Network Analysis of Cropping Practices and Injuries Profiles under Rice Agroecosystem

Sith

August 27, 2015

Abstract

Here is the abstract...........

Introduction

The use of rice crop health survey at the field levels is ground truth database that allow one identify actual constraints due to pests in rice production. The database provide the overview of the complex relationships between crop stands, their management, pest injuries, may lead to better management, and guide researchers the new research hypotheses (4; 9).

Several previous studies (11; 10; 8; 2; 5) conducted the surveys have identified the relationships of individual production situation (a set of factors that determine agricultural production) and injury profiles (combination of disease and pest injuries that may occur in a given farmers field) using nonparametric multivariate analysis (7). Performing correspondence analysis, they characterized the relationships between categorized levels of variables: actual yield, production situations and injuries profiles . the results led to the conclusions that observed injuries profiles are strongly associated with production situations(4). However, the step of transform continuous data to categorized data by clustering approach is poor reproducibility and difficult interpretability of the individual cluster (3; 1).

The components of production situation and injury profiles, such as the amount of fertilizers applied, occurrence of leaf blast are biologically related. The relationships will be more complex when the number of their components will increase. A way to systemically model and intuitively interpret such relationships is the depiction as a graph or network. This approach has been widely used and proven very useful in biological studies. Networks typically consist of nodes, usually presenting components (), while links between the node depict their interactions. Correlation network is a type of network, which two nodes are connected if their respective correlation lies above a certain threshold. The construction of this network obtain from pairwise correlation methods. By using appropriate correlation measure, correlation network can capture biologically meaningful relationships, and discover the valuable information in crop health surveys.

Selecting the suitable association methods for crop health construction is important because the method that can capture the relationships with true concordance often determined the type and amount of knowledge we can gain from survey data. we have limited prior knowledge (positive relation and negative relation) for comparing the efficiency of different association methods in discovering true functionally associated variables.

The main aim of this article s to In this study, we evaluate correlation methods including Pearson, Spearman, Kandell, Biweight to associate the components of cropping practices and the components of injuries. Furthermore, we applied network theory and model to illustrate the paiewise relasionship. Thus we hope to provide the necessay elements for a bter comprenhjesion of the methods and also the chouce of a suirtanle dependence terst method based on pracitical constrains and goals.

We inferred a interaction network by from survey project comprising 5 countries (India, Indonesia, the Philippines, Thailand, and Vietnam), 420 lowland farmers' fileds. Our study aimed to determine co-occurrence pattern among the incidence of injuries caused by animal pests and diseases and the cropping practices, potentially indicative of their occurrence relations. We thus construct the network from the surveys. The limitation of each measure are difference assumption and detach different patterns. The structure of surveys are determine for choosing the suitable measure.

Materials and Methods

Survey datasets

Crop health survey data were collected through surveys comprising 420 farmers' fields from 2010 to 2012 for wet and dry seasons in different production environments across South and South East Asia representing irrigated lowland rice growing areas of India, Indonesia, the Philippines, Thailand, and Vietnam. The survey protocol described in the IRRI publication, "A survey portfolio to chatacterize yield-reducing factors in rice", (6) was used for data collection. The variables collected included patterns of cropping practices, crop growth measurement and crop management status assessments, measurements of levels of injuries caused by pests, and direct measurements of actual yields from crop cuts. The data collected can be classified into three groups: cropping practices, injuries, and actual yield measurements.

Evaluation of association methods

Step one: Data exploratory Data were check the properties. There are three main properties to be determine before deciding the appropriate correlation measure.

Task check normality and homoscedasticity and data distribution Two assumptions, similar to those for ANOVA, are that for any value of X, the Y values will be normally distributed and they will be homoscedastic. Although you will rarely have enough data to test these assumptions, they are often violated.

Numerous simulation studies have shown that regression and correlation are quite robust to deviations from normality; this means that even if one or both of the variables are non-normal, the P value will be less than 0.05 about 5% of the time if the null hypothesis is true.

So in general, you can use linear regression/correlation without worrying about non-normality. Sometimes you'll see a regression or correlation that looks like it may be significant due to one or two points being extreme on both the x and y axes. In this case, you may want to use Spearman's rank correlation, which reduces the influence of extreme values, or you may want to find a data transformation that makes the data look more normal. Another approach would be analyze the data without the extreme values, and report the results with or without them outlying points; your life will be easier if the results are similar with or without them. When there is a significant regression or correlation, X values with higher mean Y values will often have higher standard deviations of Y as well. This happens because the standard deviation is often a constant proportion of the mean.

Task check the independence Linear regression and correlation assume that the data points are independent of each other, meaning that the value of one data point does not depend on the value of any other data point. The most common violation of this assumption in regression and correlation is in time series data, where some Y variable has been measured at different times.

Task check linearity or non–linearity Linear regression and correlation assume that the data fit a straight line. If you look at the data and the relationship looks curved, you can try different data transformations of the X, the Y, or both, and see which makes the relationship straight. Of course, it's best if you choose a data transformation before you analyze your data. You can choose a data transformation beforehand based on previous data you've collected, or based on the data transformation that others in your field use for your kind of data.

Step two: identify the most appropriate method

Although we can opt for a method based on its principle of statistical operation without paying attention the biological models in a given set, this may not lead to a coordination network that will reveal biological knowledge.

Network Construction

Co-occurrence network construction

The matrix can be viewed as an adjacency matrix of a weighted network. The matrix contains the correlation coefficient between each node (i.e., the variable). Thus the matrix can be thought of as the population average of the network structure. Because we are looking at several specific links, we control for multiple testing by controlling the False Discovery Rate (FDR method) at 5%. The generated network structure can be visualized through the R package qgraph. Only connections that surpass the significance threshold are shown in the visual representation.

Network analysis

Important information about a network can be gained by analyzing its global structure, for example by looking at the relative centrality of different nodes. In a centrality analysis, nodes are ordered in terms of the degree to which they occupy a central place in the network. Global descriptors of the modules were obtained using package qgraph in R. The neighborhood of a given node n is the set of its neighbors. The connectivity is the size of its neighborhood. The average number of neighbors indicates the average connectivity of a node in the network. A normalized version of this parameter is the network density. Density ranges between 0 and 1. It shows how densely the network is populated with edges, A network which contains no edges and solely isolated nodes has a density of 0. In contrast, the density of a clique is 1. Another related parameter is the network centralization. Networks whose topologies resemble a star have a centralization close to 1, whereas decentralized networks are characterized by having a centralization close to 0.

In undirected networks, the clustering coefficient is the number of connected pairs between all neighbors of the network. The clustering coefficient of a node is always a number between 0 and 1. The network clustering coefficient is the average of the clustering coefficients for all nodes in the network. Nodes with less than two neighbors are assumed to have a clustering coefficient of 0. We then determined network centralities on the modules obtained from network analysis. Centralities were assessed using package in R. We calculated Degree centrality and Betweenness centrality.

Pearsons product-moment correlation coefficient

The Pearsons product-moment correlation or simply Pearsons correlation is a measure of linear dependence, as the slope obtained by the linear regression of Y by X is Pearsons correlation multiplied by that ratio of standard deviations. Let $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ be the means of X and Y, respectiverly, then the Peasons's corrlation coefficient $\rho_{pearson}$ is defined as follows:

$$\rho_{pearson}(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2 \sum_{i=1}^{n}(y_i-\overline{y})^2}}$$

For joint normal distributions, Pearsons correl- ation coefficient under H0 follows a Students t-distribution with $n-2$ degrees of freedom. The t statistic is as follows:

$$t = \frac{\rho_{pearson}(X,Y)\sqrt{n-2}}{\sqrt{1-\rho^2_{pearson}(X,Y)}}$$

When the random variables are not jointly nor- mally distributed, the Fishers transformation is used to get an asymptotic normal distribution.

In the case of perfect linear dependence, we have $\rho_{pearson} = \pm 1$. The Pearson correlation is $+1$ in the case of a perfect positive (increasing) linear relationship and -1in the case of a perfect negative (decreasing) linear relationship. In the case of linearly independent random variables, $\rho_{pearson} = 0$, and in the case of imperfect linear dependence, $-1 < \rho_{pearson} < 1$. These last two cases are the ones for which misinterpretations of correlation are possible because it is usually assumed that non- correlated X and Y means independent variables, whereas in fact, they may be associated in a non- linear fashion that Pearsons correlation coefficient is not able to identify. The R function for Pearsons test is cor.test with parameter method 'pearson' (package stats). The stats package can be downloaded from the R Web page (http://www.r-project.org).

Results

Discussion

References

[1] J Avelino, H Zelaya, A Merlo, A Pineda, M Ordoñez, and S Savary. The intensity of a coffee rust epidemic is dependent on production situations. *Ecological Modelling*, 197(3-4):431–447, 2006.

[2] Kun Dong, Bin Chen, Zhengyue Li, Yan Dong, and Hailong Wang. A characterization of rice pests and quantification of yield losses in the japonica rice zone of yunnan, china. *Crop Protection*, 29(6):603–611, 2010.

[3] Daxin Jiang, Chun Tang, and Aidong Zhang. Cluster analysis for gene expression data: A survey. *Knowledge and Data Engineering, IEEE Transactions on*, 16(11):1370–1386, 2004.

[4] Twng Wah Mew, Hei Leung, Serge Savary, Casiana M Vera Cruz, and Jan E Leach. Looking ahead in rice disease research and management. *Critical Reviews in Plant Sciences*, 23(2):103–127, 2004.

[5] C S Reddy, G S Laha, M S Prasad, and D Krishnaveni. Characterizing multiple linkages between individual diseases, crop health syndromes, germplasm deployment, and rice production situations in India. *Field Crops*, 2011.

[6] S Savary and N.P Castilla. A survey portfolio to characterize yield-reducing factors in rice. *IRRI Discussion Paper No 18*, 2009.

[7] S Savary, FA Elazegui, HO Pinnschmidt, NP Castilla, and PS Teng. A new approach to quantify crop losses due to rice pests in varying production situations. *IRRI discusion paper series no20. International Rice Research Institute, PO Box*, 933, 1997.

[8] Serge Savary, Nancy P Castilla, FA Elazegui, and Paul S Teng. Multiple effects of two drivers of agricultural change, labour shortage and water scarcity, on rice pest profiles in tropical asia. *Field Crops Research*, 91(2):263–271, 2005.

[9] Serge Savary, Paul S Teng, Laetitia Willocquet, and Forrest W Nutter. Quantification and modeling of crop losses: a review of purposes. *Annual Review of Phytopathology*, 44(1):89–112, 2006.

[10] Serge Savary, Laetitia Willocquet, Francisco A Elazegui, Nancy P Castilla, and Paul S Teng. Rice pest constraints in tropical Asia: quantification of yield losses due to rice pests in a range of production situations. *Plant disease*, 84(3):357–369, 2000.

[11] Serge Savary, Laetitia Willocquet, Francisco A Elazegui, Paul S Teng, Pham Van Du, Defeng Zhu, Qiyi Tang, Shiwen Huang, Xianquing Lin, and H M Singh. Rice pest constraints in tropical Asia: characterization of injury profiles in relation to production situations. *Plant Disease*, 84(3):341–356, 2000.