

# *The Content Classification via Bayesian Model in Meteorological Teaching*

Xu Guanlei, Shao Limin  
Ocean Department of Dalian Navy Academy  
Dalian, China, 116018  
xgl\_86@163.com

Liu Yonglu, Lang Fengwang  
Ocean Department of Dalian Navy Academy  
Dalian, China, 116018  
Liuy1@163.com

**Abstract**—In meteorological teaching, in order to improve the teaching efficiency, the content classification is of much signification. After content classification, the teachers can give the detailed and better teaching plans to the students. On the other hand, these classified content can be employed for the students selection of excellent ones. In this paper, we proposed a Bayesian model based meteorological content classification. We classify the meteorological content into three classes. And through the experimental comparison with the students', our method is efficient for the meteorological content automatical classification.

**Keywords**—Meteorological content classification; Meteorological teaching; Bayesian classification

## I. INTRODUCTION

In the classroom, one of the most important things is that the students are very interested in the content taught by the teachers[1]. If this case happens, it means that the teaching efficiency will be better[1]. So how to let the students be interested in the content taught by the teachers is the teachers' research work in teaching. It is especially important for the difficult courses such as synoptic meteorology and meteorological teaching[2]. Therefore, in order to let the courses of synoptic meteorology and meteorological teaching interesting, the first work is to classify the meteorological teaching content to different classes. After classification, the teachers can give the detailed and better teaching plans to the students. On the other hand, these classified content can be employed for the students selection of excellent ones. In addition, the classified content can be used for the optimization of teaching resource allocation.

It is well known that interest is a kind of motivation in studying and teaching [1]. A college student, no matter what to do things, if this thing is he more interested in things, so he must be very willing to do it. If again no tired, difficult, he is willing to do it. Facts show that, interest, makes the student do not fear endures hardship; Interest, make the student consuming; Interest, make the student persistent and willing to put in the effort. Once the students for a certain things became interested in, then they would make every effort to engage in this matter into activity. Therefore, here we will discuss the interesting content and the uninteresting content for college students.

In this paper, we proposed a Bayesian model[3][4][5] based

meteorological content classification method. We classify the meteorological content into three classes. And finally, the experimental comparison with the students' is made to show our method's efficiency.

## II. BAYESIAN CLASSIFICATION THEORY AND METHOD

### A. Bayesian Theory

With the development and application of Bayesian theory[6][7], it has become one of the most branches in pattern recognition and statistics.

Bayesian theory and method is the one manner in statistics problems based on Bayesian principle. The key/core of Bayesian theory and method is the Bayesian formula shown as follows :

$$P(w|x) = \frac{P(x|w)p(w)}{P(x)},$$

where  $P(w)$  denotes the initial probability of  $w$ , i.e., the probability *a priori* of  $w$ , and it discloses the rule that  $w$  is the correct hypothesis.  $P(x)$  denotes the total probability of the set  $x$  to be observed.  $P(x|w)$  denotes the conditional probability of  $x$  under the hypothesis  $w$  is true.  $P(w|x)$  denotes the conditional probability  $w$  under the hypothesis  $x$  is true, which is a *posteriori* probability and discloses the confidence level of  $w$  after the given set  $x$ .

The most important application of Bayesian theory and method is the classification in various fields. The classification is the problem of identifying to which of a set of categories a new sample belongs, on the basis of a training set of samples containing ones whose category membership is yet known. Different from other procedures, Bayesian classification procedures provide a natural manner of taking into account any available information about the relative sizes of the sub-populations associated with the different groups within the total population. Bayesian classification procedures often tend to be computationally expensive and approximations for Bayesian clustering rules. A Bayesian network model is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph. Up till now, Bayesian theory and method has been widely used in internet, language and language processing, machine learning, astronomy and physics and so on via building the mathematical optimization model. In most

---

This work is sponsored by the NSFCs (61002052, 61471412).

of the applications, the naive Bayes classifier is one of the most widely used methods.

### B. The Building of Naive Bayes Model

In many cases, our focus of Bayesian classification lies in the revising of a *posteriori* probability. In order to search for the max likelihood of the hypothesis  $w \in W$  (here  $W$  is the candidate hypothesis set) once the set  $x$  is given. This kind of max likelihood of the hypothesis is called as maximum a posteriori hypothesis short as  $w_{map}$ , i.e.,

$$w_{map} = \arg \cdot \max P(w_n | x_m) . \quad (1)$$

Therefore, we have:

$$w_{map} = \arg \cdot \max \frac{P(x_m | w_n)P(w_n)}{P(x_m)} . \quad (2)$$

Since  $P(x_m)$  is the constant independent of  $w_n$ , we get the following formula:

$$w_{map} = \arg \cdot \max P(x_m | w_n)P(w_n) . \quad (3)$$

In order to make good classification, we must combine the practical problem with Bayesian theory well. In fact, the recognition of content's class is related with these factors such as the theoretical deepness, the theoretical span, the degree of difficulty and the maneuverability. That is to say, if a set of factors are given, then the significant correlation will be found through classification and data mining. Thus we can optimize the probability. We can combine the recognition of content's class with Bayes classifier through letting the independent variables denote these factors such as the theoretical deepness, the theoretical span, the degree of difficulty and the maneuverability and so on. At the same time, let the dependent variable denote the content's class. Clearly, the problem of classification can make full use of MAP hypothesis to solve.

Consider the employment of MAP hypothesis, the object of  $x$  is to obtain the most probable value of  $w(x)$  under the condition of the given set  $\langle x_1, x_2, x_3, \dots, x_m \rangle$ .

By the formula (3), we get the following

$$w(x)_{map} = \arg \cdot \max P(x_1, x_2, x_3, \dots, x_m | w_n)P(w_n) . \quad (4)$$

Now what to do is to estimate the two probable values in above formula. It is very easy to estimate every  $P(w_n)$  only if the computing of the frequency that every  $w_n$  arises in the set  $W$  is made. However, the estimation of every  $P(x_1, x_2, x_3, \dots, x_m | w_n)$  is impossible, and the reason lies in these aspects: first, the complexity of the full estimation of  $P(x_1, x_2, x_3, \dots, x_m | w_n)$  amounts to learning a Bayes network, and it is a very difficult question. Therefore, in order to obtain the reasonable estimation, we introduce the naive Bayes classifier under the hypothesis: these independent variables are strictly mutually independent, and the observed joint probability is the product of the probability of every variable.

The detailed equation is as follows:

$$P(x_1, x_2, x_3, \dots, x_m | w_n) = \prod_{j=1}^m P(x_j | w_n) . \quad (5)$$

Insert this equation into the formula (4), we can obtain the classification formula of the naive Bayes classifier:

$$w(x)_{map} = \arg \cdot \max P(w_n) \prod_{j=1}^m P(x_j | w_n) , \quad (6)$$

where,  $x_j$  is the  $j$ 'th property of the factor  $x$ . The probability of  $P(w_n)$  and  $P(x_j | w_n)$  can be computed through computing the joint frequency of different class and property in the set simply. The computing formula is as follows:

$$P(w_n) = \frac{\sum_{i=1}^n \delta(w_i, w_n)}{n} , \quad (7)$$

$$P(x_j | w_n) = \frac{\sum_{i=1}^n \delta(x_{ij}, x_j) \delta(w_i, w_n)}{\sum_{i=1}^n \delta(w_i, w_n)} , \quad (8)$$

where,  $n$  is the number of the sets.  $w_i$  is the  $i$ 'th factor in the set,  $x_{ij}$  is the  $j$ 'th property of the  $i$ 'th factor,  $\delta(w_i, w_n)$  is a two-valued function. When  $w_i = w_n$ , the value is 1, otherwise the value is 0.

In most cases, in order to avoid of the problem of zero-valued frequency, Laplace estimation is often used to smooth the above obtained probability. So, revise (7) and (8) to get:

$$P(w_n) = \frac{\sum_{i=1}^n \delta(w_i, w_n) + 1}{n + n_w} , \quad (9)$$

$$P(x_j | w_n) = \frac{\sum_{i=1}^n \delta(x_{ij}, x_j) \delta(w_i, w_n) + 1}{\sum_{i=1}^n \delta(w_i, w_n) + n_j} , \quad (10)$$

where,  $n_w$  is the number of the dependent variables,  $n_j$  is the number of the property of factors.

### III. THE APPLICATION OF NAIVE BAYES MODEL

Now we use the naive Bayes classifier to solve the classification of the meteorological teaching content: judge which class the content belongs to according to the given factors of the content. Screen a group of data set from the experimental data (see TABLE 1 for partial data), we can divide the dependent variables into two classes: Interesting content  $W_N$  and uninteresting content  $W_Q$ .

We can select the appropriate independent variables: the theoretical deepness and span  $X_1$ , the degree of difficulty  $X_2$ , the maneuverability  $X_3$ . In every class of independent variable we can divide them into three properties :

The normlized values of the theoretical deepness and span (ab. as VTDS)  $X_1: X_{11} \in [0,0.33]; X_{12} \in [0.33,0.67]; X_{13} \in [0.67,1]$ .

The normlized values of the degree of difficulty (ab. as VDD)  $X_2: X_{21} \in [0,0.33]; X_{22} \in [0.33,0.67]; X_{23} \in [0.67,1]$ .

The normlized values of the maneuverability (ab. as VM)  $X_3: X_{31} \in [0,0.33]; X_{32} \in [0.33,0.67]; X_{33} \in [0.67,1]$ .

Now for the given data  $X: (X_1=0.3871; X_2=0.4930; X_3=0.6598)$ , let's judge which class of these data belongs to, the interesting content (ab. as IC) or uninteresting content (ab. as UIC) for the college students?

Clearly, our task is to judge the class of these data belongs to ( $W_N$  or  $W_Q$ ). Now we take one kind of case to consider (the others in the same manner) and build up the naive Bayes classification network. The dependent variable  $W$  belongs to the class  $W_N$ , and the other three independent variables are  $X_1, X_2, X_3$ , which respectively denotes the theoretical deepness and span, the degree of difficulty and the maneuverability. According to the independent character of these variables, we have

$$w_n(x_{ij})_{map} = \arg \max P(x_{1j}|w_n)P(x_{2j}|w_n)P(x_{3j}|w_n)$$

where,  $j$  belongs to the property of the independent variable. Therefore, we can rewrite the above formula as follows:

$$w_n(x_{ij})_{map} = \arg \max P(w_n)P(x_{11}|w_n)P(x_{22}|w_n)P(x_{33}|w_n)$$

In order to compute the value of  $W_N$ , we need estimate the probability from the TABLE 1:  $P(W_N), P(x_{11}|W_N), P(x_{22}|W_N), P(x_{33}|W_N), P(W_Q), P(x_{11}|W_Q), P(x_{22}|W_Q), P(x_{33}|W_Q)$  and so on.

We collect all these computed probability as shown in the following TABLE 2 and 3.

After the computation, we find that the given data belong to the class  $W_Q$ . This result coincides with the real case. In this manner, we perform the comparison with the students' subjective appointting, we find that our method is efficient for the meteorological content automatical classification over 80% accuracy (see Fig.1).

TABLE I. PART OF EXPERIMENTAL DATA.

Number	VTDS	VDD	VM	Content class
1	0.4908	0.5657	0.7811	IC
2	0.1671	0.3193	0.2632	UIC
3	0.7418	0.1159	0.3698	UIC
4	0.5833	0.4465	0.5821	IC
5	0.3908	0.5655	0.7789	IC
6	0.1765	0.3233	0.2098	UIC
7	0.7234	0.1230	0.3743	UIC
8	0.5988	0.4532	0.5987	IC
9	0.7988	0.7554	0.7337	IC
10	0.1766	0.3212	0.2334	UIC
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	0.0946	0.8699	0.9213	UIC

TABLE II. THE CONDITIONAL PROBABILITY OF  $W_N$

$P(x_{ij} W_N)$	$X_{1j}$	$X_{2j}$	$X_{3j}$
$X_{11}$	0.0164	0.9199	0.0160
$X_{12}$	0.9456	0.0640	0.1268
$X_{13}$	0.0443	0.0160	0.8657

TABLE III. THE CONDITIONAL PROBABILITY OF  $W_Q$

$P(x_{ij} W_Q)$	$X_{1j}$	$X_{2j}$	$X_{3j}$
$X_{i1}$	0.9489	0.0490	0.0245
$X_{i2}$	0.0245	0.4900	0.1501
$X_{i3}$	0.0250	0.4589	0.8301

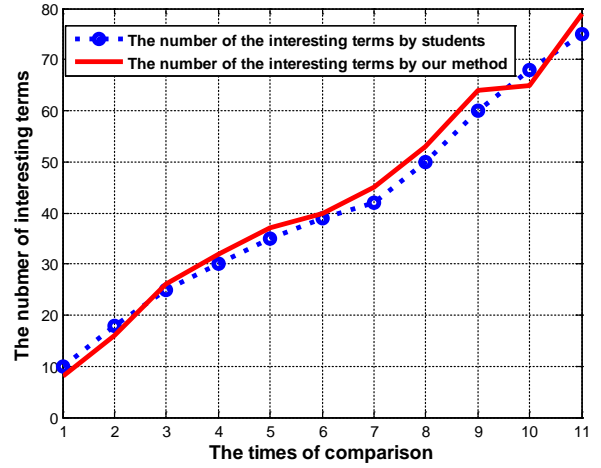


Fig. 1. The comparison between the students and our method for the interesting terms of meteorological content classification

In Fig. 1, we show the comparison between the students and our method for the interesting terms of meteorological content classification. Every time, we give one hundred of terms of meteorological content, and let the college students to determine which ones are interesting for them. Then we use our method to determine which ones are interesting. Finally, we compare the two numbers to see the difference between the students and our method. According to the eleven times' tests, we show that our method is efficient for the meteorological content automatical classification.

#### IV. CONCLUSIONS

In this paper, we proposed a Bayesian model based meteorological content classification. It is well known that in meteorological teaching, in order to improve the teaching efficiency, the content classification is very important. After the content classification, the teachers can give the detailed and better teaching plans to the college students. On the other hand, these classified content can be used for the college students selection of excellent ones. This paper classifies the meteorological content into three classes. And through the experimental comparison with the students', the proposed method is efficient for the meteorological content automatical classification.

# REFERENCES

- [1] Harlow, Essex, England: Longman Group Taxonomy of educational objectives: The classification of educational goals, 1st ed.(1956).
- [2] Zhang Yongning. Meteorology for Mariners, Dalian Marine University Press, Dalian, 2008.
- [3] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-33, no. 4, pp. 959–963, Apr. 1985.
- [4] Liangxiao Jiang, Dianhong Wang, Zhihua Cai, Xuesong Yan, Survey of Improving Naive Bayes for Classification, Advanced Data Mining and Applications, Volume 4632 of the series Lecture Notes in Computer Science, 2007, pp 134-145.
- [5] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification. Willey and Sons, Inc., New York, 2nd edition, 2001.
- [6] Xu Guanlei, Wang Xiaotong, Xu Xiaogang. Improved bi-dimensional EMD and Hilbert spectrum for the analysis of textures[J].Pattern Recognition,2009,42(5):718 – 734
- [7] Xu Guanlei, Wang Xiaotong, Xu Xiaogang. On Analysis of Bi-dimensional Component Decomposition via BEMD[J]. Pattern Recognition,2012,45(4):1617-1626.