Two-factor Interaction Effect Detection for the Generalized Linear Models

Jing Shyr [1], Jane Chu[1], Sier Han[2]
[1]IBM SPSS Predictive Analytics, 233 S. Wacker Dr. 11th Fl., Chicago, IL 60606, USA
[2] IBM SPSS Predictive Analytics, 2&3 Floor, Building C, Outsourcing Park Phase I, No. 11 Jinye 1st Rd., High Tech Zone, Xi'an, China

Abstract
This paper proposes a method of two-factor interaction effect detection for the generalized linear models. To test whether an interaction effect is significant, a likelihood ratio test is used because the traditional ANOVA type test, which is valid for linear models with the normality assumption, is not applicable. A likelihood ratio test is to compare the log-likelihood values between the full model (two main effects and an interaction effect) and the main effects only model. The log-likelihood value for the full model can be computed without estimating parameters, but parameter estimation is needed to obtain the log-likelihood value for the main effects only model. We propose to estimate parameters for the main effects only model under a linear model framework using only basic statistics. Since those basic statistics can be computed in a single data pass, the new method overcomes the drawback of many data passes needed in the traditional parameter estimation process for the generalized linear models. Implementation of such tests in a single data pass is important for the large and distributed data sources which become increasingly common in practice now.

Key words: generalized linear model, likelihood ratio test, influential combination

1. Introduction

Business Analysts like to know which factors (or categorical predictors) impact the target variable of interest and by how much they impact the target. A linear regression model is often used to answer that question. Furthermore, in many business scenarios, the interaction between factors is important though often overlooked. The patent application publication by Shyr et al. (2012) changes that by proposing an automated report discovery method. It is a linear-model-based and scalable process for discovery of a multitude of low-dimensional tabular reports exhibiting strong interaction, which describes a situation in which the simultaneous influence of two factors on the target is not additive. The interaction detection test is a traditional “ANOVA” (analysis of variance) method, but the test based on basic statistics will be applied many times to accommodate datasets with the large number of factors.

However the ANOVA method only works in linear regression models which assume the target follows a normal distribution and the linear relationship exists between the target and factors, so it won't work in more general models which have been developed and applied to solve many business problems as shown in the following examples.

A software company would like to determine which characteristics of customers will affect their decision to buy or not to buy its product, then a logistic regression model will be more appropriate because the target (buy or not to buy a product) is binary so a normal distribution is not right then Bernoulli distribution is assumed; and the mean of the target has to be between 0 and 1 so the linear assumption is not suitable then a function of the target mean is assumed to be linearly related to factors which is called a "logit link function".

A car insurance company would like to analyze what factors contribute the most to customer's claim size, then a seasoned analyst knows to fit a gamma regression to damage claims for cars because it is more appropriate to the analysis of positive range data by using a gamma distribution and an inverse link function to relate the mean of the target to a linear combination of the factors.

A shipping company concerns damage to cargo ships caused by waves and would like to determine which factors, such as ship types, years of construction, etc., are more prone to damage, then the incident counts should be modeled as occurring at a Poisson rate and a log-linear model (with a Poisson distribution and a log link function) is usually used.

Many such general models belong to so called "generalized linear models" which were first introduced by Nelder and Wedderburn (1972) and later expanded by McCullagh and Nelder (1989). The generalized linear model expands the linear regression model so that the target variable is linearly related to the predictors via a specified link function. Moreover, the model allows for the target to have a non-normal distribution.

Since the ANOVA method is not applicable in generalized linear models, a likelihood ratio test can be used to detect interaction. The likelihood ratio test is to compare log-likelihood values between the full and reduced models. For a two-way interaction, the full model includes two factors (also called "main effects") and an interaction effect while the reduced model includes only two main effects. Unfortunately, computation of log-likelihood value in the reduced model is an iterative process and requires many data passes. It might be tolerable if only one or few tests need to be performed. However, in order to apply the automated report discovery method to generalized linear models, it is not practical nor efficient to perform interaction detection many times for a large number of predictors because number of data passes will be multiplied when multiple interaction detection processes are conducted. This tremendous computation cost makes it impossible for the application under generalized linear models to handle large and distributed data sources. Hence a new efficient method for generalized linear models is needed to conduct multiple interaction detection processes based on basic statistics between categorical predictors and the target, which can be collected and computed in a single data pass.

The rest of sections are arranged as follows: Section 2 introduces the generalized linear model and the interaction detection based on log-likelihood ratio test. Log-likelihood value computation based on basic statistics is given in Section 3. Section 4 extends the two-way interaction method to m-way interaction detection. A conclusion is given in Section 5.

2. Generalized Linear Model and Interaction Detection

The multiplicative interaction detection in this paper mainly focuses on the generalized linear model. We assume readers understand the concept and terminologies of generalized linear models in general, otherwise read McCullagh and Nelder (1989) for details. Hence we first give a brief introduction of the generalized linear model then an interaction detection process is followed.

A generalized linear model of a target **y** with a set of predictors $\boldsymbol{X}$ has the form

$$\boldsymbol{\eta} = g(\mathrm{E}(\boldsymbol{y})) = g(\boldsymbol{\mu}) = \boldsymbol{X\beta}, \quad \boldsymbol{y} \sim F, \tag{1}$$

where η is the linear predictor; $g(\cdot)$ is the monotonic differentiable link function which states how the expectation of $\boldsymbol{y}$, $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\mu}$, is related to the linear predictor $\boldsymbol{\eta}$; F is the target's probability distribution. Choosing different combinations of a proper probability distribution and a link function can result in different models. The distributions we consider here and some commonly used link functions for the specific distributions are listed in Table 1. Please note that the combinations will not be limited to the list and the method in this paper can be applied to any distribution that belongs to exponential family and link function that is monotonic differentiable. However, for the ordinal multinomial distribution and its relevant cumulative link functions, the method proposed here cannot be extended straightforwardly and some cares are needed. Please contact the authors for further details.

Table 1: Distributions and Some Commonly Used Link Functions

Target distribution	**Link function**
Normal	Identity, Log, Power
Inverse Gaussian	Identity, Log, Power
Gamma	Identity, Log, Power
Poisson	Identity, Log, Power
Binomial	Logit, Probit, Complementary log-log
Nominal Multinomial	Logit

For each pair of categorical predictors, say $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, we would test whether the multiplicative interaction effect $\boldsymbol{X}_1 \times \boldsymbol{X}_2$ is significant in the following so called full model

$$\boldsymbol{\eta} = g(\mathrm{E}(\boldsymbol{y})) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \left(\boldsymbol{X}_1 \times \boldsymbol{X}_2\right)\boldsymbol{\beta}_3 \tag{2}$$

If the null hypothesis $H_0 : \boldsymbol{\beta}_3 = \mathbf{0}$ is not rejected, then the interaction effect $\boldsymbol{X}_1 \times \boldsymbol{X}_2$ should not be considered in the subsequent analyses and it becomes the following so called reduced model

$$\boldsymbol{\eta} = g(\mathrm{E}(\boldsymbol{y})) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 \tag{3}$$

Given are data for a target and a set of categorical predictors and some generalized linear model information (including the distribution and link function), the log-likelihood ratio test can be used to detect the two-way interaction, which can be described as below steps:

1) Compute the log-likelihood value for the full model, denote it as ℓ_{full}.
2) Compute the log-likelihood value for the reduced model, denote it as ℓ_{reduced}.
3) Compute the likelihood ratio test statistics: $\chi^2 = 2(\ell_{\text{full}} - \ell_{\text{reduced}})$.
4) Compute the *p*-value: $p = 1 - \Pr(\chi^2_{df} \le \chi^2)$, where χ^2_{df} is a random variable that follows a chi-squared distribution with df degrees of freedom, and df is the difference of the number of parameters between the full model and the reduced model.
5) If $p \le \alpha$, where α is a significant level (the default is usually 0.05), then the interaction effect $\boldsymbol{X}_1 \times \boldsymbol{X}_2$ is significant and it will be included in the subsequent analyses.

3. Log-likelihood Value Computation

To compute the log-likelihood value for the reduced model, Equation (3), we need to estimate parameters in the reduced model. Since there is no closed form solution unless it is a linear model (distribution is normal and link function is identity), the traditional method used is the Newton type algorithm which the first derivative (gradient) and/or second derivative (Hessian) need to be computed to update the parameters. The process is described in the Figure 1. We can see that the method is an iterative process in which each iteration means one data pass for computation of gradient and Hessian, so it is not practical to use the traditional method for each pair of predictors, in particular for large and distributed data sources. To overcome this drawback, this section discusses how to use basic statistics to compute log-likelihood values with only one data pass. Therefore, in this section, we will discuss how to use basic statistics to compute log-likelihood values. Especially, for reduced model, we proposed “recursive marginal mean accumulation” which do not need additional data pass.

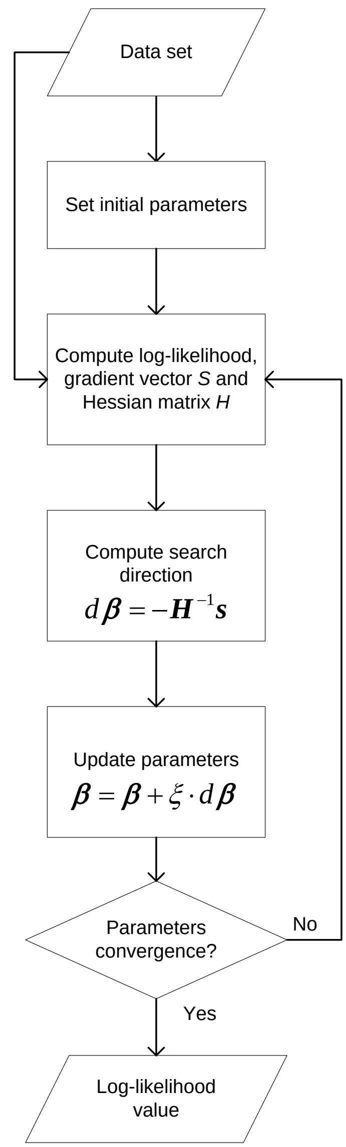

Figure 1. The traditional Newton Type Method for Log-likelihood Computation

3.1 Basic Statistics Collection

In this section, basic statistics among the target and each pair of categorical predictors for all possible predictors are computed in a single data pass.

For a pair of $\boldsymbol{X}_1$ and $\boldsymbol{X}_{2,}$ Table 2 shows a list of statistics to be collected and computed.

Table 2: Basic Statistics for Target Distribution

Target distribution	**Basic statistics**
Normal Inverse Gaussian Gamma Poisson	• R, the number of categories for predictor $\boldsymbol{X}_1$ • S, the number of categories for predictor $\boldsymbol{X}_2$ • N_{ij}, the number of records in the combination of $\boldsymbol{X}_1 = i$

Binomial	and $\boldsymbol{X}_2 = j$ • $\bar{y}_{ij}$, the target mean in the combination of $\boldsymbol{X}_1 = i$ and $\boldsymbol{X}_2 = j$
Nominal Multinomial	• R, the number of categories for predictor $\boldsymbol{X}_1$ • S, the number of categories for predictor $\boldsymbol{X}_2$ • K, the number of categories of target $\boldsymbol{y}$ • $N_{ij,k}$, the total number of records for the k^{th} target category in the combination of $\boldsymbol{X}_1 = i$ and $\boldsymbol{X}_2 = j$ • N_{ij}, the total number of records in the combination of $\boldsymbol{X}_1 = i$ and $\boldsymbol{X}_2 = j$ • $\bar{y}_{ij,k}$,the proportion of the k^{th} target category in the combination of $\boldsymbol{X}_1 = i$ and $\boldsymbol{X}_2 = j$, i.e. $\bar{y}_{ij,k} = N_{ij,k}/N_{ij}$

3.2 Pseudo Log-Likelihood functions

In reduced model and full model, there are some terms are the same in their log-likelihood function. So we exclude these terms from the original log-likelihood function and call the resulted function as pseudo log-likelihood functions which are listed in Table 3.

Table 3: Target Distribution and Pseudo Log-likelihood Function

Target distribution	Pseudo log-likelihood
Normal	$\ell = -\frac{1}{2}\sum_{i=1}^{R}\sum_{j=1}^{S} N_{ij}\left(\bar{y}_{ij} - \mu_{ij}\right)^2$
Inverse Gaussian	$\ell = -\frac{1}{2}\sum_{i=1}^{R}\sum_{j=1}^{S} N_{ij}\left(\frac{\bar{y}_{ij} - 2\mu_{ij}}{\mu_{ij}^2}\right)$
Gamma	$\ell = -\sum_{i=1}^{R}\sum_{j=1}^{S} N_{ij}\left(\ln\mu_{ij} + \frac{\bar{y}_{ij}}{\mu_{ij}}\right)$
Poisson	$\ell = \sum_{i=1}^{R}\sum_{j=1}^{S} N_{ij}\left(\bar{y}_{ij}\ln(\mu_{ij}) - \mu_{ij}\right)$
Binomial	$\ell = \sum_{i=1}^{R}\sum_{j=1}^{S} N_{ij}\left(\bar{y}_{ij}\ln(\mu_{ij}) + (1-\bar{y}_{ij})\ln(1-\mu_{ij})\right)$
Nominal multinomial	$\ell = \sum_{i=1}^{R}\sum_{j=1}^{S}\sum_{k=1}^{K} N_{ij,k} \times \ln(\mu_{ij,k})$

For the full model, the pseudo log-likelihood value will be computed by replacing the expectation of $\boldsymbol{y}$ in each cell, μ_{ij} or $\mu_{ij,k}$, by $\bar{y}_{ij}$ or $\bar{y}_{ij,k}$, respectively.

For the reduced model, μ_{ij} or $\mu_{ij,k}$ will be computed together with parameters estimation using the recursive marginal mean accumulation method which is described in section 3.3.

3.3 Recursive Marginal Mean Accumulation Method

In this sub-section, we proposed a new method called "recursive marginal mean accumulation" to compute log-likelihood value for reduced model. The method is a doubly iteratively process: updating parameters is iterative and computing parameter increments is also iterative. However, the advantage is that it doesn't need any data pass.

First, we can rewrite the reduced model in Equation (3), for all distributions except nominal multinomial (see a note in the end of this sub-section for nominal multinomial) based on each category combination of two factors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ as follows:

$$\eta_{ij} = g(\mu_{ij}) = \alpha_i + \beta_j \tag{4}$$

where α_i and β_j are the parameters for $\boldsymbol{X}_1 = i$ and $\boldsymbol{X}_2 = j$, and can be called "row parameter" and "column parameter", because their increments will be computed by the row marginal mean and column marginal mean of a two way table, respectively.

Then the doubly iterative process is described as follows:

(a) Set the initial values of α_i and β_j to be 0, for $i = 1, \ldots, R$ and $j = 1, \ldots, S$, and compute initial value of $\mu_{ij} = g(\alpha_i + \beta_j)^{-1}$, see Table 5 for the corresponding inverse forms.
(b) Compute the initial log-likelihood value by plugging initial values of μ_{ij} into formulae in Table 3.
(c) A $R \times S$ two-way table is created with the elements w_{ij} and s_{ij} in each cell, where

$$w_{ij} = \frac{N_{ij}}{V(\mu_{ij})\left(g'(\mu_{ij})\right)^2} + N_{ij}\left(\bar{y}_{ij} - \mu_{ij}\right) \times \frac{V(\mu_{ij})g''(\mu_{ij}) + V'(\mu_{ij})g'(\mu_{ij})}{\left(V(\mu_{ij})\right)^2\left(g'(\mu_{ij})\right)^3} \tag{5}$$

And

$$s_{ij} = \frac{1}{w_{ij}} \times \frac{N_{ij}\left(\bar{y}_{ii} - \mu_{ii}\right)}{V(\mu_{ij})g'(\mu_{ij})} \tag{6}$$

and $V(\mu_{ij})$ is the variance function of the target, $V'(\mu_{ij})$ is the first derivative of $V(\mu_{ij})$, $g'(\mu_{ij})$ and $g''(\mu_{ij})$ are the first and second derivatives of the link function, $g(\mu_{ij})$, respectively. Table 4 lists the variance functions and the corresponding first derivatives for distributions except nominal multinomial and Table 5 lists some commonly used link functions, the inverse forms and the first and second derivatives.

Table 4: Distribution and Its Variance Function and First Derivative

Distribution	$V(\mu)$	$V'(\mu)$

Normal	1	*0*
Inverse Gaussian	μ^3	$3\mu^2$
Gamma	μ^2	2μ
Poisson	μ	1
Binomial	$\mu(1-\mu)$	$1-2\mu$

Table 5: Link function, its inverse forms and first and second derivatives

Link function	$\eta = g(\mu)$	Inverse $\mu = g^{-1}(\eta)$	1st derivative $g'(\mu) = \frac{\partial\eta}{\partial\mu} = \Delta$	2nd derivative $g''(\mu) = \frac{\partial^2\eta}{\partial\mu^2}$
Identity	μ	η	1	0
Log	$\ln(\mu)$	$\exp(\eta)$	$\frac{1}{\mu}$	$-\Delta^2$
Power(α^*) $\begin{cases}\alpha \neq 0 \\ \alpha = 0\end{cases}$	$\begin{cases}\mu^\alpha \\ \ln(\mu)\end{cases}$	$\begin{cases}\eta^{1/\alpha} \\ \exp(\eta)\end{cases}$	$\begin{cases}\alpha\mu^{\alpha-1} \\ \frac{1}{\mu}\end{cases}$	$\begin{cases}\Delta\frac{\alpha-1}{\mu} \\ -\Delta^2\end{cases}$
Logit	$\ln\left(\frac{\mu}{1-\mu}\right)$	$\frac{\exp(\eta)}{1+\exp(\eta)}$	$\frac{1}{\mu(1-\mu)}$	$\Delta^2(2\mu-1)$
Probit	$\Phi^{-1}(\mu)$, where $\Phi(\xi) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\xi} e^{-z^2/2}dz$	$\Phi(\eta)$	$\frac{1}{\phi(\Phi^{-1}(\mu))}$, where $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$	$\Delta^2\Phi^{-1}(\mu)$
Complementary log-log	$\ln(-\ln(1-\mu))$	$1-\exp(-\exp(\eta))$	$\frac{1}{(\mu-1)\ln(1-\mu)}$	$-\Delta^2(1+\ln(1-\mu))$

(d) Compute the search directions, $d\alpha_i$ and $d\beta_j$, for row and column parameters, iteratively, by the following sub-steps:

(d-1) Set the initial values of $d\alpha_i$ and $d\beta_j$ to be 0.

(d-2) Update the search direction for row parameter by adding the marginal mean of the corresponding row:

$$d\alpha_i = d\alpha_i + s_{i\bullet}, \tag{7}$$

where $s_{i\bullet}$ is the weighted marginal mean of s_{ij} for row $i, i=1,\ldots,R,$

$$s_{i\bullet} = \frac{\sum_{j=1}^{S} w_{ij} \times s_{ij}}{\sum_{j=1}^{S} w_{ij}}. \tag{8}$$

(d-3) Update the two-way table by subtracting row marginal mean for each row:

$$s_{ij} = s_{ij} - s_{i\bullet}. \tag{9}$$

(d-4) Update the search direction for column parameter by adding the marginal mean of the corresponding column:

$$d\beta_j = d\beta_j + s_{\bullet j} \tag{10}$$

where $s_{\bullet j}$ is the weighted marginal mean of s_{ij} for column $j, j=1,\ldots,S,$

$$s_{\bullet j} = \frac{\sum_{i=1}^{R} w_{ij} \times s_{ij}}{\sum_{i=1}^{R} w_{ij}}. \tag{11}$$

(d-5) Update the two-way table by subtracting column marginal mean for each column:

$$s_{ij} = s_{ij} - s_{\bullet j}. \tag{12}$$

(d-6) Check whether the search directions converge by the following criterion

$$\max\left(\left|s_{i\bullet}\right|, \left|s_{\bullet j}\right|\right) < \varepsilon_1,$$

where ε_1 is a specified tolerance level. If the criterion is not met, go back to (d-2), otherwise go to (e).

(e) Update the row and column parameters:

$$\alpha_i = \alpha_i + \xi \times d\alpha_i, \quad \text{and}$$
$$\beta_j = \beta_j + \xi \times d\beta_j, \tag{13}$$

where ξ is a step length in a line search method.

(f) Compute the log-likelihood value with the updated mean of the target which is computed with the updated parameters.

(g) Check whether the parameters converge: the absolute difference of log-likelihood values in two successive iterations is less than a specified tolerance level, say ε_2, which can be different from ε_1.

If the criterion is not met, go back to (c), otherwise stop and output the final log-likelihood value.

Note that for nominal multinomial, computation is more complex with the following steps:

(a) The estimated expectations for each category of the target as

$$\pi_{ij,k} = \begin{cases} \dfrac{\exp\left(\alpha_{ik} + \beta_{jk}\right)}{1+\sum_{k=1}^{K-1}\exp\left(\alpha_{ik} + \beta_{jk}\right)}, & k = 1,\cdots,K-1, \\ \dfrac{1}{1+\sum_{k=1}^{K-1}\exp\left(\alpha_{ik} + \beta_{jk}\right)}, & k = K \end{cases}$$

(b) The log-likelihood value as

$$\sum_{i=1}^{R}\sum_{j=1}^{S}\sum_{k=1}^{K} N_{ij,k} \times \ln\left(\pi_{ij,k}\right)$$

(c) In the $R \times S$ two-way table, w_{ij} is extended from a scalar to a matrix and and s_{ij} to a vector as

$$\boldsymbol{w}_{ij} = N_{ij}\left(diag\left(\boldsymbol{\pi}_{ij}\right) - \boldsymbol{\pi}_{ij} \times \boldsymbol{\pi}_{ij}^{\mathrm{T}}\right) \text{ and}$$
$$\boldsymbol{s}_{ij} = N_{ij}\boldsymbol{w}_{ij}^{-1}\left(\bar{\boldsymbol{y}}_{ij} - \boldsymbol{\pi}_{ij}\right),$$

where $\boldsymbol{\pi}_{ij}^{\mathrm{T}} = \left(\pi_{ij,1},\ldots,\pi_{ij,K}\right)$ and $\bar{\boldsymbol{y}}_{ij}^{\mathrm{T}} = \left(\bar{y}_{ij,1},\ldots,\bar{y}_{ij,K}\right)$.

(d) The search directions, $d\alpha_i$ and $d\beta_j$, are extended to vectors, $d\boldsymbol{\alpha}_i$ and $d\boldsymbol{\beta}_j$.

The weighted marginal means of $\boldsymbol{s}_{ij}$ for row $i, i = 1,\ldots,R,$ and for column $j,\ j = 1,\ldots,S,$ are extended to vectors:

$$\boldsymbol{s}_{i\bullet} = \left(\sum_{j=1}^{S}\boldsymbol{w}_{ij}\right)^{-1} \times \left(\sum_{j=1}^{S}\boldsymbol{w}_{ij} \times \boldsymbol{s}_{ij}\right) \text{ and } \boldsymbol{s}_{\bullet j} = \left(\sum_{i=1}^{R}\boldsymbol{w}_{ij}\right)^{-1} \times \left(\sum_{i=1}^{R}\boldsymbol{w}_{ij} \times \boldsymbol{s}_{ij}\right), \text{ respectively.}$$

(e) The parameters, α_i and β_j, are extended to vectors, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$.

Figure 2 illustrates the proposed recursive marginal mean accumulation method and it is clear the method is a doubly iterative process.

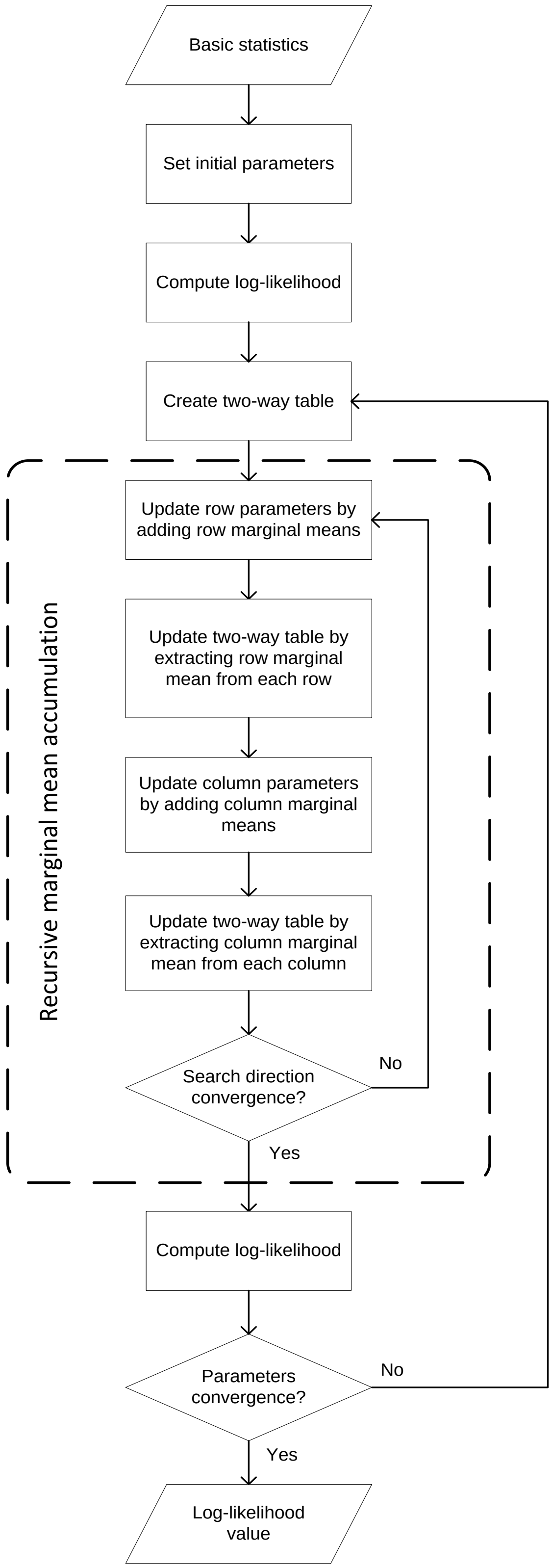

Fig 2. The Accumulative Marginal Mean Accumulation Method for Log-likelihood Computation

4. Extension to m-way Interaction Detection

The two-way interaction detection method that is introduced in previous sections can be extended to m-way interaction detection. The full model will contain all main effects, two-way interaction effects, …, and m-way interaction effect, and the reduced model is the model that the m-way interaction effect is excluded from the full model. The likelihood ratio test is still used to test if an m-way interaction effect is significant. Similar to the situation of two-way interaction detection, the basic statistics between the target and m categorical predictors are collected firstly. Then the log-likelihood value of the full model is computed based on these basic statistics. For the reduced model, the extended recursive marginal mean accumulation method for computation of the log-likelihood is described as follows:

(a) Set initial parameters corresponding to all main effects, two-way interaction effects, … , and (m-1)-way interaction effects to be 0.
(b) Compute the initial log-likelihood value based on initial parameters.
(c) Similar to the two-way interaction detection, create an m-way table.
(d) Compute the search directions iteratively
  (d-1) Set the initial search directions of one-way main effects, two-way interaction effects,..., (m-1)-way interaction effects to be 0.
  (d-2) Select one dimension in the m-way table, then update corresponding search directions of one-way main effect by adding the marginal means of this dimension, and update the m-way table by subtracting the marginal means of this dimension. Such process is repeated for other main effects.
  (d-3) Select two dimensions in the m-way table, then update corresponding search directions of two-way interaction effect by adding the marginal means of the two-dimensional table, and update the m-way table by subtracting the marginal means of the two-dimensional table. Such process is repeated for other two-way interaction
  (d-4) Similar to (d-2) or (d-3), update the search directions from three-way to (m-1)-way interaction effects.
  (d-5) Check whether the search directions converge: if the maximum absolute marginal means is less than a tolerance level. If the criterion is met, then go to the step (e), otherwise go back to step (d-2).
(e) Similar to the step (e) in the two-way interaction detection, update the parameters from one-way main effects to (m-1)-way interaction effects.
(f) Compute the log-likelihood value with the updated mean of the target which is computed with the updated parameters.
(g) Check whether the parameters converge: the absolute difference of log-likelihood values in two successive iterations is less than a specified tolerance level. If the criterion is not met, go back to (c), otherwise stop and output the final log-likelihood value.

5. Conclusion

This paper discussed two-factor interaction effect detection based on log-likelihood ratio test in the generalized linear models. To overcome the drawback of traditional method that needs many data pass to compute log-likelihood value for reduced model, we proposed a method called recursive marginal mean accumulation and extended it to m-way interaction detection situation. Since the proposed method is based on basic statistics which can be computed in a single data pass, it is very efficient to detect interaction effect in large and distributed data source.

References

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, Second Edition, London: Chapman and Hall.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society A*, 135, 370–384.

Shyr, J., Chu, Y.J., Spisic, D., Han, S. and Zhang, X.Y. (2012), *Relationship Discovery in Business Analytics*, US Patent Application Publication No. SVL920120001US1.