

モーションセンサによる異常検知と ベイジアンネットワークによる誤センシングデータの除去 A Bayesian Network based Method for Reducing Noise-data for Anomaly Detection in Motion sensors

大塚 孝信[†]
Takanobu Otsuka

伊藤 孝行[‡]
Takayuki Ito

1. まえがき

近年、携帯情報端末の普及により、標準的なワイヤレスセンサネットワークに用いられる標準的な通信モジュール、センサ類が低価格で入手できるようになった。しかし、まだまだ一般ユーザーにとって容易に用いられるものではない。我々は、日本において深刻な問題となっている高齢者の増加に対応可能な、高齢者のみまもりを目的とした設置が容易な WSNs を試作する。そして、モーションセンサを用いたセンサネットワークの実センシングデータを用いて SVM による異常検知を行った。異常検知においては、何をノイズデータと定義して、何を異常と定義するかが難しい。本研究における異常データは侵入や深夜徘徊などの通常の生活とは異なる時間帯でのセンシングデータを異常と定義した。しかし、設置場所の外的要因によるセンシングデータが多く観測され、異常検知精度を低下させていることが判明した。ノイズデータの多くは白熱灯や制御機器によるノイズにより発生していることが多く、設置環境を変更するには多大なコストが必要となる。そのため、モーションセンサや設置環境に依存しない方法でノイズ除去を行った。我々は、ベイジアンネットワークを用いた提案手法により実際には実現不可能な移動パターンをノイズデータと定義した。具体的には入り口を通過せずに発生した室内の移動パターンである。それらノイズデータをデータインスタンスより除去するとともに、SVM のパラメータを変更することで異常検知精度の向上を確認した。

2. はじめに

近年、スマートフォンに代表される情報端末が広く普及している。そのため、従来は大型だった通信モジュールや各種センサ類の小型化が進むとともに、一般ユーザーが気軽に購入できる価格で販売されている。また、通信モジュールとセンサを組み合わせた無線通信によるセンサネットワークの研究が多く行われている。代表的な例として、橋梁や高速道路などの建造物における健全性をモニタリングする構造センシングや温度や湿度を大規模に計測して可視化を行う環境センシングなどが挙げられる。更に、提供されているサービスとして潮流の状況や海面温度などを世界中に設置されたセンサにより測定することで情報を提供するサービスや個人が製作したセンサネットワークにより取得したセンシングデータを共有し、閲覧できるサービスなども存在する。しかし、センサネットワークの設置や運

用には専門的な知識が必要であり、一般ユーザーにとってはまだ敷居が高い。

また、近年日本では介護を受けながら一人暮らしをしている老人が爆発的に増えており、日本の総人口はほぼ横ばいになっているのに対し、65 歳以上の高齢者は 2010 年には総人口に対し 23.1% と増加の一途をたっている。それに伴い社会保障に必要な金額も年々増加しており、社会保障費用全体に占める高齢者関係給付金の割合は 69.5% となっている。そのため、高齢化社会が今後ますます進む現在において、「在宅介護」の負担を減らすことが求められている。我々は、身近な問題である高齢化社会を題材として一般ユーザーにも設置が簡単に行え、かつ運用も容易に行うことが出来る高齢者みまもり用途のワイヤレスセンサネットワークを開発する。そして、開発したワイヤレスセンサネットワークを研究室内で稼働させ、データを取得することで介護ドメインと研究室ドメインにおいて共通した「日常とは違う時間帯での活動という異常」を検出するため、SMOTE アルゴリズムを用いた SVM による評価実験を行う。SMOTE アルゴリズムとは、不均衡データへの対処を目的としており、負例データをオーバーサンプリングすることにより不均衡を解決する手法である [17]。異常検知においては、何をノイズデータと定義して、何を異常と定義するかが難しい。本研究における異常データは通常の生活とは異なる時間帯でのセンシングデータを異常と定義した。また、簡単に設置できるセンサユニットは専門的な知識を持たない一般ユーザーがセンサユニットを設置することが想定されるため、研究室内での実験においても外的要因を特に気にせず設置を行った。その際、人の動きに起因しない外部環境の変化によるセンサの誤検知が発生し、それらノイズが異常検出における予測精度を低下させていることがわかった。そこで、本研究ではベイジアンネットワークを使ったノイズ除去手法を提案する。我々は、ベイジアンネットワークを用いた提案手法により実際には実現不可能な移動パターンをノイズデータと定義した。具体的には入り口を通過せずに発生した室内の移動パターンである。それらノイズデータをデータインスタンスより除去することで、異常検知精度が向上していることを確認した。本稿の構成を以下に示す。まず、2 章で本研究と関連する先行研究を紹介し、本研究の位置づけを示す。そして、3 章で試作したセンサネットワークと異常検知手法について述べる。その後、4 章において提案するノイズ除去手法について説明する。5 章では、評価実験の結果と得られた知見について示す。そして最後に、6 章で本稿のまとめと今

[†]名古屋工業大学大学院 情報工学専攻

[‡]名古屋工業大学大学院 産業戦略工学専攻

後の課題を示す。

3. 関連研究

3.1. 関連サービス

防犯分野で異常を検知するシステムは多く提供されている。代表的なのがセコム株式会社の提供するセコム・ホームセキュリティシステム [7] や ALSOK 総合警備保障の提供するホームセキュリティシステム [8] である。以上のシステムは主に侵入経路と生活空間に設置された人感センサ、窓や玄関などに設置された開閉センサ、警備会社への通報ユニットで構成されている。高齢者見守りの分野でも多くのサービスや研究があり、高齢者宅に設置された電気ポットを用いて電気ポットがいつ使われたかといった情報を親族に通知するサービス [1]、人感センサを用いて高齢者の行動パターンを分析し、異常時に通報を送るシステム [2][3]、RFID や荷重センサを家庭の各箇所に設置して行動パターンを分析し、異常時に通報を送るシステム [4]、ウェアラブルセンサによる体温情報監視システム [5] また、カメラを用いた動画による異常検知システム [6]、などがある。

以上のサービス・研究の問題点として専用品を用いているため一般的にコストが高く・異常通報するシステムでは誤報が多いとシステムの信頼性が保てなくなる。また、常に動画を撮影するとプライバシーの問題があるといった点がある。

3.2. 異常検出に関する研究

近年、センサを用いてセンシングしたデータやクレジットカードの使用履歴など様々なデータを解析することにより異常検出を行う Anomaly Detection[10] と呼ばれる研究が多くの研究分野とアプリケーションドメイン内で研究されている。

Anomaly Detection の研究分野の主な技法は大半が、分類技法をもとにした技術のカテゴリ、最近傍法をもとにした技術のカテゴリ、及びクラスタリング技法をもとにした技術のカテゴリ、統計技法をもとにした技術のカテゴリに分類することができる。また、いくつかの技術は情報理論をもとにした技術のカテゴリ、スペクトル理論をもとにした技術のカテゴリに属している。以上のように様々な技法が存在しており、実際に運用されているシステムとしては、コンピュータ・ネットワーク上のハッキングされたコンピュータが不正に機密データを送信していることを意味する異常なトラフィックパターンの検出 [11]、MRI 画像中の悪性腫瘍の存在を示す異常の検出 [12]、宇宙機の幾つかの部品の障害を意味する宇宙機のセンサからの異常な測定値の検出 [13]、などが研究されている。ワイヤレスセンサネットワークからのストリーミングデータの信頼性がユーザの観点から重要なことから、センサ間の条件付き依存関係を考慮したベイジアンネットワークによるセンサの異常値を検出する研究 [14]、ワイヤレスセンサネットワークを使って位置検出を行うシステムにおいて、敵によって引き起こされる位置情報の異常を検出するため統計モデルを使う研究 [15]、気象データを収集する分散型ワイヤレスセンサネットワーク内の異なるノード

ドから集めたデータを融合し、スペクトル理論を使った異常検出を行う研究 [16] などがある。

以上の研究の特徴として異常状態と正常状態の定義がしやすいことが挙げられる。一方、我々は独居高齢者における異常は通常の生活パターンの中での異常を検知することを目標としている。そのため、通常的生活の中における境界が決定しづらい状況の中で異常と正常の境界を設定する必要がある。

本研究は、単純な人感センサのデータによる異常検知を行うため、教師あり学習を行うことが出来る SVM を用いることとした。更に、ノイズデータをベイジアンネットワークを用いた手法により除去することで、異常検知精度が向上することを確認した。

4. ワイヤレスセンサネットワークによる異常検知

4.1. 設置が容易なセンサネットワークの開発

我々は独自にワイヤレスセンサネットワークシステムの設計・開発を行った。センサユニットは電池駆動、無線通信が可能で人感センサを搭載し、置くだけで設置が完了する。受信ユニットには、マイクロプロセッサを内蔵しており、センサ ID と検知時刻を統合してサーバへ送信する。以上により、センサユニットに関しては置くだけで設置が完了するワイヤレスセンサネットワークシステムを開発した。受信ユニットはセンサユニットから送信される情報と ZigBee モジュールが個々に持つ 64bit アドレスをサーバへ送信する。サーバは受信したデータを web ページを通して表示し、ユーザがセンシング状況を閲覧することを可能としている。

システムの概要を図 1 に示す。

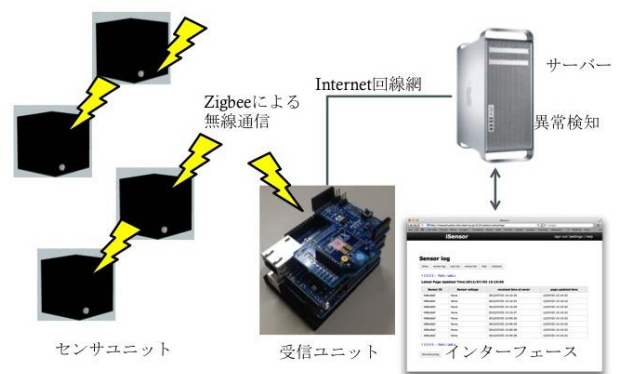


図 1: システム概要図

また実際にセンサユニットを設置する際には回路基板を筐体に格納して設置を行う。使用する筐体も事前実験での知見 [9] からアクリル製 75mm 角の黒い立方体を用いて設置が容易な構造としている。筐体の外観と内部の写真を図 2 に示す。

センサユニットの構成を以下に示す。

- Zigbee モジュール (Digi Corp. Xbee S2B)
- モーションセンサ (Panasonic Corp. Motion Sensor Napiion シリーズ)

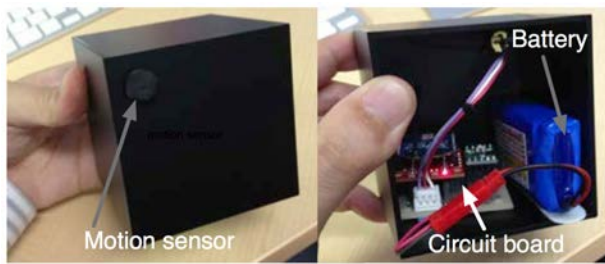


図 2: センサユニット筐体外観図, 内部図

- DC-DC コンバータ
(Linear Technology Corp. LTC3105)
- Li-バッテリー
(S.T.L Japan Corp. LI-3400SP 3.7V4000mAh)

現在はユニバーサル基板にカイロ実装を行っており、75mm 角と小型化を実現した。我々は専用の回路基板を設計することで更なる小型化を目指している。また、低消費電力の DC-DC コンバータを用いることで低消費電力を実現している。これによりバッテリーでの稼働時間を延ばすことができた。本 DC-DC コンバータは 250mV と超低電圧からの入力が可能となっており、主に環境発電（エネルギーハーベスト）や太陽光発電モジュール向けに設計されている。具体的には、振動発電や小型の太陽光発電モジュールからの超低電圧の入力にも耐えうる製品であるため、将来的には内蔵バッテリーだけでなく微小な振動発電装置や太陽光発電モジュールなどのユニットとのハイブリッド構造とすることで稼働時間を延ばせるよう採用している。

更に、センサユニットは通常スリープ状態になっておりセンサからの入力により起動し受信ユニットへセンシング情報を送信している。本機能も低消費電力を実現するために実装している。

以上の機能を実装することで現時点では2ヶ月の連続動作が可能となっているが、最低でも1年間の連続稼働を目指し Zigbee の低消費電力設定に加え太陽光パネル等の発電装置とのハイブリッドモデルを計画している。

4.2.SVM による異常検知手法

Anomaly Detection[10] の研究分野では一般に異常の定義はその異常の性質から3つのカテゴリに分類されている。1つ目は個々のデータインスタンスがデータの残りの部分に対して異常とみなすことができる場合、そのインスタンスは点異常と呼ばれる。2つ目はデータインスタンスが特定の文脈に対して異常がある場合、そのデータインスタンスは文脈異常（または条件付き異常）と呼ばれる。各データインスタンスはコンテキスト属性と行動属性を持ち、コンテキスト属性により各インスタンスの状況を行動属性で各インスタンスの文脈とは関係のない特性を定義する。3つ目は関連するデータインスタンスのコレクションがデータセット全体に対して異常である場合、そのデータインスタンスのコレクションは集団異常と呼ばれる。

本研究では介護ドメインと研究室ドメインにおいて

共通する「日常とは違う時間帯での活動という文脈異常」の検出を目的としている。また実証実験では研究室において行っているが、共通する異常を対象としている点では研究室ドメインでの異常検出手法は介護や見守りドメインにおいても同様に適用可能であると考えている。

高齢者見守り支援における異常とは、日常の活動している時間帯という文脈からデータを解析し、日常とは違う異なる行動という文脈異常を検出することを目的としている。また、機械学習により設置宅毎に通常の生活パターンを学習させることによって個々の生活パターンに応じた異常検知を行うことが可能となる。

実際にワイヤレスセンサネットワークシステムから取得されたセンサデータの例を表1に示す。取得データは、センサノードの固有 ID とバッテリー電圧、及びセンサ検知時刻である。

表 1: 取得センサデータ例

デバイス ID	電圧	検知時刻
40981d4f	3.8	2012-08-20 15:49:17 +0900
40981d4f	3.8	2012-08-20 15:49:31 +0900
40981d4f	3.8	2012-08-20 15:49:32 +0900
40981d4f	3.8	2012-08-20 15:49:36 +0900
40981d4f	3.8	2012-08-20 15:49:38 +0900
40981d4f	3.8	2012-08-20 15:49:45 +0900
40981d4f	3.8	2012-08-20 15:49:46 +0900
40981d4f	3.8	2012-08-20 15:49:47 +0900

SVM に用いる素性は、開発したワイヤレスセンサネットワークシステムにより収集したデータによって生成する。表1に示したデータにより3つの素性を生成する。1つ目はセンサデータの検知時刻の「時間」、2つ目はセンサデータの検知時刻の「分」、そして3つ目はセンサデータの検知時刻の「秒」である。

以上の特徴を用いて SVM に学習させる。素性は、先頭にラベル、続いて「時間」、「分」、及び「秒」を1行に書き込んだものを使用する。また、実験時には学習データとしての素性を生成する。学習データにおける素性は、先頭に、正例ならば「+1」、負例ならば「-1」を行の先頭に書き込む。上記に記した処理を行う事で、SVM に正常クラスと異常クラスを認識させている。

4.3. 不均衡データへの対処

SVM による異常検知手法で使用する SVM は教師あり学習を行い分類境界を設定する。教師あり学習を使用した異常検出では訓練データ内の異常データは正常なデータと比較して、はるかに少ないという問題がある。本問題は機械学習やデータマイニングの研究分野で問題として多く挙げられており、データセットが不均衡であるため、分類性能の低下を引き起こしている[18]。そのため、多くの解決法が提案されている。具体的には、ランダムなオーバーサンプリング、ランダムなアンダーサンプリング、及び誘導されたオーバーサンプリング、誘導されたアンダーサンプリング、またはこれ

らの技術を組み合わせた手法が存在する。

アルゴリズムレベルでは次のような解決法が提案されている。クラスの不均衡に対処するためそれぞれのクラスのコストを調整するもの、判定しきい値を調整するもの、デシジョンツリーの葉の確率推論を調整するものなどがある。

本研究では、負例データをランダムにオーバーサンプリングすることにより不均衡を解決する SMOTE アルゴリズム [17] により予測精度を向上する手法を採用した。更に、観測された異常データにおいてノイズが多数観測されたため、ベイジアンネットワークによるノイズ除去を行っている。次の 4 章で詳細を述べる。

5. ベイジアンネットワークを用いたノイズ除去手法

5.1. センサネットワークの設置

本章では、ベイジアンネットワークを用いたノイズ除去手法について述べる。本研究で開発した設置が容易なセンサネットワークでは、一般ユーザーが容易に設置できることを特徴としている。しかし、センサ設置の知識がないユーザーが設置した場合、白熱灯やモーター等の熱源反応を気にせず設置することも考えられる。そのため、本研究では外的要因によるノイズを除去することで異常検知の精度が向上させる。

実際のデータセットにも外部環境の変化によるセンシングデータといったノイズが含まれており、ノイズが予測精度を下げていることが判明したため、ベイジアンネットワークを用いたノイズ除去の手法を提案する。本研究では、開発したセンサネットワークを本研究室内の実験室に 3 つのセンサを設置してセンシングデータを収集した。センサユニットは実験室の動線にあわせて設置しており部屋全体をセンシング可能とすべく動線に沿った配置となっている。実験室の見取り図とセンサユニットの配置図を図 3 に示す。また図 3 では各センサユニットの検知範囲も示している。センサユニットに搭載されたモーションセンサは、距離が離れると検知範囲が線形に増加するため、センサユニットを中心として、扇状の検知範囲となる。本実験で利用したデータは 2012 年 12 月 1 日から 2012 年 12 月 8 日のデータを用いた。図 3 が示すように実験室は中央に会議机があり、通路が多く存在している。そのため、一般の住宅のように廊下に向けて設置した場合と類似した環境と考える。

5.2. ベイジアンネットワークの構築

本実験では実験室内の各センサ a, b, 及び c の検知範囲を A, B, 及び C と設定し、4 つのノード A, B, C, 及び室外を使用して実験室内の人の移動モデルを構築した。室外より実験室内に入るためには検知範囲 A を通らなければならないため、構築した移動モデルでは室外と検知範囲 A 間の移動と、室内での移動である検知範囲 A と検知範囲 B 間の移動、検知範囲 A と検知範囲 C 間の移動、及び検知範囲 B と検知範囲 C 間の移動が可能と設定している。そのため、検知範囲 B から室外、検知範囲 C から室外への移動は不可能と設定している。実際に構築した移動モデル図を図 4 に示す。

構築するベイジアンネットワークは原因となる確率

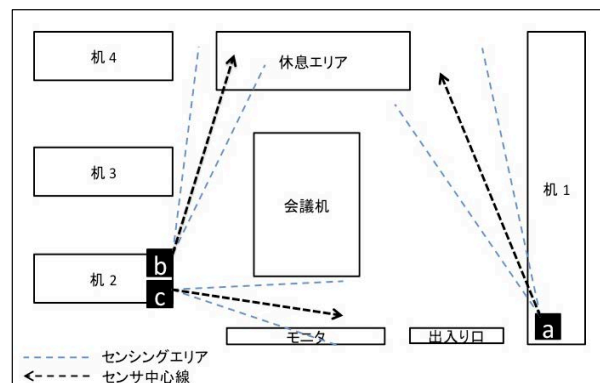


図 3: 実験室の配置図とセンサユニット配置図

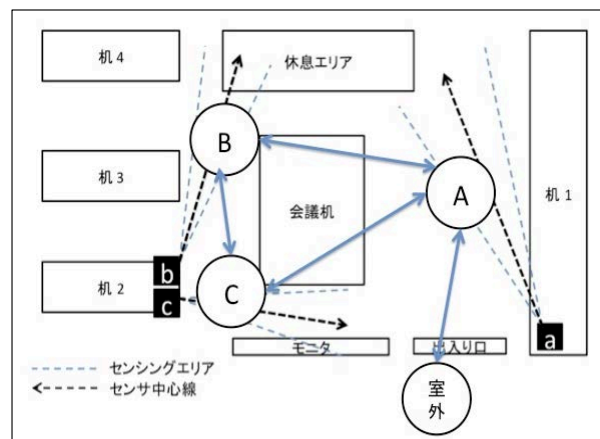


図 4: 実験室内での移動モデル図

変数をセンサを設置した実験室内のセンサ a の検知範囲からセンサ b の検知範囲への移動（以降 ab と表示する）などと設定する。

原因ノードとなる確率変数の確率設定の方法について述べる。確率設定ではセンサデータから一定時間以内に連続して発生したセンサデータの並びをカウントする。本実験では一定時間を 1 分間に設定してカウントした。1 分間という数値は、その場にいる人間をモーションセンサで検知した場合 1 分間の間に、全く動作をしないことはあり得ないため、設定した。

以上により、センサデータの並びをカウントすることで、各センサの検知範囲内からの遷移確率を求める。

本実験ではそれぞれのセンサデータの並びのカウントと遷移確率を求めた。表 2 にセンサの並びのカウントと遷移確率を示す。

表 2 の遷移確率を遷移確率行列で表す。式 (1) に遷移確率行列の定義を示す。 δ (一定時間後) にセンサが反応した場合の列については 3 つの組み合わせのみとなる。

表 2: 実験から求められた遷移確率

センサデータの並び	カウント	遷移確率
AA	3836	0.4787818
AB	1421	0.1773590
AC	2658	0.3317524
AΔ	97	0.0121068
BA	1393	0.0733312
BB	7500	0.3948200
BC	9811	0.5164772
BΔ	292	0.0153717
CA	2696	0.0945169
CB	9774	0.3426588
CC	1568	0.5498878
CΔ	369	0.0129365
ΔA	87	0.1147757
ΔB	301	0.3970976
ΔC	370	0.4881266

$$P = \begin{pmatrix} p_{aa} & p_{ab} & p_{ac} & p_{a\delta} \\ p_{ba} & p_{bb} & p_{bc} & p_{b\delta} \\ p_{ca} & p_{cb} & p_{cc} & p_{c\delta} \\ - & p_{\delta a} & p_{\delta b} & p_{\delta c} \end{pmatrix} \quad (1)$$

式 (2) にセンサデータより計算した遷移確率行列を示す。

$$P = \begin{pmatrix} 0.47878 & 0.17735 & 0.33175 & 0.01210 \\ 0.07333 & 0.39482 & 0.51647 & 0.01537 \\ 0.09451 & 0.34265 & 0.54988 & 0.01293 \\ - & 0.11477 & 0.39709 & 0.48812 \end{pmatrix} \quad (2)$$

$$\pi_1 = P(a), \pi_2 = P(b), \pi_3 = P(c), \quad (3)$$

$$\pi_4 = P(\Delta)$$

$$\pi = \pi P, \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \quad (4)$$

定常確率 π にて表すことで、各ノードの定常確率を式 (3) で定義すると式 (4) より定常確率が求められる。定常確率は計算の結果表 3 のように求められる。

表 3: 実験から求められた定常確率

π_1	0.138121243
π_2	0.331184645
π_3	0.504725888
π_4	0.025968225

求めた定常確率と遷移確率を積算し、各移動毎に生起する確率を求めた。求めた各移動の生起する確率を表 4 に示す。

表 4: 実験から求められた原因ノードの確率

$P(\delta a)$	0.0046527
$P(aa)$	0.0661299
$P(ab)$	0.0487832
$P(ac)$	0.0935272
$P(bb)$	0.1307583
$P(bc)$	0.3439981
$P(cc)$	0.2775426
$P(\text{noise})$	0.0346079

5.3. 結果ノードの設定

構築するベイジアンネットワークは実験室内を一定時間以内に連続して発生したセンサデータの並びと設定する。次に、条件付き確率 $P(X|Y)$ を求めることで、 $X \in$ 結果ノード、及び $Y \in$ 原因ノードが求められる。ここで、本実験では実験室内において検知範囲 A から検知範囲 A までの周回移動を 100 回行うことで、発生するセンサデータ列のサンプリングを行った。サンプリングにより得られたデータより原因ノードの確率を計算した。以上の結果からセンサ間の移動結果ノードを観測各原因ノードの確信度を計算した。確信度の結果は表 5 に示す。

表 5 の太字項目が各結果ノードの確率変数が観測された時最も確信度の高い項目になる。この結果からノイズが原因であると推定されるセンサデータの並びは $b\delta, c\delta, \delta b$ 、及び δ の 4 つである。この結果から図 4 で定義した実験室内の移動モデルから現実的に移動が不可能である $b\delta$ と $c\delta$ の移動が生起する確率はノイズと設定した。評価実験ではセンサデータを除去前と除去後で予測精度の比較を行う。

6. 評価実験

6.1. 実験設定

本章では、研究室内に設置されたセンサシステムからのデータを SVM を用いた異常検知により評価する。次に、ベイジアンネットワークを用いた提案手法によりノイズと推定されたセンサデータの並びを訓練データから除去し、得られた評価値の比較を行うことで提案手法の有効性を示す。

本研究において、異常検知手法の評価指標として、訓練データから異常クラスとラベル付けられているデータを分類する際の正解率、適合率、再現率、及び F 値を用いて評価を行う。

正解率を式 (5) で求め、F 値の値は、適合率 (Precision) を式 (6) で求め、再現率 (Recall) を式 (7) で求め、各値を用いて式 (8) で F 値を計算する。正解率は、異常クラスと正常クラスと分類されたデータインスタンス数のうち、どれだけ正しく異常クラスと正常クラスに分類されているかを示し、適合率は、異常クラスと分類されたデータインスタンスのうち、どれだけ正しく異常クラスと分類されているかという正確性を表している。再現率は訓練データに含まれる異常クラスと

表 5: 原因ノードの確信度

	aa	ab	ac	aδ	ba	bb	bc	bδ	ca	cb	cc	cδ	δa	δb	δc
P(NOISE X)	0.021	0.021	0.020	0.021	0.021	0.020	0.020	0.554	0.020	0.020	0.020	0.448	0.021	0.665	0.643
P(δA X)	0.008	0.003	0.003	0.427	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.427	0.003	0.003
P(AA X)	0.742	0.040	0.039	0.305	0.040	0.039	0.038	0.039	0.039	0.038	0.038	0.039	0.305	0.039	0.039
P(AB X)	0.116	1.000	0.029	0.127	1.000	0.066	0.028	0.059	0.029	0.028	0.028	0.029	0.127	0.065	0.029
P(AC X)	0.221	0.057	1.000	0.243	0.057	0.055	0.054	0.056	1.000	0.054	0.092	0.101	0.243	0.056	0.122
P(BB X)	0.080	0.080	0.079	0.080	0.080	0.677	0.076	0.237	0.079	0.076	0.077	0.078	0.080	0.079	0.079
P(BC X)	0.225	0.226	0.224	0.226	0.226	0.465	1.000	0.420	0.224	1.000	0.349	0.380	0.226	0.461	0.453
P(CC X)	0.177	0.178	0.177	0.178	0.178	0.174	0.170	0.176	0.177	0.170	0.829	0.434	0.178	0.176	0.176

ラベル付けられたデータインスタンスの内どれだけ異常クラスと分類できているかという網羅性を表し、適合率と再現率の調和平均により F 値を表す。

$$\text{正解率} = \frac{\text{異常又は正常と判断したデータインスタンス数}}{\text{分類された全データインスタンス数}} \quad (5)$$

$$\text{適合率} = \frac{\text{異常データと判断したインスタンス数}}{\text{異常と分類されたデータインスタンス数}} \quad (6)$$

$$\text{再現率} = \frac{\text{異常クラスと判断したデータインスタンス数}}{\text{テストデータ中の異常データインスタンス数}} \quad (7)$$

$$F \text{ 値} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

本研究で行う評価実験について述べる。

評価実験では、研究室内に設置し稼働実験を行ったセンサシステムによるデータに対し手動でラベル付けを行い、その学習データを 5-fold-cross-validation 及び SVM の各カーネル関数に対しパラメータ調整を行う。

また、本論文では、精度向上手法として提案した不均衡データへの対処法の 1 つの SMOTE アルゴリズムに関し、ラベル付けをした学習データに対し SMOTE アルゴリズムを適用する。その後パラメータ調整を行うとともに、5-fold-cross-validation を各カーネル関数に対して実施して分類精度の比較を行うことで予測精度が向上したことを示す。またベイジアンネットワークによってノイズが原因と推定されたセンサデータの並びをセンサシステムからのデータセットから除去して、データを手動でラベル付を行った後、学習データに対し SMOTE アルゴリズムを適用する。その後パラメータ調整を行うとともに、5-fold-cross-validation を各カーネル関数に対して実施して分類精度の比較を行うことで提案手法がノイズ除去に対し有効であることを示す。

評価実験では、本研究で開発したセンサシステムから 2012 年 8 月 1 日から 2012 年 8 月 31 日までに得られたデータインスタンスを使用した。異常クラスとラベル付けされたデータインスタンスが 2,000 件であり、正常ク

ラスとラベル付けされたデータインスタンスは 70,000 件である。異常クラスと正常クラスの合計 72,000 件のデータインスタンスを用いて、訓練データの生成を行った。正常クラス及び異常クラスのラベル付けは手動で行い、クラスを判別する基準として、前後 1 週間のデータと比較し日常的に活動していない時間帯を区別した。具体的には、日常的に移動していない時間帯に反応しているデータインスタンスは異常とラベル付けをし残りは正常とラベル付けを行った。

6.2. カーネル関数ごとの比較結果

本研究で提案した SVM によるデータインスタンスの検知時刻を素性に用いた異常検出手法による異常クラス正解率、適合率、再現率、及び F 値を表 6 に示す。

表 6: 提案手法の各カーネル関数毎の評価値

	正解率	適合率	再現率	F 値
線形カーネル	0.853	1.000	0.853	0.921
多項式カーネル	0.859	0.995	0.861	0.923
RBF カーネル	0.859	0.995	0.861	0.923
シグモイドカーネル	0.776	0.871	0.866	0.869

表 6 を見ると RBF カーネルを使った学習方法が最も良い予測精度をだしている。しかしながら用いている訓練データが異常クラスとラベル付けされたデータインスタンスが 2000 件、正常クラスとラベル付けされたデータインスタンスを 7 万件、合わせて 7 万 2 千件の訓練データで、異常正常の割合が約 9 対 1 と極端で不均衡クラス分布になっている。不均衡なデータを用いた場合、予測モデルの精度が低下することが知られているため、不均衡の解消が必要である。また、使用している訓練データは開発したシステムが蓄積したデータをそのまま使用しているので、開発したセンサユニットの誤作動や故障・検出対象としている日常の活動している時間帯という文脈とは関係のない外部環境の変化によるセンシングデータといったノイズが使用しているデータセットに含まれていると考えられる。ノイズによる予測モデルの精度低下を防ぐためにもノイズの除去が必要である。

6.3. SMOTE アルゴリズムの効果

本論文では不均衡データへの対処によって予測精度を向上していることを示すため、研究室内に設置したセン

システムからのデータを基にラベル付けを行った学習データに対し SMOTE アルゴリズムを適用した。その後パラメータ調整と 5-fold-cross-validation を各カーネル関数に対して実施することで分類精度を示す。SMOTE アルゴリズムを適用するデータは 7 万 2 千件のセンサデータに対し適用し正常クラス 9 千件、異常クラス 9 千件に調節を行った。SMOTE アルゴリズムを用いて不均衡データに対処した異常検出手法による異常クラスの正解率、適合率、再現率、及び F 値を表 7 に示す。

表 7: SMOTE アルゴリズム適用後の各カーネル関数毎の評価値

	正解率	適合率	再現率	F 値
線形カーネル	0.900	0.942	0.868	0.904
多項式カーネル	0.867	0.981	0.798	0.880
RBF カーネル	0.980	0.973	0.987	0.980
シグモイドカーネル	0.934	0.962	0.912	0.936

不均衡データの調整を行った結果、各カーネルによる学習モデルの予測精度は向上した。本結果から不均衡データへの対処を行うことで、予測精度の向上が示せた。また RBF カーネルによる学習モデルの予測精度が高いことから、現在使用しているデータ構造に適しているといえる。

6.4. ノイズ除去手法の効果

本研究では提案したベイジアンネットワークを用いたノイズ除去手法に対し、研究室に設置したセンサシステムからのデータを手動でラベル付けを行うことで学習データを作成する。その後、学習データに対し SMOTE アルゴリズムを適用している。その後、ベイジアンネットワークによってノイズが原因と推定されたセンサデータを収集したデータから除去し、データを手動でラベル付けを行い、学習データを作成する。その後、学習データに対し SMOTE アルゴリズムを適用する。その後パラメータ調整と 5-fold-cross-validation[21] を実施して分類精度を示し、比較を行うことで、提案手法の有効性を示す。

ノイズ除去を行う前の異常検出手法による異常クラスの正解率、適合率、再現率、及び F 値を表 8 に示す。

表 8: ノイズ除去前の各カーネル関数毎の評価値

	正解率	適合率	再現率	F 値
線形カーネル	0.983	0.983	0.984	0.983
多項式カーネル	0.991	0.998	0.984	0.991
RBF カーネル	0.992	0.993	0.992	0.992
シグモイドカーネル	0.979	0.979	0.980	0.979

ノイズ除去を行った後のデータを使用した異常検出手法による異常クラスの正解率、適合率、再現率、及び F 値を表 9 に示す。

ノイズ除去前の表 8 とノイズ除去後の表 9 の結果を比較して、ノイズ除去を行った結果、RBF カーネル以

表 9: ノイズ除去後の各カーネル関数毎の評価値

	正解率	適合率	再現率	F 値
線形カーネル	0.987	0.983	0.991	0.987
多項式カーネル	0.994	0.997	0.991	0.994
RBF カーネル	0.990	0.989	0.991	0.990
シグモイドカーネル	0.984	0.979	0.990	0.984

外のカーネルでは予測精度が向上した。本結果からノイズ除去による予測精度が向上した。また、ノイズ除去の結果 RBF カーネルを使った学習モデルでは精度が悪化している。理由としてはノイズ除去の際、精度の高い分類境界を作るサポートベクトルに当たるデータを誤って消去したためと考える。

本論文で行った実験の結果から、提案手法である SVM を使ったセンサデータの解析による異常検知手法により本研究の検出対象である「日常とは違う時間帯での活動という文脈異常」が検出できることを確認した。本論文で行った結果から、SMOTE アルゴリズムを適用したデータの解析結果が他のカーネル関数と比較し、予測精度が優秀であることを示した。また、ベイジアンネットワークによりノイズであると推定されるセンサデータの並びを特定し、ノイズを除去することで予測精度が向上した。本結果から提案手法はノイズ除去に対して有効であることを示した。

7. まとめと今後の課題

本論文では、設置が容易なセンサネットワークによる異常検知手法を提案した。提案手法により本研究の検出対象である「日常とは違う時間帯での活動という文脈異常」が検出できることを示した。またベイジアンネットワークを実験により構築してノイズであると推定されるセンサデータの並びを特定し、そのセンサデータを除く前と除いた後のデータセットを用いて分類評価を行った。結果、RBF カーネル以外のカーネルでは予測精度が向上した。結果からノイズ除去による予測精度の向上が示せた。RBF カーネルの性能低下理由としてはノイズ除去の際、精度の高い分類境界を作るサポートベクトルに当たるデータを誤って消去したためと考える。そこでノイズ除去の際にノイズでは無いデータの消去を防ぐため、ベイジアンネットワークの結果ノードの確率変数を 2 つのセンサ列から 3 つのセンサ列に増やして詳細化を行う。また原因ノードを多段化することでノイズ推定の精度を高めることで、より詳細にノイズ除去を行えるよう改善を図る。

謝辞

本研究の一部は、内閣府の先端研究助成基金助成金（最先端・次世代研究開発プログラム）により助成を受けている。

参考文献

- [1] 象印マホービン株式会社, "みまもりホットライン" <http://www.mimamori.net/>

- [2] 青木 茂樹, 大西 正輝, 小島 篤博, 福永 邦雄, "独居高齢者の行動パターンに注目した非日常状態の検出", 電気学会論文誌 (E), vol.125-E, no.6, pp.259-265, June 2005
- [3] 松岡 克典, "住宅内行動の長期蓄積に基づく以上検知手法の検討", 電子情報通信学会技術研究報告, Vol.102,2003
- [4] 矢崎 俊志, 松永 俊雄, 月江 伸弘, "RFID と生体センシング技術を利用した高齢者向け快適生活支援システム", 電気学会研究会資料. IIS, 産業システム情報化研究会, 2009.
- [5] 田中 仁, 中内 靖, "ユビキタスセンサによる独居高齢者見守りシステム", 日本機械学会論文誌, No.75-760, 2009.
- [6] 関 弘和, 堀 洋一, "高齢者モニタリングのためのカメラ画像を用いた異常動作検出", 電子情報通信学会総合大会講演論文集, Vol.122, 2000.
- [7] セコム株式会社. "セコム・ホームセキュリティシステム", <http://www.secom.co.jp/homesecurity/plan/kodate/>
- [8] ALSOK 総合警備保障株式会社. "ホームセキュリティシステム", <http://www.alsok.co.jp/person/hsprice.html>
- [9] Takanobu Otsuka, Tatsunosuke Tsuboi, Takayuki Ito, "Prototyping and evaluation of a wireless sensor network that aims easy installation", THE 26TH INTERNATIONAL CONFERENCE ON INDUSTRIAL, ENGINEERING & OTHER APPLICATIONS OF APPLIED INTELLIGENT SYSTEMS, 2013.
- [10] Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly Detection: A Survey," Technical Report, Department of Computer Science and Engineering University of Minnesota, TR- 07-017, 2007.
- [11] Kumar, V. 2005. Parallel and distributed computing for cybersecurity. Distributed Systems Online, IEEE 6, 2010.
- [12] Spence, C., Parra, L., and Sajda, P. 2001. Detection, synthesis and compression in mammo- graphic image analysis with a hierarchical image probability model. In Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. IEEE Computer Society, Washington, DC, USA, 3.
- [13] Fujimaki, R., Yairi, T., and Machida, K. 2005. An approach to spacecraft anomaly detection problem using kernel feature space. In Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. ACM Press, New York, NY, USA, 401410.
- [14] Janakiram, D., Reddy, V., and Kumar, A. 2006. Outlier detection in wireless sensor networks using bayesian belief networks. In First International Conference on Communication System Software and Middleware. 16.
- [15] Du, W Fang, L., and Peng, N. 2006. Lad: localization anomaly detection for wireless sensor networks. J. Parallel Distrib. Comput. 66, 7, 874886.
- [16] Chatzigiannakis, V., Papavassiliou, S., Grammatikou, M., and Maglaris, B. 2006. Hierarchical anomaly detection in distributed large-scale sensor networks. In ISCC '06: Proceedings of the 11th IEEE Symposium on Computers and Communications. IEEE Computer Society, Washington, DC, USA, 761767.
- [17] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", Journal of Artificial Intelligence Research
- [18] Rehan Akbani, Stephen Kwek, and Nathalie Japkowicz, "Applying Support Vector Machines to Imbalanced Datasets," Lecture Notes in Computer Science Volume 3201, 2004, pp. 39-50.16 (2002) 321357
- [19] 大塚 孝信, 坪井 辰之助, 吉村 卓也, 伊藤 孝行, "設置が容易なワイヤレスセンサネットワークの試作と評価", JAWS2012 予稿集, 2012
- [20] Vladimir N. Vapnik "A Training Algorithm for Optimal Margin Classifiers" 5th COLT, pp.144-152 (1992)
- [21] A Practical Guide to Support Vector Classification, <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>