A Naive Bayes Classification Algorithm Based on Improved Feature Weighting (Postprint)

Authors: Ding Yue, Wang Xueming

Date: 2018-10-11T00:00:00+00:00

Abstract

The traditional Naive Bayes classification algorithm fails to assign different importance weights to features according to their characteristics, leading to inaccurate classification results. To address this problem, Jensen-Shannon (JS) divergence is introduced to quantify the amount of information that each feature can provide. To overcome the limitations of JS divergence, adjustments and corrections are made by considering three aspects: word frequency and document frequency within and between categories, and inverse category frequency corrected by the coefficient of variation. The weight of each feature is ultimately calculated and incorporated into the Naive Bayes formula. Comparative experiments with other algorithms demonstrate that the Naive Bayes algorithm enhanced based on JS divergence and the three aspects of word, document, and category achieves the best classification performance. Therefore, the JS divergence feature-weighted Naive Bayes classification algorithm exhibits substantial improvement in classification performance compared with other classification algorithms.

Full Text

Preamble

Title: A Naive Bayes Classification Algorithm Based on Improved Feature Weighting

Authors: Ding Yue, Wang Xueming† (College of Computer Science & Technology, Guizhou University, Guiyang 550025, China)

Abstract: The traditional Naive Bayes classification algorithm does not differentiate the importance of features according to their characteristics, leading to inaccurate classification results. To address this problem, we introduce Jensen-Shannon (JS) divergence to represent the amount of information provided by each feature term. To overcome the limitations of JS divergence, we adjust

and correct it from three perspectives: term frequency and document frequency both within and between categories, and inverse category frequency corrected by the coefficient of variation. We then calculate the weight of each feature term and incorporate these weights into the Naive Bayes formula. Comparative experiments with other algorithms demonstrate that the Naive Bayes algorithm improved by JS divergence and enhanced from the three dimensions of term, document, and category achieves the best classification performance. Therefore, compared with other classification algorithms, the proposed JS divergence-based feature-weighted Naive Bayes classification algorithm significantly improves classification performance.

Keywords: text classification; Naive Bayes; Jensen-Shannon divergence; term frequency; document frequency; class frequency

0 Introduction

With the rapid development of the Internet, information is emerging on a massive scale. How to filter target information from this vast amount of data has become a critical research topic in information mining. Various text classification algorithms in data mining group and categorize information, improving both accuracy and efficiency. Commonly used classification algorithms include decision tree classification, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes classification. Experimental studies have shown that when processing large-scale datasets, KNN incurs substantial computational overhead, SVM has high time costs despite its high accuracy, and decision tree efficiency decreases as data volume increases. In contrast, Naive Bayes classification maintains relatively stable efficiency during classification, and its required parameters are easily obtained and computable even with limited data. Overall, the algorithm is relatively simple and efficient, supported by robust mathematical theory. However, its classification accuracy is relatively poor, leaving room for improvement.

Currently, there are two main categories of improvement methods for traditional Naive Bayes classification. The first relaxes the coupling degree of the independence assumption, improving classification accuracy by reducing the independence constraint. However, this approach significantly increases computational cost. For example, Tree-Augmented Naive Bayes (TAN) improves classification accuracy but substantially increases computational difficulty. The second method amplifies the influence of important features in text classification by assigning weights to feature terms. This approach is both simple and effective at improving classification accuracy. Therefore, this paper proposes a feature weighting algorithm based on JS divergence that assigns different weights to features according to their varying contributions to classification results, thereby improving the Naive Bayes algorithm.

1 Related Research

1.1 Naive Bayes Classification Algorithm

Naive Bayes is an effective classification algorithm based on Bayesian theory that assumes conditional independence among features. Given a category set $C(c_1, c_2, c_3, ..., c_n)$ and a text to be classified with feature terms $X(x_1, x_2, x_3, ..., x_n)$, Naive Bayes calculates the probability $P(c_n|X)$ of the feature terms belonging to each category under the assumption that features are independent between categories c_a and c_b. The category with the maximum probability is the classification result. The Naive Bayes classification formula is:

$$\mathrm{NB} = \arg\max_{c_n} P(c_n) \prod_{m=1}^{j} P(x_m|c_n) \tag{1}$$

where $P(c_n)$ represents the probability that the text to be classified belongs to category c_n, and $P(x_m|c_n)$ represents the probability that category c_n contains feature term x_m.

However, the formula assumes complete independence of each feature vector and identical weights for all feature terms, which does not reflect reality and leads to suboptimal classification results.

1.2 Feature Weighting Algorithms

Many researchers have studied and improved Naive Bayes classification models using attribute weighting algorithms. Commonly used feature weighting algorithms include Term Frequency (TF), Inverse Document Frequency (IDF), Information Gain (IG), Mutual Information (MI), and Expected Cross Entropy (ECE). Shan et al. compared TFIDF, MI, IG, and ECE, proposing improvements and applying them to travel-related text classification, concluding that absent terms cause more interference than contribution to classification. Rao et al. improved the traditional TFIDF algorithm by incorporating the distribution of features within and between classes, proposing the TFIDF-FC algorithm that improved classification performance when applied to Naive Bayes. Wang et al. focused on the relationship between terms and categories, proposing that if a feature appears in most categories, its weight should be reduced to improve accuracy. Shi et al. addressed IG' s limitation of ignoring term frequency, proposing to incorporate within-class and between-class term frequencies into IG to improve accuracy. Peng et al. proposed an improved IG algorithm based on relative document frequency distribution, further enhancing classification performance.

Although these methods have improved classification accuracy to some extent, they only consider single, partial aspects without comprehensively evaluating

the information content carried by features, term frequency, document frequency, category frequency, and their distributions within and between categories. Therefore, this paper uses JS divergence to represent feature information content and proposes a new JS divergence-based feature-weighted Naive Bayes classification algorithm that comprehensively considers features from three dimensions: term, document, and category.

---

2 Improved Feature-Weighted Naive Bayes Classification Algorithm

Traditional Naive Bayes treats all features as equally important, but in practice, features contribute differently to classification, reducing accuracy. Therefore, it is necessary to use text feature selection algorithms to weight features and improve performance. The improved Naive Bayes formula is:

$$\mathrm{NB} = \arg\max_{c_n} P(c_n) \prod_{m=1}^{j} \omega(m,n) \times P(x_m|c_n) \tag{2}$$

where $\omega(m,n)$ is the weight of feature term x_m in category c_n, measuring its importance in classification. Accurately calculating $\omega(m,n)$ is key to improving Naive Bayes classification accuracy.

2.1 JS Divergence and Its Limitations

Entropy was originally a physics term until Shannon introduced it to information theory in 1948 as a measure of information uncertainty. Information Gain (IG) is the difference in information entropy when a feature term is present versus absent, representing the amount of uncertainty reduced by the feature's presence —the information provided by the feature term.

KL (Kullback-Leibler) divergence, also called cross-entropy, is similar to IG. It measures the distance difference between two probability distributions of a feature term' s presence and absence in documents, representing the information contributed by the feature. Unlike IG, KL divergence only considers the impact of feature presence on classification, not absence. While feature absence also affects classification, its interference outweighs its contribution, making KL divergence more accurate than IG for weight calculation. The KL divergence formula is:

$$\mathrm{KL}(P||Q) = \sum_{n} P(c_n|x_m) \log \frac{P(c_n|x_m)}{P(c_n)} \tag{3}$$

where $P(c_n|x_m)$ represents the probability that a document containing feature term x_m belongs to category c_n, and $P(c_n)$ represents the proportion of category

c_n in the entire training set.

In probability theory and statistics, JS divergence is a method for measuring the similarity between two probability distributions based on divergence, offering advantages over KL divergence. KL divergence has limitations: (a) it lacks symmetry and is not a true distance metric despite appearing to measure distance; (b) its unbounded results make comparison difficult. JS divergence, a variant based on KL divergence, inherits its advantages while addressing these defects. JS divergence results always range between 0 and 1, making similarity judgments more definite and comparisons easier than with KL divergence. JS divergence is symmetric and represents a true distance measurement standard. The JS divergence formula is:

$$\mathrm{JS}(P||Q) = \frac{1}{2}\mathrm{KL}\left(P\middle\|\frac{P+Q}{2}\right) + \frac{1}{2}\mathrm{KL}\left(Q\middle\|\frac{P+Q}{2}\right) \tag{4}$$

The expanded form is:

$$\mathrm{JS}(P||Q) = \frac{1}{2}\sum_n P(c_n|x_m)\log\frac{2P(c_n|x_m)}{P(c_n|x_m)+P(c_n)} + \frac{1}{2}\sum_n P(c_n)\log\frac{2P(c_n)}{P(c_n|x_m)+P(c_n)} \tag{5}$$

Introducing JS divergence into Naive Bayes allows representing the information carried by features through the distance between two distributions. Therefore, the larger the JS entropy of feature term x_m, the greater its assigned weight.

However, the JS divergence formula has three shortcomings when evaluating feature importance:

a) **Ignores the impact of term frequency on weight.** If a feature term exists in all texts of a category but appears only a few times in each text—widely distributed but with low frequency—it cannot well represent the category despite having high between-category document frequency and thus receives an inaccurately high weight. For example, consider two categories with three texts each, containing two feature terms. As shown in , both feature terms t_1 and t_2 appear in three texts of category c_1 and one text of category c_2. JS divergence calculates equal weights for t_1 and t_2, yet t_1 appears more frequently than t_2 in each text of c_1 and less frequently in c_2. Clearly, t_1 better represents c_1 and contributes more to classification, deserving a higher weight. This is not reflected in the JS divergence calculation, causing errors.

b) **Does not reflect within-category concentration of documents containing the feature.** The $P(c_n|x_m)$ term in JS divergence reflects the aggregation degree of documents containing the feature term across categories but not their concentration within a specific category. If a feature term is uniformly distributed across texts of a category, it indicates

representativeness for that category and should receive greater weight, requiring consideration of the distribution proportion within categories.

c) **Neglects the impact of category frequency on text classification.** When two feature terms have identical term frequency and document frequency, the ratio of total categories to categories containing the feature term also indicates importance. As shown in , feature terms t_3 and t_4 have identical term frequency and document frequency in the training set, but t_3 appears only in categories c_3 and c_4, while t_4 appears in all four categories. Feature term t_3 is more concentrated and thus has greater discriminative power between categories.

Therefore, to differentiate feature importance, we must comprehensively consider term frequency, document frequency, and inverse category frequency.

2.2 Feature Term Frequency TF

Term frequency refers to the frequency of feature term x_m in text d. Since text classification aims to categorize texts into classes rather than distinguish between individual texts, we must first 统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计

统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计
统计统
计统计计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计

统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统
计统计
统计统计统计统计统计统计

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.