Computers and Electrical Engineering 000 (2017) 1–12

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

A Bayesian assumption based forecasting probability distribution model for small samples☆

Zonglei Lu[a,b,*], Xiaohan Geng[a], Guoming Chen[b]

[a] Department of Computer Science & Technology, Civil Aviation University of China, Tianjin, China
[b] Information Technology Research Base of Civil Aviation Administration of China, Tianjin, China

A R T I C L E I N F O

Article history:
Received 15 July 2017
Revised 20 November 2017
Accepted 20 November 2017
Available online xxx

Keywords:
Small sample size problem
Probability distribution forecasting
Naïve Bayes
Bayesian network

A B S T R A C T

In this work, a novel forecasting probability distribution model is presented. Probability distribution plays a role in the function of probability values. Therefore, forecasting the probability distribution function is a challenging process. To that end, the method described in this work loosens the control conditions of the given data set. Subsequently, statistical methods can be applied to the resulting sample data. The distribution functions are then fitted using the cubic spline interpolation method. In this work, the naive Bayes and the Bayesian network methods are adjusted to handle the small sample problem. In addition, the maximal extension clusters are used to determine the conditional function. Two data sets from the UCI repository and a custom data set are used to validate the forecasting model. The experiments show the proposed method can generate an accurate distribution function.

1. Introduction

Estimating the probability distribution function is a classical machine learning problem. Probability distribution is a useful tool for describing random variables because a single probability value does not adequately describe the variable. Probability distribution is a mathematical description of a random phenomenon. The base of this description is the probability of events. Thus, probability distribution requires a more complicated model than the probability of events. Moreover, probability distribution has been widely applied in many fields, such as lightning current amplitude [1], wind speed [2], and multimeric systems [3]. Obtaining the probability distribution function of a large sample set is a relatively easy task. However, in the case of a small sample set, this traditional method may not be applicable.

The small sample size problem (SSSP) is a hot topic in current academic research. For example, in tasks pertaining face recognition [4] or speech emotion [5], the lack of large sample size is a challenge. To that end, loosen control condition (LCC) [6] has been applied as valid method for a small sample set. Virtual sample generation (VSG) [7] has also been used to address the SSSP. Li Der-Chiang proposed a genetic algorithm based on virtual sample generation, derived from LCC and VSG [8]. In [9], Zhang Cui-Cui proposed an ensemble framework to generate new data from the distribution of the original samples.

The bootstrap method has been widely used to address the SSSP. First, several bootstrap samples are generated by resampling the original data set. Then, the probability distribution of each bootstrap sample is calculated, thereby enabling the

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. Huimin Lu.

* Corresponding author.
E-mail addresses: zllv@cauc.edu.cn (Z. Lu), xhgengyjs_16@cauc.edu.cn (X. Geng), gmchenyjs_13@cauc.edu.cn (G. Chen).

https://doi.org/10.1016/j.compeleceng.2017.11.025
0045-7906/

estimation of the probability distribution of the original dataset. Thus, the bootstrap method can be utilized for estimating the parameters in the case of small samples.

The conventional small sample learning methods mainly focus on a possible value of the target attribute; however, this value is only related to the probability of events. Even so, forecasting probability distribution of in the case of small samples is a challenging problem. To that end, the proposed model, which is based on LCC and the Bayesian learning model, attempts to address this challenge.

The remainder of this paper is organized as follows. Related work is reviewed in Section 2. The proposed method is introduced in Section 3. Experiments and results of the experimentation are described in Section 4. The paper is concluded and the findings are presented in Section 5.

2. Basic concepts and principles

Probability is a measure of the likelihood of an event occurring, and as such, there are two different theoretical explanations about probability. The first is a frequentist view that defines probability as an objective concept. The probability of an event is the limiting proportion of times that the event occurs from a long series of independent identical opportunities. The second explanation is the Bayesian probability view (also called subjective probability view), wherein the probability is regarded as an inner state rather than an objective property of the outside world [10]. The probability only denotes the degree of personal beliefs that the event occurred.

According to the Bayesian view, all probabilities are conditional probabilities. The definition of conditions is related to subjective recognition. Moreover, different conditions are likely to result in different outcomes. Thus, several different schemes can be generated to select instances from the data set. Consider a data set of seismic records from all regions of the world. The task is to obtain the distribution of intermediate-focus earthquakes in a certain region. Therefore, the sample should consist of seismic records from the same location. However, the probability of finding an earthquake instance that matches the exact condition is very low. The probability improves, however, when the definition of the location extends to a specific area surrounding the specific location. If the definition extends to across the world, it is equivalent to encompassing all the data sets. As the definition extends, the set contains seismic data with different characteristics. Therefore, the extension should be confined to meet the statistical requirements.

According to the Bayesian probability view, the sample is the result of selection. The sample set can be extended by omitting some selected attribute value. Let us consider a universal set, and the sample set is just a subset of the universal set. The generation of the samples mainly depends on the selection criteria. Thus, a small sample may be generated by a highly strict selection criterion. In this work, *selection* [11] is defined as follows:

Definition 1. Selection produces a horizontal subset of a given data set D, which consists of all the instances that satisfy the condition set C. The selection by C from D is denoted as $\sigma_{\hat{C}}(D) = \{d | d \in D \wedge \hat{C}(d) = T\}$, where D is a subset of the Cartesian product of conditional attributes and a target attribute, namely $D = X \times Y$, $X = X_1 \times X_2 \times \ldots \times X_k$ and $\hat{C}$ denotes a selection condition generated by the conjunction of conditions in C.

Here, C is a group of constraints on X. Therefore, $\hat{C}(\cdot)$ is a logical expression with a value of "T" or "F". For example, if $C = \{A_1 = a, A_2 = b, A_3 = c\}$, then $\hat{C}(d) = (d.A_1 = a) \wedge (d.A_2 = b) \wedge (d.A_3 = c)$.

Theorem 1. *If* $D_1 = \sigma_{\hat{C}_1}(D)$, $D_2 = \sigma_{\hat{C}_2}(D)$ *and* $C_1 \subseteq C_2$, *then* $D_2 \subseteq D_1$.

Proof. 1) If $C_1 = C_2$, then we establish that $D_1 = D_2$.
2) If $C_1 \subset C_2$, and because $D_2 = \sigma_{\hat{C}_2}(D)$,
We have for each $d \in D_2$, $\hat{C}_2(d) = T$.
Moreover, given that $C_1 \subset C_2$, we have $\hat{C}_1(d) = T$.
Thus, $d \in D_1$ is established.
Therefore, $D_2 \subseteq D_1$.

Corollary 1. *If* $D_1 = \sigma_{\hat{C}_1}(D)$, $D_2 = \sigma_{\hat{C}_2}(D)$ *and* $C_1 \subseteq C_2$, *then* $|D_1| \geq |D_2|$.

Here, Corollary 1 shows that when the conditions become more stringent, the number of data will be reduced, and vice versa. Therefore, the data set may be expanded or shrunk by adjusting the condition set.

Expansion of the data set is a conventional method applied to solving the SSSP. Therefore, the approach utilized to expand the data set, i.e., to loosen the conditions, is significant. Each constraint in the given condition set can generate a new single-constraint condition set. Moreover, each single-constraint condition can generate a conditional data set. Thus, the generated dataset may have more instances than the original one. Therefore, for sufficient instances, there may exist a function that fits the single-constraint conditions. Moreover, these conditional functions contain the complete information about the final distribution function.

Parameter learning has been commonly applied to predict the conditional distribution. Parameter learning depends on the prior distribution type in the corresponding field. Unfortunately, in some cases, there is little information about the prior distribution type. According to the analytic geometry, the process of forecasting distribution is a process of fitting the plane curve. Therefore, the interpolation function can fit the distribution. Moreover, the cubic spline interpolation function