Inferring Transportation Mode and Human Activity from Mobile Sensing in Daily Life

Jonathan Liono
Computer Science and Information Technology
School of Science
RMIT University
Melbourne, Victoria, Australia
jonathan.liono@rmit.edu.au

Zahraa S. Abdallah*
School of Computing and Digital Technology
Birmingham City University
Birmingham, United Kingdom
zahraa.s.abdallah@gmail.com

A. K. Qin
Department of Computer Science and Software
Engineering
School of Software and Electrical Engineering
Swinburne University of Technology
Melbourne, Australia
kqin@swinburne.edu.au

Flora D. Salim
Computer Science and Information Technology
School of Science
RMIT University
Melbourne, Victoria, Australia
flora.salim@rmit.edu.au

ABSTRACT

In this paper, we focus on simultaneous inference of transportation modes and human activities in daily life via modelling and inference from multivariate time series data, which are streamed from off-the-shelf mobile sensors (e.g. embedded in smartphones) in real-world dynamic environments. The transportation mode will be inferred from the structured hierarchical contexts associated with human activities. Through our mobile context recognition system, an accurate and robust solution can be obtained to infer transportation mode, human activity and their associated contexts (e.g. whether the user is in moving or stationary environment) simultaneously. There are many challenges in analysing and modelling human mobility patterns within urban areas due to the ever-changing environments of mobile users. For instance, a user could stay at a particular location and then travel to various destinations depending on the tasks they carry within a day. Consequently, there is a need to reduce the reliance on location-based sensors (e.g. GPS), since they consume a significant amount of energy on smart devices, for the purpose of intelligent mobile sensing (i.e. automatic inference of transportation mode, human activity and associated contexts). Nevertheless, our system is capable of outperforming the simplistic approach that only considers independent classifications of multiple context label sets on data streamed from low-energy sensors.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computer systems organization** → **Embedded and cyber-physical systems**;

*Also with Information Technology, Monash University, Australia.

KEYWORDS

ubiquitous computing, transportation mode, human activity recognition, context modelling

1 INTRODUCTION

Rapid development and population growth in urban areas have gained tremendous attention from researchers to study human dynamics via mobile sensing technology. In order to close the gaps towards a smarter city, it is also important for the local authorities to discover the relevant knowledge from the observation of human mobility patterns within various transportation contexts. The human movements within the cities are becoming more dynamic and complex due to the high availability of the public transportation services and dynamic environmental factors, especially with the involvement of surrounding spatial and temporal elements that could affect the accuracy in mobile sensing applications.

In daily commuting routines, smartphone users travel between urban areas via certain transportation modes. The dynamic mobility of the humans in each journey involves activities across multiple transportation modes. Hence, the mode of transportation associated with human activities always changes over the time, depending on user's spatial and temporal contexts. For example, in the morning a mobile user travels from home to the office using a bus, train and light rail while he could be standing or sitting while inside those transportation mediums. Furthermore, during the transition between these transportation mediums, the user might be required to walk for a certain distance in order to get onboard on different transportation mode (e.g. train or light rail). At a certain time slot, the user could be relaxing while sitting on a couch at home, which means having no transportation mode for the given contexts of this mobile user. It should be noted that a mobile user in this paper is

defined as a user that frequently moves (i.e. commute) and utilises mobile devices (not limited to smartphones) in daily life.

In major metropolitan areas, smartphones are heavily used and carried everywhere by individuals in order to perform their daily tasks. These smartphones are embedded with various built-in mobile sensors that are capable of perceiving the user contexts, such as transportation modes, human activities and types of environment the mobile users are contained in, from the signals captured within the vicinity of smart devices. In this case, the smart devices are not only limited to first-person perspective (smartphone users), but also to the ubiquitous devices that could be deployed throughout the city. In this paper, the scope of our study is focused on a non-trivial issue to perform simultaneous inference of human activities, transportation modes and their related contexts from mobile users' perspective for the purpose of intelligent applications. For instance, let us imagine the future intelligent assistant that can automatically understand human commuting behaviour and recommend the best route based on user contexts (including the past behaviour from user's mobility patterns).

In this paper, transportation mode, human activity and their related contexts (e.g. environment type) can be inferred from the data streamed from the ubiquitous sensors embedded in a smartphone. Such independent assumption of modelling human activities and transportation mode may result in a wrong conceptual interpretation of an action, which is unfavourable for applications such as intelligent assistants. For example, it may be less practical for a human to "stand" in a car while the actual activity in such transportation could be "driving" or "sitting". Furthermore, we focus on the simultaneous inference of mobile user contexts (transportation modes, human activities and associated environment contexts) based on non-location based sensors. As a result, a less reliant solution to location-based sensors (e.g. GPS sensor) can be leveraged for low-energy mobile sensing applications in dynamic urban environments. Inherently, this type of solutions will be an energy efficient option as GPS sensors typically consume more energy in most mobile sensing applications.

Hence, the contributions of this paper include:

(1) More accurate inference of transportation mode and human activities without relying on the locations of mobile users.
(2) Simultaneous inference of transportation modes, human activities and mobile user's environmental contexts (e.g. whether the user is in a moving environment or not) based on the dependency of contextual labels of user's annotations.
(3) A robust system for mobile context recognition on data streaming from low-energy ubiquitous sensors.

2 RELATED WORKS

For many years, major works have been performed in order to understand the human behavior and mobility patterns in various interdisciplinary studies. To understand the human behavior in social interactions, previous study in [36] includes empirical analysis and prediction that incorporates the essence of human mobility patterns to network measure such as connectivity in social networks. They derived the mobility patterns from similarities in movements and interactions via trajectories and communication records, which

subsequently be correlated with the social connectedness in order to produce models for social link prediction.

In regards to human activity recognition, Su et al. [32] recently presented a general overview of techniques and challenges in performing activity recognition from the mobile phone sensors. Mainly, the research in activity recognition is often treated as a classification problem (for example: [7, 10, 12, 14, 26]) since there are labels associated to certain human activities for training and testing phases. Other problem in this domain is related to adaptability to perform real-time activity recognition on continuous streaming of sensor data. The challenges are dominant in such live environment is due to concept of variability and evolving user's behavior. These changes can be affected by the presence of concept drift [39]. Therefore, adaptive approaches such as [1, 37] are recommended to perform incremental learning in order to adjust with the evolving sensor streams and user's behaviors.

Furthermore, Lane et al. [15] has produced a survey in regards to the significance of the sensors embedded in mobile phones to be integrated within the space of personal, group and community sensing applications. It is aligned with the fact that mobility patterns can be inferred from these powerful mobile phones through variety of built-in sensors. Especially in transportation domain, they noted that traffic remains as a major problem, which is visible in our society given with growing population and limited services that are able to fully satisfy the transport needs. Therefore, congestion becomes prominent and crucial in affecting urban planning and traffic management. It brings a significant motivation for our study to help extending the horizon of human mobility research. As one of the benefits, the congestion can be reduced to a certain extent with improved quality of service delivery and transport resource allocation. Since our communities have been emerged with the usage of mobile phones in their daily operations, the potential is limitless to extract more knowledge from human behavior and mobility patterns. Many works such as [19, 30] included the tracking of user positioning characteristic in conjunction to activities in order to offer location-based model of mobility patterns.

In the context of human activity recognition based on mobile sensors, segmentation on live streaming data is often required in order to extract the summary for further analysis and prediction tasks. Within the field of activity recognition, relevant solutions such as [8, 20, 35] have been proposed for time segmentation purposes. However, these did not address the problem where each of the sensor can be sampled at different rate, thus, it is feasible to have a sparse segment when size of time window is small. However, this problem was then addressed in the previous work [18] for optimal windowing of multivariate sensor data, especially in the scenario of multi-activity recognition. In a larger scenario of human activities in daily life, multi-context recognition is a prominent problem, especially when different environments have their intrinsic sensing patterns, given the characteristics that can be captured and identified from smart devices. Therefore, the immediate challenges that we face in multi-context recognition in an uncontrolled environment are associated to the noise of streamed sensor data that are subject to different types of user's environments and their activities. Such in-the-wild sensing settings [34] require non-trivial consideration of many factors associated with the userâĂŹs daily life, which increases the magnitude of real challenges and needs

in multi-context recognition exponentially. For example, the users may turn off the location tracking (e.g. GPS sensor) in order to preserve the operational time of smartphone in their daily life (i.e. minimising battery usage as highlighted in [34]). Hence, another immediate challenge is addressed in this paper to design and construct a mobile context recognition system that relies less on the location data of users. Consequently, our hypothesis initially assumed that these multiple contextual label sets are associated with each other, reflected by multi-faceted contexts of human activities in the daily life of these mobile users.

The scope of this paper includes the study of human activity recognition from the mobile sensors that are induced within the context of transport mobility patterns. Therefore, the problems addressed in our study are related to transportation mode discovery and the associated human activities within the temporal domain. The mobility patterns presented are derived from overlapping labels in temporal segmentation of sparse sensor data streams. To the best of our knowledge, there is a lack of study in terms of inferring mobility patterns for recognizing both transportation mode and human activity.

3 PROBLEM DEFINITION

In this section, we present the problem definition based on the applications of smart sensing to infer multiple contextual labels simultaneously from raw activity annotations in daily life. Inherently, these multiple contextual labels are related to mobile user contexts such as transportation modes, human activities and their environmental contexts (i.e. mobility of corresponding transportation modes).

Let $S = \{S_1, S_2, ..., S_d\}$ be the set consisting of d number of mobile sensors of a user and S_i ($1 \leq i \leq d$) is a sensor identifer within the range of d sensor streams. Each arrival of sensor reading instance point I_{ji} is associated to unique timestamp t_j that is naturally continuous in a sensor stream S_i. It should be noted that sensor streams set S correponds to a segment dedicated to a given raw user annotation a. In this case, the raw user annotation refers to the short description (typically in a diary-based study) of a human activity and its related contexts (e.g. transportation mode and the type of environment). Given the temporal set $T = \{t_1, t_2, t_3, t_4, ..., t_m\}$ in a segment, the timestamp $t_j \in T$ corresponds to a particular sensor data instance I_{ji}. Each instance I_{ji} consists of a collection of values corresponding to multiple channels of a given sensor S_i.

Given the feature vectors that can be computed from all segments of human activities in daily life, we formulate the following problem to infer multiple contexts simultaneously from a raw user annotation.

Definition 3.1. Let a be the instance of a set of raw human annotations $Annotations = \{a_1, a_2, a_3, a_4, ..., a_r\}$ in a mobile sensing application.

For a general overview of the inferring mobile user contexts, Figure 1 describes that an instance of user annotation a can be decomposed into multiple contextual labels consisting of human activity A and its related mobile contexts C_{mobile} (including transportation mode).
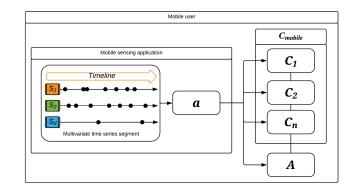

Figure 1: Inferring multiple contexts (human activity and related mobile contexts) on multivariate time series data of mobile sensors

Definition 3.2. Therefore, we can define $a = \{A, C_{mobile}\}$ where A is the instance of human activity and C_{mobile} is the mobile contexts set associated to A. In this case, a transportation mode *trans* is a member of $C_{mobile} = \{C_1, C_2, ..., C_n\}$. Hence, the aim of a mobile sensing application is to infer A and each member in C_{mobile} simultaneously, assuming there are dependencies between the instances of A, and members in C_{mobile}.

4 MOBILE CONTEXT RECOGNITION SYSTEM

In this paper, our mobile context recognition system is used to tackle the problem of simultaneous inference of human activities, transportation modes (including their environments) of mobile users in their daily life. Therefore, we designed this system to be generic for mobile context recognition purposes. In overall, this system is composed of two overarching sub-modules (refer to Figure 2): contextual modelling and multi-contexts inference.

In the first module, the process of constructing structured hierarchical contexts is performed using the proposed modelling approach (defined in Section 4.1) by generalising, decomposing and extracting the contexts associated with human activities in a mobile sensing application. The second module is dedicated to a simultaneous inference of the human activity and related contexts of the mobile user.
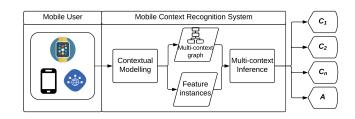

Figure 2: General overview of mobile context recognition system

In the contextual modelling module, there are two major components associated: 1) feature construction and extraction, and 2)

modelling human activities, transportation modes and their related contexts in a hierarchical structure. The first component relates to typical process to produce important features of a segment, which is related to a raw user annotation a. The second component will be elaborated in Section 4.1.
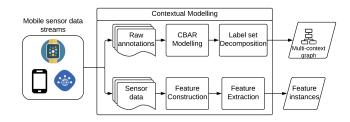

Figure 3: Process workflow in contextual modelling component of mobile context recognition system

4.1 Context-based Activity Recognition (CBAR) Modelling

Raw annotations can be noisy and inconsistent in a diary-based study, especially the ones collected through Experience Sampling Method (ESM) [9]. In this section, context-based activity recognition (CBAR) modelling is presented as a conceptual modelling approach that can be used to interlink the raw annotations to human activities and mobile contexts defined in Definition 3.2. In other cases, entity recognition [28] can be performed on textual data of the raw user annotation, which can then be automatically associated to a relevant human activity and its mobile contexts (including transportation mode).

Since the human activities are being contained in a space, there are several common properties that can characterise the transportation modes and environments of the mobile users. In this paper, we refer these as the **properties of sensing space**, which consist of:

(1) **Mobility of the sensing space.** For instance, the environment can move in spatial and temporal domains. In other words, the mobile users will experience dynamic mobility extrapolated by their sensing space (environment) by moving from one location to another location. In this paper, the nature of a sensing space can defined whether it is: 1) moving or 2) stationary.

(2) **Motorisation of the sensing space.** In this case, the environment (space) where mobile users are contained in can be characterised whether it is motorised or not. Consequently, further categorisations (e.g. vehicle vs non-vehicle) of the environments can be derived from this property.

As a result, the human activities can be modelled based on the above properties, which could produce a significant distinction of where the mobile users are contained in (including the environmental contexts associated to the contained space). It should be noted that the above properties can be expanded further to characterise human activities in the wild. In this paper, these properties are presented in our initial framework for the purpose of multi-context recognition of mobile users.
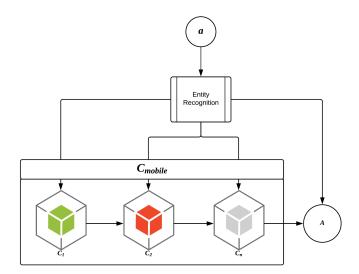

Figure 4: CBAR Modelling from raw user annotation

As shown in Figure 4, a given raw user annotation a can be decomposed through a pre-defined entity recognition process. The objective of this conceptual process is to first derive relevant transportation modes and human activities from all user annotations. In C_{mobile} set, the last item (i.e. C_n) should correspond to the transportation mode (i.e. $C_n = trans$) where a human activity A is performed in. Consequently, the commonalities (based on **properties of sensing spaces**) of all transportation modes in the system can be derived into a sequence of C_1 to C_{n-1}. Since there is an assumption of dependencies between all mobile contexts and human activities, the modelling can be expressed as: $C_1 \rightarrow C_2 \rightarrow \cdots \rightarrow C_n \rightarrow A$.

For a typical implementation of simultaneous recognition of all contextual labels ($C_1, C_2, ..., C_n, A$), a simplistic approach is to build an independent classifier per each member of the contextual labels. This typical approach is used as the baseline in our experiment.

Based on our conceptual modelling approach, there are two approaches to address the simultaneous inference issue:

(1) Multi-stage inference from multiple classifiers that are modelled hierarchically. In other words, inference is made from C_1 sequentially to C_n (transportation mode), then A.

(2) Inference from a multi-target classifier. In particular, multi-target classification [16] is a special case of multi-label classification in terms of modelling the dependencies between target classes from multiple label sets.

As a result of CBAR modelling process in our system, all raw annotations can be then be forwarded to the label set decomposition process, where multi-context graph (schema) is produced as an output for the next phase (models construction in multi-context inference module of our mobile context recognition system).

4.2 Multi-context Inference of Transportation Mode and Human Activity

Given the multi-context graph (schema) and feature instances produced from the previous module, various machine learning models can be constructed (refer to Figure 5).
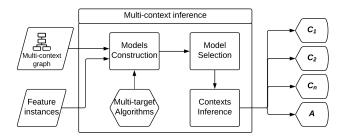
Figure 5: Process workflow in multi-context inference of mobile context recognition system

The construction of models are based on two approaches that have been previously mentioned in Section 4.1. The first approach refers to construction of a classifier set, where inferred context (output) becomes the input for next inference stage. In other words, the classifiers are modelled in a hierarchical structure until the final inference is achieved. The second approach of constructing a model is by training a classifier based on multi-target algorithms. Each model is constructed based on the input of training set (from feature instances). Ultimately, contexts inference sub-component takes the best model to perform simultaneous recognition of $C_1, C_2, ..., C_n$ and A. Consequently, the best model is selected through an internal evaluation process in our system based on appropriate metrics (e.g. the proportion of exact match).

5 EXPERIMENTS AND EVALUATION

5.1 Crowdsignal Dataset

In order to validate our mobile context recognition system, Crowdsignal dataset [38] is used in the experiment. This dataset consists of rich crowdsourced mobile sensor data from smartphones and wearable devices through in-the-wild data collection campaign. Hence, each annotation is labelled by real participants in their daily life. In particular, we leverage the time-interval labels (refer to the ground truth annotations of scripted behaviour) in this dataset, due to the presence of exact start and end time of specific and ongoing activities, events and situations. Although this dataset consists of the lifelog mobile sensor data from more than 30 participants, we use the Android smartphone data of five representative participants for the experiment and evaluation in this paper. From these five mobile users, we identified the following unique raw annotations:

(1) Walking
(2) Riding in a car
(3) Bus riding
(4) Playing video game
(5) Stairs
(6) Light rail riding
(7) Escalator
(8) Elevator
(9) Drinking
(10) Riding scooter

For a typical solution of activity recognition, these ten annotations can be used directly to classify the human activity based on

single-label-classification mechanism. However, it should be noted that from the recognised human activities, their transportation modes need to be inferred further. Through the CBAR modelling approach, we can decompose these annotations to three different mobile context label sets $C_{mobile} = \{C_1, C_2, C_3\}$ and an activity label set A as shown in Figure 6. In other words, each raw annotation a above will be associated to all decomposed contextual labels (i.e. $a = \{C_1, C_2, C_3, A\}$). In this case, the ultimate aim of a mobile sensing application is to infer the all decomposed labels accurately.

Although the decomposition of a raw user annotation can be performed through entity recognition, the schema graph (refer to Figure 6) is manually defined in our experiment. Hence, a mobile context recognition system should be able to perform an accurate simultaneous inference of mobile user contexts from the following label sets derived from Crowdsignal dataset:

(1) **Environment type of transportation mode** *(C_1)* label set:
$$C_1 = \{\text{Moving, Stationary}\}$$

(2) **Type of sensing space** *(C_2)* label set:
$$C_2 = \begin{Bmatrix} \text{Vehicle Riding, Non-vehicle Riding,} \\ \text{Stairs, Others} \end{Bmatrix}$$

(3) **Transportation mode** *(C_3)* label set:
$$C_3 = \begin{Bmatrix} \text{Bus, Light Rail,} \\ \text{Car, Scooter,} \\ \text{Escalator, Elevator} \end{Bmatrix}$$

(4) **Human activity** *(A)* label set:
$$A = \begin{Bmatrix} \text{Riding,} \\ \text{Walking,} \\ \text{Drinking,} \\ \text{Playing video game} \end{Bmatrix}$$

Table 1: Activity segments and their annotations for sampled mobile users of Crowdsignal dataset

User ID	Smartphone	Annotated segments	Annotations
A	Asus Zenfone 2	56	1, 2, 3, 4, 5
B	Samsung Galaxy S7	20	1, 2, 3, 5, 6, 7, 8, 9
C	Samsung Galaxy S7 Edge	131	1, 3, 4, 5, 6, 7, 8, 9, 10
D	Samsung Galaxy S6 Edge	212	1, 2, 3, 4, 5, 6, 7, 8, 9
E	Samsung SM-A800F	325	1, 2, 3, 4, 5, 6, 7, 8, 9

In terms of the coverage of activity segments, the annotations are not always given by the participants at all time. This evidence is also shown in Table 1 and Figure 7.

As described in Table 1, the sampled mobile users have different types of Android smartphones. Consequently, the accuracy of multi-context inference could also be affected by the quality of mobile sensors assembled by the device manufacturers. Moreover, each smartphone model has different capabilities for mobile sensing due to availability of sensor streams and limited types of embedded mobile sensors. To build a generic model (i.e. person-independent model), we leverage the following sensor streams that exist for the five sampled mobile users:
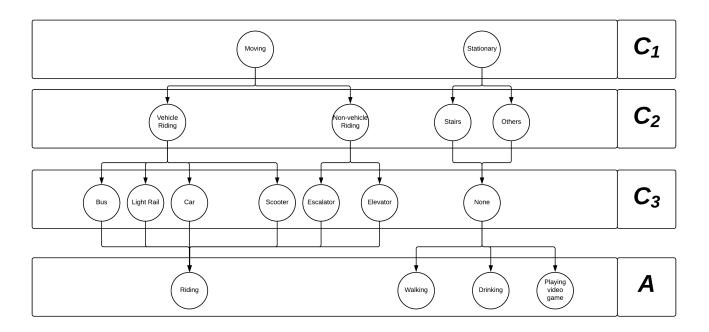
(1) Accelerometer

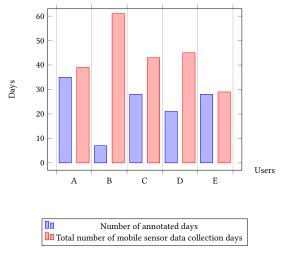**Figure 6: Multi-context graph from CBAR modelling of Crowdsignal dataset**

Figure 7: Number of annotated days and total days for sampled mobile users of Crowdsignal dataset

(2) Magnetic field

(3) Light

(4) Screen status

In the Crowdsignal dataset, these sensors may be sampled at different frequencies. In particular, accelerometer, magnetic field and light sensors are sampled with **SENSOR_DELAY_FASTEST** setting during the mobile data collection on Android smartphones. This setting corresponds to the default setting of the associated smartphone model for reading the sensor data as fast as possible.

Hence, the frequency might also be affected by the quality of sensors embedded in these smartphones. For instance, an accelerometer sensor may be sampled at 20Hz for a particular smartphone while another model would be sampled at 50Hz. For the screen status, the smartphone listens to the callback events when the screen is on or off. Consequently, the granularity of the data used in our experiment is shown in Table 2 in terms of counts of sensor data points.

Table 2: Granularity (count) of raw sensor data points from Android smartphones

User ID	Accelerometer	Magnetic field	Light	Screen status
A	10,270,550	10,149,507	51,896	371
B	1,044,436	247,243	24,981	49
C	17,318,445	4,786,078	447,907	460
D	22,554,947	9,627,477	521,762	1,214
E	40,040,845	20,814,032	40,755,734	1,664

5.2 Feature Construction and Extraction

For each annotated segment, we applied time-interval based temporal segmentation (1 second window size τ with 50% overlap) and extract the features for each sensor stream. In this case, we extracted the statistical features (mean, median, maximum, minimum, standard deviation, interquartile range, root-mean-square) from each window of sensor streams. In other words, a statistical feature is computed via a given function for all feature values in a window, bounded by τ.

For accelerometer and magnetic field sensor streams, the features are extracted from the magnitude value computed from tri-axial sensor readings. In other words, the magnitudes for accelerometer mg_{acc} and magnetic field mg_{mag} are calculated as follows:

(1)
$$mg_{acc} := \sqrt{x_{acc}^2 + y_{acc}^2 + z_{acc}^2}$$

where x_{acc}, y_{acc}, z_{acc} are the tri-axial sensor values of phone's acceleration.

(2)
$$mg_{mag} := \sqrt{x_{mag}^2 + y_{mag}^2 + z_{mag}^2}$$

where x_{mag}, y_{mag}, z_{mag} are the tri-axial sensor values of magnetic field measurement.

The purpose of computing and leveraging the magnitude values is to construct a more robust model, which is invariant towards smartphone's orientation [40]. For light sensor stream, the statistical features are extracted from the raw reading of illuminance (measured in lx) within a given window. On the other hand, the screen status corresponds to whether the screen on the smartphone is on or off, which may indicate how users are actively engaged with their smartphones. Furthermore, it is known that there are distinct variations of accelerometer and magnetometer reading inside various indoor spaces [33]. Hence, it is justified that a robust and effective model could be build to infer the transportation mode of a mobile user at a given point in time. Unlike the systems such as [6, 25] that require GPS sensor to determine the transportation mode, our mobile context recognition system do not need to rely on the user's locations. In fact, sampling from the GPS sensor is known to have significant energy consumption compared to other sensors (also proven in the study by Hemminki et al. [12]). It should be noted that the gyroscope sensor stream is excluded during the modelling stage due to its unavailability for user E's smartphone model.

5.3 Experiment Setup

A pre-defined schema (refer to Figure 6) is used to map the predicted raw annotation to the associated environment type (C_1) of transportation mode, further categorisations (C_2) of C_1, transportation mode (C_3) and human activity (A).

Our experiment is composed of two distinct sets. The first set refers to inference made by independent classifiers for each contextual label (shown in Figure 8). In this case, the classifiers are independently trained based on decomposed label sets (defined in Section 5.1). The first experiment set is used as the baseline of a typical mobile sensing application.

The second set of our experiment is associated with the inference of all contextual labels using our mobile context recognition system. We implemented two approaches of models construction of multi-context inference. The first one is referred as **multi-stage inference** where a classifier is trained on each target label set. Subsequently, the output of inference will be appended to the feature instances that are used the next stage of inference (refer to Figure 9). For instance, the training process of classifier for C_2 relies on feature instances (consist of feature vectors and ground-truth vector of C_1).
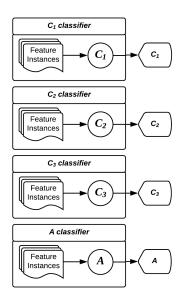

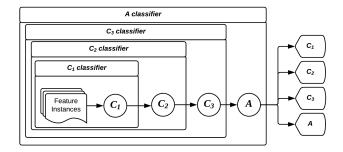Figure 8: Inference from independent classifiers

Figure 9: Multi-stage inference from classifiers

The second approach of models construction is based on the pre-defined multi-target algorithms (refer to Figure 10). Hence, the multi-target classifiers are deployed and evaluated on the following algorithms:

(1) Class Relevance (CR), which is the multi-target version of the Binary Relevance (BR) method for multi-label classification.
(2) Classifier Chains (CC)
(3) Nearest Set Replacement (NSR), which is the multi-target version of Pruned Set (PS) method for multi-label classification.
(4) Ensemble of Classifier Chains of boosted classifiers (EN-CC-AdaBoost). For the boosting implementation, AdaBoost algorithm is used. The setup of EN-CC-AdaBoost could require significant memory resources in order to operate. This fact is also proven during our experiment.

For both experiment sets, we leverage the following algorithms as the base classifiers:
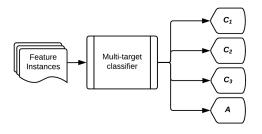
(1) Naïve Bayes (NB).
(2) Support Vector Classifier (SVC).

Figure 10: Multi-target inference from a classifier

(3) Multilayer Perceptron (MLP).
(4) Decision Tree (DT).
(5) Random Forests (RF).
(6) k-Nearest Neighbor with k=1 (1NN).

For the baseline (first experiment set) and multi-stage classification experiment (in our mobile context recognition system), we implemented the inferences using WEKA (version 3.8.1) data mining software library [11]. It contains the various functions needed for data exploration and data mining tasks. For base classifiers, the default configurations are used imperatively in WEKA. In this case, decision tree classifier is named as J48 in WEKA. Default parameter for k-Nearest Neighbor classifier is set where $k = 1$. In particular, we use the SMO algorithm [13] for SVC in WEKA. Furthermore, the maximum number of trees for Random Forests is set to 100.

For the multi-target inferences (in our mobile context recognition system), they are evaluated based on our experiment using MEKA (version 1.9.0) [24], a multi-label/multi-target extension to WEKA.

5.4 Evaluation

In order to validate the accuracy of multi-context inferences, 10-fold cross-validation was applied to all runs in our experiment.

As mentioned previously, the experiment is evaluated on following two experiment sets: inference from independent classifiers (baseline) and mobile contexts inference in our system (using both multi-stage and multi-target algorithms approaches in models construction process).

For the evaluation, we leverage exact match and accuracy metrics. Accuracy is computed for each inference from the corresponding contextual label set, while exact match measure refers to the proportion of correct prediction of all contextual labels (simultaneous inference of all contextual labels). Table 5 shows the overall performance for independent inference, given the base classifiers. Although the accuracy performance between independent inference and multi-stage inference is competitive, the exact match (simultaneous inference of mobile contexts) through multi-stage models shows a significant difference, proving that the dependencies between human activities and transportation mode (including related contexts) are not to be neglected for a reliable inference in a mobile sensing application.

From the performance comparison in Table 3, 4, 5, 6 and 7, we can conclude that best base classifier is dedicated to Random Forests in most cases (except the performance in Table 8), thus maintaining itself as the state-of-art algorithm for a real mobile sensing application. Although the basic tree based classifier (i.e. Decision Tree) has

Table 3: Performance of independent classifiers (baselines)

Metric	Baseline					
	NB	SVC	MLP	DT	RF	1NN
Exact Match	1.4%	45.5%	34.5%	70.8%	**89.4%**	88.7%
Accuracy: C_1	66.0%	66.9%	68.5%	88.1%	**93.4%**	92.1%
Accuracy: C_2	11.6%	65.3%	67.3%	87.5%	**93.0%**	91.6%
Accuracy: Transportation Mode (C_3)	5.4%	49.4%	56.4%	83.5%	**91.7%**	89.3%
Accuracy: Activity	16.8%	67.0%	69.2%	87.5%	**93.0%**	91.6%

Table 4: Performance of Multi-stage models

Metric	Multi-stage classification					
	NB	SVC	MLP	DT	RF	1NN
Exact Match	2.6%	48.0%	49.7%	81.0%	**90.7%**	88.7%
Accuracy: C_1	66.0%	66.9%	68.5%	88.1%	**93.4%**	92.1%
Accuracy: C_2	12.3%	65.2%	67.0%	87.2%	**92.8%**	91.6%
Accuracy: Transportation Mode (C_3)	5.8%	48.2%	52.7%	82.4%	**91.1%**	89.3%
Accuracy: Activity	22.4%	66.8%	65.9%	87.1%	**93.1%**	91.6%

Table 5: Performance of Class Relevance (CR) models

Metric	Class Relevance					
	NB	SVC	MLP	DT	RF	1NN
Exact Match	1.4%	45.3%	36.7%	70.8%	**92.9%**	88.7%
Accuracy: C_1	66.0%	66.9%	68.8%	88.1%	**93.5%**	92.1%
Accuracy: C_2	11.5%	65.3%	66.8%	87.5%	**93.1%**	91.6%
Accuracy: Transportation Mode (C_3)	5.4%	49.6%	56.0%	83.4%	**91.8%**	89.3%
Accuracy: Activity	16.8%	67.0%	68.5%	87.5%	**93.2%**	91.6%

Table 6: Performance of Classifier Chains (CC) models

Metric	Classifier Chains					
	NB	SVC	MLP	DT	RF	1NN
Exact Match	2.6%	48.8%	49.1%	82.1%	**91.4%**	88.7%
Accuracy: C_1	66.0%	67.2%	67.4%	88.3%	**93.7%**	92.1%
Accuracy: C_2	12.2%	65.7%	65.9%	87.4%	**93.3%**	91.6%
Accuracy: Transportation Mode (C_3)	22.4%	49.6%	56.0%	83.4%	**91.8%**	89.3%
Accuracy: Activity	5.8%	66.6%	61.2%	87.3%	**93.4%**	91.6%

Table 7: Performance of NSR models

Metric	NSR					
	NB	SVC	MLP	DT	RF	1NN
Exact Match	11.4%	49.5%	50.4%	82.4%	**91.6%**	88.9%
Accuracy: C_1	38.1%	68.1%	69.0%	88.2%	**93.8%**	92.4%
Accuracy: C_2	32.7%	66.5%	67.4%	87.3%	**93.4%**	91.9%
Accuracy: Transportation Mode (C_3)	27.3%	49.8%	56.2%	83.3%	**91.8%**	89.6%
Accuracy: Activity	22.5%	67.9%	63.8%	87.4%	**93.6%**	91.9%

the best performance in Table 8, its overall performance is yet lower than our proposed multi-stage inference models (Table 4). Hence, this result suggests that the applications of ensemble and boosting on multi-target algorithms may not increase the performance of a mobile context recognition system. However, this approach can still be used for model selection (in the internal evaluation

Table 8: Performance of mobile context recognition system using En-CC-AdaBoost

Metric	EN-CC-AdaBoost					
	NB	SVC	MLP	DT	RF	1NN
Exact Match	1.6%	48.7%	50.1%	**90.1%**	89.8%	86.9%
Accuracy: C_1	65.9%	67.4%	70.9%	**93.1%**	92.9%	91.3%
Accuracy: C_2	10.5%	65.8%	69.4%	**92.7%**	92.4%	90.7%
Accuracy: Transportation Mode (C_3)	4.9%	49.2%	57.1%	**90.8%**	90.4%	87.9%
Accuracy: Activity	18.0%	67%	67.5%	**92.8%**	92.5%	90.6%

process), in case if it outperforms other inference approaches. Inherently, the applications of both ensembles and boosting consume significant resources (in terms of training time, memory and space requirements).

Through our mobile context recognition system, its performance for the exact match is increased by 2.2% (from the baseline), which is ultimately achieved by a multi-target inference approach of NSR algorithm. On the other hand, the most significant increase of 19.3% can be noticed (for exact match in Table 10) through the usage of ensemble, multi-target classifier chains and boosting strategy on decision tree (as base classifier). Furthermore, the significant improvement is also shown for the accuracy of each contextual label (e.g. inference of transportation mode depicted in Table 9).

Table 9: Accuracy of transportation mode inference for baseline and mobile context recognition system

System	NB	SVC	MLP	DT	RF	1NN
Baseline	5.4%	49.4%	**56.4%**	83.4%	91.7%	89.3%
Proposed system	**27.3%** (NSR)	**49.8%** (NSR)	56.0% (CR, CC)	**90.8%** (EN-CC-AdaBoost)	**91.8%** (NSR)	**89.6%** (NSR)

Table 10: Exact match for baseline and mobile context recognition system

System	NB	SVC	MLP	DT	RF	1NN
Baseline	1.4%	45.6%	34.5%	70.8%	89.4%	88.7%
Proposed system	**11.4%** (NSR)	**49.5%** (NSR)	**50.4%** (NSR)	**90.1%** (EN-CC-AdaBoost)	**91.6%** (NSR)	**88.9%** (NSR)

By being able to recognise multiple contexts of mobile users simultaneously, many interactive and ubiquitous applications can be enabled. Let us consider the scenario of a smart device to be able to recognise the user contexts in daily life ubiquitously. Given the accurate simultaneous recognition of both user environments and activities, it can be used by the assistive technologies to improve the quality of life for these mobile users. For example, intelligent reminders of user activities and notifications for major transportation delays due to the current situation of the users. This outcome can also be leveraged for the applications of discovering user routines [21, 22, 27] based on personal contexts of mobile users. In an intelligent healthcare scenario, a robust and simultaneous recognition of multiple user contexts would be important to be considered for elderly and disabled people, while travelling through various

accessible paths [23]. In a large and pervasive sensing scenario (e.g. [17]), mobile context recognition would be beneficial for crowd-sensing from smart devices (e.g. Internet of Things) if the model can be used for robust and simultaneous user contexts recognition. In smart city applications, simultaneous contexts recognition can be used to enable smarter citizen violation prediction based on real-time inference from multiple sources, such as parking violation prediction [29] that can be improved by considering the ubiquitous contexts from both parking officer and IoT devices on the parking spaces. In the case of analysing the space utilisation (e.g. [2–5]), simultaneous recognition of contexts from multiple rooms can be leveraged for better occupancy counting and recommendation of places to occupy, also by considering the dynamic environment and personal user contexts. In a smart home environment, intelligent electricity consumption prediction (e.g. [31]) can be improved by considering multiple contexts recognition of tenant activities.

6 CONCLUSION AND FUTURE WORKS

In this paper, a mobile context recognition system is presented for simultaneous inference of transportation modes and human activities, including the environmental contexts of transportation modes. Our system and study aim to enable smarter mobile sensing applications by being able to provide intelligent assistance from the accurate inference of related mobility contexts associated with human activities in daily life. Through our hierarchical and contextual modelling approach (combined with multi-target algorithms), our system is capable of outperforming the traditional approach of independent inference for multiple contexts of mobile users. Through the exact match of all contextual labels of the mobile users, 19.3% improvement is noticeable for a solution with decision tree as the base classifier. On the other hand, the best performance of exact match is acquired by random forests based solution in multi-context recognition by far.

Furthermore, our system is robust towards the intelligent inference for the data streamed from low-energy sensors, which reduces the overall reliance on location-based sensors. Nevertheless, many challenges should be addressed to improve our system further. For example, one of the immediate challenges in real-world is related to coping context recognition with subjective and in-situ user annotations (e.g. in a mobile crowdsensing scenario). Consequently, our future works will include the adaptation towards the new label that emerges from mobile users and accurate prediction of human annotation in a diary-based study, within the scope of intelligent assistant applications.

REFERENCES

[1] Zahraa Said Abdallah, Mohamed Medhat Gaber, Bala Srinivasan, and Shonali Krishnaswamy. 2015. Adaptive mobile activity recognition system with evolving data streams. *Neurocomputing* 150 (2015), 304–317.

[2] Irvan Bastian Arief-Ang, Margaret Hamilton, and Flora Dilys Salim. 2018. RUP: Large Room Utilisation Prediction with carbon dioxide sensor. *Pervasive and Mobile Computing* 46 (2018), 49–72.

[3] Irvan Bastian Arief-Ang, Margaret Hamilton, and Flora Dilys Salim. 2018. A Scalable Room Occupancy Prediction with Transferable Time Series Decomposition of CO2 Sensor Data. *ACM Transactions on Sensor Networks* (2018).

[4] Irvan Bastian Arief-Ang, Flora Dilys Salim, and Margaret Hamilton. 2016. Human occupancy recognition with multivariate ambient sensors. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 1–6.

[5] Irvan Bastian Arief-Ang, Flora Dilys Salim, and Margaret Hamilton. 2017. DA-HOC: Semi-Supervised Domain Adaptation for Room Occupancy Prediction using CO2 Sensor Data. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys '17)*. ACM, New York, NY, USA, Article 1, 10 pages.

[6] Young-Ji Byon and Steve Liang. 2014. Real-time transportation mode detection using smartphones and artificial neural networks: Performance comparisons between smartphones and conventional global positioning system sensors. *Journal of Intelligent Transportation Systems* 18, 3 (2014), 264–272.

[7] Liming Chen, Jesse Hoey, Chris D Nugent, Diane J Cook, and Zhiwen Yu. 2012. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (Nov 2012), 790–808.

[8] Hyunjeong Cho, Jihoon An, Intaek Hong, and Younghee Lee. 2015. Automatic Sensor Data Stream Segmentation for Real-time Activity Prediction in Smart Spaces. In *Proceedings of the 2015 Workshop on IoT Challenges in Mobile and Industrial Systems (IoT-Sys '15)*. ACM, New York, NY, USA, 13–18.

[9] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology*. Springer, 35–54.

[10] Tao Feng and Harry J.P. Timmermans. 2013. Transportation mode recognition using {GPS} and accelerometer data. *Transportation Research Part C: Emerging Technologies* 37 (2013), 118 – 130.

[11] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.* 11, 1 (Nov. 2009), 10–18.

[12] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. 2013. Accelerometer-based Transportation Mode Detection on Smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys '13)*. ACM, New York, NY, USA, Article 13, 14 pages.

[13] S. Sathiya Keerthi, Shirish Krishnaj Shevade, Chiranjib Bhattacharyya, and Karuturi Radha Krishna Murthy. 2001. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural computation* 13, 3 (2001), 637–649.

[14] Youngsung Kim, F.C. Pereira, Fang Zhao, A. Ghorpade, P.C. Zegras, and M. Ben-Akiva. 2014. Activity Recognition for a Smartphone Based Travel Survey Based on Cross-User History Data. In *22nd International Conference on Pattern Recognition (ICPR), 2014*. 432–437.

[15] Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T. Campbell. 2010. A survey of mobile phone sensing. *Communications Magazine, IEEE* 48, 9 (Sept 2010), 140–150.

[16] Mark Last. 2016. Multi-target Classification: Methodology and Practical Case Studies. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 280–283.

[17] Jonathan Liono, Prem Prakash Jayaraman, A.K. Qin, Thuong Nguyen, and Flora D. Salim. 2018. QDaS: Quality driven data summarisation for effective storage management in Internet of Things. *J. Parallel and Distrib. Comput.* (2018).

[18] Jonathan Liono, A. K. Qin, and Flora D. Salim. 2016. Optimal Time Window for Temporal Segmentation of Sensor Streams in Multi-activity Recognition. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MOBIQUITOUS 2016)*. ACM, New York, NY, USA, 10–19.

[19] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. 2010. The Jigsaw Continuous Sensing Engine for Mobile Phone Applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys '10)*. ACM, New York, NY, USA, 71–84.

[20] George Okeyo, Liming Chen, Hui Wang, and Roy Sterritt. 2014. Dynamic Sensor Data Segmentation for Real-time Knowledge-driven Activity Recognition. *Pervasive Mob. Comput.* 10 (Feb. 2014), 155–172.

[21] Mohammad Saiedur Rahaman, Margaret Hamilton, and Flora D. Salim. 2017. Predicting Imbalanced Taxi and Passenger Queue Contexts in Airport. In *Proc. of the Pacific Asia Conf. on Info. Systems (PACIS)*.

[22] Mohammad Saiedur Rahaman, Margaret Hamilton, and Flora D. Salim. 2018. *Using Big Spatial Data for Planning User Mobility*. Springer International Publishing, Cham, 1–6.

[23] Mohammad Saiedur Rahaman, Yi Mei, Margaret Hamilton, and Flora D. Salim. 2017. CAPRA: A contour-based accessible path routing algorithm. *Information Sciences* 385-386 (2017), 157 – 173.

[24] Jesse Read. 2012–2015. MEKA: a multi-label extension to WEKA. http://meka.sourceforge.net/. [Online; accessed 02-December-2015].

[25] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. 2010. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)* 6, 2 (2010), 13.

[26] Daniele Riboni and Claudio Bettini. 2011. COSAR: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing* 15, 3 (2011), 271–289.

[27] Amin Sadri, Yongli Ren, and Flora D. Salim. 2017. Information gain-based metric for recognizing transitions in human activities. *Pervasive and Mobile Computing* 38 (2017), 92 – 109.

[28] Immanuel Schweizer and Benedikt Schmidt. 2014. Kraken.me: multi-device user tracking suite. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 853–862.

[29] Wei Shao, Flora D. Salim, Tao Gu, Ngoc-Thanh Dinh, and Jeffrey Chan. 2018. Traveling Officer Problem: Managing Car Parking Violations Efficiently Using Sensor Data. *IEEE Internet of Things Journal* 5, 2 (April 2018), 802–810.

[30] Timothy Sohn, Alex Varshavsky, Anthony LaMarca, MikeY. Chen, Tanzeem Choudhury, Ian Smith, Sunny Consolvo, Jeffrey Hightower, WilliamG. Griswold, and Eyal de Lara. 2006. Mobility Detection Using Everyday GSM Traces. In *UbiComp 2006: Ubiquitous Computing*, Paul Dourish and Adrian Friday (Eds.). Lecture Notes in Computer Science, Vol. 4206. Springer Berlin Heidelberg, 212–224.

[31] Hui Song, A. K. Qin, and Flora D. Salim. 2017. Multi-resolution Selective Ensemble Extreme Learning Machine for Electricity Consumption Prediction. In *Neural Information Processing*, Derong Liu, Shengli Xie, Yuanqing Li, Dongbin Zhao, and El-Sayed M. El-Alfy (Eds.). Springer International Publishing, Cham, 600–609.

[32] Xing Su, Hanghang Tong, and Ping Ji. 2014. Activity recognition with smartphone sensors. *Tsinghua Science and Technology* 19, 3 (June 2014), 235–249.

[33] Melania Susi, Valérie Renaudin, and Gérard Lachapelle. 2013. Motion mode recognition and step detection algorithms for mobile phone users. *Sensors* 13, 2 (2013), 1539–1562.

[34] Yonatan Vaizman, Nadir Weibel, and Gert Lanckriet. 2018. Context Recognition In-the-Wild: Unified Model for Multi-Modal Sensors and Multi-Label Classification. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 168 (Jan. 2018), 22 pages.

[35] Jie Wan, Michael J. O'grady, and Gregory M. O'hare. 2015. Dynamic sensor event segmentation for real-time activity recognition in a smart home context. *Personal and Ubiquitous Computing* 19, 2 (2015), 287–301.

[36] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. 2011. Human Mobility, Social Ties, and Link Prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*. ACM, New York, NY, USA, 1100–1108.

[37] Zhelong Wang, Ming Jiang, Yaohua Hu, and Hongyi Li. 2012. An Incremental Learning Method Based on Probabilistic Neural Networks and Adjustable Fuzzy Clustering for Human Activity Recognition by Using Wearable Sensors. *Information Technology in Biomedicine, IEEE Transactions on* 16, 4 (July 2012), 691–699.

[38] Evan Welbourne and Emmanuel Munguia Tapia. 2014. CrowdSignals: a call to crowdfund the community's largest mobile dataset. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 873–877.

[39] Gerhard Widmer and Miroslav Kubat. 1996. Learning in the Presence of Concept Drift and Hidden Contexts. *Machine Learning* 23, 1 (1996), 69–101.

[40] Aras Yurtman and Billur Barshan. 2017. Activity recognition invariant to sensor orientation with wearable motion sensors. *Sensors* 17, 8 (2017), 1838.