Statistics and Probability Letters 143 (2018) 37–42

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Best look-alike prediction: Another look at the Bayesian classifier and beyond

Hanmei Sun [a], Jiming Jiang [b,*], Thuan Nguyen [c], Yihui Luan [a]

[a] *Shandong University, China*
[b] *University of California, Davis, USA*
[c] *Oregon Health & Science University, USA*

A R T I C L E   I N F O

Article history:
Received 23 September 2017
Received in revised form 13 July 2018
Accepted 17 July 2018
Available online 1 August 2018

Keywords:
BLAP
Categorical outcome
Mixed logistic model
Zero-inflated random variable
Small area estimation

A B S T R A C T

A criterion of optimality in prediction is proposed that requires the predictor to assume the same type of values as the random variable it is predicting. In the case of categorical responses, the method leads to the Bayesian classifier with a uniform prior. However, the method extends to other cases, such as zero-inflated observations, as well. The method, called best look-alike prediction (BLAP), justifies an "usual practice" from a theoretical standpoint. Application of BLAP to small area estimation is considered. A real-data example is discussed.

1. Introduction

Many practical problems are related to prediction, where the main interest is at subject (e.g., precision medicine) or sub-population (e.g., small area estimation) level. The traditional concept of best prediction (BP) is in terms of mean squared prediction error (MSPE). Under such a framework, the best predictor (BP) is known as the conditional expectation of the random variable to be predicted, say, α, given the observed data, say, Y, that is, $\mathrm{E}(\alpha|Y)$. Based on the BP, a number of more specialized prediction methods have been developed, such as best linear prediction (BLP) and best linear unbiased prediction (BLUP). See, for example, sec. 2.3 of Jiang (2007). In particular, mixed model prediction, that is, prediction based on mixed effects models, has a fairly long history starting with Henderson's early work in animal breeding (Henderson, 1948). See, for example, Robinson (1991), Jiang and Lahiri (2006), Jiang (2007, sec. 2.3), and Rao and Molina (2015).

In spite of its dominance in prediction theory, and overwhelming impact in practice, the BP can have a very different look than the random variable it is trying to predict. This is particularly the case when the random variable is discrete, categorical, or has some features related to a discrete or categorical random variable. For example, if α is a binary random variable taking the values 1 or 0, its BP, $\mathrm{E}(\alpha|Y)$, is typically not equal to 1 or 0; instead, the value of $\mathrm{E}(\alpha|Y)$ usually lies strictly between 0 and 1. Such a feature of the BP is sometimes unpleasant, or inconvenient, for a practitioner because the values 1 and 0 correspond directly to outcomes of scientific, social, or economic interest; there is no such an outcome that corresponds to, say, 0.35, or at least not directly.

Let α be an unobserved, possibly vector-valued random variable for which we wish to predict. The prediction will be based on the observed data, denoted by Y. A predictor, say $\tilde{\alpha}$, is said to be *look-alike* with respect to α if it has the same set

* Corresponding author.
E-mail address: jimjiang@ucdavis.edu (J. Jiang).

https://doi.org/10.1016/j.spl.2018.07.014

of possible values as α. We derive the best predictor under this framework and a suitable criterion of optimality, which is different than the BP. We refer the method as best look-alike prediction, or BLAP.

It should be noted that, in the case of predicting categorical outcomes, BLAP leads to the same solution as what is known as the Bayesian classifier under a uniform prior (e.g., Nurty and Devi, 2011). However, under our framework there is no prior; instead, the unknown parameters involved in BLAP are estimated from the data, leading to the empirical BLAP, or EBLAP. Furthermore, the BLAP method applies to other situations, such as zero-inflated observations, for which there is no Bayesian classifier.

The derivation of BLAP is given in Section 2 for the case of prediction of a discrete or categorical random variables. In Section 3, we derive BLAP for zero-inflated random variables. In Section 4 we consider an application of BLAP to small area estimation (e.g., Rao and Molina, 2015) with zero-inflated random effects. Some real-data applications are discussed in Section 5. Further details and results can be found in an online supplement.

2. BLAP for discrete/categorical random variable

Let α denote a discrete or categorical random variable (r.v.) that we wish to predict. Without loss of generality, we can assume that α is a discrete r.v. whose values are nonnegative integers. Let S denote the set of possible values of α. Let $\tilde{\alpha}$ be a predictor of α based on the observed data, Y. $\tilde{\alpha}$, is look-alike (with respect to α) if it also has S as its set of possible values. The performance of $\tilde{\alpha}$ is measured by the probability of mismatch:

$$p(\tilde{\alpha} \neq \alpha) = \sum_{k \in S} P(\tilde{\alpha} = k, \alpha \neq k). \tag{1}$$

$\tilde{\alpha}$ is said to be the best look-alike predictor, or BLAP, if it minimizes the probability of mismatch, (1). The following theorem defines a BLAP.

Theorem 1. *A BLAP of α is given by*

$$\tilde{\alpha}^* = \min\left\{ i \in S : P(\alpha = i|Y) = \max_{k \in S} P(\alpha = k|Y) \right\}, \tag{2}$$

provided that the right side of (2) is computable.

Note. In the context of BP, it is well known that the expression of BP is very similar to the Bayesian posterior mean. Similarly, the expression of BLAP in this case, that is, (2), resembles that of the Bayesian classifier under a uniform prior (e.g., Nurty and Devi, 2011). Of course, there is no prior distribution in our consideration.

Proof of Theorem 1. Note that the right side of (1) can be expressed as

$$\begin{aligned} \sum_{k \in S} E\left\{ 1_{(\tilde{\alpha}=k)} P(\alpha \neq k|Y) \right\} &= \sum_{k \in S} E\left[1_{(\tilde{\alpha}=k)} \{1 - P(\alpha = k|Y)\} \right] \\ &\geq \sum_{k \in S} E\left[1_{(\tilde{\alpha}=k)} \left\{ 1 - \max_{k \in S} P(\alpha = k|Y) \right\} \right] \\ &= \sum_{k \in S} E\left[1_{(\tilde{\alpha}^*=k)} \{1 - P(\alpha = k|Y)\} \right] \\ &= P(\tilde{\alpha}^* \neq \alpha). \end{aligned} \tag{3}$$

The second-to-last equation in (3) is because, when $\tilde{\alpha}^* = k$, one has, by the definition, $P(\alpha = k|Y) = \max_{k \in S} P(\alpha = k|Y)$; the last equation in (3) is, again, due to (1) and the first equation in (3). This completes the proof.

A special case of Theorem 1 is the binary case, which we state as a corollary.

Corollary 1. *Suppose that δ is a binary r.v. taking the values 1 and 0. Then, the BLAP of δ is given by $\tilde{\delta}^* = 1_{\{P(\delta=1|Y) \geq 1/2\}}$, provided that $P(\delta = 1|Y)$ is computable.*

Typically, the conditional probability, $P(\alpha = k|Y)$, depends on some unknown parameters, say, θ. It is customary to replace the θ by $\hat{\theta}$, a consistent estimator of θ, on the right side of (2). The result is called an empirical BLAP, or EBLAP, denoted by $\hat{\alpha}^*$.

3. BLAP of a zero-inflated random variable

A zero-inflated random variable, α, has a mixture distribution with one mixture component being 0 and the other mixture component being an a.s. nonzero random variable. Suppose that $\alpha = \delta\xi$, where δ is a binary random variable such that