

## PERSPECTIVE OF DIMENSION REDUCTION IN THE STUDY OF GENOTYPE- ENVIRONMENT INTERACTION USING ROUGH SET THEORY

CHETAN, RATNA RAJ LAXMI & NISHA

Department of Statistics, M.D. University, Rohtak, Haryana, India

### ABSTRACT

In agricultural experimentation, a large number of genotypes are generally calibrated over a wide range of environments. The genotypic values under different environment may increase or decrease as per there conditions. In plant breeding programs the effects of different genotypes are tested under the multi-environment such type of combined investigation is termed as the Genotype x Environment (G x E) Interaction. In the present study, the concept of Rough Set Theory (RST) is adopted to reduce the number of environments without loss of generality of original experimental dataset.

**KEYWORDS:** Discernibility Function, Discernibility Matrix, Genotypic Variation, Genotype x Environment Interaction, Dimension Reduction etc

### INTRODUCTION

Many situations in agricultural experimentation, where the large number of genotypes are examined over the wide range of environment. There genotypic values may vary from one environment to another which might cause genotypes to even rank differently between environments. Genotype x Environment (GE) interaction is a common phenomenon in plant breeding experiments as it results in inconsistent performance between the genotype across environment such as locations, years, growing seasons etc., as well as the genetic constitution of genotype, influence the phenotypic expression of a traits.

In the last fifty years the researches put their eyes on investigating the performance of different genotypes under various environments. Finlay and Wilkinson (1963), Eberhart and Russell (1966), Freeman and Perkins (1971) and Shukla (1972) contributed the remarkable work in the study of GE interaction. They use the regression concept of stability and suggested the use of two stability parameters. The new stability measure is proposed by Laxmi (1992) by giving momentousness to environmental condition. A stability measures are proposed by Raju (2002) which are equivalent to biplot with first PCA axes. In addition, Saini et al. (2015) presented the combined study and calculated the stability measures of different genotypes. Laxmi et al. (2015) observed that an integrated likelihood approach for studying genotypic variation obtained by using a conditional reference prior is equal to the marginal likelihood approach for parameter obtained by the non-centrality parameter. It was observed that not so much of the efforts have been given to study the reduction of dimensionality of the dataset use for obtaining the GE interaction. In the present study, an attempt is made to reduce the number of environments with the loss of generality of experimental data by using rough set theory (RST).

RST has proved its effectiveness in drawing conclusion from experimental data [Skowron (2001)]. Pawlak (1982) introduce the concept of RST to deal with uncertain, incomplete or vague information. RST is the diffuseness of set theory and has an inherent feature of abbreviate the dataset. This abbreviation is only because of classes which are based on

indiscernibility relations and it is use to eliminate the meaningless attributes. The stuffy approach is to obtain the discernibility matrix which seeks to determine its corresponding discernibility function or in other words, it is way to get the reduce dataset. For any information system, say S, with n elements and  $n \times n$  discernibility matrix, the entries of the matrix are given by

$$c_{ij} = \{a \in A | a(x_i) \neq a(x_j)\} \text{ for } i = 1, 2, \dots, n; i \neq j$$

Where each entry contains the subset of attributes that distinguishes element  $x_i$  and  $x_j$ , being the diagonal entries null, according to the definition. Then the corresponding discernibility function  $f_A$  is a Boolean function of m attributes  $(a_1, a_2, \dots, a_n)$  and given by

$$f_A(a_1^*, a_2^*, \dots, a_n^*) = \bigwedge \{ \bigvee c_{ij} | 1 \leq j \leq i \leq n, c_{ij} \neq \emptyset \}$$

This classical approach is used to reduce the dimensionality of data set contain the information regarding GE interaction.

**METHODOLOGY**

Consider the data set of pod yield of 15 varieties  $(G_1, G_2, \dots, G_{15})$  of ground nut crop raised at 20 locations  $(L_1, L_2, \dots, L_{15})$ . The experimental design used is RCBD at each location with three replications [Sainiet al. (2015)]. If the experimenter want to reduce the dimensionality of the dataset using RST, the attribute condition should be want to reduce the number of locations without loss of generality, attribute Yes(Y) is assign if the yield in the particular location is greater than the average yield of location other No(N). This will gives us the intermediate dataset (Table 1) with 20 conditional attributes.

**Table 1**

	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$	$L_6$	$L_7$	$L_8$	$L_9$	$L_{10}$	$L_{11}$	$L_{12}$	$L_{13}$	$L_{14}$	$L_{15}$	$L_{16}$	$L_{17}$	$L_{18}$	$L_{19}$	$L_{20}$
$G_1$	Y	N	Y	Y	N	N	N	N	N	N	Y	Y	N	Y	Y	Y	Y	N	N	N
$G_2$	Y	N	Y	Y	N	N	N	N	Y	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_3$	N	N	Y	N	N	N	N	N	N	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_4$	Y	N	Y	Y	Y	Y	N	N	N	N	Y	N	N	N	Y	Y	N	N	N	N
$G_5$	N	Y	Y	Y	N	Y	N	N	N	N	Y	Y	N	N	Y	Y	N	N	N	N
$G_6$	Y	N	Y	Y	N	N	N	N	N	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_7$	Y	N	Y	Y	N	N	N	N	Y	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_8$	N	N	N	Y	N	N	N	N	Y	N	Y	Y	N	Y	Y	Y	Y	N	N	N
$G_9$	Y	N	Y	Y	N	N	N	N	Y	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_{10}$	Y	N	Y	Y	N	Y	N	N	Y	N	Y	Y	N	N	Y	Y	Y	N	N	N
$G_{11}$	Y	N	Y	Y	Y	N	N	N	N	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_{12}$	Y	N	Y	Y	Y	N	N	Y	N	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_{13}$	Y	N	Y	Y	Y	N	N	N	N	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_{14}$	Y	N	Y	Y	N	N	N	N	N	N	Y	Y	N	Y	N	Y	Y	N	N	N
$G_{15}$	Y	N	Y	Y	N	N	Y	N	Y	N	N	Y	N	Y	N	N	Y	N	N	N

The discernibility matrix for the concern situation is given in the annexure. The Boolean simplification of  $f_A$  yields the set of reducts of A is given by

$$f(x) = [(L_9 \vee L_{15}) \wedge (L_1 \vee L_{15}) \wedge \dots \wedge (L_7 \vee L_9 \vee L_{15})] \wedge \dots \wedge [(L_5) \wedge (L_5 \vee L_9 \vee L_{15})] \wedge [(L_9)]$$

The above logical function is expressed by the conjunction of many terms with each one corresponding to column of the related discernibility matrix. The Boolean simplifications of this function yields then a single reduct

$$f(x) = L_1 \wedge L_2 \wedge \dots \wedge L_9 \wedge L_{11} \wedge L_{12} \wedge L_{14} \wedge \dots \wedge L_{17}$$

Thus by theory, this reduct is composed of location  $L_1, L_2, \dots, L_9, L_{11}, L_{12}, L_{14}, \dots, L_{17}$  and the original dataset with 20 locations which can be reduced to 15 locations without the loss of generality. Comparison between genotype variation for original dataset and reduced dataset is done by the concept of rank correlation coefficient

$$\eta_s = 1 - \frac{6 \sum_{i=1}^n d_i}{n(n^2-1)} = 0.8002$$

where  $d_i$  is the difference between two ranks of investigated stability measure.

**Table 2: Genotype Variation for Different Genotypes**

Genotype	Genotype Variation (Original Dataset)	Genotype Variation (New Dataset)
G1	1.18206	1.02127
G2	1.06242	0.95133
G3	1.31187	1.08635
G4	1.06961	1.09240
G5	1.12011	1.10417
G6	1.12261	1.09998
G7	1.00935	0.99108
G8	0.75720	0.90941
G9	1.11915	1.06111
G10	1.00262	1.08932
G11	0.88775	0.90778
G12	1.10923	1.06620
G13	0.82707	0.82325
G14	0.96580	0.91035
G15	0.81812	0.88601

## DISCUSSIONS AND CONCLUSIONS

The recognition of the importance of dealing with the effects of genotype x environment(G x E) interaction in multi-environment testing of genotype is given in plant breeding programs. There has been substantial development in the area of analytical methodology to quantify and describe this interaction. The one of the major are is to deal with dimensionality of the multi-environment data. In many practical problems, the data under consideration is high-dimensional but experimenters may have reason to believe that the data lay near a lower-dimensional manifold without loss of geniality. RST plays an important role in this direction. In the present study, RST is used to reduce the number of environments. Table 2 shows the genotype variation for original as well as reduced dataset. It is found that the rank correlation between genotype variations of different genotypes, for two types of data set is 0.8002 which is on the higher side.

## REFERENCES

1. Eberhart, S.A and Russel, W.A. (1966). Stability parameters for comparing varieties. Crop Sci., 6, 36-40.

2. Finlay, K.W. and Wilkinson, G.N. (1963). The analysis of adaptation in a plant breeding programme, Australian Journal of Agricultural Research, 14,142-754.
3. Freeman, G.H. and Perkins, J.M. (1971). Environmental and genotype-environmental components of variability VII, Relation between genotype grown in different environments and measurement of these environments, Heredity, 27, 15-23.
4. Laxmi, Ratna Raj, Sandhya and Cheta (2015). Concept of Conditional reference prior in the study of genotypic variation, International Journal of Agricultural Science and Research, 5(6), 283-288.
5. Laxmi, Ratna Raj. Genotype –Environment Interaction: Its Role in Stability of crop varieties. Unpublished Thesis, submitted to IASRI, New Delhi, India, (1992).
6. Pawlak,Z, (1982). Rough Set, International Journal of Information and Computer Science, 11(5), 341-356.
7. Raju, B.M.K. (2002). A study of AMMI model and its biplots, Journal of Indian Society of Agricultural Statistics, 55(3), 297-322.
8. Saini, Parul, Chetan and Sandhya (2015). A review on genotype environment interaction and its stability measures, International Journal of Science and Research, 4(1), 1210-1213.
9. Shukla, G.K. (1972). Some Statistical aspects of partitioning genotype-environmental components of variability, Heredity, 29,237-245.
10. Skowron, A. (2001). Rough set in KKD, Proceeding of the Conference on Intelligent Information Processing (IIP 2000), Publishing House of Electronic Industry, Beijing, 1-17.

**ANNEXURE**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	–														
2	$L_9, L_{15}$	–													
3	$L_1, L_{15}$	$L_1$	–												
4	$L_5, L_2, L_6, L_{12}, L_{14}, L_{17}$	$L_5, L_6, L_{12}, L_{14}, L_{15}, L_{17}$	$L_1, L_5, L_6, L_{12}, L_{14}, L_{15}, L_{17}$	–											
5	$L_1, L_2, L_6, L_{14}, L_6$	$L_1, L_2, L_6, L_{14}, L_{15}, L_{17}$	$L_2, L_6, L_{14}, L_{15}, L_{17}$	$L_1, L_5, L_{12}$	–										
6	$L_{15}$	$L_9$	$L_1$	$L_5, L_6, L_{12}, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_6, L_{14}, L_{15}, L_{17}$	–									
7	$L_9, L_{15}$	∅	$L_1, L_9$	$L_5, L_6, L_{12}, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_6, L_9, L_{14}, L_{15}, L_{17}$	$L_9$	–								
8	$L_1, L_9$	$L_1, L_{15}$	$L_9, L_{15}$	$L_1, L_5, L_6, L_{12}, L_{14}, L_{15}, L_{17}$	$L_2, L_6, L_9, L_{14}, L_{17}$	$L_1, L_9, L_{15}$	$L_1$	–							
9	$L_9, L_{15}$	∅	$L_1, L_9$	$L_5, L_6, L_9, L_{12}, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_6, L_9, L_{14}, L_{15}, L_{17}$	$L_9$	$L_1, L_{15}$	$L_1, L_{15}$	–						
10	$L_6, L_9, L_{14}$	$L_6, L_{15}, L_{14}$	$L_1, L_9, L_{14}, L_{15}$	$L_5, L_6, L_{12}, L_{17}$	$L_1, L_2, L_9, L_{17}$	$L_6, L_9, L_{14}, L_{15}$	$L_6, L_{14}, L_{15}$	$L_1, L_{14}$	$L_6, L_{14}, L_{15}$	–					
11	$L_5, L_{15}$	$L_5, L_9$	$L_1, L_{15}$	$L_1, L_2, L_5, L_6, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_5, L_6, L_{14}, L_{15}, L_{17}$	$L_5$	$L_5, L_9$	$L_1, L_5, L_9, L_{15}$	$L_5, L_9$	$L_1, L_5, L_9, L_{15}$	–				
12	$L_5, L_{15}$	$L_5, L_9$	$L_1, L_{15}$	$L_1, L_2, L_5, L_6, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_5, L_6, L_{14}, L_{15}, L_{17}$	$L_5$	$L_5, L_9$	$L_1, L_5, L_9, L_{15}$	$L_5, L_9$	$L_1, L_5, L_9, L_{15}$	∅	–			
13	$L_5, L_{15}$	$L_5, L_9$	$L_1, L_{15}$	$L_1, L_2, L_5, L_6, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_5, L_6, L_{14}, L_{15}, L_{17}$	$L_5$	$L_5, L_9$	$L_1, L_5, L_9, L_{15}$	$L_5, L_9$	$L_1, L_5, L_9, L_{15}$	∅	∅	–		
14	$L_{15}$	$L_9$	$L_1$	$L_1, L_2, L_6, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_6, L_{14}, L_{15}, L_{17}$	∅	$L_9$	$L_1, L_9, L_{15}$	$L_9$	$L_1, L_9, L_{15}$	$L_5$	$L_5$	$L_5$	–	
15	$L_7, L_9, L_{15}$	$L_7$	$L_1$	$L_1, L_2, L_6, L_7, L_9, L_{14}, L_{15}, L_{17}$	$L_1, L_2, L_6, L_7, L_9, L_{14}, L_{15}, L_{17}$	$L_7, L_9$	$L_7$	$L_1, L_7, L_{15}$	$L_7$	$L_1, L_7, L_{15}$	$L_5, L_7, L_9$	$L_5, L_7, L_9$	$L_5, L_7, L_9$	$L_9$	–