

センサで取得した人流情報に基づく 異常検知システム

熊谷 兼太郎*, 小野 憲司**

Machine learning system for anomaly detection of pedestrian flow

Kentaro KUMAGAI and Kenji ONO

Abstract:

In this study, a practical system which consists of infrared sensors and computer programs for machine learning was proposed for anomaly detection of pedestrian flow. Utilizing time-series data collected by the sensors at an entrance of a building in a university as training data for machine learning, the proposed system successfully classified the data into two groups, based on characteristics of flow of visitors to the building. Subsequently, the system was applied to data set which is collected in real-time, and it was shown that the system can detect properly abnormal events that occur only with low probability.

Keywords : Anomaly detection, pedestrian flow, infrared sensor, machine learning

要旨:

公共施設の建物の入口などに設置したセンサで来訪者の行動を測定し、離れた場所からもリアルタイムでその情報を取得できるようなシステムを提案するとともに、機械学習の手法を用いて、毎日の建物入場者の人流を特徴に基づいて2群程度に分類し、未知の日についても類似度によってそのどちらかの群にリアルタイムに仕分けたうえで、低い確率でしか発生しない異常な事象を検知する実用的な手法を構築した。

キーワード : 異常検知, 人流, 赤外線センサ, 機械学習

1. 研究の背景と目的

公民館、体育館などの公共施設は、例えば災害発生直後などに通常と異なった利用をされることがある。こうした施設を適切に管理するため、利用状況を常に把握し、異常が発生した場合には速やかに対処する必要がある。しかし、そのために人間が24時間監視するのは非効率的である。

1.1. 既往の研究

近年、常時インターネットに接続した人流センサを用いて、人流を自動感知し、リアルタイムに情報を取得する技術の提供が進んでいる¹⁾。また、自動的処理を行うコンピュータ・プログラムを用いて、センサの観測結果に基づいて建物内などの滞在者の人

数をリアルタイムに集計することが比較的容易に実現できる²⁾。従って、このような技術の組み合わせにより、センサで取得した人流情報に基づく異常検知システムの構築は、十分に実現可能と考える。

センサで物体検知をするには、いくつかの方式がある³⁾。そのうち、例として赤外線を使用したセンサについて述べる。その仕組みは、発光部から受光部にむけて赤外線が断続的に発射されていて、その間を人などが通ると受光部に届くはずの赤外線が途切れるので、人流があったと判定されるものである³⁾。発光部と受光部は数 m~10 m 程度離すことが可能なので⁴⁾、建物の入口を横断するように設置できる。赤外線は目に見えない可視光外の電磁波であり、通過する人の行動にほとんど影響を与えない。また、単に人流の有無を判定しているだけで年齢、性別などの属性は収集できないというデメリットがあるものの、この同じ特性は、個人の特定に直接つながる情報を収集しないという意味でプライバシーをある程度保つことができるメリットとして解釈すること

*湘南工科大学 工学部 コンピュータ応用学科
准教授, **京都大学 経営管理大学院 客員教授

もできる。そこで、センサの物体検知に赤外線を使用したシステムの構築が考えられる。

1.2. 研究の目的

本研究では、人流センサ及び異常検知プログラムを組み合わせた異常検知システムを提案することを目的とした。

そのための基礎的検討として、人流センサ及び異常検知プログラムから構成される試行システムを構築するとともに、その検証 (verification) 及び妥当性確認 (validation) を行った。以下、2.1 節で試行システムの構築について述べ、2.2 節で同システムの検証・妥当性確認をした結果について述べる。3 章は、結論である。

2. 異常検知システム

人流センサ及び異常検知プログラムを組み合わせた異常検知システムを考える。具体的には、赤外線センサによって人流をモニタリングし、測定結果をウェブ上のストレージに転送し、コンピュータが自動でリアルタイムに異常検知を行って、施設管理者の判断を補助するようなシステムである。

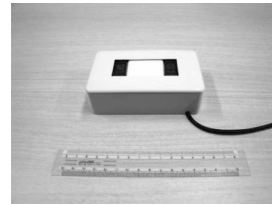
2.1. 試行システムの構築

a) 人流センサの設置

まず、建物への来訪者をモニタリングするために、人流センサを以下のとおり設置した。

センサの機材は、スキャンマチック販売株式会社の「i トレンド・ミニ」⁴⁾を用いた。発光部と受光部は最大で 10 m 離すことが可能である。最も細かい時間解像度は 0.3 s である。従って、1 秒間に 2 人が等時間間隔で通過してもそれぞれ検知できる分解能があることになる。また、安全性について、LED 機器の光生物学的安全性に関する代表的規格の一つである EU EN62471 に基づいて Exempt (何らの光生物学的傷害も起こさない) と評価されている。従って、仮に通過する人が裸眼で発光部を覗き込んだとしても、安全性は十分に確保されていると考えられる。図-1(1)に、使用したセンサの写真を示す。

センサの設置場所は、京都大学吉田キャンパス北部構内 (京都府京都市左京区北白川追分町) の北部総合教育研究棟 4 階フロア入口である。図-2 に、同フロアの平面図を示す。図中の黒丸印の位置に、2 基で 1 組のセンサを配置した。センサ間は距離が 4.7 m で、2 測線の赤外線が平行に射出されている。ここで、フロア外からフロア内へむかう方向を入場、その反対方向を出場と定義する。フロア外側測線で最初に



(1) センサ



(2) ロガー



(3) センサの設置場所



(4) センサを設置した建物

図-1 センサ、ロガー及び設置場所

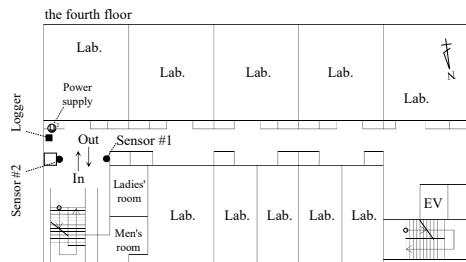


図-2 フロア平面図

物体を検知し、続いて内側測線で検知した場合は入場と判断する。反対に、内側測線で最初に検知し、続いて外側測線で検知した場合は出場と判断する。センサは床面から 1 m の高さに設置した。これは、1 人だけ通過したときに複数人と誤って検出する可能性を極力小さくするため、大人の腹から腰にかけての体幹部の高さにセットしたものである。

図-1(2)に、ロガーの写真を示す。そして、図-2 の黒四角印の位置にロガーを配置した。ロガーとセンサとはケーブルを用いて接続されている。データは、ロガーに内蔵された SD カードに記録される。記録される項目はイベントの発生時刻及び入出場方向である。なお、最終的に目指しているシステムを実現するためには、測定データをリアルタイムに外部へ送信することが必要となる。ただし、今回は基礎的検討のため、SD カードにいったん記録しておいて検証・妥当性確認を行うこととした。今後は、携帯電話回線を介してリアルタイムにインターネット上へ送信するなどの仕組みの検討が必要である。また、災害時の停電、人為的理由による平常時の停電などに備えて、センサ及びロガーは無停電電源装置 (UPS) に接続しておくことが望ましい。ただし、今回は基礎的検討として極力簡易に構築することを優先したため、UPS は組み込まなかった。

センサの設置期間は、2017 年 12 月～2019 年 2 月である。異常検知を行うためには、異常が発生していない平常時のデータが一定期間必要である (以下、これを訓練データと呼ぶ)。そこで、2018 年 1 月 1 日～12 月 31 日のあいだ、訓練データを取得した。ただし、同年 11 月 2 日～5 日は、建物において電源設備のメンテナンス作業がありシステムを停止したため、欠測期間である。すなわち、訓練データの日数 n は $n = 361$ である。また、一日の単位を午前 6:00 から翌日の午前 5:59 までと定義した (以下、特に断りの無い限りこの定義を用いる)。その理由は、センサの設置場所が大学内の研究棟であり深夜まで棟内に長時間留まっている人がいるので、未明から

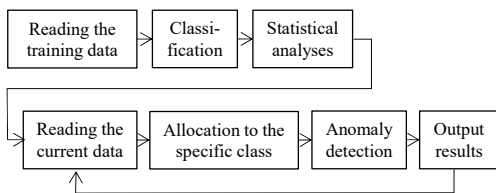


図-3 異常検知の手順概要

早朝にかけて発生するイベントのほとんどは前夜から留まっている人が行き来する行動であると考えたためである。

b) 異常検知プログラム

次に、異常検知プログラムを開発した。図-3に、手順概要を示す。同プログラムは、訓練データの読み込み、訓練データの仕分けをする階層的クラスター分析 (以下、「クラスタリング」と言う。)、及び、統計的指標の計算を行う。続いて、リアルタイムで測定されているデータを逐次読み込んで異常検知を行う。

訓練データの読み込みのプログラミング言語は Fortran であり、Intel Parallel Studio XE 2017 Update 2 for Fortran を Microsoft Visual Studio Professional 2013 に組み込んだ環境で作成した。元々の訓練データは、イベントが発生するたびに入出場方向及び時刻を記録したテキスト形式のファイルである。そのなかから入場だけを抽出し、累積入場者数 (Cumulative number of Visitors, 以下 CV と記す。) を午前 6:00 から翌日の午前 5:59 までの 5 分ごとに集計し、行列データ X のかたちに変換した。ここで、24 時間 $\times 60 \div 5 = 288$ より $T = 288$ として、 X は

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{n,1} \\ \vdots & \ddots & \vdots \\ x_{1,T} & \cdots & x_{n,T} \end{pmatrix} \quad (1)$$

という T 行 n 列の行列である。行列の要素 x の添え字のうち 1 番目はデータ取得日を示す数値 (1, 2, ..., n) である。2 番目は、6:00～6:04 が 1, 6:05～6:09 が 2, ... という順に翌日 5:55～5:59 まで割り振られた、時間帯を示す数値である。

訓練データの仕分けをするクラスタリングのプログラムは、金 (2017) ⁵⁾ を参考に、汎用的な統計解析ソフトウェアである R Ver. 3.5.2 を用いて作成した。訓練データは、1 日の CV も、1 日のなかでの増加の様

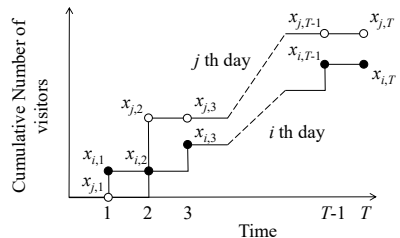


図-4 ユークリッド距離

子も、日によって大きくなばつきがあった。例えば、1日のCVが100人超の日もあれば0人の日もあった。そのような状況で、訓練データの全部を用いて統計的指標（例えば、平均値）を計算すると、データが本来有している特徴が失われてしまう可能性がある。そこでCVの時系列データを、類似度に基づいて適当な数のグループに仕分けることとした。まず、CVの時系列データ同士のユークリッド距離を求める(図-4)。i日目のデータ $x_{i,t}$ ($t \in \{1, 2, \dots, T\}$)とj日目のデータ $x_{j,t}$ とのユークリッド距離 $d_{i,j}$ は以下の式で算出した。

$$d_{i,j} = \sqrt{\sum_{t=1}^T (x_{j,t} - x_{i,t})^2} \quad (2)$$

$d_{i,j} = d_{j,i}$ であり、また、 $i = j$ のとき $d_{i,j}$ はゼロであるから、 $d_{i,j}$ を行列で表現した距離行列 D は対角成分がゼロの n 次対称行列である。

$$D = \begin{pmatrix} 0 & d_{1,2} & \dots & d_{1,n} \\ d_{2,1} & 0 & \dots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n,1} & d_{n,2} & \dots & 0 \end{pmatrix} \quad (3)$$

続いて、クラスタリングを行った。ここで、 $d_{i,j}$ が小さいときに類似度が大きいと考え、個別のデータ同士をまず小グループにまとめ、次に小グループ同士をより大きなグループにまとめ、という手順を繰り返す。その際、個別のデータとグループとのユークリッド距離、あるいは、グループ同士のユークリッド距離の計算方法は複数の手法が提案されている。そして、そのうちのどの手法を用いるべきか、定まった評価方法が提案されているとは言えず、「探索的にさまざまな角度でデータを眺めて、総合的に判断することが必要」とされている⁶⁾。そこで、主要6手法（群平均法、ワード法、重心法、最近隣法、メディアン法、最遠隣法）を用い、得られた結果を樹形図（デンドログラム）でそれぞれ描いた。デンドログラムは、逆さにした木の構造に似たグラフである。ある高さのところで直線を引いて切断することにより、元の大きな1つの木を複数の房に切り分けることが出来る。今回は、2つの比較的大きな房が形成される切り方のうちで樹形図の最も高い（木の根元に最も近い）位置で切り分けることとした。このような作業を、6つのデンドログラムについてそれぞれ行った。

統計的指標を計算するプログラムは、先述と同じFortranを用いて作成した。

また、リアルタイムで異常検知を行うプログラムは、先述と同じFortranを用いて作成した。井出(2015)⁷⁾によると、ホテリング理論⁸⁾により以下の定理が成立する。すなわち、1次元の観測データ $\{x_1, x_2, \dots, x_N\}$ がある正規分布に独立に従い、新たな観測値 x' も同じ分布に独立に従うとする。 $\hat{\sigma}^2$ は標本分散、 $\hat{\mu}$ は標本平均とし異常度 $a(x')$ を以下で定義したとき、

$$a(x') \equiv \frac{1}{\hat{\sigma}^2} (x' - \hat{\mu})^2 = \left(\frac{x' - \hat{\mu}}{\hat{\sigma}} \right)^2 \quad (4)$$

$N \gg 1$ ならば $a(x')$ は自由度 $k = 1$ のカイ二乗分布に従う。

$$a(x') \sim \chi^2(k | k=1) \quad (5)$$

この定理を参考に、それぞれの変数に添字 t をつけた次の式(4')及び(5')も成立すると考える。すなわち、今回の観測データは任意の時間帯 t について、ホテリングの理論による定理が成立すると考える。

$$a(x'_t) \equiv \frac{1}{\hat{\sigma}_t^2} (x'_t - \hat{\mu}_t)^2 = \left(\frac{x'_t - \hat{\mu}_t}{\hat{\sigma}_t} \right)^2 \quad (4')$$

$$a(x'_t) \sim \chi^2(k | k=1) \quad (5')$$

(5')のときのカイ二乗分布の確率密度関数 f は、

$$f(a(x'_t) | k=1) = \frac{1}{2\Gamma\left(\frac{1}{2}\right)} \cdot \left(\frac{a(x'_t)}{2}\right)^{\frac{1}{2}} \exp\left(-\frac{a(x'_t)}{2}\right) \quad (6)$$

ここで、 Γ はガンマ関数であり、

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

より、(6)は

$$f(a(x'_t) | k=1) = \frac{1}{2\sqrt{\pi}} \cdot \left(\frac{a(x'_t)}{2}\right)^{\frac{1}{2}} \exp\left(-\frac{a(x'_t)}{2}\right) \quad (6')$$

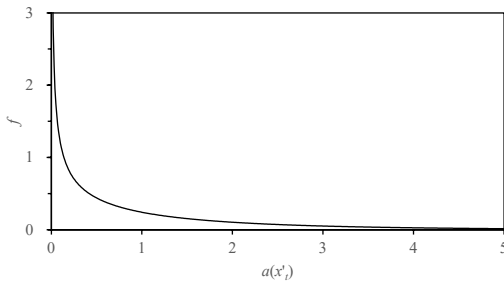


図-5 異常度と確率密度関数の関係

表-1 発生確率 (p) とそれに対応する異常度 (a(x'))

p	.995	.990	.985	.975	.970	.950	...
a(x')	.00004	.00016	.00035	.00098	.00141	.00393	...

...	.050	.030	.025	.015	.010	.005
	3.8415	4.7093	5.0239	5.9165	6.6349	7.8794

図-5に、 $a(x')$ と f の関係を描く。Aをある実数としたとき、 $a(x') = A$ となる事象の発生確率 p は、図中で曲線、横軸 $f = 0$ 及び直線 $a(x') = A$ に囲まれた範囲の面積に相当する。最後に、求められた p と、あらかじめ決めておいた閾値とを比較して、前者が後者よりも小さければ異常と判定する。表-3に、 p とそれに対応する $a(x')$ のうち代表的な組合せを示す。これによれば例えば、 p が5%以下となるのは $a(x')$ が3.8415以上のときである。式(5)より

$$\left(\frac{x'_i - \hat{\mu}_i}{\hat{\sigma}_i} \right)^2 \geq 3.8415$$

$$\therefore \frac{x'_i - \hat{\mu}_i}{\hat{\sigma}_i} \geq \pm 1.959 \dots = \pm 2 \quad (7)$$

となり、観測された値が $\hat{\mu}$ より $2\hat{\sigma}$ 以上大きいまたは小さいとき、通常では5%以下の確率でしか発生しないような異常が検知されたことになる。

2.2. 検証・妥当性確認

a) 訓練データの特徴

図-6に、毎日のCVの時間変化を5分刻みの階段状グラフのかたちで示す。横軸は午前6:00から翌日午前5:59までの時刻をとり、縦軸はCVの値をとった。

1日のCVは、最大116人で最少0人であり、日によ

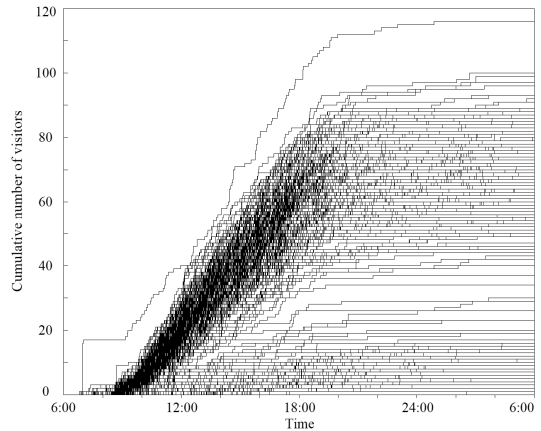


図-6 毎日のCVの時間変化

ってばらつきがあった。図によると、40人程度を境に比較的多い日と少ない日とがある。1日のCVが比較的多い日の日付を調べたところ、ほとんどが平日であった。またCVの変化は、ばらつきがあるものの8:45頃から大きくなり、朝から人流が活発になっている。CVの変化が大きい状態は19時頃まで続き、その前後の時間帯は小さい。すなわち、日中の時間帯の利用が中心と分かる。それに対し、1日のCVが比較的小さい日は、CVの変化について明瞭な立ち上がりは見られない。

このように、訓練データは2つのグループで構成されている特徴をもったデータであった。

b) クラスタリングによる仕分け

図-7に、主要6手法のうち例として群平均法を用いてクラスタリングした結果のデンドログラムを示す。図中の下側の、木の枝先に相当する位置には n 日分のデータが並んでいる。縦軸はユークリッド距離であり、個別のデータ同士、個別のデータとグループ、または、グループ同士の類似度を示す。任意の2つのデータについて、枝分かれの位置が図中で上にある(木の根元に近い)ほど、類似度が小さい。また、図中の点線は、2つの比較的大きな房が形成されるように切分けた高さを示している。

仕分け精度を評価するため、まず、京都大学の年間行事予定表などに基づいて、校休日及び平日のリストを作成した。校休日は、土曜・日曜・祝日に加えてお正月(1/1~3)、創立記念日(6/18)、お盆(8/13~15)及び年末(12/31)で、欠測日(11/3~4)を除外して122日である。平日はそれ以外で、欠測日(11/2及び11/5)を除外して239日である。そして、クラスタリングの結果と比較することにより、本来

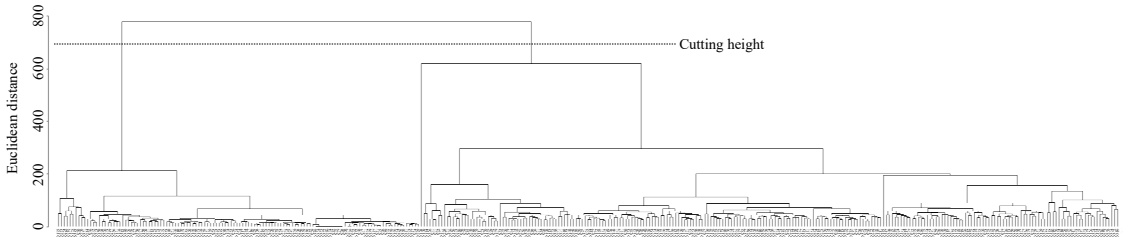


図-7 クラスタリング結果（群平均法）のデンドログラム

表-2 主要6手法の仕分け精度

手法	n_{e1}	n_{e2}	n	r
群平均法	2	2	361	0.011
ウォード法	2	2		0.011
重心法	2	3		0.014
最近隣法	6	2		0.022
メディアン法	9	2		0.030
最速隣法	41	2		0.119

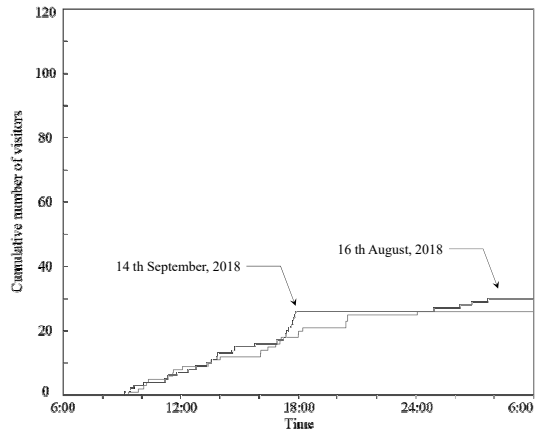
は平日であるがクラスタリングの結果では校休日
に仕分けされた日数 n_{e1} と、本来は校休日であるが平日
に仕分けされた日数 n_{e2} とをそれぞれ求めたうえで、
仕分けエラー率 r を算定した。

$$r = (n_{e1} + n_{e2}) / n \quad (8)$$

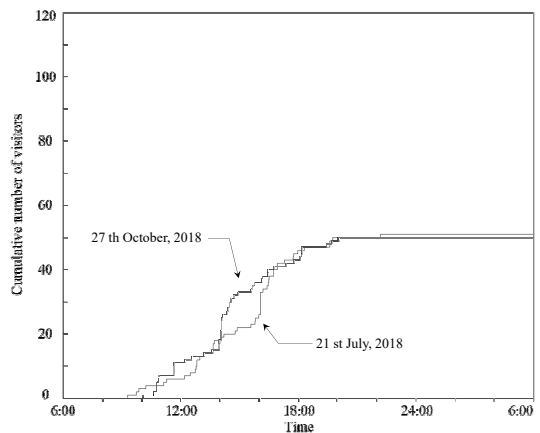
表-2に、その結果を示す。主要6手法のうちで仕
分け精度が最も良かったのは群平均法及びウォード法
で、どちらも $r \approx 0.011$ であった。また、2手法とも仕
分けした結果に違いは無かった。ここからは、群平
均法及びウォード法の結果に限定して述べる。

n_{e1} または n_{e2} に該当した日を調べた。まず、8/16(木)
及び9/14(金)はいずれも本来は平日であるが、校
休日に仕分けされていた(図-8(1))。その理由は、8/16
はお盆(8/11~15)の直前で1日を通して来訪者が比
較的少なかったこと、夜に近隣で五山送り火の「大」
の字の点火があって、建物から漏れる人工光の抑制
のため建物利用を通常よりも控えたことが考えられ
る。また、9/14は連休(9/15~17)の直前であるた
め18時以降の来訪者が少なかったことが考えられ
る。

次に、7/21(土)及び10/27(土)はいずれも本来
は校休日であるが、平日に仕分けされていた(図
-8(2))。その理由は、7/21は建物の1階会議室で開か



(1) 本来は平日であるが校休日に仕分け



(2) 本来は校休日であるが平日に仕分け

図-8 仕分けエラー

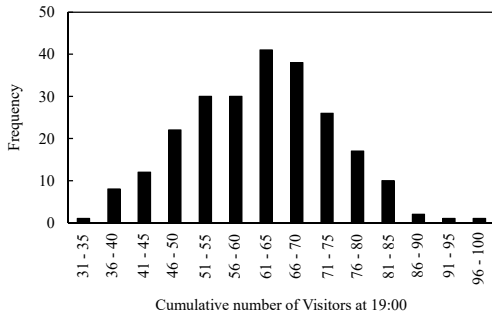


図-9 19:00時点のCVの度数分布

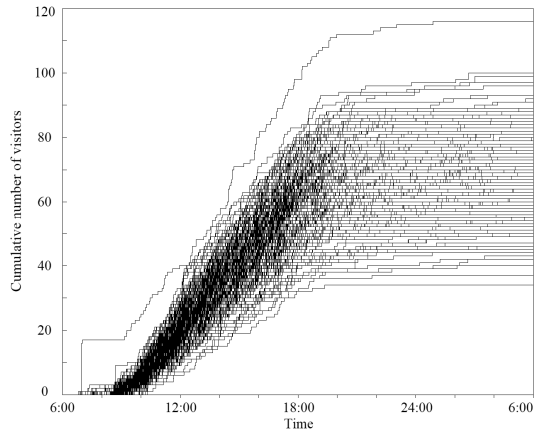
れた会合、10/27は同じく1階ホールで開かれた講演会の影響がそれぞれ考えられる。

ここでは基礎的検討として、 r が最も小さかった群平均法及びウォード法を用いてクラスタリングした結果をそのまま、「平日」及び「校休日」として扱うことにした。ただし、 r の許容範囲の適切な決め方については今後の課題である。

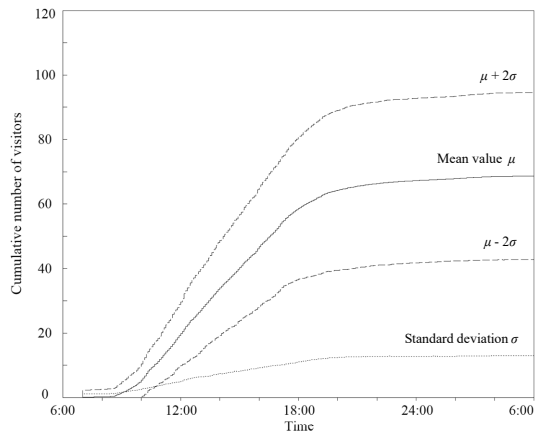
c) センサの検定

センサの検出誤差を調べるため、目視調査で検定した。調査は、2019年2月6日の10:00～11:00、14:00～15:00及び16:00～17:00の計3時間に、センサ設置場所の通過人数、方向及び時刻を目視で記録した。その結果、調査時間内の入場者数は計31人であった。それに対し、センサで記録された入場者数は34人であった。目視で記録した実際の人流の31人分はいずれも正確に記録されており、それに加え3人分多かった。すなわち、今回使用したセンサは人数を約9%多く検出する傾向があったと言える。実際の人流よりも多かった3人分のイベントは、個別かつ調査時間内でランダムに発生していた。それぞれ実際の入場者があった直後の1～3秒後に記録されていたので、荷物や歩き方などに影響を受けていると考えられる。このように個別にランダムに発生する検出誤差は、異常検知への影響は比較的小さいと考えて無視することにした。なお、入場者1名と出場者1名が目視ではほぼ同時に通過するイベントが調査時間内に1回発生したが、センサでは入場者と出場者がそれぞれ正確に検出されていた。

ただし、同時に複数の人が同じ方向に入場または出場するイベントや、数秒程度のあいだに多数の人が連続的に入場するイベントは時間内に発生しなかったため、これらのイベントについては現段階では十分に検定できていない。今後、模擬的に人流を生



(1) 毎日のCVの時間変化



(2) 平均値及び標準偏差

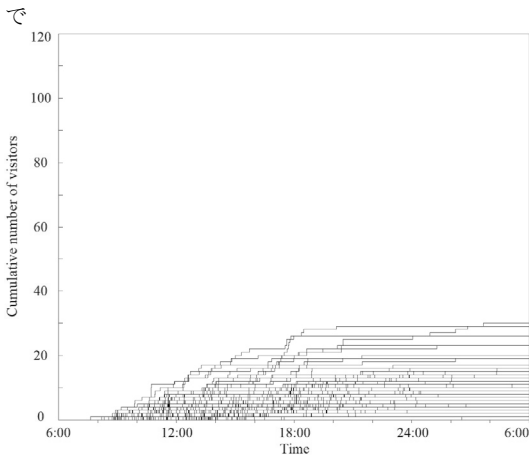
図-10 訓練データ（平日）

成するなどして検定を行う必要がある。

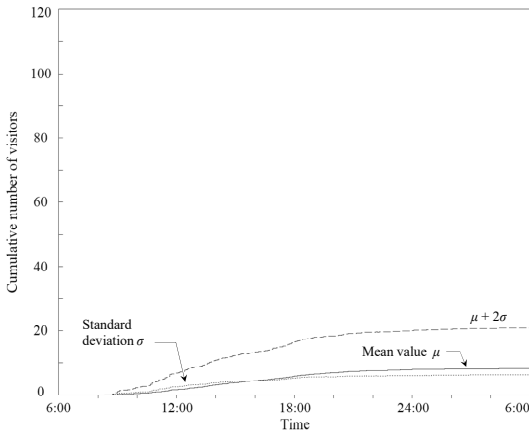
d) 平均値と標準偏差

図-9に、平日の訓練データについて毎日のCVの19:00時点での度数分布を示す。図に描かれた分布は、一つの峰があり概ね左右へと対称に下る形状をしている。他の時刻についても同様である。そこで、データは正規分布に従っているとみなした。

図-10(1)は、b)項で仕分けしたうち平日の239日分の訓練データについて、毎日のCVの時間変化を示している。すなわち、図-6は全部の訓練データを図化したものであったが、そのうち平日だけを抽出したもの



(1) 毎日のCVの時間変化

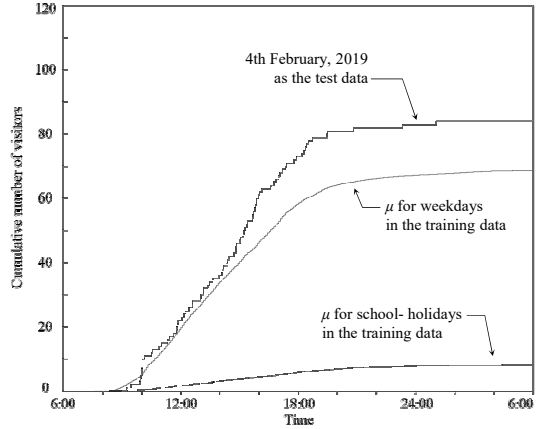


(2) 平均値及び標準偏差

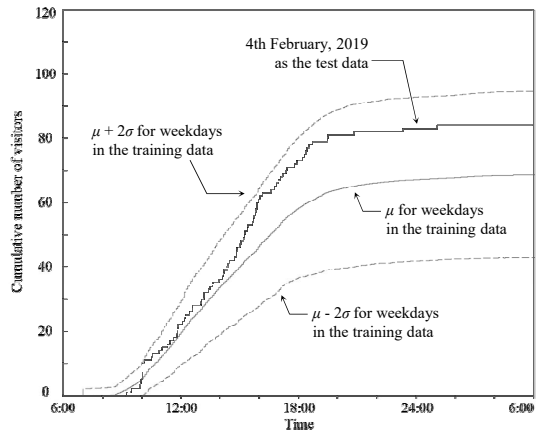
図-11 訓練データ（校休日）

ある。この平日の訓練データについて、毎日のCVの平均値 μ 及び標準偏差 σ を5分刻みでそれぞれ求めた。図-10(2)に、その時間変化を階段状グラフのかたちで示す。図中に実線で描いた μ は、8:45以前は1人未満であるが8:50に1.0人となり、その後は19:00ころまで概ね一定の傾きで増加する。そして、傾きは小さくなり翌日朝までに68.7人となった。また、点線で描いた σ は、6:55以前はゼロであったが7:00に1.0人となり、19:00ころまで概ね一定の傾きで増加する。そして、傾きは小さくなり翌日朝までに13.0人

となった。



(1) CVの時間変化



(2) 平日の $\mu \pm 2\sigma$ の時間変化との重ね合わせ

図-12 テストデータを用いた妥当性確認

図-11(1)に、校平日の122日分の訓練データについて、毎日のCVの時間変化を示す。また、図-11(2)は、このデータについて μ 及び σ をそれぞれ描いた。 μ は、11:00まで1人未満であるが11:05に1.0人となり、その後は緩やかに増加し、翌日朝までに8.2人となった。また σ は、9:00以前は0.5人未満で、9:05に0.5人となったあと翌日朝までに6.3人となった。

e) テストデータを用いた妥当性確認

2019年2月の第1月曜日である2月4日に、構築したシステムの妥当性確認を行うためのデータ（以下、

これをテストデータと呼ぶ。)を取得した。図-12(1)に、同日のCVの時間変化を示す。また、図中にd)項で得られた平日と校休日の μ を重ねて示した。

テストデータはリアルタイムに得られるデータを模擬している。従って、時刻 t_p ($t_p \in \{1, 2, \dots, T\}$)の時点では $1 \sim t_p$ の範囲のデータしか得られておらず、 $t_p \neq T$ であれば $t_p + 1 \sim T$ の範囲のデータは利用できないことになる。

そこで、まず $t_p = 1$ として、テストデータと平日の μ の時系列データのそれぞれ1番目のデータ(すなわち、6:00~6:04の5分間のデータ)だけを用い、両者のユークリッド距離 d_w を求めた。また同様に、テストデータと校休日の μ のそれぞれ1番目のデータだけを用い、両者のユークリッド距離 d_h を求めた。そのうえで、 d_w と d_h とを比較して、 d_w が小さければ平日と判定した。または、 d_h が小さければ校休日と判定した。

続いて、 $t_p = 2$ として、1番目と2番目のデータ(すなわち、6:00~6:04と6:05~6:10の10分間のデータ)だけを用いて d_w と d_h を計算し、平日か校休日かの判定を行った。以後、 $t_p = T = 288$ まで逐次的にこの過程を繰り返した。

このようにして、テストデータの取得日が平日か校休日かの逐次的な判定を行った。その結果、9:29 ($t_p = 41$)までは判定不能($d_w = d_h$)あるいは校休日と仕分けされていた。2月4日は平日であり特別に考慮すべき事情もなかったようであるから、本来は平日であるが誤って校休日とその時間帯は仕分けされる状態が発生したことになる。それ以降の時間帯は平日として正しく仕分けされていた。

以上のとおり、来訪者が少ない時間帯について仕分けの精度を高めるための検討が必要ではあるものの、毎日の建物入場者の人流をその特徴に基づいて2群程度に分類し、未知の日についても類似度によってそのどちらかの群にリアルタイムに仕分けすることができる実用的な手法を構築することが出来た。

図-12(2)に、テストデータを描いたうえにd)項で得られた平日の $\mu \pm 2\sigma$ の時間変化(図-10(2)で破線を用いて描いたもの)を重ねて示した。図中の上側の破線が $\mu + 2\sigma$ であり、下側の破線が $\mu - 2\sigma$ である。式(7)より、2つの破線で囲まれた領域よりも外側に観測値がある場合は、通常では5%以下の確率でしか発生しない異常が検知されたことになる。テストデータでは異常は検知されなかったが、例えば、12:00のテストデータの値は22人であった。それに対して、同時刻の $\mu = 19.8$ 、 $\sigma = 4.9$ で $\mu + 2\sigma = 29.6$ より、上側の破線は29.6人の位置に描かれている。 $29.6 - 22 = 7.6$ 人であるから、仮に、8人以上の人が

正午から5分間のあいだに入場する人流が発生すれば、リアルタイムに異常を検知できる。

以上のとおり、テストデータでは異常は検知されなかったが、訓練データが正規分布に従うという条件のもとで、例えば5%以下程度の、低い確率でしか発生しない異常な事象をリアルタイムに検知する実用的な手法を構築することが出来た。

f) プログラムの処理時間

計算に用いたPCのプロセッサはIntel Core i7-5960X (8 core, 3.00 GHz, 20 MB L3 Cache)、メモリは64 GB RAM、オペレーティングシステムはWindows 10 Pro 64bitである。今回の訓練データの規模(361日分)であればプログラムの処理時間は数秒程度の短い時間であった。

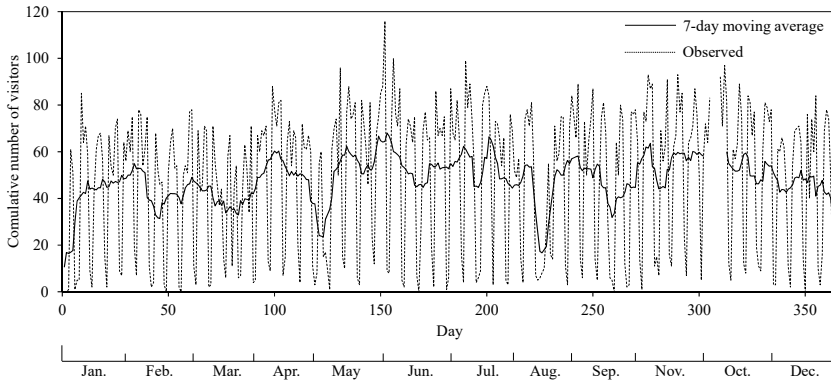
2.3. 議論

訓練データに季節変動などの長周期の変動が含まれているか、そして含まれていない場合、データ取得期間を長くすれば仕分け精度が向上するのか、という点を議論する。

図-13に、破線でCVの日変化を示す。横軸は経過日数、縦軸は1日あたりCVである。土曜日及び日曜日には入場者が比較的少ないので1週間周期の変動が含まれていることは当然に予想されるが、その他にも季節変動などの1週間よりも長周期の変動が含まれているかを調べることをした。汎用プログラミング言語の一つであるPython Ver. 3.7.2を用いて1日あたりCVの日変化について、既往の手法⁹⁾を参考にスペクトルを計算した(図-14)。スペクトルは7.0日について最大のピークがあった。また、7日よりも大きい周期のスペクトルは顕著には示されなかった。

そこで、7日間移動平均値を計算した。ただし、ある日の7日間移動平均値を計算するためには、当該日のデータとともに前後それぞれ3日間のデータが必要となる。センサの設置にあたって、訓練データの取得期間(2018年1月1日~12月31日)の前後の期間も予備的に情報を収集してあった。そこで、2017年12月29日~31日及び2019年1月1~3日のデータもあわせて利用して、7日間移動平均値を求めた。図-13に、7日間移動平均値を実線で示す。それによると、一部に欠測期間があるものの、1月初め・12月末の年末年始、5月初めの大型連休、及び、8月中旬のお盆の休暇期間にそれぞれ7日間移動平均値が30人を下回る結果となった。こうした期間に含まれるデータは、2.2のb)項のクラスタリングによって校休日に仕分けされていて、同d)項で平日の統計的指標を算出する際に適切に除去できている。

以上より、数日~100日程度のスケールに関して



Note: 1) The monitoring data from December 29 to 31 of 2017 was used for calculating the 7-day moving average values of January 1st to 3, 2018. And, the data from January 1 to 3 of 2019 was used for the 7-day moving average values of December 29 to 31, 2018. 2) The monitoring data from November 2 to 5 was not available, due to power failure for maintenance of the building which the sensors were installed.

図-13 CVの日変化及び7日間移動平均

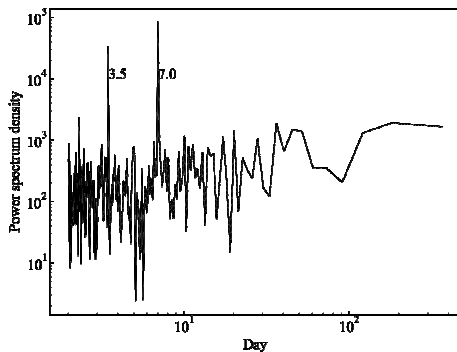


図-14 1日あたりCVのスペクトル

は、7日より長周期の変動は含まれていないものと判断した。従って、訓練データの取得期間をより長くしても、このスケールの周期変動についての精度向上という観点では期待できない。ただし、それより長周期の、半年あるいは年単位の周期変動についてはこの限りではない。例えば季節変動は年単位の周期なので、数年のあいだ継続して人流の測定を行えば、季節変動について検討でき、仮に季節変動が含まれていればその効果について定量的に分析できる可能性があると考えられる。

3. 結論

本稿は、公民館、体育館などの施設を適切に管理するため、利用状況を常に把握し、異常が発生した

場合には速やかに対処するように人流センサ及び異常検知プログラムを組み合わせた異常検知システムについて基礎的検討を行った。その結論は以下のとおりである。

公共施設の建物の入口などに設置したセンサで来訪者の行動を測定し、離れた場所からもリアルタイムでそうした情報を取得できるシステムを提案するとともに、機械学習の手法を用いて、毎日の建物入場者の人流を特徴に基づいて2群程度に分類し、未知の日についても類似度によってそのどちらかの群にリアルタイムに仕分けしたうえで、低い確率でしか発生しない異常な事象を検知するような実用的な手法を構築した。

謝辞

京都大学経営管理大学院港湾物流高度化寄附講座・渡部富博先生、小川雅史先生、平岡美里さん、同大学院官民協働まちづくり実践講座・吉田恭先生、高橋裕美さんに、人流センサの設置・データ取得にあたり支援を頂きました。渡部先生には、特異日についての情報も頂きました。京都大学北部総合教育研究棟の関係者の皆さんに、人流センサのデータ取得にあたりご理解を頂きました。また、東京大学先端科学技術研究センター・柳澤大地先生、Qounts GmbH社・Stefan Leitmannslehnerさんに、それぞれクラスタリングの手法について、及び、人流センサの検出誤差についてご意見を頂きました。ここに記して謝意を表します。

参考文献

- 1) 例えば, LiDAR を使った人流計測サービス「スキア」, <https://www.nikkei-r.co.jp/service/lifestyle/lidar/>, 2021年6月1日確認.
- 2) 入退場者をカウントして混雑状況を表示するセンサー, <https://xtech.nikkei.com/atcl/nxt/column/18/00120/00411/>, 2021年6月1日確認.
- 3) 動線計測ソリューション, <https://www.hitachi-ite.co.jp/products/o-tracking/>, 2021年6月1日確認.
- 4) スキャンマチック販売株式会社 : https://www.scanmatic.co.jp/itrend_mini/, 2021年6月1日確認.
- 5) 金明哲 : R によるデータサイエンス (第2版), 森北出版株式会社, pp. 97-125, 2017.
- 6) Christensen L. and Ruch C. E.: The effect of social influence on response to hurricane warnings, *Disasters*, Vol. 4, No. 2, pp.205-210, Pergamon Press Ltd., 1980.
- 7) 井出剛 : 入門 機械学習による異常検知-R による実践ガイド, コロナ社, pp. 17-36, 2015.
- 8) Hotelling H.: The generalization of student's ratio, *Annals of Mathematical Statistics*, 2(3), pp. 360-378, 1931.
- 9) Qiita, <https://qiita.com/AnchorBlues/items/5497ee68c3a3d64875d4>, 2021年6月1日確認.