MT Method for Anomaly Detection and Classification using EM-λ Algorithm

DOI: 10.12776/qip.v28i1.1964

Katsuhiko Tateishi, Hiroki Iwamoto, Shinto Eguchi, Yasushi Nagata

Received: 2023-12-21 Accepted: 2024-03-06 Published: 2024-03-31

ABSTRACT

Purpose: In this paper, we propose a method to classify and detect normal, known anomalies, and unknown anomalies by combining the expectation–maximisation (EM-λ) algorithm and the Mahalanobis–Taguchi (MT) method.

Methodology/Approach: The proposed method learns normal data that are expected to be homogeneous and known abnormal data and performs classification and detection by parameter estimation using the EM-λ algorithm. Conventional methods perform analysis based on parameter estimation using the EM algorithm. However, the EM algorithm can degrade classification accuracy if it does not assume that the data fits the model's generative process.

Findings: We verify the performance of the proposed method using artificially generated data and real-world bean data for classification as data that do not satisfy this assumption. The validation results show up to 6% improvement over the conventional method in classification accuracy and unknown anomaly discrimination accuracy.

Research Limitation/implication: We try various patterns for the parameter of the proposed method in the verification. However, this way is computationally expensive.

Originality/Value of paper: Conventional methods perform analysis based on parameter estimation using the EM algorithm. Our proposal method seeks to improve accuracy by using the EM-λ algorithm for parameter estimation, which is expected to improve classification accuracy when the data do not conform to the generative assumptions of the EM algorithm's model.

Category: Conceptual paper

Keywords: MT method; Anomaly detection; Semisupervised Learning, Classification, EM-λ algorithms

1 INTRODUCTION

The Mahalanobis–Taguchi (MT) method is a representative method in the MT system (Taguchi and Jugulum, 2002), which is a system of pattern recognition and prediction techniques proposed by Taguchi. Examples of MT applications include cooling fan anomaly detection (Jin and Chow, 2013) and financial crisis prediction (Lee and Teng, 2009).

In the MT method, a homogeneous population is defined as a unit space with respect to the objective, and the analysis is performed based on the Mahalanobis distance from the center of the unit space to the target data. Homogeneity can be defined as a normal or steady state (Tatebayashi et al., 2008). For example, when using the MT method for anomaly detection, data that do not belong to a unit space can be identified as abnormal by forming a unit space with normal data.

In the MT method, the normal state is often used as the unit space. However, according to Tatebayashi et al. (2008), the abnormal state can also be defined as the unit space if information on all abnormalities is available. For example, ball bearing anomalies can be classified into several types of anomalies, such as flaking (fatigue failure), seizure, and cracking (Soylemezoglu et al., 2010). Thus, when homogeneity can be expected for each anomaly (known anomalies), a unit space can be defined for each anomaly. Moreover, based on the Mahalanobis distance from each unit space, the normality of the target data and the type of anomaly can be simultaneously determined. If the Mahalanobis distance is far from any of the unit spaces, the anomaly may be an "unknown anomaly" without a known pattern.

Therefore, the purpose of this study is to propose a new MT analysis process that can classify normal, known abnormal, and unknown abnormal data when normal data and a specific pattern of abnormal data are obtained.

The remainder of this paper is organised as follows: Section 2 describes the related work. Section 3 presents the algorithm of the proposed method, and Section 4 discusses the simulation using artificial data. Section 5 discusses the simulation using real-world data. Finally, Section 6 summarises the conclusions and future work.

2 RELATED WORK

2.1 EX-MT Method

Honma et al. (2022) proposed an extended MT (EX-MT) method that can classify and detect "normal", "known abnormalities", and "unknown abnormalities", when homogeneity can be expected for multiple normal and abnormal patterns.

The EX-MT method defines an "extended unit space" that handles multiple unit spaces simultaneously. Assuming K unit spaces, the density around x in the extended unit space follows the probability density function of the Gaussian

mixture model (GMM), and we detect and classify anomalies based on the model. The extended unit space model is as follows:

$$p(x) = \sum_{k=1}^{K} \pi_k \, N(x|\mu_k, \Sigma_k), \tag{2.1}$$

where $N(x|\mu_k, \Sigma_k)$ is a multivariate normal distribution that generates individuals with class label k, μ_k is the population mean vector of class k, Σ_k is the population covariance matrix of class k, and π_k is the population mixture ratio of class k.

In the abnormality detection phase, the abnormality $a(x)$ of individual x is calculated to detect "unknown abnormalities." In the classification phase, the clustering function is used to classify x, which is determined to be an individual belonging to the extended unit space, into K classes of "normal" and "known abnormalities." The formulas for the abnormality $a(x)$ and clustering function are given in Equations (3.2) and (3.3), respectively, in Section 3.2.

Next, we explain the parameter estimation method of the EX-MT method, which requires the estimation of the parameter $\theta = \{\pi_k, \mu_k, \sum_k\}_{k\in\{1,2,\ldots,K\}}$. Two estimation methods have been proposed: one using only labelled data and the other using labelled and unlabelled data. The EX-MT method based on supervised learning is called the EX-MT(SL) method, and the EX-MT method based on semisupervised learning is called the EX-MT(SSL) method. In the EX-MT(SL) method, parameter θ is estimated to maximise the likelihood of the training data. The specific formulas are described in Equations (3.4)–(3.7) in Section 3.2. Meanwhile, the EX-MT(SSL) method specifies parameter θ by maximum likelihood using the expectation–maximisation (EM) algorithm based on Zhu and Goldberg (2009). The specific formula is given by substituting $\lambda = 1$ in Equations (3.8)–(3.13) in Section 3.2.

2.2 EM-λ algorithm

Nigam et al. (2000) proposed two improvements to the EM algorithm for semisupervised learning in text classification. One is the EM-λ algorithm, which assigns a weight λ only to unlabelled data in the EM algorithm, thus allowing the influence of unlabelled data to be adjusted.

The background for this weighting is related to a problem of the EM algorithm. The EM algorithm in semisupervised learning assumes that labelled and unlabelled data are generated from the same model. However, this assumption is often not true in the real world (e.g., in text classification). Therefore, we assign a weight λ to real data for which the process of the EM algorithm does not hold to make the unlabelled data valid.

3 PROPOSED METHOD

3.1 Analysis Procedure

In the proposed method, as in Honma et al. (2022), we assume conditions under which homogeneity can be expected for each of multiple normalities and known anomalies. We define an "extended unit space" that treats K unit spaces simultaneously. We assume that the marginal probability density of x in the extended unit space follows the probability density function of the GMM, and we detect and classify the anomalies based on the model. The model of the extended unit space is as follows:

$$p(x) = \sum_{k=1}^{K} \pi_k \, N(x|\mu_k, \Sigma_k), \tag{3.1}$$

where $N(x|\mu_k, \Sigma_k)$ is a multivariate normal distribution that generates individuals with class label k, μ_k is the population mean vector of class k, Σ_k is the population covariance matrix of class k, and π_k is the population mixture ratio of class k. The specific analysis procedure is described below. The analysis consists of the following two steps.

Step1: Detection of unknown anomalies

Here, individuals with a small probability of belonging to the extended unit space (outliers) are detected as "unknown anomalies." An individual x' is judged as an outlier when its abnormality value calculated using the following Equation is large:

$$a(x') = -\log \sum_{k=1}^{K} \hat{\pi}_k N\left(x' \middle| \hat{\mu}_k, \hat{\Sigma}_k\right). \tag{3.2}$$

Here, $\hat{\theta}_k = \{\hat{\pi}_k, \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k\}$ is the parameter estimator in class k of the extended unit space. In addition, when $K = 1$, $a(x')$ is an anomaly measure equivalent to the Mahalanobis distance, which is an anomaly measure in the MT method.

Step2: Classification in the extended unit space

Here, for each individual x' determined to belong to the extended unit space in Step1, a *K*-class classification of "normal" and "known abnormality" is performed. This classification is based on the clustering function shown in the following Equation:

$$p(y = k|\boldsymbol{x}') = \frac{\hat{\pi}_k N\left(\boldsymbol{x}' \middle| \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k\right)}{\sum_{j=1}^{K} \hat{\pi}_j N\left(\boldsymbol{x}' \middle| \hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\Sigma}}_j\right)}. \tag{3.3}$$

3.2 Parameter Estimation

The proposed method estimates the parameters $\theta = \{\pi_k, \mu_k, \sum_k\}_{k\in\{1,2,\ldots,K\}}$ of the extended unit space. The estimation method is based on the EX-MT method

proposed by Honma et al. (2022) and the EM-λ algorithm proposed by Nigam et al. (2000).

The estimation consists of two main steps. First, parameter θ is estimated by maximum likelihood estimation using only labelled data. Second, using this value as the initial value, the estimated value is calculated based on the EM-λ algorithm using labelled and unlabelled data.

We assume that the sample size l and $(p+1)$ -dimensional data set $D_l = \{x_i, y_i\}_{i=1}^{l}$ with the class label indicating the "normal" or "known anomaly" labels and the sample size u and p-dimensional data set $D_u = \{x_i\}_{i=l+1}^{l+u}$ with no label are observed as the training data. In this case, the estimation process of the proposed method is as follows.

Step 1: Estimation of initial values using labelled data

The likelihood of training data D_l with labels can be expressed as follows using the simultaneous density $p(x_i, y_i|\theta)$ and marginal likelihood $p(x_i|\theta)$:

$$p(D_l|\theta) = \prod_{i=1}^{l} p\left(x_i, y_i\middle|\theta\right) = \prod_{i=1}^{l} \pi_{y_i} N(x_i|\mu_{y_i}, \sum_{y_i}), \tag{3.4}$$

where $\theta = \{(\pi_k, \mu_k, \sum_k) : k = 1, \dots, l\}$.

The parameter $\theta_k = \{\pi_k, \mu_k, \sum_k\}$ for class k that maximises the log-likelihood in Equation (3.4) can be estimated as follows:

$$\hat{\mu}_k = \frac{1}{l_k} \sum_{i=1}^{l_k} x_i\,, \tag{3.5}$$

$$\hat{\sum}_k = \frac{1}{l_k} \sum_{i=1}^{l_k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T, \tag{3.6}$$

$$\hat{\pi}_k = \frac{l_k}{l}. \tag{3.7}$$

Step 2: Estimation using labelled and unlabelled data

The EM-λ algorithm is executed using the labelled data D_l and unlabelled data D_u, with the estimates calculated from the labelled training data D_l as initial values. The training data consist of $D = \{D_l + D_u\}$ with sample size $l + u$.

The log-likelihood of training data D is as follows:

$$\log p(D|\theta) = \sum_{i=1}^{l} \log p(y_i|\theta) p(x_i|y_i, \theta) + \lambda \sum_{i=l+1}^{l+u} \log \sum_{k=1}^{K} \pi_k N(x_i|\mu_k, \Sigma_k)\,. \tag{3.8}$$

The EM-λ algorithm is used to estimate the parameter θ that maximises the log-likelihood. First, the value of the maximum likelihood estimator $\theta^{(0)} =$

$\{\pi_k^{(0)}, \mu_k^{(0)}, \sum_k^{(0)}\}_{k\in\{1,2,\dots,K\}}$ calculated using labelled data D_l is obtained as the initial value ($t = 0$).

Second, for the E-step, we compute the following burden ratios $r_{ik}, i \in \{l+1, l+2, \dots, l+u\}$, $k \in \{1, 2, \dots, K\}$ for unlabelled data D_u using the current parameters.

$$\gamma_{ik} = \frac{\pi_k^{(t)} N\left(x_i \middle| \mu_k^{(t)}, \sum_k^{(t)}\right)}{\sum_{j=1}^{K} \pi_j^{(t)} N(x_i | \mu_j^{(t)}, \sum_j^{(t)})}. \tag{3.9}$$

Note that the burden ratio $r_{ik}, i \in \{1,2, \dots, l\}$ of labelled data D_l is 1 if $y_i = k$ and 0 otherwise.

Third, in the M-step, the parameter $\theta^{(t+1)}$ is updated using the burden rate γ_{ik} of training data D as follows:

$$l_k = \sum_{i=1}^{l} \gamma_{ik} + \lambda \sum_{i=l+1}^{l+u} \gamma_{ik}\,, \tag{3.10}$$

$$\mu_k^{(t+1)} = \frac{1}{l_k}\left(\sum_{i=1}^{l} \gamma_{ik} x_i + \lambda \sum_{i=l+1}^{l+u} \gamma_{ik} x_i)\right), \tag{3.11}$$

$$\Sigma_k^{(t+1)} = \frac{\begin{matrix}\sum_{i=1}^{l} \gamma_{ik}(x_i - \mu_k^{(t+1)})(x_i - \mu_k^{(t+1)})^T + \\ \lambda \sum_{i=l+1}^{l+u} \gamma_{ik}\,(x_i - \mu_k^{(t+1)})(x_i - \mu_k^{(t+1)})^T\end{matrix}}{l_k}, \tag{3.12}$$

$$\pi_k^{(t+1)} = \frac{l_k}{l + \lambda u}. \tag{3.13}$$

The E-step and M-steps are repeated until the log-likelihood $\log p(D|\theta)$ converges. Then, the analysis process described in Section 3.1 is performed using the converged θ.

4 ACCURACY COMPARISON USING ARTIFICIAL DATA

In this section, we verify the effectiveness of the proposed method by comparing its accuracy with that of conventional methods using artificial data. The conventional method is the EX-MT method proposed by Honma et al. (2022).

4.1 Simulation Settings

We generate the dataset used in this experiment with the following settings and perform 1000 simulations. For the training data, we use labelled data generated to follow a multivariate normal distribution and unlabelled data generated to follow a multivariate t-distribution. We obtain the t-distribution with location parameter μ, positive definite inner product matrix $\sum$, and γ degrees of freedom,

$$(y;\ \mu, \sum, \gamma) = \frac{\Gamma\left(\frac{\gamma + p}{2}\right) |\sum|^{-\frac{1}{2}}}{(\pi\gamma)^{\frac{1}{2}}\Gamma\left(\frac{\gamma}{2}\right)}\left\{1 + \frac{1}{\gamma}(y - \mu)^T\sum^{-1}(y - \mu)\right\}^{-\frac{\gamma+p}{2}}, \quad (4.1)$$

where $y \in \mathbb{R}^p$ (Peel and McLachlan, 2000). In this experiment, the multivariate t-distribution data are generated based on the method described by Genz et al. (2023).

The parameters of the labelled data are as follows: $p = 10, K = 3, \mu_1 = (0,0, \dots ,0)^T, \mu_2 = (1.5,1.5, \dots ,1.5)^T, \mu_3 = (-1.5, -1.5, \dots , -1.5)^T, \sum_1 = \sum_2 = \sum_3$. The diagonal components are set to 1, while the rest of the components are set to 0.3. $\pi_1 : \pi_2 : \pi_3 = \frac{9}{12} : \frac{2}{12} : \frac{1}{12}$. In this experiment, we assume that "normal," "known anomaly," and "known anomaly 2" are generated when $k = 1$, $k = 2$, and $k = 3$, respectively.

The parameters $\mu, \sum$ and π of the unlabelled data are the same as those for the labelled data, and the number of data is assumed to be 3,600. Note that the covariance matrix of the t-distribution is $\frac{\gamma}{\gamma - 2}\sum$ with γ degrees of freedom.

Three types of test data are prepared: threshold, classification and evaluation, and unknown abnormality discrimination data. All variables are generated to follow a multivariate normal distribution. The parameters of the threshold data and data for classification and evaluation are the same as those of the data with labels. Meanwhile, the parameters of the data for discriminating unknown abnormalities are the same as those of the data with labels, that is, mother mixture ratio, mother mean vector, and mother covariance matrix multiplied by 4. The number of data for all the test data is 1,000.

4.2 Evaluation Criteria

The evaluation indices used are the "unknown abnormality discrimination rate," which is the percentage of unknown abnormalities detected; and $macroF1$, which checks whether normal or known abnormal data can be classified. Each indicator takes a value between 0 and 1; the closer it is to 1, the better.

$macroF1$ is calculated by the following Equation:

$$macroF1 = \frac{1}{K}\sum_{k=1}^{K} F1_k\,. \quad (4.2)$$

$F1_k$ of class k is calculated as follows:

$$F1_k = \frac{2 \times Precision_k \times Recall_k}{Precision_k + Recall_k}. \quad (4.3)$$

Note that $Precision_k$ indicates the proportion of individuals classified as class k that are actually class k, and $Recall_k$ indicates the proportion of individuals that are actually class k that can be correctly classified as class k. See Yang and Liu, 1999 for details on $macroF1$.

The unknown abnormality discrimination rate is the percentage of "test data for the unknown abnormality discrimination" that are correctly detected as an unknown abnormality.

The threshold value is set such that the normal discrimination rate in the extended unit space of the test data (for the threshold setting) is 99%. In other words, the rate of misjudgment of the extended unit space as "unknown abnormality" is 1%.

4.3 Impact of the Number of Labelled Data

In this experiment, we compare the accuracies of the proposed and conventional methods by varying the number of labelled data. The data settings are those described in Section 4.1. The number of data with labels (l) is set to 180, 360, 630, 900, and 1200, while the degree of freedom of the t-distribution of the unlabelled data is fixed at 20.

Under each condition, the parameter of the proposed method, that is, the "unlabelled data weight λ," is moved by 0.1 from 0 to 1, where $\lambda = 0$ and $\lambda = 1$ are respectively synonymous with the EX-MT(SL) and EX-MT(SSL) methods proposed by Homma et al. (2022). In other words, if the evaluation index is maximised when the weight λ is in $\{\lambda : 0 < \lambda < 1\}$, we can conclude that the proposed method is effective.

The average values of each indicator over 1,000 simulations are shown in Figures 1 and 2.

First, considering Figure 1, in the case of $l = 360$, the proposed method shows $macroF1$ value approximately 6% higher than the comparative methods ($\lambda = 0{,}1$), making it the most improved condition. Furthermore, for other values of l, improvements of over 1% are observed at $l = 630, 900$, while $l = 180$ and $l = 1200$ show performance levels similar to the comparative methods. Based on these results, the proposed method is considered more effective in terms of classification accuracy than the comparative methods. One possible factor contributing to this improvement is the alteration of the occurrence distribution of labelled and unlabelled data. Since the assumption of the EM algorithm regarding the generation from the same distribution was not met, the EM-λ algorithm, which can adjust the influence of unlabelled data, was able to estimate parameters more effectively, leading to the improvement in $macroF1$ values.

Additionally, when comparing the degree of improvement for the number of labelled data (l), the proposed method showed smaller improvements when l was too large or too small. This is attributed to the fact that with a large l, estimation using only labelled data is effective, whereas, with a small l, sufficient information cannot be obtained from labelled data, hindering the effective utilisation of unlabelled data.

Next, considering Figure 2, a trend is observed across all patterns of labelled data (l) in which the unknown anomaly detection rate remains largely unchanged when λ is not equal to 0. This suggests that the EM-λ algorithm employed by the

proposed method is primarily effective in improving classification accuracy. Consequently, under the experimental conditions of introducing homogenous unknown anomalies in all directions around the extended feature space, the proposed method performs at a level comparable to the comparative methods.

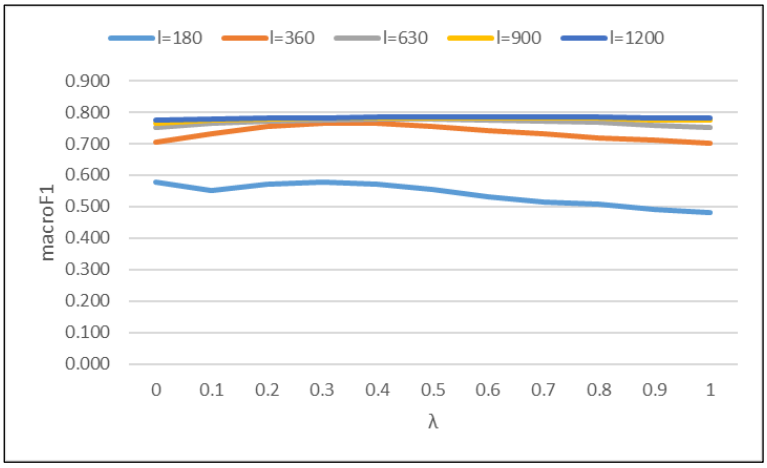

Figure 1 – $macroF1$ by number of labelled data

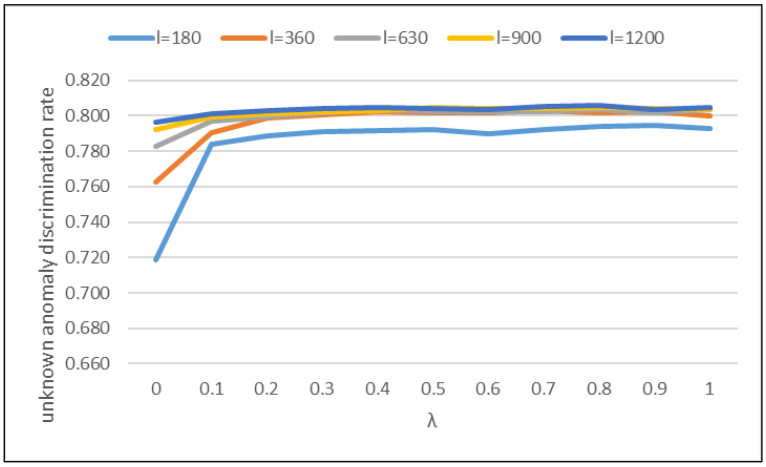

Figure 2 – Unknown anomaly discrimination rate by the number of labelled data

4.4 Impact of Degrees of Freedom of t-Distribution

In this experiment, we compare the accuracies of the proposed and conventional methods by varying the degrees of freedom of the t-distribution of unlabelled data. The data settings used in this experiment are the same as those described in Section 4.1. The number of data with labels is fixed at 360, and the degrees of freedom df of the unlabelled t-distribution is set to 5, 10, 20, 30, and ∞ (normal distribution).

For each condition, the parameter of the proposed method, that is, the "weight of unlabelled data λ," is moved by 0.1 from 0 to 1.

The average values of each indicator over 1,000 simulations are shown in Figures 3 and 4.

Figure 3 shows that the proposed $macroF1$ method is at most 6% better than the EX-MT method ($\lambda = 0{,}1$) for unlabelled data with 20 degrees of freedom, indicating the greatest improvement. When the degrees of freedom are too small or too large, a minimal difference is noted between the proposed and EX-MT methods. This result may be due to the nature of the t-distribution, which approaches a normal distribution as the degrees of freedom increase. When the degree of freedom is 30, the unlabelled data following the multivariate t-distribution are much closer to the multivariate normal distribution, which is the true distribution, and the process of the EM algorithm to treat the unlabelled data as coming from the same model with and without labels is approximately valid. Therefore, the weights of the unlabelled data need not be adjusted, and the proposed method is not completely effective when the degree of freedom is 30. Conversely, when the degree of freedom is 5, the unlabelled data are so far from the true distribution that reducing the influence of unlabelled data does not facilitate classification. When the degree of freedom of the unlabelled data is 5, the value of $macroF1$ decreases as the value of λ increases.

Figure 4 shows almost no performance difference between the proposed method and the EX-MT method ($\lambda = 0{,}1$) in terms of the unknown anomaly discrimination rate. In particular, as in the experiment described in Section 4.3, the assignment of weight λ has little effect on the outer boundary of the entire extended unit space and is not important for the detection of unknown anomalies. Therefore, when unlabelled data are used ($\lambda \neq 0$), the unknown anomaly discrimination rate is close to the value when the weight λ is changed.

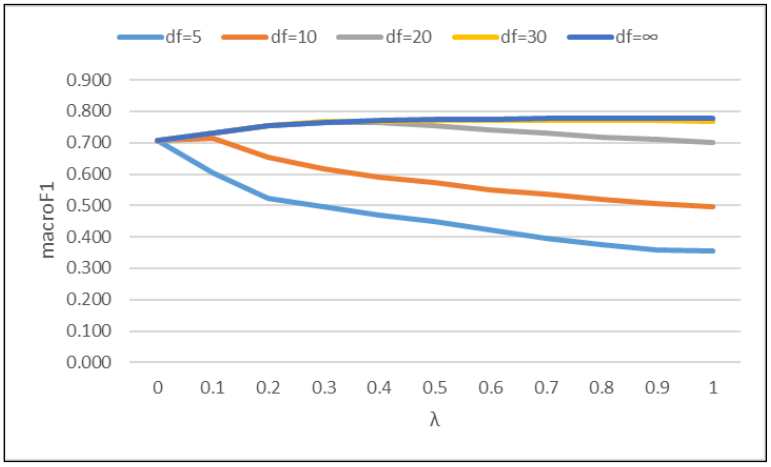

Figure 3 – $macroF1$ by degrees of freedom of t-distribution

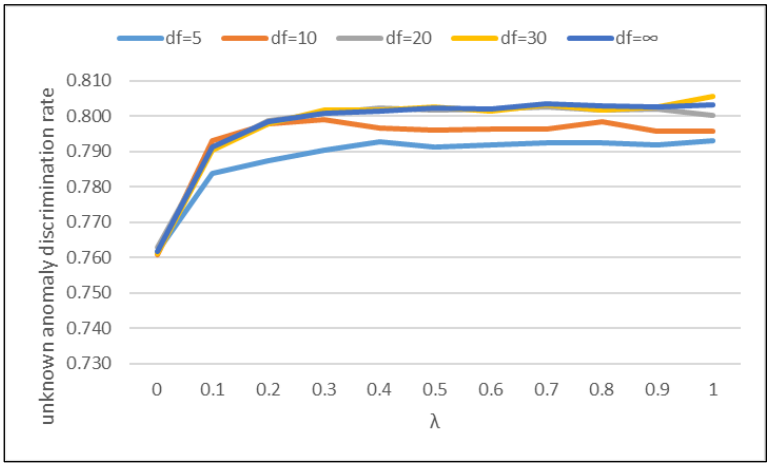

Figure 4 – Unknown anomaly discrimination rate by degrees of freedom of t-distribution

5 ACCURACY COMPARISON USING REAL-WORLD DATA

5.1 Simulation Settings

We use The UCI dry bean data set (UCI, 2020). Thirteen thousand six hundred eleven dry beans of seven types were photographed with a high-resolution camera, and variables related to bean dimensions were obtained. The objective variable is the type of bean data, and the explanatory variables are 16 variables. In this simulation, beans "SIRA" are treated as normal, beans "CALI" as known abnormality 1, and beans "BARBUNYA" as known abnormality 2. For the other beans, "DERMASON" is treated as unknown anomaly 1, "SEKER" as unknown anomaly 2, "HOROZ" as unknown anomaly 3, and "BOMBAY" as unknown anomaly 4.

For the training data, the number of data with labels is 250, and the number of data without labels is 1200, and the data are randomly selected with a population mixing ratio of $\pi_1 : \pi_2 : \pi_3 = \frac{9}{12} : \frac{2}{12} : \frac{1}{12}$. For the test data, the number of data for threshold and classification evaluation is set to 1000 each, and the data are extracted with the same ratio of mother mixing ratios as the training data. The unknown anomaly data are extracted from four types of unknown anomaly beans, each with 500 unknown anomalies.

The evaluation method is the average of the $macroF1$ and unknown anomaly discrimination rates over 1000 simulations. The unknown abnormality discrimination rate is calculated for each of the four types of beans. The threshold for discrimination is set so that the normal discrimination rate in the extended unit space of the test data (for threshold setting) is 99%. The comparison methods are the EX-MT(SL) method ($\lambda = 0$) and the EX-MT(SSL) method ($\lambda = 1$).

5.2 Experimental Results

Table 1 shows the experimental results. Table 1 shows that the proposed method performs best when λ is 0.6. Under this condition, the proposed method outperformed the conventional method in all evaluation indices. In particular, some of the beans showed an improvement of about 4% for unknown anomalies. Unlike the case of artificial data, the unknown anomaly data were distributed in a specific direction, and it is considered that the proposed method was effective for the unknown anomaly.

Table 1 – Comparison of accuracy by λ in real data

λ	Unknown anomaly 1 discrimination rate	Unknown anomaly 2 discrimination rate	Unknown anomaly 3 discrimination rate	Unknown anomaly 4 discrimination rate	$macro\ F1$
0	**0.325**	**0.526**	**0.498**	**0.806**	**0.841**
0.1	0.377	0.578	0.554	0.801	0.842
0.2	0.400	0.598	0.561	0.810	0.843
0.3	0.410	0.614	0.577	0.831	0.842
0.4	0.405	0.603	0.563	0.808	0.839
0.5	0.414	0.614	0.569	0.814	0.837
0.6	**0.436**	**0.642**	**0.600**	**0.852**	**0.844**
0.7	0.407	0.601	0.558	0.808	0.835
0.8	0.422	0.627	0.580	0.835	0.837
0.9	0.419	0.614	0.572	0.824	0.836
1.0	**0.417**	**0.614**	**0.565**	**0.810**	**0.831**

6 CONCLUSION AND FUTURE WORK

In this study, we proposed a method that can classify and detect normal, known anomalies, and unknown anomalies based on parameter estimation using the EM-λ algorithm. Simulations using artificial and real data showed the effectiveness of the proposed method.

Two major issues need to be addressed in the future. The first is how to determine the parameter λ. In this study, we tried various patterns for the parameter λ. The problem is that the parameter λ is computationally expensive. Second, we would like to increase the number of validation data patterns to understand the conditions under which the proposed method is effective. We would like to test other dimensional patterns on artificial data and other data patterns on real data to verify the effectiveness of the proposed method.

ACKNOWLEDGEMENTS

I would like to thank the anonymous referees for their valuable comments.

REFERENCES

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., Maechler, M., and Hothorn, T., 2023. mvtnorm: Multivariate normal and t distributions. [online] Available at: https://cran.r-project.org/web/packages/mvtnorm/mvtnorm.pdf, [Accessed 01 June 2023].

Honma, K., Ohkubo, M., Eguchi, S., and Nagata, Y., 2022. Mahalanobis-Taguchi Method for Anomaly Detection and Classification. *Total Quality Science*, 8(1), 1-13. https://doi.org/10.17929/tqs.8.1.

Jin, X., and Chow, T. W., 2013. Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis–Taguchi system. *Expert Systems with Applications*, 40(15), 5787-5795. https://doi.org/10.1016/j.eswa.2013.04.024.

Lee, Y. C., and Teng, H. L., 2009. Predicting the financial crisis by Mahalanobis–Taguchi system– Examples of Taiwan's electronic sector. *Expert Systems with Applications*, 36(4), 7469-7478. https://doi.org/10.1016/j.eswa.2008.09.037.

Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T., 2000. Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39, 103-134. https://doi.org/10.1023/A:1007692713085.

Peel, D., and McLachlan, G. J., 2000. Robust mixture modelling using the t distribution. *Statistics and computing*, 10, 339-348. https://doi.org/10.1023/A:1008981510081.

Soylemezoglu, A., Jagannathan, S., and Saygin, C., 2010. Mahalanobis Taguchi system (MTS) as a prognostics tool for rolling element bearing failures. *Journal of Manufacturing Science and Engineering*, 132(5). https://doi.org/10.1115/1.4002545.

Taguchi, G., and Jugulum, R., 2002. *The Mahalanobis-Taguchi strategy: A pattern technology system*. Hoboken: John Wiley & Sons.

Tatebayashi, K., Teshima, S., and Hasegawa, R., 2008. *Introduction to MT System*. JUSE Publishing Co. (in Japanese).

UCI, 2020. Dry Bean Dataset. Machine Learning Repository. [online] Available at: https://doi.org/10.24432/C50S4B, [Accessed 2023-12-10].

Yang, Y., and Liu, X., 1999. A re-examination of text categorisation methods. In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 42-49. https://doi.org/10.1145/312624.312647.

Zhu, X., and Goldberg, A. B., 2009. *Introduction to semi-supervised learning*. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Brachman, R.J., Dietterich, T. ed. Kentfield: Morgan and Claypool Publishers, 3(1), 1-130.

ABOUT THE AUTHOR / ABOUT AUTHORS

Katsuhiko Tateishi[0009-0000-7942-7717] (K.T.) – Waseda University, Tokyo, Japan, e-mail: k.tateishi@akane.waseda.jp

Hiroki Iwamoto[0000-0003-2267-6610] (H.I.) – Waseda University, Tokyo, Japan, e-mail: h_iwamoto@aoni.waseda.jp

Shinto Eguchi[0000-0002-3974-618X] (S.E.) – The Institute of Statistical Mathematics, Tokyo, Japan, e-mail: eguchi@ism.ac.jp

Yasushi Nagata[0000-0002-4947-1204] (Y.N.) – Waseda University, Tokyo, Japan, e-mail: ynagata@waseda.jp

AUTHOR CONTRIBUTIONS

Conceptualisation, K.T. and S.E.; Methodology, K.T. and S.E.; software, K.T.; Validation, K.T.; Formal analysis, K.T.; Investigation, K.T.; Resources, K.T.; Data curation, K.T.; Original draft preparation, K.T.; Review and editing, H.I., S.E. and Y.N.; Visualization, K.T.; Supervision, H.I., S.E. and Y.N.; Project administration, Y.N.; Funding acquisition, Y.N.

CONFLICTS OF INTEREST

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.