Article

Modeling Categorical Random Fields via Linear Bayesian Updating

Xiang Huang and Zhizhong Wang *

Department of Statistics, Central South University, Changsha 410012, China; huangxiang@csu.edu.cn

* Correspondence: wzz8713761@163.com; Tel.: +86-136-2731-1056

Abstract: Categorical variables are common in spatial data analysis. Traditional analytical methods for deriving probabilities of class occurrence, such as kriging-family algorithms, have been hindered by the discrete characteristics of categorical fields. This study introduces the theoretical backgrounds of linear Bayesian updating (LBU) approach for spatial classification through expert system. Transition probabilities are interpreted as expert opinions for updating the prior marginal probabilities of categorical response variables. The main objective of this paper is to present the solid theoretical foundations of LBU and provide a categorical random field prediction method which yields relatively higher classification accuracy compared with conventional Markov chain random field (MCRF) approach. A real-world case study has also been carried out to demonstrate the superiority of our method. Since the LBU idea is originated from aggregating expert opinions and not restricted to conditional independent assumption (CIA), it may prove to be reasonably adequate for analyzing complex geospatial data sets, like remote sensing images or area-class maps.

Keywords: Bayesian updating; expert opinion; spatial classification; transition probability

PACS: 02.50.Sk

1. Introduction

Categorical spatial data, such as lithofacies, land-use/land-cover (LULC) classifications, mineralization phases, are widely investigated geographical and geological data sources. They are typically represented by mutually exclusive and collectively exhaustive (MECE) classes and visualized as area-class maps [1]. Traditionally, indicator variograms or indicator covariances are commonly used in geostatistics to measure spatial continuity in categorical fields [2]. Different from (but related to) those two indicator approaches, transition probabilities [3], or equivalently, transiograms [4], are naturally designed to quantify spatial variability of categorical variables. The advantages of transition probability-based methods over indicator covariances or variograms had been well discussed [3]. The critical and most challenging step here is to model the data redundancy between these elementary probabilities or interactions among multiple variables. Several methods have been developed in consideration of data interaction, e.g., Tau (τ) model [5,6], Nu (ν) expression [7], generalized linear mixed model (GLMM) [8], etc. The paradigm of latent random fields and spatial covariance functions used in GLMM was also discussed in [9], the results show that the class occurrence probability at a target location can be formulated as a multinomial logistic function of spatial variable covariances between the target and sampled locations. The interaction terms were included into logistic regression models in [10] to account for data dependency in terms of Markov random fields [11,12]. In the Geo-information context, the Rao's quadratic diversity was used in [13] to measure the scale-dependent landscape structure. In the geological counterpart, a spatial hidden Markov chain model was employed in [14] for estimation of petroleum reservoir categorical variables. Most often, the step of considering data interaction is avoided, or at least simplified, by assuming some form of conditional independent assumption (CIA), such as the Markov chain random field (MCRF) method, introduced by [15]. The recently proposed MCRF co-located

cosimulation algorithm [16] is an enrichment of the previous MCRF theory by integrating auxiliary data.

Generally speaking, the spatial classification problem can be regarded as combing two-point transition probabilities into a multi-point conditional probability. A formal review of most of the available techniques to aggregate probability distributions in geoscience can be found in [17]. What we focus on is to use probability aggregation method for spatial classification based on the pioneering work of [18] and [19]. This paper is organized as follows. We start by introducing the basic forms of the linear Bayesian updating (LBU) method in Section 2. We then make detailed proofs in Section 3 for some propositions introduced by [18] and [19] while have not yet been proven. Finally, to further investigate the performance of the proposed method in real cases, the lithology types in the well-known Jura data set [20] are used for case study (Section 4). The contribution is completed by conclusions and acknowledgments in Section 5.

2. Linear Bayesian updating model

We use A and $D_1, \dots, D_n$ to represent the events in sample spaces of categorical random variable $C(x_0)$ and $C(x_1), \dots, C(x_n)$ respectively and $\bar{A}$ denotes the complementary event of A . We wish to assess the probability of an event A , conditional on the occurrence of a set of data events $D_i, i = 1, 2, \dots, n$. For categorical events, the formal probabilistic set-up is the following. We need to consider a sample space Ω such that all events A and D_i are subsets of Ω . In the case of categorical data, let $\mathbf{A}$ be the finite set of events in Ω such that the events $A_1, A_2, \dots, A_K$ of $\mathbf{A}$ are MECE, that is $\mathbf{A}$ forms a finite partition of Ω . Obviously,

$$A_j = \begin{cases} 1 & \text{if } C(x_0) = j, \quad j = 1, 2, \dots, K \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Without loss of generality, we will use A as a short notation for A_j . The event A can for example be a lithofacies category at a specified location x_0 , while the data D_i can represent information provided by core samples at surrounding wells $x_1, x_2, \dots, x_n$. This means that we wish to approximate the probability $P(A | D_1, D_2, \dots, D_n)$ on the basis of the simultaneous knowledge. This multi-point posterior probability can be obtained by combining prior probability $P_0(A)$ with the two-point transition probabilities $P(A | D_i), i = 1, 2, \dots, n$. This prior probability is independent on any other probabilities $P(A | D_i)$. It can be thought of as arising from an abstract and never specified information D_0 with $p = P(A | D_0)$. In geoscience, such a prior probability could be for example the proportion of a kind of lithofacies varying in space [17]. The n neighboring events $D_1, D_2, \dots, D_n$ are regarded as experts, thus the conditional probability $P(A | D_i)$ can be considered as expert opinion Q_i for the occurrence of A , an event of interest.

We then examine the situation in which a decision maker (DM) uses the opinions of $n \geq 1$ expert individual $\mathbf{Q} = (Q_1 = q_1, Q_2 = q_2, \dots, Q_n = q_n)$ to revise his/her own prior probability $P_0(A)$, denoted p , for the occurrence of A , and it is desired to use this information to form the DM's posterior probability for A . This process is called Bayesian updating. The experts' opinions are treated as random variables Q_i whose values $q_i, 1 \leq i \leq n$, are to be revealed to the DM. The posterior probability of A given $\mathbf{Q} = \mathbf{q}$ is then $p^*(\mathbf{q})$.

The LBU method was proposed by [18] of the form

$$p^*(\mathbf{q}) = p + \sum_{i=1}^n \lambda_i (q_i - \mu_i) \quad (2)$$

with possibly negative weights, λ_i , expressing the amounts of correlation between each Q_i and A . μ_i denotes the mathematical expectation of Q_i . When $\mu_1 = \dots = \mu_n = p$ and $\lambda_i \geq 0$ for all i , (2) reduces to the linear opinion pool [21]

$$p^*(\mathbf{q}) = \lambda_0 p + \sum_{i=1}^n \lambda_i q_i \quad \text{subject to} \quad \sum_{i=0}^n \lambda_i = 1 \quad (3)$$

where the DM is considered as one of the experts, thereby lending support to a supposition of [22] to the effect that (3) sometimes corresponds to an application of Bayes' theorem. To obtain a legitimate probability, λ_i must obey a number of inequalities [18]. If, for example, the DM feels that all λ_i are positive, the most common case, then they must be chosen so that

$$\max \left\{ \sum_{i=1}^n \lambda_i \mu_i / p, \sum_{i=1}^n \lambda_i (1 - \mu_i) / (1 - p) \right\} \leq 1, \quad (4)$$

which can be considered as a sufficient but not necessary condition for LBU model. The proof of (4) is given in Section 3.1.

The LBU method presented above has some parameters that need either to be estimated or specified by the DM. In the context of aggregating expert opinion, [23] introduced four ways of assessing the weights for the linear pool: the first suggestion is to set equal weights when there is no element which allows to prefer one source of information to another, or when symmetry of information justifies it; the second and third propositions are setting weights proportional to a ranking based on expert's advice and self-rating, respectively. The last suggestion is to set weights based on some comparison of previously assessed distributions with actual outcomes. An algorithm [24] is based on the minimization of a Kullback-Leibler distance for selecting weighting factors in logarithmic opinion pools. The optimal weights are found by solving a quadratic programming problem. We can also find approach based on ordinary kriging, a well-known framework for accounting for data interaction in a spatial context [25]. In their discussion, information interaction between D_i and D_{i-1} is encapsulated by the exponent weight factor τ_i , but the concept of distance between source of information and that of variogram of probabilities is not obvious [17]. Our derivation provides an interpretation of the coefficients λ_i in LBU (2) as the linear regression coefficients of $p^* - p$ on $\mathbf{Q} - \boldsymbol{\mu}$ so long as the experts are not linearly dependent. The proof of this proposition can be found in Section 3.2. As in multiple regression, each λ_i can thus be thought of as a measure of the additional information that the i th expert provides over and above the other experts and what the DM already knows.

3. Theoretical foundations of linear Bayesian updating model

3.1. Parameter ranges for linear Bayesian updating

Since $p^*(\mathbf{q})$ is a probability, it must satisfy $0 \leq p^*(\mathbf{q}) \leq 1$, that is to say,

$$\begin{cases} p + \sum_{i=1}^n \lambda_i (q_i - \mu_i) \geq 0 \\ p + \sum_{i=1}^n \lambda_i (q_i - \mu_i) \leq 1. \end{cases} \quad (5)$$

Through algebra transformation, (5) can be simplified as

$$\begin{cases} 1 + \frac{\sum_{i=1}^n \lambda_i q_i}{p} \geq \frac{\sum_{i=1}^n \lambda_i \mu_i}{p} \\ \frac{\sum_{i=1}^n \lambda_i q_i - \sum_{i=1}^n \lambda_i \mu_i}{1-p} \leq 1. \end{cases} \quad (6)$$

Suppose that all λ_i are positive, since $0 \leq q_i \leq 1$, as long as

$$\begin{cases} \frac{\sum_{i=1}^n \lambda_i \mu_i}{p} \leq 1 \\ \frac{\sum_{i=1}^n \lambda_i (1 - \mu_i)}{1-p} \leq 1, \end{cases} \quad (7)$$

(6) can be satisfied. Therefore, (7) can be considered as a sufficient but not necessary condition for LBU model (2). Only when (7), or equivalently (4), is satisfied, can (2) be a valid probability model.

3.2. Interpreting parameters as regression coefficients for linear Bayesian updating

Let $\Sigma_{\mathbf{Q}}$ denote the covariance matrix of $\mathbf{Q}$ and $\sigma_{p^* \mathbf{Q}}$ be the vector of covariances between p^* and $\mathbf{Q}$. Let t denote matrix transposition and

$$\mathbf{Q} = (Q_1, Q_2, \dots, Q_n) \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n) \quad \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)^t,$$

using the definition of expectation, it is

$$E(A | \mathbf{Q}) = p^*, \quad E(E(A | \mathbf{Q})) = E(A) = E(p^*) = p.$$

In addition, (2) can be given as

$$p^* = p + (\mathbf{Q} - \boldsymbol{\mu}) \cdot \boldsymbol{\lambda} \quad \text{and} \quad \boldsymbol{\mu} = E(\mathbf{Q}),$$

i.e.,

$$p^* = p + (Q_1 - \mu_1, Q_2 - \mu_2, \dots, Q_n - \mu_n) \cdot (\lambda_1, \lambda_2, \dots, \lambda_n)^t.$$

Taking the expectations on both sides of the equation after transformation yields

$$\begin{aligned}
E[(\mathbf{Q}-\boldsymbol{\mu})^t \cdot (p^* - p)] &= E[(Q_1 - \mu_1, Q_2 - \mu_2, \dots, Q_n - \mu_n)^t \cdot (Q_1 - \mu_1, Q_2 - \mu_2, \dots, Q_n - \mu_n)] \cdot (\lambda_1, \lambda_2, \dots, \lambda_n)^t \\
&= \text{cov}(Q_1, Q_2, \dots, Q_n) \cdot (\lambda_1, \lambda_2, \dots, \lambda_n)^t \\
&= \Sigma_{\mathbf{Q}} \cdot \boldsymbol{\lambda}
\end{aligned}$$

we have

$$\boldsymbol{\lambda} = \Sigma_{\mathbf{Q}}^{-1} \sigma_{\mathbf{p}^* \mathbf{q}}$$

provided that the covariance matrix $\Sigma_{\mathbf{Q}}$ is invertible.

Suppose we have m samples in the training set, consider the regression model

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{B} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{Y} = \begin{pmatrix} p_1^* - p \\ p_2^* - p \\ \vdots \\ p_m^* - p \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & q_{11} - \mu_1 & \dots & q_{1n} - \mu_n \\ 1 & q_{21} - \mu_1 & \dots & q_{2n} - \mu_n \\ \vdots & \vdots & \dots & \vdots \\ 1 & q_{m1} - \mu_1 & \dots & q_{mn} - \mu_n \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix},$$

$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)^t$ denotes the random errors. As long as the experts are not linearly dependent, the least squares estimation of the regression coefficients is

$$\hat{\mathbf{B}} = (\mathbf{X}^t \mathbf{X})^{-1} (\mathbf{X}^t \mathbf{Y}),$$

which can be written in its matrix form

$$\begin{aligned}
\hat{\mathbf{B}} &= \left[\begin{pmatrix} 1 & 1 & \dots & 1 \\ q_{11} - \mu_1 & q_{21} - \mu_1 & \dots & q_{m1} - \mu_1 \\ q_{12} - \mu_2 & q_{22} - \mu_2 & \dots & q_{m2} - \mu_2 \\ \vdots & \vdots & \dots & \vdots \\ q_{1n} - \mu_n & q_{2n} - \mu_n & \dots & q_{mn} - \mu_n \end{pmatrix} \cdot \begin{pmatrix} 1 & q_{11} - \mu_1 & \dots & q_{1n} - \mu_n \\ 1 & q_{21} - \mu_1 & \dots & q_{2n} - \mu_n \\ \vdots & \vdots & \dots & \vdots \\ 1 & q_{m1} - \mu_1 & \dots & q_{mn} - \mu_n \end{pmatrix} \right]^{-1} \cdot \begin{pmatrix} 1 & \dots & 1 \\ q_{11} - \mu_1 & \dots & q_{m1} - \mu_1 \\ q_{12} - \mu_2 & \dots & q_{m2} - \mu_2 \\ \vdots & \dots & \vdots \\ q_{1n} - \mu_n & \dots & q_{mn} - \mu_n \end{pmatrix} \cdot \begin{pmatrix} p_1^* - p \\ p_2^* - p \\ \vdots \\ p_m^* - p \end{pmatrix} \\
&= \begin{pmatrix} m & m(\overline{Q_1} - \mu_1) & \dots & m(\overline{Q_n} - \mu_n) \\ m(\overline{Q_1} - \mu_1) & \sum_{i=1}^m (q_{i1} - \mu_1)^2 & \dots & \sum_{i=1}^m (q_{i1} - \mu_1)(q_{in} - \mu_n) \\ m(\overline{Q_2} - \mu_2) & \sum_{i=1}^m (q_{i2} - \mu_2)(q_{i1} - \mu_1) & \dots & \sum_{i=1}^m (q_{i2} - \mu_2)(q_{in} - \mu_n) \\ \vdots & \vdots & \dots & \vdots \\ m(\overline{Q_n} - \mu_n) & \sum_{i=1}^m (q_{in} - \mu_n)(q_{i1} - \mu_1) & \dots & \sum_{i=1}^m (q_{in} - \mu_n)^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} m\overline{p^*} - mp \\ \sum_{i=1}^m (p_i^* - p)(q_{i1} - \mu_1) \\ \sum_{i=1}^m (p_i^* - p)(q_{i2} - \mu_2) \\ \vdots \\ \sum_{i=1}^m (p_i^* - p)(q_{in} - \mu_n) \end{pmatrix}.
\end{aligned}$$

When the sample size is large enough, the sample mean is approximately equal to the total expectation, we have

$$\begin{aligned}\hat{\mathbf{B}} &= \frac{1}{m} \cdot \begin{pmatrix} 1 & E(Q_1) - \mu_1 & E(Q_2) - \mu_2 & \cdots & E(Q_n) - \mu_n \\ E(Q_1) - \mu_1 & E[(Q_1 - \mu_1)^2] & & & \\ E(Q_2) - \mu_2 & E[(Q_2 - \mu_2)(Q_1 - \mu_1)] & E[(Q_2 - \mu_2)(Q_n - \mu_n)] & & \\ \vdots & & & & \\ E(Q_n) - \mu_n & E[(Q_n - \mu_n)(Q_1 - \mu_1)] & \cdots & E[(Q_n - \mu_n)^2] \end{pmatrix}^{-1} \cdot m \cdot \begin{pmatrix} E(p^*) - p \\ E[(p^* - p)(Q_1 - \mu_1)] \\ E[(p^* - p)(Q_2 - \mu_2)] \\ \vdots \\ E[(p^* - p)(Q_n - \mu_n)] \end{pmatrix} \\ &= \frac{1}{m} \cdot \begin{pmatrix} 1 & E(Q_1) - \mu_1 & E(Q_2) - \mu_2 & \cdots & E(Q_n) - \mu_n \\ E(Q_1) - \mu_1 & & & & \\ E(Q_2) - \mu_2 & & \text{cov}(Q_1, Q_2, \dots, Q_n) & & \\ \vdots & & & & \\ E(Q_n) - \mu_n & & & & \end{pmatrix}^{-1} \cdot m \cdot \begin{pmatrix} E(p^*) - p \\ \sigma_{p^*Q} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \Sigma_Q^{-1} \cdot \sigma_{p^*Q} \end{pmatrix}_{(n+1) \times 1}.\end{aligned}$$

Therefore, the parameters λ in LBU model can be interpreted as the linear regression coefficients $\mathbf{B}$ of $p^* - p$ on $\mathbf{Q} - \boldsymbol{\mu}$.

4. Case study

To investigate the performance of LBU method in real cases, the lithology types in the well-known Swiss Jura data set [20] are used, in which four rock types are sampled in a 14.5 km^2 area and 259 samples are selected as training data set (Figure 1). The class proportions of these four categories are [20.46%, 32.82%, 24.32%, 22.39%], respectively. We consider 10 nearest

Figure 1. Jura lithology data set with four classes.

samples as the neighbors of target locations, thus we have 10 experts for consultation. Correlation analysis (Figure 2) shows that the 10 experts are mutual correlated to some extent. The two nearest neighbors have the highest correlation coefficient (greater than 0.9). Therefore, a variable-reduction

technique is needed. We use principal component analysis (PCA) to convert those correlated experts into a set of values of linearly uncorrelated principal components. Before doing this, the well-known measures of appropriateness of factor analysis, KMO and Bartlett's test, have been performed. Table 1 shows several important parts of the output: the Kaiser-Meyer-Olkin measure of

Table 1. KMO and Bartlett's test.

	KMO measure of sampling adequacy	Bartlett's test of sphericity		
		Chi-Square	Degree of freedom	Significance
Argovian	0.911	3935.104	45	<0.0001
Kimmeridgian	0.909	2996.234	45	<0.0001
Sequanian	0.855	2482.901	45	<0.0001
Quaternary	0.850	2951.061	45	<0.0001

Figure 2. Correlation analysis of the 10 experts for each lithology type: Argovian (upper-left), Kimmeridgian (upper-right), Sequanian (lower-left) and Quaternary (lower-right).

sampling adequacy and Bartlett's test of sphericity. The KMO statistic of each lithology type is greater than 0.5, so we should be confident that PCA is appropriate for these data. Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix. For these data, Bartlett's test is highly significant ($p < 0.0001$), and therefore PCA is appropriate. Combined, components 1 and 2 account for 82.7% and 80.9% of the total variance for Argovian and Kimmeridgian, and the first three components account 81.5% and 83.2% for the last two lithology types (Table 2). Corresponding factor loadings are presented in Table 3. The first component of

Table 2. Accumulative contribution rate of the 10 principle components. Prin1 is short for the first principle component.

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10
Argovian	0.726	0.827	0.912	0.941	0.961	0.975	0.987	0.996	0.999	1.000
Kimmeridgian	0.610	0.809	0.872	0.906	0.934	0.958	0.977	0.992	0.997	1.000
Sequanian	0.504	0.702	0.815	0.870	0.913	0.944	0.974	0.993	0.997	1.000
Quaternary	0.501	0.685	0.832	0.880	0.920	0.955	0.979	0.997	0.999	1.000

Table 3. Factor loadings of the retained principle components.

	Argovian		Kimmeridgian		Sequanian			Quaternary		
	Prin1	Prin2	Prin1	Prin2	Prin1	Prin2	Prin3	Prin1	Prin2	Prin3
Q_1	0.335	-0.356	0.321	-0.402	0.367	-0.355	-0.153	0.381	-0.340	0.120
Q_2	0.335	-0.371	0.320	-0.406	0.367	-0.358	-0.143	0.379	-0.344	0.127
Q_3	0.336	-0.356	0.320	-0.407	0.373	-0.355	-0.122	0.378	-0.349	0.138
Q_4	0.323	-0.244	0.308	-0.309	0.351	-0.154	0.023	0.319	-0.231	-0.015
Q_5	0.319	0.133	0.322	0.214	0.309	0.350	-0.163	0.333	0.310	-0.242
Q_6	0.337	0.109	0.350	0.204	0.337	0.341	-0.093	0.369	0.213	-0.218
Q_7	0.330	0.144	0.329	0.272	0.287	0.394	-0.198	0.340	0.330	-0.181
Q_8	0.325	0.250	0.328	0.285	0.275	0.437	0.027	0.292	0.392	-0.118
Q_9	0.260	0.387	0.307	0.285	0.263	-0.079	0.533	0.110	0.379	0.561
Q_{10}	0.248	0.536	0.247	0.298	0.180	0.067	0.762	0.077	0.208	0.695

Argovian, for example, is a comprehensive indicator reflected by expert 1 to 8, the eight nearest neighbors. The second component can thus be regarded as a indicator mainly explained by the last two experts.

We then perform logistic regression for each lithology type with regard to corresponding retained principal components. The purpose is to obtain the continuous posterior occurrence probability p^* of the binary event A for estimating parameter λ_i . Stepwise selection results (Table 4) show that several principal components have been removed from the logistic regression model. Max-

rescaled R-square provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model. As we know, the parameters λ in LBU model can be interpreted as the vector of coefficients of the linear regression of $p^* - p$ on $Q - \mu$, so long as the experts are not linearly dependent. Firstly, we want to find some statistical indicators to estimate the expectation of expert opinions. Sample means can be good approximation when the data accord with normal distribution. If the data distribution is skewed, sample medians rather than means will be chosen as the estimator of μ . Table 5 displays the descriptive statistics of transition probabilities (expert opinions). Jarque-Bera test is used for goodness-of-fit to a normal distribution. We reject the hypothesis if the test is significant at the 5% level. The normal probability plot (Figure 3) shows that, with a p-value of 0.002, we can reject the hypothesis that the distribution of P_{11} , for example, is normal. Therefore, the median 0.201 is used as the estimated value of corresponding expert opinion's expectation. We then use the global class proportions of these four categories as the prior probabilities p , the parameters λ is the coefficients

Table 4. Stepwise selection results of the logistic regression model.

	Intercept	Prin1	Prin2	Max-rescaled R-square
Argovian	-5.361	4.328	-2.683	0.849
Kimmeridgian	-3.662	2.924	-3.244	0.729
Sequanian	-3.959	3.263	-3.053	0.699
Quaternary	-6.770	6.667	-	0.792

Table 5. Descriptive statistics of transition probabilities (expert opinions).

Transition probability	Mean	Median	Maximum	Minimum	Standard deviation
P_{11}	0.270	0.201	1.000	0.000	0.250
P_{12}	0.290	0.292	0.632	0.000	0.173
P_{13}	0.254	0.239	0.769	0.000	0.140
P_{14}	0.186	0.193	0.436	0.000	0.123
P_{21}	0.227	0.239	0.563	0.000	0.134
P_{22}	0.332	0.282	1.000	0.091	0.198
P_{23}	0.254	0.256	0.436	0.000	0.090
P_{24}	0.187	0.192	0.387	0.000	0.088
P_{31}	0.220	0.200	0.636	0.000	0.141
P_{32}	0.313	0.309	1.000	0.000	0.155
P_{33}	0.294	0.262	1.000	0.000	0.203
P_{34}	0.173	0.185	0.469	0.000	0.119
P_{41}	0.219	0.182	0.720	0.000	0.154
P_{42}	0.312	0.318	0.562	0.000	0.142
P_{43}	0.256	0.238	0.586	0.000	0.123
P_{44}	0.213	0.184	1.000	0.000	0.208

of stepwise regression in LBU model. All the estimated λ_i are greater than zero, which means that Q_i is more likely to be high (and less likely to be low) when $A = 1$ than when $A = 0$.

We also use MCRF method for comparison. As a non-linear and parameter-free method, the MCRF is based on the product of probabilities, which will dramatically amplify higher probability and reduce lower probability close to zero. The lithology types of corresponding 259 locations can be predicted based on the maximum a posterior (MAP) probability criterion (Figure 4). Together with Table 6, we find that LBU is superior to MCRF in overall prediction accuracy by comparing the prediction results of MCRF with LBU model.

5. Conclusions

In this work, we consummate the theoretical foundations of LBU method for the prediction of categorical spatial data. The philosophy underlying this approach is originated from combining probability distributions in expert system. The Bayesian argument embodied in LBU is both normative and logical. With a linear expression, the probabilities of class occurrence of target locations can be assessed by neighboring sample types. Weights in this model are interpreted as linear regression coefficients. We have enriched our previous findings [19] by adding some rigorous theoretical proofs of the LBU method. Experimental results show that our method performs better

Figure 3. Normal probability plot and Jarque-Bera test for goodness-of-fit to a normal distribution.

Table 6. Prediction accuracy comparison.

	Argovian	Kimmeridgian	Sequanian	Quaternary	overall
LBU	86.79% (46/53)	85.88% (73/85)	73.02% (46/63)	84.48% (49/58)	82.63% (214/259)
MCRF	88.68% (47/53)	78.82% (67/85)	77.78% (49/63)	82.76% (48/58)	81.47% (211/259)

in terms of classification accuracy compared with traditional MCRF method. As pointed out by [19], our method can also be generalized to nonlinear cases, where more confident probability forecasting results can be obtained.

Acknowledgments: This study is supported by the Fundamental Research Funds for the Central Universities of Central South University (No. 2016zzts011). The authors are indebted to Ying Chen, Danhua Chen, Ruizhi Zhang and Wuyue Shen for their critical reviews of the paper.

Author Contributions: Xiang Huang and Zhizhong Wang conceived and designed the experiments; Xiang Huang performed the experiments; Zhizhong Wang analyzed the data; Xiang Huang wrote the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Figure 4. Predicted lithology types of corresponding 259 locations based on MAP probability criterion.

References

1. Cao, G.; Kyriakidis, P.C.; Goodchild, M.F. Combining spatial transition probabilities for stochastic simulation of categorical fields. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1773-1791.
2. Deutch, C.V.; Journel, A.G. *GSLIB: geostatistical software library and user's guide*. Oxford University Press, New York, **1992**.
3. Carle, S.F.; Fogg, G.E. Transition probability-based indicator geostatistics. *Math. Geol.* **1996**, *28*, 453-476.
4. Li, W. Transiogram: A spatial relationship measure for categorical data. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 693-699.
5. Journel, A. Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. *Math. Geol.* **2002**, *34*, 573-596.
6. Krishnan, S. The tau model for data redundancy and information combination in earth sciences: theory and application. *Math. Geosci.* **2008**, *40*, 705-727.
7. Polyakova, E.I.; Journel, A.G. The Nu expression for probabilistic data integration. *Math. Geol.* **2007**, *39*, 715-733.
8. Cao, G.; Kyriakidis, P.C.; Goodchild, M.F. A multinomial logistic mixed model for the prediction of categorical spatial data. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 2071-2086.
9. Cao, G.; Yoo, E.-H.; Wang, S. A statistical framework of data fusion for spatial prediction of categorical variables. *Stoch. Env. Res. Risk. A.* **2014**, *28*, 1785-1799.
10. Schaeben, H. A mathematical view of weights-of-evidence, conditional independence, and logistic regression in terms of Markov random fields. *Math. Geosci.* **2014**, *46*, 691-709.
11. Besag, J. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc. B* **1974**, *2*, 192-236.
12. Norberg, T.; Rosén, L.; Baran, A.; Baran, S. On modelling discrete geological structures as Markov random fields. *Math. Geol.* **2002**, *34*, 63-77.
13. Ricotta, C.; Carranza, M.L. Measuring scale-dependent landscape structure with Rao's quadratic diversity. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 405-412.
14. Huang, X.; Li, J.; Liang, Y.; Wang, Z.; Guo, J.; Jiao, P. Spatial hidden Markov chain models for estimation of petroleum reservoir categorical variables. *J. Petrol. Explor. Prod. Technol.* **2016**, doi: 10.1007/s13202-016-0251-9.
15. Li, W. Markov chain random fields for estimation of categorical variables. *Math. Geol.* **2007**, *39*, 321-335.
16. Li, W.; Zhang, C.; Willig, M.R.; Dey, D.K.; Wang, G.; You, L. Bayesian Markov Chain Random Field Cosimulation for Improving Land Cover Classification Accuracy. *Math. Geosci.* **2015**, *47*, 123-148.
17. Allard, D.; Comunian, A.; Renard, P. Probability aggregation methods in geoscience. *Math. Geosci.* **2012**, *44*, 545-581.
18. Genest, C.; Schervish, M.J. Modeling expert judgments for Bayesian updating. *Ann. Stat.* **1985**, *13*, 1198-1212.
19. Huang, X.; Wang, Z.; Guo, J. Prediction of categorical spatial data via Bayesian updating. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1426-1449.
20. Goovaerts, P. *Geostatistics for natural resources evaluation*. Oxford university press, New York, **1997**.
21. Stone, M. The opinion pool. *Ann. Math. Stat.* **1961**, *32*, 1339-1342.

22. French, S. Consensus of opinion. *Eur. J. Oper. Res.* **1981**, *7*, 332-340.
23. Winkler, R.L. The consensus of subjective probability distributions. *Manage. Sci.* **1968**, *15*, B-61-B-75.
24. Heskes, T. Selecting weighting factors in logarithmic opinion pools, in: Jordan, M., Kearns, M., Solla, S. (Eds.), *Advances in neural information processing systems*. MIT Press, Cambridge, **1998**, pp. 266-272.
25. Cao, G.; Kyriakidis, P.; Goodchild, M. Prediction and simulation in categorical fields: a transition probability combination approach, *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, New York, **2009**, pp. 496-499.

© 2016 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).