Mobile Activity Recognition for a Whole Day: Recognizing Real Nursing Activities with Big Dataset

Sozo INOUE
Kyushu Institute of Technology
1-1 Sensuicho, Tobata-ku, Kitakyushu, 804-8550, Japan
sozo@mns.kyutech.ac.jp

Naonori UEDA
NTT Communication Science Laboratories
2-4 Hikaridai, Seika, Sagara-gun, Kyoto, 619-0237, Japan
ueda.naonori@lab.ntt.co.jp

Yasunobu NOHARA and Naoki NAKASHIMA
Kyushu University Hospital
3-1-1, Maidashi, Higashi-ku, Fukuoka, 812-8582, Japan
{nnaoki|y-nohara}@med.kyushu-u.ac.jp

ABSTRACT
In this paper, we provide a real nursing data set for mobile activity recognition that can be used for supervised machine learning, and big data combined the patient medical records and sensors attempted for 2 years, and also propose a method for recognizing activities for a whole day utilizing prior knowledge about the activity segments in a day. Furthermore, we demonstrate data mining by applying our method to the bigger data with additional hospital data. In the proposed method, we 1) convert a set of segment timestamps into a prior probability of the activity segment by exploiting the concept of importance sampling, 2) obtain the likelihood of traditional recognition methods for each local time window within the segment range, and, 3) apply Bayesian estimation by marginalizing the conditional probability of estimating the activities for the segment samples. By evaluating with the dataset, the proposed method outperformed the traditional method without using the prior knowledge by 25.81% at maximum by balanced classification rate. Moreover, the proposed method significantly reduces duration errors of activity segments from 324.2 seconds of the traditional method to 74.6 seconds at maximum. We also demonstrate the data mining by applying our method to bigger data in a hospital.

Author Keywords
Mobile activity recognition; nursing activity; domain-specific activity recognition; dataset

ACM Classification Keywords
I.5.4. Pattern Recognition: Signal processing

INTRODUCTION
In the field of healthcare, the standardization of care processes, termed *Clinical* or *Critical pathways*, has been attempted [37, 40, 12, 41, 29, 21]. In meeting such an objective, the recognition and data mining of nursing activities can lead to a better understanding and improvements in medical care, and they can help prevent unnecessary activities and excessive work. At the same time, these approaches are beneficial to patients because the overall care process is optimized, thus resulting in shorter hospitalization times and lower costs.

UbiComp '15, September 7-11, 2015, Osaka, Japan.

http://dx.doi.org/10.1145/2750858.2807533

Recently, researchers have explored the possibility of human activity recognition with mobile sensors; for example, accelerometers, gyroscopes, and low-frequency audio have been explored [8, 44, 50, 26, 28, 31, 24, 3, 30, 1, 39]. In addition, several researchers have applied such technology to domain-specific applications in nursing activities [34, 45, 36]. However, in the available methods, several unclear points still remain:

The nature of the real activities is not clear
In the application of nursing activity recognition, the *activity classes* — the types of activities — are defined in a domain-specific manner (as listed in Table 1. Here, the activities are not always easy to recognize because the table includes feature value varieties even for single classes, such as blood pressure measurements starting by attaching the corresponding equipment to a patient, followed by pushing air pumps periodically, and finishing with detaching the equipment. Moreover, such activities have imbalance varieties, such as the number of occurrences among classes, starting times in a day, and duration. For example, complex activities, such as capturing X-ray, require dozens of minutes, whereas other activities are completed more quickly. Because the traditional approach normally assumes that activity classes have similar probabilities of being performed, similar probabilities any time in a day, and similar durations, the way in which accuracy changes when we consider such imbalances is not known.

The application is not clear
In the application of nursing activity analysis, we can set up clear goals, such as improving nursing activities effectively for timing, duration, and patient satisfaction, or optimizing the costs of the nursing process. For such goals, the technical objective is not only improving recognition accuracy each time, derived from the traditional recognition from the current time window or those in the vicinity (called *local time windows*),but also estimating the *segment* — the range where the activity is performed continuously — attached with correct timestamps and durations. Thus, by clarifying the appli-

cation, we could choose the recognition aspects to which to assign importance, but this is not case with the existing work.

No dataset with clear goals

To overcome the aforementioned challenges, we require real data to evaluate or input into a machine learning algorithm. However, there is extreme shortage of such open datasets obtained from multiple subjects, and a set of entire days with densely annotated labels. In the literature, there are several datasets, such as [2, 5, 38] that provide data with longer times, but because they do not intend clear application, it is not clear what accuracy aspects to pursue.

For this paper, we collected 1) (*labled data*) actual activities from nurses wearing accelerometers in a hospital for approximately 2 weeks and combined them with training labels, which resulted in 25 activity classes with 5,743 labels from 22 nurses, and 2) (*unlabeled data*) the open big data for 60 nurses for 442 [days x people] in the trial for almost 2-years with the duty days which could obtain agreements from the nurses and up to 100 patients, combined with patients' wearable, vital, and environmental sensor data and medical records. From the obtained labeled data, we observed that the activities have imbalances in the number of occurrences for each activity class, the starting times in a day, and the duration of each activity class, as explained in Section "Sensor Data Collection for Nursing Activities".

Then, we propose a method for recognizing wholeday activities using prior knowledge on the information of a sequence of activity segments which are obtained from wholeday training dataset, such as the daily timestamps, duration, and imbalances among activity classes, as explained in Section "Activity Recognition for a Whole Day".

In the proposed method, we 1) convert the set of timestamps of the training data into the prior probability of the activity segment by exploiting the concept of importance sampling, 2) obtain the likelihood for the test data with a traditional recognition methods for each local time window within the range of the segments, and 3) apply Bayesian estimation by marginalizing the conditional probability of estimating the activities for the segment samples.

By evaluating with the nursing dataset in Section "Evaluation", the proposed method outperformed the naive method without using prior knowledge by 25.81% at maximum through the balanced classification rate. Moreover, the proposed method significantly reduces the duration of errors of activity segments from 324.2 seconds of the naive method to 74.6 seconds in k-NN, from 173.5 seconds to 90.33 seconds in NaiveBayes, and from 122.2 seconds to 7.88 seconds in RandomForest.

In order to demonstrate research probabilities with ubiquitous healthcare research to the community, we introduce an analysis of the unlabeled data utilizing the machine-learning result of the labeled data, combined with nurses profiles and medical records.

The contribution of our paper is four-fold: 1) provide the real dataset[1] of nursing activities that can be used for supervised machine learning, and also big data combined with patient medical records and sensors, 2) propose a method for utilizing prior knowledge on activity segments in a day, 3) evaluate the proposed method for improvements to the accuracy of activity recognition and the durations of activity segments, and 4) demonstrate data mining by applying our method to bigger data in a hospital merged with additional hospital data.

SENSOR DATA COLLECTION FOR NURSING ACTIVITIES

We collected mobile-sensor data from the nurses of a hospitals cardiovascular center [35]. The experiment was exclusive to those nurses who agreed to usage of the sensor data, and to the duties related to patients who consented to participate in the experiment.

It includes labeled data for 2 weeks, and unlabelled data for the duty days which we could obtain agreements from up to 100 patients in 2 years. In this section, we describe the protocols for data collection and review both of the labelled and the unlabelled datasets.

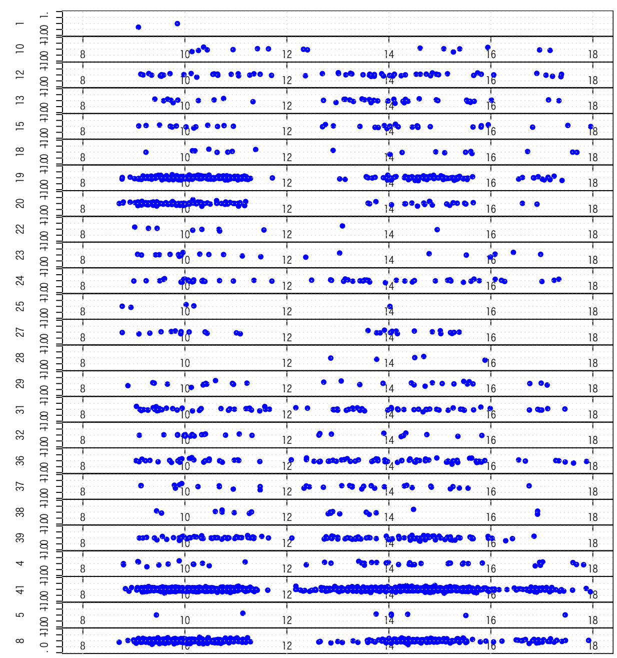

Figure 1. Start time for each activity in a day range. Each row corresponds to an activity class (the number corresponds to the No. in Table 1), and the x-axis is the hour in a day. The dots are the recorded starting time of an activity. We can see imbalances between activity classes and times in a day.

Protocol

We requested the nurses to wear mobile devices (iPod touches) that record accelerations in their breast pockets in a generally fixed direction. They also attached a small accelerometer device on their right wrist, and another on the back of their waist. Fig. 3 illustrates the attachments. Each sensor measured accelerations on three axes in the range of $\pm 2G$.

[1] http://nurseact.sozolab.jp

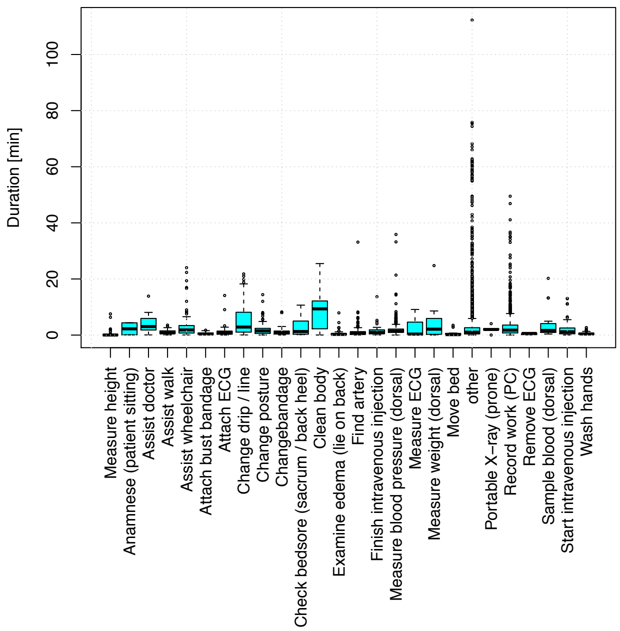

Figure 2. Durations of each activity label in the dataset

Labelled data collection

The daytime duties of 22 selected nurses over the period over two weeks on Feb. 2014 were labeled with mobile tablets by other nurses who acted as observers. Before the trial, we defined 41 activity classes from the clinical path, and asked the observers to record them.

Annotating labels for real activities requires careful design. In real nursing activities, nursing the patient has the highest priority, and there occur a lot of missing labels or incorrect timestamps. Therefore, another nurse acted as an observer, and operated another iPod device to record the activities of the subject nurse. On the software on the iPod, the observer selects the activity class which the subject nurse is about to start, and pushes the finish icon when the subject finishes.

Figure 3. Nurses with three accelerometers: one on their right wrist, one attached to their breast pocket, and one on the back hip.

In reality, if the observer waits for the subject nurse to start the activity, the start timestamp will have a latency than the correct one. Therefore, they collaborated with each other to have correct start timestamps, such that the subject nurse declares the observer the activity before s/he will start it.

Unlabelled data collection

In the same department of the hospital as above, we collected unlabeled sensor data for 2 years from the nurses who wear three accelerometers in the same way as the labelled data collection.

Since we also collected the patients' sensor and medical data associated with the nurses' mobile sensor data, — which are out of the main scope of this paper — we specifically collected the nurses' sensor data for the duty days which could obtain agreements. The data we used are collected carefully to be able to be open data, by obtaining agreements from the subject nurses and the patients.

Formatting the dataset

To interoperate the data sets for labelled and unlabeled data, they were formatted uniformly as well as possible. The ID for the nurses are consistent, then an ID for a nurse is the same for both data sets.

Moreover, even the each sensor on each position on the body stores their sensor data separately on the device, it is useful for data analysis to be merged into one multi-column table. Therefore, we joined the data for 3 devices' data of a duty date to a single table in an off-line manner. We firstly generated timestamps increasing by 20 Hz, which means 0.05 seconds, and adopted the closest sample within 0.025 seconds for each timestamp. If there are no samples within 0.025 seconds, we reused the last timestamp value.

Since each device has its own clock and they have no interaction for time synchronization with each other, there is a risk that the clock is not synchronized. To avoid this, we shook the devices together periodically — once in a day on average — as a reference timestamp, and used the relative time from the shaking time as well as possible.

Overview of the dataset

As the result of the experiment, we collected 346.5 [hours x people] of sensor data from 22 nurses by the labeled data collection, and 1,655 [day x people] from 60 nurses by the unlabeled data collection.

To review the collected labels, we review the labels obtained by the labelled data collection in the following.

After the trial, the activity classes actually observed were 25, listed in Table 1. The total number of labels was 5,743. The labels for each activity class are also listed in Table 1.

Fig. 1 shows the plot of the start times for each activity in a day range. Fig. 2 shows the duration of each activity class.

As shown in Fig. 1, the number of activities varies among activity classes. Moreover, we can see that not all the activities

Table 1. Observed activity classes and numbers of labels

No.	Activity class	# labels
1	Anamnese (patient sitting)	2
4	Measure height	45
5	Measure weight (dorsal)	8
8	Measure blood pressure (dorsal)	529
10	Sample blood (dorsal)	16
12	Start intravenous injection	61
13	Finish intravenous injection	40
15	Change drip / line	38
18	Assist doctor	19
19	Find artery	257
20	Examine edema (lie on back)	118
22	Check bedsore (sacrum / back heel)	10
23	Measure ECG	22
24	Attach ECG	54
25	Remove ECG	5
27	Attach bust bandage	29
28	Portable X-ray (prone)	5
29	Changebandage	30
31	Change posture	77
32	Clean body	27
36	Assist wheelchair	86
37	Assist walk	35
38	Move bed	19
39	Wash hands	117
41	Record work (PC)	912

occur at any time uniformly. Some activities, such as No. 27, occur only during several hours in the morning or afternoon, and others occur continuously, such as No. 12. Compared with traditional experiment settings where the training data are collected in a balanced way or in a short time without considering the time of day, this may result in difficulties during activity recognition.

Moreover, as shown in Fig. 2, activity duration also varies considerably. For example, the maximum median duration in the dataset we collected was 9.35 minutes for clean body, whereas the minimum was 0.03 minutes for "measure height." The variances within a class are large, such that measure weight has a standard deviation of 8.40 minutes, and other has 8.09. These phenomena are considered more significant than other research fields, such as segmentation in voice recognition [49, 33, 32, 6], and chunking in natural language processing [4, 11].

In summary, the real activity dataset attempted for several entire days has imbalances in several aspects, such as class-wise, times of day, and activity duration. If such information is obtained in the training phase, we can expect it to be instructive for improving activity recognition.

ACTIVITY RECOGNITION FOR A WHOLE DAY

In this section, we propose a method for recognizing the activities of an entire day. propose a method for recognizing activities of a whole day.

Approach

As shown in Fig. 1, nursing activities have different possibilities, depending on the time of day. If we have training data with labels and timestamps, we can convert the set of timestamps into the prior probability of the activity being performed. In addition, if we use both the starting and ending times of an activity, we can obtain information on the activity's duration. As explained in Section "Sensor Data Collection for Nursing Activities", such information of when and how long nursing activities are performed is important for analysis. In our approach, in addition to the traditional method for estimating activities from the sensor input of neighborhood time windows, we exploit the timestamp information in order to construct a prior probabilistic distribution on the activities of an entire day, implement them based on importance sampling, and utilize them for the Bayesian estimation of activities.

Preliminary

As a preliminary step, we introduce the mathematical expressions used throughout this paper. Table 2 provides a summary of expressions, and Fig. 4 shows an overview of the expression for a single activity class c.

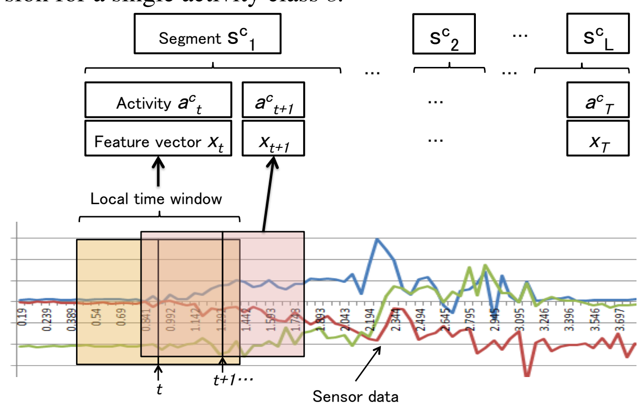

Figure 4. Overview of one-day activities for single activity class $c \in C$.

Table 2. Basic expressions used in the paper

Symbol	Summary
C	The set of activity classes to be recognized
$1 : T := (1, 2, \cdots, T)$	The time sequence in a day.
x_t	The feature vector at time t ($t \in 1 : T$).
a^c_t	Whether the activity at time t is c or not ($c \in C$).
L^c	The number of segments for activity $c \in C$.
$s^c_l := (b(l), e(l))$	The l'th segment ($l \in 1 : L^c$).
$b(l) \in 1 : T$	The start time of the l'th segment.
$e(l) \in 1 : T$	The end time of the l'th segment.

For simplicity, we assume that the time of day is expressed as an integer between one and T. We abbreviate the sequence $(1, 2, \cdots, T)$ as $1 : T$. For each t, we assume that a feature vector is extracted t. For each t, we assume that a *feature vector* is extracted that contains several statistic values from the time window of the sensor input around t. We specify the sequence of feature vectors $(x_1, x_2, \cdots, x_T)$ as $x_{1:T}$.

Moreover, C refers to the set of activity classes to be recognized. We assume that at any time t multiple activities might be included, either because the nurse is performing several activities concurrently, or because the activity-recognition algorithm conducts fuzzy estimations. Therefore, we define whether the activity at time t is $c \in C$ or not as the binary value a_t^c.

In the remainder of the section, we focus on the recognition of a single activity $c \in C$. In reality, we could apply the proposed method for each activity $c \in C$, and adopt either the most probable class $\arg_c \max \mathbf{P}(a_t^c)$, or adopt all classes estimated for a time t. In Sec , we evaluated the accuracy using the latter strategy.

We use the term *segment* as the continuous time range where the activity c is performed, and represent it as a pair of start and end times. When we assume that L^c segments are repeated for activity c in a day, the lth segment from time $b(l)$ to $e(l)$ is defined as:

$$s_l^c := (b(l), e(l)), \quad \text{where} \quad 1 \leq b(l) \leq e(l) \leq T.$$

Traditional activity recognition such as that from Bao et al.[1], can be modeled as the problem of obtaining the maximum argument $c \in C$ of

$$\mathbf{P}(a_t^c | x_t) \tag{1}$$

for the local time window t only. Note that obtaining $\mathbf{P}(x_t | a_t^c)$ is easy following Bayes' theorem. For the rest of this paper, we call $\mathbf{P}(x_t | a_t^c)$ a *local time likelihood.*

In contrast, our goal can be represented as the problem of obtaining the probability of an entire day's activities

$$\mathbf{P}(a_{1:T}^c | x_{1:T}).$$

For the remainder of this section, we describe the method used to conduct this.

Proposed method

We assume the Bayesian network as shown in Fig. 5.

Fig. 5 represents the conditional probabilities for one segment s_l^c. We assume that the probabilities between any segments s_l^c and $s_{l'}^c$ $(l \neq l')$ are independent.

The marginal probability of the figure is written as

$$\mathbf{P}(x_{b(l):e(l)}, a_{b(l):e(l)}^c, s_l^c)$$

$$= \mathbf{P}(s_l^c) \prod_{t \in b(l):e(l)} \mathbf{P}(x_t | a_t^c) \mathbf{P}(a_t^c | s_l^c)$$

When s_l^c is fixed, then a_t^c for $b(l) \leq t \leq e(l)$ is straightforward, and we can eliminate $\mathbf{P}(a_t^c | s_l^c)$. Accordingly,

$$= \mathbf{P}(s_l^c) \prod_{t \in b(l):e(l)} \mathbf{P}(x_t | a_t^c)$$

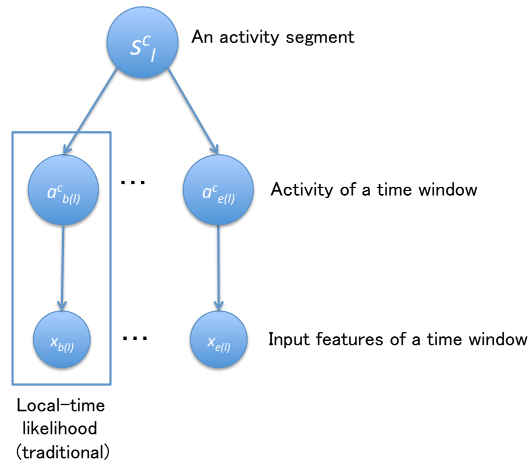

Figure 5. Overview of the proposed method

To obtain the conditional probability between $a_{b(l):e(l)}^c$ and $x_{b(l):e(l)}$, we marginalize s_t^c, then

$$\mathbf{P}(a_{b(l):e(l)}^c, x_{b(l):e(l)}) = \sum_{s_l^c} \mathbf{P}(s_l^c) \prod_{t \in b(l):e(l)} \mathbf{P}(x_t | a_t^c). \tag{2}$$

Next, we divide the time sequence $1 : T$ to the segments

$$\{b(1) : e(1)\}, \ \{b(2) : e(2)\}, \ \cdots, \ \{b(L^c) : e(L^c))\}$$

and consider the marginal probability for all the times $1 : T$ as

$$\begin{aligned} &\mathbf{P}(a_{1:T}^c, x_{1:T}) \\ = \mathbf{P}(\quad &a_{b(1):e(1)}^c, x_{b(1):e(1)}, \\ &a_{b(2):e(2)}^c, x_{b(2):e(2)}, \\ &\cdots, \\ &a_{b(L^c):e(L^c)}^c, x_{b(L^c):e(L^c)} \quad) \end{aligned}$$

Assuming any pairs of segments are independent each other, The formula is written as the product of the segment marginal probabilities, as

$$= \prod_{l \in 1:L^c} \mathbf{P}(a_{b(l):e(l)}^c, x_{b(l):e(l)}).$$

Substituting (2),

$$= \prod_{l \in 1:L^c} \{ \sum_{s_l^c} \mathbf{P}(s_l^c) \prod_{t \in b(l):e(l)} \mathbf{P}(x_t | a_t^c) \}$$

Therefore, given the input $x_{1:T}$,

$$\mathbf{P}(a_{1:T}^c | x_{1:T})$$

$$\propto \prod_{l \in 1:L^c} \{ \sum_{s_l^c} \mathbf{P}(s_l^c) \prod_{t \in b(l):e(l)} \mathbf{P}(x_t | a_t^c) \} \tag{3}$$

This formula utilizes not only the local time likelihood $\mathbf{P}(x_t | a_t^c)$ as the traditional approach in Eq. (1), but also the prior probability of the segments $\mathbf{P}(s^c_l)$. We use the local time likelihood $\mathbf{P}(x_t|a^c_t)$ from the result of the naive method, and also prepare and utilize the prior probability $\mathbf{P}(s^c_l)$ using the samples from the training data. Because $\mathbf{P}(s^c_l)$ can be informative when we obtain training data for an entire day, our method can lead to accuracy improvement for activity recognition of an entire day.

Implementation

In the implementation, we calculate Eq. (3) according to the following steps, where we adopt the logarithmic probability to avoid underflows, and exploit the idea of importance sampling to obtain those samples weighted by the prior knowledge of the segments.

1. Train local time log likelihood
$$\log \mathbf{P}(x_t|a^c_t) \quad \text{for each } t \in 1:T$$
with the naive method, and store the results.
2. Construct $\mathbf{P}(s^c_l)$ from the training data. This probability is implemented as a set of k samples from the training data. We represent the i'th sample as $s^c_{l[i]}$, where $1 \leq i \leq k$.
3. For each $s^c_{l[i]}$ in Step 2, calculate
$$\exp\big(\sum_{t\in b(l[i]):e(l[i])} \log \mathbf{P}(x_t|a^c_t)\big) \qquad (4)$$
using the result of Step 1).
4. The average of Eq. (4) for $l[1], l[2], \cdots, l[k]$ can be regarded as importance sampling for Eq. (4), and as an approximation of Eq. (2) accordingly. Because Eq. (2) is the same function for any $l \in 1 : L^c$, we can utilize this function directly to estimate s^c_l rather than completely calculate Eq. (3). In practice, to simplify the calculation of the average for day wise, we can pick up some $l \in 1 : L^c$ with larger Eq. (2) values by a threshold such as the average of Eq. (2).

Note that $\log \mathbf{P}(x_t|a^c_t)$ can be used multiple times for different $s^c_{l[i]}$ in Step 3, and thus they are pre-calculated and stored in Step 1 to avoid redundant calculations.

EVALUATION

In this section, we describe the dataset collected from actual nurses wearing accelerometers in a hospital for approximately two weeks, and we evaluate our proposed method by applying it to this collected data.

Objective

The goal of the evaluation is to answer the following questions:

1. Can the proposed method improves the recognition accuracy?
2. Can the proposed method estimate better segments?
3. Can we obtain knowledge about nursing activities or clinical pathways from the real data?

For Question 1, we evaluate accuracy compared with the naive method indicated in points 1 and 2 of Section "Results". For Question 2, we evaluate the activity durations indicated in point 3 of the same section. Moreover, for Question 3, we discover knowledge about the nursing activities by applying our method to the two years of data collected, and explore correlations with the medical data.

Preprocessing

We extracted feature vectors from the three axes using the accelerometer data. For the sensor data, time windows of 5 seconds were extracted, shifting every 2.5 seconds, as in Bao et al. [1]. For each time window, we calculated 47 feature values, following the literature of [52, 53].

We reduced the 47 feature variables to 27 by applying stepwise-feature selection [13] to 1,000 randomly sampled vectors over ten iterations. The feature variables that were selected are listed in Table 3.

Applying the method

In order to evaluate our proposed method, we compared the *proposed* method with the prior knowledge about $\mathbf{P}(s^c_l)$, and the *naive* method without the prior knowledge. As underlying machine learning algorithms for $\mathbf{P}(x_t|a^c_t)$, which is the same as the naive method after applying the Bayes' theorem, we adopted k-Nearest Neighborhood *(k-NN)*, naive Bayes (*NaiveBayes*) and RandomForest, and evaluated each of them. We adopted a Gaussian distribution for the naive Bayes method, which is a parametric model of probabilities. Because it assumes specific probability function, it may lead to incorrect modelling of the probability. Therefore, we also adopt k-NN, which can non-parametrically approximate the probability by using the powered inverse of distances with the k'th samples, as addressed in many literatures. Random forest does not have such proven approximation, as far as we know, but we can apply Bayes' rule to the majority rate obtained from each weak-decision tree. Random forest is popular and achieves better accuracy in many papers, then we adopted this to demonstrate to use high-performance baseline.

The detail of the methods are described in the following: In order to evaluate the accuracy of real usage where the training and usage data are different, we applied 1-*duty-day*-left-out cross validation, which means testing each nurse's working day with the model trained with the data that have either different days or different nurses.

Evaluation method

To evaluate the proposed method, precisions, recalls, and F-measures for each time window are not necessary for the following reasons:

1. The targeted real data are imbalanced, as discussed previously. Standard measures, such as precision and F-measure, are affected by these imbalances. Thus, it is preferable to use imbalance-independent measures.
2. Activity duration is also important. The traditional measures do not consider the fragmentation of estimated activities. If fragmentation remains in the estimated activity

Table 3. List of feature variables after feature selection

No.	Feature	Sensor	Axis (if any)
1	Mean intensity	Chest	
2	Mean intensity	Right wrist	
3	Mean	Chest	Y
4–6	Mean	Waist	X, Y, and Z
7–9	Mean	Right wrist	X, Y, and Z
10	Variance of intensity	Right wrist	
11–13	Variance	Right wrist	X, Y, and Z
14–15	Variance	Chest	Y and Z
16	Variance	Right wrist	Z
17–18	Mean FFT-domain energy	Chest	Y and Z
19–20	Mean FFT-domain energy	Right wrist	X and Z
21	Mean sum of the absolute values of each axis	Chest	
22	Mean sum of the absolute values of each axis	Waist	
23	Number of samples out of mean intensity $\pm 0.1G$	Right wrist	
24	Number of samples out of mean intensity $\pm 0.1G$	Waist	
25	Number of crosses of the zone of the mean intensity $\pm 0.1G$	Waist	
26	Number of crosses of the zone of the mean intensity $\pm 0.1G$	Right wrist	
27	Covariance between intensities	Chest and Waist	

sequences, this can result in many segments of shorter durations. In order to analyze nursing activities, duration is one of the critical values.

To overcome these problems, we adopted the evaluation methods introduced in this section.

With regard to point 1, we adopted BCR, a measure used by [10, 9] and define as follows:

$$\text{BCR} = \frac{\text{TP-rate} + \text{TN-rate}}{2}$$

where TP-rate is defined as $TP/(FN + TP)$, and TN-rate is defined as $TN/(TN + FP)$, where TP (FP, TN, FN) is the number of true positives (false positives, true negatives, or false negatives, respectively). In contrast with other measures such as recall and F-measure, these values are not affected by imbalanced positive and negative samples at the ground truth level, and accordingly, BCR — the mean of them — is also imbalance independent.

With regard to point 2, we evaluate the difference between the mean durations of the estimated and true labels for each activity. If the value is smaller, the estimated segments have closer durations to the true segments.

Results

Following the evaluation approach discussed above, we explain the results shown in Fig. 6 and Fig. 7. From here, to easily visualize the result, we omit the result of the activity classes for no more than 5 labels (activity class No. 25 and 28) and "Other" class. Note that these samples were used in the evaluation for reality, but just removed when showing the result.

Accuracy by the balanced classification rate

Fig. 6 shows the results for k-NN, NaiveBayes, and RandomForest as the underlying machine learning algorithm.

As we can see from the figure, most of the activity classes improve with our method. Averaging all activity classes, when we adopt k-NN as the underlying algorithm, BCR for naive method is 56.10% ($\sigma = 9.6$), and for the proposed method, it is 73.18% ($\sigma = 14.2$). When we adopt Naive Bayes as the underlying algorithm, BCR for the naive method is 55.15% ($\sigma = 15.8$), and for the proposed method, it is 80.96% ($\sigma = 14.5$). Moreover, when we adopt RandomForest as the underlying algorithm, BCR for the naive method is 59.03% ($\sigma = 17.3$), and for the proposed method, it is 67.83% ($\sigma = 13.4$).

Accuracy of activity durations

Fig. 7 shows the mean error for activity durations for the naive and proposed methods for each activity. Because the y-axis is the error, the smaller the y-axis, the better is the accuracy. From the figure, in any activity class, the proposed method greatly outperforms the naive method. The mean errors are 324.2 seconds for the naive method and 74.6 seconds for the proposed method, with k-NN. When using NaiveBayes, they are 173.5 seconds for the naive method and 90.33 seconds for the proposed method. Moreover, when we use RandomForest, they are 122.2 seconds for the naive method and 7.88 seconds for the proposed method.

Discussion

As a result of the evaluations with BCR, our proposed method outperformed the naive method by 17.08 (25.81, 8.8)% with k-NN (NaiveBayes, Random Forest, respectively). Although the best absolute accuracy of the naive methods is by RandomForest, the best improvement and absolute accuracy of the proposed method is by NaiveBayes, and the second is k-NN. The reason why RandomForest was not improved so much would be because the probability modeling of it is not perfect, compared with other probabilistic methods.

On the other hand, the accuracy of activity durations was the best in RandomForest even in the absolute error value. Compared with the prior knowledge about the timestamps of segments, that about the activity durations seems to be effective in any underlying algorithms.

Although we achieved improvements for BCR, further work for other types of improvements such as precision and F-measure, are important. This is inherently difficult to achieve, for example, prediction of disasters or diseases that hardly

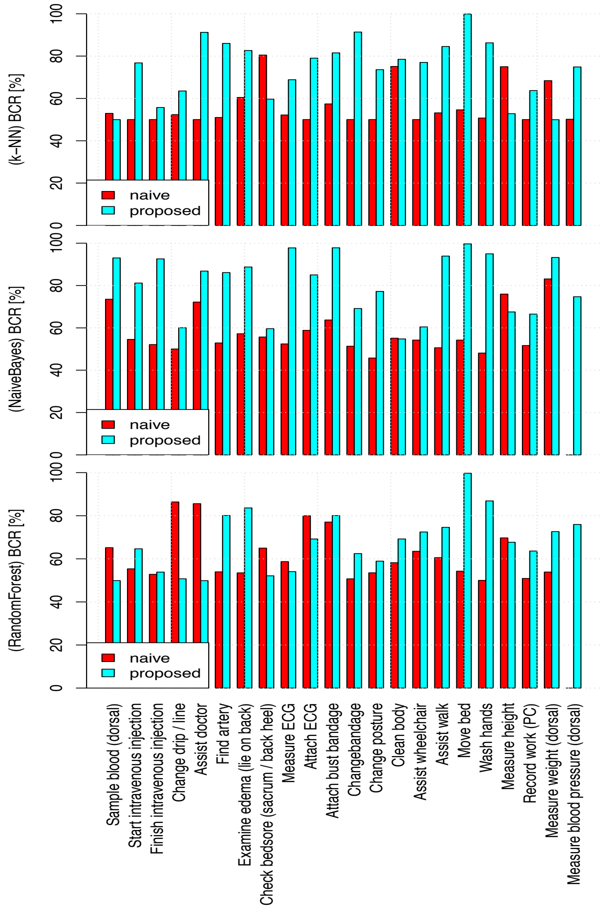

Figure 6. BCR for naive / proposed methods for each activity with k-NN (NaiveBayes, RandomForest) (Average: 56.10 (55.15, 59.03)% for the naive method and 73.18 (80.96, 67.83, respectively)% for the proposed method.)

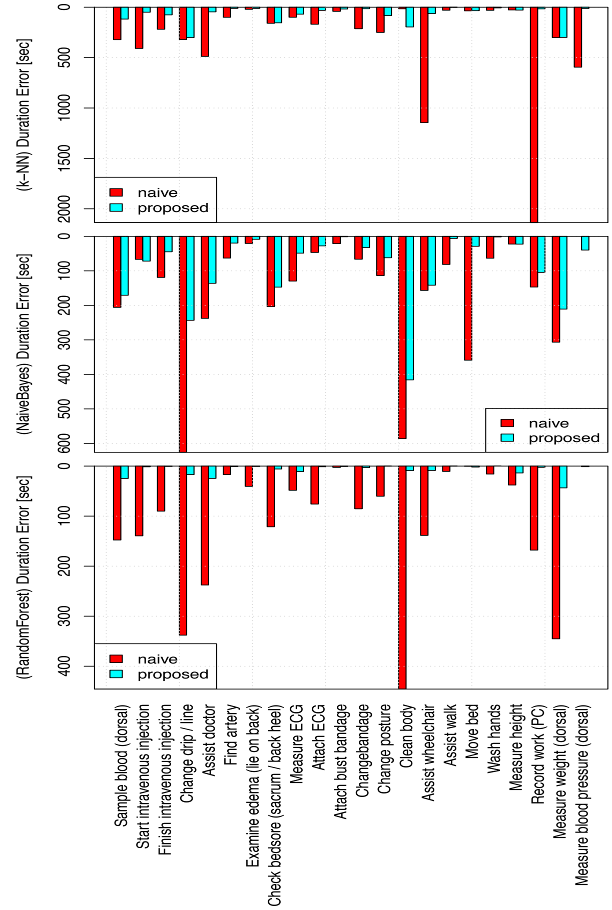

Figure 7. Errors for activity durations for k-NN (NaiveBayes, RandomForest) (Average: 324.2 (173.5, 122.2) seconds for the naive method and 74.6 (90.33, 7.88, respectively) seconds for the proposed method.)

occur, but other approaches, such as feature engineering, and considering state-transition probabilities, such as [48] can be effective.

Instead of the prior knowledge about the timestamps, it is possible to use the time-of-day as a feature. However, the prior knowledge about the activity duration cannot be utilized. Since the activity duration is only known when the segment is defined, it is not applicable for the features in the traditional method. The activity durations are drastically improved on our method as in Fig. 7, it would be an advantage of our method.

We assume that we can obtain a multiple activity classes simultaneously. If we assume that we can restrict to a single activity class at a time, the problem is more difficult. Approaches such as optimizing multi-class ROC [14, 43] can be the candidates for solving this problem.

In this paper, we adopt k-NN, NaiveBayes, and RandomForest as the underlining algorithms. Nonetheless, our approach can be used as a post-process of any type of estimation algorithm that can output local-time likelihood.

Applying to bigger data

Using our method, we demonstrate an example of recognizing and analyzing bigger datasets, such as correlation with nurses' experience, correlation with patients' levels of nursing needs, and the relationship between delays of discharges.

For the unlabelled data, we extracted 265,002 time windows, which corresponded to 771 duty days x nurses, and applied our proposed method in order to estimate the real activities involved in nursing duties. Although RandomForest can be used to deduce higher accuracy, to balance the accuracy of BCR and activity durations, and to make the probability distribution of the result natural, we adopted the k-NN algorithm instead of NaiveBayes and RandomForest.

Nursing times in a day

For 658 daytime duties, the average time for the defined care time is 277.8 minutes with $\sigma = 55.7$.

Fig. 8 is the estimated care times for each activity class in one daytime. From the figure, we can see the types of activity on which the nurses spend more time, such as "Examine edema", and "Measure blood pressure". We can also see that the nurses spend significant time recording their work on a

PC, which were introduced after the electronic medical record system was introduced, and hence there is an opportunity for reducing this time.

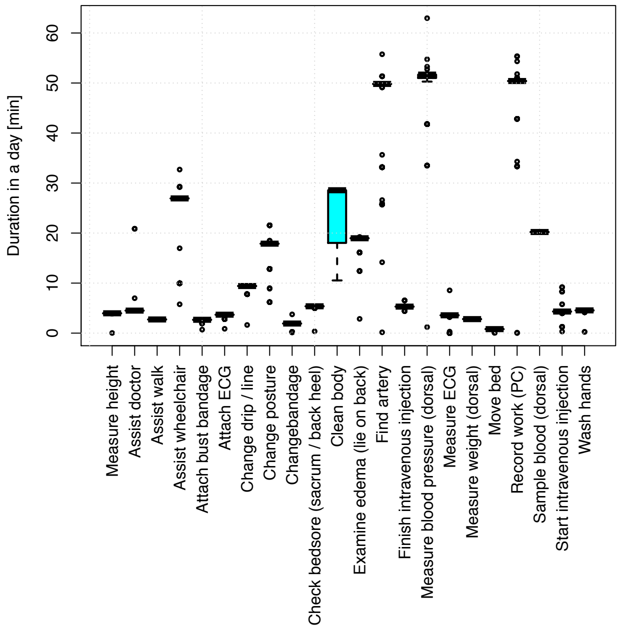

Figure 8. The nursing times for each activity class in one daytime

Correlation with the nurse' experience

If we join the results with additional data, such as nurse' profile, we can data mine further knowledge. To demonstrate this, we joined the results with the number of experienced years for each nurse. The mean experienced years is 7.36 years with $\sigma = 5.65$, the minimum is 1, and maximum is 25.

Then, we applied Poisson regression to explain the experienced years with the duration of each activity in a day. Table 4 lists the activities whose total time in a day is statistically significant by Poisson regression under the level of $p = 0.05$. From the table, "Measure blood pressure" can explain the experienced years with a coefficient of $\exp(0.002) = 1.002$, "Examine edema" can explained with a coefficient of $\exp(0.046) = 1.005$, and "Record work (PC)" can explain with $\exp(-0.003) = 0.997$. Statistically, this implies that experienced nurses measure blood pressures and examin edema slightly longer, and spend slightly less time for recording work with a PC.

Table 4. Nursing activities whose total time in a day is statistically significant by Poisson regression under the level of $p = 0.05$.

	Coefficient	Std. Error	p-value
Measure blood pressure	0.002	0.00	0.00006
Examine edema	0.046	0.02	0.00393
Record work (PC)	-0.003	0.00	0.00471

Correlation with the level of nursing needs

In a hospital, patient status is assessed in a standardized manner, based on the level of *nursing needs*. After joining the patient nursing needs data with the amount of time spent by nurses for each activity by dates, we applied Poisson regression to explain the nursing needs with the duration of each activity in a day. The number of patients is 55 and the number of samples after joining with the nurse activities is 189.

As a result, "Measure blood pressure" could explain the nursing needs with coefficient of $\exp(0.017) = 0.98$. This implies that patients with high nurseing needs take slightly less time for blood pressure and could spend more time for other activities.

Differences with regard to discharge delays

We compared the amount of time spent by nurses for each activity between those who had duty dates when there are patients with delay in the discharges (more than 4 inpatient days) and those who had duty dates when there are patients with no delay. The number of patients after joining with the patient record is 24 and the number of samples is 155.

As a result, the amount of time for "Finish intravenous injection" in a day has statistically significant difference between delayed patients and those with no delay with $p = 0.019$. The mean time spent in this activity is 221.0 minutes ($\sigma = 0$) for delayed patients, and 203.6 minutes ($\sigma = 58.64$).

We can observe that intravenous injection takes longer times for patients who cause delays of discharges.

As shown in this section, by linking our proposed method with additional data which already exist in hospitals, we can produce a valuable knowledge for reflecting and improving medical processes.

RELATED WORK

In the literature, many works attempted mobile activity recognition [8, 44, 50, 26, 28, 31, 24, 3, 30, 1, 39]. A few papers also attempted to apply nursing activity recognition and application [34, 45, 36].

Because activity recognition manages sequential data, techniques for sequential data such as Hidden Markov Model (HMM) [7, 24] and Conditional Random Fields (CRF) [17, 27, 25, 7], used in speech recognition and natural language processing, are related. Some works have attempted to apply these techniques to mobile activity recognition [46, 47, 42, 51]. Here, we claim that using HMM and CRF are independent of our contribution. Basically, HMM and CRF are not segmenting wise, but the time window wise if we use them straightforwardly. Then, we can apply our method to utilize prior knowledge independently. Applying HMM and CRF for segment wise is not straightforward since they are not determined from the first. And, HMM and CRF are complex to estimate the parameter, but our method can simply integrate and utilize other popular methods of non-sequential machine learning.

Another approach that is applicable to sequential data is Bag-of-Features (BoF), which makes histograms of feature values and utilizes their statistic features [53, 52]. However, this can only be applicable to data that is already segmented. The segmentation technique is common in speech recognition [11, 4] and natural language processing [33, 6, 49, 32].

However, among the aforementioned work, to the best of our knowledge, none addresses the challenges real-world applications, nor try to utilize the prior knowledge on the daily basis, such as our method. Class-wise prior probability, timestamps in a day, and activity durations have large variances. These can difficulties in activity recognition when applying the existing work.

With regard to activity durations, [48] adopts the CRF model that can integrate the knowledge of activity duration using Semi-CRF, which learns segmentation in addition to the Markov transitions, as well as the traditional CRFs. Moreover, it improves computation costs by considering omitting "other" activities. This work generates promising results in accuracy, although the computation and parameter estimation often becomes complex in such high-dimensional approach. Our method manages the duration and segments as a prior knowledge obtained from the training dataset, and infers the activities considering them by Bayesian network and importance sampling approaches, which is demonstrated to be tractable in real nursing big data.

In addition, the challenge lies in recognizing complex domain-specific activities such as nursing activities which, we resolve in our paper.

For machine learning from imbalanced data, problems and approaches are addressed in the literature. Classification for imbalanced data is highly important in the area of risk management, such as medical decision domains, where a positive instance, such as a specific disease, hardly occurs. [16] introduced several assessment metrics, such as ROC that is robust for imbalanced data, and reviewed several approaches, such as importance sampling, cost-sensitive methods, and active learning. It also addresses the effectiveness of *one-class learning*, a binary classification of positive or negative. [22] applied empirical evaluation of RandomForest algorithm for imbalanced data. [18] proposed sampling method for bagging, and evaluated their method using AUC. [23] evaluated several boosting and bagging algorithms comprehensively for noisy and imbalanced data, and concluded that bagging generally outperforms boosting. In this paper, we incorporate the robustness of one-class learning to our method.

In the literature, several datasets for mobile activity recognition are available. [15], with their large-scale activity collection, collected over 35,000 activities from more than 200 people over approximately 13 months. [19] provided a dataset that consists of 28 days of sensor data from a single person with annotations added by their proposed system. [20] was a unique trail to collect activity recognition datasets from the laboratories of multiple universities. In the 5 years, the total number of activities reached over 50,000 samples. [5] provided a dataset with varieties of sensor displacement status for 33 fitness activities from 17 participants. [2, 38] provided an activity dataset with sensor-rich environment where the subjects wore multiple sensors on the body, with more than 27,000 activities from 12 subjects. Among them, [2, 5, 38] provided an entire day data / multi-day data as a part of them. However, the activity classes are common types, such as those that appear in Activity in Daily Life (ADL) records, and not similar to our dataset, which is closely coupled with the application domain and domain data, such as medical records.

CONCLUSION

In this paper, we collected a real nursing dataset for mobile activity recognition that can be used for supervised machine learning, and proposed a method for recognizing activities for an entire day utilizing prior knowledge about the activity segments in a day. The results showed accuracy improvement compared with the baseline method that did not employ our method; in particular, there were significant improvements in activity durations. It implies that the dataset are valid, and that the proposed method is confident.

We also demonstrated data mining by applying our method to bigger data combined with 2 years of patient medical records, and demonstrated the value of linking with additional day. The future work includes expanding the data mining in order to explore the knowledge about clinical paths, such as finding important activities that lead to earlier discharge from the hospital.

Because activity recognition in nursing domain is new and challenging, there is no statement or reference how much accuracy is required, and our method cannot be proven whether they have enough accuracy. However, we believe the result of the paper can be a reference of how challenging it is, and moreover, we claim that we could achieve non-negligible improvement for the durations of activities, and demonstrated the durations could be used for nursing activity analysis.

The data we used were collected carefully to be used as open data by obtaining agreements from the subject nurses and patients. The data are also related to RFID tag data in order to recognize nurses' entry into patients' rooms, vital data from hospitalized patients (e.g., cardiograms, bed sensors to measure heart rate/breathing/body movements), accelerometer, in-room sensors, and medical information recorded in the electronic clinical pathways, and indirectly, inpatient sensor data. As future work, data mining these whole data combined with the activity recognition result and extracting valuable knowledge which contributes to efficient clinical pathways and better health care will be important.

Acknowledgment

The part of this research are supported by Funding Program for World-leading Innovation R&D on Science and Technology (FIRST) "Developement of the Fastest Database Engine for the Era of Very Large Database and Experiment and Evaluation of Strategic Social Services Enabled by the Database Engine (Main Researcher: Masaru Kitsuregawa)", and Grant-in-Aid for Scientific Research (B) "Sensor Context Estimation Techniques by Semantic and Physical Layer (Principal Investigator: Sozo Inoue)". The authors would like to thank their support. We also appreciate the cooperation for experiment by the staff of Saiseikai Kumamoto Hospital, Japan.

REFERENCES

1. Bao, L., and Intille, S. S. *Pervasive Computing*, vol. 3001 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
2. Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Troster, G., Millan, J. D. R., and Roggen, D. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters 34* (2013), 2033–2042.
3. Chen, L., Hoey, J., Nugent, C. D., Cook, D. J., and Yu, Z. Sensor-based activity recognition, 2012.
4. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. Natural Language Processing (almost) from Scratch. *Journal of Machine Learning Research 12* (2011), 2493–2537.
5. Damas, M., Rojas, I., Amft, O., Toth, A. M., Baños, O., Pomares, H., and Tóth, M. A. A benchmark dataset to evaluate sensor displacement in activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12* (2012), 1026.
6. Demuynck, K., and Laureys, T. A comparison of different approaches to automatic speech segmentation. In *Text, Speech and Dialogue* (2002), 277–284.
7. Dietterich, T. Machine learning for sequential data: A review. *Structural, syntactic, and statistical pattern recognition* (2002), 1–15.
8. Farringdon, J., Moore, A., Tilbury, N., Church, J., and Biemond, P. Wearable sensor badge and sensor jacket for context awareness. *Digest of Papers. Third International Symposium on Wearable Computers* (1999).
9. Fawcett, T. ROC Graphs : Notes and Practical Considerations for Researchers. *ReCALL 31* (2004), 1–38.
10. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters 27* (2006), 861–874.
11. Franco, A., and Destrebecqz, A. Chunking or not chunking? How do we find words in artificial language learning? *Advances in Cognitive Psychology 8* (2012), 144–154.
12. Gooch, P., and Roudsari, A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *Journal of the American Medical Informatics Association : JAMIA 18* (2011), 738–48.
13. Guyon, I., and Elisseeff, A. An introduction to variable and feature selection. *Journal of Machine Learning Research 3* (2003), 1157–1182.
14. Hand, D., and Till, R. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning* (2001), 171–186.
15. Hattori, Y., Inoue, S., and Hirakawa, G. A Large Scale Gathering System for Activity Data with Mobile Sensors. *2011 15th Annual International Symposium on Wearable Computers* (2011), 97–100.
16. He, H., and Garcia, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering 21* (2009), 1263–1284.
17. He, X. H. X., Zemel, R., and Carreira-Perpinan, M. Multiscale conditional random fields for image labeling. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. 2* (2004).
18. Hido, S., Kashima, H., and Takahashi, Y. Roughly balanced Bagging for Imbalanced data. *Statistical Analysis and Data Mining 2* (2009), 412–426.
19. Kasteren, T. V., Noulas, A., Englebienne, G., and Kr, B. Accurate Activity Recognition in a Home Setting. In *UbiComp '08 Proceedings of the 10th international conference on Ubiquitous computing* (2008), 1–9.
20. Kawaguchi, N., Ogawa, N., and Iwasaki, Y. Hasc challenge: gathering large scale human activity corpus for the real-world activity understandings. In *Proceedings of the 2nd Augmented Human International Conference* (2011), 27.
21. Kelder, T., Conklin, B. R., Evelo, C. T., and Pico, A. R. Finding the right questions: Exploratory pathway analysis to enhance biological discovery in large datasets. *PLoS Biology 8* (2010), 11–12.
22. Khoshgoftaar, T. M., Golawala, M., and Hulse, J. V. An Empirical Study of Learning from Imbalanced Data Using Random Forest. *19th IEEE International Conference on Tools with Artificial Intelligence(ICTAI 2007)* (Oct. 2007), 310–317.
23. Khoshgoftaar, T. M., Van Hulse, J., and Napolitano, A. Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans 41* (2011), 552–568.
24. Kim, E., Helal, S., and Cook, D. Human Activity Recognition and Pattern Discovery. *Pervasive Computing, IEEE 9* (2010), 48–53.
25. Klinger, R. Classical Probabilistic Models and Conditional Random Fields. *Entropy 51* (2007), 282–289.
26. Kwapisz, J. R., Weiss, G. M., and Moore, S. A. Activity Recognition using Cell Phone Accelerometers. *Human Factors 12* (2010), 74–82.
27. Lafferty, J., McCallum, A., and Pereira, F. C. N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning* (2001), 282–289.
28. Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., and Campbell, A. T. A survey of mobile phone sensing. *IEEE Communications Magazine 48* (2010), 140–150.

29. Lemmens, L., Van Zelm, R., Vanhaecht, K., and Kerkkamp, H. Systematic review: Indicators to evaluate effectiveness of clinical pathways for gastrointestinal surgery, 2008.

30. Lymberopoulos, D., Bamis, A., and Savvides, A. Extracting spatiotemporal human activity patterns in assisted living using a home sensor network, 2011.

31. Mannini, A., and Sabatini, A. M. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors 10* (2010), 1154–1175.

32. McQueen, J. M. Segmentation of continuous speech using phonotactics. *Journal of Memory and Language 39* (1998), 21–46.

33. Mohri, M., Moreno, P., and Weinstein, E. Discriminative Topic Segmentation of Text and Speech. *Artificial Intelligence 9* (2010), 533–540.

34. Naya, F., Ohmura, R., Takayanagi, F., Noma, H., and Kogure, K. Workers' Routine Activity Recognition using Body Movements and Location Information. *2006 10th IEEE International Symposium on Wearable Computers* (2006), 105–108.

35. Nohara, Y., Sozo, I., Nakashima, N., Naonori, U., and Kitsuregawa, M. Large-scale Sensor Dataset in a Hospital. In *International Workshop on Pattern Recognition for Healthcare Analytics* (Tsukuba, Japan, 2012), 4 pages.

36. Osmani, V., Balasubramaniam, S., and Botvich, D. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of Network and Computer Applications 31*, 4 (Nov. 2008), 628–655.

37. Panella, M., Marchisio, S., and Di Stanislao, F. Reducing clinical variations with clinical pathways: Do pathways work? *International Journal for Quality in Health Care 15* (2003), 509–521.

38. Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Forster, K., Troster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., Doppler, J., Holzmann, C., Kurz, M., Holl, G., Chavarriaga, R., Sagha, H., Bayati, H., Creatura, M., and Millan, J. d. R. Collecting complex activity datasets in highly rich networked sensor environments. *Networked Sensing Systems (INSS), 2010 Seventh International Conference on* (2010).

39. Roggen, D., Troster, G., Lukowicz, P., Ferscha, a., Millan, J. D. R., and Chavarriaga, R. Opportunistic human activity and context recognition. *Computer 46* (2013), 36–45.

40. Rotter, T., Kinsman, L., James, E., Machotta, A., Gothe, H., Willis, J., Snow, P., and Kugler, J. Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs. *The Cochrane database of systematic reviews* (2010), CD006632.

41. Saraiya, P., North, C., and Duca, K. Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Information Visualization 4* (2005), 191–205.

42. Sun, X., Kashima, H., Tomioka, R., and Ueda, N. Large scale real-life action recognition using conditional random fields with stochastic training. *Advances in Knowledge ...* (2011), 222–233.

43. Tang, K., Wang, R., and Chen, T. Towards Maximizing the Area Under the ROC Curve for Multi-Class Classification Problems. *AAAI*, Elkan 2001 (2011), 483–488.

44. Tapia, E. M., Intille, S. S., Haskell, W., Larson, K., Wright, J., King, A., and Friedman, R. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *Proceedings - International Symposium on Wearable Computers, ISWC* (2007), 37–40.

45. Tentori, M., and Favela, J. Monitoring behavioral patterns in hospitals through activity-aware computing. In *2008 Second International Conference on Pervasive Computing Technologies for Healthcare*, IEEE (Jan. 2008), 173–176.

46. Vail, D., Veloso, M., and Lafferty, J. Conditional random fields for activity recognition. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, vol. 5 (2007), 1–8.

47. Vail, D. L., Lafferty, J. D., and Veloso, M. M. Feature selection in conditional random fields for activity recognition. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2007), 3379–3384.

48. Vinh, L. T., Lee, S., Le, H. X., Ngo, H. Q., Kim, H. I., Han, M., and Lee, Y. K. Semi-Markov conditional random fields for accelerometer-based activity recognition. *Applied Intelligence 35* (2011), 226–241.

49. Wang, D., Lu, L., and Zhang, H.-J. Speech segmentation without speech recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03).*, vol. 1 (2003), 468–471.

50. Ward, J. A., Lukowicz, P., Tröster, G., and Starner, T. E. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28* (2006), 1553–1566.

51. Zhan, K., Faux, S., and Ramos, F. Multi-scale Conditional Random Fields for first-person activity recognition. *Pervasive Computing and ...* (2014).

52. Zhang, M., and Sawchuk, A. Motion primitive-based human activity recognition using a bag-of-features approach. *Proceedings of the 2nd ACM SIGHIT ...*, 1 (2012), 631.

53. Zhang, M., and Sawchuk, A. A. A feature selection-based framework for human activity recognition using wearable multimodal sensors. In *Int. Conf. Body Area Networks* (2011), 92–98.