Method for forecasting the volume of applications using probabilistic classifier based on Bayesian theorem for recruitment in the government organizations

Sunil Kumar Chhillar
Centre for Development of Advanced Computing,
Mohali City, Punjab India
E-mail: sunilkc@cdac.in

Baljit Singh Khehra
Baba Banda Singh Bahadhur Engineering College,
Fatehgarh Sahib City, Punjab India
E-mail: Baljit.singh@bbsbec.ac.in

Abstract — In this paper, we propose a practical method to estimate or forecast the volume of applications based on Bayesian classifier which is a probability method that makes optimal decision based on known probability distribution and recently observed data.

By using the Bays estimate method, which selects spectrum analysis, the weight of forecasting model is obtained by carefully analyzing the sample space (various recruitments in the government organizations of Punjab, India). The sample calculation shows that the proposed method is highly reliable and improves the precision with a prior and likelihoods post probabilities of defined sample space. i.e. (i) development of an approximate distribution of likely volume of applications, (ii) prior probabilities of the volume of applications in various mutually exclusive events and (iii) likelihoods relating the confidence of forecast results for obtaining the Bayesian forecast of volume of applications.

Earlier the forecasts are available, better will be their utility for government i.e. provides valuable information with regard to procedure fixation, improve service delivery, reduce costs, redefine administrative processes.

Keywords- Bays theorem; Spectrum analysis

I. INTRODUCTION

True Bayesians actually consider conditional probabilities as more basic than joint probabilities. It is easy to define P (A|B) without reference to the joint probability P (A, B). To see this note that we can rearrange the conditional probability formula to get:

P (A|B) P (B) = P (A, B)

By symmetry: P (B|A) P (A) = P (A, B)

It follows that:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

A. Bayes' theorem.

Let A1, A2, ... , An be a set of mutually exclusive events that together form the sample space S. Let B be any event from the same sample space, such that P(B) > 0. Then,

$$P(Ak | B) = P(Ak \cap B) / P(A1 \cap B) + P(A2 \cap B) + . . . + P(An \cap B)$$

Note: Invoking the fact that P(Ak ∩ B) = P(Ak)P(B | Ak), Baye's theorem can also be expressed as

$$P(Ak | B) = P(Ak) P(B | Ak) / P(A1) P(B | A1) + P(A2) P(B | A2) + . . . + P(An) P(B | An)$$

B. Forecasting Bayeian Probabilty

Model based forecasts arise from rule of model which formalizes relationship between variables of interest. Subjective forecasts are based on guesses, judgment, experience or institution and accumulated knowledge. They do not follow clear rules and depends on processing information in an informal way i.e usually inside someone's head. These forecasts are not necessarily inaccurate since in some situations experience can be the best source of information.

For forecasting the volume of applications based on Bayesian theorem involves the collection of expert opinion data of organizations that are actually engaged in recruitment or data collection and processing for recruitments in the government organizations about the likely results. In this connection, the data collection was conducted by carefully analyzing the sample space (various recruitments in the government organizations of Punjab, India) for (i) development of an approximate distribution of likely volume of applications, (ii) prior probabilities of the volume of applications in various mutually exclusive events and (iii) likelihoods relating the confidence of forecast results for obtaining the Bayesian forecast of volume of applications.

II. RESEARCH AND ILLUSTRATION IN BAYESIAN CLASSIFIER

Bayesian network classifier is a typical classification model based on statistical method. It is the theoretical basis for the Baye's theorem, and the priori probability of the events cleverly linked to the posterior probability, using priori information and sample data to determine the posterior probability of events.

The volume of applications diagnosis makes a mapping function from a departments or organizations of a sample space to the type of posts and category of posts or vacancy and its essence is the classification process that determines the types of variables based on variable attributes.

The difference between the different Bayesian network classifier is that it has different ways to strike the value P(Ak | B).

Assuming that all the condition attributes are independent, and the condition attributes have a parent node that also calls class attribute node, this assumption is called condition-independent assumption. The model icon shown in Figure 1：

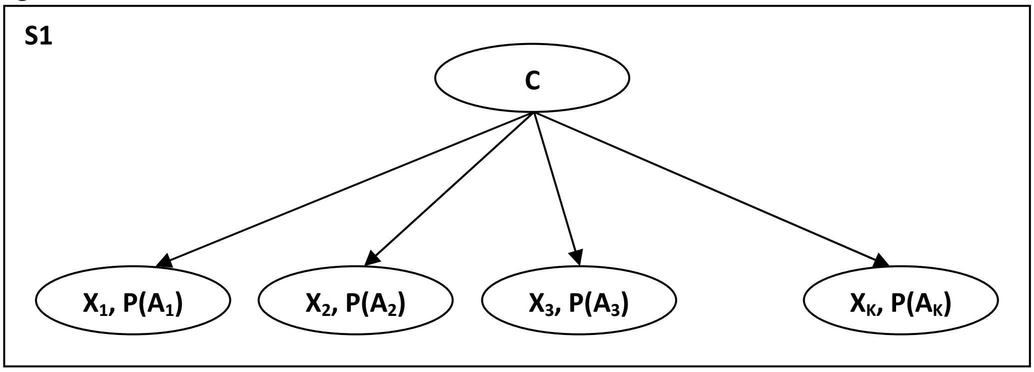

Fig.1 Bayesian Classifier Model

For obtaining the distribution regarding the recruitment, 'n' departments or organizations are selected from the sample space (various recruitments in the government organizations of Punjab, India) as per the sampling design adopted in the sample collection. Based on expert opinion data of organizations that is actually engaged in recruitment or data collection and processing for recruitments in the government organizations about the likely results. The information or data collected in the first round are summarized in various K mutually exclusive classes viz. $<y_1$, $(y_1\text{-}y_2)$, $(y_2\text{-}y_3)$ … … $(y_{k-1}\text{-}y_k)$ where $y_1<y_2<y_3\ldots\ldots<y_k$ are number of applications are received by online e-Form submission techniques regarding recruitments.

These responses are obtained by carefully analyzing such as (i) number of application against type of post, (ii) number of applications against type of department or organization, (iii) number of applications against category of posts etc. and also (i) development of an approximate distribution of likely volume of applications, (ii) prior probabilities of the volume of applications in various mutually exclusive events and (iii) likelihoods relating the confidence of forecast results for obtaining the Bayesian forecast of volume of applications

A. *Application of Baye's theorem for obtaining Bayesian forecast*

The results relating to the prior probabilities of falling likely number of applications in the i^{th} classified group are revised in the light of averages prior forcast using Bayes' theorem. First of all averages prior forcast of volume of application for the current year or department wise is obtained as

$$\overline{y}_c = \sum_{i=1}^{k} p_i y'_i \quad (1)$$

where p_i is the prior probability of obtaining the number of applications as y'_i (the mid value of the number of applications corresponding to the i^{th} class). These values of p_i and y'_i are computed on the basis of the response of current year or recent department data. These revised probabilities are then revised in the light of average prior volume of applications forecast in calculated above formula. The revised probabilities known as posterior or Bayesian probabilities are computed using Bayes' theorem as follows.

$$P(y'_i \mid \bar{y}_c) = \frac{p_i . P(\bar{y}_c \mid y'_i)}{\sum p_i . P(\bar{y}_c \mid y_i)} \quad (2)$$

Where P($\bar{y}_c$ | y'$_i$) is the probability relating to the likelihood that the group would forecast the volume of applications as $\bar{y}_c$ when actually the true number of applications lies in the ith class. These likelihoods are taken from the previous year(s) or previous department or organizations based data corresponding to the class in which the forecast $\bar{y}_c$ lies. The Bayesian probabilities P are then used in obtaining the Bayesian forecast of the volume of applications for specific posts or specific departments/organizations $\bar{y}_b$ and its variance as follows.

$$\bar{y}_B = E(y'_i) = \sum_{i=1}^{k} P(y'_i \mid \bar{y}_c) . y'_i$$

$$V(\bar{y}_B) = E\left[y'_i - E(y'_i)\right]^2 = \sum_{i=1}^{k} P(y'_i \mid \bar{y}_c) . {y'_i}^2 - \bar{y}_B^{\,2} \quad (3)$$

In theory, if it meet the conditions for independence that the variable conditions is the best. However, Bayesian forecast classifier unrealistic conditions for the independence of the restrictions so that the classification performance affected to a certain extent [6].

III. SPECTRUM ANALYSIS[5]

Data P(t) fainted Fourier decomposition. After being time series decomposition, characteristics of a certain frequency executed combination, there are five components:

$$P(t) = a_0 + D(t) + W(t) + L(t) + H(t)$$

In type, a_0 is component, D(t),W(t),L(t),H(t) are the sum of harmonic.

D(t) is component of 24h; a_0 + D (t) is component of day; W(t) is component of week; L(t) are the sum of low frequency components; The cycle is longer than 24h and reflected influence of meteorological factors, H(t) are sum of high frequency.

IV. ANALYSIS UNDER VARIOUS DATA CONDITIONS

The amount of data collected to perform a population variability analysis can be characterized based on the number of subpopulations as well as the exposure level.

TABLE I. VARIOUS DATA COLLECTION OF SAMPLE SPACE S1

Department	Post Type	Year	Post Category	Observation Data
D2	X2	2011	RG	0.6816
D2	X3	2011	RG	1.5069
D1	X1	2011	RG	12.2172
D2	X4	2011	RG	0.1414
D2	X5	2011	RG	0.3821
D2	X6	2011	RG	0.0593
D2	X7	2011	RG	0.1446
D3	X8	2009	CL	2.2235
D3	X9	2009	CL	14.6050
D3	X10	2009	CL	9.2290
D4	X3	2011	RG	6.9579
D5	X3	2011	RG	2.0621
D5	X11	2011	RG	0.6098
D5	X6	2011	RG	0.0028
D6	X12	2011	RG	0.5001
D6	X3	2011	RG	1.0833
D6	X6	2011	RG	1.0576
D6	X12	2011	RG	0.0868
D7	X2	2011	RG	1.5494
D7	X1	2011	RG	6.7018
D7	X12	2011	RG	0.1600
D9	X3	2011	RG	37.0174
D8	X8	2010	RG	0.6088
D9	X13	2011	RG	0.0177
D9	X4	2011	RG	1.9007

A. Development of prior probabilities of the volume of applications in various mutually exclusive events

The estimated average prior probabilities for various classes for sample space S1 is shown in Table-II. Using these prior probabilities, the number of application forecast in year-1 is 26.91%, the number of applications in year – 2 is 43.54% and similarly the number of applications in year-3 is 31.22%.

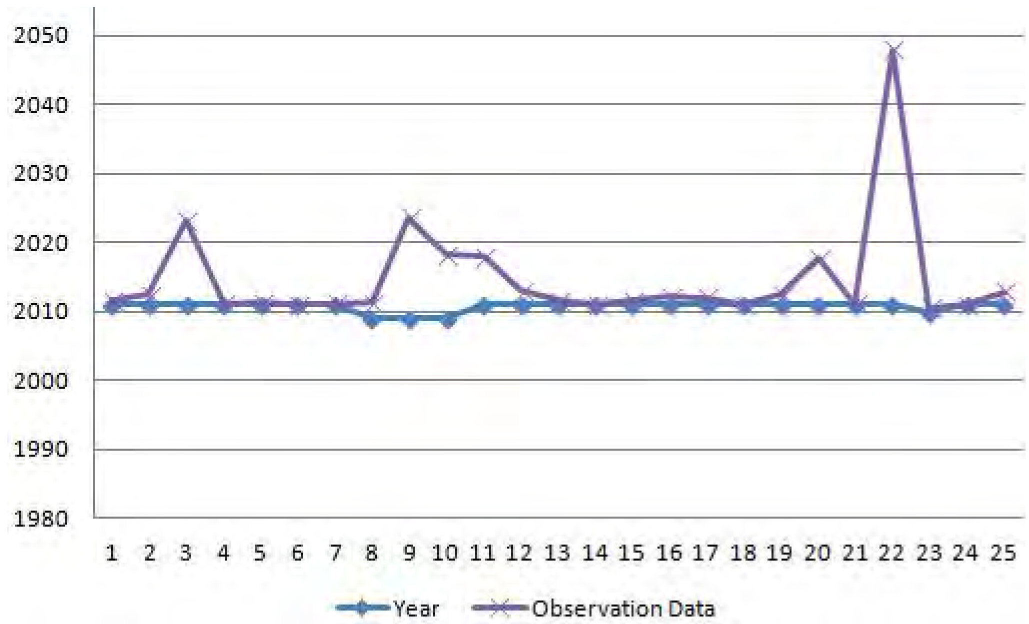

Fig.2. Prior probabilities for obtaining forecast in different number of applications classess

B. Development of likelihoods relating the confidence of forecast results for obtaining the Bayesian forecast of volume of applications

After the recruitment process completion of a particular year i.e. year-1, year-2 and year-3, the data are collected from each department or organizations having same level from sample space S1. The actual volume of applications is classified into various classes such on the basis of post type, post category and department or organizations and number of responses are recorded as per the forecast made earlier for every class of the true value of the results. These numbers of responses are transformed into probabilities which are likelihoods that indicate the accuracy of the group forecasts of the actual results.

C. *Application of Baye's theorem for obtaining Bayesian forecast*

TABLE II. THE BAYESIAN FORECAST OF VOLUME OF APPLICATIONS FOR DIFFERENT YEARS IS SUMMARIZED AS BELOW;

Forecast For	**Using Liklihoods For**	**Forecast (in %age)**	**Actual Results (in %age)**
Year-2	Year-1	35.96%	43.54%
Year-3	Year-1	26.92%	31.22%
Year-3	Year-2	29.04%	31.22%
Year-3	Year-1 & 2	28.33%	31.22%

V. CONCLUSION

This paper put forward the Model by using the Bays estimate method, which selects spectrum analysis, the weight of forecasting model is obtained by carefully analyzing the sample space (various recruitments in the government organizations of Punjab, India). The sample calculation shows that the proposed method is highly reliable and improves the precision with a prior and likelihoods post probabilities of defined sample space. i.e. (i) development of an approximate distribution of likely volume of applications, (ii) prior probabilities of the volume of applications in various mutually exclusive events and (iii) likelihoods relating the confidence of forecast results for obtaining the Bayesian forecast of volume of applications.

So the mind of portfolio diagnosis and the methods in the paper is an effective way to resolve such problems. Hence Earlier the forecasts are available, better will be their utility for government i.e. provides valuable information with regard to procedure fixation, improve service delivery, reduce costs, redefine administrative processes.

REFERENCES

[1] Gong Xiujun. Bayesian learning theory and its application:[PhD thesis].Beijing: Graduate School of the Chinese Academy of Sciences,2002.

[2] Maarten Veger, "How does Internet recruitment have effect on recruitment performance?" Fourth Twente Student Conference on IT, 2006 - referaat.cs.utwente.nl

[3] Roger Munger, "Technical Communicators Beware: The Next Generation of High-Tech Recruiting Methods.", Appears in: Professional Communication, IEEE Transactions, Volume: 45, Issue: 4, Page(s): 276-290, Dec 2002, Type: Journals & Magazines

[4] Sholom M. Weiss, Nitin Indurkhya "Predictive Data Mining: A Practical Guide (The Morgan Kaufmann Series in Data Management Systems)", ISBN 1-55860-478-2.

[5] Mu Gang,Hou Kai-yuan,Yang You-hong.Load forecasting method based on intrinsic error evaluation[J].Automation of Electric Power Systems,2001,25(22):37-40.

[6] Wang Yingming. Based on the relevance of the combination forecasting method[J].2002,21(2):58~ 62

[7] Cheeseman, P., and Stutz, J. 1995. Bayesian Classification (AutoClass): Theory and Results. Knowledge Discovery in Data Bases II. Menlo Park, Calif.: AAAI Press / The MIT Press.