Letter to the Editor

Comment on “Modified Weights-of-Evidence Modeling with Example of Missing Geochemical Data”

Helmut Schaeben

Department of Geophysics and Geoinformatics, TU Bergakademie Freiberg, Freiberg, Germany

Correspondence should be addressed to Helmut Schaeben; schaeben@tu-freiberg.de

Received 2 May 2019; Accepted 26 May 2020; Published 22 July 2020

Academic Editor: Carlos Gershenson

Copyright © 2020 Helmut Schaeben. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Weights-of-Evidence, Modified Weights-of-Evidence, and Logistic Regression in Prospectivity Modeling

Even when restricted to a particular method, the undertaking by Zhang and Agterberg [1] seems surprising to review a topic as complex as estimating probabilities [2–5] without a single mathematical definition or formula. Thus, it becomes almost inevitable that this review provokes comments for reasons of mathematical clarification, in particular by this author because he has fragmentarily been referred to at several instances.

This comment intends to clarify the required mathematical modeling assumptions and the mutual relationship of mathematical methods, namely, weights-of-evidence, modified weights-of-evidence, and logistic regression of prospectivity modeling, and in particular, their possibilities to handle missing data as addressed by Zhang and Agterberg [1]. Their communication [1] solely argues with practical applications. However, what do practical applications of mathematical methods convey about the methods and their relationships? The mathematics of methods, e.g., properties of estimators, cannot be compared by the way of practical applications, i.e., by estimates. Case studies may be useful to compare the performance of methods in terms of cpu time and storage consumption.

Joint conditional independence of predictor variables, given the target variable, is a property of a set of random variables, and it cannot be controlled by any method applied

to the variables. If it is satisfied, contrasts of weights-of-evidence and logistic regression parameters with indicator variables are the same; in particular, they are independent of one another. If it is not satisfied, the logistic regression parameters are different from the contrasts of weights-of-evidence, while the latter is still independent and the former is dependent to counterbalance the violation of conditional independence. Including interaction terms of those random variables which cause the violation of conditional independence in the logistic regression model exactly compensates for the lack of conditional independence. Weights-of-evidence cannot reasonably include interaction terms. Thus, in case of lacking conditional independence, there is no need to turn weights-of-evidence into modified weights-of-evidence [6] by successively applying first weights-of-evidence to the initial indicator variables even though the application is not mathematically authorized, and secondly, logistic regression to the corrupted weights to yield corrected weights.

Moreover, this comment elaborates on the case of missing data; it reveals the details of the strategy numerically realized by weights-of-evidence, and it shows that logistic regression can easily be adjusted. Assuming that logistic regression and modified weights-of-evidence are “equivalent” as Zhang et al. claim [1], then why not use logistic regression in the first place? Since logistic regression does not only apply to random indicator variables as weights-of-evidence but also to continuous random variables, an indicator transformation according to a user defined threshold becomes obsolete.

2. A False Model Cannot Decrease Bias

How can a method [7] “significantly reduce the bias” (p. 2) and “diminish lack of CI (conditional independence) bias in WofE (Weights-of-Evidence) results” (p. 2) that was not just illustrated by way of a counter-example but explicitly proven to be mathematically false [8] by grossly mistaking that the logit of a sum is the sum of logits ([8]; p. 406)?

3. Boosting Does Not Remedy Lack of Joint Conditional Independence

How can “. . . better results generally (can) be obtained by other methods such as boosting [22–24] . . .” (p. 2) as put forward by Cheng [9]; in [24] by Zhang and Agterberg, even though it was explicitly shown and concluded [10] that “Boost weights-of-evidence does not generally reduce not to mention significantly reduce the effect of lacking joint conditional independence. Its application yields corrupted predicted conditional probabilities and corrupted spatial patterns of prospectivity as weights-of-evidence does. In particular, its results depend on the sequential processing order of predictors.” [10]; (p. 404). “. . . criteria for the user decision how to choose the sequence are not provided.” [10]; (p. 404). This publication [10] is not referred to nor included in the authors’ references.

4. Lack of Conditional Independence Corrupts Weights-of-Evidence

Joint conditional independence of predictor variables, given the target variable, is required to simplify multifoldly conditional probabilities involved in Bayes theorem for several variables to simply conditional probabilities. It is a required modeling assumption of weights-of-evidence to ensure its features and proper estimates of probabilities. If this assumption is relaxed to “. . . approximately conditional independence . . .” (p. 1), then weights-of-evidence results in approximately proper estimates of probabilities. “Logistic regression prevents lack of CI (conditional independence) bias in situations where WofE does not” (p. 2), and it does perfectly so by including proper interaction terms [11]. Actually, logistic regression including proper interaction terms compensates lack of joint conditional independence completely [11]. Weights-of-evidence cannot reasonably include interaction terms.

Should the simple but in case of lacking joint conditional independence mathematically unauthorized and erroneous weights-of-evidence model be preferentially applied because its weights and contrasts are easy to interpret? In contrary to “Logistic regression prevents lack of CI bias in situations where WofE does not. . . LR (logistic regression) coefficients then (accentuation by this author) cannot be used individually for interpretation or prediction purposes. This in

itself can be a good reason to prefer usage of WofE.” (pp. 2-3) it is actually not a good reason to prefer weights-of-evidence, because (i) in case of lacking joint conditional independence, weights-of-evidence must not be applied, and (ii) in case of joint conditional independence of all B_ℓ , given T , as required by weights-of-evidence logistic regression parameters β_ℓ like contrasts C_ℓ can be interpreted in terms of conditional odds ratios:

$$\exp(\beta_\ell) = \exp(C_\ell) = \frac{O(T = 1|B_\ell = 1)}{O(T = 1|B_\ell = 0)}, \quad \ell = 1, \dots, m. \quad (1)$$

The odds ratio is a measure of association; it approximates how much more likely (or unlikely) it is for the target to be present ($T = 1$) among those with $B_\ell = 1$ than among those with $B_\ell = 0$.

Moreover, since “the WofE contrast [C_ℓ] measures the strength of spatial correlation between a point pattern and an indicator pattern,” (p. 2) so do the logistic regression parameters β_ℓ as they are identical for jointly conditionally independent indicator predictor variables discussed as follows.

Logistic regression applies to continuous random predictor variables and does not require an indicator transformation by thresholding. Logistic regression is the canonical generalization of weights-of-evidence, or the other way round, weights-of-evidence is the special case of logistic regression if the predictor variables are dichotomous and jointly conditionally independent, given the dichotomous target variable.

5. How Missing Data Are Accounted for?

The “category” of missing data is technically often referred to as not available (NA) whether the data are categorical or not. The authors’ review is restricted to dichotomous predictor variables indicating the presence or absence of a given phenomenon. The “powerful tool to deal with missing data” (p. 1) comes down to “Pixel or unit area input for MWofE (Modified Weights-of-Evidence) consists of positive and negative weights for the presence and absence of a pattern plus zeros for missing data” (p. 1). Why is “logistic regression as such (is) not capable of handling missing data” (p. 1)?

Assuming m jointly conditionally independent indicator predictor variables $B_1, \dots, B_m$, given the dichotomous target variable T the weights-of-evidence are individually determined by counting events, e.g., the ordinary weights,

$$W_\ell(i) = \ln \frac{P(B_\ell = i|T = 1)}{P(B_\ell = i|T = 0)}, \quad i = 0, 1, \quad (2)$$

of B_ℓ are estimated, provided that B_ℓ is not missing for any unit area as follows:

$$\widehat{W}_\ell(1) = \ln \frac{\sum_{i=1}^n B_{\ell,i} T_i / \sum_{i=1}^n T_i}{\left(\sum_{i=1}^n B_{\ell,i} (1 - T_i) \right) / \sum_{i=1}^n (1 - T_i)} = \ln \frac{\sum_{i=1}^n B_{\ell,i} T_i / \sum_{i=1}^n T_i}{\left(\sum_{i=1}^n B_{\ell,i} - \sum_{i=1}^n B_{\ell,i} T_i \right) / (n - \sum_{i=1}^n T_i)}, \quad (3)$$

$$\widehat{W}_\ell(0) = \ln \frac{\left(\sum_{i=1}^n (1 - B_{\ell,i}) T_i \right) / \sum_{i=1}^n T_i}{\left(\sum_{i=1}^n (1 - B_{\ell,i}) (1 - T_i) \right) / \sum_{i=1}^n (1 - T_i)} = \ln \frac{\left(\sum_{i=1}^n T_i - \sum_{i=1}^n B_{\ell,i} T_i \right) / \sum_{i=1}^n T_i}{\left(n - \sum_{i=1}^n B_{\ell,i} - \sum_{i=1}^n T_i + \sum_{i=1}^n B_{\ell,i} T_i \right) / (n - \sum_{i=1}^n T_i)}, \quad (4)$$

where n denotes the total number of unit areas, usually the total number of pixels of a digital map image. Provided that all predictor variables $B_\ell, \ell = 1, \dots, m$, are not missing for

any unit area, the practical ordinary weights-of-evidence model is

$$\begin{aligned} \widehat{\text{logit}}(T = 1 | B_1, \dots, B_m) &= \widehat{\text{logit}}(T = 1) + \sum_{\ell: B_\ell=1} \widehat{W}_\ell(1) + \sum_{\ell: B_\ell=0} \widehat{W}_\ell(0) \\ &= \widehat{\text{logit}}(T = 1) + \sum_{\ell=1}^m (\widehat{W}_\ell(1) B_\ell + \widehat{W}_\ell(0) (1 - B_\ell)) \\ &= \widehat{\text{logit}}(T = 1) + \widehat{W}(0) + \sum_{\ell=1}^m \widehat{C}_\ell B_\ell \\ &= \widehat{\text{logit}}(T = 1) + \widehat{W}(0) + \sum_{\ell: B_\ell=1} \widehat{C}_\ell, \end{aligned} \quad (5)$$

with $\widehat{W}(0) = \sum_{\ell=1}^m \widehat{W}_\ell(0)$ and contrasts $\widehat{C}_\ell = \widehat{W}_\ell(1) - \widehat{W}_\ell(0), \ell = 1, \dots, m$, which are identical with the parameters of the corresponding logistic regression model [11, 12] as we assumed joint conditional independence of all B_ℓ given T . Otherwise, a linear relationship between weights-of-evidence and logistic regression parameters does not exist [11, 12]. It should be noted that the total number of events $B_\ell = 1$ or $B_\ell = 0$ is n , which is the total number of pixels of the digital map image.

To keep the notation light, just the indicator variable B_{ℓ_0} is assumed to be missing for some $0 < n_{\text{NA}} < n$ pixels. Then, in equations (3) and (4), n has to be replaced by $n' = n - n_{\text{NA}}$, thus leading generally to a biased estimate $\widehat{W}_{\ell_0}'$ of W_{ℓ_0} only. The corresponding practical “modified” weights-of-evidence model is as follows:

(i) For pixels where B_{ℓ_0} is missing,

$$\begin{aligned} \widehat{\text{logit}}(T = 1 | B_1, \dots, B_{\ell_0} = \text{NA}, \dots, B_m) &= \widehat{\text{logit}}(T = 1) + \sum_{\ell: B_\ell=1} \widehat{W}_\ell(1) + \sum_{\ell: B_\ell=0} \widehat{W}_\ell(0) + \sum_{\ell: B_\ell=\text{NA}} \widehat{W}_\ell(\text{NA}) \\ &= \widehat{\text{logit}}(T = 1) + \sum_{\substack{\ell: B_\ell=1 \\ \ell \neq \ell_0}} \widehat{W}_\ell(1) + \sum_{\substack{\ell: B_\ell=0 \\ \ell \neq \ell_0}} \widehat{W}_\ell(0) + \underbrace{\widehat{W}_{\ell_0}(\text{NA})}_{\equiv 0} \\ &= \widehat{\text{logit}}(T = 1) + \sum_{\substack{\ell: B_\ell=1 \\ \ell \neq \ell_0}} \widehat{W}_\ell(1) + \sum_{\substack{\ell: B_\ell=0 \\ \ell \neq \ell_0}} \widehat{W}_\ell(0) \\ &= \widehat{\text{logit}}(T = 1) + \widetilde{W}(0) + \sum_{\substack{\ell: B_\ell=1 \\ \ell \neq \ell_0}} \widehat{C}_\ell, \end{aligned} \quad (6)$$

where $\widetilde{W}(0) = \sum_{\substack{\ell=1 \\ \ell \neq \ell_0}}^m \widehat{W}_\ell(0)$.

(ii) For pixels where B_{ℓ_0} is not missing,

$$\begin{aligned}
\widehat{\text{logit}}(T = 1 | B_1, \dots, B_{\ell_0} \neq \text{NA}, \dots, B_m) &= \widehat{\text{logit}}(T = 1) + \sum_{\substack{\ell: B_\ell=1 \\ \ell \neq \ell_0}} \widehat{W}_\ell(1) + \sum_{\substack{\ell: B_\ell=0 \\ \ell \neq \ell_0}} \widehat{W}_\ell(0) + \widehat{W}'_{\ell_0}(1) - \widehat{W}'_{\ell_0}(0) \\
&= \widehat{\text{logit}}(T = 1) + \widetilde{W}(0) + \sum_{\substack{\ell: B_\ell=1 \\ \ell \neq \ell_0}} \widehat{C}_\ell + \widehat{C}'_{\ell_0},
\end{aligned} \tag{7}$$

with $\widetilde{W}(0) = \sum_{\substack{\ell=1 \\ \ell \neq \ell_0}}^m \widehat{W}_\ell(0) + \widehat{W}'_{\ell_0}(0)$ and $\widehat{C}'_{\ell_0} = \widehat{W}'_{\ell_0}(1) - \widehat{W}'_{\ell_0}(0)$.

Thus, for pixels where B_{ℓ_0} is missing, it is just omitted by setting its weights to 0; for pixels where B_{ℓ_0} is not missing, its weights W'_{ℓ_0} and its contrast C'_{ℓ_0} are biased as their support does not cover the entire map image. Due to the assumption of joint conditional independence of all $B_\ell, \ell = 1, \dots, m$, given T their weights and contrasts are independent whatever their common support in terms of total numbers of pixels. Thus, the contrasts $C_\ell, \ell \neq \ell_0$, remain unaffected. Nevertheless, equation (7) represents a mixture of two different models while equation (6) represents a restricted model.

A similar procedure applies to ordinary logistic regression, i.e., the two logistic regression models are as follows:

(i)

$$\begin{aligned}
\widehat{\text{logit}}_n(T = 1 | B_1, \dots, B_{\ell_0} = \text{NA}, \dots, B_m) \\
= \widehat{\beta}_0 + \sum_{\substack{\ell=1 \\ \ell \neq \ell_0}}^m \widehat{\beta}_\ell B_\ell,
\end{aligned} \tag{8}$$

supported by all pixels of the digital map image, and

(ii)

$$\begin{aligned}
\widehat{\text{logit}}_n(T = 1 | B_1, \dots, B_{\ell_0} \neq \text{NA}, \dots, B_m) \\
= \widehat{\beta}'_0 + \sum_{\ell: B_\ell=1} \widehat{\beta}'_\ell B_\ell,
\end{aligned} \tag{9}$$

supported by those pixels for which B_{ℓ_0} is not NA.

In contrast to modified weights-of-evidence, all parameters of the latter model, equation (9), are affected by the fact that B_{ℓ_0} is missing for some pixels. Since logistic regression does not require a mathematical modeling assumption as restrictive as joint conditional independence of all B_ℓ , given T , its parameters are not independent which allows to some extent to account for any kind of interdependencies. Including interaction terms turns logistic regression nonlinear and allows to compensate any lack of joint conditional independence exactly, i.e., logistic regression for dichotomous variables including proper interaction terms recovers the true conditional probability exactly; thus, it is optimum [11].

Depending on what caused realizations of B_{ℓ_0} to be missing at some locations, an alternative option to resolve that problem might be spatial interpolation as the data are georeferenced.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

- [1] D. Zhang and F. Agterberg, “Modified weights-of-evidence modeling with example of missing geochemical data,” *Complexity*, vol. 2018, Article ID 7945960, 12 pages, 2018.
- [2] C. S. Peirce, “The probability of induction,” in *Popular Science Monthly 1878, Republished the World of Mathematics 2*, J. R. Newman, Ed., pp. 1341–1354, Simon and Schuster, New York, NY, USA, 1956.
- [3] M. Minsky and O. G. Selfridge, “Learning in random nets,” in *Information Theory: Fourth London Symposium*, C. Cherry, Ed., pp. 335–347, Butterworths, London, UK, 1961.
- [4] I. J. Good, *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, Research Monograph No. 30, The MIT Press, Cambridge, UK, 1968.
- [5] G. Matheron, *Estimating and Choosing—An Essay on Probability in Practice*, Springer, Berlin, Germany, 1989.
- [6] F. Agterberg, “A modified weights-of-evidence method for regional mineral resource estimation,” *Natural Resources Research*, vol. 20, no. 2, pp. 95–101, 2011.
- [7] M. Deng, “A conditional dependence adjusted weights of evidence model,” *Natural Resources Research*, vol. 18, no. 4, pp. 249–258, 2009.
- [8] H. Schaeben and K. G. van den Boogaart, “Comment on “A conditional dependence adjusted weights of evidence mode” by Minfeng Deng in natural resources research 18 (2009), 249–258,” *Natural Resources Research*, vol. 20, no. 4, pp. 401–406, 2011.
- [9] Q. Cheng, “BoostWofE: A new sequential weights of evidence model reducing the effect of conditional dependency,” *Mathematical Geosciences*, vol. 47, no. 5, pp. 591–621, 2015.
- [10] H. Schaeben and G. Semmler, “The quest for conditional independence in prospectivity modeling: Weights-of-evidence, boost weights-of-evidence, and logistic regression,” *Frontiers of Earth Science*, vol. 10, no. 3, pp. 389–408, 2016.
- [11] H. Schaeben, “A mathematical view of weights-of-evidence, conditional independence, and logistic regression in terms of Markov random fields,” *Mathematical Geosciences*, vol. 46, no. 6, pp. 691–709, 2014.
- [12] H. Schaeben, “Targeting: Logistic regression, special cases and extensions,” *ISPRS International Journal of Geo-Information*, vol. 3, no. 4, pp. 1387–1411, 2014.