


修士論文

離散量時系列データの  
隠れマルコフモデルによる  
異常検知システムの検討

平成21年2月13日 提出

指導教員 酒井 信介 教授 

76205 牧野 尚人

## 目次

|          |                        |           |
|----------|------------------------|-----------|
| <b>1</b> | <b>序論</b>              | <b>9</b>  |
| 1.1      | 研究の背景                  | 9         |
| 1.2      | 本研究の目的                 | 10        |
| 1.3      | 対象機器                   | 11        |
| 1.4      | 本報告書の構成                | 12        |
| <b>2</b> | <b>HMM(隠れマルコフモデル)</b>  | <b>14</b> |
| 2.1      | 緒言                     | 14        |
| 2.2      | HMM(隠れマルコフモデル) 概略      | 15        |
| 2.3      | 基礎理論                   | 16        |
| <b>3</b> | <b>異常検知フロー</b>         | <b>20</b> |
| 3.1      | 緒言                     | 20        |
| 3.2      | 異常情報の分類                | 21        |
| 3.2.1    | 動作原因                   | 21        |
| 3.2.2    | 検知すべき異常                | 22        |
| 3.2.3    | 異常発生パターン               | 23        |
| 3.3      | 異常検知フロー                | 25        |
| 3.3.1    | フローチャート                | 25        |
| 3.4      | 異常の検出フロー               | 27        |
| 3.5      | 結言                     | 31        |
| <b>4</b> | <b>フィルターによるスクリーニング</b> | <b>32</b> |
| 4.1      | 緒言                     | 32        |
| 4.2      | フィルター 1st              | 33        |
| 4.3      | フィルター 2nd(統計的フィルター)    | 34        |
| 4.3.1    | 動作時間・動作頻度              | 34        |
| 4.3.2    | 動作時間フィルター              | 35        |
| 4.3.3    | 動作頻度フィルター              | 38        |
| 4.4      | パラメータの設定方法             | 43        |
| 4.4.1    | 分布パラメータの設定             | 43        |
| 4.4.2    | 最尤推定                   | 43        |
| 4.4.3    | ベイズ推定                  | 44        |
| 4.4.4    | 結果、考察                  | 45        |
| 4.5      | 情報の違いによる動作時間、頻度の違いについて | 51        |
| 4.6      | 結言                     | 54        |

|          |                                    |           |
|----------|------------------------------------|-----------|
| <b>5</b> | <b>統計的フィルターの検証</b>                 | <b>55</b> |
| 5.1      | 緒言                                 | 55        |
| 5.2      | 異常時の異常情報                           | 56        |
| 5.3      | 検出限界の検討                            | 58        |
| 5.4      | 検証結果                               | 64        |
| 5.4.1    | 正常踏切                               | 64        |
| 5.4.2    | 異常踏切                               | 66        |
| 5.5      | 結言                                 | 68        |
| <b>6</b> | <b>HMM を用いた異常検知方法</b>              | <b>69</b> |
| 6.1      | 緒言                                 | 69        |
| 6.2      | HMM の選定理由                          | 70        |
| 6.3      | 本研究への適用方法、解析手順                     | 71        |
| 6.3.1    | 適用方法、解析手順                          | 71        |
| 6.3.2    | 適用例                                | 72        |
| 6.3.3    | 閾値設定方法                             | 74        |
| 6.4      | パラメータの設定                           | 76        |
| 6.4.1    | AIC(Akaike's information Criteria) | 76        |
| 6.4.2    | 検証結果、考察                            | 76        |
| 6.5      | HMM での検出限界                         | 78        |
| 6.5.1    | 異常時の異常情報                           | 78        |
| 6.5.2    | 統計的フィルターで検知できない異常                  | 81        |
| 6.6      | 検証結果                               | 84        |
| 6.6.1    | 動作時間異常                             | 84        |
| 6.6.2    | 動作間隔異常                             | 89        |
| 6.6.3    | 動作時間・間隔異常                          | 90        |
| 6.7      | 考察                                 | 94        |
| 6.8      | 結言                                 | 98        |
| <b>7</b> | <b>検知フローに従った検証結果</b>               | <b>99</b> |
| 7.1      | 緒言                                 | 99        |
| 7.2      | 検知の流れ                              | 100       |
| 7.3      | 検証結果                               | 102       |
| 7.3.1    | 統計的フィルター閾値 50%                     | 102       |
| 7.3.2    | 統計的フィルター閾値 90%                     | 102       |
| 7.3.3    | 異常の検知                              | 105       |
| 7.4      | 考察                                 | 110       |
| 7.5      | 結言                                 | 111       |

|      |     |
|------|-----|
| 目次   | 4   |
| 8 結論 | 112 |
| 9 謝辭 | 113 |

目次

|    |   |    |
|----|---|----|
| 1  | probability of independence of observations . . . . .                                   | 17 |
| 2  | learning of HMM . . . . .   | 17 |
| 3  | calculate probability of independence of observation sequence . . . . .                 | 17 |
| 4  | parameter of HMM . . . . .  | 19 |
| 5  | assortment of real anomaly . . . . .  | 24 |
| 6  | flowchart . . . . .   | 26 |
| 7  | pattern0(first anomaly information) . . . . .   | 28 |
| 8  | pattern0(not first,pass statical filter) . . . . .                                      | 28 |
| 9  | pattern0(not first,pass HMM filter) . . . . .   | 29 |
| 10 | pattern1(no indication) . . . . .   | 29 |
| 11 | pattern2(first anomaly information) . . . . .   | 30 |
| 12 | pattern3(caught by statical and HMM filters) . . . . .                                  | 30 |
| 13 | length of working time and interval of working . . . . .                                | 34 |
| 14 | length of working time(Crossing C) . . . . .  | 35 |
| 15 | histogram of length of working time(Crossing B) . . . . .                               | 36 |
| 16 | histogram of length of working time(Crossing D) . . . . .                               | 37 |
| 17 | histogram of length of working time(Crossing G) . . . . .                               | 37 |
| 18 | histogram of interval of working (Crossing B) . . . . .                                 | 39 |
| 19 | histogram of interval of working (Crossing D) . . . . .                                 | 39 |
| 20 | histogram of interval of working (Crossing G) . . . . .                                 | 40 |
| 21 | histogram of inverse of interval (Crossing B) . . . . .                                 | 40 |
| 22 | histogram of inverse of interval (Crossing D) . . . . .                                 | 41 |
| 23 | histogram of inverse of interval (Crossing G) . . . . .                                 | 41 |
| 24 | histogram of inverse of interval (Crossing C) . . . . .                                 | 42 |
| 25 | rate of mis-detection using maximum likelihood estimation . . . . .                     | 46 |
| 26 | rate of detection using Bayesian estimation . . . . .                                   | 47 |
| 27 | rate of detection using maximum likelihood estimation and Bayesian estimation . . . . . | 48 |
| 28 | comparing rate of detection about length of working time . . . . .                      | 49 |
| 29 | comparing rate of detection about interval of working . . . . .                         | 49 |
| 30 | change of parameter of exponential distribution . . . . .                               | 50 |
| 31 | difference of length of working time among some type of information . . . . .           | 52 |
| 32 | difference of length of interval of working among some type of information . . . . .    | 53 |
| 33 | length of working time on Crossing A . . . . .  | 56 |
| 34 | interval of working on Crossing A . . . . .   | 57 |

|    |   |     |
|----|---|-----|
| 35 | length of working time 250%   | 59  |
| 36 | length of working time 500%   | 59  |
| 37 | length of working time 1000%  | 60  |
| 38 | interval of working 1%  | 60  |
| 39 | interval of working 2.5%  | 61  |
| 40 | interval of working 5%  | 61  |
| 41 | interval of working 10%   | 62  |
| 42 | length of working time on Crossing C  | 62  |
| 43 | interval of working on Crossing C   | 63  |
| 44 | after applying filter(length of working time)   | 65  |
| 45 | after applying filter(interval of working)  | 65  |
| 46 | HMM score at Crossing G   | 73  |
| 47 | HMM score at Crossing H   | 73  |
| 48 | HMM score at Crossing G   | 74  |
| 49 | HMM score and threshold level   | 75  |
| 50 | AIC   | 77  |
| 51 | length of working time and theoretical distribution about Crossing D                      | 80  |
| 52 | inverse of interval of working and distribution about Crossing D                          | 80  |
| 53 | type of anomaly   | 81  |
| 54 | anomaly about length of working time  | 82  |
| 55 | anomaly about interval of working   | 82  |
| 56 | histogram of working time of anomaly about length of working time and interval of working | 83  |
| 57 | histogram of interval of anomaly about length of working time and interval of working     | 83  |
| 58 | working time and interval of working  | 86  |
| 59 | working time and interval of working  | 87  |
| 60 | working time and interval of working  | 88  |
| 61 | working time and interval of working  | 91  |
| 62 | working time and interval of working  | 92  |
| 63 | working time and interval of working  | 93  |
| 64 | working time and interval of working  | 95  |
| 65 | working time and interval of working  | 96  |
| 66 | working time and interval of working  | 97  |
| 67 | flowchart   | 100 |
| 68 | Flow of detection system(50%)   | 104 |
| 69 | number of detection   | 104 |
| 70 | rate of detection   | 105 |

|    |                                      |     |
|----|--------------------------------------|-----|
| 71 | Flow of detection system . . . . .   | 106 |
| 72 | number of detection . . . . .        | 106 |
| 73 | rate of detection . . . . .          | 107 |
| 74 | Flow of detection system . . . . .   | 107 |
| 75 | score of HMM at Crossing A . . . . . | 108 |
| 76 | score of HMM at Crossing B . . . . . | 109 |

## 表目次

|    |  |     |
|----|--|-----|
| 1  | Crossing C : compare of length of working time . . . . .               | 35  |
| 2  | Crossing C : compare interval of working time . . . . .                | 39  |
| 3  | length of working time about 2 type information . . . . .              | 51  |
| 4  | interval of working about 2 type of information . . . . .              | 51  |
| 5  | result about length of working time . . . . .                          | 58  |
| 6  | result about interval of working . . . . .                             | 58  |
| 7  | data of Crossing C . . . . .   | 63  |
| 8  | result of normal crossing . . . . .                                    | 64  |
| 9  | about simbol . . . . .   | 66  |
| 10 | result of pattern1 . . . . .   | 66  |
| 11 | result of pattern2 . . . . .   | 67  |
| 12 | result of abnormal detection(pattern3) . . . . .                       | 67  |
| 13 | statistics value about Crossing D . . . . .                            | 79  |
| 14 | Anomaly about length of working time . . . . .                         | 85  |
| 15 | Anomaly about interval of working . . . . .                            | 89  |
| 16 | Anomaly about length of working time and interval of working . . . . . | 90  |
| 17 | result of applying detection flow(threshold level = 50%) . . . . .     | 103 |
| 18 | result of applying detection flow(threshold level = 90%) . . . . .     | 103 |
| 19 | result of fleow of detecting system about anomal Crossing . . . . .    | 108 |



# 1 序論

## 1.1 研究の背景

現在、東日本旅客鉄道株式会社(以下 JR 東日本)では定常状態監視システムと呼ばれる集中管理システムを用い、列車や線路の状況を監視・管理し、保全活動を実施している。このシステムは東京の通信司令室にあり、関東近郊の JR 東日本在来線の列車情報を一括して監視している。このシステムは列車の故障や、事故情報をリアルタイムで監視し、各メンテナンスセンターなどに指示ができるようにしたシステムである。定常状態監視システムはセンサ系、伝送系、表示制御系で構成されている。センサ系では各種センサがレールや駅、踏切、信号機などに取り付けられ、そのセンサで電流値や電圧値を検知することにより故障情報、列車の位置情報、速度情報や信号の点灯状況などをチェックしている。伝送系はその情報を表示制御系へ、また表示制御系からの制御データをセンサ系に伝達するものである。表示制御系は各センサから収集したデータを蓄積し、表示や分析を行うとともに、システムの機能変更などの制御を行う。この定常状態監視システムには1日に膨大な量の異常情報が挙がってくる。しかし実際には何の故障もないのにセンサ系の誤作動が原因で異常情報として挙がってしまう誤情報がほとんどである。これはセンサ系が電流値・電圧値をセンシングし、その値が閾値を超えた場合に異常と判断しているため、雨の影響や計測ノイズ等により、一瞬でも電流値が閾値を超えてしまうと異常であると判断し、定常状態監視システムに異常情報として挙げてしまうためである。その結果、実際の故障件数は1カ月に極わずかであるのに対し異常情報は膨大な量に達する。このため、単一の情報からでは本当に異常や故障が発生したのかは判断できず、指令員が複数の異常情報を検討し経験的に異常があるかどうか、出動する必要があるかどうかを判断している。判断基準が複雑なため精度高く判断するためにはもちろん豊富な経験が必要となり、また、全ての異常情報に対する判別が必要であり、個別の事象への対応が薄く、あるいは短くなり、それが正確な判断の妨げとなる。このような状況では故障や事故の発見が遅れ修理にも遅れが生じ、最悪の場合、更なる事故にもつながると考えられる。そのため、誤情報を減らし簡便に出動・修理の必要があるかを判断できるようにすることが望まれる。

そこで本研究では統計的な手法を用いてこの警報情報から真の異常情報を抽出する手法の検討を行った。具体的には0, 1(警報の有無)で出力される定常監視システムの2種の時系列データを用いてその動作時間、動作間隔といったパラメータから診断を行っている。正常時と異常時におけるこのパラメータの違いを利用し、実際には異常が検出されていないデータに関する統計解析から発生確率に基づき閾値を設定し、自動的な抽出を行っている。

## 1.2 本研究の目的

本研究では定常監視システムに大量に挙がってくる誤情報を多く含む異常情報の中から真の異常情報を抽出し、指令員に対応の指示を促すシステムの提案を目的とする。現在の定常監視システムは誤情報を含む大量の異常情報が挙がってきてしまい、すべての異常情報に対応しきれていない。

そこで本研究では、まず異常情報を分析することによりこのシステムで検知すべき異常という状態を定義し、次にその定義に基づき異常状態を検知できるような異常検知フローを提案し、これにより司令員が大きな負担を負うことなく確認できる程度の異常情報数まで減らすことを目的としている。現状のシステムでは誤検知でも全て挙がってきてしまう情報に対してフィルターをかけてスクリーニングを行い、情報を減らし、さらに通過した異常情報に対してHMMという確率的モデルを用いた診断手法を適用することにより精度の高い診断を行い、異常を抽出する。診断には閾値設定など各種パラメータの設定があるが本研究ではその設定を踏切ごとに自動的に設定することでより高い精度で診断できるようにする。

以上の段階を踏み、統計的な手法を用いた異常検知システムの提案を目的とする。

### 1.3 対象機器

本研究では昨年度に引き続き、踏切内に障害物がないかを検知する踏切障害物検知装置を対象機器とする。この踏切障害物検知装置はレーザーの送信機と受光器の2つの機器から成り、踏切が閉まるとレーザーを射出し、それを受光器で受けるようになっている。もしこの送信機と受光器の間に何らかの障害物があるとレーザーがその障害物によって遮断され、受光器側で受光できなくなってしまう。このように受光器でレーザーを受光できなくなった場合に障害物があると判断し、定常監視システムに異常情報が挙がるようになっている。この踏切障害物検知装置は本来、閉まっている踏切内に車などの障害物が立ち往生して動けなくなってしまうときにそれを検知し、列車に停止信号を出すための装置である。しかし実際には人や動物が閉まっている踏切を無理に横断した場合や、伸びてきた植物の枝や葉によってレーザーが遮断されるたびに異常を検知して定常監視システムに異常情報が挙がってしまう。この定常監視システムに挙がってきた異常情報から、現場に出動する必要がある異常情報を抽出することを目的とする。

## 1.4 本報告書の構成

第1章「序論」では鉄道定常監視システムの概要、そのシステムにおける現状での問題点などといった研究背景、そしてその問題点に対する本研究の目的、本研究での対象機器について述べた。

第2章「基礎理論」では本研究で適用したHMM(隠れマルコフモデル)という統計的な確率手法について解説する。まずマルコフ課程やHMMの概要について説明した後、基礎理論や実際の計算上に関する問題、その解決アルゴリズムについて解説する。続いてパターン認識や状態推定といわれる推定にHMMを適用する手法について説明を行う。

第3章「異常検知フロー」では実際の異常情報と状況報告書から異常情報の分析を行い、正常・異常の定義を行い、検知すべき異常状態を明確にした。次に異常の発生パターンを調べることにより発生パターンの分類を行った。そしてその発生パターンに基づき異常を検知するための異常検知フローについて検討した。

第4章「フィルターによるスクリーニング」では統計的な手法を用いて異常情報のスクリーニングを行う方法について検討した。これは実際の現場で司令員の方が定常監視システムのモニタ上で目視によって簡単に判断している方法を統計的な手法を用いることにより自動化したものである。フィルターは2段階に分けて掛けている。2段階目のフィルターはさらに2種類ある。動作時間と動作間隔に関するフィルターである。このフィルターの設定に関しパラメータ推定などの手法を説明し、検討を行った。

第5章「統計的フィルターの検証」では第4章で検討したフィルターを用いて検証を行った。まず、異常状態のシミュレーションデータを作成しフィルターの検出限界について検討し、続いてフィルターを実機に適用、検証した結果を示す。

第6章「HMMを用いた異常検知方法」ではHMMを本研究にどのように適用し、どのような手順で解析を行うのかについて説明する。第2章で説明したHMMを本研究への具体的な適用方法、解析手順について解説する。次にその手法を用いてシミュレーションデータに対して検出限界の検証を行った結果を示し、異常検知のための最適な適用方法について検討した。

第7章「異常検知フローに従った検証結果」では検知フローに基づいて異常の検知を行った。第4章、第6章ではそれぞれの手法で単独に異常検知を行っていたが、本章ではそれらをすべて用いた検知フローで、どの程度の異常検知が可能であるの

か検証を行った。

第8章「結言」では本研究を通して得られた結論を総括する。

## 2 HMM(隠れマルコフモデル)

### 2.1 緒言

本研究では、異常情報の動作時間および動作頻度の二つのパラメータに関し、その正常時の平均および変動幅、実際に異常が発生した際の平均および変動幅から両者を分割する閾値を決定し、それを元に診断を実施している。今回の対象においては各動作時間・頻度の発生率分布が明確に分かれており分割の閾値の決定が容易であった。しかしながら、ノイズ（この場合、意図しない異常情報を示す）等の多い環境では正常・異常ともに変動幅が大きくなり、分布の重複部が大きくなり閾値の決定が困難な場合が想定される。そこで本章では単純に動作時間ではなく時系列波形パターンの出現率を元に異常の検知を行う手法としてHMM(隠れマルコフモデル)を用いるフィルターによる検証について示す。

## 2.2 HMM(隠れマルコフモデル) 概略

HMM は記号列の出現確率を計算するモデルである。

HMM とは隠れマルコフモデル (Hidden Markov Model) のことであり確率的な手法である。HMM はパターン認識やバイオインフォマティクスなどの研究で使われ、成果を上げている。パターン認識の分野では文字認識、音声認識、バイオインフォマティクスの分野では遺伝子発見、タンパク質のモデル化や検索などの研究がある。

HMM は確率モデルで表わされるため、統計や確率過程論による論理展開がしやすいという特徴がある。また時系列データなどにおける時間軸方向に伸縮するデータに強いという特徴をもつ。

### 2.3 基礎理論

今回の場合、記号列は1秒ごとの障検動作の有無を1,0として、その100秒間分の1,0の列で表している。したがって  $\{0,0,0,0,0,0,1,1,0,0,0,0 \dots 0,0,0,0\}$  のような記号列になる。長時間動作しているときは1がその秒数分続き、回復するとまた0が続く。頻度が上がると1が何回も出てくることになる。

HMMは記号列  $o = \{0,0,0,1,1,0,0,0\}$  といった記号列がどのくらいの確率で現れるのかといった具合に計算するモデルである。

仮に今、図1のHMM1のモデルを考えると記号0が出現する確率が0.9、記号1が出現する確率は0.1という状態を表している。したがって上に書いた赤文字の記号列は0が3回出現し、次に1が2回続き、その後0が3回続くという記号列なので、この記号列がこのHMM1から生成される確率は

$$P(\text{記号列}) = 0.9 \times 0.9 \times 0.9 \times 0.1 \times 0.1 \times 0.9 \times 0.9 \times 0.9 \quad (1)$$

$$= 0.00531441 \quad (2)$$

となる。

今、HMM1で記号出現確率  $p(0)$ 、 $p(1)$  を適当に決めたが、これは学習によって決定する。学習は記号列を与えて、その記号列を生成する確率が高くなるように記号出現確率を修正していく。たとえばHMM1に上の記号列を学習させると、8個の記号のうち0が6回、1が2回なので

$$p(0) = 0.75 \quad (3)$$

$$p(1) = 0.25 \quad (4)$$

としたときに記号列を生成する確率が最大になるので、この値に近づくように修正する。

このようにあるパターンの記号列を複数与え、記号出現確率を修正させていくことにより、そのパターンを生成しやすいようにHMMのモデルを作成する(図2)。

こうしてあるパターンに対応したHMMが作成できたら、そのパラメータ(記号出現確率)を用いて未知のパターンの記号列の生成確率を計算する。今回の場合だと未知のパターンが診断したいデータにあたり、異常情報が挙がってくるが正常であるときのパターンを学習させることにより正常時のパターンときは確率が高く、真の異常情報のような今までと違ったパターンに対しては低い確率となるHMMができる(図3)。

実際に計算に使っているHMMは先ほどの1,0の出現確率が1通りではなく数通りあり、何種類かの状態として与えられている。それが図4で表しているようなHMMで、この場合4つの状態  $S$  があり、それぞれの状態  $S$  がそれぞれ  $p(0)$ 、 $p(1)$  を持っている。



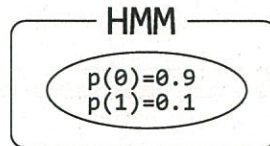


fig. 1: probability of independence of observations



fig. 2: learning of HMM

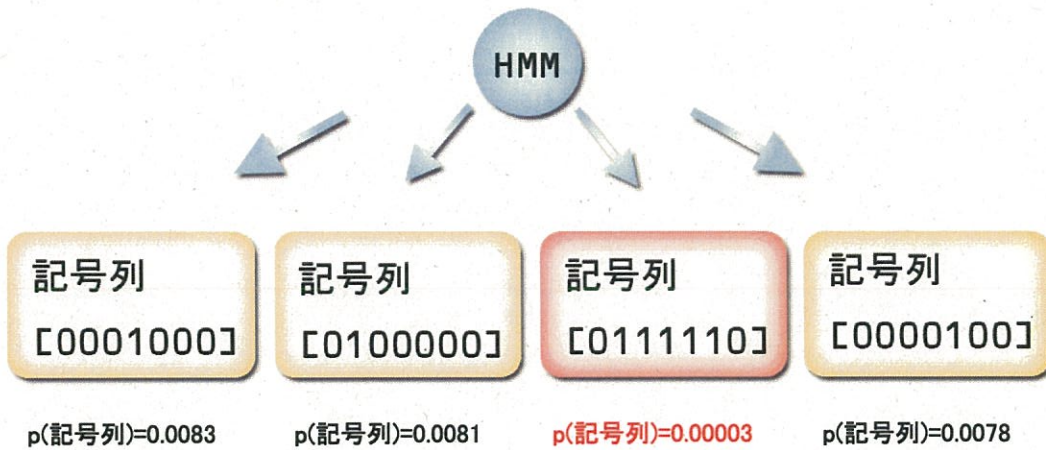


fig. 3: calculate probability of independence of observation sequence

そしてある状態で記号を1つ出現させると次の状態に移ってまた記号を出現させていく。たとえば状態  $S1$  で記号を出現させると、次に状態  $S2$  や  $S3$ 、 $S4$  に移る、またはそのまま状態  $S1$  に留まって、次の記号を出現させる。どの状態に移るかは確率によって決まる：状態遷移確率 (図中の  $a_{12}$  は状態  $S1$  から  $S2$  に移る確率)。

さらに今回は複数の状態を持っているため、どの状態から記号列の生成が始まるのか、という初期状態確率がある。これも初期状態としてどの状態になるかを確率で決めるためのものである。

したがって記号列の生成確率は

$$P(\text{記号列}) = \text{初期状態確率 } i \times \text{状態 } S_i \text{ での記号出現確率} \times \\ \text{状態 } S_i \text{ から } S_u \text{ への遷移確率} \times \text{状態 } S_u \text{ での記号出現確率} \times \\ \text{状態 } S_u \text{ から } S_w \text{ への遷移確率} \times \dots$$

として計算される。

学習は3つのパラメータ (記号出現確率、状態遷移確率、初期状態確率) を、学習させたいパターンの記号列が高い確率で出現するように修正していき、診断ではそのパラメータを用いて記号列の生成される確率を計算する。

学習では状態遷移の仕方 (状態数が4、記号の数が100のとき、遷移の仕方は  $4^{100}$ ) など膨大な計算量となるのでアルゴリズムを用いて計算量を減らしてしく。

HMMのモデル化はこの学習段階にあたり、診断するときを使う値はそのHMMを使った記号列の生成される確率である。この値は正常と異なるほど小さい値となる。そこで異常度が高いほど高い値になるように、単にHMMの計算結果の対数を取り、それにマイナスをつけた ( $-\log$  を取った) ものがスコアである。

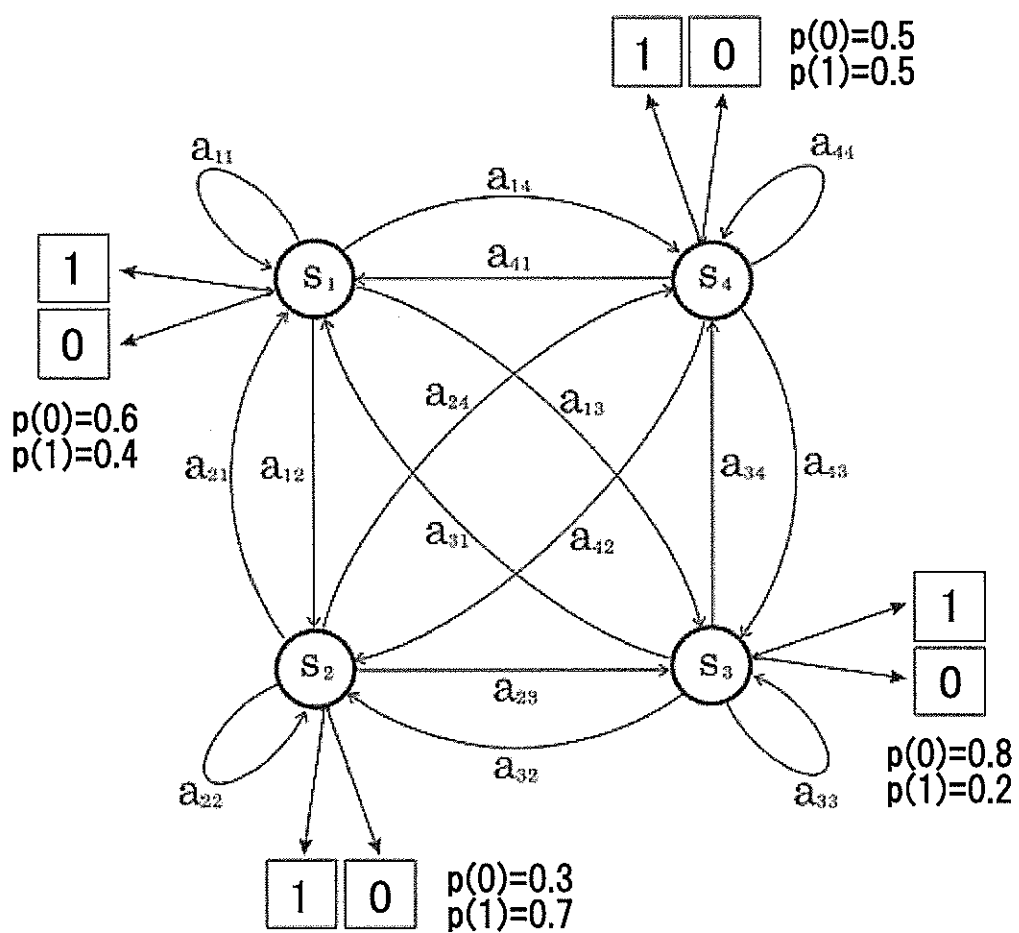


fig. 4: parameter of HMM

## 3 異常検知フロー

### 3.1 緒言

本章では障害物検知装置の動作原因について述べ、それに基づく正常・異常の定義付けを行う。さらにその中から抽出すべき異常と、その異常情報の挙がり方に関するパターン分けについて説明する。そしてその異常情報のパターンに基づきこれらのパターンに応じた真の異常を検出できるようなフィルターの作成方法について検討した。そして異常の検出の流れをフローチャートで示し、第4章以降で各種フィルターについて詳しい説明を行っている。

## 3.2 異常情報の分類

### 3.2.1 動作原因

踏切障検が動作する原因にはいくつかの種類がある。以下にその例を挙げる。踏切障検の動作原因には機器故障以外では確実な原因が分からない場合もあり、ここで挙げる動作原因とは状況報告書によるものや一般的な推測であり必ずしも原因が特定されているわけではない。

- 閉まっている踏切での人や動物の横断
- 車の立ち往生(列車の停止)
- 植物の枝や葉(伐採)
- ケーブルの断線(ケーブルの交換)
- 動物の装置への衝突による光軸のずれ(光軸の調整)

( )内は必要な対応

このように実際の事故ではない発生原因による異常情報というものが多発している。異常情報は1日に1000件弱挙がってくるが、そのうち実際に異常が発生していたのは月に数件であり、実にそのほとんどが誤検知となっている。

### 3.2.2 検知すべき異常

上で挙げた例のうち、閉まっている踏切での人や動物の横断には特に対応する必要がなく、それ以外の原因に対してはそれぞれ列車の停止、伐採、ケーブルの交換、光軸の調整、といった対応が必要となる。定常監視システムでは本来対応すべき異常情報を検知すべきシステムであり、本研究ではこの対応すべき異常情報を「異常」と定義する。

すなわち

正常

- 閉まっている踏切での人や動物の横断

異常

- 車の立ち往生(列車の停止)
- 植物の枝や葉(伐採)
- ケーブルの断線(ケーブルの交換)
- 動物の装置への衝突による光軸のずれ(光軸の調整)

というように「正常」・「異常」を定義し、この「異常」を検知することを目的とし、次節ではこの「異常」の発生と異常情報の挙がり方に関してもう少し細かく分類し、検知すべき異常を明確にする。

また全てを異常情報と呼ぶとどちらの異常情報のことか分かりにくくなるため、以下では正常時に挙がってくる異常情報を「誤情報」、実際に異常が発生しているときに挙がってくる異常情報を「真の異常情報」と呼ぶことにする。

誤情報 正常時に挙がってくる異常情報(実際には異常ではない)

真の異常情報 実際に異常が発生しているときに挙がってくる異常情報

この障検装置に関する異常情報のうち、ここで挙げた「正常」な異常情報、すなわち誤情報は先ほど述べたように1日に約1000件弱ほどある。これに対して「異常」な異常情報、すなわち真の異常情報は月に約数件ほどと極わずかである。したがって確実に真の異常情報を抽出し、誤情報が挙がってこないように出来れば大幅に異常情報を減らすことが出来、システム管理の面や、異常かどうかを見極める負担が減り、より効率的なシステムとなる。

### 3.2.3 異常発生パターン

ここで定常監視システムの活用という面から考えると、異常が発生して電車で遅延が発生してしまう状態になってから異常を検知するのでは対応が出来ず、列車遅延に対する何の対策も出来ない。したがって定常監視システムで異常が発生する前の「異常の兆候状態」で異常を検知することが出来ることが望ましい。そのため本研究では兆候状態で異常を検知できるかどうかで異常状態の分類を行った。なお本研究における兆候状態とは、機器故障などが発生し、異常情報が挙がってきているがまだ列車停止を行っていない状態である。つまり定常監視システムまたは現場で異常に気がつき(異常情報が挙がりっ放しになる等)、列車を停止させるように対応した場合、その列車停止の指示が出る前に挙がってきた異常情報である。またその異常情報が誤情報なのか、真の異常の兆候を表す異常情報なのかはデータから予想したものであり、確実に誤情報ではなく兆候状態であるとは言えないが、実際に異常が発生した踏切においては誤情報も少なく、兆候状態と考えてほぼ間違いないと考えられる。

分類としてはまず兆候があるかどうかで分類する。この兆候がある異常状態に関してはその兆候状態での異常検知を目的とする。兆候がない場合には突然真の異常情報が挙がってくるため、事前の対処ではなく現場で列車停止等の対応の必要がある。

次に誤情報があるかどうか、つまり正常時に異常情報が挙がってきていたかどうかで分類する。正常時に異常情報が挙がってきていない場合には、真の異常情報が初めての異常情報となる。これは正常時には異常情報は挙がっていないが異常時には異常情報が挙がってくる、という差異によって異常を検出できる。正常時に異常情報が挙がってきた場合には、誤情報と真の異常情報との異常情報のあがり方の差異によって異常を検出する必要がある。

以上のように分類しこれらのパターンを検出できるようにフィルターを作成する。

上で挙げた異常状態をそれぞれパターン1、パターン2、パターン3とパターン分けしてものをツリーとして表したのが図5である。

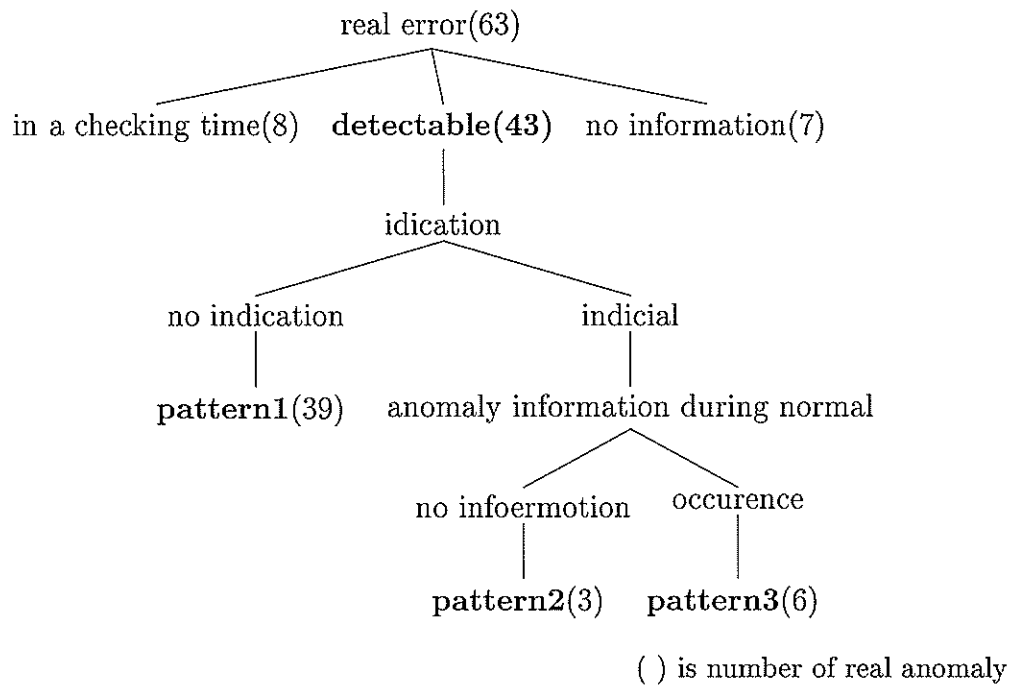


fig. 5: assortment of real anomaly

パターン1 兆候なしで異常情報が挙がる

パターン2 兆候はあるが初めての異常情報

パターン3 兆候があり、正常時にも異常情報が挙がっている



### 3.3 異常検知フロー

以上の異常発生パターンを踏まえ、すべての異常を確実に検知できる異常検知フローに関して検討する。

#### 3.3.1 フローチャート

実際の異常検出には複数のフィルターを用いて診断を行う。その流れを表したのが図6に示すフローチャートである。今回の異常検出では3段階のステップを踏む。

1段階目は「以前に異常情報があったか」のフィルターである。フィルターの詳細については4.2節で説明する。

2段階目は「統計的フィルター2種類」である。これは「動作時間フィルター」「動作頻度フィルター」の2種類のフィルターである。この2種類のフィルターについては4.3節で詳細な説明を行う。

3段階目は「HMMを用いたフィルター」である。これは2段階目の統計的フィルターでスクリーニングした異常情報にさらにフィルターをかけ、より高精度に診断するためのフィルターである。詳細に関しては6章で解説する。

3段階のフィルターを通してでも正常でないと判断された場合、「指令員が確認、対応の指示」という対応を必要とする。これらのフィルターを用いて図6に従った異常の検知を行う。

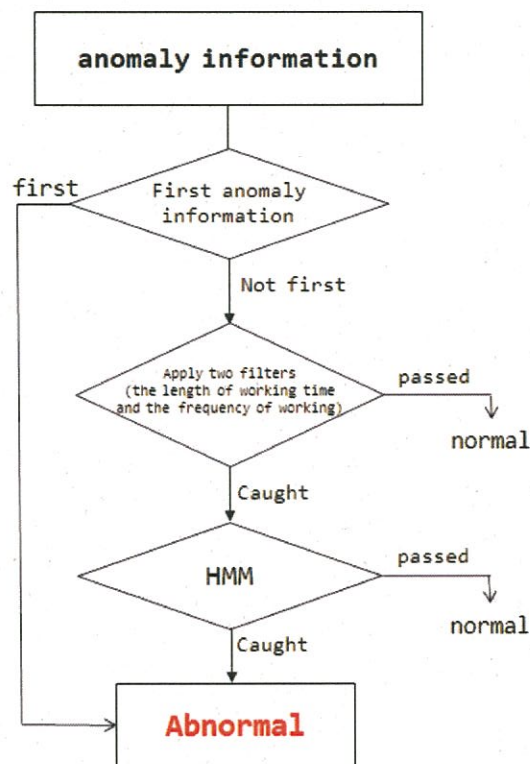


fig. 6: flowchart

### 3.4 異常の検出フロー

異常が発生した場合に図6のフローチャートに従ってフィルターをかけていくとそれぞれのパターンがどのフィルターで異常として検出されるかを図7から図12に示す。ここでは実際には異常が発生していないが異常情報の挙がってくる、すなわち誤情報であるパターンについても確認する。このパターンを『パターン0』とする。『パターン0』に関しては挙がってきた異常情報が、その踏切において初めての異常情報である場合には異常と判断し(図7)、初めてでない場合には正常と判断する(図8、図9)。

兆候がなく、異常情報が回復せずに挙がったままの状態である『パターン1』の場合は図10のように異常を検知できる。これは、フローチャートに従うのではなく、実際には障害物検知装置が動作したままになるので同時に「特発制御器故障」も挙がり、列車に対して停止信号が提示されるためメンテナンスセンターからの指示よりも現場での対応が先に行われる。そのためフローチャートとは別の経路で異常と判断されるが、その情報をフローチャートに従わせた場合も検知は可能である。

兆候があるが正常時に異常情報の挙がってこない『パターン2』に対しては実際に異常が発生したときの異常情報がその踏切における最初の異常情報となるため「以前に異常情報があるかどうか」のフィルターで異常を検出することが可能である(図11)。

兆候があり、しかも正常時にも異常情報が挙がってくる『パターン3』に対しては「以前に異常情報があるかどうか」のフィルターでは検出できないが「動作時間フィルター」または「動作間隔フィルター」によって異常の検出が出来る(図12)。

このフローに従い、異常がないと診断された場合には正常とし、それ以外の場合には正常でないとして判断し「実際に指令員が確認、対応の指示」を行う。

以上のようにして全てのパターンの異常を検出することが出来る。

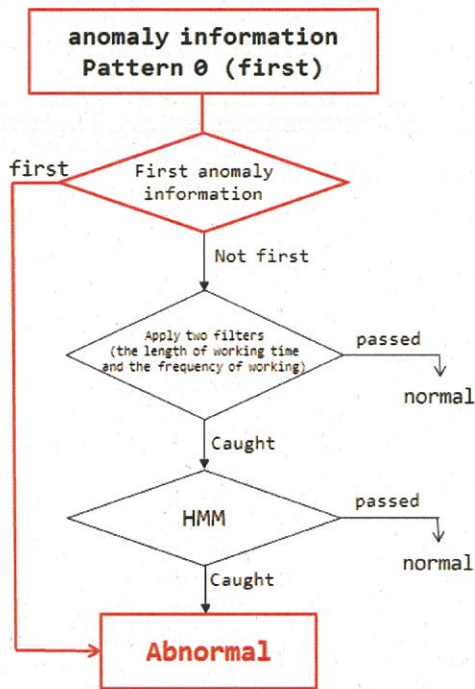


fig. 7: pattern0(first anomaly information)

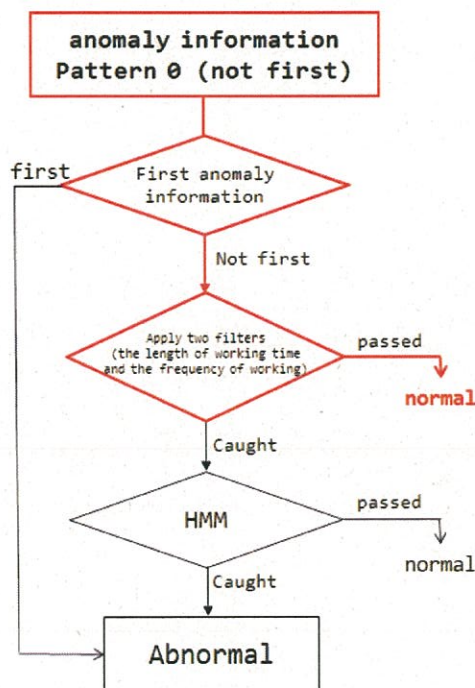


fig. 8: pattern0(not first, pass statical filter)

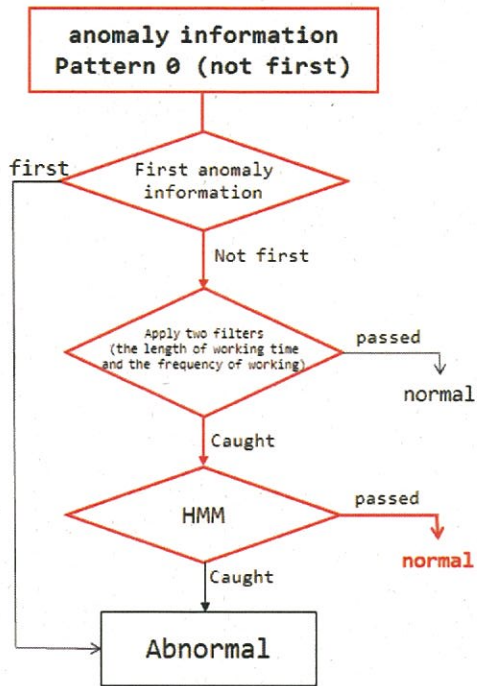


fig. 9: pattern0(not first,pass HMM filter)

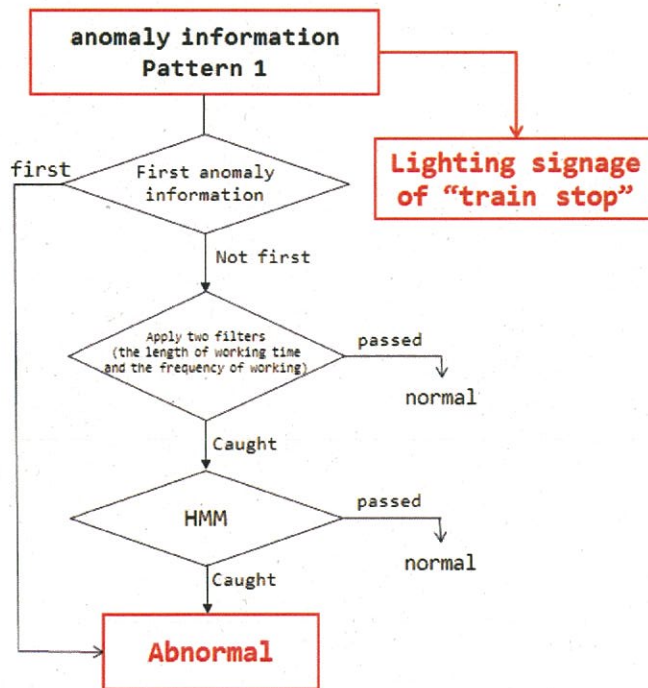


fig. 10: pattern1(no indication)

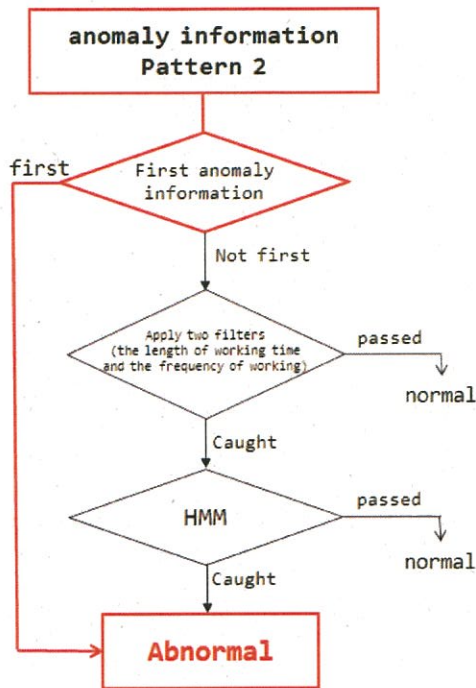


fig. 11: pattern2(first anomaly information)

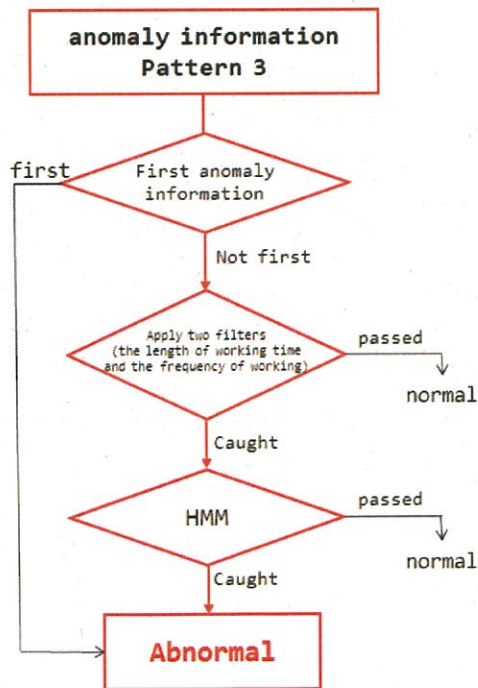


fig. 12: pattern3(caught by statical and HMM filters)

### 3.5 結言

本章では踏切障検の動作原因、異常情報に関する「正常」「異常」に関する定義から検出すべき異常情報の発生パターンについて見てきた。これにより検出すべき異常情報が明確になったので、次章以降ではこれらの異常を検出するためのフィルターについて検討していく。

## 4 フィルターによるスクリーニング

### 4.1 緒言

本章では異常検知フローにおける第一段階のフィルターと第二段階の統計的フィルターに関して述べ、その有効性について検討する。また第二段階で設定する統計的フィルターは踏切の異常発生をモデル化するため、そのモデル化に必要な各種パラメータの設定方法についての検討を行った。最尤推定、ベイズ推定の両者を比較し、パラメータを決定するために適している推定手法を採用することとした。



## 4.2 フィルター1st

フローチャートで最初のフィルターに設定されているフィルター1stは以前に異常情報があったかどうか、すなわち「誤情報」があったかどうかを見るためのフィルターである。以前に異常情報が挙がってきていない踏切で初めて異常情報が挙がったときには、その異常情報を真の異常情報として扱う。このフィルターは短時間回復するような異常情報に対しても異常と判断するようになっている。

### 4.3 フィルター 2nd(統計的フィルター)

#### 4.3.1 動作時間・動作頻度

2段階目のスクリーニングとして2種類の統計的なフィルターを適用する。動作時間と動作頻度に関するフィルターである。正常時でも異常情報が挙がってきてしまう場合(誤情報)と真の異常時に挙がってくる異常情報(真の異常情報)には挙がり方に何らかの差異があると考えられる。そこでこの差異を用いて正常でも異常が挙がってきてしまう場合と、真の異常情報の判別を行うためのフィルターを作成する。

今回は時系列データから差異があると考えられる「動作時間」と「動作頻度」に関するフィルターを作成した。「動作時間」「動作頻度」のフィルターに用いるパラメータには「動作時間」「動作間隔」の2つがある。

ここで「動作時間」「動作間隔」を以下のように定義する(図13参照)。図は横軸に時間、縦軸に障検の作動状態(1)、回復状態(0)を表している。

**動作時間** 障害物検知装置が作動してから回復すまでの時間

**動作間隔** 障害物検知装置が作動し回復してから次の作動までの時間

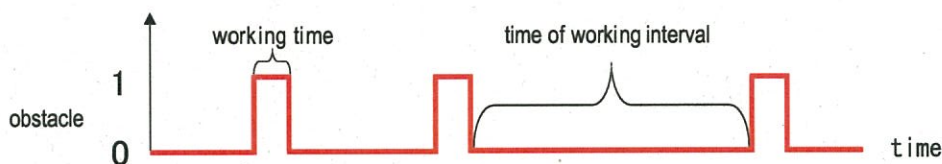


fig. 13: length of working time and interval of working

### 4.3.2 動作時間フィルター

まず動作時間に関するフィルターについて説明する。動作時間に関するフィルターとは誤情報における動作時間と真の異常情報における動作時間の差異からその判別を行うためのフィルターである。

まず動作時間のフィルターを作成するために、誤情報と真の異常情報での動作時間の違いについて調べる。誤情報であっても踏切ごとにその動作時間はことなるため、誤情報と真の異常情報の動作時間の違いを見るためには両方の異常情報が挙がっている踏切で調べる必要がある。そこで今回はパターン3に該当する「踏切C」の異常情報から調べた。踏切Cの誤情報と真の異常情報の動作時間をヒストグラムにして表したのが図14、基本統計量を表1にまとめている。このヒストグラム、基本統計量から真の異常情報と誤情報における動作時間はその長さを用いることによって判別できると考えられる。

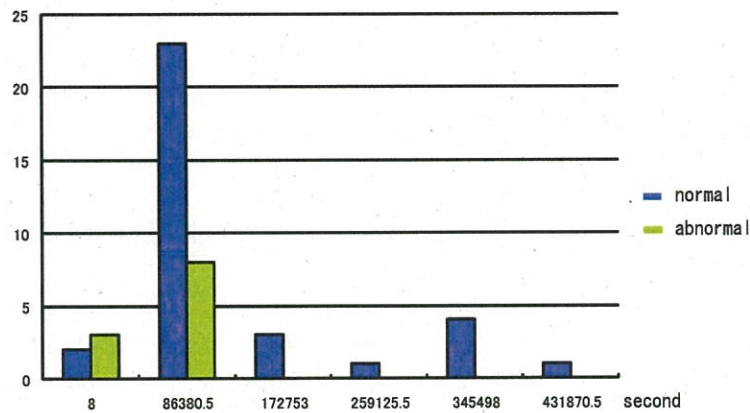


fig. 14: length of working time(Crossing C)

table 1: Crossing C : compare of length of working time

| not real error(working time) |         | real error(working time) |          |
|------------------------------|---------|--------------------------|----------|
| average                      | 23.96   | average                  | 142.50   |
| standard variation           | 35.85   | standard variation       | 185.03   |
| variance                     | 1285.39 | variance                 | 34237.91 |

また、真の異常情報が挙がっていない踏切でも、異常時の特徴から、差異が生じる場合には真の異常情報は誤情報の場合よりも長い動作時間となると考えられるため、全ての踏切でこの方法が適用できると考えられる。

誤情報と真の異常情報との差異があることが分かったので、動作時間に関するフィルターではある閾値を用いて、それ以上の動作時間にはその異常情報は真の異常であるとみなすようなフィルターとする。この閾値を決定するために動作時間に関するモデル化を行う。ここでは異常情報の分布を調べ、その分布から何秒以上だと異常であると考えられるのかを統計的に定める。これにより情報が挙がってくる都度、フィルターにかけることにより即座に異常であるかどうかの判断が出来、速報性を持たせることが出来る。

動作時間の分布を求めるために、踏切ごとに動作時間のヒストグラムを作成した。図 15、図 16、図 17 に示すのが踏切 B、踏切 D、踏切 G の例である。これらのヒストグラムから動作時間の分布を指数分布と仮定する。

指数分布とは正のパラメータ  $\lambda$  に対して確率密度関数が

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad (5)$$

で与えられる分布である。動作時間を指数分布と仮定すると統計的な信頼区間を与えることにより、その値を閾値とし、その値以上の動作時間の異常情報が発生した場合に真の異常と判断することが可能である。

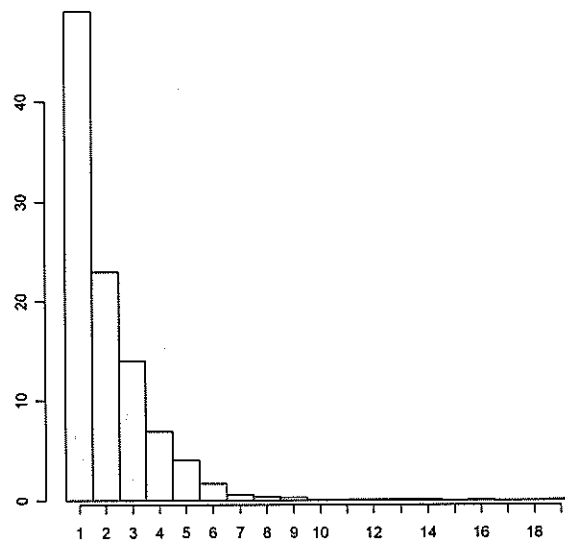


fig. 15: histogram of length of working time(Crossing B)

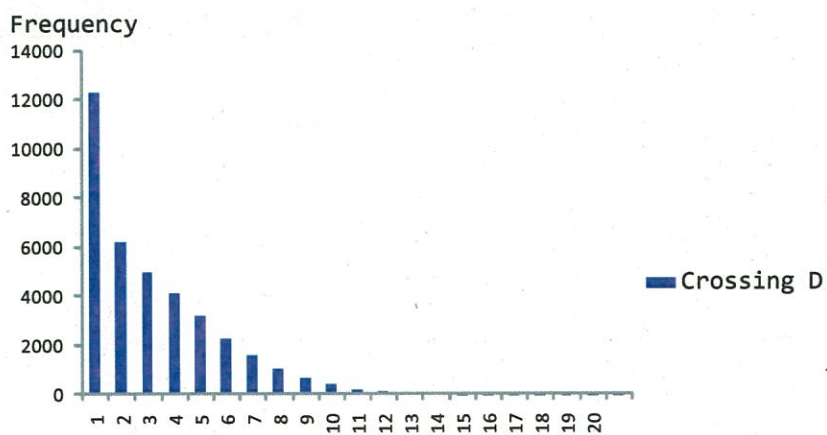


fig. 16: histogram of length of working time(Crossing D)

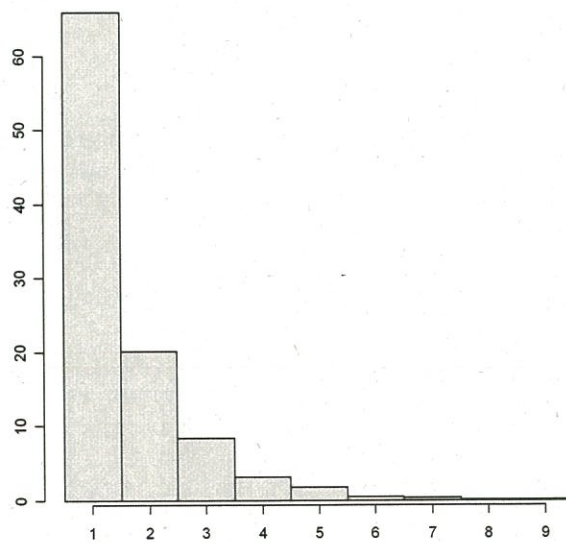


fig. 17: histogram of length of working time(Crossing G)

### 4.3.3 動作頻度フィルター

次に動作時間に関するフィルターについて説明する。動作頻度も同様に誤情報と真の異常時における動作頻度の差異をもって判別できるようにする。一般に信頼性工学において、機器の故障がランダムに発生するとき、その単位時間当たりの故障回数である故障率は一定となる。そしてこの場合、機器の故障間隔は指数分布に従うことが分かっている。今回の異常情報を信頼性工学でいう「故障」とみると、挙がってくる異常情報はランダムに発生するため異常情報の挙がってくる間隔が指数分布に従うと考えられる。そこでまずは実際の異常情報が指数分布に従っているかを調べる。先ほどの動作時間のときと同様に踏切 B、踏切 D、踏切 G における動作間隔のヒストグラムを図 18、図 19、図 20 に示す。

これより動作間隔は指数分布に従っていると考えられる。ここで今回も踏切 C の動作間隔を誤情報、真の異常情報と分けた場合のヒストグラムを図 18 に、基本統計量を表 2 示す。今回の場合、ヒストグラムから真の異常情報の分布が誤情報の分布に含まれてしまっているため、単純な比較ではなく、指数分布のパラメータの変化を見なければならぬ。これは例えば、誤情報による異常情報から推定した指数分布のパラメータと診断時の異常情報から推定した指数分布のパラメータの検定を行う必要がある。しかし指数分布のパラメータの推定を行うためには 5 回ほどの異常情報が必要になる。そのためパラメータの傾向が変わってきた場合などには異常の検出がしやすいが、単発的な異常情報の変化は検出できなくなってしまう。さらに複数のデータが必要であるため、異常が発生してから数データ分待たなければならず、速報性が失われてしまうという問題がある。

そこで今回は速報性を持たせるために動作間隔をそのまま用いるのではなく、動作時間の逆数を用いることにする。まず動作間隔の逆数の分布を示したグラフが図 21、図 22、図 23 になる。この分布から動作間隔の逆数の分布を指数分布に従うと仮定する。真の異常状態で挙がる異常情報は正常時の誤情報と比べて発生頻度は高くなるため、動作間隔は短くなり、その逆数は大きくなる。したがってこの差異を用いることにより誤情報と真の異常情報を判別できると考えられる。動作時間のフィルターと同様に閾値を用いてある値より大きな動作間隔の逆数、すなわち短い動作間隔となった場合に正常ではないと診断する。これにより 1 データごとに診断が可能であるため速報性を持たせることが出来る。

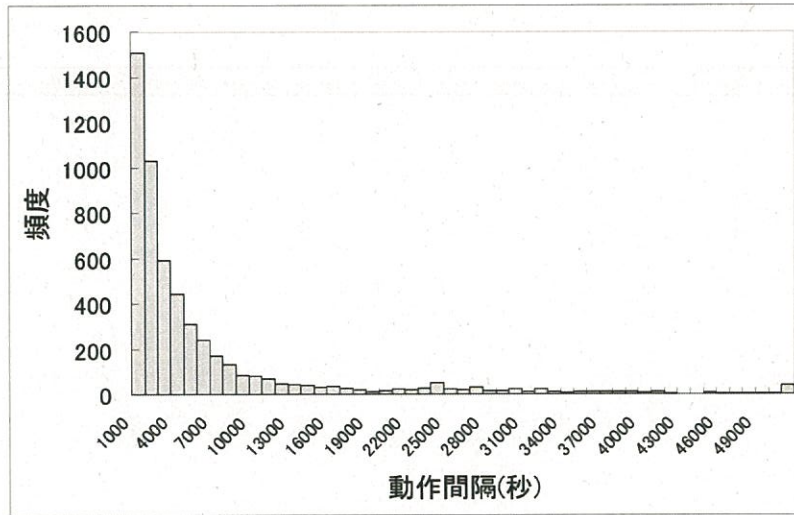


fig. 18: histogram of interval of working (Crossing B)

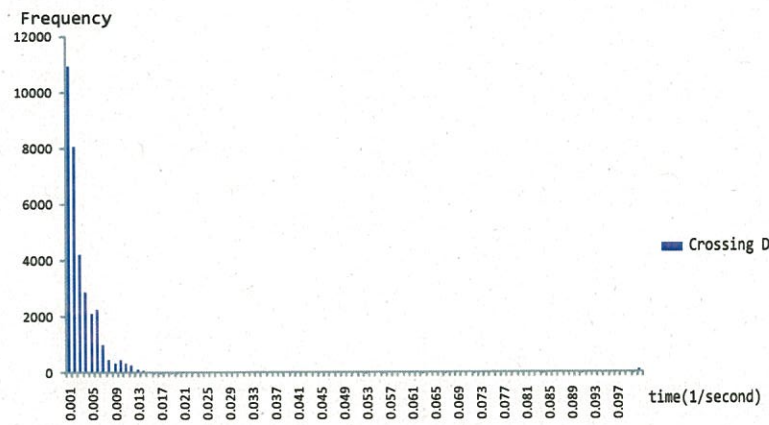


fig. 19: histogram of interval of working (Crossing D)

table 2: Crossing C : compare interval of working time

| not real error(interval) |                | real error(interval) |         |
|--------------------------|----------------|----------------------|---------|
| average                  | 88497.81       | average              | 29.55   |
| standard variation       | 154420.08      | standard variation   | 33.58   |
| variance                 | 23845560875.19 | variance             | 1127.47 |

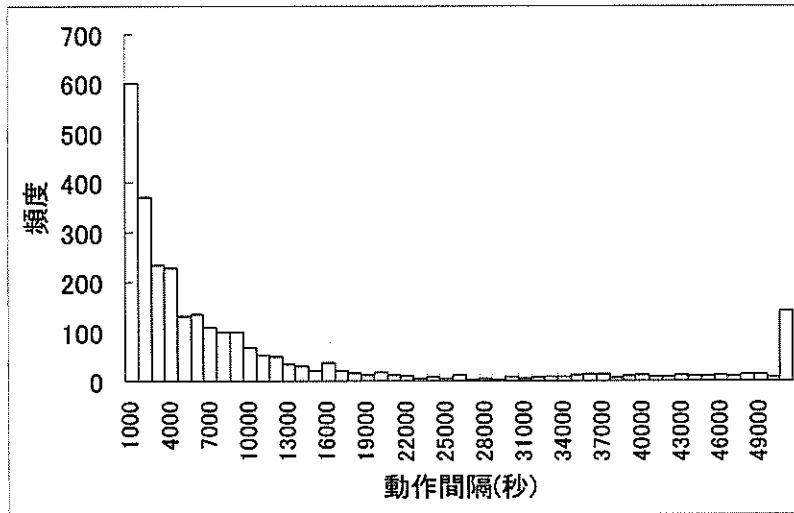


fig. 20: histogram of interval of working (Crossing G)

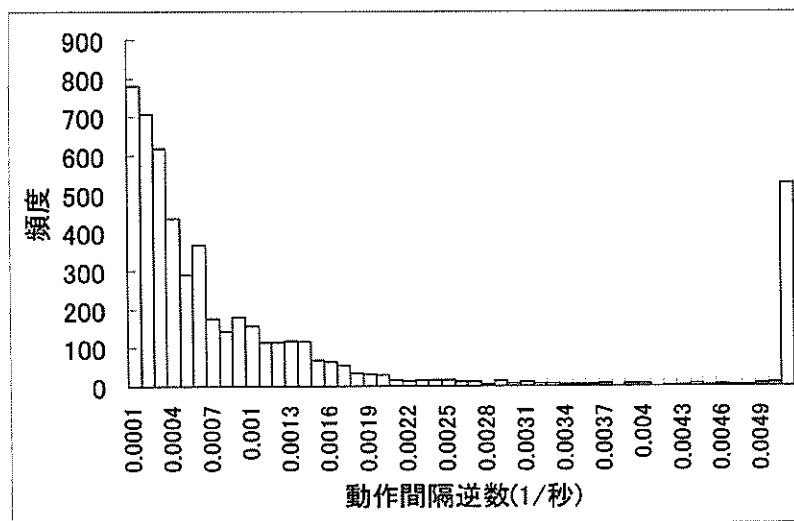


fig. 21: histogram of inverse of interval (Crossing B)



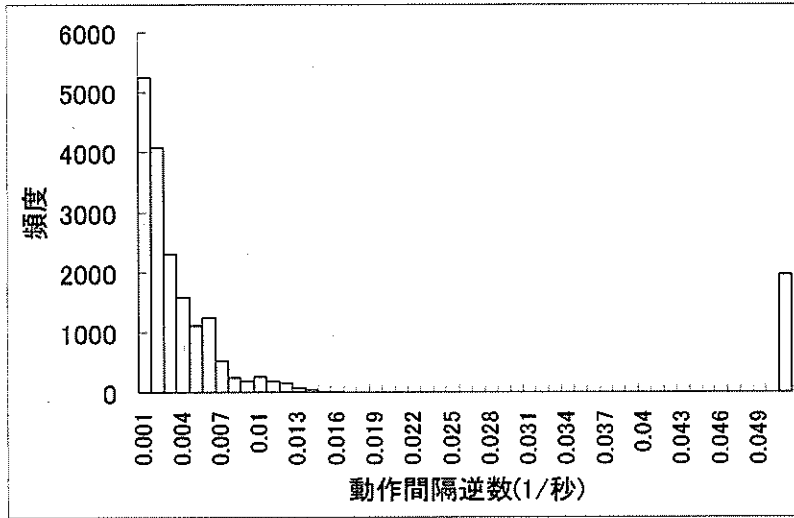


fig. 22: histogram of inverse of interval (Crossing D)

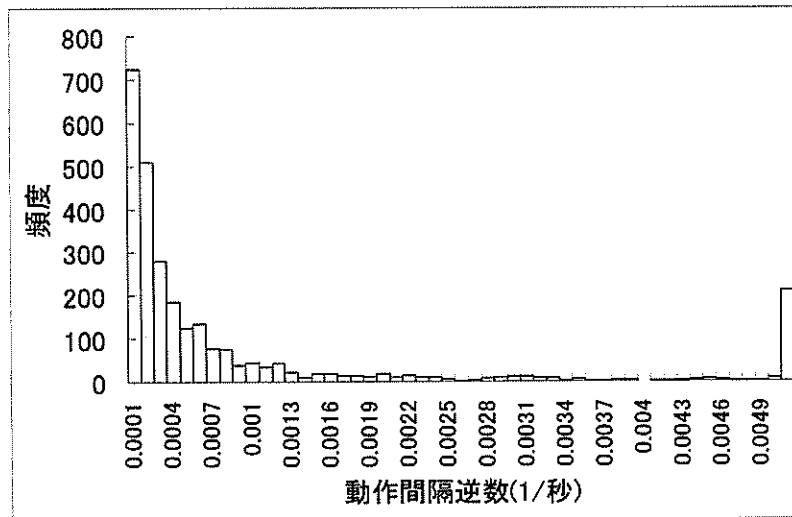


fig. 23: histogram of inverse of interval (Crossing G)

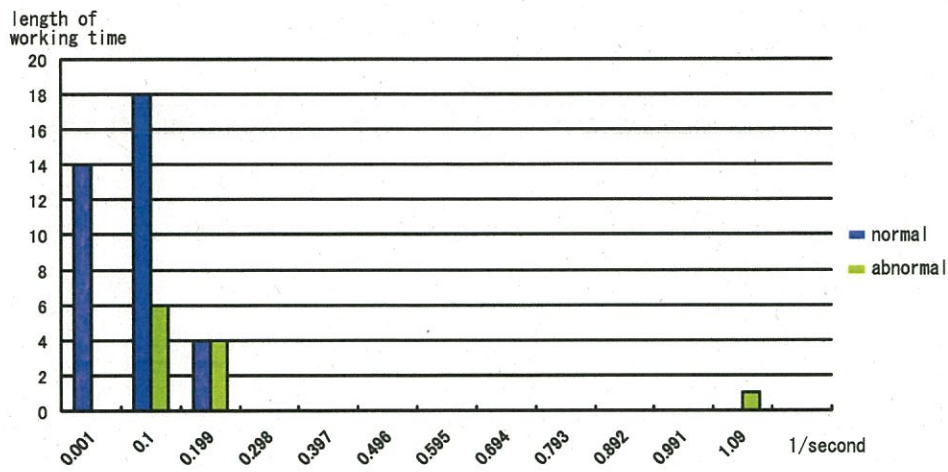


fig. 24: histogram of inverse of interval (Crossing C)

## 4.4 パラメータの設定方法

### 4.4.1 分布パラメータの設定

フィルターの作成に際し、パラメータ  $\lambda$  の決定にはベイズ推定または最尤推定によって推定する。ここではベイズ推定によってパラメータを推定した場合と最尤推定によってパラメータを推定した場合を比較し、どちらの推定手法がパラメータ設定に適切か判断する。今回は正常な状態で挙がってくる異常情報を、真の異常情報と判断してしまう誤検知の割合で比較した。例として2箇所の踏切を挙げる。

ここでは各踏切の正常時のデータを用いて、そのデータが蓄積されるごとにパラメータを推定する。そして推定するたびに正常時の全データをこのフィルターにかけて誤検知率を計算した。つまり全データが100データあるとすると、そのうちの最初の1データ目が蓄積されるとパラメータの推定を行いフィルターを作成し全100データをフィルターにかけて誤検知率を計算する。次に2データ目が蓄積されると再びパラメータの推定を行い、全100データをフィルターにかけて誤検知率を計算する。こうして誤検知率の推移を見る。フィルターの精度が高くなると誤検知率は低くなると考えられる。

この検証を先に挙げた2箇所の踏切に対して、最尤推定、ベイズ推定の2つの推定方法によって推定する。

### 4.4.2 最尤推定

最尤推定法とは尤度関数  $L(\theta|x)$  が与えられたとして、 $\theta$  に対する最尤推定値 (MLE) は次式を満たす値  $\hat{\theta}$  である。

$$L(\hat{\theta}|x) = \max_{\theta} L(\theta|x) \quad (6)$$

指数分布は

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad (7)$$

と表されるため、 $n$  観測値が  $x_1, \dots, x_n$  のとき尤度関数  $L(\lambda|x)$  は

$$L(\lambda|x) = \lambda e^{-\lambda x_1} \times \lambda e^{-\lambda x_2} \times \dots \times \lambda e^{-\lambda x_n} \quad (8)$$

$$= \lambda^n e^{-\lambda \sum x_i} \quad (9)$$

となる。このとき  $\lambda$  の MLE 推定値  $\hat{\lambda}$  は  $L(\lambda|x)$  を最大にする  $\lambda$  なので  $L(\lambda|x)$  を  $\lambda$  で偏微分して

$$\frac{\partial L(\lambda|x)}{\partial \lambda} = 0 \quad (10)$$

となる  $\lambda$  を求めればよい。ここでは (9) 式の  $\log$  をとり

$$\log L(\lambda|x) = n \log \lambda - \lambda \sum x_i \quad (11)$$

とした上で  $\lambda$  で偏微分をすると

$$\frac{\partial \log L(\lambda|x)}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i \quad (12)$$

より、求める  $\hat{\lambda}$  は

$$\frac{n}{\hat{\lambda}} - \sum x_i = 0 \quad (13)$$

$$\hat{\lambda} = \frac{n}{\sum x_i} \quad (14)$$

と求まる。

#### 4.4.3 ベイズ推定

ベイズ推定はベイズ統計学の考え方の下で母数推定を行う手法である。ベイズ統計学は「主観確率」という独特の定義を行っている。ここで簡単にベイズ推定について説明する。

ベイズ推定はベイズの定理を用いた客観的手続きによる更新により標本による情報を取り入れ、事前分布より精度を向上させた事後分布を評価するという立場から、少数の標本であっても有効に推定に活用することができる。推定対象の母数  $\theta$  は確率変数として扱われ、ある分布に従う。この分布は観測データの蓄積に伴い更新される。このとき、更新を行う前に与えられるのが事前分布である。この事前分布  $f_{pri}(\theta)$  は事前情報として以前の観測データや解析者の知見、同機種に対する事例データベースなどを用いて設定される。この事前分布に観測データを与えることにより事後分布として更新していき、より適切な推定値を得る。

本研究ではこの事前分布として、事前に全く情報がない場合である無情報性事前分布 (式 (15)) を設定した。

$$f(\lambda; k, \theta) = \theta^k x^{k-1} \frac{\exp(-\theta x)}{\Gamma(k)} \quad (15)$$

この分布を更新させ推定値を求めた。推定値は式 (16) で表わされる。

$$\hat{\lambda}_{\text{ベイズ}} = E(\lambda) = \frac{0.5 + n}{T} \quad (T = \sum t_i) \quad (16)$$

この値を閾値設定のためのパラメータとして採用している。

## 4.4.4 結果、考察

図 25、図 26 にパラメータの推定に最尤推定、ベイズ推定を用いた場合の誤検知率の推移を示している。図 27 は両者を同じグラフ上に示したものである。傾向がほぼ同じためこのグラフでは重なってしまう。

蓄積されたデータ数が少ないときの両者を比較したグラフが図 28、図 29 に示してある。このグラフはデータ数が 30 点までの最尤推定およびベイズ推定を用いた場合の誤検知率である。どちらも 15 件程度のデータ蓄積後は誤検知率が低くなっていることが分かる。

一般にベイズ推定は適切な事前分布を与えた場合には、データ数が少なくても最尤推定より真値に近い値を推定することが出来る。しかし今回は蓄積データが 0 の時点から考えるため、事前分布は無情報事前分布となり、最尤推定の場合とあまり変わらなくなっている。

グラフよりベイズ推定よりも最尤推定を用いた場合のほうが誤検知率が高くなっているが、これはベイズ推定が無情報事前分布としてガンマ (0.5, 0) を用いているためパラメータの推定値  $\hat{\lambda}_{\text{ベイズ}}$  が

$$\hat{\lambda}_{\text{ベイズ}} = E(\lambda) = \frac{0.5 + n}{T} \quad (T = \sum t_i) \quad (17)$$

となり、最尤推定による推定値  $\hat{\lambda}_{\text{最尤}}$

$$\hat{\lambda}_{\text{最尤}} = \frac{n}{T} \quad (T = \sum t_i) \quad (18)$$

よりも大きくなり、図 30 のように指数分布が左側にずれたような分布になり、閾値がゆるくなるためである。

ベイズ推定と最尤推定ではほとんど差がないため、どちらを採用しても良いことが分かる。したがって今回は同じ%点を閾値とした場合に閾値がきつい側に設定される最尤推定によるパラメータの設定方法を採用する。

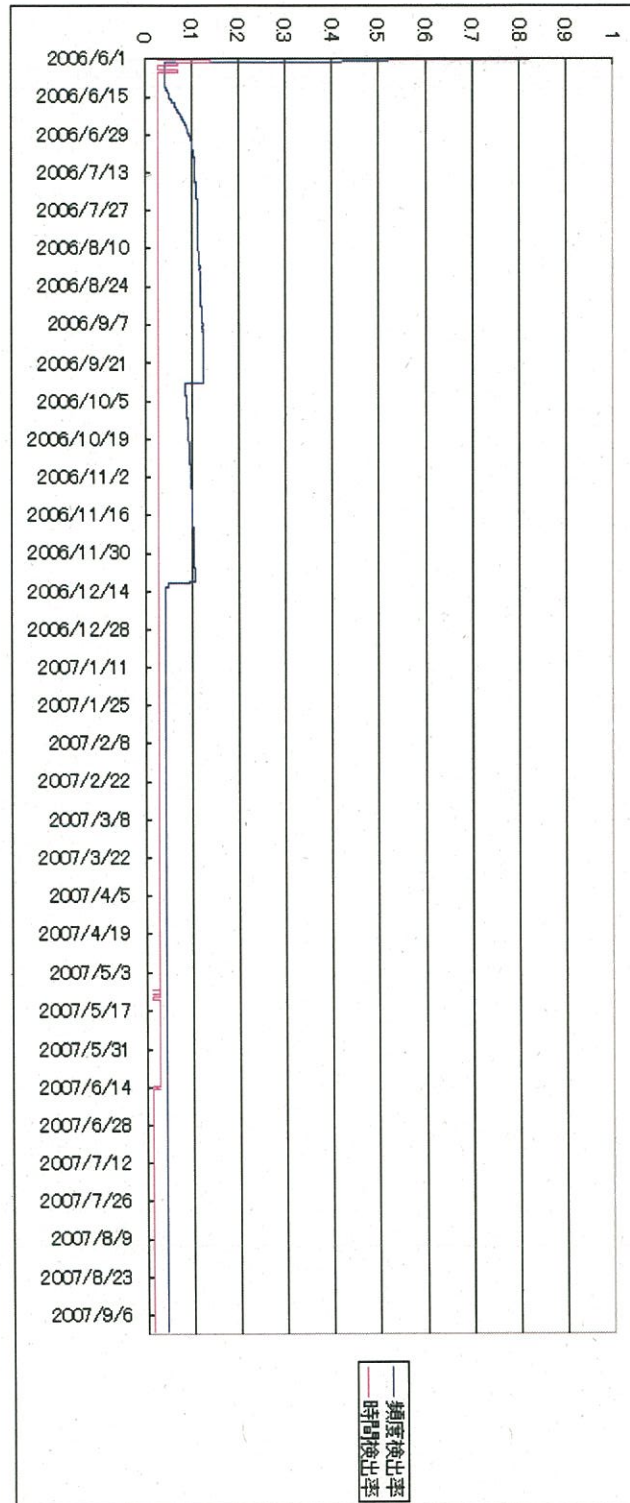


fig. 25: rate of mis-detection using maximum likelihood estimation

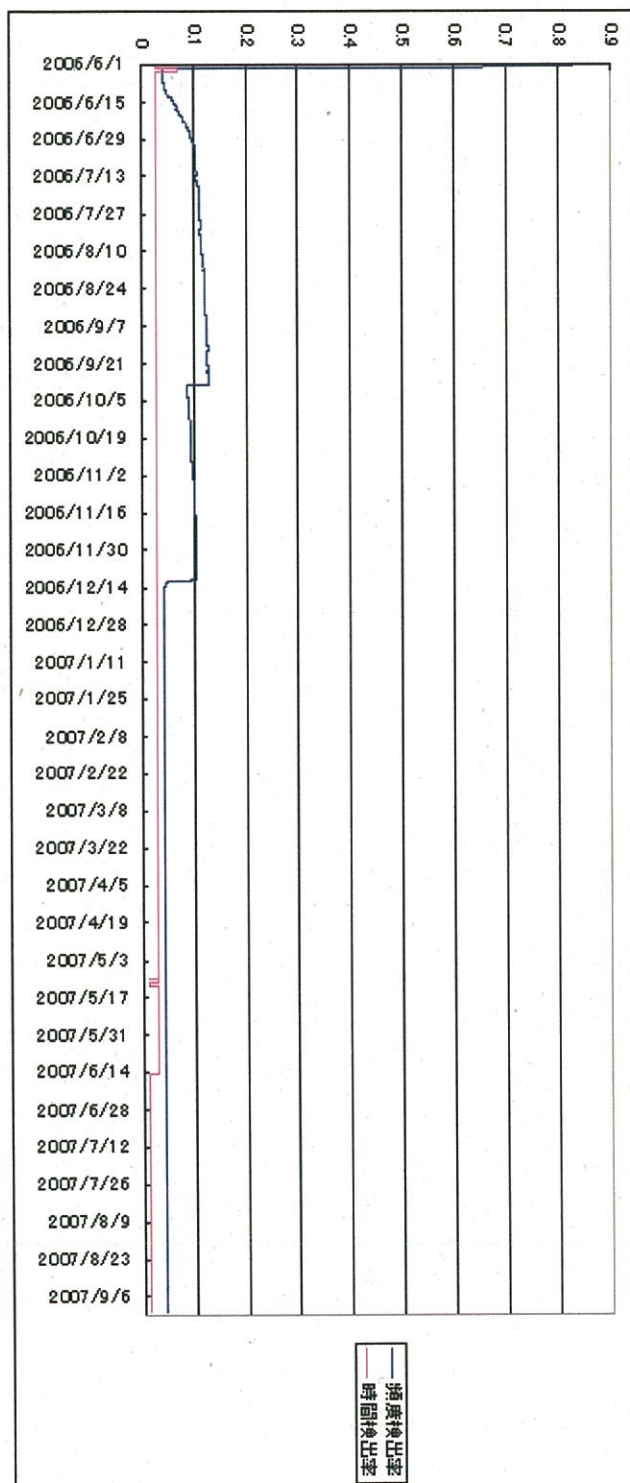


fig. 26: rate of detection using Bayesian estimation

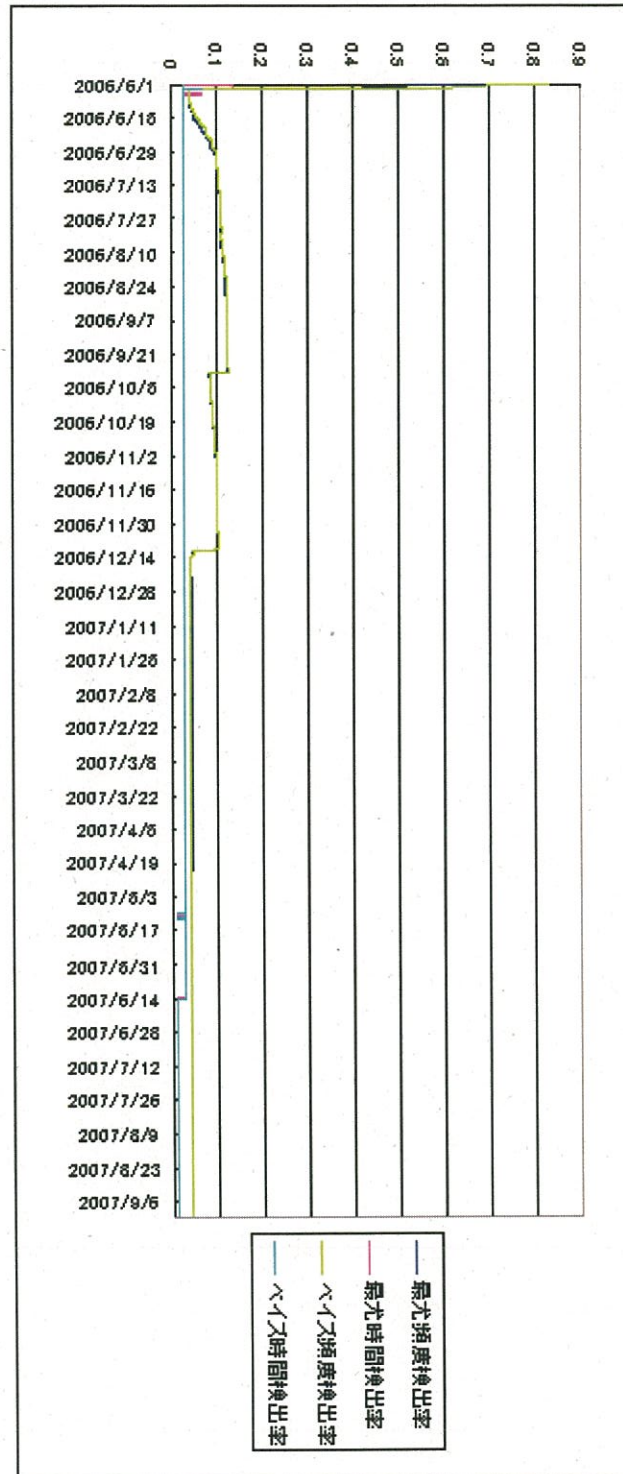


fig. 27: rate of detection using maximum likelihood estimation and Bayesian estimation



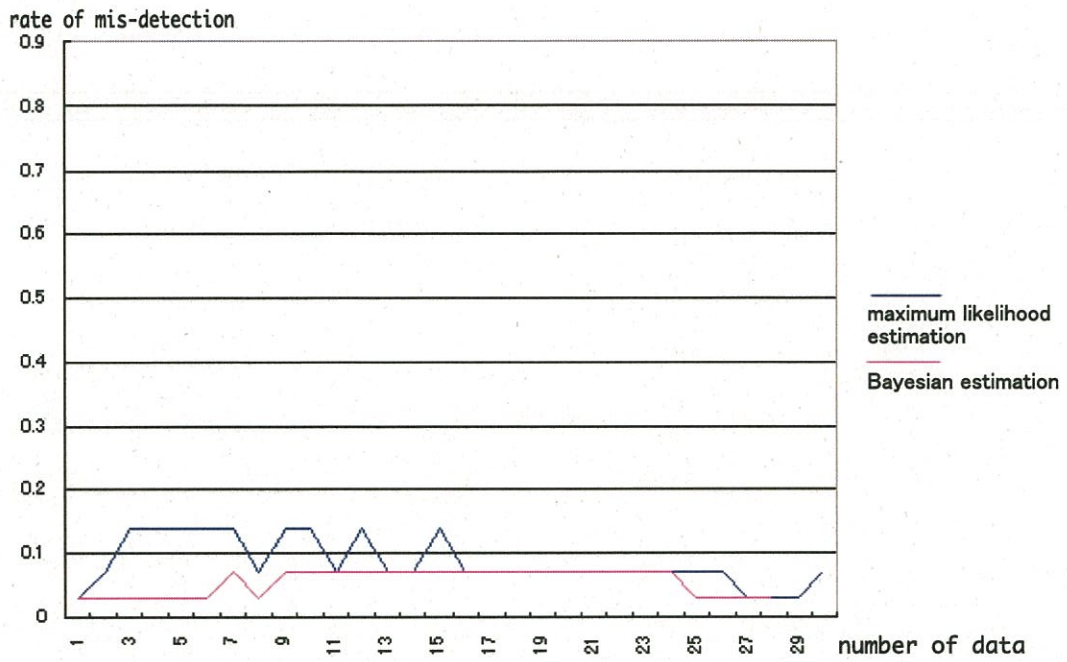


fig. 28: comparing rate of detection about length of working time

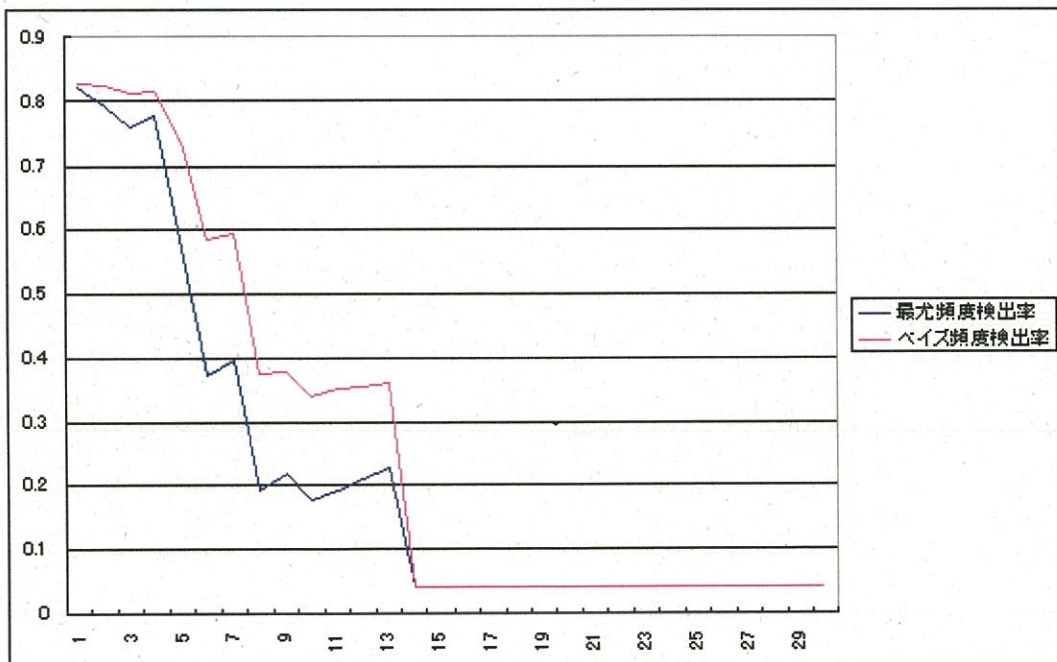


fig. 29: comparing rate of detection about interval of working

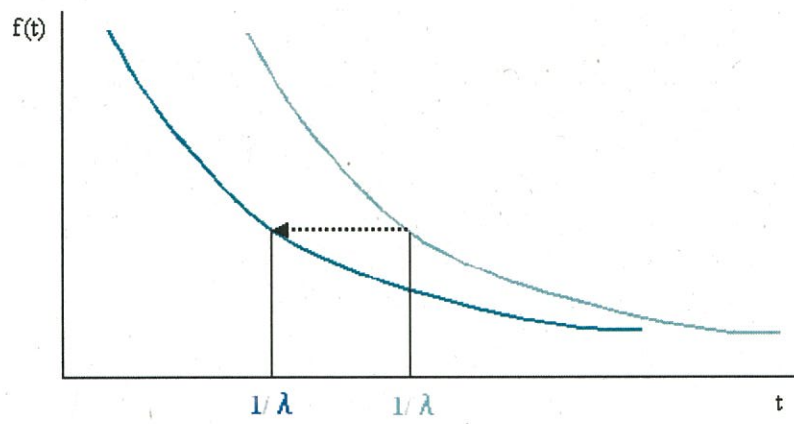


fig. 30: change of parameter of exponential distribution

#### 4.5 情報の違いによる動作時間、頻度の違いについて

定常監視システムに挙がってくる踏切障検関連の異常情報には大別して2種類ある。「障検動作」と「障故R」である。両者の違いは受光レベルの違いによるものである。「障検動作」は機器が故障してしまったり、間に障害物がありレーザーが完全に遮断されたときに挙がる。これに対して障故Rは受光器の窓が汚れているなど、受光するレーザーが完全に遮断されるわけではないが受光レベルがある値に達しなくなると挙がる異常情報である。

**障検動作** 障検本体が故障した、障検の光線が遮られた、など

**障故R** 障検投受光器のお互いの光軸がずれたり、受光器側である一定のレベルが取れない、など

原因が同一であっても受光レベルによって挙がってくる情報が異なるのだが、この異常情報を一連の異常情報として一緒くたにして扱うのか、それとも情報の種類ごとに別々に扱うのかを検討する。図31、図32、表3、表4に踏切Dに関して調べた異常情報ごとの基本統計量とそのヒストグラムを示す。ヒストグラムは縦軸に頻度、横軸にそれぞれ動作時間、回復時間をとっている。頻度に関しては「障故R」の件数が少ないため「障故R」のみ5倍してある。

table 3: length of working time about 2 type information

| length of working time "障検動作" |       | length of working time "障故R" |          |
|-------------------------------|-------|------------------------------|----------|
| average                       | 3.34  | average                      | 36.69    |
| standard variation            | 4.39  | standard variation           | 142.15   |
| variance                      | 19.29 | variance                     | 20207.27 |

table 4: interval of working about 2 type of information

| interval of working "障検動作" |             | interval of working "障故R" |              |
|----------------------------|-------------|---------------------------|--------------|
| average                    | 1600.70     | average                   | 6620.31      |
| standard variation         | 4459.78     | standard variation        | 14165.34     |
| variance                   | 19889625.25 | variance                  | 200656718.02 |

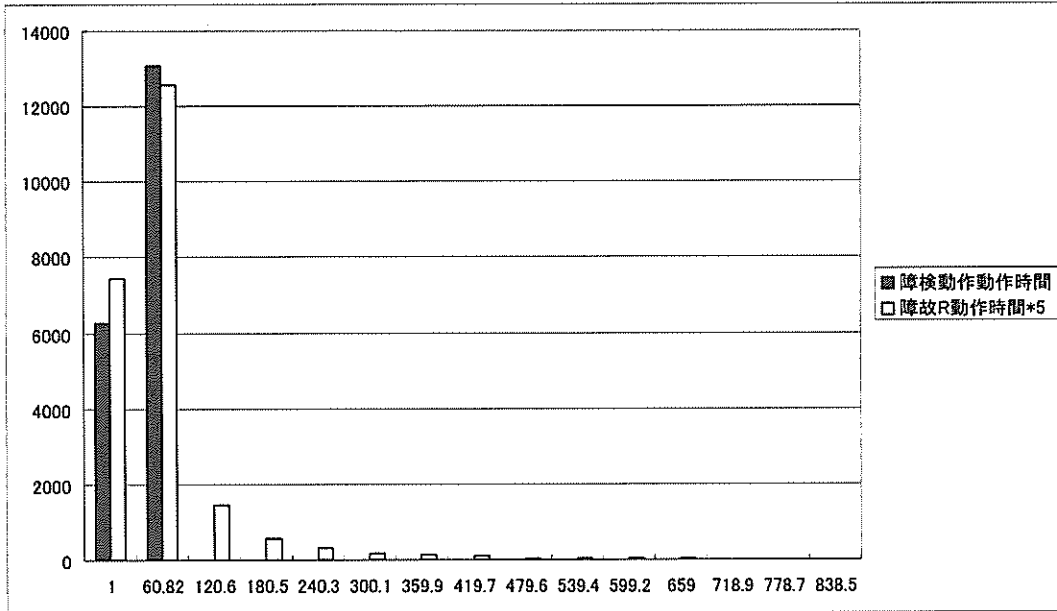


fig. 31: difference of length of working time among some type of information

これから分かるようにそれぞれの異常情報は動作間隔では同じような特性を示すが、動作時間に関しては障検動作は短い障故Rは長くなっていて特性が異なるため、一緒くたに扱えば正しい閾値の設定が出来ず、正確な異常検知が出来なくなってしまうと考えられる。したがって「障検動作」「障故R」は別々の情報として扱う必要があることが分かる。このため2種類の異常情報それぞれについて閾値を設定して診断する。

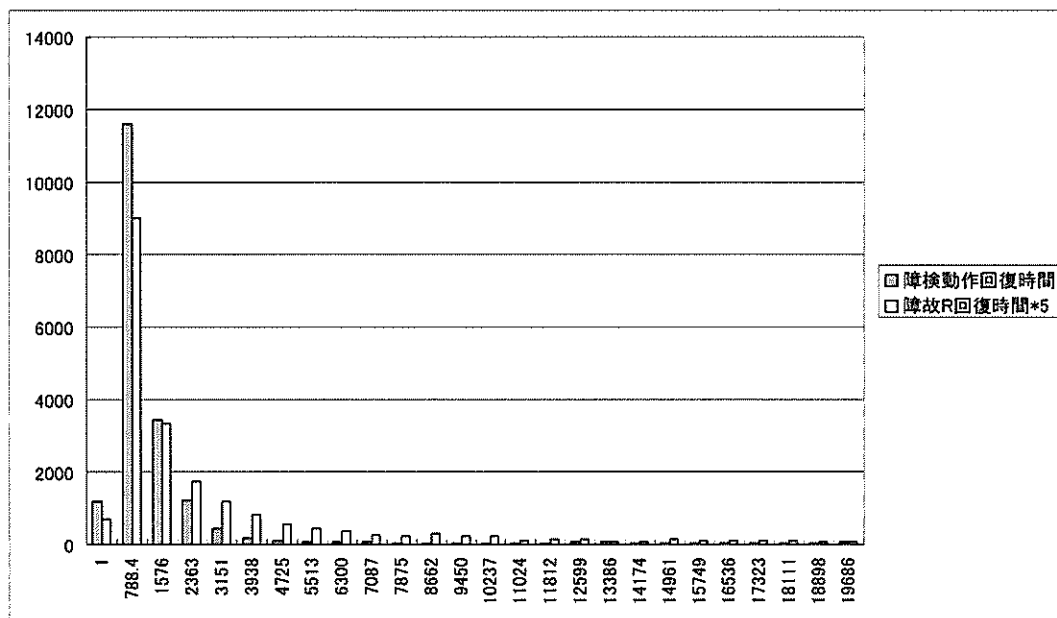


fig. 32: difference of length of interval of working among some type of information

#### 4.6 結言

本章では異常情報をモデル化し、そのモデルに従って2段階のフィルターを作成した。第1段階としては以前に異常があったか(誤情報があるか)、第2段階では挙がってきた異常情報が誤情報のときと異なった挙がり方をしているかどうかを見るフィルターである。そしてそのフィルターを用いて各種パターンの異常が検出できることを示した。またフィルターに用いる指数分布のパラメータの推定方法に関する考察を行った。

## 5 統計的フィルターの検証

### 5.1 緒言

本章では作成したフィルターを用いて実際にはどの程度の異常が発生すれば異常として検出できるのかという異常の検出限界について検証した。なお実際に発生した異常が少ないため、今回は異常状態のシミュレーションデータを作成して検証を行っている。異常情報のシミュレーションデータの作成方法を説明し、その後、検証を行った結果を示す。続いて実機データへ適用、検証を行った。実際には異常が発生していない「正常踏切」に対しては、フィルターを適用した結果、どの程度誤検知を低減できるのかを検討、実際に異常の発生した「異常踏切」に対してはフィルターを適用し異常が確実に検知できているかを検討した結果を示す。

## 5.2 異常時の異常情報

今回は実際に真の異常があった踏切 A の異常情報を元に擬似的に異常情報を作成した。踏切 A の異常情報を用いたのは異常時の異常情報が多く、分布の推定をしやすかったためである。ここから得られた分布を元に、真の異常情報の少ない踏切に対して同じ分布を仮定すると大体当てはまっていると考えられる。このときのヒストグラムを図 33、図 34 に示す。

この結果から、動作時間は正規分布、動作間隔は指数分布にしたがっていると考え異常時の分布を数種類作成し、フィルターを用いてどの程度までの異常なら検出できるのかを調べた。

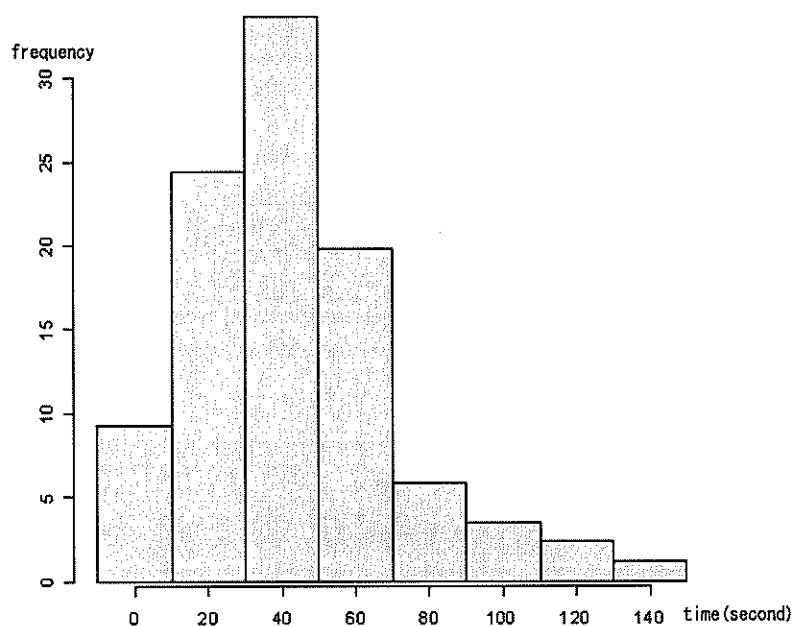


fig. 33: length of working time on Crossing A



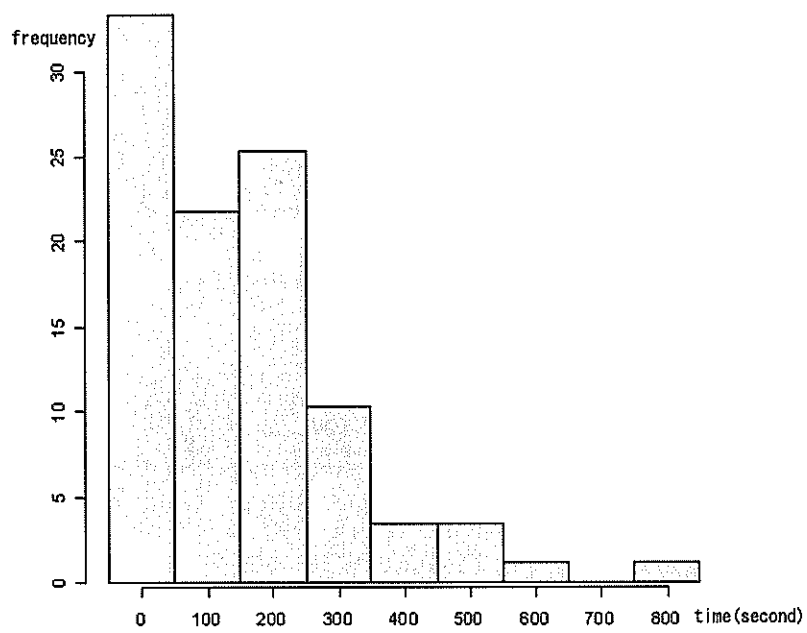


fig. 34: interval of working on Crossing A

### 5.3 検出限界の検討

実際に作成した異常時の分布とは以下の通りである。

まず異常状態の動作時間は誤情報に比べて長くなるため、誤情報のときよりも長い動作時間を発生させる分布として、平均値が誤情報のときの250%、500%、1000%の3種類の正規分布を作成し、その分布から異常情報を作成している。なおこのとき分散は誤情報のときと同じにしてある。

次に動作頻度に関してである。この場合も同様に誤情報のときよりも短い動作間隔になったときを異常状態としている。そのため、誤情報における動作間隔の平均値の1%、2.5%、5%、10%とした4種類の指数分布を作成し、その分布から異常情報を作成している。

以上のように作成した異常情報を用いて検証を行った。今回は検証に用いる踏切を踏切Bとした。その結果を図35から図41、表5、表6に示す。図、表からも分かる通り、動作時間に関するフィルターでは動作時間が250%ほどの時はまだ検知率が低い。500%ほどになるとほぼ100%の検知が来ている。動作時間が250%のときは19%であったが閾値を分布の上側5%としているのでその4倍くらいの検知は来ている。

また動作時間に関するフィルターにおいては動作間隔10%で12%の検知率、5%で34%、2.5%で54%、1%で88%となっている。動作時間が5%ほどになると閾値である5%の7倍ほどの検知率となっている。

実際のデータから比べると誤情報と真の異常情報のあった踏切Cでは動作時間は異常時に596%、動作間隔は0.03%となっているため、検証を行ったケースよりも検知はしやすい状況である(表7)。また複数

挙がってくる真の異常情報の場合、その全てを検知する必要はなく、1件でも検知できれば「正常ではない」という判断に至るので十分検知できると考えられる。

table 5: result about length of working time

| length of working time | 250% | 500% | 1000% |
|------------------------|------|------|-------|
| rate of detection      | 19%  | 100% | 100%  |

table 6: result about interval of working

| interval of working | 1%  | 2.5% | 5%  | 10% |
|---------------------|-----|------|-----|-----|
| rate of detection   | 88% | 54%  | 34% | 12% |

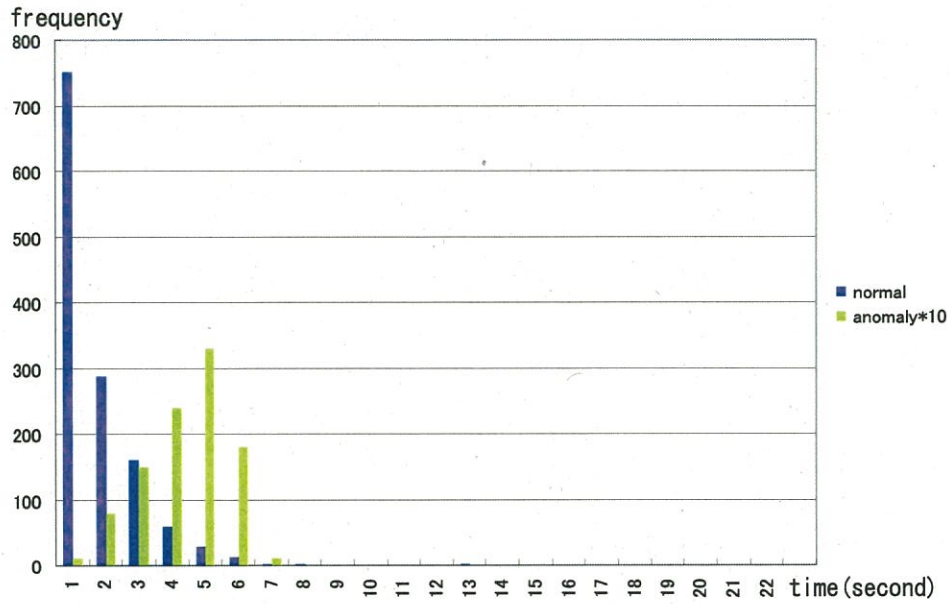


fig. 35: length of working time 250%

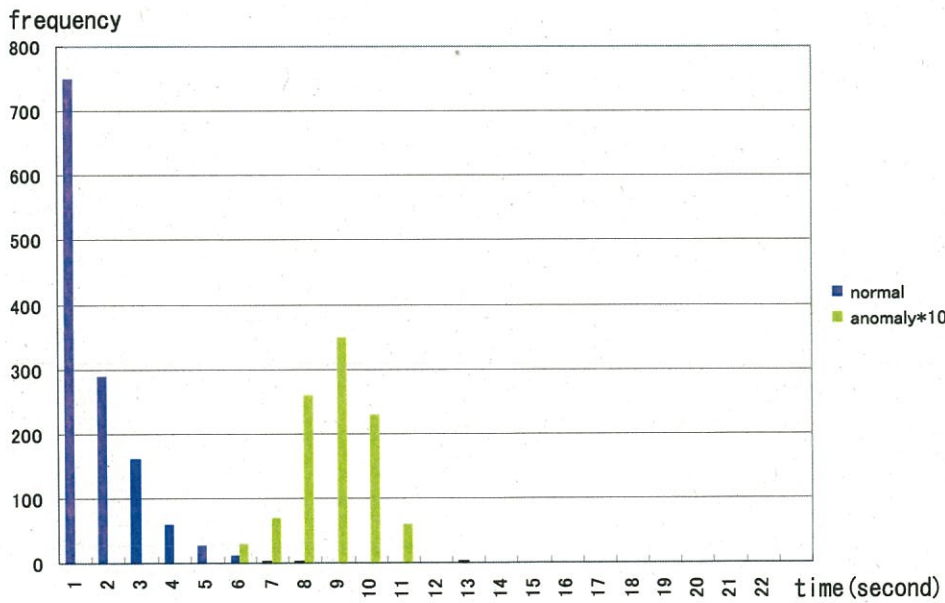


fig. 36: length of working time 500%

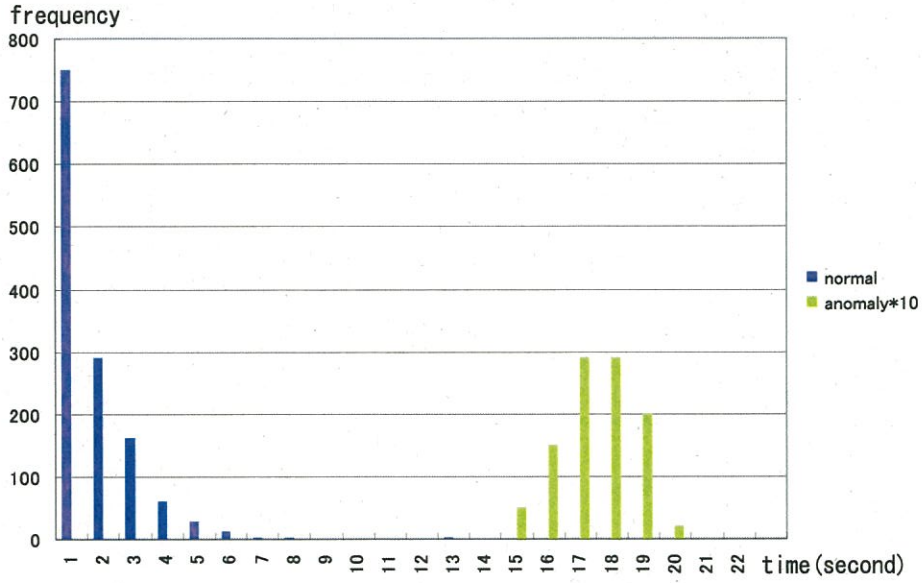


fig. 37: length of working time 1000%

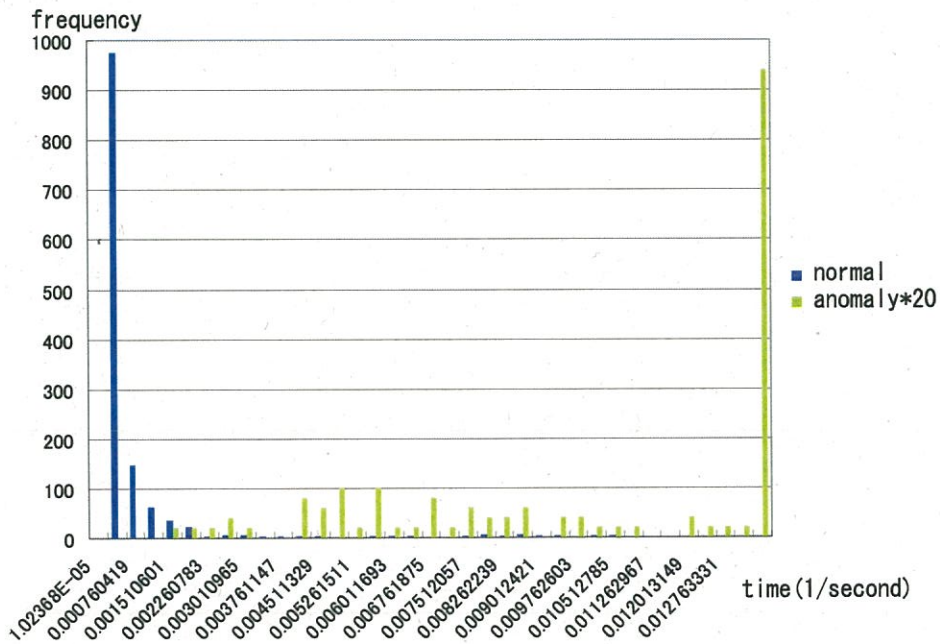


fig. 38: interval of working 1%

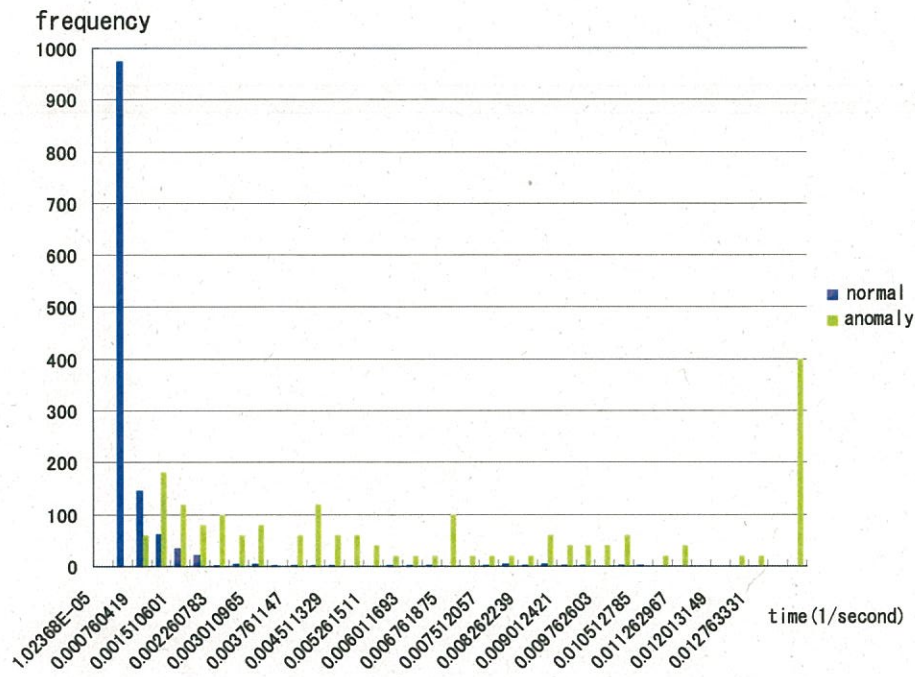


fig. 39: interval of working 2.5%

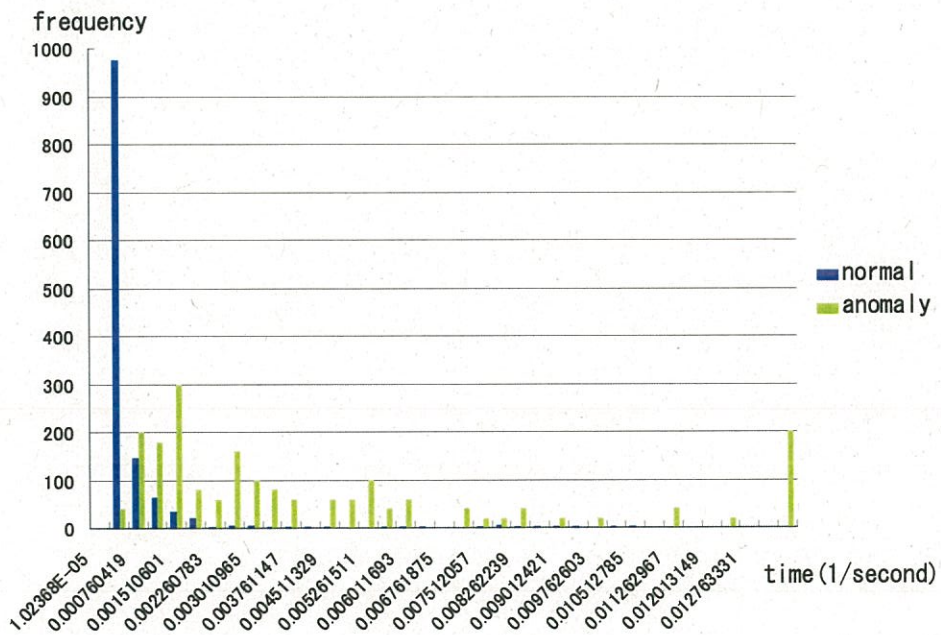


fig. 40: interval of working 5%

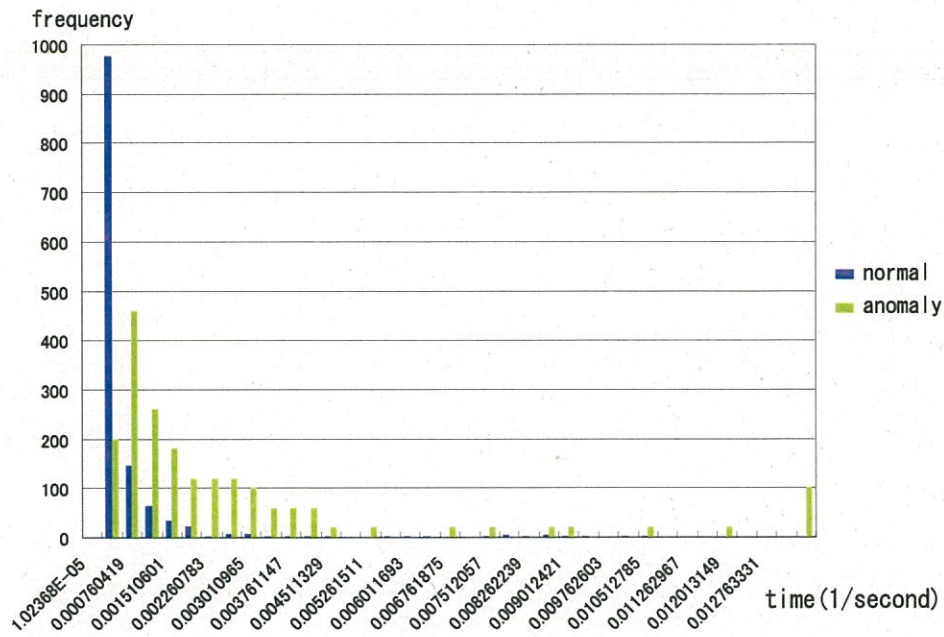


fig. 41: interval of working 10%

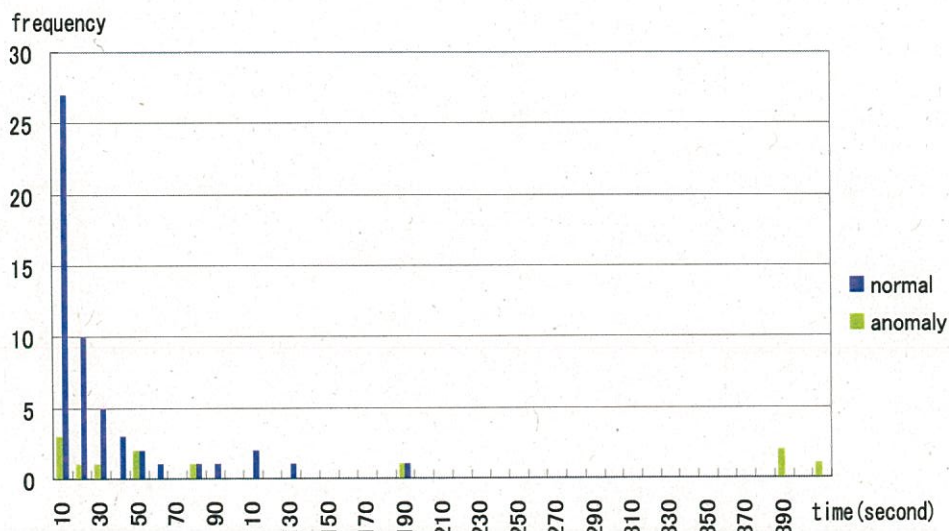


fig. 42: length of working time on Crossing C

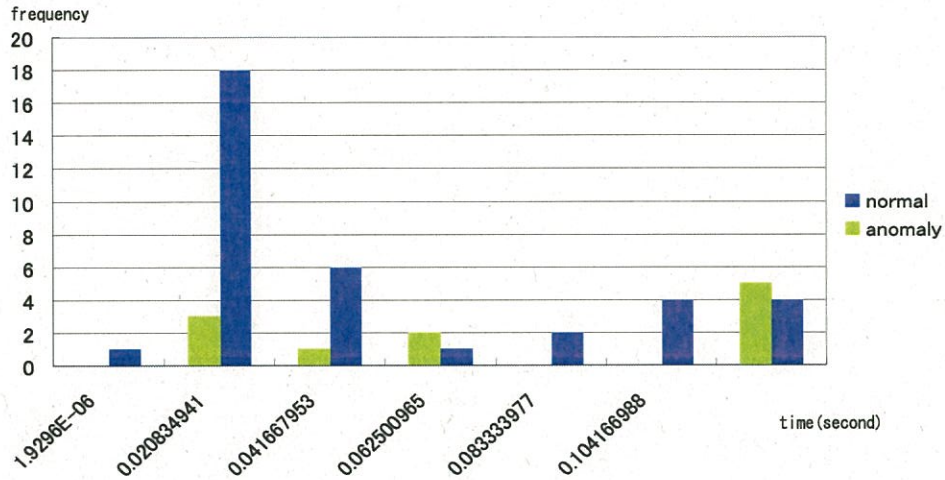


fig. 43: interval of working on Crossing C

table 7: data of Crossing C

|                        | average (normal) | average (abnormal) | rate(abnormal/normal) |
|------------------------|------------------|--------------------|-----------------------|
| length of working time | 23.9 second      | 142.5 second       | 596%                  |
| interval of working    | 88947 second     | 29.55 second       | 0.03%                 |

## 5.4 検証結果

### 5.4.1 正常踏切

本節では実際にフィルターを用いた異常検知を誤情報のみの踏切に対して行った結果を示す。誤情報のみの踏切であるため、今まで挙がってきた異常情報がこのフィルターを適用するとどの程度まで減らせるか検証した結果を示す。

管轄内の踏切 1500 箇所余りに対して検証を行った。実際にフィルターを適用した結果のうち、3 箇所を例として表 8 の CrossingD から CrossingF まで、e 右は時の列に、全踏切に対して行った結果を示す。表は上から、全件数、動作時間フィルターでの検知件数、誤検知率、動作間隔のフィルターでの検知件数、誤検知率となっている。

全踏切に対して検証を行った結果、誤検知率が動作時間フィルターで 7.21%、頻度フィルターで 6.23% だった。すなわち現在検知している異常情報がフィルターを適用することにより 7% ほどに減らせることが分かる。実際には同時に 2 つのフィルターにかけて診断を行うため両方のフィルターにかかっている異常情報もある。

table 8: result of normal crossing

|                           |                          | Crossing D | Crossing E | Crossing F | All crossing |
|---------------------------|--------------------------|------------|------------|------------|--------------|
| Total                     |                          | 37418      | 16971      | 13983      | 291499       |
| Filter of<br>working Time | Number of<br>detection   | 3252       | 1214       | 420        | 21017        |
|                           | Rate of<br>mis-detection | 8.69%      | 7.15%      | 3.00%      | 7.21%        |
| Filter of<br>interval     | Number of<br>detection   | 2190       | 1583       | 275        | 18160        |
|                           | Rate of<br>mis-detection | 5.85%      | 9.33%      | 1.97%      | 6.23%        |



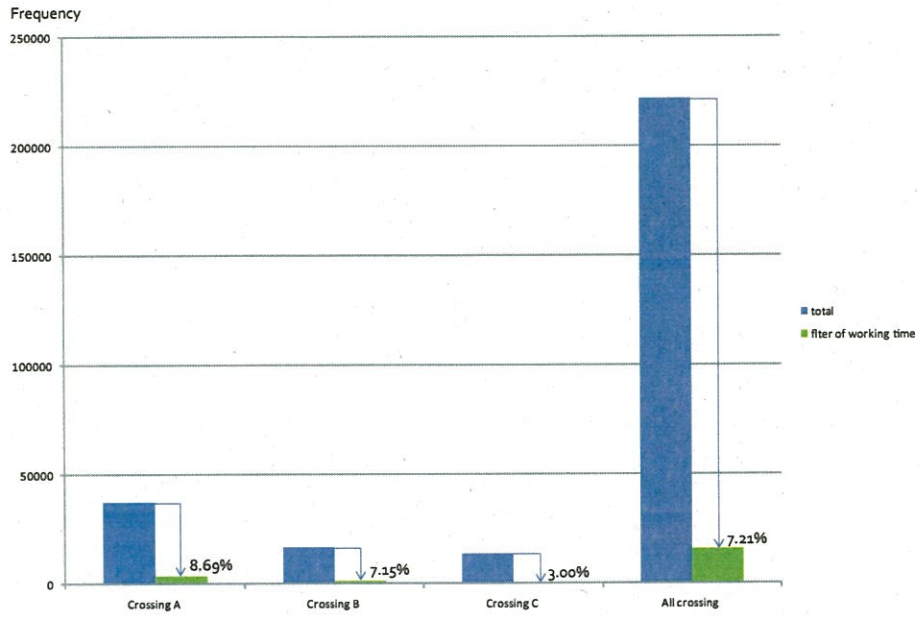


fig. 44: after applying filter(length of working time)

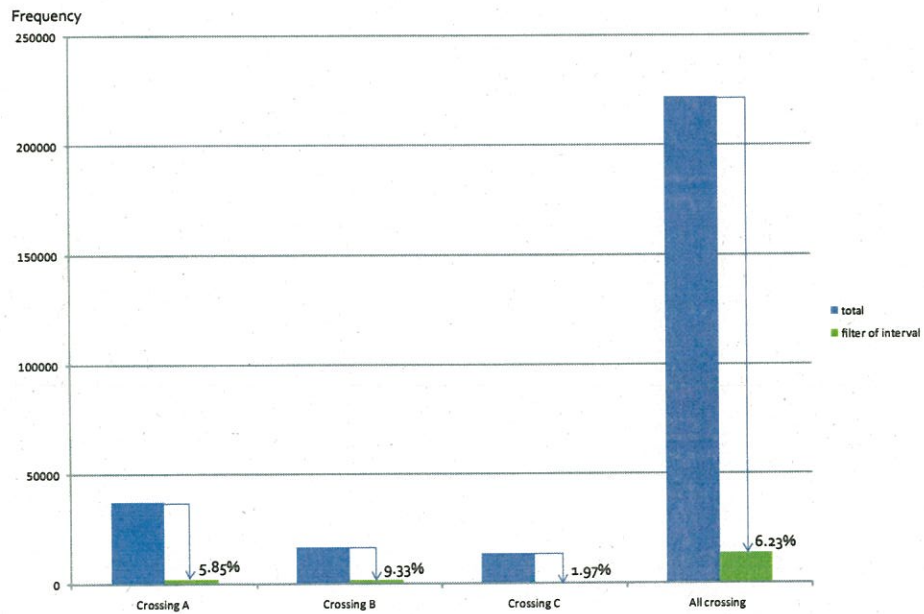


fig. 45: after applying filter(interval of working)

## 5.4.2 異常踏切

次に実際に異常が発生した踏切件に対してフィルターを用いて異常検知が出来るかどうかを行った。実際に異常が発生している踏切であるため確実に異常を検知できる必要がある。各踏切の異常情報に対して検知フローにかけた結果をパターンごとにまとめた。

パターン1に対して検証を行うと兆候状態がないので兆候状態での検知は出来ないが、挙がってくる情報に対しては全て異常の検知が出来ている。この結果を表10に示す。表中の「○」は検知が可能であったこと、「×」は検知できなかったこと、「-」は検知に用いていないことをそれぞれ表す。踏切名のすぐ下の行に異常検知が出来たかどうかを示してある。その下にはどのフィルターによって異常の検知が出来たのかを示してある。

table 9: about simbol

|   |              |
|---|--------------|
| ○ | detectable   |
| × | undetectable |
| - | not use      |

table 10: result of pattern1

|           |              | Crossing A | Crossing B | Crossing C | Crossing D |
|-----------|--------------|------------|------------|------------|------------|
| detection |              | ○          | ○          | ○          | ○          |
| filter    | filter1st    | ×          | ×          | ×          | ×          |
|           | working time | ○          | ○          | ○          | ○          |
|           | interval     | ○          | ×          | -          | ×          |

|           |              | Crossing E | Crossing F | Crossing G | Crossing H |
|-----------|--------------|------------|------------|------------|------------|
| detection |              | ○          | ○          | ○          | ○          |
| filter    | filter1st    | ×          | ×          | ×          | ×          |
|           | working time | ○          | ○          | ○          | ○          |
|           | interval     | -          | -          | -          | -          |

パターン2に対して検証を行うと以前に異常情報があるかどうかのフィルターで全て検知できる。この結果を表11に示す。

パターン3に対しては以前に異常があるかどうかのフィルターだけでは検知できず、2段階目の2種類のフィルターを適用して異常検知を行った。パターン3に該当するのは1件のみであったのでこの1件に対して検知フローを適用した結果を表12に示す。この踏切においては動作時間に関するフィルターと動作頻度に関するフィルターの両方で検知できていることが分かる。

table 11: result of pattern2

|           |              | Crossing I | Crossing J | Crossing K | Crossing L |
|-----------|--------------|------------|------------|------------|------------|
| detection |              | ○          | ○          | ○          | ○          |
| filter    | filter1st    | ○          | ○          | ○          | ○          |
|           | working time | -          | -          | -          | -          |
|           | interval     | -          | -          | -          | -          |

table 12: result of abnormal detection(pattern3)

|                                      |              | Crossing C |
|--------------------------------------|--------------|------------|
| detection                            |              | ○          |
| threshold level(working time:second) |              | 71.78      |
| threshold level(interval)            |              | 9.07       |
| filter                               | 1st          | ×          |
|                                      | working time | ○          |
|                                      | interval     | ○          |

## 5.5 結言

第4章で設定したフィルターを用いた場合の異常検出のシミュレーションと実際の異常のデータの比較を行い、十分異常検出が出来ることを示した。異常時にはすべてのパターンにおいて異常の検知ができています。

## 6 HMM を用いた異常検知方法

### 6.1 緒言

本章では HMM(隠れマルコフモデル) について説明する。

はじめに隠れマルコフモデルの概略について説明し、その適用分野について説明する。隠れマルコフモデルの理論的説明に関しては 2 章で行っているため、ここでは本研究での適用方法について具体的に説明する。

また本研究で隠れマルコフモデルが必要となる異常情報について具体的に示し、隠れマルコフモデルの適用理由について説明する。

次に各種パラメータの設定方法、HMM での検出限界について検討した上で実機データを用いて検証を行った結果を示す。

## 6.2 HMM の選定理由

本研究において異常検知の手法としてHMMを採用した。隠れマルコフモデルは離散量を扱うことができ、時系列モデルのパターン認識に対して強いという特徴がある。本研究では正常、異常の1, 0の離散量を扱い、その時系列データに関して診断を行う。

したがってHMMが適していると判断してHMMを本研究における異常検知のための手法として選定した。

### 6.3 本研究への適用方法、解析手順

#### 6.3.1 適用方法、解析手順

本研究では HMM を用いて異常診断を行う。前節で述べた通り、正常でも異常情報の挙がってくる状態のデータを学習させると、このパターンに近いデータ、すなわち誤情報のデータに対しては記号出現確率が高くなり、そのスコアは低くなる。また真の異常情報など、誤情報とは異なったパターンのデータに対しては記号出現確率は低くなり、そのスコアは高くなる。したがってスコアを見ることにより正常時のパターンからどのくらい離れているか分かり、大きく離れている場合には異常と判断できる。

次に記号列の作成方法について説明する。記号には 1 秒ごとの障検の動作状態を使う。障検が作動しているとき (動作状態)、作動していないとき (回復状態) をそれぞれ 1/0 と表す。この 1/0 が記号となる。そしてその記号を並べたものである記号列は、200 秒間、すなわち 200 個の 1/0 を並べたものを 1 つの記号列とし、以下のような記号列を考える。具体的に数字を入れた記号列を式 20 に示す。

$$O = \{x_1, x_2, x_3, x_4, \dots, x_{197}, x_{198}, x_{199}, x_{200}\} \quad (19)$$

$$O = \{0, 0, 1, 0, \dots, 0, 0, 1, 0\} \quad (20)$$

このようにして 1 つの記号列を作成する。

この記号列を毎秒 1/0 のデータを取得し記号列を作成する。この記号列を  $m$  秒間に渡って作成した  $m$  個の記号列  $\{O_1, O_2, O_3, \dots, O_m\}$  の集合を表すと以下のようになる。

$$\begin{aligned} O_1 &= (x_{t_1} \ x_{t_2} \ x_{t_3} \ x_{t_4} \ \dots \ x_{t_{197}} \ x_{t_{198}} \ x_{t_{199}} \ x_{t_{200}}) \\ O_2 &= (x_{t_2} \ x_{t_3} \ x_{t_4} \ x_{t_5} \ \dots \ x_{t_{198}} \ x_{t_{199}} \ x_{t_{200}} \ x_{t_{201}}) \\ O_3 &= (x_{t_3} \ x_{t_4} \ x_{t_5} \ x_{t_6} \ \dots \ x_{t_{199}} \ x_{t_{200}} \ x_{t_{201}} \ x_{t_{202}}) \\ &\vdots \\ O_m &= (x_{t_m} \ x_{t_{m+1}} \ x_{t_{m+2}} \ x_{t_{m+3}} \ \dots \ x_{t_{m+196}} \ x_{t_{m+197}} \ x_{t_{m+198}} \ x_{t_{m+199}}) \end{aligned}$$

この  $m$  秒分のデータ、 $m$  個の記号列を用いて HMM の学習を行う。

記号列の集合を与えて学習を行った HMM に、診断時には記号列を入力させ、その記号出現確率を求める。記号出現確率は学習させたデータに類似したデータほど高い値となり、学習させたデータと異なるほど小さな値となる。したがってこの値の大小を見ることにより、診断時のデータが、学習データの正常時とどの程度離れているかを測ることができる。これにより正常・異常の診断を行う。ただし記号出現確率は 0 から 1 の確立値であり、異常の度合いが大きくなるほど小さな値を取る。

そこで異常の度合いが大きいほど、大きな値を取るようなスコアを記号出現確率の対数をとることによって以下のように定義する。

$$score = -\log\{p(O|\theta)\} \quad (21)$$

今回の検証ではまず誤情報が挙がってくる状態で記号列を作成し、その記号列を HMM に学習させていった。すなわち 0 のみの記号列は学習用データには入れていない。これにより誤情報と同じ様なパターンの異常情報に対してはスコアは低くなり、誤情報と異なったパターン、すなわち異常があると考えられる異常情報に対しては高いスコアとなる。

診断時には、時刻  $t$  と、その時刻から 1 秒前、2 秒前、...、199 秒前の障検の発生・回復状態を 1、0 で表した 200 個から記号列を作成 (式 22) し、先ほど作成させた HMM によってスコアを計算する。このスコアがある閾値を超える場合は正常ではないと判断が出来る。これらの HMM は踏切ごとに作成して診断を行う。

$$O = (x_t \ x_{t-1} \ x_{t-2} \ \dots \ x_{t-198} \ x_{t-199}) \quad (22)$$

### 6.3.2 適用例

この HMM を用いて実際に計算を行った例を図 46 から図 48 に示す。50 件前後の誤情報を学習させて踏切ごとに HMM を作成し、その HMM を用いて診断を行っている。横軸には時刻 (今回の場合、診断開始時の記号列を 1 番目とし、その 1 秒後の記号列を 2 番目の記号列、として表しているため横軸は 1 からの通し番号となっている) を取り、縦軸にはスコアを取っている。図 46 において赤い線で区切りを入れているところまでは 750 の辺りと、1500 の辺りで異常情報が挙がってきて記号列出現確率を計算したためスコアが挙がっている。このように完全に時系列に即したグラフとすると時間軸が膨大な長さになるため、以下では異常情報が挙がっていない時間に関しては低いスコアのみが与えられるだけであるため省略して示す。図 46 における 2000 以降の示し方とする。

図 47 は実際には異常の発生していない踏切である。学習データには誤情報のデータを用いているため、実際には異常がなく、誤情報のみの踏切ではスコアは低いままとなっているが、図 48 に示した実際に異常のあった踏切では異常のあった異常情報の時刻のスコアが高くなっている。図中で赤丸で囲ってある部分が異常のあった異常情報によるスコアである。このように異常が発生すると HMM のスコアは高くなるため、このスコアを用いて異常の検知をすることは十分可能であると考えられる。



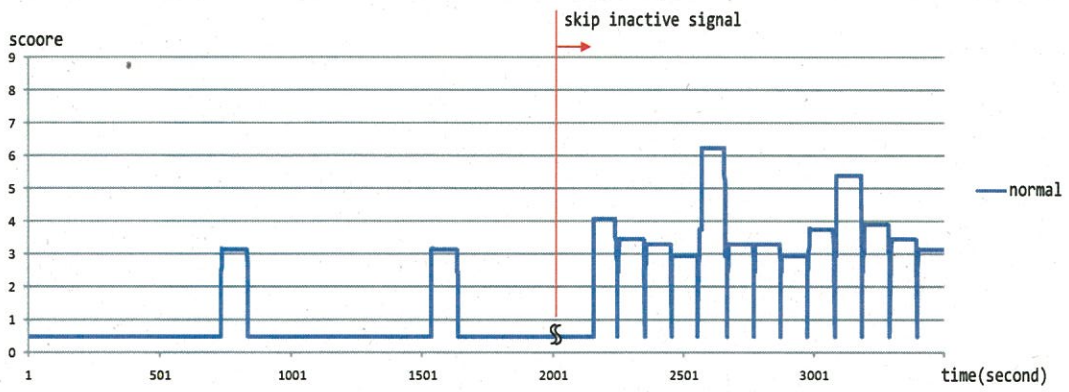


fig. 46: HMM score at Crossing G

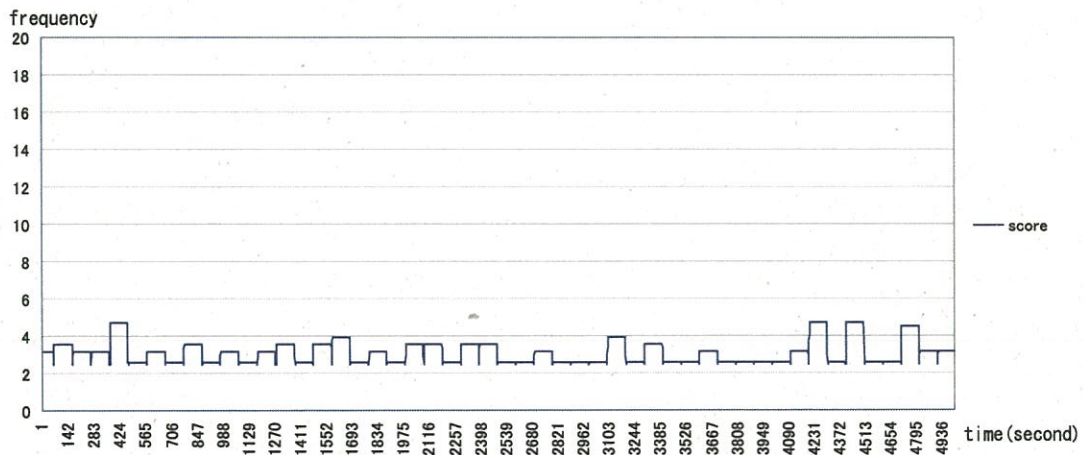


fig. 47: HMM score at Crossing H

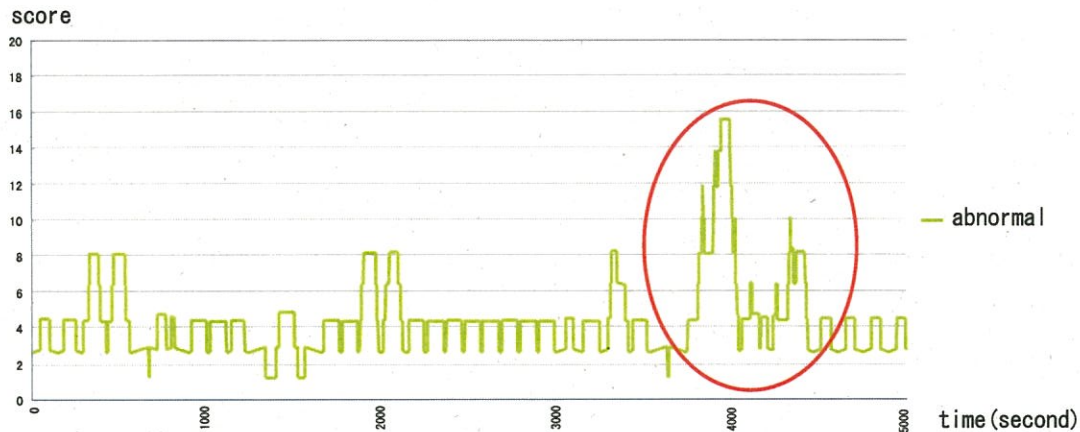


fig. 48: HMM score at Crossing G

### 6.3.3 閾値設定方法

蓄積されたデータ群の中から、全体の傾向から逸れた外れ値を検出することによって異常値としてとらえる手法がある。これは異常値の中にノイズだけではなく異常行動や新しい傾向を示す重要な情報を持ったデータが含まれているため、外れ値を検出することには大きな意味があると考えられている。このように異常値を検出する技術を「外れ値検出」という。

外れ値検出の応用分野として不正検出がある。これは不正を行った際の行動におけるデータは、正しい行動におけるデータ群とはことなデータが得られるという考えに基づく。そのため正しい行動のデータ群に対して、不正を行った際には外れ値として検出されるものであり、ネットワークの不正侵入検出、携帯電話のなりすまし利用の検出、クレジットカードの不正利用の検出、医療や保険業界における不正請求検出など応用されている。

この外れ値検出エンジンとして有名な手法に Yamanishi *et al.*(2000) により提案されている統計的外れ値検出エンジン SmartShifter がある。これはデータの入力ごとにデータのスコアを計算し、そのスコアが高いほど、外れ値の度合いが高いとみなす。ある一定期間データを蓄積し、そのデータのスコアをソートすることにより、外れ値の度合いが高いデータを簡単に探すことができ、高い確率で不正データを検出することができる。外れ値検出にはこのような手法がある。

しかしこのような手法は常に一定期間データを蓄積しなければならず、異常検出としての速報性は失われる。そこで本研究では一定期間のデータを蓄積させた後はそれまでのデータを用いて統計的に閾値を設定し、その値異常のスコアが得られた場合に異常と判断する手法を取る。閾値設定の手法としては前述の統計的フィルターと同様にスコアの分布から統計的な信頼区間を取り、その時のスコアの値を閾値とする。そして正常・異常の診断をする際、本研究ではその閾値を超えたかどうかで

判断する。

閾値設定のためには全踏切について分布を推定する必要がある。ここでは1か所の踏切での閾値設定を例として挙げる。この踏切で正常時のデータを学習させた後、あるまとまった数の診断を行いスコアを大量に取得する。そうして得られたスコアのヒストグラムを描いて分布を調べたのが図 49 である。

この分布は2母数指数分布と仮定した。2母数指数分布とは

$$f(t) = \frac{1}{\mu} e^{-(t-\gamma)/\mu} \quad (23)$$

と表わされる分布であり、信頼性工学の分野で、時刻  $\gamma$  以前には故障は発生せず、かつ  $\gamma$  以降のみ通常の指数分布に従う寿命分布を有するとする寿命値  $T$  の分布  $p.d.f$  である。ここで

$\mu$  : 尺度母数,  $\gamma$  : 位置母数

と呼ぶ。この分布を仮定しパラメータの推定を行い、その95%点となる値を閾値とした。この値と診断時のデータから得られたスコアを比較することにより診断を行う。この判定方法であればデータの蓄積を待たずに速報性を持って診断を行うことができる。

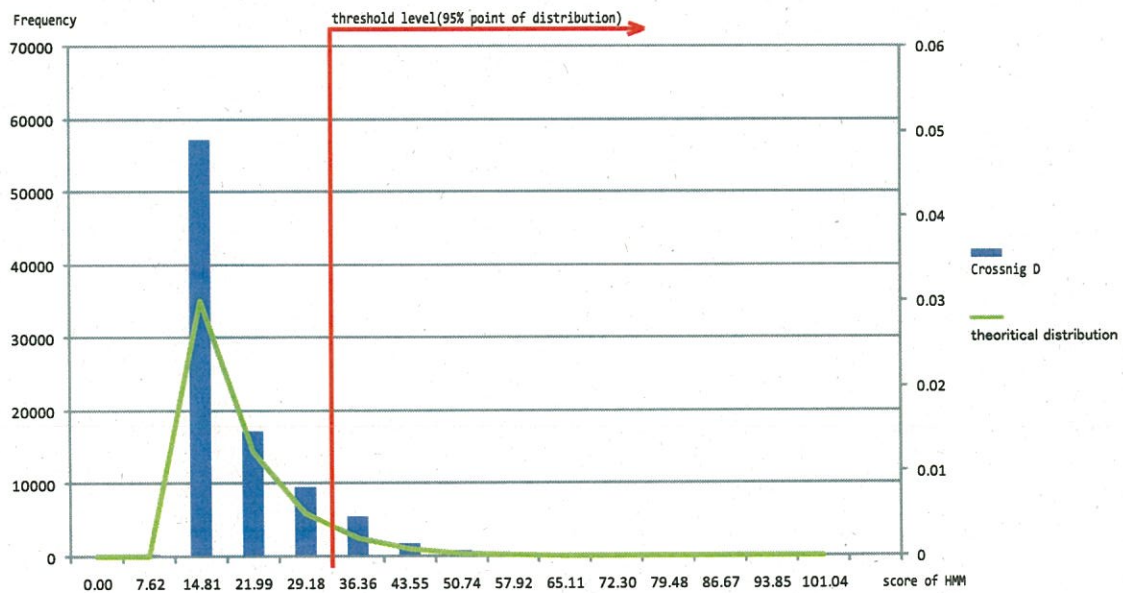


fig. 49: HMM score and threshold level

## 6.4 パラメータの設定

### 6.4.1 AIC(Akaike's information Criteria)

HMM は内部パラメータとして状態数、状態遷移確率、初期状態確率、記号出現確率などを持つ。このうち状態遷移確率、初期状態確率、記号出現確率に関しては学習によってモデルが最適になるように決定される。そこで本節では状態数の決定方法に関して検討を行う。

HMM で学習を行い診断を行うためには HMM のモデルが診断を行うにあたって良いモデルである必要がある。ここで状態数を決定するためにモデルの良さについて検討する。「良いモデル」として尤度のみをモデルの良さとするならば状態数を増やしていくほど複雑にモデルを表せるようにあるため学習データに対する尤度は挙がる。しかしこの場合、状態数が増えるために推定すべきパラメータの数が増えてしまい汎用力が低くなり、診断データに対する診断精度が下がってしまう。そこで「モデルの良さ」を表すための基準が必要となる。本研究では AIC(Akaike's Information Criteria) という情報量基準を用いる。これはモデルの当てはまりの良さを表す項とモデルの複雑さを表す項から成り立つ。実際の分布  $p(y)$  とモデルの確率分布  $f(y|\theta^*)$  (ただし  $\theta^*$  は最尤推定された点であるとする) との距離を Kullback 情報量、

$$D(p, \int) = \int p(x) \log \frac{p(x)}{f(y|\theta)} \quad (24)$$

$$= \int p(x) \log p(x) d\mu - \int p(x) \log f(y|\theta^*) d\mu \quad (25)$$

を用いて定義する。式 25 における第一項はモデルとは関係のない項である。そこで第二項の期待値を推測することによって AIC は次式で定義される。

$$AIC = (-2) \sum_i \log f(y|\theta) + 2dk \quad (26)$$

ここで第一項は最大対数尤度を表しモデルの当てはまりの良さを、第二項はフリーパラメータの数でモデルの複雑さを表す。パラメータを変化させていき、この値が最小となるパラメータ数が最適なパラメータの数となる。これを用いて状態数の決定を行う。

### 6.4.2 検証結果、考察

検証には誤情報の頻発するある特定の踏切を用いて行った。状態数と AIC の値の関係を図 52 に示す。グラフは AIC の値、第一項(尤度)、第二項(パラメータ数)を表している。横軸に状態数、縦軸に AIC の値、尤度、パラメータ数を取っている。なおパラメータ数はグラフ右側に縦軸をとっている。「良いモデル」とはこの AIC

が最小となる状態数が最適な状態数をとったモデルと考えられる。したがってグラフより状態数が3~5のときが「良いモデル」となる。今回は状態数を4として以下検証を行っている。

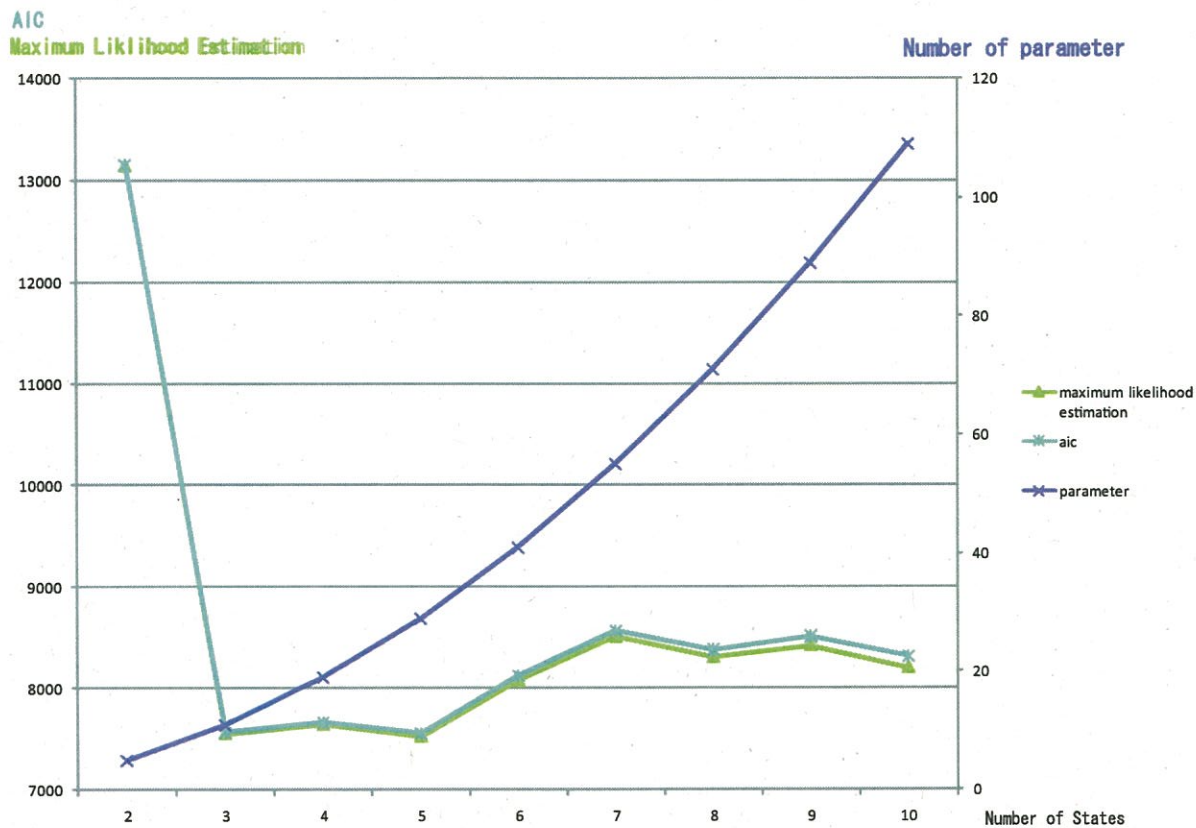


fig. 50: AIC

## 6.5 HMM での検出限界

次に HMM での検出限界について検討し、HMM の有効性について述べる。真の異常情報が少ないため、4 章の統計的フィルター同様、異常状態のシミュレーションデータを作成して検証を行った。ここでは誤情報の頻発していた踏切 D をモデルとして HMM の検出限界に関する検証を行っている。

### 6.5.1 異常時の異常情報

異常時のシミュレーションデータ作成のために踏切 D に関する統計量に関してまとめた表が表 13 である。図はこの踏切に関する動作時間とその理論分布、動作間隔の逆数とその理論分布である。

この分布をもとに異常状態について考える。今回は動作時間に関する異常、動作間隔に関する異常、動作時間・間隔共に異常、という 3 つのパターンについて検証する。さらにそれぞれな異常パターンにおいても、異常の大きさを 3 種類シミュレーションデータを作成した。動作時間、動作間隔ともに正規分布を仮定して作成している。正規分布のパラメータとして用いた平均、標準偏差の一覧を図 53 に示す。またこのときの動作時間、動作間隔の分布を図 54、図 55 に示す。

図 53 は異常のパターンごとに色分けしてある。上から動作時間に関する異常、2 段目に動作間隔に関する異常、一番下が動作時間・間隔共に異常の場合である。異常の大きさは表 13 の正常時の値から考えられる異常を、程度に差をつけ 3 つほど作成している。その時の平均、標準偏差を示している。表中で赤字となっている箇所が異常時の値である。黒文字部分は正常時と同じ値を用いている。

正常時に挙がる誤情報と、真の異常情報の動作時間、動作間隔の大きさの違いについては 5.3 の表 7 の値を参考にし、動作時間に関しては異常小が、動作間隔に関しては異常中がそれぞれ近い異常状態となっている。

図 54 は動作時間にもみ異常を発生させた場合の分布である。横軸に動作時間、縦軸に頻度を取っている。左端の分布が正常時の分布である。図中の threshold のラインは正常時の分布の 95% 点を閾値としたときの統計的フィルターの閾値となる点である。異常が発生するとそのときの分布が normal の分布に対して右側に移っている。この異常に対して HMM を用いたフィルターでも検知可能かどうか検証した。「動作時間の異常」とは動作時間にもみ任意の分布を与え、動作間隔に関してはオリジナルの誤情報と考えられる分布を用いて作成したシミュレーション上の異常状態である。

図 55 は動作時間にもみ異常を発生させた場合の分布である。横軸に動作間隔の逆数、縦軸に頻度を取っている。こちらも動作時間の場合と同様に左端に正常時の動作間隔逆数は分布している。異常が発生してくるとその動作間隔は短くなるのでそ

の逆数は大きくなり分布は右に移動してくることが分かる。この異常状態に関して異常検知可能かどうかを検証した。

図 56、図 57 は動作時間・動作間隔共に異常を発生させた場合の動作時間、動作間隔の逆数の分布である。こちらも同様に異常が発生し、その程度が大きいほど分布は右に移動する。ただし上記の 2 つと異なるのは動作時間、動作間隔単独で見た場合、異常の程度が小さいということである。そのため統計的フィルターでは検出しにくいという特徴がある。

ここで統計的フィルターの 95% を示しているのは一般的な統計量で判断の基準になり、また実際の司令員の方の判断基準よりも厳しい値であるため、この値以下で検知できることがフィルターの一種の有効性の基準となる。

table 13: statistics value about Crossing D

| length of werking time |       | interval of working |              |
|------------------------|-------|---------------------|--------------|
| average                | 3.291 | average             | 1947.030     |
| standard variation     | 2.636 | standard variation  | 5219.747     |
| variance               | 6.946 | variance            | 27245761.254 |

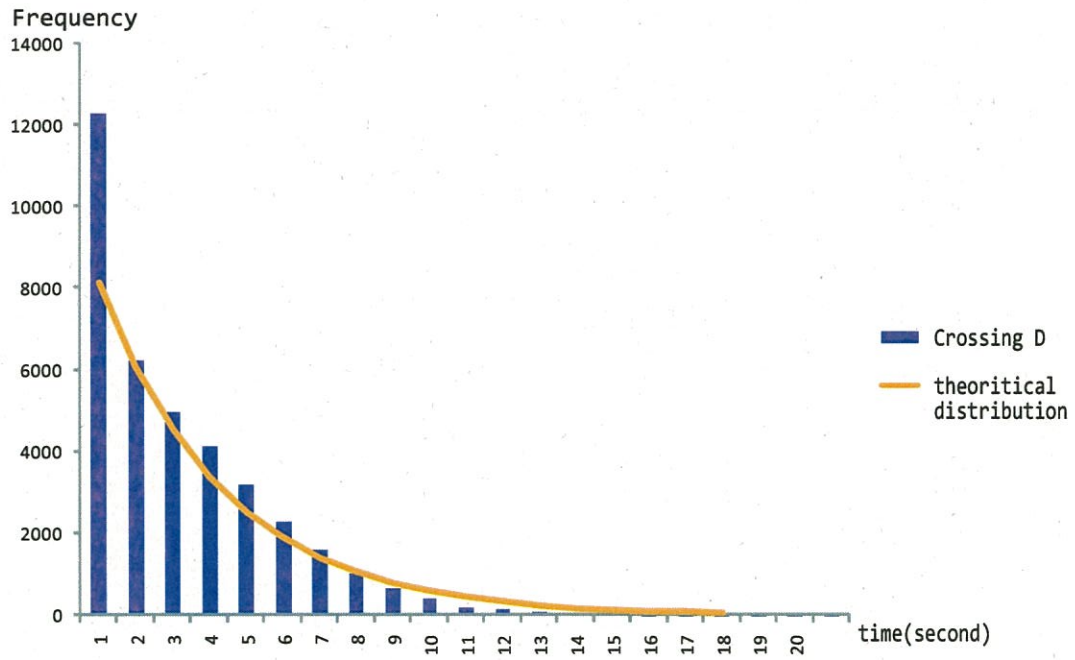


fig. 51: length of working time and theoretical distribution about Crossing D

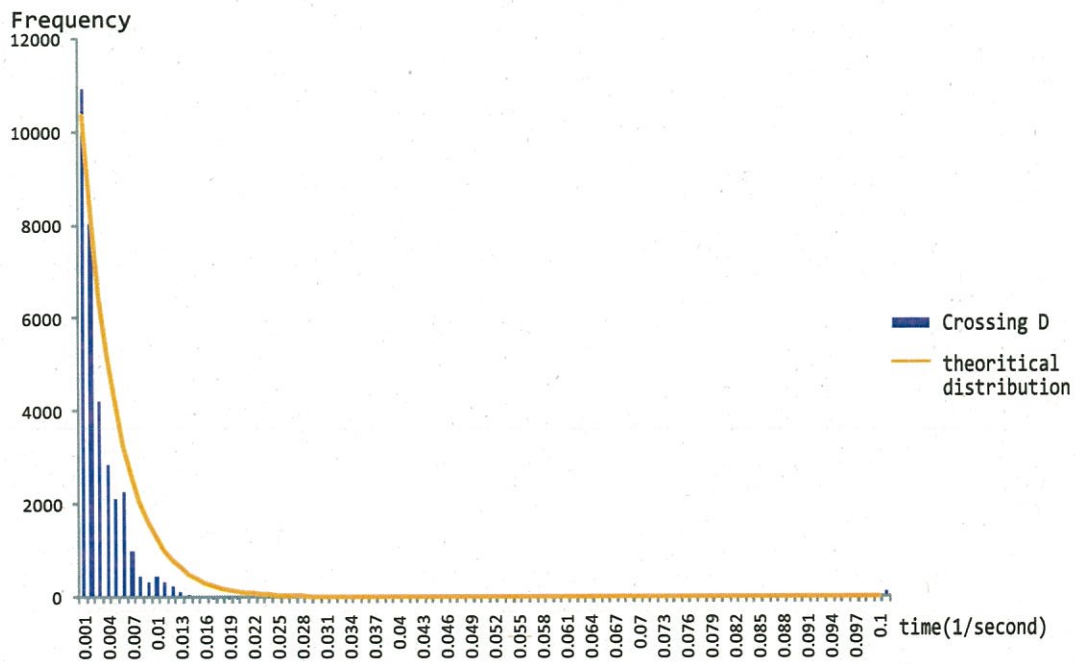


fig. 52: inverse of interval of working and distribution about Crossing D



fig. 53: type of anomaly

| Anomaly about length of working time                         |               |          |                  |          |                |          |
|--|---------------|----------|------------------|----------|----------------|----------|
|  | Anomaly (Low) |          | Anomaly (Middle) |          | Anomaly (High) |          |
|  | Working time  | Interval | Working time     | Interval | Working time   | Interval |
| Average  | 15            | 1947.03  | 30               | 1947.03  | 60             | 1947.03  |
| Standard variation   | 5             | 5219.747 | 10               | 5219.747 | 20             | 5219.747 |
| Anomaly about interval of working                            |               |          |                  |          |                |          |
|  | Anomaly (Low) |          | Anomaly (Middle) |          | Anomaly (High) |          |
|  | Working time  | Interval | Working time     | Interval | Working time   | Interval |
| Average  | 3.291412      | 500      | 3.291412         | 50       | 3.291412       | 15       |
| Standard variation   | 2.635573      | 150      | 2.635573         | 15       | 2.635573       | 5        |
| Anomaly about length of working time and interval of working |               |          |                  |          |                |          |
|  | Anomaly (Low) |          | Anomaly (Middle) |          | Anomaly (High) |          |
|  | Working time  | Interval | Working time     | Interval | Working time   | Interval |
| Average  | 4             | 150      | 10               | 100      | 15             | 50       |
| Standard variation   | 1             | 25       | 2                | 20       | 4              | 15       |

### 6.5.2 統計的フィルターで検知できない異常

図 54 から図 57 の分布図より、統計的フィルターでは検知できない異常について検討する。本異常検知システムでは司令員の方が確認できる件数まで減らすことが求められるが、仮に統計的フィルターのみで異常の検知を行うとすると 95%点を閾値として大幅に減らすことが必要である。その場合の閾値を図 54 から図 57 の分布図に載せてある。統計的フィルターのみで診断を行う場合には、この 95%点以下の情報に関しては安全と判断してしまう。しかし何らかの異常が発生する場合には、正常状態から異常状態へと徐々にシフトしていくと考えられる。この異常の兆候状態で異常を検知しようとするならば、異常状態へとシフトしていく早い段階で検知することが望まれるが、統計的フィルターではその検知を行うことが難しい。また変化の小さな異常を検知しようと閾値を下げると今度は誤情報を大量に含み、また真の異常情報が埋没してしまうと考えられる。実際にフィルターで閾値の値を変化させた場合の検知率を表 に示す。この表から、閾値を下げると真の異常の検知は難しくなることが分かる。

このことから HMM を用いたフィルターには統計的なフィルターでは検知できない異常状態を検知することが望まれる。

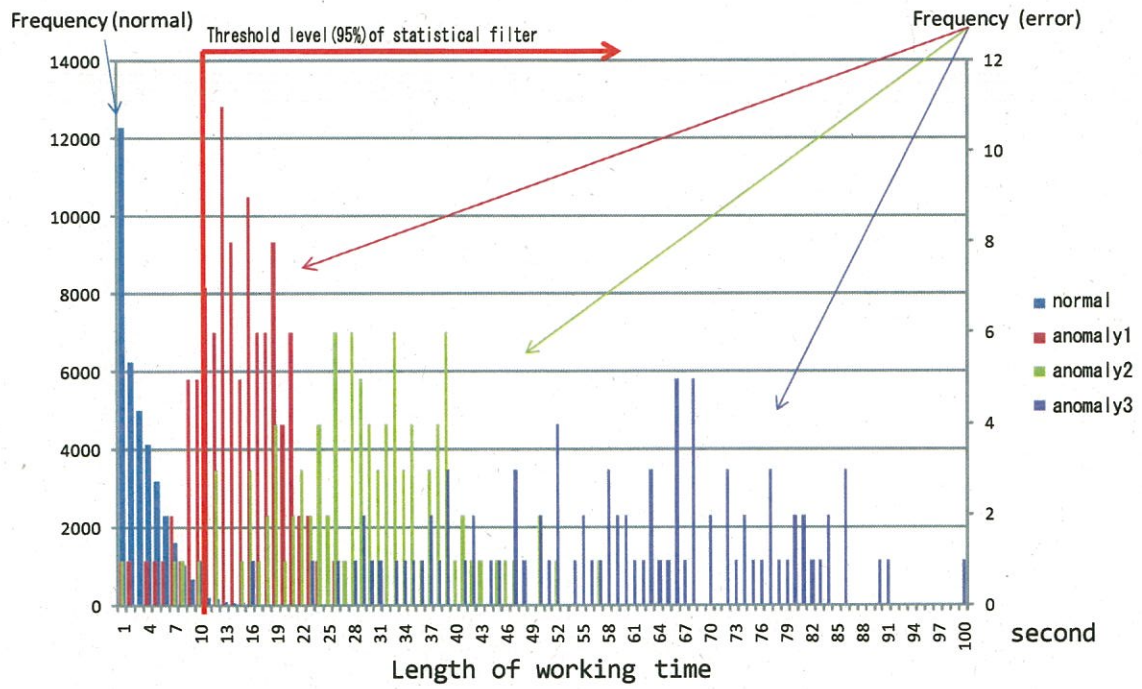


fig. 54: anomaly about length of working time

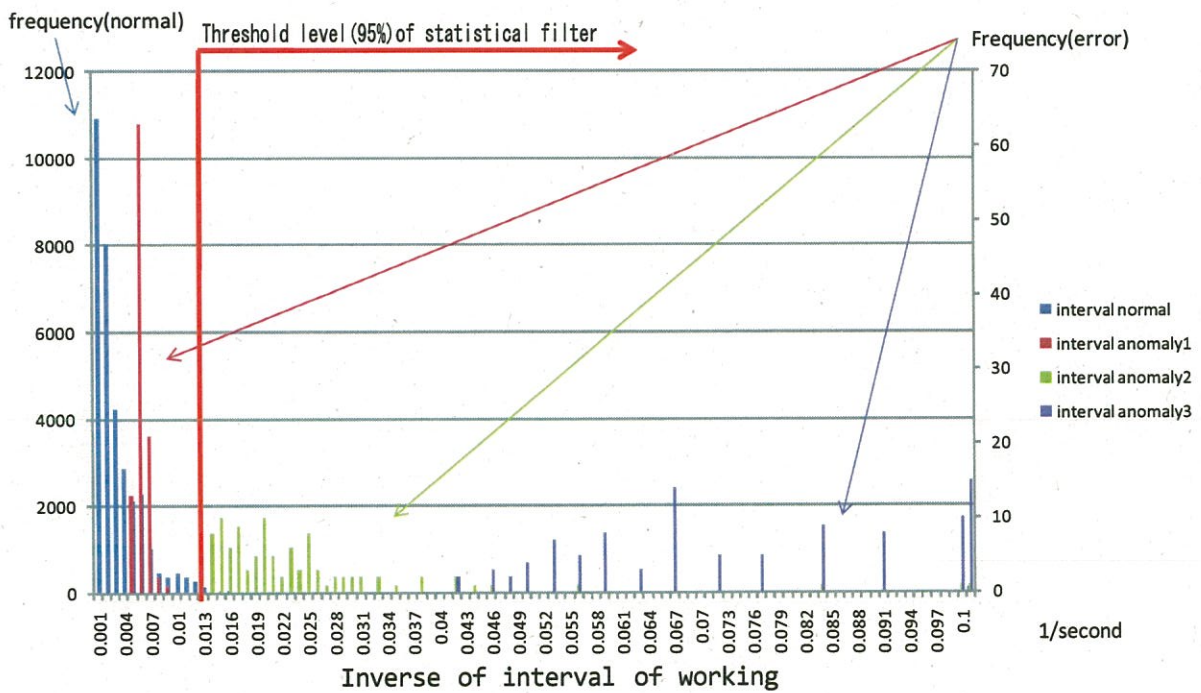


fig. 55: anomaly about interval of working

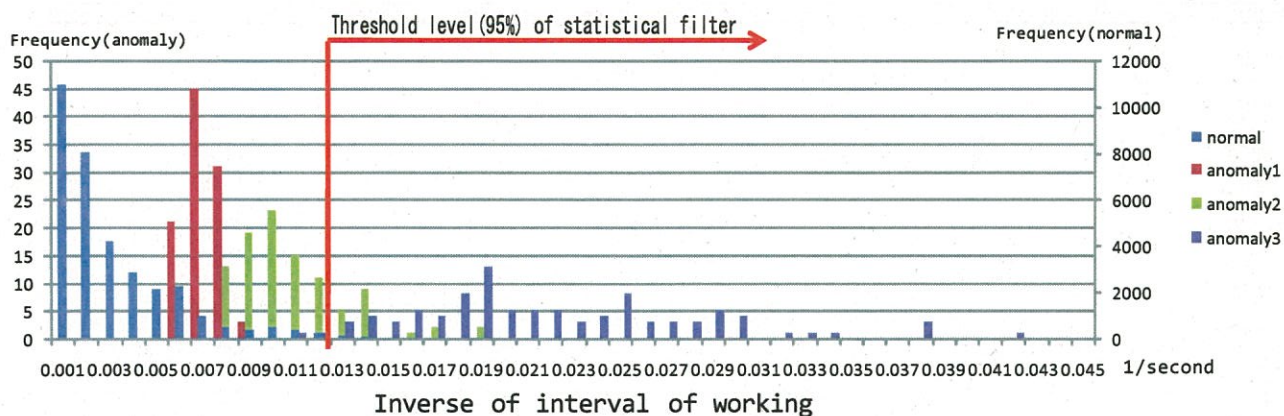


fig. 56: histogram of working time of anomary about length of working time and interval of working

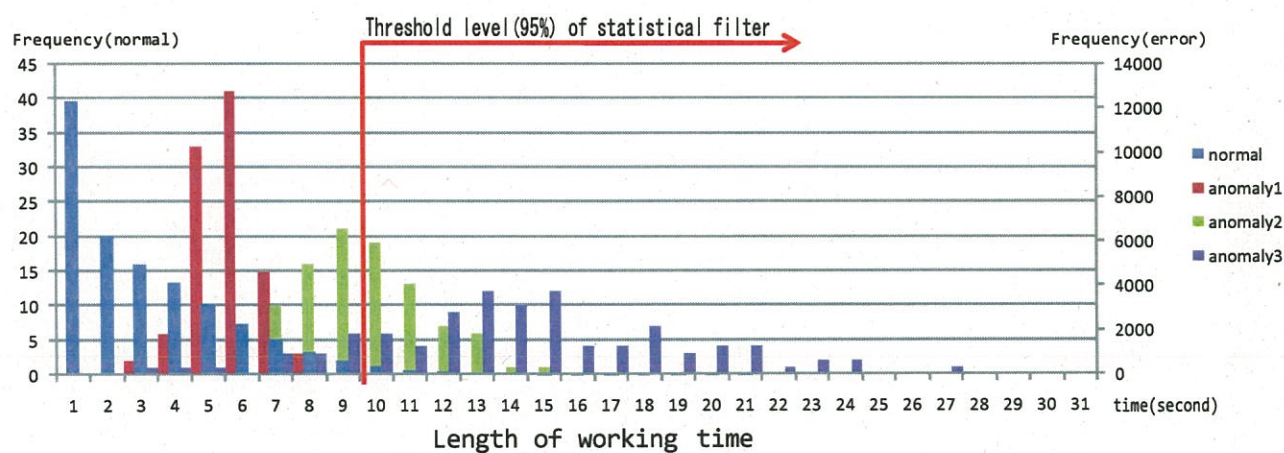


fig. 57: histogram of interval of anomary about length of working time and interval of working

## 6.6 検証結果

検証結果は次章で載せればよいそれか HMM で 95% 点の判定を行った結果を載せるかストーリー的には載せない方が分かりやすいかも。

動作間隔の小ささは完全にかぶってしまい、動作時間も正常時と同じ値ということで検知することは目視でも不可能である。異常が発生しているとは考えられない。

検証には正常時に挙がってくる誤情報で HMM を学習させ、その HMM を用いて異常状態のシミュレーションデータを診断させた。異常情報は 100 件の異常情報を生成し、どの程度検知できるかを検討した。動作間隔が短くなった場合には 5 件の異常情報を 1 つのまとまりとし、このまとまりのうち 1 回でも検知ができていれば異常診断されるため、検知ができているとする。

### 6.6.1 動作時間異常

動作時間に関しては 1 つの記号列に複数の異常情報が入ってくることはないため、100 件の異常情報に対して個々に正常・異常の診断を行う。診断を行った結果を図 58 から図 60、表 14 に示す。

図は横軸に時間 (秒)、縦軸には HMM のスコアを取っている。青いグラフが HMM のスコアを表し、赤い線が閾値となる HMM のスコアを表わしている。したがって赤い線よりスコアが高くなると異常と判断する。異常情報が入ってくるとスコアが高くなり、単発の場合、約 200 秒間ほとんど変わらない記号列に対してスコアを計算するため、スコアはほぼ一定となる。動作時間が長くなった場合や 200 秒間の記号列に複数の異常情報が入ってくるとスコアはさらに高くなる。

表には異常の程度を 3 段階に分けて検証を行った結果の正常・異常と判断したそれぞれの件数を示している。

動作時間の異常として異常 (高) を与えたときには 100 件すべてを異常と検知している。スコアが最高値 100 となっているのは計算により記号出現確率が 0 となり、その対数が取れないのでその場合のスコアを 100 としたためである。異常 (高) は動作時間平均値の 18 倍となっている。だいぶ高めの設定のようだが、実際に異常が発生した場合にはこれ以上の異常情報が挙がってくることは多々ある。

異常 (中) 及び異常 (低) に関してはそれぞれ 3 件、8 件の見落としがあった。しかしこの異常情報を見てもいずれも動作時間の短い単発の異常情報であり、仮に統計的フィルター 95% 点を設定していたとしても拾えないような情報であった。このように単発で動作頻度に変化がない場合の情報においては統計的なフィルターと同程度の検知能力である。

table 14: Anomaly about length of working time

|                   | Anomaly(Low) | Anomaly(Middle) | Anomaly(High) |
|-------------------|--------------|-----------------|---------------|
| normal            | 8            | 3               | 0             |
| anomaly           | 92           | 97              | 100           |
| rate of detection | 92%          | 97%             | 100%          |

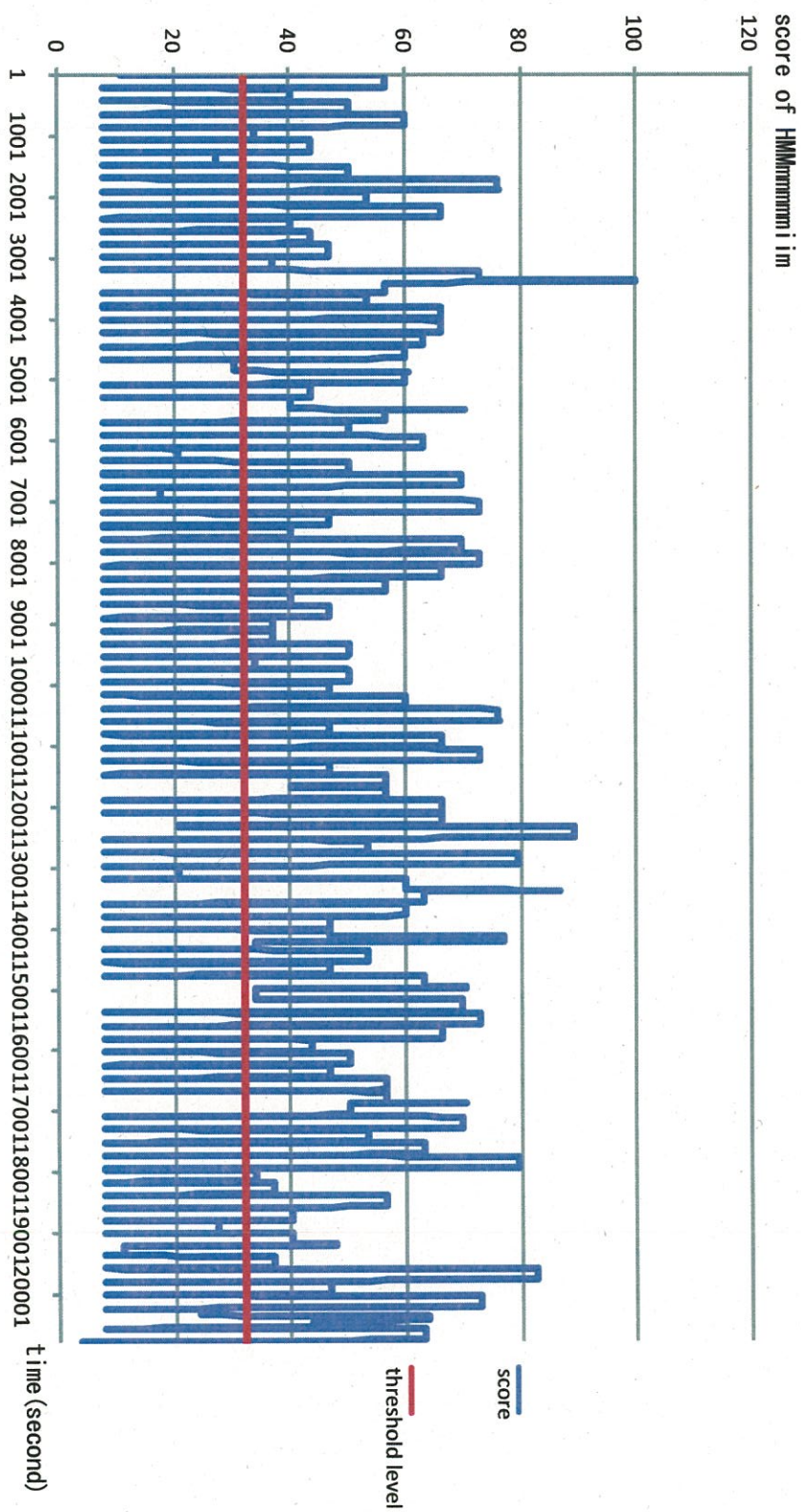


fig. 58: rking time and interval of working

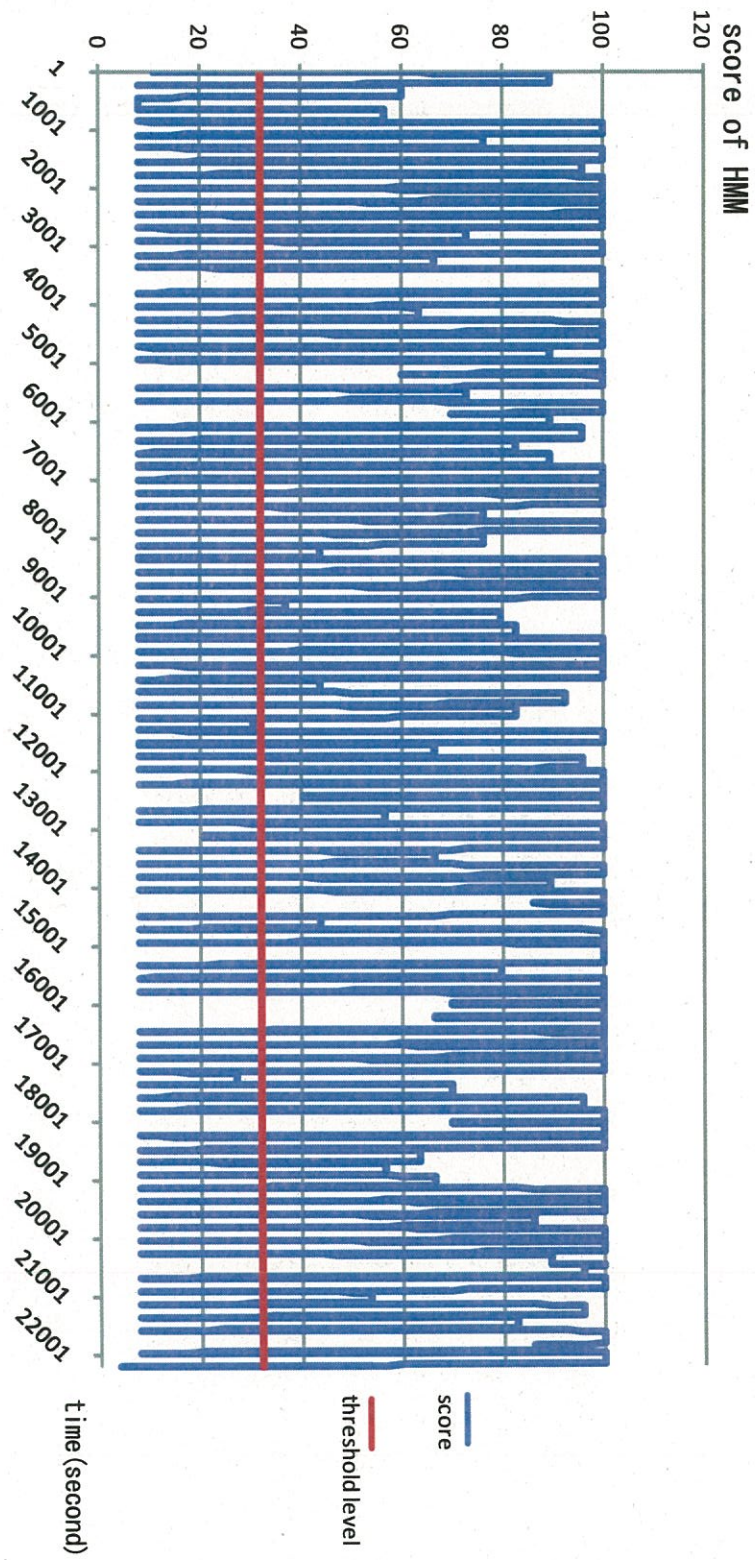


fig. 59: rking time and interval of working

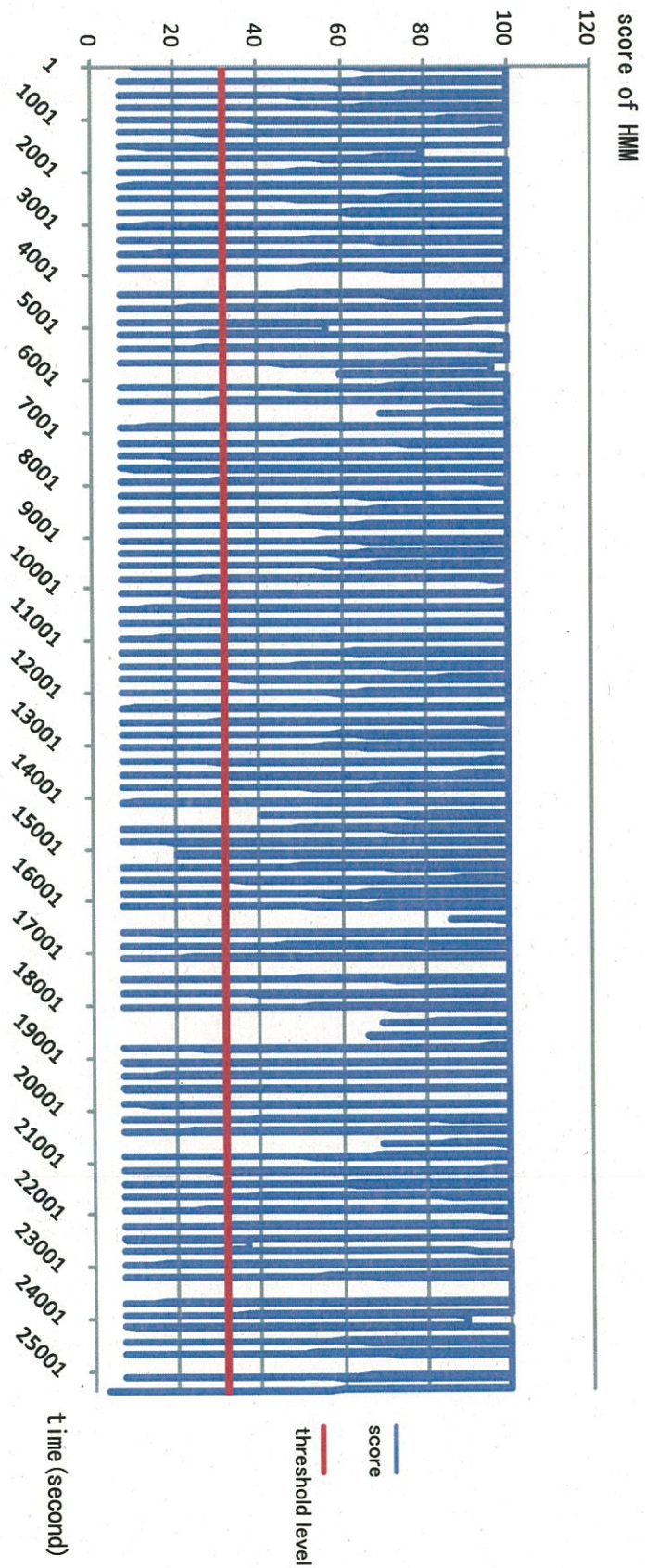


fig. 60: rking time and interval of working



### 6.6.2 動作間隔異常

動作間隔が挙がった場合の異常に関しては複数の異常情報をまとめて考える。この一連の情報のうち1回でも異常と判断できれば異常が発生していると診断する。今回の検証では5件の異常情報を一連の異常情報と設定し、5回ごとに頻度の高い異常情報が挙がってくる状態をシミュレートし検証を行っている。結果は図 61 から図 63、表図 15 に示す。図中の縦の補助線が一連の異常情報の区切り目となっている。

異常(低)では図 45 から、統計的フィルターの95%よりも低く動作時間も通常と変わらないため、この情報はほとんど正常状態と変わらず、見分けが困難な状態ではあるが、その情報にHMMのフィルターを適用すると検知率は55%となる。しかし100%の検知はできていない。

次に異常(中)に関してである。図??から判断して統計的フィルターでも十分判断できる範囲であるが、HMMを用いたフィルターでも異常の検知ができています。1件の見逃しは診断に用いる一連の情報にごく短時間しか動作しない異常情報しか入っていないため検知できなかったと思われる。実際の異常情報ではデータのばらつきはもっと大きくなるため、十分異常の検知は可能であると思われる。

異常(大)に関しては異常情報のまとまりに関しては全て異常と診断している。

table 15: Anomaly about interval of working

|                   | Anomaly(Low) | Anomaly(Middle) | Anomaly(High) |
|-------------------|--------------|-----------------|---------------|
| normal            | 9            | 1               | 0             |
| anomaly           | 11           | 19              | 20            |
| rate of detection | 55%          | 95%             | 100%          |

### 6.6.3 動作時間・間隔異常

最後に動作時間、動作間隔の両方に異常が発生した場合である。実際の異常も動作時間、動作間隔の両方が正常時とは異なった挙動を示すため、異常のモデルとしては実際の異常に最も近いと思われる。

異常状態としては前の2つの状態と同様に3段階の異常の大きさを検討しているがそれぞれの状態で、動作時間、動作間隔はより小さな異常状態となっている。

今回の異常状態も動作間隔について検討するため5件の異常情報を一連の異常情報として、このうち1件でも異常と検知できればよいこととする。

異常(低)に関しては図??、図??よりともに統計的フィルターの95%点以下であるため検知はできない領域である。この異常情報にHMMを用いたフィルターを適用すると異常の検知率は100%となる。統計的フィルターのように単発の異常情報で考えるのではなく複数の異常情報から異常の検知を行うことにより、より小さな異常から検知できると思われる。実際にはこの程度の異常は兆候状態であると考えられるため、その兆候状態を捉えることによりいち早く対応し、列車遅延を防ぐことができると思われる。

異常(中)、異常(大)に関しても異常の検知率は100%であり兆候状態から徐々に以上の状態が変わってきてもその変化の途中で異常を検知することができると思われる。

table 16: Anomaly about length of working time and interval of working

|                   | Anomaly(Low) | Anomaly(Middle) | Anomaly(High) |
|-------------------|--------------|-----------------|---------------|
| normal            | 0            | 0               | 0             |
| anomaly           | 20           | 20              | 20            |
| rate of detection | 100          | 100             | 100           |

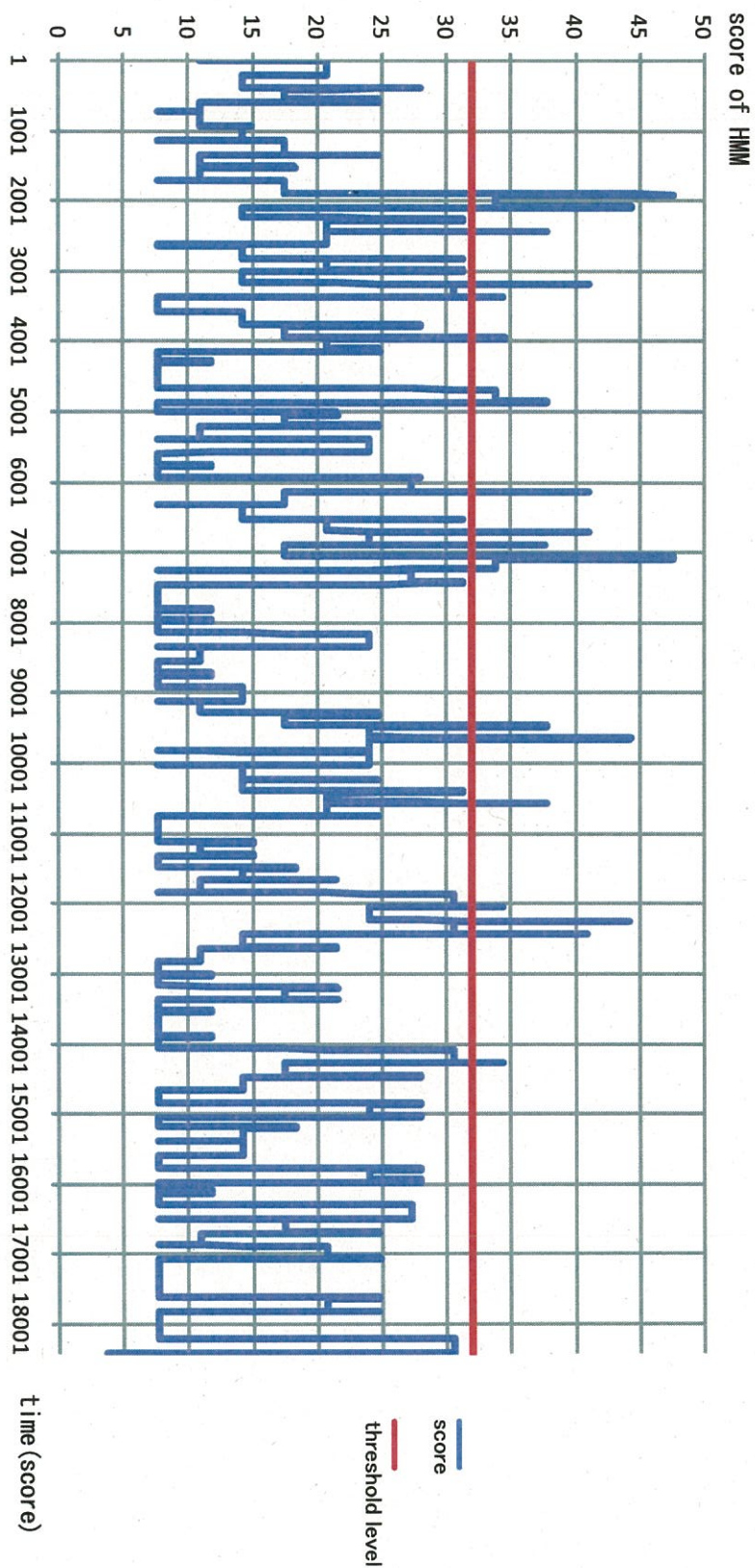


fig. 61: rking time and interval of working

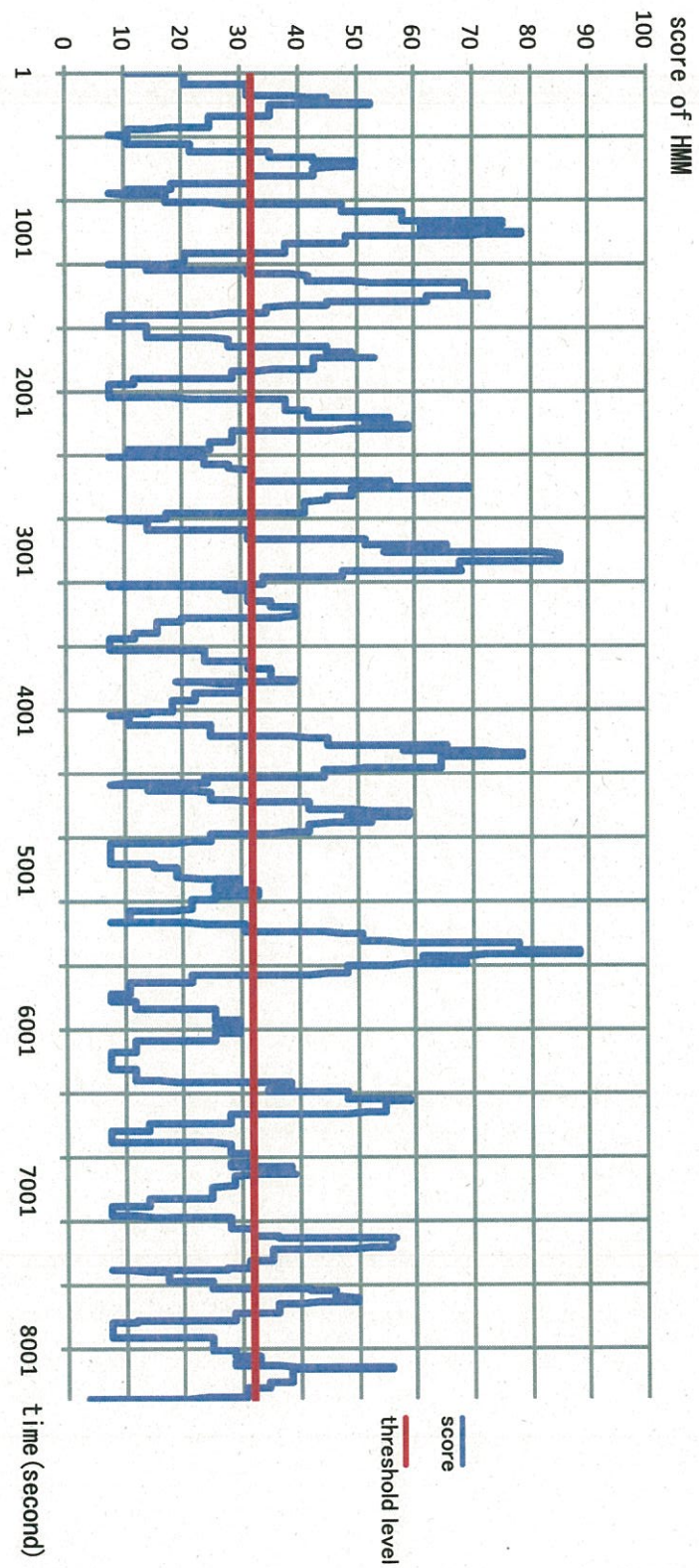


fig. 62: rking time and interval of working

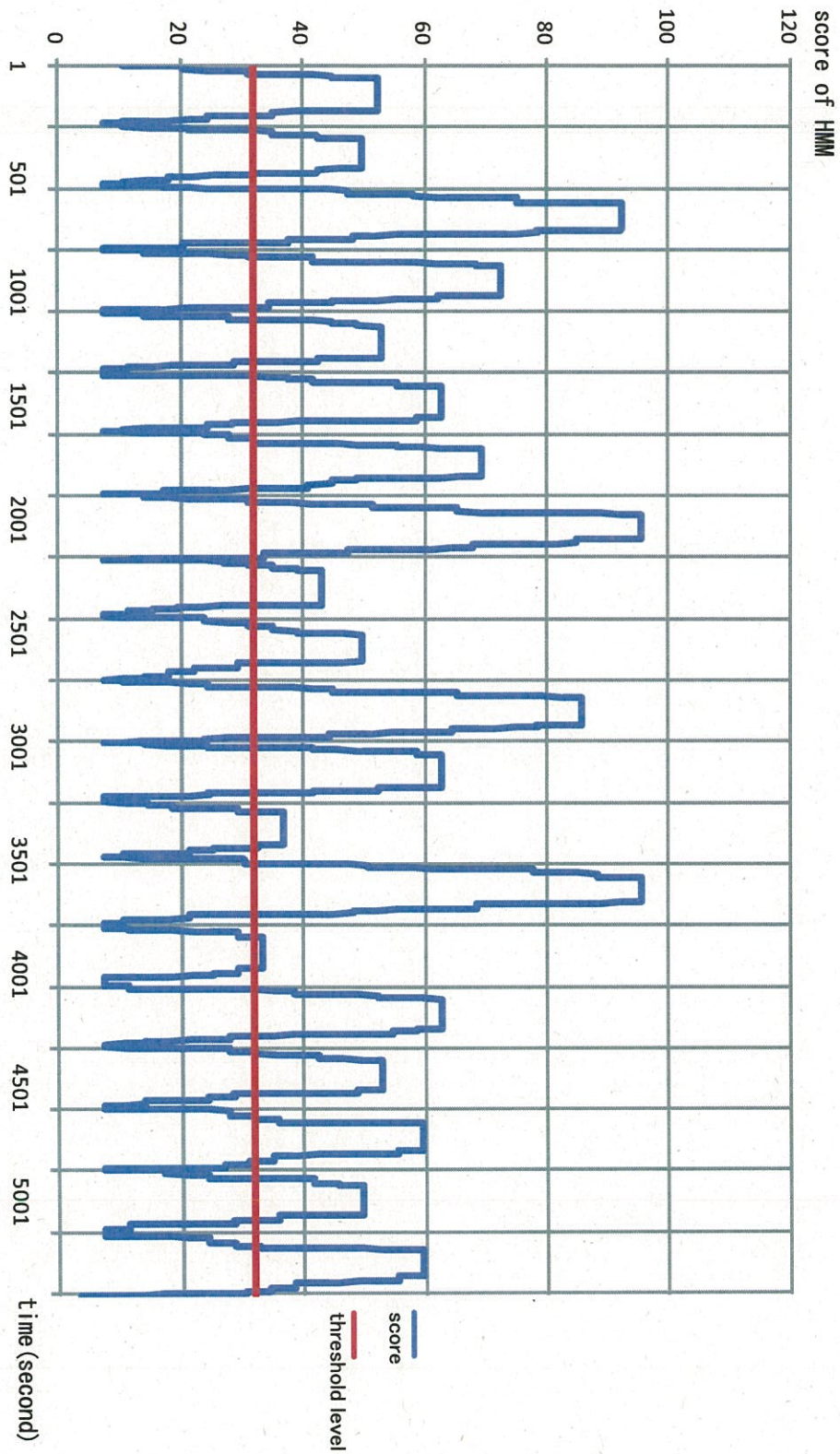


fig. 63: rking time and interval of working

## 6.7 考察

動作時間と動作間隔に異常が発生した場合にはともに小さな異常であっても高いスコアとなり、確実に異常を検知できることを示した。またここで設定した動作時間、間隔両方に異常を発生させた状態のうち、異常(小)は??で述べた統計的フィルターでは検知できない異常情報である。したがって統計的フィルターは設定する閾値を低くし、確実に誤情報である異常情報をスクリーニングし、その後に残った異常情報に対してHMMを用いたフィルターを適用することにより、効率的に、かつ確実に異常が検知でき、また兆候状態での異常の検知が可能であると考えられる。

実際の異常が、動作時間や動作間隔どちらかのみではなく両方に対して正常時とは異なった挙動を示し、小さな異常状態から大きな異常状態へと変化していくと考えるならば十分に早い段階での以上の検知が可能であると考えられる。

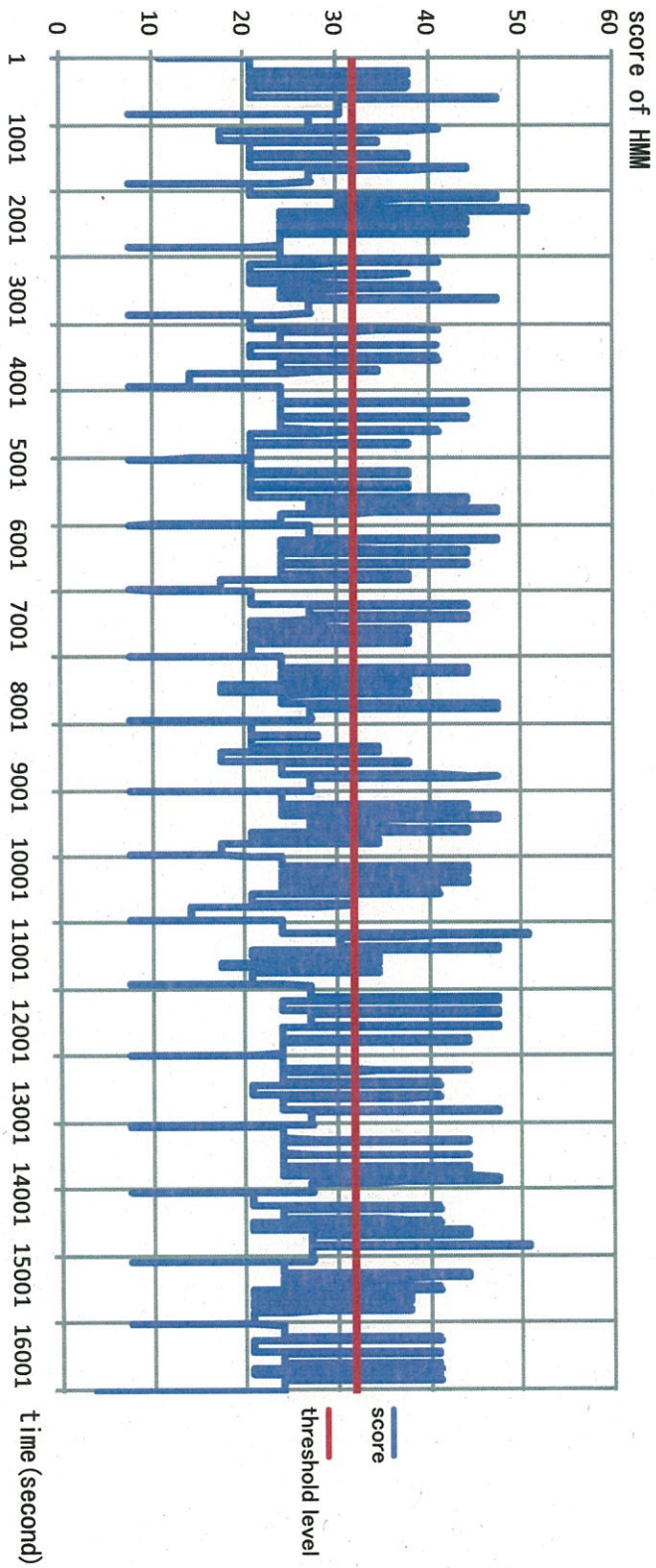


fig. 64: rking time and interval of working

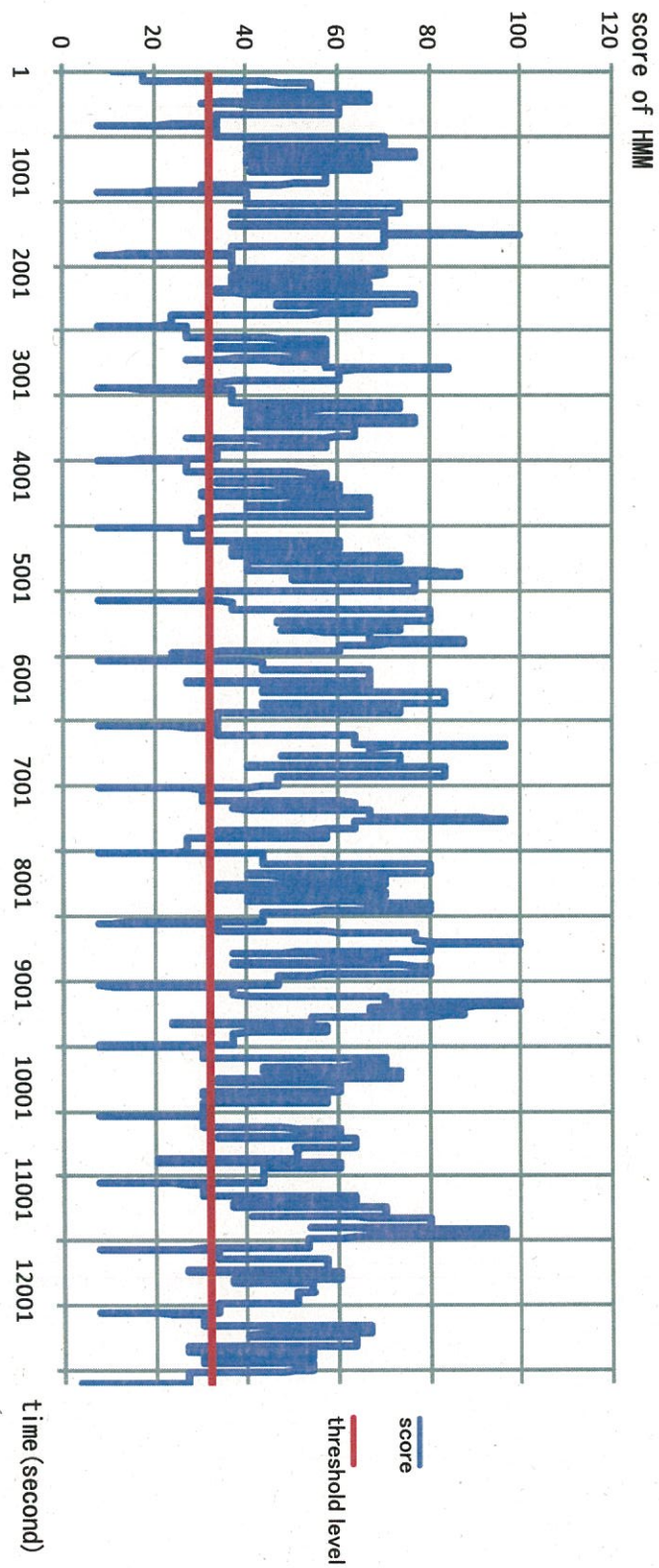


fig. 65: rking time and interval of working



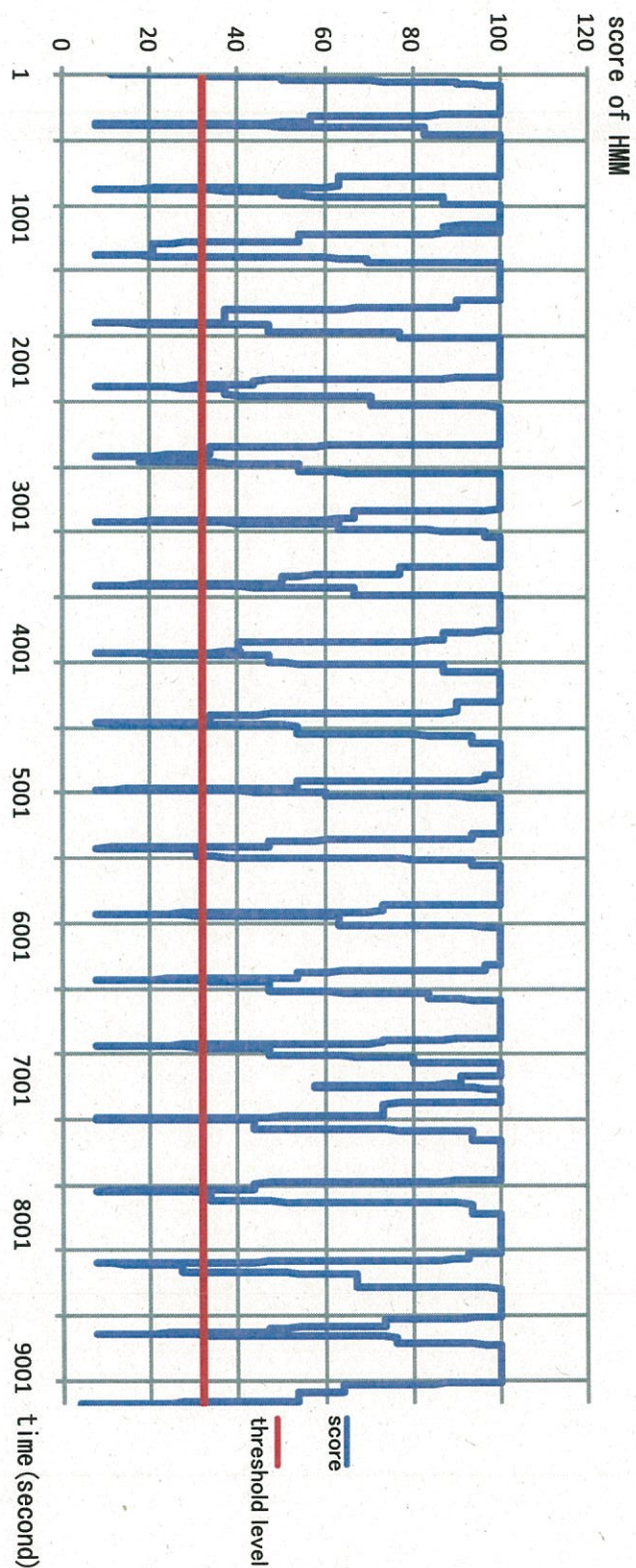


fig. 66: rking time and interval of working

## 6.8 結言

本章では HMM を用いた異常検知方法について述べた。本研究で HMM を採択した理由について、どこが適しているのかを述べた上で本研究への適用方法を説明した。続いてこの HMM を用いたフィルターがどの程度の検出能力があるのかをシミュレーションデータを用いて検証を行った。異常状態を複数設定し検証を行うことによって HMM を用いたフィルターの有効性について検証した。

## 7 検知フローに従った検証結果

### 7.1 緒言

本章では4章で検証した統計的フィルター、6章で検証したHMMを用いたフィルターに関して、それぞれ単独ではなく、図6に示すフローチャートに従って異常検知を行った場合の検証を行った。ここではHMMの閾値を一定にしたまま統計的フィルターの閾値を変えて検証を行っている。

## 7.2 検知の流れ

ここで再び図6で示した異常検知フローを示す。

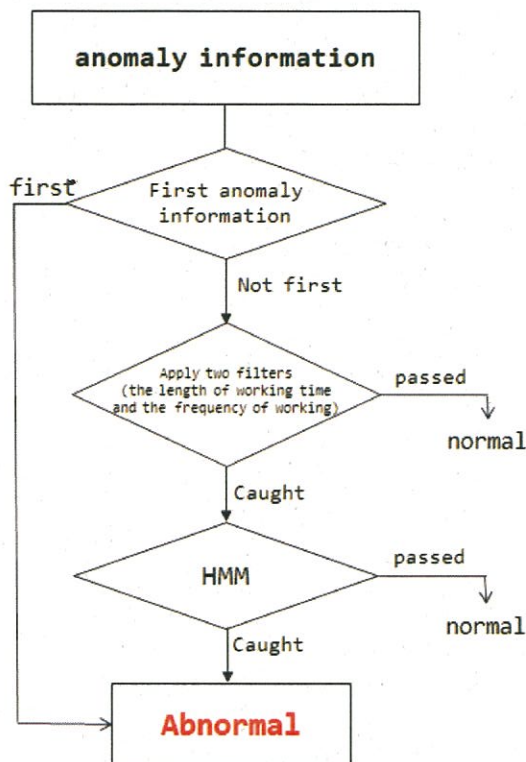


fig. 67: flowchart

今回検証の対象とするのは関東地域の踏切、2年分のデータである。

診断に際し、統計的フィルター及びHMMフィルターのパラメータを決定し、閾値を設定する必要があるため、異常検知フローにかける前に各フィルターの設定を行った。統計的フィルターの設定は4章で示した方法で設定した。今回は各踏切で100件の誤情報を学習用データとしてパラメータ推定を行い分布を推定、閾値の設定を行った。また100件に満たない踏切については全件を用いて推定を行っている。検証用のデータとしては2年間のデータ全件を用いた。すなわち学習データと診断データが重複する踏切も存在する。HMMを用いたフィルターのパラメータの決定は、同様に2年間の誤情報のうち数十件から数百件を用いて行った。ここでHMMを用いたフィルターは統計的フィルターで正常ではないと診断された情報に対して診断を行うが、学習に用いるデータは統計的フィルターで診断を行う前の誤情報を用いている。

HMMの閾値は95%で一定としたが統計的フィルターの閾値は90%と50%の2通りで検証を行った。50%は6章で示したようにHMMで検知可能なごく小さな異常

が発生していると仮定した場合、それらを統計的フィルターで安全側に排除してしまわないための閾値である。それに対し90%は実際の異常の発生状況から考えた効率的な閾値である。

## 7.3 検証結果

### 7.3.1 統計的フィルター閾値 50%

統計的フィルターの閾値を 50% に設定した場合の結果を表 17、図 68～図 70 に示す。表には各フィルターを適用した時点での、正常ではないと判断された件数を示している。表には関東圏の 7 つの管轄地域のうちサンプルとして 4 か所と全 7 か所の合計を示している。数字は、上から、挙がってきた異常全件数、第一段階のフィルター、統計的フィルター、HMM フィルターをそれぞれ適用した後の件数である。またこの異常の検知状況を図 6 のフローに沿って示した図が図 68 になる。

図 68 ではまず最初に、「初めての異常情報」かどうかのフィルターに掛ける。このフィルターは各踏切の異常情報の種類ごとに 1 回しか挙がらないため 0.5% となっている。残りの大半の異常情報は統計的フィルターに掛けられる。

統計的フィルターは動作時間、動作間隔のそれぞれについてフィルターを、どちらか一方でも正常ではないと判断された場合にはその異常情報は正常でないとされる。つまり 2 つのフィルターで OR の判断をしている。そのためそれぞれのフィルターで 50% の閾値を設定しているが検知される異常情報は合計では 50% より多くなる。今回は合計で、全体の 73% ほどが正常ではないとして検知された。

図 69 は異常検知フローによる誤検出の低下を示すグラフである。横軸に各フィルターを取り縦軸に検出された異常情報数を示す。この図より、各フィルターを適用していくことにより誤検出が減っていることが分かる。図 70 は図 69 の件数をパーセント (%) で表わしたグラフである。このグラフからどの管轄も同様に誤検出が減っていることが分かる。

この異常検知フローにより全体では 94% を正常と判断し 6% を異常と判断しており、大幅な誤検知の削減が見込める。

### 7.3.2 統計的フィルター閾値 90%

統計的フィルターの閾値を 90% に設定した場合の結果を表 18、図 73 に示す。またこの異常の検知状況を図 6 のフローに沿って示した図が図 71 になる。

これらのグラフの読み方については前述の通りである。

最終的な HMM を用いたフィルターには 57000 件ほどの異常情報が入り、その 20% ほどを異常として検知している。

この異常検知フローにより全体では、95% を正常と判断し、4.6% を異常と判断しており、大幅な誤検知の削減が見込める。これは図 71、図 73 から見て取れる。

図 74 は両者の誤検知率を比較した図である。このグラフから分かるように統計的フィルターで検出する情報数には差があるものの、どちらも最終的にはほぼ同じ値となる。これは HMM の学習データが統計的フィルターを掛ける前のデータを用いて学習し、その分布の 95% 点を閾値としているため、最終的には全体の 5% ほどまで

は削減されるためであると考えられる。ただし6章で述べたとおり、統計的フィルターの低い位置にもHMMで検出できる異常情報はあるためそのような情報の分は増えると考えられる。

table 17: result of applying detection flow(threshold level = 50%)

|                 | Area 1 | Area 2 | Area 4 | ... | Area 7 | All Area      |
|-----------------|--------|--------|--------|-----|--------|---------------|
| Total           | 145726 | 13668  | 52571  | ... | 3688   | 215653(100%)  |
| filter 1st      | 145368 | 13563  | 52085  | ... | 3635   | 214651(99.5%) |
| statical filter | 109765 | 7951   | 37084  | ... | 1991   | 156791(73.5%) |
| HMM filter      | 5063   | 1537   | 5921   | ... | 361    | 12882(5.97%)  |

table 18: result of applying detection flow(threshold level = 90%)

|                 | Area 1 | Area 2 | Area 3 | ... | Area 7 | All Area      |
|-----------------|--------|--------|--------|-----|--------|---------------|
| Total           | 145726 | 13668  | 60744  | ... | 3688   | 291499(100%)  |
| filter 1st      | 145368 | 13563  | 60322  | ... | 3635   | 289965(99.5%) |
| statical filter | 26200  | 2239   | 14822  | ... | 734    | 56934(19.5%)  |
| HMM filter      | 4789   | 732    | 3165   | ... | 463    | 12047(4.13%)  |

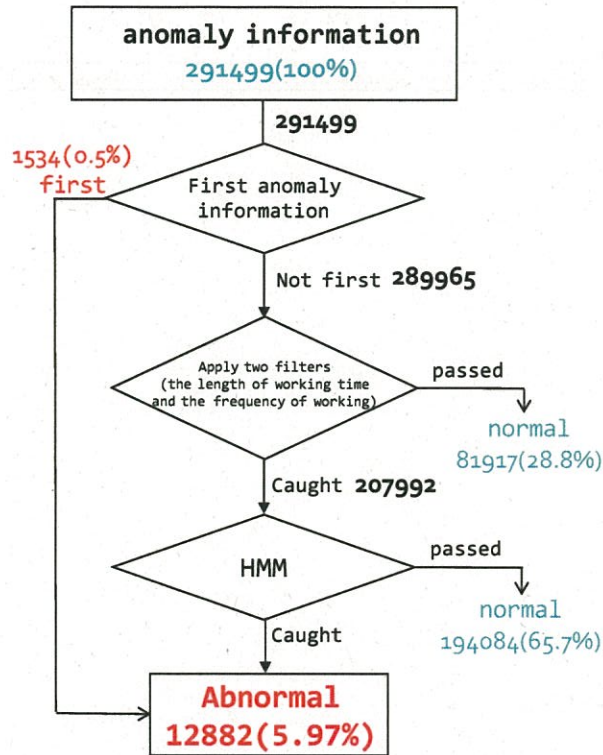


fig. 68: Flow of detection system(50%)

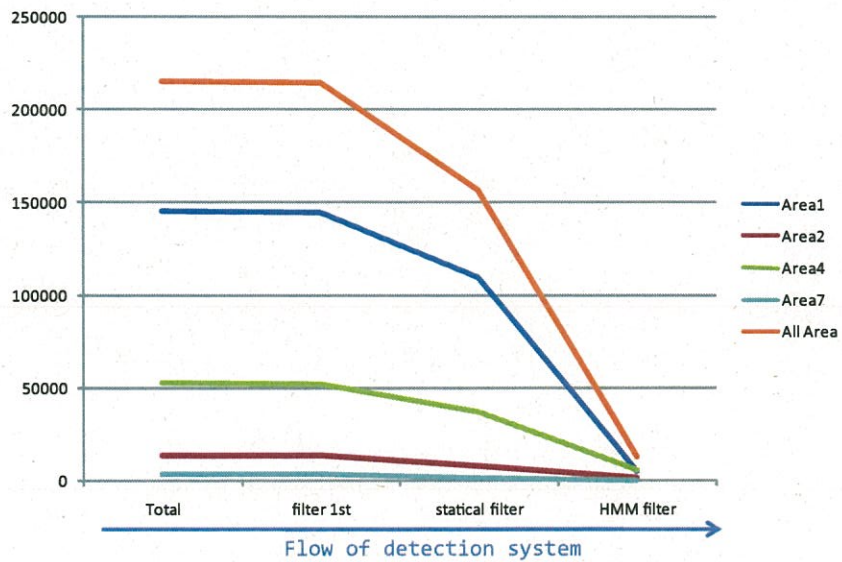


fig. 69: number of detection



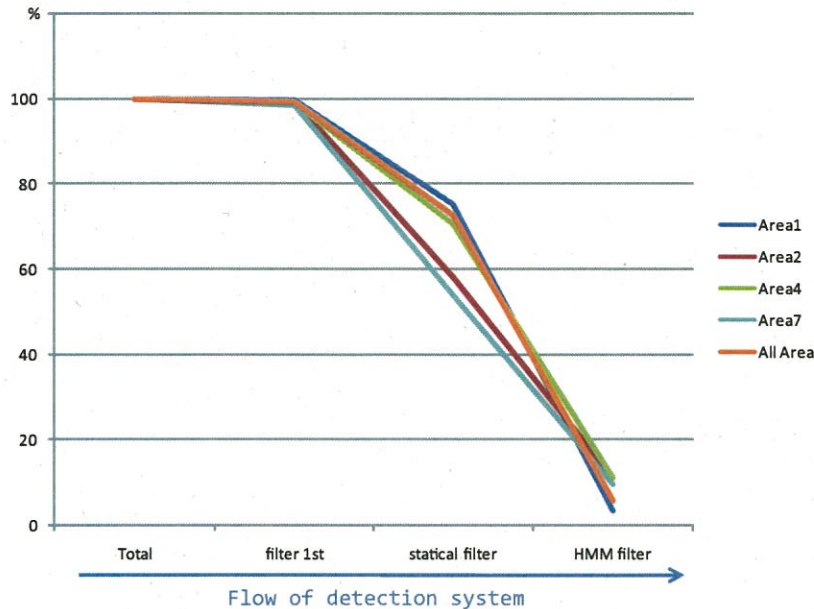


fig. 70: rate of detection

### 7.3.3 異常の検知

次に実際に異常が発生していた場合に異常を確実に検知できているかどうかの検証を行う。実際に異常が発生した踏切の異常情報を HMM のフィルターにかけ、確実に検知できているか確認した。今回は実際の事故事例から、兆候があったと考えられる機器故障などの異常情報 5 件に対して検証を行った。

その結果を表??、図 75、図 76 に示す。HMM スコアの図は 2カ所のみ掲載する。閾値を 50%に設定した場合も 90%に設定した場合も異常検知に関しての差は見られなかった。これは異常の兆候と考えられる異常情報に関しては、挙がった兆候状態が多きく、すべて 90%点を超えるような異常情報だったため、フィルターにかかった異常情報が同じであり、HMM で計算する記号列がどちらも同じであったためと考えられる。

そこで真の異常情報に関する結果については統一してまとめる。

表は各踏切で挙がってくる異常情報から異常が検知できていたかどうか、またその兆候状態で検知できていたかどうかについて示す。表には各踏切における障検動作関連の 2 種類の異常情報が示してあり、このどちらの情報で検知が可能かを示した。「Type1」は「障検動作」関連、「Type2」は「障故 R」関連の異常情報であることを示す。また表中の「○」は異常、兆候が検知できていたことを、「-」はその時に情報が挙がっていなかったことを示す。表より、異常情報の種類は異なるが、すべての踏切で異常の検知が可能であり、また何らかの兆候状態で異常を検知できていたことが分かる。

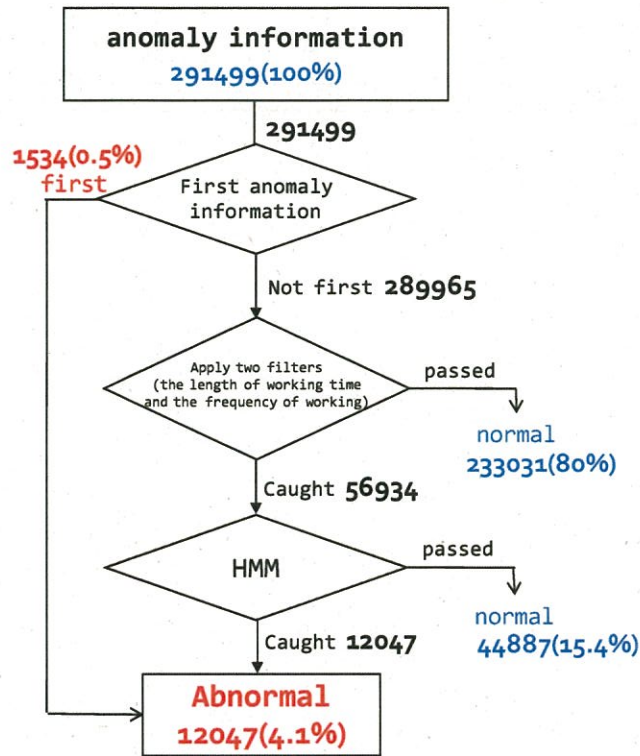


fig. 71: Flow of detection system

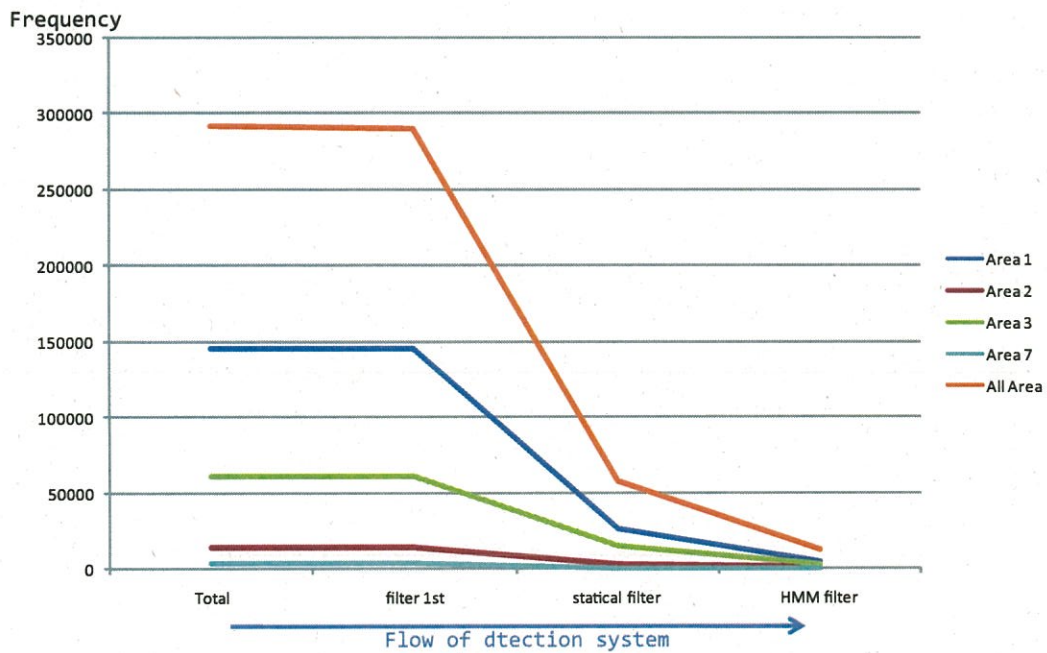


fig. 72: number of detection

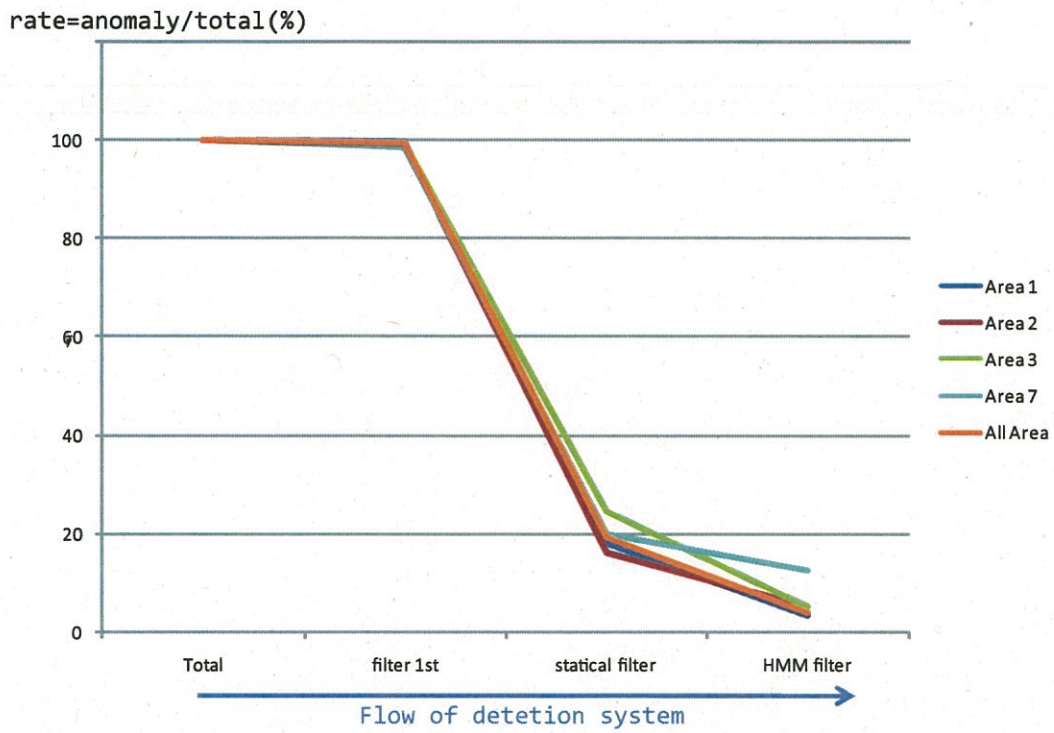


fig. 73: rate of detection

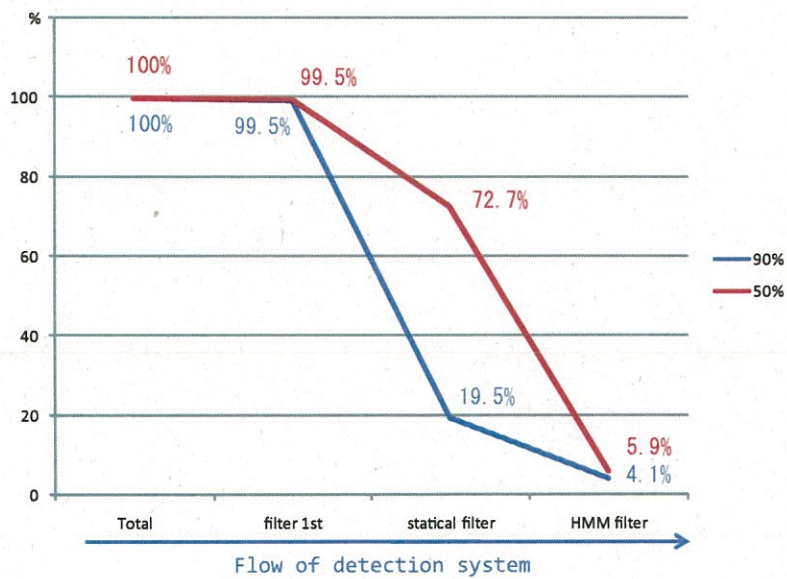


fig. 74: Flow of detection system

図はHMMの時系列データを示す。横軸が時間(秒)、縦軸がHMMのスコアを表す。図中の赤矢印が異常を発見し、対処またはその指示が出された時間である。いずれも異常を発見した時刻以前に異常を検知できていることが分かる。特に踏切Bに関しては前日から予兆があり、HMMを用いたフィルターではその異常を検知している。

table 19: result of flow of detecting system about anomal Crossing

|            | Crossing A |       | Crossing B |       | Crossing C |       | Crossing D |       | Crossing E |       |
|------------|------------|-------|------------|-------|------------|-------|------------|-------|------------|-------|
|            | Type1      | Type2 | Type1      | Type2 | Type1      | Type2 | Type1      | Type2 | Type1      | Type2 |
| detection  | -          | ○     | ○          | ○     | ○          | ○     | ○          | ○     | ○          | ○     |
| indication | -          | ○     | -          | ○     | ○          | ○     | ○          | ○     | -          | ○     |

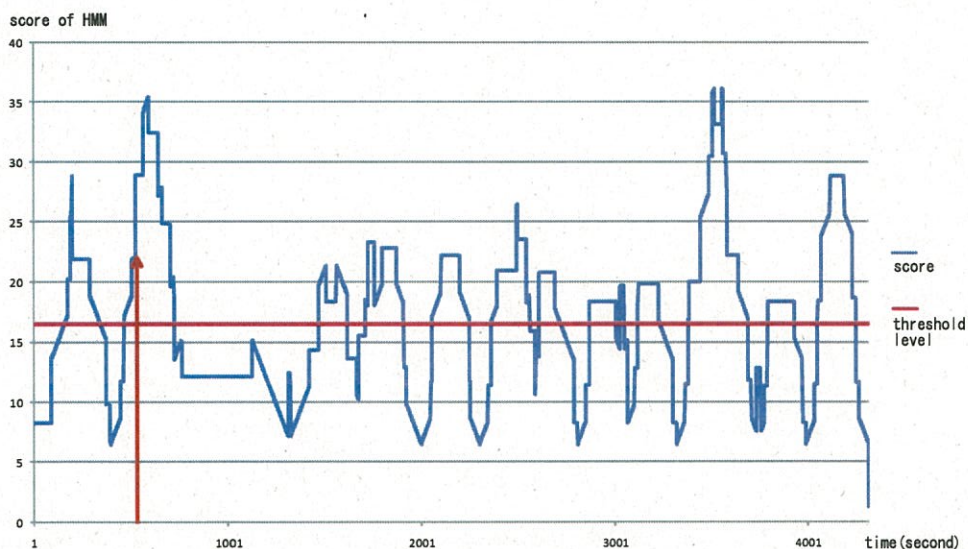


fig. 75: score of HMM at Crossing A

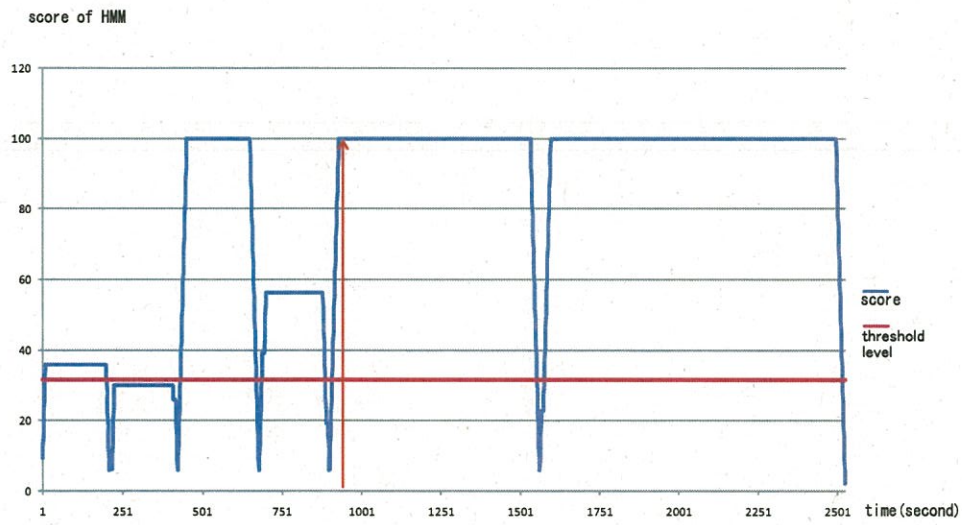


fig. 76: score of HMM at Crossing B

#### 7.4 考察

統計的フィルターを50%、90%のいずれに設定しても大量の誤情報を減らし、実際に異常が発生した場合には確実に異常が検知できることを示した。ここで統計的フィルターの閾値について考察する。

閾値を50%にした場合、統計的フィルターで検知される異常情報は増えてしまうが、その分兆候状態として小さな変動も残したままにすることができる。それにより何らかの異常が発生した場合には早くその異常を検知できる可能性がある。しかし実際の異常情報から調査した結果、2年間の真の異常情報の中に兆候状態と考えられる情報でそのように微小な変化から始まる異常情報はなく、統計的フィルターを90%に設定しても検知できるような異常情報ばかりであった。

しかし、今後そのような微小変化から始まる異常の可能性を考えると、閾値を低く設定する必要もある。

本研究で適用した統計的フィルターの閾値は90%、50%であった。これによりどちらも約1万2千件ほどまで誤検知を減らすことが可能となった。異常情報が24ヶ月で1万2千件ほどというのはひと月当たり500件、1日当たり17件程度である。列車の運行時間を考えると1時間に1件程度の異常情報となり、無理なく監視することができるレベルであり、本研究の目的を達成しうる件数である。

## 7.5 結言

本章では異常検知フローに従いフィルターを掛けた場合にどのように正常・異常の判断がおこなわれるかを具体的な検証を通して示し、大幅な誤検知の低下、異常情報の確実な検知を示し、その有効性について述べた。また統計的なフィルターの閾値を変化させることによる誤検知率の変化について考察し、本研究で提案する異常検知フローにより実運用可能な異常情報数まで誤情報を減らせることを示した。

## 8 結論

本論文では鉄道定常監視システムに挙がってくる誤情報を多く含んだ異常情報から真の異常情報を検出する手法に関する提案を行った。まず挙がってくる異常情報から実際に対処が必要な異常情報と、対処のいらぬ異常情報という観点から正常・異常の定義をした。

次に本来、このシステムで検出したい異常情報として実際に異常が発生して列車遅延が発生する前の兆候状態にあたる異常情報の有無という観点から先の定義における異常を分類した。それにより異常時の異常情報の挙がり方が大きく分けて3パターンあることが分かった。そしてこれらの定義、分類に基づいて異常の検知フローを考えた。異常検知フローは図6の通りである。この異常検知フローでは2段階のフィルターとHMMという手法を用いて「正常である」「正常ではない」という判断を行う。実際に「異常がある」とはフィルターの性質上言い切れないため「正常時とは異なっている」という判断である。

第1段階のフィルターでは「以前に異常情報(誤情報)があったかどうか」のフィルターであり、第2段階のフィルターは「動作時間」「動作頻度」に関する2種類の統計的フィルターである。「動作時間」「動作頻度」のフィルターは誤情報のときの「動作時間」「動作頻度」が、実際に異常が発生しているときに上がる異常情報の「動作時間」「動作頻度」と異なるという点を用いている。今まで経験的に判断していた「動作時間」「動作頻度」の異常をこのフィルターによって判断することが可能である。このフィルターによりスクリーニングを行い、大幅に異常情報を削減し、次にかけるのがHMMを用いたフィルターである。このHMMを用いたフィルターを適用することにより統計的フィルターだけでは検知できない異常情報も検知し、高精度に異常を検知ができることを示した。HMMでは正常時における誤情報の発生パターンと診断時における異常情報の発生パターンという複数の異常情報を元に比較することによって診断時の異常情報が正常であるのかどうかを判断し、異常の検知の検知が可能であることを示した。

以上の分析、検証により図6の異常検知フローを用いることにより、誤情報を大幅に減らし、また真の異常情報に関しては確実に検知が可能であることを示した。この異常検知フローによる誤情報の削減により、司令員の方が確認するなど異常情報に実際に対応することができるようになる。



## 9 謝辞

本研究を行うにあたり、機器情報や異常情報を提供して下さった(株)JR東日本の山崎さん、杉本さん、石渡さん、ありがとうございました。特に杉本さんには3年間お世話になり、研究の分野のみならず、飲み会の席での最年少者の働き方まで示していただきました。

研究を行う上で丁寧に指導して下さった群馬大学の岩崎先生には大変お世話になりました。毎回、ギリギリでしか動かないため多々迷惑をおかけし、また色々と至らない点があったにもかかわらず、3年間丁寧に指導していただきました。本論文は岩崎先生のお力添えの下に成り立っています。

隠れマルコフモデルや研究に関する適切なアドバイスをして下さった情報理工系中村研究室の小野さん、九乗さん、倉貫さん、ありがとうございました。

お忙しい中、研究分野から極私的なことまで、公私に渡ってお世話をいただいた上海蟹の皆さま、本当にありがとうございました。

研究の大詰めになる1, 2月、寒い中にもかかわらず風邪をひかないようにと全力でその役目を果たしていただいた研究室の加湿器「FE-14KFW」には本当に助けていただきました。妥協することなく常に全力で働き続けるその姿勢には、研究で挫折しそうになるたびに励まされてきました。これから社会人となる身としてはそのどんな困難な状況にも全力で挑んでいく姿はとても参考になりました。今後も挫折しそうになったときには加湿機が全力で加湿している姿を思い出し、励みとしたいと思います。末筆にはなりませんが、加湿機への感謝を述べて本論文の締め言葉とさせていただきます。

本当にありがとうございました。

## 参考文献

- [1] L.R.Rabiner, Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proc. IEEE, Vol77 No2, pp.257-284, 1989
- [2] A.Krogh "An introduction to Hidden Markov models for biological sequences," In S.L.Salzberg, D.B.Searls and S.Kasif, editors, Computational Methods in Molecular Biology, chapter 4, pp.45-63. Elsevier, Amsterdam, 1988
- [3] A.Krogh, M.Brown, I.S.Mian, K.Sjolander and D.Haussler "Hidden Markov-models in computational biology: Applications to protein modering," textitJournal of Molecular Biology, 235:1501-1531, Feb, 1994
- [4] 前村一哉, 小野廣隆, 定兼邦彦, 山下雅史, "隠れマルコフモデルによる DNA 配列設計," 電子情報通信学会論文集, pp.39-46, 2006
- [5] 森武俊, "生活パターンを覚えて助ける知能住宅," 電子情報通信学会論文集, pp.1-5, 2005
- [6] 沼田成弘, "隠れマルコフモデルを用いた高齢者の異常検知システムの構築", 電機情報通信学会論文集, pp.49-52, 2003
- [7] 中川, "確率モデルによる音声認識", 電子情報通信学会, 1998
- [8] 野村健, 柳哲, 後藤富朗, 徳田恵一, 北村正, "情報量基準に基づいた HMM ジェスチャー認識の状態数の検討", 電子情報通信学会論文集, pp.283
- [9] 大川和宏, 中村仁彦, "統計的モデルを用いた運動認識用階層構造の自動構築"