IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

# Power Loss Classification on Shifts Based on SMS (Singlemode-Multimode-Singlemode) Structured Fiber Optic Using Gaussian Naïve Bayes Method

To cite this article: D H Sulaksono and A C P Siregar 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **462** 012024

IOP Conf. Series: Materials Science and Engineering 462 (2019) 012024 doi:10.1088/1757-899X/462/1/012024

# Power Loss Classification on Shifts Based on SMS (Singlemode-Multimode-Singlemode) Structured Fiber Optic Using Gaussian Naïve Bayes Method

**D H Sulaksono**[1] **and A C P Siregar**[1]

[1]Computer Science Department, Institut Teknologi Adhi Tama Surabaya, Arief Rahman Hakim 100, Surabaya, Indonesia

danang_h_s@yahoo.com

**Abstract.** Singlemode-multimode-singlemode based optical fiber can be used as a good communication medium, but energy or power carried by light will be weakened (losses) due to leakage or due to lack of clarity or shift in optical fibers. In this study, power losses in fiber optics based on SMS will be classified based on changes in the values of power losses to shifts and divided them according to three classes, there are good, average, and bad. The shift will be used as a classification variable that is between 0 µm to 450 µm with an increment of 50 µm for every interval. The SMS optical fiber structure used is 5.5 with 25 attempts on different optical fibers. The classification method used is Naïve Bayes with a Gaussian distribution. Gaussian distribution is used in Naïve Bayes because the dataset will be processed in the form of continuous values. From the results of testing based on TP+TN=6, FP=6 FN =6 on confusion matrix, the classification accuracy value was 42.86%. This indicates that this classification method is still less effective for classifying fiber optic power losses with an SMS structure. For further study, another classification methods can be used in the power loss classification to get better results.

## 1. Introduction

An optical fiber is a flexible, transparent fiber made by making the diameter of the glass (silica) or plastic material smaller than human hair. There are several types of optical fibers, one of which is that the structure has two singlemode fibers (SM) which are axially connected to the end of a fiber optic parabolic multimode (MM) core. Optical fiber with this structure is also called SMS (singlemode-multimode-singlemode). The feasibility of optical fiber is a power loss. Power loss can occur for numbers of reasons, such as the presence of a leak or lack of speed when the light is delivered. A good power loss change is continuous, in other words, when the power loss value goes up and down it will interfere with the optical fiber to communicate when it's used for transmission.

Fiber optic quality against power loss can be determined by a decision-making system, one of which is data mining. Data mining itself is a process for finding patterns from a data set and involves methods that bring machine learning, statistics and databases together. One branch of data mining is classification, where this branch compares a sample data compared to training data. In this study, a classification method with a Naïve Bayes with Gaussian distribution is used to make decisions on the quality of an optical fiber based on power loss. This method is used because the dataset used is a continuous data value.

The dataset itself will be divided into three classes, there are good, average, and bad. The division of these classes is a representation of the value of the change in power loss value at each shift. For the shift variables used in this classification use a distance between 0 µm, 50 µm, 100 µm, 150 µm, 200 µm, 250 µm, 300 µm, 350 µm, 400 µm and 450 µm with a multimode width of 5.5 cm which will be tested on 25 different optical fibers.

## 2. Methods

### *2.1 Naïve Bayes Classifier*

Naive Bayes is a simple probabilistic classification that calculates a set of probabilities by adding up the frequency and combination of values from a given dataset. The algorithm uses Bayes theorem and assumes all independent or non-interdependent attributes given by values in class variables [8]. Another definition says Naive Bayes is a classification with probability and statistical methods proposed by British scientist Thomas Bayes, which predicts future opportunities based on previous experience [1].

Naive Bayes is based on the assumption of simplification that attribute values are conditionally free when given output value. In other words, given the output value, the probability of observing together is a product of individual probabilities [8]. The advantage of using Naive Bayes is that this method only requires a small amount of training data to determine the parameter estimates required in the classification process. Naive Bayes often works much better in most complex real-world situations than expected [2]. Naïve Bayes formula is written as follows :

$$\boldsymbol{p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}} \tag{1}$$

where $p(x|C_k)$or can be called posterior probability is the probability of the $C_k$ hypothesis based on condition $x$. $p(C_k)$ is also called prior, the probability of a hypothesis in class $C_k$. In this study, prior uses three classes determined from the results of fiber optic power loss values, which are good, medium and bad. For $p(x|C_k)$ called likelihood, the likelihood value is the result of the vector between all variables with each class. There are 10 variables used, where all of these variables are the result of shifting optical fibers, namely 0 µm, 50 µm, 100 µm, 150 µm, 200 µm, 250 µm, 300 µm, 350 µm, 400 and 450 µm. Whereas for $p(x)$ is usually referred to as marginal likelihood or evidence, which is the sum of all $p(C_k)p(x|C_k)$.

### *2.2 Gaussian Distribution*

As the value of features, the power loss data used is continuous, so it can also be assumed that the continuous values associated with each class will be distributed by using Gaussian distribution. This Gaussian distribution is used to find the value of $p(x|C_k)$. with the following equation:

$$p(x = v \mid Ck) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}} \tag{2}$$

Where, $x$ is a training data set that has a continuous attribute, in this case are the values of power loss. Whereas, µk is the average value of the $x$ value that is assosiated in each class of $C_k$, with the following equation:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

For $\sigma_k^2$ is a variance value that calculates how far a set of numbers is scattered from the average value. The result $\sigma_k^2$ is obtained from the value of x which is assosiated by class $C_k$, the variance formula is as follows :

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \pi)^2 \quad (4)$$

In Gaussian distribution, $v$ value itself is the sample data that is tested against training data, so the probability of the distribution of the $x$ value to the class $C_k$ that is $p(x = v \mid Ck)$, can be calculated by entering the value $v$ into a normal distribution which is the value of $\mu_k$ and $\sigma_k^2$ are used as parameters.

## 3. Results And Discussions

### *3.1 Training Data*

There was a set of tests to get the power loss values which are used as training data. The variables used in this study are the value of shifting in optical fibers with a distance between 0µm to 450µm at intervals of 50µm at each shift.

**Table 1.** Power Loss Training Data

| **0µm** | **50µm** | **100µm** | **150µm** | **200µm** | **250µm** | **300µm** | **350µm** | **400µm** | **450µm** | **$C_k$** |
|---|---|---|---|---|---|---|---|---|---|---|
| 3.5704 | 3.7543 | 3.8642 | 3.8789 | 4.1258 | 4.3584 | 4.3698 | 4.4568 | 4.477 | 4.632 | good |
| 3.629 | 3.6185 | 3.7541 | 3.7642 | 3.5575 | 3.5499 | 3.4864 | 3.5815 | 3.6914 | 3.5889 | bad |
| 3.3302 | 3.469 | 3.407 | 3.4536 | 3.4017 | 3.5543 | 3.5106 | 3.6344 | 3.5129 | 3.6598 | bad |
| 3.2584 | 3.2687 | 3.2989 | 3.3112 | 3.3189 | 3.3338 | 3.3476 | 3.3491 | 3.4589 | 3.5686 | average |
| 3.6777 | 3.7123 | 3.7521 | 3.8195 | 3.8721 | 3.9013 | 4.0781 | 4.012 | 4.1445 | 4.152 | good |
| 3.6734 | 3.6772 | 3.6851 | 3.7121 | 3.7431 | 3.7486 | 3.8112 | 3.9112 | 3.9211 | 3.9389 | average |
| 3.6471 | 3.6511 | 3.6576 | 3.7214 | 3.7354 | 3.7231 | 3.7395 | 3.7415 | 3.7613 | 3.7781 | average |
| 3.715 | 3.7421 | 3.7561 | 3.7588 | 3.8471 | 3.851 | 3.8563 | 3.8761 | 3.8845 | 3.895 | average |
| 3.5671 | 3.621 | 3.7612 | 3.8121 | 3.8812 | 4.0214 | 4.135 | 4.2158 | 4.2568 | 4.3541 | good |
| 3.5891 | 3.5196 | 3.5527 | 3.5247 | 3.5595 | 3.638 | 3.4998 | 3.5931 | 3.4094 | 3.4135 | bad |
| 3.3164 | 3.1236 | 3.4187 | 3.7392 | 3.7658 | 3.6905 | 3.655 | 3.7828 | 3.7823 | 3.8705 | bad |
| 3.6143 | 3.6517 | 3.6778 | 3.6989 | 3.7124 | 3.731 | 3.7342 | 3.7421 | 3.748 | 3.7681 | average |
| 3.5761 | 3.5833 | 3.6122 | 3.6816 | 3.7151 | 3.792 | 3.9511 | 3.9621 | 3.978 | 4.0321 | good |
| 3.514 | 3.5251 | 3.5326 | 3.5387 | 3.5443 | 3.5124 | 3.5161 | 3.5312 | 3.5452 | 3.571 | average |
| 3.5925 | 3.6834 | 3.9877 | 4.2511 | 3.9734 | 3.9553 | 3.9355 | 3.9549 | 4.1345 | 3.9981 | bad |
| 3.6122 | 3.7011 | 3.7612 | 3.7732 | 3.8122 | 3.8418 | 3.912 | 4.0181 | 4.081 | 4.1225 | good |
| 3.512 | 3.587 | 3.4126 | 3.4137 | 3.523 | 3.5112 | 3.5145 | 3.6541 | 3.6127 | 3.5124 | bad |
| 3.7127 | 3.7124 | 3.701 | 3.8212 | 3.7816 | 3.7986 | 3.7812 | 3.7932 | 3.7986 | 3.8124 | bad |
| 3.8112 | 3.8162 | 3.821 | 3.841 | 3.8472 | 3.851 | 3.8612 | 3.8713 | 3.8811 | 3.8826 | average |
| 3.6315 | 3.6608 | 3.543 | 3.5337 | 3.5911 | 3.6138 | 3.662 | 3.2142 | 3.3336 | 3.4867 | bad |
| 3.512 | 3.5262 | 3.5413 | 3.5512 | 3.584 | 3.6121 | 3.6268 | 3.6321 | 3.6326 | 3.6413 | average |
| 3.5126 | 3.5268 | 3.6171 | 3.6577 | 3.7124 | 3.7223 | 3.7431 | 3.7851 | 3.8781 | 3.9511 | good |
| 3.6471 | 3.6321 | 3.721 | 3.8454 | 3.8732 | 3.9413 | 4.351 | 4.4588 | 4.5126 | 4.5324 | good |
| 3.6788 | 3.7246 | 3.8412 | 3.8962 | 3.9131 | 3.9463 | 3.9821 | 4.0511 | 4.121 | 4.1322 | good |
| 3.5172 | 3.611 | 3.621 | 3.6289 | 3.7031 | 3.7358 | 3.7481 | 3.7921 | 3.797 | 3.813 | good |

### *3.2 Mean And Variance For Each Class*

After the training data is obtained, the calculation of the mean $\mu_k$ and variance $\sigma_k^2$ is done on the $x$ values that associated with the values of each class $C_k$. These values are used to provide parameters to the set of training data $x$ against the values of the test data $v$. The results of the mean and variance in each class $C_k$ are shown in table 2, table 3, and table 4.

**Table 2.** Mean And Variance for $C_k$=good

| | 0µm | 50µm | 100µm | 150µm | 200µm | 250µm | 300µm | 350µm | 400µm | 450µm |
|---|---|---|---|---|---|---|---|---|---|---|
| **mean** | 3.595 | 3.6518 | 3.7279 | 3.777 | 3.8453 | 3.9178 | 4.030 | 4.083 | 4.138 | 4.191 |
| **variance** | 0.00786 | 0.0113 | 0.0178 | 0.0193 | 0.0354 | 0.0748 | 0.1040 | 0.1242 | 0.1205 | 0.1432 |

**Table 3.** Mean And Variance for $C_k$=average

| | 0µm | 50µm | 100µm | 150µm | 200µm | 250µm | 300µm | 350µm | 400µm | 450µm |
|---|---|---|---|---|---|---|---|---|---|---|
| **mean** | 3.5931 | 3.607 | 3.621 | 3.641 | 3.6665 | 3.6703 | 3.686 | 3.706 | 3.7290 | 3.7554 |
| **variance** | 0.0492 | 0.0496 | 0.0462 | 0.04901 | 0.05513 | 0.0548 | 0.0568 | 0.06607 | 0.0504 | 0.0379 |

**Table 4.** Mean And Variance for $C_k$=bad

| | 0µm | 50µm | 100µm | 150µm | 200µm | 250µm | 300µm | 350µm | 400µm | 450µm |
|---|---|---|---|---|---|---|---|---|---|---|
| **mean** | 3.539 | 3.5467 | 3.597 | 3.6876 | 3.6441 | 3.663 | 3.6306 | 3.6510 | 3.6594 | 3.667 |
| **variance** | 0.0365 | 0.0629 | 0.0737 | 0.135 | 0.0812 | 0.0388 | 0.0457 | 0.0561 | 0.0517 | 0.0747 |

### *3.3 Test Data*

After getting the mean $\mu_k$ and variance values $\sigma_k^2$, the next step is to calculate the likelihood value $p(x|C_k)$ for each $C_k$class using Gaussian distribution. A set of power loss test data is substituted into the variable $v$ in the Gaussian distribution equation. The likelihood value $p(x|C_k)$ of each class that produced by Gaussian distribution is used to classify the sample data to produce a classified test data. Test data that will be used for classification testing are shown in table 5.

**Table 5.** The Results Of Classified Test Data

| 0µm | 50µm | 100µm | 150µm | 200µm | 250µm | 300µm | 350µm | 400µm | 450µm | $C_k$ | Actual |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.6471 | 3.6511 | 3.6576 | 3.7214 | 3.7354 | 3.7231 | 3.7395 | 3.7415 | 3.7613 | 3.7781 | average | average |
| 3.6016 | 3.6014 | 3.601 | 3.7101 | 3.6705 | 3.6875 | 3.6701 | 3.6821 | 3.6875 | 3.7014 | average | bad |
| 3.5360 | 3.5210 | 3.6209 | 3.7343 | 3.7621 | 3.8302 | 4.2408 | 4.3477 | 4.4015 | 4.4214 | good | good |
| 3.2053 | 3.0125 | 3.3077 | 3.6281 | 3.6547 | 3.5811 | 3.5454 | 3.6717 | 3.6712 | 3.7641 | average | bad |
| 3.6016 | 3.6015 | 3.610 | 3.7101 | 3.6705 | 3.6875 | 3.6701 | 3.6832 | 3.6875 | 3.7011 | average | bad |
| 3.604 | 3.6310 | 3.6450 | 3.6477 | 3.7360 | 3.7401 | 3.7452 | 3.7650 | 3.7634 | 3.784 | average | average |
| 3.5704 | 3.7543 | 3.8642 | 3.8789 | 4.1258 | 4.3584 | 4.3698 | 4.4568 | 4.477 | 4.632 | good | good |
| 3.5623 | 3.5661 | 3.5740 | 3.6510 | 3.6320 | 3.6374 | 3.7001 | 3.8001 | 3.8122 | 3.8278 | bad | average |
| 3.5172 | 3.611 | 3.621 | 3.6289 | 3.7031 | 3.7358 | 3.7481 | 3.7921 | 3.797 | 3.813 | good | good |
| 3.7127 | 3.7124 | 3.701 | 3.8212 | 3.7816 | 3.7986 | 3.7812 | 3.7932 | 3.7986 | 3.8124 | bad | bad |

### 3.4 *Classification Accuracy and precission*

Confusion matrix is used for measuring the effectiveness of classification. By using Gaussian Naïve Bayes classification, Classification Accuracy and precission calculation is performed for 10 test data against training data.

**Table 6.** Confusion Matrix

| | | Prediction | | | |
|---|---|---|---|---|---|
| | | **Good** | **Average** | **Bad** | **FN** |
| **Actual** | **Good** | 3 | 0 | 0 | 0 |
| | **Average** | 0 | 2 | 1 | 1 |
| | **Bad** | 0 | 3 | 1 | 3 |
| | **FP** | 0 | 3 | 1 | |

From the table above we have obtained the value of TP+TN=6, FP=6 FN =6, so that the accuracy can be calculated as :

$$Accuracy = \frac{6}{6+4+4} * 100\% = 42.86\% \quad (5)$$

## 4. Conclusions

The classification accuracy calculation using Gaussian Naïve Bayes classifier method is less effective when it used for automatic classification of power loss data on optical fiber. The accuracy is still below 50% using the effectiveness measurement Confusion Matrix. For further study, another classification methods can be used in the power loss classification to get better results.

## 5. References

[1] Bustami 2013 Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. TECHSI : Jurnal Penelitian Teknik Informatika. Vol. 3. No.2. Hal. 127-146.

[2] Pattekari S A and Parveen A 2012 Prediction System for Heart Disease Using Naive Bayes. International Journal of Advanced Computer and Mathematical Sciences. ISSN 2230-9624. Vol. 3. No3, 290-294.

[3] Lou W, Wang X, Chen F, Chen Y, Jiang B and Zhang H 2014 Sequence Based Prediction of DNA-Binding Proteins Based on Hybrid Feature Selection Using Random Forest and Gaussian Naïve Bayes. PLoS ONE 9(1): e86703.

[4] Ghaith A, Siti S K and Husniza H 2018 Conceptual Framework for Stock Market Classification Model Using Sentiment Analysis on TwitterBased on Hybrid Naïve Bayes Classifiers. International Journal of Engineering & Technology. 7(2.14) (2018) 57-61.

[5] Dishant K and Arunima S 2018 Kernel-Based Naive Bayes Classifier for Medical Predictions. Advances in Intelligent Systems and Computing. Volume 695.

[6] Raizada RDS and Lee Y-S 2013 Smoothness without Smoothing: Why Gaussian Naive Bayes Is Not Naive for Multi-Subject Searchlight Studies. PLoS ONE 8(7): e69566.

[7] Ziemann and Olaf et al 2008 POF Handbook-Optical Short Range Transmission Systems. Springer : Berlin.

[8] Nian Z, Jiang X, Jing Z and Keenan L 2018 Gaussian Process Regression Method for Classification for High-Dimensional Data with Limited Samples. IEEE. ISSN: 2573-3311. DOI: 10.1109/ICIST.2018.8426077.

[9] Aslam C P S and Danang H S 2017 Perancangan Sensor Suhu Menggunakan Metode Interpolasi Lagrange Berbasis Serat Optik Berstruktur Sms (Singlemode-Multimode-Singlemode). JEEE UMSIDA. Vol. 1 No. 2. DOI: https://doi.org/10.21070/jeee-u.v1i2.976.

[10] Aslam C P S and Danang H S 2017 Perancangan Sensor Pergeseran Menggunakan Metode Interpolasi Lagrange Berbasis Serat Optik Berstruktur Sms (Singlemode-Multimode-Singlemode). SNTEKPAN. C-123. ISBN : 978-602-98569-1-0.