Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection

Asma Belhadi [a], Youcef Djenouri [b], Gautam Srivastava [c,d], Djamel Djenouri [e],
Jerry Chun-Wei Lin [f,*], Giancarlo Fortino [g]

[a] Department of Technology, Kristiania University College, Oslo, Norway
[b] Department of Mathematics, and Cybernetics, SINTEF Digital, Oslo, Norway
[c] Department of Mathematics and Computer Science, Brandon University, 270 18th Street, Brandon, Canada R7A 6A9
[d] Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan, Republic of China
[e] Computer Science Research Centre, Department of Computer Science & Creative Technologies, University of the West of England, Bristol, UK
[f] Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, 5063, Norway
[g] Department of Informatics, Modeling, Electronics and Systems (DIMES) University of Calabria, Via P. Bucci, 87036 Rende (CS), Italy

ARTICLE INFO

ABSTRACT

This paper introduces a new model to identify collective abnormal human behaviors from large pedestrian data in smart cities. To accurately solve the problem, several algorithms have been proposed in this paper. These can be split into two categories. First, algorithms based on data mining and knowledge discovery, which study the different correlation among human behavioral data, and identify the collective abnormal human behavior from knowledge extracted. Secondly, algorithms exploring convolution deep neural networks, which learn different features of historical data to determine the collective abnormal human behaviors. Experiments on an actual human behaviors database have been carried out to demonstrate the usefulness of the proposed algorithms. The results show that the deep learning solution outperforms both data mining as well as the state-of-the-art solutions in terms of runtime and accuracy performance. In particular, for large datasets, the accuracy of the deep learning solution reaches 88%, however other solutions do not exceed 81%. Additionally, the runtime of the deep learning solution is below 50 seconds, whereas other solutions need more than 80 seconds for analyzing the same database.

1. Introduction

Our cities and environments are revolutionized by the current information and communication (ICT) technologies including wireless sensing and the Internet of Things (IoT) that enable interconnecting smart objects via machine-to-machine communication protocols as well as through the Internet, advanced machine learning (ML) and knowledge mining approaches, pervasive and mobile computing, and high performance computing (HPC). All of these technologies make our cities and premises smarter, with augmented capacities through cyber–physical systems in which machines and humans interact and act on the environment [1–3]. One of the fundamental services needed in many smart city applications is that of automatic human behavior analysis. Human behavior in cities is a complex process on both spatial and temporal dimensions, which is governed by semantic actions, and latent factors [4–6]. Human behavior depends on many factors such as

intention, need, belief, and environmental aspects. For example, when standing at a red traffic light, people can wait or walk depending on whether there is a car approaching, a police car, or they are looking at their mobile phone when walking. Several areas of smart city applications have been studied in recent years, which is contributing not only to the sophistication of citizens' daily services and quality of life, but also in the development of the economy. Moreover, many social aspects of mankind can be analyzed through the understanding and modeling of human behavior [7,8]. A smarter city provides urban planners with tools to exploit various sources of information about human behaviors to aid in the allocation of land, water, transportation, and many others [9–13]. This work is inline with the current efforts to use cutting edge information and communication technologies (ICT) to realize a real smart city, in which we study the modeling of human behaviors by exploring pedestrian data in a smart city environment [14].

---

* Corresponding author.
  E-mail addresses: asma.belhadi.17@gmail.com (A. Belhadi), youcef.djenouri@sintef.no (Y. Djenouri), srivastavag@brandonu.ca (G. Srivastava), djamel.djenouri@uwe.ac.uk (D. Djenouri), jerrylin@ieee.org (J.C.-W. Lin), giancarlo.fortino@unical.it (G. Fortino).

1.1. Motivation

This research is inline with the ongoing efforts to use cutting edge ICTs for providing advanced smart city applications, and its aim is to improve modeling human behavior for different smart city applications. The existing smart city systems uses human behavior to enhance smart agents. A human behavior knowledge base can be collected from different sources using sensorial data, and Internet of Things (IoT) technologies, social media tools such as Twitter, Facebook, Flickr, Instagram, etc, which enable different representations, including texts, images, videos, time series, and trajectories.

Our goal in this paper is to design a new framework of modeling human behavior, which integrates data mining and deep learning models to understand and analyze human behavior better. The strategy used in this work is based on collective abnormal human behaviors. The smart agents might also interact with the current human behavior knowledge base for decision making. For instance, if a transportation agent detects a grouping of anomalous trajectory of people in some given circumstance, (e.g., low number of Bikes for a given day), then the agent could interact with the human behavior knowledge base to make an informative decision on whether the observation is a real outlier or just related to a particular behavior of people in a certain circumstance. In the healthcare domain for example, there is a high correlation between human activities such as biking in a city and smart healthcare activities. Understanding collective human health behavior helps the smart healthcare agents to improve and predict healthcare activities.

1.2. Contributions

This paper studies collective abnormal human behaviors in a smart city environment and proposes a new model for collective abnormal human behavioral data. Our main contributions can be summarized as follows:

- We formally define the collective abnormal human behaviors problem by considering the individual abnormal human behavioral data as potential candidates. We introduce a new concept of density of a collection of abnormal human behaviors. A collection of abnormal human behavioral data is thus defined as a set of individual abnormal human behaviors that are highly correlated, i.e., with a high number of shared locations.
- We propose solutions based on data mining, and deep learning to accurately identify the collective abnormal human behaviors data. Data mining based solutions explore the nearest neighbors, the density computation, and the feature selection in deriving anomalies from human behavioral data. In addition, two improvements based on ensemble learning, and GPU computing are shown to boost the performance of these solutions in terms of both runtime, and accuracy of the returned anomalies. The deep learning based solution applies a convolution neural network to learn and extract relevant features, and then retrieve anomalies from human behavior data.
- We evaluate the performance of the proposed algorithms using different real human-behavioral databases. To the best of our knowledge, this is the first work that explores collective abnormal human behavior, and there is no relevant candidate in the literature to compare with. We therefore compare the proposed solutions with general anomaly detection solutions (see Section 2 for an overview). The results of our experiments demonstrate that our algorithms outperform the baseline algorithms. Moreover, they also show the effectiveness of the deep learning solution over the data mining based solutions.

2. Related work

Human behavior has recently studied in some smart city applications, notably smart buildings [15,16]. Several solutions have been proposed ranging from technical solutions based on pervasive computing for optimal monitoring [17–19], energy consumption analysis [20], and the use of ML for human for occupancy estimation [21], and activity recognition [22,23]. A detailed review on all these solutions is available in [8]. We are interested in this paper in another category of smart city applications that received relatively less attention amongst the research community; pedestrian behavior analysis. We are particularly interested in analysis from the perspective of collective anomaly detection.

Solutions to collective anomaly detection employ statistical models to derive the group of outliers [24]. Chalapath et al. [25] considered the use of a deep generative model and test it on various image applications. The outlierness for each group in the input data is estimated by group reference function using a standard back-propagation algorithm. Tang et al. [26] defined contextual outlier detection as small group of points that share similarity on some attributes but largely deviates on other attributes. Contextual outlier detection has also been explored in other works such as [27,28]. Li et al. [27] assigned feature weights for each group outlier, that were used to compute chain rule entropy and determine correlation between different feature groups. The use of HPC to deal with contextual outlier detection in high and sparse dimensional space was dealt with by Zhao et al. [28].

Other approaches use clustering strategies on individual outliers to group them into similar clusters, while each cluster is considered as a group of outliers [26,29]. Soleimani et al. [29] proposed a supervised learning approach that groups anomalous patterns when memberships are previously unknown. The salient features are extracted from an appropriate training set with discrete data inputs. This approach applies to topic document modeling and is able to discover irregular topic mixtures from a collection of documents. Sun et al. [30] proposed an approach to detect abnormal group-based joint medical frauds. The abnormal group problem has been converted to the maximal clique enumeration problem by considering the set of patients as a set of vertices, where each edge indicates that the two connected patients are similar. The similarity between patients can be obtained by computing their identical joint behavior. Given that the maximal clique enumeration is an NP-hard problem, the authors investigated methods to reduce the size of the graph.

We conclude from this literature review that existing group anomaly algorithms proposed in different contexts do not apply to human behavioral data. They focus on finding a group of outliers from the set of candidate groups, and not from the individual outliers. Candidate groups are not always known a priori in human behaviors such as pedestrian trajectories [31–36]. Moreover, existing solutions for trajectory outlier detection [37–39] are only limited to identifying single outliers. In this paper, we propose the first algorithm to detect collective abnormal human behaviors from individual outliers directly. We then demonstrate the application of our solutions on analyzing human behaviors.

3. Collective abnormal human behavior

In order to introduce the collective abnormal human behavior problem, we need a few preliminary definitions. We represent human behavioral data as a trajectory which is a sequence of location points in space. Each single spatial location point is a tuple of two values—the latitude and the longitude of this location.

Definition 3.1 (*Human Behaviors Database*). We define a human behavior database $\mathcal{HB} = \{\mathcal{HB}_1, \mathcal{HB}_2 ... \mathcal{HB}_m\}$, where each element $\mathcal{HB}_i$ is a sequence of spatial location points $(p_{i1}, p_{i2}...p_{in})$. Each point is represented by the latitude, and the longitude values, respectively.

As common in the literature [40], the location points which are similar enough are aggregated into regions. Let us denote R as a location region in space.

Definition 3.2 (*Mapped Human Behaviors Database*). We define a mapped human behavior database $\Lambda = \{\Lambda_1, \Lambda_2...\Lambda_m\}$, where each mapped human behavior Λ_i is a sequence of spatial location regions $(R_{i1}, R_{i2}...R_{in})$, obtained by mapping each point in $\mathcal{HB}_i$ to the closest region R_i. We note $R = \{R_1, R_2...R_{|R|}\}$, by the set of all regions.

We define the dissimilarity between any two human behaviors data as the distance between them.

Definition 3.3 (*Human Behaviors Dissimilarity*). We define the distance between two human behaviors data $d(\Lambda_i, \Lambda_j)$ by the number of all regions minus the number of shared regions between the two human behaviors data Λ_i, and Λ_j, as

$$d(\Lambda_i, \Lambda_j) = n - |\{(R_{il}, R_{jl})|R_{il} = R_{jl}, \forall l \in [1..n]\}| \tag{1}$$

We define the collective abnormal human behaviors candidates, i.e., the set of potential human behavioral data belong to the collective abnormal human behaviors. This human behavioral data is retrieved from the individual abnormal human behavioral data.

Definition 3.4 (*Abnormal Human Behaviors Candidate*). We define an abnormal human behaviors candidate $\mathcal{G}$ by the set of individual abnormal human behaviors retrieved from the set of individual abnormal human behaviors I, i.e.,

$$\mathcal{G} = \{\Lambda_i | \Lambda_i \in I\} \tag{2}$$

The density of collective human behavioral data is an important concept in our analysis. Intuitively, it is defined as the ratio between the number of human behavioral data in the collection and the number of shared regions among the human behavioral data of such a collection.

Definition 3.5 (*Density of Collective Human Behaviors Data*). We define the density of the candidate collection of human behaviors data $\mathcal{G}$ as

$$Density(\mathcal{G}) = \frac{|\mathcal{G}|}{|\{R_j | \Lambda_i \in \mathcal{G}, R_j \in \Lambda_i\}|} \tag{3}$$

To normalize the density function, we divide the result by the density of the collection having maximum density value, this ensures that we can obtain values ranging from 0 to 1. We call this function *NormalizedDensity*.

Now, we are ready to formally define the collective abnormal human behaviors problem.

Definition 3.6 (*Collective Abnormal of Human Behaviors Problem*). The collective abnormal human behavior Problem (CaHB) aims to discover from the set of all mapped human behavioral data, the set of all collections of abnormal human behavioral data, where each collective abnormal human behavior, $\mathcal{G}$, is determined as follows,

$$\mathcal{G} = \begin{cases} \mathcal{G} \subseteq I \\ NormalizedDensity(\mathcal{G}) \geq \gamma \end{cases} \tag{4}$$

Note that γ is the density threshold varied from $[0...1]$.

A trivial approach for solving the collective abnormal human behaviors problem considers all possible combinations between the mapped human behavioral data, and evaluates each subset separately using Definition 3.5. This method requires high computational cost as well as memory resources, to evaluate the candidate sets and save the potential abnormal human behaviors. The theoretical complexity of this approach is $\mathcal{O}(2^{|\Lambda|})$. To address this issue, we propose in the next section, an alternative framework to improve the collective abnormal human behaviors process.

4. CaHB solutions

Let us begin by describing the key element of our data mining and deep learning based solutions to the CaHB problem (see Fig. 1 for an

overview). Generally speaking, our approach builds upon data mining and deep learning. In particular, we use data mining techniques represented by clustering, feature selection, and neighborhood computation. We can boost the performance of such an approach using ensemble learning as well as GPU computing. The deep learning solution is represented by the convolution neural network, where an object detection is adopted to study the collective abnormal of human behaviors data. As shown in Fig. 1, our approach can be divided into distinct parts.

1. **Mapping (Pre-processing):** Typically, human behavioral data in most applications consists of noisy GPS data points where errors can exceed several meters. This can negatively influence the final output of many algorithms. Hence, first we should use a map-matching step to project GPS data points of each of the human behaviors onto a road network. In this paper, we use the probabilistic model based on a Hidden Markov Model [41]. In particular, we represent each road segment as a hidden state in the Markov chain—it has an emission probability which is the likelihood of observing the GPS point conditional to the candidate road segment being a true match. A higher probability to a road segment is assigned if the observed human behaviors points are close to it. The maximum likelihood path over the Markov chain that has highest probability is then determined, and the corresponding road segment is associated to the observed human behaviors point. In this way, we create the mapped human behaviors database, where each observed human behavior is assigned to the associated road segment.

2. **Processing:** After constructing the mapped trajectory database, a processing step is performed to find the collective abnormal human behaviors. In this context, we propose two ways to find the collective abnormal human behaviors, where we investigate data mining and deep learning solutions. Furthermore, we present various ways for improving the performance of our techniques by incorporating ensemble learning and high performance computing.

5. Data mining-based solutions

In the remainder of this section, we describe the details of the data mining based algorithms (Sections 5.1–5.3), ensemble learning (Section 5.4), and the high performance computing framework (Section 5.5).

5.1. The DBSCAN-CaHB algorithm

To present the adaptation of the DBSCAN algorithm, we need the following definitions:

Definition 5.1 (*Human Behavior Data Neighborhoods*). We define the neighborhoods of human behavioral data Λ_i, $\mathcal{N}_{\Lambda_i}$, for a given threshold ϵ by

$$\mathcal{N}_{\Lambda_i} = \{\Lambda_j | d(\Lambda_i \Lambda_j) \geq \epsilon \vee j \neq i\}. \tag{5}$$

Definition 5.2 (*Core Human Behavior Data*). A human behavioral data Λ_i is defined as a core human behavioral data if there is at least a minimum number of human behavioral data MinPts such that $|\mathcal{N}_{\Lambda_i}| \leq MinPts$.

Definition 5.3 (*Micro Cluster*). A cluster of human behavioral data C_i is defined as a micro cluster if and only if $0 < |C_i| \leq \mu$, where μ is a user threshold.

In general, solutions to sequence and human behavioral data clustering [42–44] are able to derive clusters with different densities. However, these algorithms do not explore the *micro clusters* property for anomaly detection. This section presents our approach for identifying collective abnormal human behaviors, *DBSCAN-CaHB*, that
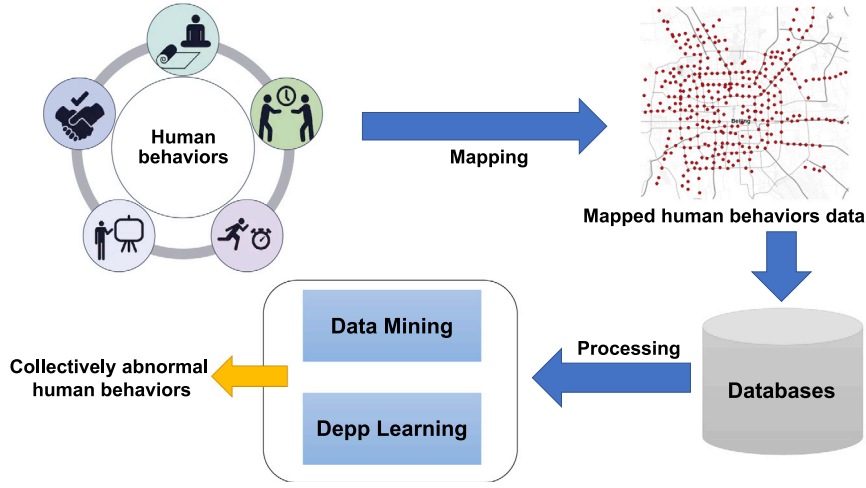
Fig. 1. CaHB solutions.

uses the DBSCAN algorithm to search for clusters by checking the ϵ-neighborhood of each human behavioral data (See Definition 5.1). The core trajectories are determined using Definition 5.2. DBSCAN-CaHB then iteratively collects density-reachable human behavioral data from these core human behaviors directly, which may involve merging a few density-reachable clusters. The process terminates when no new human behavioral data can be added to any cluster. Initially, the set of human behaviors are grouped (as in DBSCAN). This generates several clusters with different sizes. Each micro cluster (See Definition 5.3) is considered as group candidates. For each group, the density of each group is determined using Definition 3.5, if the density exceeds γ threshold, then the group is selected as abnormal.

5.2. The kNN-CaHB algorithm

Let us begin the presentation of our adaptation of the kNN Algorithm, with the following definition:

Definition 5.4 (*kNN*). We define kNN of a human behaviors data Λ_i, denoted by kNN(Λ_i), as

$$kNN(\Lambda_i) = \{\Lambda_j \in \Lambda \setminus \{\Lambda_i\} | d(\Lambda_i, \Lambda_j) \le k_{dist}(\Lambda_i)\}, \quad (6)$$

where $k_{dist}(\Lambda_i) = d(\Lambda_i, \Lambda_l)$ is the k-distance human behaviors defined by the set of k human behaviors $\Lambda' \in \Lambda$, such that $d(\Lambda_i, \Lambda_l) \ge d(\Lambda_i, \Lambda')$.

The following proposition holds:

Proposition 5.1. *Let us consider two human behavioral data Λ' and Λ''. Let $\mathcal{G}^*(t)$ be a collective abnormal of human behaviors at the iteration t such that:*

$$\Lambda' \in \bigcup_{\Lambda_i^* \in \mathcal{G}^*(t)} kNN(\Lambda_i^*) \vee \Lambda'' \notin \bigcup_{\Lambda_i^* \in \mathcal{G}^*(t)} kNN(\Lambda_i^*).$$

Then, the following holds:

$$\Lambda' \notin \mathcal{G}^*(t+1) \Rightarrow \Lambda'' \notin \mathcal{G}^*(t+1).$$

Proof. We have that:

$$\Lambda' \in \bigcup_{\Lambda_i^* \in \mathcal{G}^*(t)} kNN(\Lambda_i^*) \vee \Lambda'' \notin \bigcup_{\Lambda_i^* \in \mathcal{G}^*(t)} kNN(\Lambda_i^*)$$

$$\Rightarrow Density(\mathcal{G}^*(t) \cup \{\lambda''\}) \le Density(\mathcal{G}^*(t) \cup \{\lambda'\}) \dots \quad (7)$$

$$\Lambda' \notin \mathcal{G}^*(t+1) \Rightarrow Density(\mathcal{G}^*(t) \cup \{\lambda'\}) \le \gamma \dots \quad (8)$$

From (7) and (8) we have: $Density(\mathcal{G}^*(t) \cup \{\lambda''\}) \le \gamma \Rightarrow \Lambda'' \notin \mathcal{G}^*(t+1)$

It follows from the above proposition that if a human behavioral data Λ_i belongs to the k-nearest neighbors of at least one human behavioral data in the current collective abnormal human behaviors, and Λ_i is not in the collective abnormal human behaviors of the next iteration, then, any human behavioral data that belongs to the k-nearest neighbors of Λ_i will not be in the collective abnormal human behaviors of the next iteration. Consequently, it is judicious to prune the search into k-nearest neighbors of the individual human behavioral data. In particular, it considers that as input the set of the first p individual human behavioral data $\mathcal{G}^+ = \{\Lambda_1^+, \Lambda_2^+ \cdots \Lambda_p^+\}$, ranked according to the kNN value, i.e., $\forall i \ge j, k$NN$(\Lambda_i^+) \ge k$NN(Λ_j^+). The process aims to enumerate the sets of collective abnormal human behaviors, $\mathcal{G}^*$, by exploring a search tree of $\mathcal{G}^+$. It starts by adding the individual abnormal human behaviors ranked first, Λ_1^+, to the collective abnormal human behaviors, denoted by $\mathcal{G}_1^*$. It then generates all potential candidates from Λ_1^+. A human behavioral data t is a potential candidate from Λ_1^+, if and only if, $t \in \mathcal{G}^+ \vee t \in kNN(\Lambda_1^+)$. The density of $\mathcal{G}_1^*$ is updated by adding the potential candidates to $\mathcal{G}_1^*$, one by one. Only the potential candidates respecting the density threshold are saved, and the remaining ones are removed. Once the potential candidate is added to $\mathcal{G}_1^*$, it is removed from $\mathcal{G}^+$. If $\mathcal{G}_1^*$ contains less than two elements, it is removed from $\mathcal{G}^*$. The same process is recursively applied to all potential candidates added to $\mathcal{G}_1^*$, and the overall process is repeated for all abnormal human behaviors in $\mathcal{G}^+$.

5.3. FS-CaHB

We begin by defining the transformation of the CaHB problem to the feature selection problem.

Definition 5.5 (*Transformation to FS Problem*). Consider CaHB problem $\left\langle \mathcal{G}_{\mathcal{A}}^+, \mathcal{G}^* \right\rangle$. We transform it to the feature selection problem, represented by the set of all features F and the subset of selected features F^*, as follows: $F = \mathcal{G}_{\mathcal{A}}^+$, and $F^* = \mathcal{G}^*$. We evaluate F^* as follows:

$$Eval(F^*) = Quality(F^*) - \frac{|F^*|}{|F|}, \quad (9)$$

where Quality(F^*) is computed as per Definition 3.5.

We consider each individual abnormal human behaviors as one feature, and our aim is to select the most relevant features from the set of all features. This set then becomes the collective abnormal human behaviors (see Definition 5.5). The evaluation of the selected set of features (human behavioral data) is computed using the group

density measure, see Eq. (9). The process starts by applying the feature selection algorithm on the set of individual abnormal human behaviors, The output of this step is a ranking of individual anomalies in the descending order in terms of score feature relevance. A ranking vector is created, where a search enumeration tree is generated in the breadth-first-search (BFS) manner. If the quality of the current group candidate does not reach the criteria from Definition 3.5, a backtracking procedure is launched by taking the next human behavioral data in the ranking vector. While exploring the enumeration tree of individual anomalies, the aim is to maximize the function reported in Eq. (9).

5.4. Ensemble learning

Each of the previously proposed techniques – clustering (DBSCAN-CaHB), neighborhood computation (KNN-CaHB) and feature selection (FS-CaHB) – returns potential collective abnormal human behaviors. Out of these collections, some collections are usually good while others may not be useful. In order to improve the accuracy of the detected collections, we propose the use of ensemble learning [45]. In particular, we first launch our algorithms, i.e., DBSCAN-CaHB, kNN-CaHB, and FS-CaHB, in this context called learners. The three outputs – sets G^{DBSCAN}, G^{kNN}, and G^{FS} – are then merged to derive the final collections. Hence, the main challenge of this approach is to find an efficient merging strategy. We propose the following. For each collection, we determine the number of the occurrences of the three learners, where we look for the collections that are highly frequent. For instance, if there are two collections: the first collection $\{\Lambda_1, \Lambda_2, \Lambda_4\}$ appears two times – ones in the output of the G^{DBSCAN} algorithm, and another time in the output of the G^{kNN} algorithm – and the second collection $\{\Lambda_1, \Lambda_2, \Lambda_3\}$ appears only one time in the output of G^{FS}, then the first collection is better ranked than the second one.

5.5. High performance computing

The map is first partitioned on k grids $\{g_1, g_2 \ldots g_k\}$ using the map partitioning step. The set of designed grids are then sent to the GPU. Each block of threads is mapped onto one grid, where the CaHB solutions are applied on each block in parallel. If we consider the size of the shared memory of each block is sm, the first sm human behavioral data of the grid g_i are allocated to the shared memory of the block, and the remaining human behavioral data of the grid g_i is allocated to the global memory of the GPU host. It defines a *local table*, $table_i$, to store the collections of abnormal human behaviors data of the grid g_i. The local table of each grid is sent to CPU for further processing. In this context, CPU host performs a merging step to find the global collections, where the union of all sets of collections in the local tables is computed. From a theoretical standpoint, high performance computing improves the CaHB solutions by exploiting the massively threaded computing of GPUs while mining the grids of human behavioral data. It also minimizes the CPU/GPU communication, by defining only two points of CPU/GPU communication. The first one takes place when the grids are loaded into the GPU host, and the second one when the local tables are returned to the CPU. It also provides an efficient memory management by using different levels of memories including global and shared memories. However, it may suffer from synchronization between the GPU blocks. This takes place when the GPU blocks process grids with different number of human behaviors data. This issue degrades the performance of the GPU-based implementation of the CaHB solutions. In real scenarios, different number of human behavioral data per grid may be obtained, this depends to the way of the human behavioral data are placed into the map, as the size of the grids are different, as the synchronization cost of the GPU-based implementation will be high. All these statements will be clearly explained in the performance evaluation section.

6. Deep learning-based solutions

This section presents the proposed CNN-CaHB framework, which use a convolutional neural network (CNN) for identifying collective abnormal human behaviors. CNN-CaHB consists of two stages:

1. **Data Collection**: The aim of this step is to collect human behavioral data and build the images database. A visual strategy is used, in which each image contains a set of human behavior trajectories. The process starts by recording video frames from cameras. The frames are then transformed to images, and different distortion techniques such as mapping, re-sizing are used to correct the images. The whole trajectory images are stored into the database for training.

2. **Training**: The convolution neural network is applied to design a training model, which it is considered as a powerful vision machine to learn from the different features of the image trajectory data. The collective abnormal human behaviors is transformed to the object detection problem by setting the input of the object detection model to the trajectories image database, and the output of the object detection model to the collective abnormal human behaviors in each image. We use the regional convolution neural network [46,47]. For every image in the input, regions of interests are determined and passed to the hidden layer where the Relu activation function is performed and a similar process to the convolution neural network is followed. In fast RCNN, the model performs better and quicker as the regions of interest are found using a selective search method, and all the regions of interests of an image are found at once. This is different from CNN that finds ROI (regions of interest) and applies Relu on each ROI separately which is slower. The process is repeated for a given number of epochs, or until the training stop providing improvement for a given number of iterations. The weight initialization is done using the pre-trained ImageNet model.[1] The trained model is stored on the central workstation.

3. **Inference**: The aim here is to derive the collective abnormal human behaviors of the input image using the trained model of the previous step. Therefore, a propagation of the different weights of the trained model is performed to detect the objects of the image. The detected objects are considered as the collective abnormal human behaviors. In this step, different kinds of inference are generated. We send the trained model to the computers and infer the model for each new trajectories image data. We can also use smartphones which support Android, and GPU computing to infer the collective abnormal of human behaviors in real time processing. In this context, several technologies could be integrated such as TensorflowLite.[2]

In practice, the training images have a very high resolution in today's IoT world, from 5000 pixels to 100,000 pixels. As a result, millions to billions of region proposals have been generated. This makes the whole system very hungry in time-processing and memory. In some cases, the system will be bluntly blocked after several days and weeks of processing. To deal with this problem, we propose a strategy to prune and filter the number of bounding boxes. Two collections in the same frame should not be close to each other, and thus two bounding boxes in the same image should not be close to each other. We then compute the similarity between each new bounding box generated and the bounding boxes that have already been generated. The similarity between two bounding boxes is determined by the number of pixels that separate these bounding boxes. We only keep the bounding boxes, which gives high diversity of the image. That is, the minimal set of bounding boxes which cover the maximum number of pixels in the image.

---

[1] http://www.image-net.org/.
[2] https://www.tensorflow.org/lite.

7. Performance evaluation

The CaHB framework and its different components are evaluated in this section. The parameters of both data mining and deep learning based CaHB solutions are first tuned, and then the scalability performance of the best implementation is investigated and compared with the state-of-the-art solutions.

7.1. Experimental setup

The implementation of the different components of the CaHB framework.[3] has been integrated on the SPMF data mining library [48] The experimental evaluation has been performed on a computer with 64−bit core $i7$ processor running Windows 10 and 16 GB of RAM, coupled with a GPU device. The GPU device is an NVIDIA Tesla C2075 with 448 CUDA cores (14 multiprocessors with 32 cores each) and a clock speed of 1.15 GHz. It has 2.8 GB of global memory, 49.15 kB of shared memory, and a warp size of 32. Both the CPU and GPU are used in single precision. We used HUMBI dataset, which is a large corpus of high fidelity models of behavioral signals in $3D$ from a diverse population measured by a massive multi-camera system [49]. The dataset contains human behavior from 164 subjects across gender, ethnicity, age, and physical condition at a public venue. High models of five elementary parts are designed including: gaze, face, hands, body, and cloth. As a byproduct, the 3D model provides geometrically consistent image annotation via $2D$ projection, e.g., body part segmentation. In general, a common problem of anomaly detection techniques lies in the evaluation step, typically in emerging applications such as CaHB where the ground truth is typically unknown. To facilitate a quantitative evaluation, for CaHB detection techniques, we inject synthetic collective abnormal of human behaviors data. The evaluation is performed using runtime and ROCAUC, which are the common measures for the evaluation of anomaly detection methods.

7.2. Parameters setting

The first part of the experiment analysis focuses on tuning the parameters of different proposed CaHB solutions. Several tests have been performed by varying the user threshold from 1 to 10 for DBSCAN-CaHB, the number of neighborhood from 1 to 10 for kNN-CaHB, the tree depth from 1 to 10 for FS-CaHB, the number of epochs from 100 to 1000 for CNN-CaHB. For all the number of human behaviors data used as input, the accuracy of DBSCAN-CaHB and kNN-CaHB increases with the increase of the corresponding parameter values up to reaching the optimal point, and then starts decreasing. In case of FS-CaHB, and CNN-CaHB, the accuracy increases with values of the parameter and stabilizes when reaching the optimal point. The best parameter values obtained in this step are used in the remaining of the experiments. The user threshold is set to 4 for DBSCAN-CaHB, k is set to 6 for kNN-CaHB, tree depth is set to 6 for FS-CaHB, and epochs is set 700 for CNN-CaHB.

7.3. Comparison with state-of-the-art solutions

The aim of this experiment is to compare the proposed solutions with state-of-the art algorithms in terms of accuracy and processing time. Since this is the first work that explores collective abnormal of human behaviors (to the best of our knowledge), we compare the proposed solutions with general collective abnormal of human behaviors based solutions (see Section 2). For this purpose, we adapted two algorithms, i.e., Das et al. [50], and Sun et al. [30] and used them as base-line. Figs. 2 and 3 present runtime, and accuracy respectively, of the proposed solutions in comparison with baseline algorithms.

Fig. 2 presents the runtime of the proposed solutions against the baseline algorithms. Each plot correspond to a specified percentage of regions. Overall, we consider 1000 regions, the first plot represents results of exploring 100 regions, the second plot represents results exploring 200 regions, and so on until the last plot where all the regions are explored. At each pass of the experiment, the percentage of human behaviors data, represented by x-axis have been varied from 10% to 100%. The y-axis represents the runtime in seconds for identifying the collective abnormal of human behaviors data. The results show that the runtime of our GPU-based, as well as, the convolution neural network solutions outperform the baseline algorithms. However solution based on neighborhood computation required huge computational resources, but still outperform the baseline algorithms. For instance, for mining 1000 regions, and the whole human behaviors data, the CNN-CaHB, and the GPU-CaHB need 42 s, and 81 s, respectively, however, other solutions exceeds 110 s for mining the same number of regions, and the same number of human behaviors data. This is explained by the fact that the convolution neural network only needs time to train the data and find the weighted of the model used in the inference step. Further, the GPU-based approach benefits from the massively threaded and the efficient mapping between the human behaviors data and the GPU blocks. On the other hand, solutions based on neighborhood computation and density estimation have high cost for determining the similarity between each new human behaviors data and the whole human behaviors database.

Fig. 3 presents the accuracy represented by the AUC value of the proposed solutions against the baseline algorithms. Each plot correspond to a specified percentage of regions, in overall, we consider 1000 regions, the first plot represents results of exploring 100 regions, the second plot represents results exploring 200 regions, and so on until the last plot where all the regions are explored. At each pass of the experiment, the percentage of human behaviors data, represented by x-axis have been varied from 10% to 100%. The y-axis represents the AUC value for identifying the collective abnormal of human behaviors data. The results show that the AUC value of our convolution neural based approach outperform the baseline algorithms. However solution based on neighborhood computation, ensemble learning, and GPU computing are less accurate, and converges to the baseline algorithms. For instance, for mining 1000 regions, and the whole human behaviors data, the accuracy of CNN-CaHB is 85%, however, other solutions does not reach 75%, except Sun et al. algorithm, where its accuracy is 81%. This comes from the fact that the convolution neural network used more advanced techniques in identifying the collective abnormal of human behaviors data such as the convolution, the pooling operators, intensive number of layers, computing global image features, while the other solutions use very basic concepts in finding anomalies.

8. Future work

As far as future perspectives are concerned, we plan to explore other representation of human behavioral data such as time series, events log, and tweets. These representations will help identifying collective abnormalities. This research area needs to investigate other data mining algorithms and other deep learning architectures for maximizing its potential. For instance, the recurrent neural network is more suitable to deal with time series and textual data. Hybrid architectures based on both convolution and recurrent neural networks are needed to study the correlation of human behavioral data gathered from heterogeneous sources. Solutions to human behavioral analysis are very time consuming in particular when dealing with large scale data. Therefore, technologies from different domains could be adapted such as:

1. HPC aims at using parallel frameworks to speed up the sequential solutions. Some of the most well-known architectures apply the multi-core CPU or GPU and perform on MapReduce or Spark platforms
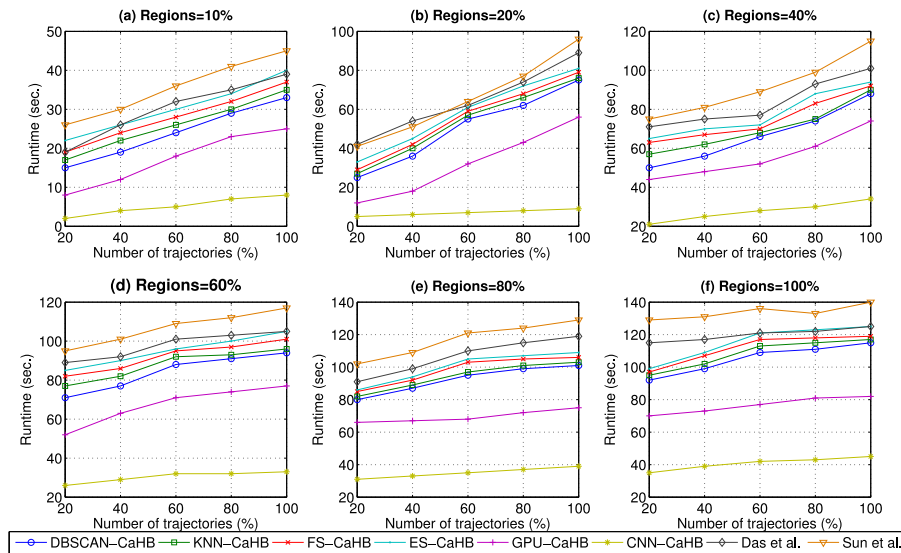
---

[3] https://github.com/YousIA/GTOD.

Fig. 2. Runtime of the proposed solutions and the state-of-the-art algorithms.
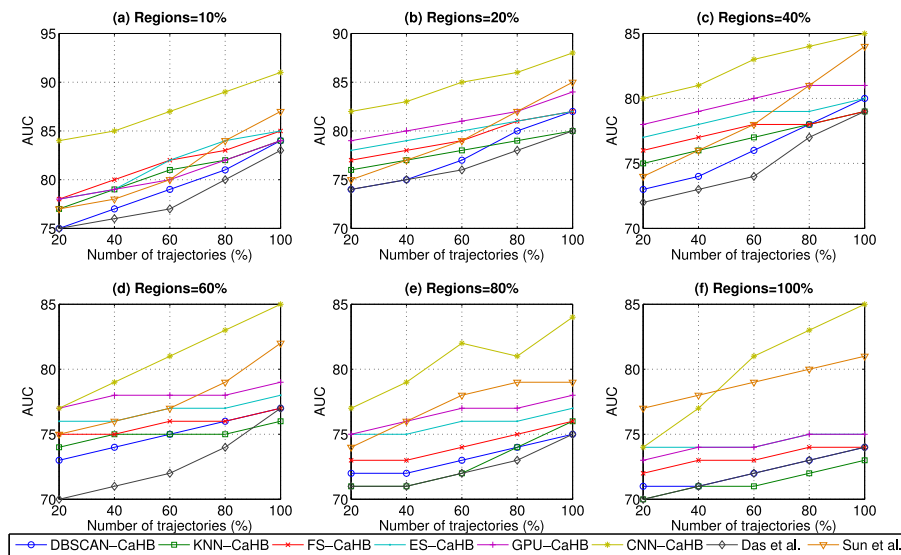

Fig. 3. Accuracy of the proposed solutions and the state-of-the-art algorithms.

2. Computational intelligence is a collection of intelligent methods aiming at optimizing complex problems with strategies like meta-heuristics

3. Database systems provide techniques to efficiently store, update, and search human behavior data, such as query optimization and index optimization.

Adapting, combining, and optimizing technologies in human behavioral analysis may also provide many open research questions and future directions. Furthermore, in this research, two measures (runtime, ROCAUC) have been considered in the evaluation step. Evaluating the proposed solutions on other measures such as F-measure, F1-score, and computing the number of grouped human behavioral data is also in our future agenda.

9. Conclusion

A new model for identifying abnormal human behavior from trajectory data has been introduced in this paper, where data mining and deep learning algorithms have been proposed to tackle the highly

important problem. Data mining techniques explore the correlation, the neighborhood computation, and the density estimation among the input. We proposed adaptations of DBSCAN, kNN, feature selection, and ensemble learning, by considering group and collective behavior of people to identify abnormalities. We also boosted the runtime performance of these solutions by using HPC approaches and benefiting from the massively threaded of the GPU architecture. We also proposed a deep learning approach based on the convolution neural network. The DL approach aims to learn from both training as well as historical trajectory data, and be able to identify the abnormal human behavior. The work has been validated through intensive experimentation analysis of both the data mining and deep learning solutions. The results confirm that the deep learning approach outperforms the data mining approaches, as well as the state-of-the-art solutions in terms of runtime and accuracy.

CRediT authorship contribution statement

Asma Belhadi: Conceptualization, Data curation, Formal analysis. **Youcef Djenouri:** Funding acquisition, Investigation. **Gautam**

References

[1] R. Khatoun, S. Zeadally, Smart cities: Concepts, architectures, research opportunities, Commun. ACM 59 (8) (2016) 46–57.

[2] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.

[3] S. Otoum, B. Kantarci, H.T. Mouftah, On the feasibility of deep learning in sensor network intrusion detection, IEEE Netw. Lett. 1 (2) (2019) 68–71.

[4] J. Bütepage, H. Kjellström, D. Kragic, A probabilistic semi-supervised approach to multi-task human activity modeling, 2018, arXiv preprint arXiv:1809.08875.

[5] A.L. Alfeo, M.G. Cimino, B. Lepri, A.S. Pentland, G. Vaglini, Assessing refugees' integration via spatio-temporal similarities of mobility and calling behaviors, IEEE Trans. Comput. Soc. Syst. 6 (4) (2019) 726–738.

[6] B. Ivanovic, M. Pavone, The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 2375–2384.

[7] D. Jiang, The construction of smart city information system based on the internet of things and cloud computing, Comput. Commun. 150 (2020) 158–166.

[8] D. Djenouri, R. Laidi, Y. Djenouri, I. Balasingham, Machine learning for smart building applications: Review and taxonomy, ACM Comput. Surv. 52 (2) (2019) 1–36.

[9] G. Pan, G. Qi, W. Zhang, S. Li, Z. Wu, L.T. Yang, Trace analysis and mining for smart cities: issues, methods, and applications, IEEE Commun. Mag. 51 (6) (2013) 120–126.

[10] S.Y. Tan, A. Taeihagh, Smart city governance in developing countries: A systematic literature review, Sustainability 12 (3) (2020) 899.

[11] Y. Djenouri, A. Belhadi, J.C.-W. Lin, A. Cano, Adapted k-nearest neighbors for detecting anomalies on spatio–temporal traffic flow, IEEE Access 7 (2019) 10015–10027.

[12] Y. Djenouri, A. Zimek, M. Chiarandini, Outlier detection in urban traffic flow distributions, in: IEEE International Conference on Data Mining, 2018, pp. 935–940.

[13] V. Balasubramanian, S. Otoum, M. Aloqaily, I. Al Ridhawi, Y. Jararweh, Low-latency vehicular edge: A vehicular infrastructure model for 5g, Simul. Model. Pract. Theory 98 (2020) 101968.

[14] G. Ciaparrone, F.L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, F. Herrera, Deep learning in video multi-object tracking: A survey, Neurocomputing 381 (2020) 61–88.

[15] M.M. Hassan, M.G.R. Alam, M.Z. Uddin, S. Huda, A. Almogren, G. Fortino, Human emotion recognition using deep belief network architecture, Inf. Fusion 51 (2019) 10–18.

[16] J. Gong, Y. Huang, P.I. Chow, K. Fua, M.S. Gerber, B.A. Teachman, L.E. Barnes, Understanding behavioral dynamics of social anxiety among college students through smartphone sensors, Inf. Fusion 49 (2019) 57–68.

[17] N. Lasla, M. Doudou, D. Djenouri, A. Ouadjaout, C. Zizoua, Wireless energy efficient occupancy-monitoring system for smart buildings, Pervasive Mob. Comput. 59 (2019).

[18] R. Laidi, D. Djenouri, UDEPLOY: User-driven learning for occupancy sensors deployment in smart buildings, in: IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops, pp. 209–214, 2018.

[19] A. Chelli, M. Bagaa, D. Djenouri, I. Balasingham, T. Taleb, One-step approach for two-tiered constrained relay node placement in wireless sensor networks, IEEE Wirel. Commun. Lett. 5 (4) (2016) 448–451.

[20] M. Ringel, R. Laidi, D. Djenouri, Multiple benefits through smart home energy management solutions-a simulation-based case study of a single- family house in Algeria and Germany, Energies 12 (2019).

[21] D. Bales, P.A. Tarazaga, M. Kasarda, D. Batra, A.G. Woolard, J.D. Poston, V.V.N.S. Malladi, Gender classification of walkers via underfloor accelerometer measurements, IEEE Internet Things J. 3 (6) (2016) 1259–1266.

[22] H.M.S. Hossain, M.A.A.H. Khan, N. Roy, Active learning enabled activity recognition, Pervasive Mob. Comput. 38 (2017) 312–330.

[23] Y. Chiang, C. Lu, J.Y. Hsu, A feature-based knowledge transfer framework for cross-environment activity recognition toward smart home applications, IEEE Trans. Hum.-Mach. Syst. 47 (3) (2017) 310–322.

[24] E. Toth, S. Chawla, Group deviation detection methods: A survey, ACM Comput. Surv. 51 (4) (2018) 77.

[25] R. Chalapathy, E. Toth, S. Chawla, Group anomaly detection using deep generative models, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2018, pp. 173–189.

[26] G. Tang, J. Pei, J. Bailey, G. Dong, Mining multidimensional contextual outliers from categorical relational data, Intell. Data Anal. 19 (5) (2015) 1171–1192.

[27] J. Li, J. Zhang, N. Pang, X. Qin, Weighted outlier detection of high-dimensional categorical data using feature grouping, IEEE Trans. Syst. Man Cybern.: Syst. (99) (2018) 1–14.

[28] X. Zhao, J. Zhang, X. Qin, J. Cai, Y. Ma, Parallel mining of contextual outlier using sparse subspace, Expert Syst. Appl. 126 (2019) 158–170.

[29] H. Soleimani, D.J. Miller, Atd: Anomalous topic discovery in high dimensional discrete data, IEEE Trans. Knowl. Data Eng. 28 (9) (2016) 2267–2280.

[30] C. Sun, Z. Yan, Q. Li, Y. Zheng, X. Lu, L. Cui, Abnormal group-based joint medical fraud detection, IEEE Access 7 (2019) 13589–13596.

[31] L. Zhang, F. Yang, Y.D. Zhang, Y.J. Zhu, Road crack detection using deep convolutional neural network, in: IEEE International Conference on Image Processing, 2016, pp. 3708–3712.

[32] S. Lu, Z. Lu, Y.D. Zhang, Pathological brain detection based on alexnet and transfer learning, J. Comput. Sci. 30 (2019) 41–47.

[33] S. Zafar, S. Jangsher, O. Bouachir, M. Aloqaily, J.B. Othman, Qos enhancement with deep learning-based interference prediction in mobile iot, Comput. Commun. 148 (2019) 86–97.

[34] M. Aloqaily, S. Otoum, I. Al Ridhawi, Y. Jararweh, An intrusion detection system for connected vehicles in smart cities, Ad Hoc Netw. 90 (2019) 101842.

[35] S. Kumar, A. Pandey, K.S.R. Satwik, S. Kumar, S.K. Singh, A.K. Singh, A. Mohan, Deep learning framework for recognition of cattle using muzzle point image pattern, Measurement 116 (2018) 1–17.

[36] S. Dey, G. Kalliatakis, S. Saha, A.K. Singh, S. Ehsan, K. McDonald-Maier, Mat-cnn-sopc: Motionless analysis of traffic using convolutional neural networks on system-on-a-programmable-chip, in: The NASA/ESA Conference on Adaptive Hardware and Systems, 2018, pp. 291–298.

[37] A. Belhadi, Y. Djenouri, J.C.W. Lin, Comparative study on trajectory outlier detection algorithms, in: The International Conference on Data Mining Workshops, 2019, pp. 415–423.

[38] Y. Djenouri, A. Zimek, Outlier detection in urban traffic data, in: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, 2018, pp. 1–12.

[39] P. Chao, W. Hua, R. Mao, J. Xu, X. Zhou, A survey and quantitative study on map inference algorithms from gps trajectories, IEEE Trans. Knowl. Data Eng. (2020).

[40] Y. Djenouri, A. Belhadi, J.C.-W. Lin, D. Djenouri, A. Cano, A survey on urban traffic anomalies detection algorithms, IEEE Access 7 (2019) 12192–12205.

[41] A. Prokhorchuk, J. Dauwels, P. Jaillet, Estimating travel time distributions by Bayesian network inference, IEEE Trans. Intell. Transp. Syst. 21 (5) (2019) 1867–1876.

[42] S. Zhou, E. Zhu, X. Liu, T. Zheng, Q. Liu, J. Xia, J. Yin, Subspace segmentation-based robust multiple kernel clustering, Inf. Fusion 53 (2020) 145–154.

[43] Z. Hu, F. Nie, R. Wang, X. Li, Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding, Inf. Fusion 55 (2020) 251–259.

[44] H. Li, J. Liu, K. Wu, Z. Yang, R.W. Liu, N. Xiong, Spatio-temporal vessel trajectory clustering based on data mapping and density, IEEE Access 6 (2018) 58939–58954.

[45] J.N. van Rijn, G. Holmes, B. Pfahringer, J. Vanschoren, The online performance estimation framework: heterogeneous ensemble learning for data streams, Mach. Learn. 107 (1) (2018) 149–176.

[46] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

[47] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, 2015, pp. 91–99.

[48] P. Fournier-Viger, A. Gomariz, T. Gueniche, A. Soltani, C.-W. Wu, V.S. Tseng, Spmf: a java open-source pattern mining library, J. Mach. Learn. Res. 15 (1) (2014) 3389–3393.

[49] Z. Yu, J.S. Yoon, P. Venkatesh, J. Park, J. Yu, H.S. Park, Humbi 1.0: Human multiview behavioral imaging dataset, 2018, arXiv preprint arXiv:1812.00281.

[50] M. Das, S. Parthasarathy, Anomaly detection and spatio-temporal analysis of global climate system, in: Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data, 2009, pp. 142–150.