Naive Bayesian Classification Algorithm for Infrared Remote Sensing Image Based on Elastic Model

Liu Shangzheng, Liu Bin

Nanyang Institute of Technology, Nanyang, 473004, Henan, China

Keywords: Infrared remote sensing image; Elastic model; Naive Bayesian algorithm

Abstract: The naive Bayesian classification algorithm of infrared remote sensing image is affected by the similarity of uniform eigenvectors of different categories, which leads to the decline of classification accuracy. Therefore, the naive Bayesian classification algorithm of infrared remote sensing image with elastic model is proposed. The spring elongation distance and the elastic coefficient decibel analog sample size and classification criteria are used to exclude the Naive Bayesian classification algorithm from being affected by the conditional independent hypothesis, avoiding noise interference, and thus achieving the goal of improving the accuracy of the classification algorithm. The experiments on infrared remote sensing images show that the naive Bayesian classification algorithm based on the elastic model has strong operability and can improve the classification accuracy.

1. Introduction

The rapid development of Internet technology has enabled human society to enter the era of "information explosion". The channels for receiving information are increasingly diversified, and the types of receiving information are increasingly diversified. How to classify massive information resources for people's use has become the current important topic. The Bayesian algorithm starts from the classical mathematics theory, and the classification hypothesis is based on the assumption that the classification sample follows a certain probability distribution. By calculating the probability of the sample category for the observed sample data, the corresponding optimal classification decision is formed. Bayesian classification algorithm has become a widely used algorithm with its clear logic, convenient algorithm and universality for the same type of data. However, the algorithm is applied in the process of multi-level logic data processing, which is easy to be complicated. Too high leads to limited classification efficiency, which in turn leads to the naive Bayesian classification algorithm.

The naive Bayesian classification algorithm has made great progress in both classification accuracy and efficiency. However, the limitation of conditional independent hypothesis and the influence of noise interference on the classification accuracy rate make the classification efficiency of naive Bayesian algorithm still exist. Certain limitations. In order to make up for the problems of the naive Bayesian classification algorithm in the above aspects, this paper proposes a naive Bayesian classification algorithm for infrared remote sensing images based on the elastic model, which uses the elastic model in the field of physics to simulate the naive Bayesian classification probability. The experimental results show that the algorithm can effectively overcome the shortcomings of the above problems and improve the accuracy of algorithm classification.

2. Modeling Based on Elastic Model

2.1 Overview of Naive Bayesian Classification Algorithm

The Bayesian network, also known as the reliability network, belongs to a model that graphically depicts the causal relationship between data variables. Based on the comprehensive consideration of the prior information of the data, the sample-related data is analyzed. The specific method is The data is categorized and the graphical model presents the potential relationships between the

variables to make them more understandable. In addition, the Bayesian network has a very strong knowledge representation and reasoning ability, which can express uncertain knowledge to a certain extent, and the processing ability of incomplete data is also advanced.

The naive Bayesian classification algorithm adds a conditional independence hypothesis to the Bayesian network, which significantly reduces the difficulty in solving the conditional probability, and becomes the most popular classification algorithm. The hypothesis holds that the probability of being at any node of the naive Bayesian network has no relationship with the set of non-child nodes in the established context of the parent node, namely:

$$p(a_i)|A(a_i), P(a_i) = p(a_i)|P(a_i) \tag{1}$$

Among them, $p(a_i)$ represents node probability; $A(a_i)$ represents arbitrary non-children node set probability of a_i, $P(a_i)$ represents parent node set probability.

The research assumes that the training data set has a total of m classification marks, and any one of the data to be classified has a total of n attributes. Now assume that X is an unknown sample of the classification mark, and obtain the following calculation formula according to the classification attribute of X:

$$P(X|C_i) = \prod_{k=1}^{m} P(x_k|C_i) \tag{2}$$

The main purpose of classification is to distinguish multi-category infrared remote sensing images. The technical route followed is mainly to extract the typical features of the images to be classified. It is assumed that the features to be extracted by the algorithm have N categories, and each feature type has potential values in M. . In order to be able to distinguish images in a significant way, the selection of classification features is crucial. Specifically, the following requirements must be met: First, the variance of the variance between the values of the same feature is not large, and it is convenient to judge the value range of the test data, thereby enhancing the credibility of the algorithm classification; The value gap should be significantly reduced to reduce the coincidence degree of the classification range and to resolve the untrustworthy problem caused by the fuzzy data type.

In the naive Bayesian classification algorithm, the prior probabilities of the sample attributes are:

$$P(x_k|C_i) = \frac{S_{ik}}{S_i} \tag{3}$$

Among them, S_i is the number of samples represented by C_i in the training data sample, S_{ik} is the number of sample instances in which the C_i classifier and the attribute value is A_k. The value method focuses on the consideration of sample weight, and the sample value change gradient is rarely involved, while the "length" relies on the distribution characteristics of the test data itself, highlighting the part with higher attribute classification efficiency and reducing the fuzzy data. The impact on classification, which in turn improves classification efficiency.

In addition, the continuous sample value is discretized, and the continuous attribute value is replaced by a discrete interval, and the continuous attribute of the sample can be converted into a descriptor attribute. The naive Bayesian classification algorithm of infrared remote sensing image is aimed at reducing the difficulty of probability calculation. However, the assumption of conditional independent attribute makes the application of the algorithm in the actual situation limited. Therefore, the naive Bayes classification algorithm is performed by a certain strategy. Improvements, so that the dependencies between attributes can be realistically necessary in the interval where the algorithm can be classified.

2.2 Naive Bayesian Classification Algorithm for Infrared Remote Sensing Image Based on Elastic Model

In the field of physics, the formula of the elastic model is:

$$F = Kx \qquad (4)$$

$$K = \frac{G \times d^4}{8 \times N_c \times D_m^3} \qquad (5)$$

Among them, K is the stiffness coefficient, x is the length of the spring extension, G is wire rigid modulus, N_c is Effective number of turns, d and D_m representing the wire diameter and the medium diameter respectively.

In the naive Bayesian classification algorithm of infrared remote sensing image, the classification index and the sample are regarded as the elastic coefficient and the spring elongation length respectively, and the elastic formula in physics can be analogized to the infrared remote sensing image classification and attribute, that is, two The probability between the two increases with the increase of the elastic force, and the elastic strength can be used as the classification probability measure.

The number of samples is equal to the sum of the "lengths" of the attributes, and the sample M（M_1, M_2...M_n）has n attributes. Therefore, the sample length can be expressed as $L_A = \sum_{i=1}^{n} L_i$,the same as the elastic direction between the same category, and the principle of the force is obtained:

$$F_{MC_i} = \sum_{j=1}^{n} M_j C_i = \sum_{j=1}^{n} \frac{1}{r_{ij}^2} L_i l_{ij} \qquad (6)$$

Among them, l_{ij} represents classification mark C_i ,sample M（M_1, M_2...M_n）'s attribute value of attribute M_i corresponds to the prior probability of C_i .

3. Result Analysis

Based on the purpose of verifying the effectiveness of the algorithm, infrared remote sensing images were acquired from the relevant platforms as experimental data, and compared with NB (traditional Bayesian classification algorithm) and SVM. The selected infrared remote sensing images were mainly farmland, forest and beach. Each type of image was selected for analysis. The image examples are shown in Figures 1, 2 and 3.
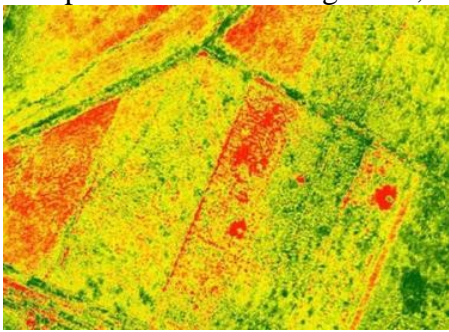

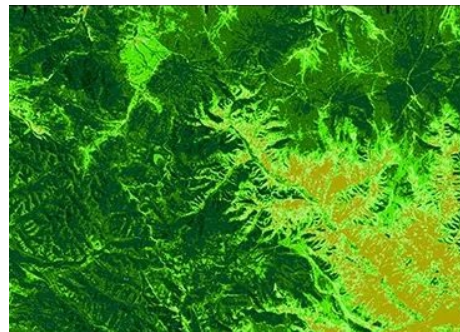Figure 1 Farmland infrared remote sensing image

Figure 2 Forest infrared remote sensing image

Figure 3 Beach infrared remote sensing image

30 kinds of infrared remote sensing images from farmland, forest and beach were selected to form redundant dictionary, 30 pieces were used for training purposes, and the remaining 40 images were selected for testing according to the randomness principle. Then, the training and test images are respectively represented in the redundant dictionary in the sparse form, and then the sparse representation feature vector of the infrared remote sensing image is obtained, and input into the classifier for classification.

The classification feature of the image data is extracted by the feature vector. The extraction position is mainly the centroid and the maximum part of the vector. At the same time, in order to eliminate the influence of noise on the extraction, after the descending method is sorted, the first three value positions are selected. The maximum value is divided into three feature points to eliminate the influence of noise, thereby increasing the confidence. Table 1 compares the accuracy of the classification of image data by the infrared image naive Bayesian classification (E-NB) and naive Bayesian classification (NB), SVM algorithm and R-NB algorithm.

Table 1 Recognition rate under different sample sizes under algorithm difference

Number of test samples	Number of categories	NB	SVM	E-NB
60	3	78.44%	85.12%	83.46%
120	3	80.91%	86.34%	87.69%
Mean	3	79.68%	85.31%	85.43%

Source of data: authors make their own based on the collected data

It can be found from Table 1 that in the background of different test samples, E-NB is significantly higher in classification efficiency than NB and R-NB algorithms. In addition, the classification efficiency increases as the number of test samples increases. When the number of samples is small, the E-NB and SVM algorithms are relatively flat in terms of classification efficiency. However, the SVM algorithm has the limitation of large sample data processing capability, and the E-NB algorithm classification efficiency gradually increases with the increase of the number of test samples. Higher than the SVM algorithm. Furthermore, the E-NB algorithm can continuously improve the classification efficiency by combining with other attribute weighting methods. It can be seen that the naive Bayesian classification algorithm based on the elastic model constructed by the Institute has operational characteristics. Feasibility, and can significantly improve image classification efficiency.

4. Conclusion

The naive Bayesian algorithm of infrared image based on elastic model proposed in this study starts from the characteristics of training data and considers the difference of infrared remote sensing image recognition under different characteristic conditions. It is proposed to simplify the probability calculation complexity with elastic model. An improved algorithm to improve classification accuracy. The experimental results show that the improved naive Bayesian classification algorithm presented in this study has strong operability and overcomes the influence

of the conditional independence hypothesis on the classification efficiency in the naive Bayesian classification algorithm. It is an effective image. The data classification method can be used as a reference for the classification of current infrared remote sensing images.

References

[1] Cui Liangzhong, Guo Fuliang, Song Jianxin. Research on Naive Bayesian Data Classification Algorithm Based on Map/Reduce[J].Journal of Naval University of Engineering,2019(04):7-10.

[2] Ye Hanxiao, Wu Ziwei. Design and implementation of prescription song learning system based on Bayesian classification algorithm [J]. China Digital Medicine, 2019, 14 (07): 49-51.

[3] Wu Jiazhen, Zhou Hongyu, Liu Linfeng. Vehicle Routing Algorithm Based on Semi-simple Bayesian Classifier[J].Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition), 2019,39(03):76-81.

[4] Li Honglei, Cong Yuliang, Ren Baihan. Vertical Switching Algorithm Based on Hierarchical Bayesian Classification Method[J]. Journal of Jilin University: Information Science Edition, 2019, 37(03): 238-244.