others (fair or poor) are removed. Therefore, Table 2 was retabulated as Table 3 so that the efficiency of the models became clearer in the practical sense. Table 3 shows that the total efficiency of the fuzzy logic evaluator is slightly better than that of the discriminant analysis.

The next step taken was to modify the fuzzy logic evaluator and add another variable identified as important to shape by the breeders. That variable was what is referred to as heavy head (YM in Ninomiya and Shigemori, 1991), where too much of the plant material is in the upper half of the plant. When this was done, there was little change in the results (for either fuzzy or discriminant analysis). Therefore, attention was returned to the original evaluator.

The results of the application of the fuzzy logic shape evaluation in this study were encouraging in contrast with the discriminant analysis used in the previous study (Ninomiya and Shigemori, 1991), because of the advantages of the fuzzy logic mentioned above. However, further improvement is necessary before the evaluator can be used as a tool to assist plant breeders practically.

In order to obtain further improvement in the fuzzy model results, a detailed investigation of plants where the evaluator and breeder ratings were mismatched was performed consulting the plant breeder who had given the original judgments. This investigation was performed on the first 60 plants from the data set. The breeders were asked why a particular rating was given to each plant in the first 60 plants where the breeder rating did not match the evaluator rating. In the breeder consultation, 7 reasons were identified for the 18 mis-

Table 2. Results of the fuzzy logic shape evaluation using the rule sets shown in Table 1 and Fig. 2 (above) and the discriminant analysis with cross-validation (below)

			Class Observed by Breeders		
			Poor	Fair	Good
Class Predicted by Model	Fuzzy Logic	Poor	161 (0.58)	89 (0.22)	10 (0.05)
		Fair	95 (0.34)	178 (0.44)	35 (0.18)
		Good	24 (0.09)	137 (0.34)	146 (0.76)
	Discriminant Analysis	Poor	197 (0.70)	114 (0.28)	12 (0.06)
		Fair	59 (0.21)	159 (0.39)	42 (0.22)
		Good	24 (0.09)	131 (0.32)	137 (0.72)
Total Number of Observation in Each Class			280 (1.00)	404 (1.00)	191 (1.00)

Each pair of the values indicates the number of the plant classified and its ratio.

Table 3. Results of the fuzzy logic shape evaluation and the discriminant analysis. The case that the poor and the fair shape plants were combined into one class is shown.

			Class Observed by Breeders	
			Poor of Fair	Good
Class Predicted by Model	Fuzzy Logic	Poor or Fair	523 (0.76)	45 (0.24)
		Good	161 (0.24)	146 (0.76)
	Discriminant Analysis	Poor or Fair	529 (0.77)	54 (0.28)
		Good	155 (0.23)	137 (0.72)
Total Number of Observation in Each Class			684 (1.00)	191 (1.00)

matches in the first 60 plants of the data set.

From these 7 categories of reasons for mismatch, a strategy for improvement of the rule sets can be developed. For two of the seven categories of mismatch, it may be possible to eliminate the mismatch by adding variables to the rule set. These two categories are: too much vertical skew; and plant leaf distribution is not tight. Vertical skew is a variable already given by Ninomiya and Shigemori (1991). Tightness of leaf distribution is not directly related to any one of the 18 shape indicators and so may be a little more difficult to accommodate. For three of the seven categories of mismatch, it may be possible to eliminate the mismatch by modifying the rules. These three categories are: the stem is bent but all other characteristics are good; the plant is too slim; and the plant has too few leaves. For the three plants in the unknown category, the action required cannot be determined. The final category is mismatch caused by a data problem. The stem is bent but the corresponding variable (XD1) does not reflect this. This is generally caused by stems which are bent which curve back upon their main axis. Because a linear approximation is done to estimate the location of the stem axis, the net bending can be too low if the stem curves back upon itself. It may be necessary to solve this problem by adding another variable to measure stem curvature.

The next step in the development of the fuzzy logic soybean plant shape evaluator will be to attempt to eliminate the problems identified by the analysis described above. The goal is to develop a program that can be used by soybean plant breeders to evaluate plant shape. However, addition of more variables will result in increased difficulty in selecting and adjusting the fuzzy rules. It has been found in a number of other cases (Ambuel, *et al.* 1994) that as the complexity of the system increases, the number of variables increases and the development of the rule sets becomes more complicated and time consuming. Some automatic method of rule set adjustment or some alternative method, such as neural networks, in which automatic adjustment is inherent, is required.

Literature Cited

Ambuel, J., T. S Colvin and D.L. Karlen (1994) A fuzzy logic yield simulator for prescription farming. Trans. ASAE.37: 1999-2009.

Kosko, B. and S. Isaka (1993) Fuzzy logic. Sci. Amer. July: 76-81.

Ninomiya, S. and I. Shigemori (1991) Quantitative evaluation of soybean (*Glycine max* L. Merrill) plant shape by image analysis. Jpn. J. Breed. 41: 485-497.

SAS Institute Inc. (1985) "SAS User's Guide: Statistics Version 5 Edtion", Cary, SAS Institute Inc.

Williams, T. (1991) Fuzzy logic simplifies complex control problems. Comput. Design. March: 90-102.

Zadeh, L.A. (1965) Fuzzy Sets. Info. Control. 8: 338-353.