Myopia prediction and early warning model based on naive Bayesian model

Junfeng Hu

Xidian University, Xi'an, China

171778195@qq.com

Keywords: Myopia evolution mechanism; naive Bayes classifier; principal component analysis; statistical regression analysis; Gaussian distribution

Abstract: In recent years, the problem of myopia in children and adolescents in China has been significantly intensified and approached to a younger age. This paper analyzes the causes of myopia and the early warning conditions of myopia by establishing a reasonable mathematical model. After collecting data, we obtained age, genetic factors, eye time for over-loading, education level, and sleep time as the main influencing indicators. Different quantitative models were established through a large number of statistical laws, and the eyeball structure was studied. With the formula and the lens, The ciliary body-based regulation mechanism and the characteristics of true and false myopia, we have summarized the evolution mechanism model of human myopia. Through Principal Component Analysis (PCA) and SPSS factor analysis, the weights of each feature on myopia are obtained. The features are brought into the target equation to solve the predicted value Y, and the Y* (standard value) calculated by the survey and brought into the model. For comparison, the suitable solution for the teenager is proposed according to different deviations.

1. Introduction

The problem of myopia in adolescents has become a major problem at this stage. The proportion of myopia among Chinese primary school students is 45.7% that of junior high school students is 74.4%, that of high school students is 83.3%, and that of college students is 87.7%. Moreover, this trend has already affected regional conscription. At this stage, factors affecting myopia and refractive errors in adolescents, such as high-intensity learning eyes, are not the main influencing factors.

2. Related work

2.1 Restatement of the Problem

(1) Analyze factors affecting vision and establish a quantitative model, giving data and its possible sources

(2) Analyze the mechanism of myopia evolution and establish a model

2.2 Overview of our work

- Part one

We analyzed the genetic factors, age, high-intensity eye time education level, and sleep time a total of five major influencing factors.

For high-intensity eye time, outdoor exercise, sleep time and education time, we adopt statistical regression model, adopt Gaussian fitting for different indicators, calculate the sum of mean squared deviations of actual points and regression points to obtain the loss function to be optimized. The least square method and the normal equation method are used to solve the problem, and finally the equation for fitting the scatter plot is obtained. In view of the influence of genetic factors on the results, we established a probabilistic model and used Bayesian and other formulas to obtain the quantitative relationship between parents' myopia and children's myopia.

- Part two

We studied the structure of the eyeball and consulted relevant data. It is concluded that changes in

  DOI: 10.25236/isrme.2019.116

eye vision are mainly due to changes in axial length and refractive power of the eye. Therefore, we derived the formula for the relationship between the degree of myopia and the length of the eye axis and the diopter of the eye. For the analysis of evolutionary mechanism models, we explored the process of true myopia and pseudomyopia, studied the changes of ciliary body and lens, and summarized the evolution mechanism model.

2.3 Assumptions

(1) Assume that all surveyed people have no trauma to the eyes and other diseases that can affect myopia.

(2) Assume that all the surveyed adolescents have the same daily eye habits and no bad habits related to myopia.

(3) It is assumed that there is no difference between the different regions in the eyes of adolescents.

(4) It is assumed that the probability of myopia in ordinary parents is equal to the proportion of myopia in adults.

(5) Assume that light from 5 meters outside the lens is parallel rays

3. The Model

3.1 Part one

- Influence model of genetic factors

The study found that genetic factors have a huge impact on children's myopia, which is a congenital effect, that is, the prevalence of myopia in the offspring is related to the prevalence of the parents. For the overall Chinese environment, the degree of myopia in adults is almost It will change, so a probabilistic model is used to represent the effect of heredity on offspring myopia.

1) Parents suffer from myopia

The proportion of male and female myopia in the crowd is approximated to indicate the probability of myopia in ordinary adolescent parents. The formula is as follows:

$$P\ (\text{father}) = P\ (A)$$

$$P\ (\text{mother}) = P\ (B)$$

This index reflects the myopia probability of ordinary parents, and under certain rules, the myopia of the parents will lead to the child's myopia, and is related to the degree of myopia of the parents. The application conditional probability formula is as follows

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

In the formula, P(AB) is the joint probability of event AB, P(A|B) is the conditional probability, which indicates the probability of A under B condition, and P(B) is the probability of event B. Let the probability that the mother is not nearsighted is P(D), then

$$P(D)=1-P(B)$$

Similarly, the probability of father not being nearsighted is P(E), which is available.

$$P(E)=1-P(A)$$

The probability of myopia in children with myopia is P(C|A), and the probability of myopia in the case of mother myopia is P(C|B). The probability of myopia in the case of both parents. For P(C|AB), the probability of myopia in children with neither parents being nearsighted is P(C|DE).

2) Bayesian formula

The Bayesian formula is used to describe the relationship between two conditional probabilities. According to the multiplication rule, the formula is expressed as

The above formula can be used to express the influence of parental illness in congenital myopia.

Under the premise of facial myopia, the probability P(C) of myopia is obtained, because the myopia in the crowd is independent of each other, and the probability formula is as follows

$$P(C)=P(C|B)P(B)+P(C|A)P(A)$$

$$P(AB)=P(A)P(B)$$

$$P(ABC)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC)$$

From the above formula, the probabilistic model of the child with myopia is obtained:

$$P(C|AB)=\frac{P(ABC)}{P(AB)}$$

This probability formula combined with the three formulas shown above can account for the proportion of male and female myopia in the known adult population. This data can be obtained from social surveys, according to the Xinqin law of large numbers.

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} a_i-\mu\right|<\varepsilon\right)=1$$

It can be expressed as an approximate probability when the amount of statistical data is large. In addition, the probability of myopia in the case of a single-parent myopia can be obtained through statistical data, thereby obtaining the probability that the child will suffer from myopia in a specific group. And the extent to which genetic factors influence myopia.

● Model of other factors affecting myopia

(1) Statistical regression model using Gaussian fitting

After a large number of investigations, we have mastered a large amount of social data, including eye habits for adolescents, high-intensity eye time, average outdoor exercise time (number of times), education level and sleep time are factors that mainly affect young people's myopia, and these data are formed. The statistical laws can be applied to different intrinsic connections and mathematical laws using different linear fitting methods.

Taking sleep as an example, the horizontal axis of the data set is the daily sleep time, and the vertical axis is the proportion of people with myopia.

1) Establish a regression equation model

A scatter plot of sleep time and percentage of myopia by a large amount of statistical data is as follows:

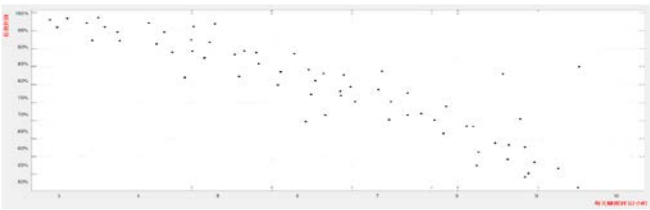

Figure 1 Education level and myopia scatter plot

As can be seen from the figure, as the sleep time x increases, the percentage of myopia y has a decreasing trend of decreasing rate. Here, we use curves to fit these data points, in this way, from curves or lines to The difference in distance between data points is minimal. The data is not complicated, so Gaussian fitting is used, that is, a Gaussian curve is used to fit a function curve, as follows

$$Gi(x)=A_i*e^{\frac{(x-B_i)^2}{C_i^2}}$$

This Gaussian regression can be compared to the polynomial fitting. The difference is that the polynomial fitting is a power function system, and the Gaussian fitting is a Gaussian function system. You can choose a combination of many group features, such as a combination of multiple independent variables x1, x2, we only consider the combination of two features for regression, that is, only three parameters Ai, Bi, Ci.

2) Error function: least squares model

The actual scatter data statistic minus the square of the result of the polynomial function to represent the error relationship between f(x) and the polynomial function, ie

When Loss is reduced to a minimum in the iterative calculation, it can be considered that the parameter can best fit the points of the scatter plot and form a function curve.

$$Loss = \sum_{k=1}^{N} (f(x_i) - [G(i)])^2$$

3) Solution of the model: normal equation method

The least squares processing procedure of linear parameters can be summarized as follows: firstly, the error equation is listed according to the specific problem; then according to the principle of least squares, the method of extremum is used to transform the error into a normal equation; then the normal equation is solved to obtain the desired equation. Estimate; finally give an accuracy estimate. For nonlinear parameters, they can be linearized first and then processed according to the least squares processing procedure of the above linear parameters. Therefore, the establishment of the normal equation is the basic link of the parameter least squares processing.

Our data set is relatively small, and we can use the normal equation method once.

$$\theta = (X^T X)^{-1} X^T y$$

The above formula is a normal equation expression. Assuming that our training set feature matrix is X (contains x0=1) and our training set result is vector y, the normal equation is used to solve the vector. The superscript T represents matrix transposition, the superscript -1 represents the inverse of the matrix, and the equations of the following figure are solved (only ai, bi, ci for this problem)

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{m+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum x_i^m & \sum x_i^{m+1} & \sum x_i^{m+2} & \cdots & \sum x_i^{2m} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \cdots \\ \sum x_i^m y_i \end{pmatrix}$$

In our expression, the matrix represented by θ is the parameter matrix, which is our optimization goal. Because the polynomial we set is 2 times and 3 terms, our parameter is the vector composed of Ai, Bi, and Ci, that is, the following formula

$$\theta = \begin{pmatrix} Ai \\ Bi \\ Ci \end{pmatrix}$$

Our data set is relatively small. We can solve it without iteration and training in one time by using the normal equation method. It has higher running efficiency. We sort the points in the scatter plot into [X, Y] matrix and substitute it into the normal equation.

$$\begin{cases} Ai = 99.4 \\ Bi = 2.764 \\ Ci = 8.251 \end{cases}$$

The curve equation and fitting result constructed by the parameter matrix are as follows

$$Gi(x) = 99.4 * e^{\frac{(x-2.764)^2}{8.251^2}}$$

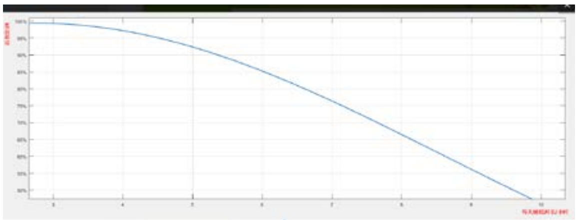

Figure 2 Gaussian fitting curve of daily sleep time on myopia

By the above method, the statistical regression method is applied to the sleep time, the exercise amount, the high intensity eye time, and the age-to-adolescent myopia data, and the quantitative influence curves of these features on myopia can be obtained, and the specific solution is obtained by the least squares method. Solution, by contrast, from the scatter plot fitting effect, the Gaussian fitting effect on the effect of age on vision is the best, and the solution process is the same.

The same applies Gaussian fitting, the solution of the effect of age on myopia, the horizontal axis is the age, and the vertical axis is the proportion of myopia in this age group.

$$Gi(x) = 92.44 * e^{\frac{(x-21.29)^2}{11.21^2}}$$

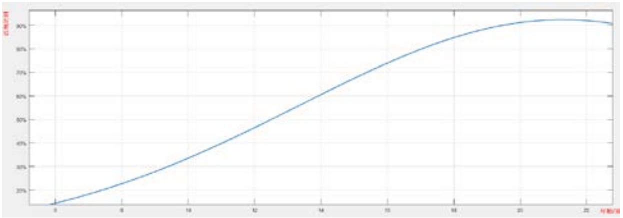

Figure 3 Gaussian fitting curve of age on myopia

(2) Statistical regression model using Fourier series fitting

After trying various fitting methods such as Gaussian fitting, polynomial fitting, etc., we found that in the effect of education on myopia, the Fourier fitting results have the least loss value, that is, the best fitting effect.

The core principle is to use the Fourier series to represent the fitted curve in the objective function. The Fourier fitting method is a discrete Fourier transform based harmonic analysis method. For any function, it can be expanded into The sum of the trigonometric functions whose angular frequency is nω, the formula is as follows

$$S_N x = \frac{a_0}{2} + \sum_{n=1}^{N} \left(a_n \cos \frac{2\pi nx}{P} + b_n \sin \frac{2\pi nx}{P} \right)$$

Where A0 is a constant component, P is a period, and an, bn is defined as the Fourier coefficient as follows

$$a_n = \frac{2}{P} \int_{x_0}^{x_0+P} s(x) \cdot \cos(\tfrac{2\pi nx}{P})\, dx,$$

$$b_n = \frac{2}{P} \int_{x_0}^{x_0+P} s(x) \cdot \sin(\tfrac{2\pi nx}{P})\, dx,$$

Set when using Fourier series as the objective function

$$f(x) = a_0 + a_1 \cos(x\omega) + b_1 \sin(x\omega)$$

The loss function and the solution process do not need to change, still using the least squares method and the normal equation

$$\theta = \begin{pmatrix} a_0 \\ a_1 \\ b_1 \\ \omega \end{pmatrix}$$

The difference is that θ (parameter matrix) becomes the Fourier parameter as follows

Then apply the above method to the normal equation method to solve the matrix solution as follows

The Fourier series equation and the fitted curve thus obtained are as follows

$$f(x) = -9.721\times10^8+9.721\times10^8\cos(2.928\times 10^{-5}x) + 5.427 \times 10^5 \sin(2.928 \times 10^{-5}x)$$

$$\begin{cases} a_0 = -9.721 \times 10^8 \\ a_1 = 9.721 \times 10^8 \\ b_1 = 5.427 \times 10^5 \\ \omega = 2.928 \times 10^{-5} \end{cases}$$

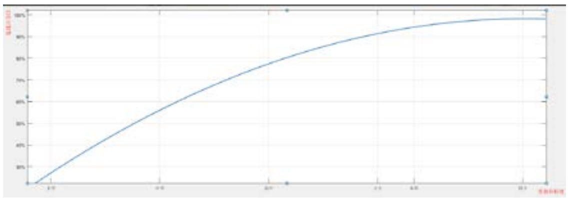

Figure 4 The influence of education level on myopia

Similarly, the fitting function and fitting curve of continuous eye time and myopia are as follows

$$f(x) = 286.8e^{0.182x} - 361.3$$

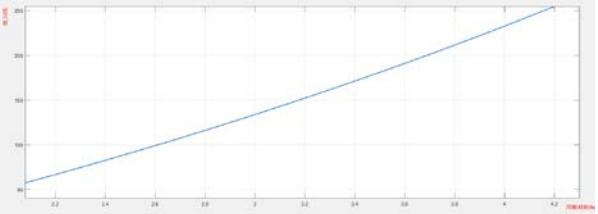

Figure 5 Daily continuous eye time and myopia

- Data and its access

For the congenital effects of genetics on myopia, the data we need to obtain a probabilistic model is the proportion of men and women who are suffering from myopia in a specific population. The access can be a network survey, a large-scale questionnaire, and an official health report.

For the analysis of the impact of sleep time, education level, age, etc. on myopia, the statistical regression model requires data on specific characteristics in the population. It can be obtained through the reports of non-profit organizations such as the Red Cross Health Report and large youth social journals. Myopia-related data, matching the teenage myopia statistics in the region, can generate scatter plots for statistical regression analysis.

3.2 Part two

- Model preparation

Principle and degree formula of myopia

By reviewing the data, we learned that myopia can be divided into two types: refractive myopia (due to the refractive index enhancement of the eye) and axial myopia (the front long axis of the eyeball exceeds the normal range). The refractive index of the refractive index is mainly determined by the refractive power. The length of the anterior-posterior axis of the eyeball, that is, the length of the axial axis (ie, the center of the cornea to the center of the retina).

The positive vision we know refers to the regulation of the eye when it is still (the lens of the lens, the ciliary muscle intervention). After the external parallel light is refracted by the refractive system, it can be focused at the center of the macular of the retina. The effect of the right eye is also corrected by the glasses to the person with myopia. effect.

According to the data, the lens degree C1=eye diopter D1×(-100), the glasses are concave lenses, the diopter is negative, the cornea and the lens can be regarded as convex lenses, and when the two lenses are combined, the focal length formula is

$$f = \frac{f_1 f_2}{f1 + f2 - s}$$

Where f1, f2 are the focal lengths of the two lenses, and s is the distance between the optical centers of the two lenses (eye lens and glasses). We compare the cornea and the lens to obtain the equivalent diopter. Our calculation results are 58.1D and the scientific results are 58.6. D is very close. By referring to the following table, we know that the distance between the eyeglasses (the distance from the glasses to the cornea) is preferably 12~13mm, but considering the human myopia, the shape of the anterior chamber and the shape of the lens are unpredictable, and the corneal diopter is large. In the lens, we have the distance from the glasses to the cornea equivalent to 13mm. After the correction of the glasses, the combined focal length should be equal to the length of the eye axis.

$$f = \frac{1}{D}$$

Where D is the equivalent lens diopter

It is known that the refractive power of the cornea is about 43.1, the diopter of the lens is about 19.7, the anterior chamber is about 3~5mm, and the length of the axial length is about 24.2mm. Therefore, it is assumed that the refractive power of the lens is d2, and the refractive power of the eye is d1, the eye. The length of the axis is l, you can get the following equations

$$\begin{cases} f_1 = \dfrac{1}{d_1} \\ f_2 = \dfrac{1}{d_2} \\ \dfrac{1}{d_1 d_2} = (\dfrac{1}{d_1} + \dfrac{1}{d_2} - s)l \end{cases}$$

The equation for obtaining the degree of the glasses from this equation is as follows

$$C_1 = -100 d_2 = \frac{100(d_1 l - 1)}{(1 - d_1 s)l}$$

S is known, so we get the relationship between myopia and axial length, corneal diopter.

- Establishment of an evolutionary model

Myopia is often caused by bad eye habits on weekdays. The principle lies in the long-term use of the eye, the ciliary muscle is often in a state of contraction, resulting in adjustment of tension or regulation of paralysis. When the lens loses the traction of the ciliary muscle, it thickens and becomes convex. A thickened and convex lens can cause the light to not focus on the retina when it enters the eye. This stage is in the early stage of myopia, except for the congenital myopia, mostly pseudo-myopia.

False myopia can be 100% restored to myopia by the changes in eye habits and proper training mentioned later. If you wear glasses because of pseudo-myopia or if you have pseudo-myopia that is not treated in time, it will become true myopia. True myopia, also known as axial myopia, is due to congenital or acquired factors, which makes the anteroposterior diameter of the eyeball (ie, the axial length) longer than the normal average of 24 mm, causing parallel rays to enter the eyeball, and the focus falls in front of the retina and cannot be imaged clearly.

Prolonged use of the eye for a long time causes the ciliary muscle to often be in a contracted state, resulting in adjustment of tension or regulation of spasm. When the lens loses the traction of the ciliary muscle, it thickens and becomes convex. When the light enters the eye, it focuses on point D, and there is a distance from the point C of the retina. So at this time we can't see it clearly. The most real performance is to see the objects in the distance become blurred. In general, adolescents with early myopia abstain from congenital myopia, one is pseudo-myopia caused by eye muscle tension, and the other is true myopia caused by the axis of the eye.

Normally, when our eyes look far away, the ciliary muscles relax and the lens becomes thinner. When we look close, the ciliary muscles contract and the lens becomes thicker and convex. Myopia is caused by prolonged looking objects, causing ciliary muscle spasm, and can not effectively adjust the traction of the lens, so pseudo-myopia occurs at this time. There was no fundamental change in the axial axis of patients with pseudo-myopia. Proper rest and exercise can be recovered

The process of true myopia occurs when there is no proper treatment in the case of pseudo-myopia. The contraction of the lens thickens to produce two forces on the eye. One force is the pressure in the axial direction due to the thickening of the lens, which can be pressed against the vitreous and thus the eye shaft. The other force is the tensile force perpendicular to the axial direction, which is equivalent to the pulling force of the lens when the muscle contracts. This pulling force can flatten the eyeball, thereby indirectly making the eye axis longer. These two forces work together to flatten and lengthen the eye, from a spherical shape to an olive shape, and the curvature of the cornea is increased, which is the deformation of myopia caused by the adjustment force of the lens. The so-called true myopia should be the long-term force that causes the eye axis to grow, and the eye is deformed to the point where it is difficult to return to a spherical shape. The higher the degree of myopia, the more the lens shrinks and the greater the force, which is enough to change the shape of the eyeball and cause damage. At this time, true myopia is formed.

In addition to the influence of the length of the eye axis, after consulting a large amount of data, we found that the influence of eye diopter is not negligible. Generally, for every 1 mm increase in the axial length, the degree of myopia is increased by 300 degrees. cornea

However, the static diopter of other refractive materials does not change much, so the corneal curvature and the eye diopter are positively correlated. Through experimental research, the relationship between diopter and corneal curvature, in myopic refractive error, low myopia is affected by the curvature value of the cornea, and the medium height is less affected by this. At the same time, several experiments have indicated that the axial changes are positively correlated with the increase in myopic diopter. Therefore, as the eye axis continues to lengthen, the eye diopter will also increase. According to the formula we derived, the increase of both will cause the myopia degree to deepen.

$$d_2 = \frac{1 - d_1 l}{(1 - d_1 s) l}$$

In the case of myopia, d1s<1<d1l, the absolute value of the molecule increases, the denominator becomes smaller, and d2 increases).

4. The evaluation of the model

4.1 Advantages of the model

1) The statistical regression model has a higher degree of fit to the real data, and is more applicable to the real scene. The obtained data fitting curve reflects the actual situation of the society and can reflect the changing trend. Solving the parameter matrix in the regression equation by the normal equation method, replacing the commonly used gradient descent method, improving the efficiency of the model

2) In the early warning model, the innovative method of principal component analysis is used to obtain the weight between different features, eliminating the complicated training and backtracking steps of other machine learning algorithms, and the obtained model can match the actual data, and the persuasive power is better. Strong.

3) Naïve Bayesian classifier model applied to this problem has a high classification accuracy of 87%, and the classification is relatively stable, which is very good for small-scale data, suitable for incremental training, not for missing data. Sensitive, the implementation of the algorithm is relatively simple, has a higher degree of fit to solve the target task of this question, and greatly improves the efficiency.

4.2 Disadvantages of the model

1) Because real data is difficult to collect comprehensively, our model lacks analysis of other features, and deviations may occur in some specific cases.

2) The Bayesian classifier needs to know the prior probability, and the prior probability often depends on the hypothesis. There are many models that can be assumed, so in some cases the prediction effect will be poor due to the hypothetical prior model. .

References

[1] Wang Renwei, Zhang Xiaoyun, Li Qing.A Survey of Physical Exercise and Vision, Learning and Sleep Time of Youth in Shanghai[J].Sports Science Research,2008(03):91-93.

[2] Yan Guangli. Relationship between axial length and corneal curvature and refractive power of adolescent myopia [J]. China Journal of Optical Technology, 2017 (09): 137-138.

[3] WANG Zhirong, HE Qizhou, LIU Hanyu, LI Fulin. Relationship between myopic diopter and axial length [J]. Health Vocational Education, 2010, 28(04): 154-155.

[4] Xiao Guoshi. The Causes and Types of Refractive Errors [J]. China Optical Technology Journal, 1996 (02): 65-68.

[5] Li Liang, Xu Jianfang, Lu Yili, Feng Lianshi. Research progress in outdoor activities and physical exercise prevention and control of myopia in children and adolescents [J/OL]. China Physical Education: 1-10 [2019-05-03].https: //doi.org/10.16470/j.csst.2019029.