Model Selection for Poisson Regression via Association Rules Analysis

Pannapa Changpetch[1], & Dennis K. J. Lin[2]

[1] Department of Mathematical Sciences, Bentley University, Waltham, MA, USA

[2] Department of Statistics, Pennsylvania State University, University Park, PA, USA

Correspondence: Pannapa Changpetch, Department of Mathematical Sciences, Bentley University, Waltham, MA 02452, USA. Tel: 1-781-891-2267. E-mail: pchangpetch@bentley.edu

Abstract

This study integrates association rules analysis, a methodology for selecting potential interactions, with Poisson regression modeling. Though typically ignored in conventional Poisson regression, interactions are very common in practice. However, selecting a Poisson regression model when many main effects and interactions are involved is problematic. In this study, we develop a model selection framework to address this problem. Specifically, we focus on building an optimal Poisson regression model by (1) discretizing the response and quantitative attributes into levels; (2) exploring via association rules analysis combinations of input variables that have a significant impact on response; (3) selecting potential (low- and high-order) interactions; (4) converting these potential interactions into new variables; and (5) selecting variables from all the input variables and the newly created variables (interactions) to build the optimal Poisson regression model. Our model selection procedure is the first approach to enable a global search for potential interactions and the first to establish the optimal combination of main effects and interaction effects in the Poisson regression model. A real-life example is given for illustration. It is shown that the proposed method finds the optimal model including important interactions that cannot be found by other existing methods.

Keywords: association rules analysis, interaction effect, model selection, poisson regression model

1. Introduction

The Poisson regression model is one of the most important models for count data. It has been used to explore relationships among variables in a wide variety of areas, including economics, health care, demography, business, and manufacturing. However, selecting a Poisson regression model when many main effects and interactions are involved is complicated. Interactions are very common in practice. Yet, in the conventional Poisson regression, they are typically ignored.

In this paper, we propose a systematic procedure designed to consider the interactions among variables and thereby produce models that are superior to those developed using established methods. We consider both low-order and high-order interactions among variables. Note that we also consider interactions among categorical variables, which are usually neglected in the literature. However, it is both impractical and ineffective to include all the possible interactions in a model selection process. Instead, we apply (a data-mining technique) association rule analysis to find potential interactions for the model selection procedure.

Association rules analysis is a methodology that aids in the selection of potential interactions among categorical variables from a large pool of possibilities. Using this method, we are able to narrow the field of possible combinations by eliminating interactions that are unlikely to contribute to the fit of the Poisson model.

Our model selection framework overpowers the ability of classical model building to consider potential interactions. Association rules analysis streamlines the process of selecting important rules, converts the selected rules into interaction variables, and determines the optimal model for Poisson regression by implementing a subset selection method that considers all the main effects and the potential interactions. The key advantages of the proposed framework include its ability to deal with a large number of interactions, its ability to select potential interactions, and its provision of alternative setups for interactions. In the proposed method, interactions are incorporated into the Poisson regression model. Higher-order interactions can also be included in the

model—a feature that differentiates our model selection approach from other Poisson regression modeling approaches.

This paper is organized as follows. Section 2 reviews the relevant academic literature pertaining to Poisson regression modeling and association rules analysis. Section 3 presents the framework and discusses the proposed method in detail. Section 4 presents an application of our framework to a real dataset. Section 5 offers a discussion and concluding remarks.

2. Poisson Regression and Association Rules Analysis

The Poisson regression model is perhaps one of the most important models for count data. It has been widely used in many academic areas and applications for count data (Cameron & Trivedi, 1998; Winkelmann, 2008). For example, it is used to analyze accident rates (Kahn, 2005; Hahn & Prieger, 2006) and crime rates (Kelly, 2000; Plassmann & Tideman, 2001). Other application areas include health economics (Schellhorn, 2001; Doorslaer, Koolman & Jones, 2004; Winkelmann, 2004; Winkelmann, 2006), demography (Winkelmann & Zimmermann, 2000), and marketing (Greene, 2000).

A Poisson regression model uses a log link to explain the relationship between the response, i.e., count data, and its explanatory variables (Agresti, 2002). The Poisson regression model with only main effects is written as:

$$\log \mu = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$$

where μ is the expected value of response, the $X's$ are the main effects, and the $\beta's$ are the coefficients of those main effects.

However, selecting a Poisson regression model when many main effects and interactions are involved is complicated. Interactions, especially those between categorical variables, can be addressed in different ways. In the literature on the categorical regression model, interactions among categorical variables are generally neglected. Given that the number of interactions increases at an accelerated rate as the main effects increase in number, the subset selection method is extremely inefficient when all the interactions are considered at the same time. Therefore, an efficient method for selecting potential interaction variables is required. Here, we develop a methodology for implementing association rules analysis for exactly this purpose. As will be seen, the proposed method is able to find the optimal model when the interaction effect is present.

Association rules analysis is a popular data-mining technique that was introduced in the early 1990s (Agrawal & Srikant, 1994; Liu, Hsu & Ma, 1998). A large number of studies use association rules analysis in a wide variety of areas: biology (Becquet, Blachon, Jeudy, Boulicaut & Gandrillon, 2002; Marghny & El-Semman, 2005), business and marketing (Changchien & Lu, 2001), geography (Appice, Ceci, Lanza, Lisi & Malerba, 2003; Ding, Ding & Perrizo, 2002; Lee, Hong, Ko, Tsao & Lin, 2007), agriculture (Matsumoto, 1998), education (Garcia, Romero, Ventura & Calders, 2007; Garcia, Amandi, Schiaffino & Campo, 2007), photography (Liu, Weng, Tseng, Chuang & Chen, 2008), economics (Dopfer & Potts, 2004; Kuo, Lin & Shih, 2007), and so on.

Association rules analysis is a methodology for exploring relationships among items in the form of rules. Each rule has two parts: the first part pertains to left-hand side item(s) or the condition(s), and the second part to the right-hand side item or the result. The rule is always represented as a statement: If *condition*, then *result* (Berry & Linoff, 1997). Two measurements are attached to each rule. The first measurement, support (*s*), is computed by $s = \text{Prob}(condition \text{ and } result)$. The second measurement, confidence (*c*), is computed by $c = \text{Prob}(result \mid condition)$. Association rules analysis finds all the rules that satisfy both of these key thresholds: minimum support and minimum confidence (Agrawal & Srikant, 1994).

This set of rules can be used for other purposes, including classification. A technique called classification rule mining (CRM), a subset of association rules analysis, was developed to form an accurate classifier by finding a set of rules in a database (Liu, Hsu & Ma, 1998; Quinlan, 1992). This technique uses an item to represent a pair consisting of a main effect and its corresponding integer value. More specific than association rules analysis, CRM has only one target, which must be specified in advance. In general, the target of CRM is the response, which means the result of the rule (the right-hand item) can only be the response and its class. Therefore, the left-hand item (the condition) consists of the explanatory variable and its level. For example, assume that there are k categorical factors, $X_1, X_2, …, X_k$, and a categorical response Y. Many rules can be generated by CRM. One such rule could be "If $X_1 = 1$, then Y = 1" with $s = P(X_1 = 1 \text{ and } Y = 1)$ and $c = P(X_1 = 1 \text{ and } Y = 1)/P(X_1 = 1)$. Another rule could be "If $X_1 = 1$, then Y = 0" with $s = P(X_1 = 1 \text{ and } Y = 0)$ and $c = P(X_1 = 1 \text{ and } Y = 0)/P(X_1 = 1)$.

In the present study, we use CRM to screen out insignificant or irrelevant interactions and consider only those that are potentially significant for the Poisson regression model. This methodology is a major aspect of the selection of variables in our process. To our knowledge, there are no studies linking association rules analysis

and classification rules mining to Poisson regression modeling. The methods proposed by Changpetch and Lin (2013a) and Changpetch and Lin (2013b) are limited to binary response variable and multinomial response cases. Poisson regression model is very important and different from the logistic regression model and the multinomial logit model. For the Poisson regression model, there can be myriads of different numbers for the response. Therefore, it is necessary to discretize the response before applying association rules analysis. This is an additional work that is particularly important for the Poisson regression model. Moreover, we also discretize the quantitative attributes in the same step to improve our approach's ability to handle the interactions between the categorical and quantitative variables as well. Therefore, we include the extra step of discretization from the four-step approach proposed by Changpetch and Lin (2013a, 2013b).

Several studies use Poisson regression in conjunction with other data-mining techniques in support of achieving other objectives. For example, Chaudhuri, Lo, Loh, and Yang (1995) used decision tree analysis to find subnodes for a decision tree as a basis for estimating the Poisson regression model for each subnode. Here, we focus on searching for potential interactions for the Poisson regression model. The decision tree is the other data-mining technique that can be used to find interactions. However, the decision tree has the disadvantage of a hierarchical structure. Therefore, as association rules analysis enables a global search, more potential interactions can be located and thus considered than is possible with the decision tree structure.

3. The Proposed Method

The proposed framework for building a Poisson regression model for count data consists of five key steps. As shown in Figure 1, the five steps in our framework are as follows:

Step 1: Discretize the response and the quantitative variables into categorical variables.

Step 2: Generate the rules from association rules analysis.

Step 3: Select the rules based on confidence.

Step 4: Generate the variables for each rule from step 2.

Step 5: From the variables in step 3 and all the main effects, search for the optimal model.
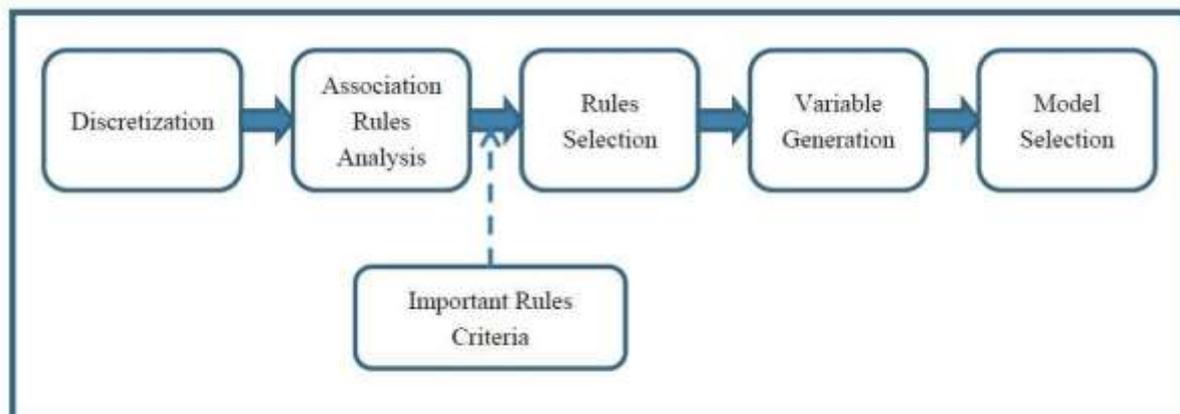
Figure 1. Framework for building the proposed model.

Step 1: Discretization

Given that association rules analysis works with categorical variables, we first discretize the response and the quantitative attributes into categories. For recommended level for discretization, we recommend between 3-5 levels since we need at least three levels for response and quantitative variables, i.e. low level, medium level, and high level. Based on our empirical studies, if we divided into just two levels, we will lose potential significant interactions. If we divide into too many levels, a large number of rules with small support will be resulted. We will then face a screening problem, some difficulties finding the rules (among so many potential rules) that significantly contribute to the model. The practitioner can divide the data into levels based on distribution or available criteria. When there is no existing reference, the simplest way to discretize the response is to divide all the observations equally for each level. For example, if we discretize the response into 4 levels, each level will contain 25% of the observations.

For the quantitative variable, after dividing it into levels, we will convert it into dummy variables, which is similar to the way in which we convert categorical variables into dummy variables.

Step 2: Association Rules Analysis

First, we use association rules analysis to create rules from datasets. Specifically, we perform CRM. For each rule, the condition (left-hand items) represents the combination of explanatory variables and their levels, whereas the result (the right-hand item) is the response and its class. To perform CRM, we use the CBA program (Liu, Hsu & Ma, 1998) developed by the Department of Information Systems and Computer Sciences at the National University of Singapore (website: http://www.comp.nus.edu.sg/~dm2/). With this program, we are able to obtain all the active rules using the given minimum support and minimum confidence. For Poisson regression, we do not restrict the level of minimum support but recommend a minimum confidence level of 80%. The expected results from this step are the rules in the form "If X_i's = x_i's, then $Y = y$," where x_i is the level of variable X_i and where y is the level of response Y. With each rule, the respective support and confidence are attached. All active rules become inputs for the second step.

Step 3: Rule Selection

In this step, we select the rules to convert into interaction variables for the next step. The rules selection criterion used here is confidence. Therefore, rules with the highest level of confidence are selected from the active rules obtained in the first step. We call the rules selected at this stage *potential rules*. Note that the number of rules selected at this stage is relatively small compared to the total number of possible interactions for the dataset. In this work, we set the number of potential rules at between 30 and 50 (the same number recommended in Changpetch and Lin (2013a) and Changpetch and Lin (2013b)). The higher the number of variables, the higher the number of potential rules we select. All the potential rules are inputs for the third step.

Step 4: Variable Generation

In this step, we generate the variables for the Poisson regression model from the potential rules. To convert a rule into an interaction, we separate the rule into two cases. In the first case, the variables on the left-hand side are originally categorical variables. We generate interactions among the main effects on the left-hand side with the same settings that appear in the rule. Assume that the selected rule has three predictors with the form "If $X_i = x_i$, $X_j = x_j$, and $X_k = x_k$, then $Y = y$," where x_i is the level of variable X_i, x_j is the level of variable X_j, x_k is the level of variable X_k, and y is the level of response Y. We generate an interaction among X_i, X_j, and X_k by labeling this interaction as 1 if $X_i = x_i$, $X_j = x_j$, and $X_k = x_k$, and as 0 otherwise. This interaction is denoted by $X_i(x_i)X_j(x_j)X_k(x_k)$. For example, for the rule If $X_1 = 0$, $X_2 = 1$, and $X_3 = 1$, then $Y = 0$, we create an interaction among X_1, X_2, and X_3 denoted by $X_1(0)X_2(1)X_3(1)$. We have $X_1(0)X_2(1)X_3(1) = 1$ if $X_1 = 0$, $X_2 = 1$, and $X_3 = 1$, and 0 otherwise. Note that the level of Y does not play any role in generating the variables.

The second case is when variable(s) on the left-hand side include any variable that has been converted from the quantitative variable(s). For example, the selected rule is "If $X_i = x_i$, $X_j = x_j$, $X_k = x_k$, and $X_q = x_q$, then $Y = y$," where X_q is the variable converted from the quantitative variable and x_q is the level of variable X_q. We generate an interaction among X_i, X_j, X_k, and X_q by multiplying the original variable of X_q with the generated interaction $X_i(x_i)X_j(x_j)X_k(x_k)$, as in the previous example. Assume that the original variable of X_q is Q and that the generated interaction is referred to as $X_i(x_i)X_j(x_j)X_k(x_k)Q$. The results from this step are inputs for the fifth step.

Step 5: Model Selection

In principle, any model selection criterion can be used. Here, the Bayesian information criterion (BIC) is used for illustration (Schwarz, 1978). The best subset selection method is performed by testing all the possible combinations of variables and selecting the one that gives the optimal BIC. In other words, the model selected is the one that gives the lowest BIC among all the models. We use the original response (before discretization) for the model fit. The variables in this step consist of all the interactions generated in Step 4 plus all the main effects. Note that there are many choices in this model selection step. Other advanced model selection methodologies, e.g., the least absolute shrinkage and selection operator (lasso) (Tibshirani, 1996), the smoothly clipped absolute deviation (SCAD) (Fan, 1997), the adaptive lasso (Zou, 2006), the least-angle regression algorithm (LARS) (Efron, Hastie, Johnstone & Tibshirani, 2004), the Dantzig selector (Candes & Tao, 2007), or another kind of selection criteria, e.g., the Akaike information criterion (AIC) (Akaike, 1974), the Corrected Akaike information criterion (AICc) (Hurvich & Tsai, 1993), the Corrected Bayesian information criterion (BICc) (Tremblay & Wallach, 2004), the significance of each individual variable criterion; can be applied here. Note that each association rule can be converted into only one interaction. The final selected model will be the optimal model with a combination of main effects and potential interactions.

4. Application: Patent Applications Dataset

In this section, we use a patent applications dataset (Cincera, 1997) to demonstrate how to apply our proposed

method. We select the model for this dataset using our framework. Note that we use a portion of the dataset used in Cincera (1997), which contains records of 181 firms with variables, as shown in Table 1. The objective is to construct the model in order to explain and predict the number of patent applications in 1991 in reference to the available variables including industry sector, firm location, availability of patent applications in 1990, and R&D expenditure.

Table 1. Attributes for the Patent Applications Dataset

Attribute	Type of attribute	Levels	Categorical variables
Patent applications (1991)	Categorical	4 levels	$Y = 1$ if patent applications in 1991 is 0
			$Y = 2$ if patent applications in 1991 is 1–10
			$Y = 3$ if patent applications in 1991 is 11–100
			$Y = 4$ if patent applications in 1991 is > 100
Sector	Categorical	15 levels	$X_1 = 1$ if aerospace, and $X_1 = 0$ otherwise
			$X_2 = 1$ if chemistry, and $X_2 = 0$ otherwise
			$X_3 = 1$ if computers, and $X_3 = 0$ otherwise
			$X_4 = 1$ if drugs, and $X_4 = 0$ otherwise
			$X_5 = 1$ if electricity, and $X_5 = 0$ otherwise
			$X_6 = 1$ if food, and $X_6 = 0$ otherwise
			$X_7 = 1$ if fuel and mining, and $X_7 = 0$ otherwise
			$X_8 = 1$ if glass, and $X_8 = 0$ otherwise
			$X_9 = 1$ if instruments, and $X_9 = 0$ otherwise
			$X_{10} = 1$ if machinery, and $X_{10} = 0$ otherwise
			$X_{11} = 1$ if metals, and $X_{11} = 0$ otherwise
			$X_{12} = 1$ if other, and $X_{12} = 0$ otherwise
			$X_{13} = 1$ if paper, and $X_{13} = 0$ otherwise
			$X_{14} = 1$ if software, and $X_{14} = 0$ otherwise
			$X_{15} = 1$ if vehicles, and $X_{15} = 0$ otherwise
Area	Categorical	4 levels	$X_{16} = 1$ if the EU, and $X_{16} = 0$ otherwise
			$X_{17} = 1$ if Japan, and $X_{17} = 0$ otherwise
			$X_{18} = 1$ if the USA, and $X_{18} = 0$ otherwise
			$X_{19} = 1$ if the rest of the world, and $X_{19} = 0$ otherwise
Patent applications (1990)	Categorical	2 levels	$X_{20} = 1$ if patent applications in 1990 is > 0, and $X_{20} = 0$ if patent applications in 1990 is $= 0$
R&D expenditure (1991)	Quantitative	4 levels (after discretization with 25% of observations for each level)	$X_{R1} = 1$ if R&D expenditure is in the first 25%, and $X_{R1} = 0$ otherwise
			$X_{R2} = 1$ if R&D expenditure is in the second 25%, and $X_{R2} = 0$ otherwise
			$X_{R3} = 1$ if R&D expenditure is in the third 25%, and $X_{R3} = 0$ otherwise
			$X_{R4} = 1$ if R&D expenditure is in the last 25%, and $X_{R4} = 0$ otherwise

We applied the proposed method to this dataset and obtained the following results:

Step 1: Discretize the response into four levels. We followed the four levels for patent applications (Schwalbach

& Zimmermann, 1991) as summarized in Winkelmann (2008). We converted R&D expenditure, the quantitative variable, into a four-level categorical variable and thereby generated four dummy variables, as shown in Table 1.

Step 2: Use CBA to obtain the active rules. We used a minimum confidence value of 80% to generate the active rules. Note that this dataset has a large number of variables compared to number of instances.

Step 3: Select the 50 rules with the highest confidence values from among all the active rules. Examples of the selected rules:

- Rule 1: If $X_4 = 1$, $X_{20} = 1$, $X_{23} = 1$, then $Y = 2$ with $s = 3.315\%$, $c = 83.333\%$.
- Rule 2: If $X_5 = 1$, $X_{20} = 1$, $X_{22} = 1$, then $Y = 2$ with $s = 3.315\%$, $c = 83.333\%$.

Step 4: Convert the 50 selected rules into variables. For example, rule 1 was converted into the new variable referred to as $X_4(1)X_{20}(1)X_R$. Note that X_{23} was generated from X_R, which is the quantitative variable. Therefore, the conversion followed the second case of the variable generation (Step 4 in Section 3). To generate the new variable, we multiplied the original quantitative variable X_R with the dummy variable:

$$X_4(1)X_{20}(1) = \begin{cases} 1 & if \ X_4 = 1, X_{20} = 1 \\ 0 & otherwise \end{cases}.$$

Step 5: Apply the subset selection method to find the proposed model. Note that the candidate variables comprised the main effects (X_1, X_2, …,X_{24}) and all the potential interactions from Step 3.

We selected the ten-variable model which is the model with the best BIC to represent this dataset. The model is

$$\log \mu = -4.7131 - 1.8724X_1 + 1.5946X_5 + 0.5071X_9 - 2.1970X_{15} + 0.2660X_{17} - 0.4106X_{18}$$
$$+ 3.9275X_{20} + 0.9374X_R - 0.0871X_4(1)X_{20}(1)X_R - 0.3124X_5(1)X_{20}(1)X_R$$

Some major findings from the above model can be explained as follows:

  i. Among all sectors, the estimated mean for patent applications is highest in electricity sector and lowest in vehicle sector.

 ii. Among all areas, the estimated mean for patent applications from Japan is highest and the estimated mean for patent applications from USA is lowest.

iii. The estimated mean for patent applications in 1991 when there was at least one *patent applications in 1990* is higher compared to the estimated mean for patent applications in 1991 when there was no *patent applications in 1990*.

 iv. The estimated mean for patent applications increases with an increase in *R&D expenditure*.

  v. If the area is drugs and the *patent applications in 1990* is more than 0, an increase in *R&D expenditure* will decrease the estimated mean for patent applications.

 vi. If the area is electricity and the *patent applications in 1990* is more than 0, an increase in *R&D expenditure* will decrease the estimated mean for patent applications.

We stopped the search process at ten variables because the ten-variable model outperforms all the classical models, based upon the main effects only. Moreover, the proposed model also outperforms all the classical models based on the AIC criteria and deviance. Note that the AIC and BIC for this model are 8,047 and 8,112 whereas the AIC and BIC for the optimal classical model are 8,111 and 8,202, respectively. The deviance for this model is 8,025 whereas the optimal deviance among all the classical models is 8,075.

The model selected through the process described has two interactions, which when added to the Poisson regression model can elevate its performance tremendously as compared with the classical models. Therefore, compared to the classical model, the proposed method better explains the patents dataset and improves the prediction of patent applications. Note that the two interactions cannot be found by any available variable selection methods for the Poisson regression model.

The selected ten-variable model is as shown earlier. The three-way interactions are found from the proposed method. However, for those who believe in the hierarchical principle, the main effects and the two-way interactions of the variables that are involved in the three-way interactions can be added to the model as well.

5. Discussion and Conclusion

A central problem in Poisson regression modeling is accounting for interaction effects. In this study, we apply association rules with Poisson regression modeling with the central purpose of capturing possible low-order and

high-order interactions. Our study confirms that the methodology proposed herein can be used effectively to select interactions that improve model fit and to further our understanding of any given dataset. As shown via the real dataset, the results affirm the effectiveness of the proposed method used with Poisson regression modeling: specifically, the proposed method can provide a better explanation and a better fit for Poisson regression modeling than the classical method does. Note that it is not possible to select the two interactions in the proposed model using any conventional approach.

Among the data-mining techniques, the decision tree is capable of finding interactions among variables. However, the decision tree has the disadvantage of a hierarchical structure. Association rules analysis has an important advantage over the decision tree: association rules analysis enables a global search such that more potential interactions can be considered than is possible with a decision tree structure.

This paper shows that we can apply the association rules analysis to help finding important interactions not only from categorical variables but also from quantitative variables and with response that has myriads of different numbers by adding discretization into the process. This also leads us to believe that the proposed approach is also effective for any other linear regression models.

References

Agrawal, R., & Srikant, S. (1994). Fast algorithms for mining association rules. *Proceedings of the ACM SIGMOD Conference on Management of Data*, 487–499.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/0471249688

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705

Appice, A., Ceci, M., Lanza, A., Lisi, F. A., & Malerba, D. (2003). Discovery of spatial association rules in georeferenced census data: A relational mining approach. *Intelligent Data Analysis*, *7*(6), 541–566.

Becquet, C., Blachon, S., Jeudy, B., Boulicaut, J. F., & Gandrillon, O. (2002). Strong association rule mining for large gene expression data analysis: A case study on human SAGE data. *Genome Biology*, *12*(3), 1-16.

Berry, M. J. A., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support*. New York, NY: John Wiley & Sons.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511814365

Candes, E., & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, *35*, 2313–2351. http://dx.doi.org/10.1214/009053606000001523

Changchien, W., & Lu, T. (2001). Mining association rules procedure to support on-line recommendation by customers and products fragmentation. *Expert Systems with Applications*, *20*, 325–335. http://dx.doi.org/10.1016/S0957-4174(01)00017-3

Changpetch, P., & Lin, D. K. J. (2013a). Model selection for logistic regression via association rules analysis. *Journal of Computational Statistics and Simulations*, *83*(8), 1415-1428. http://dx.doi.org/10.1080/00949655.2012.662231

Changpetch, P., & Lin, D. K. J. (2013b). Selection for multinomial logit models via association rules analysis. *WIREs Computational Statistics*, *5*(1), 68–77. http://dx.doi.org/10.1002/wics.1242

Chaudhuri, P., Lo, W.-D., Loh, W.-Y., & Yang, C.-C. (1995). Generalized regression Trees. *Statistica Sinica*, *5*, 641–666.

Cincera, M. (1997). Patents, R&D, and technological spillovers at the firm level: Some evidence from econometric count models for panel data. *Journal of Applied Econometrics*, *12*, 265–280. http://dx.doi.org/10.1002/(SICI)1099-1255(199705)12:3<265::AID-JAE439>3.0.CO;2-J

Ding, Q., Ding, Q., & Perrizo, W. (2002). Association rule mining on remotely sensed images using p-trees. *Proceedings of Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD'02)*.

Dopfer, K., & Potts, J. (2004). Evolutionary realism: A new ontology for economics. *Journal of Economic Methodology*, *11*, 195–212. http://dx.doi.org/10.1080/13501780410001694127

Doorslaer, E. V., Koolman, X., & Jones, A. M. (2004). Explaining income-related inequalities in doctor utilization in Europe: A decomposition approach. *Health Economics*, *13*, 629–647. http://dx.doi.org/10.1002/hec.919

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*, 407–499. http://dx.doi.org/10.1214/009053604000000067

Fan, J. (1997). Comments on 'Wavelets in Statistics: A Review' by A. Antoniadis. *Journal of the Italian Statistical Association*, *6*, 131–138. http://dx.doi.org/10.1007/BF03178906

Garcia, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computer and Education Journal*, *49*, 794–808. http://dx.doi.org/10.1016/j.compedu.2005.11.017

Garcia, J., Romero, C., Ventura, S., & Calders, T. (2007). Drawbacks and solutions of applying association rules mining in learning management systems. *Proceedings of the International Workshop on Applying Data Mining in e-learning (ADML'07)*. Crete, Greece, 13–22.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). New York, NY: Prentice Hall.

Hahn, R. W., & Prieger, J. E. (2006). The impact of driver cell phone use on accidents. *Advances in Economic Analysis and Policy*, *6*(1), Article 9.

Hurvich, C., & Tsai, C. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis*, *14*, 271-279. http://dx.doi.org/10.1111/j.1467-9892.1993.tb00144.x

Kahn, M. E. (2005). The death toll from natural disasters: The role of income, geography, and institutions. *Review of Economics and Statistics*, *87*, 271–184. http://dx.doi.org/10.1162/0034653053970339

Kelly, M. (2000). Inequality and crime. *Review of Economics and Statistics*, *82*, 530–539. http://dx.doi.org/10.1162/003465300559028

Kuo, R. J., Lin, S. Y., & Shih, C. W. (2007). Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan. *Expert Systems with Applications*, *33*(3), 794–808. http://dx.doi.org/10.1016/j.eswa.2006.08.035

Lee, A. J. T., Hong, R. W., Ko, W. M., Tsao, W. K., & Lin, H. H. (2007). Mining spatial association rules in image databases. *Information Sciences*, *177*(7), 1593–1608. http://dx.doi.org/10.1016/j.ins.2006.09.018

Liu, B., Hsu, W., & Ma, Y. (1998). Integrating classification and association rule mining. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining* (*KDD'98*). New York, NY.

Liu, K. H., Weng, M. F., Tseng, C. Y., Chuang, Y. Y., & Chen, M. S. (2008). Association and temporal rule mining for post-processing of semantic concept detection in video. *IEEE TMM*, *10*(2), 240–251.

Marghny, M. H., & El-Semman, I. E. (2005). Extracting logical classification rules with gene expression programming: Microarray case study. *Proceedings of the International Conference on Artificial Intelligence and Machine Learning (AIML'05)*. Cairo, Egypt.

Matsumoto, K. (1998). An experimental agricultural data mining system. *Proceedings of the First International Conference on Discovery Science* (*DS'98*). Japan, 439–440.

Plassmann, F., & Tideman, T. N. (2001). Does the right to carry concealed handguns deter countable crimes? Only a count analysis can say. *Journal of Law and Economics*, *44*, 771–798. http://dx.doi.org/10.1086/323311

Quinlan, J. R. (1992). *C4.5: Programs for machine learning*. San Francisco, CA: Morgan Kaufmann.

Schellhorn, M. (2001). The effect of variable health insurance deductibles on the demand for physician visits. *Health Economics*, *10*, 441–456. http://dx.doi.org/10.1002/hec.630

Schwalbach, J., & Zimmermann, K. F. (1991). A Poisson model of patenting and firm structure in Germany," in Z. Acs and D. Audretsch (Eds.) *Innovation and technological change*, 109–120 New York, NY: Harvester Wheatsheaf.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. http://dx.doi.org/10.1214/aos/1176344136

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B, 58*, 267–288.

Tremblay, M., & Wallach, D. (2004). Comparison of parameter estimation methods for crop models. *Agronomie*, *24*, 351–365. http://dx.doi.org/10.1051/agro:2004033

Winkelmann, R. (2004). Health care reform and the number of doctor visits: An econometric analysis. *Journal of Applied Econometrics*, *19*, 455–472. http://dx.doi.org/10.1002/jae.764

Winkelmann, R. (2006). Reforming health care: Evidence from quantile regressions for counts. *Journal of Health Economics*, *25*, 131–145. http://dx.doi.org/10.1016/j.jhealeco.2005.03.005

Winkelmann, R. (2008). *Econometric analysis of count data* (5th ed.). Berlin Heidelberg: Springer-Verlag.

Winkelmann, R., & Zimmermann, K. F. (2000). Editorial: Fertility studies using count data models. *Journal of Population Economics*, *13*, 171–172. http://dx.doi.org/10.1007/s001480050131

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. http://dx.doi.org/10.1198/016214506000000735

Copyrights