Exploring Combinations of Missing Data Complement for Fault Tolerant Activity Recognition

Ren Ohmura
Toyohashi University of Technology
15-1, Hibari-ga-oka, Tenpaku-cho,Toyohashi, Aichi, Japan
ren@tut.jp

Ryoma Uchida
OMRON SOFTWARE Co., Ltd.
96, Maeda, Higashiachiwa-cho, Okazaki, Aichi, Japan
uchida.ryoma@gmail.com

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA

http://dx.doi.org/10.1145/2638728.2641311

Abstract
Disrupting the transmission of sensor data due to sensor failure or connection loss significantly decrease accuracy in existing activity recognition techniques. We introduce an approach towards managing missing sensor data which operates at each step of the standard activity recognition, beginning with raw sensor data, feature calculation, classification, and result, as well as their combination methods. Our evaluation showed that the F_1-score increased from 0.61 in the case of sensor data loss to 0.68 with the combination of all methods. Moreover, by selecting the combination of methods according to the failed sensor position, the F_1-score increased to 0.69.

Author Keywords
Activity Recognition; Fault tolerance; Missing data complement;

ACM Classification Keywords
H.1.2. [User/Machine Systems]: Human Information Processing

Introduction
Accurate activity recognition with wearable sensors is crucial for enabling context-aware services. Since sensors equipped on several body part can detect the movement of each, usage of multiple sensors can enhance the

applicability and improve its accuracy. However, existing techniques are insufficient in its fault tolerance. When sensor or communication failure happens and sensor data cannot be obtained, the accuracy of activity recognition is significantly decreased[4], or the recognition result itself cannot be produced because the set of sensor data differs from the one that train the system. In our experience of obtaining nurses' activity data in a real hospital, in which four inertial sensors are used for each nurse, there were one or more sensor data missing in 330 minutes of 810 minute data in total, which is equivalent to 40% approximately.

One of the possible approaches to cope with this problem is to complement missing data. This is possible at the each step of activity recognition, raw data, feature values, classification, and the result. For instance, Murao *et al.* proposed a complement method of raw sensor data[5]. However, the purpose of this study is to decrease power consumption by intentional turning off of a sensor. Sagha *et al.* proposed a method to complement the output of weak classifiers trained with missing data in an ensemble learning method[6] for enabling activity recognition with several sensor settings.

These methods are in just one step and do not consider a combination with complement in other steps. Combining several types of method in each step, it is expected to improve accuracy of missing data complement. Moreover, it is expected to broaden applicability of complement to various errors, such as temporally consecutive (burst) error and intermittent error.

Thus, this study explores complement methods in each step of general activity recognition and their combinations. More concretely, ARAR method[3] based on temporal correlation in raw data step, kernel regression method based on spatial correlation in feature calculation step, simplified Sagha's method[6] based on spatial correlation in classification step, transition probability of activities, which is equivalent to temporal correlation, in result step are evaluated. This study also introduces a method to combine these methods and shows results of sixteen types of combinations for coping with sensor data missing in activity recognition.

From next section, this paper describes our approach and concrete methods. Then, this paper shows the results of evaluations and discusses them.

Approach

In general, current activity recognition techniques consist of four steps[1], obtaining sensor data, calculating feature values, classification (comparing calculated feature values with a model created with labelled sensor data), and classification result. Methods to cope with missing data can be put along four these steps, raw data, feature extraction, classifier, and result.

The basics of complement is to estimate a missing (or new) value based on a model, which describes the relations between the target value and one or more existing values. Since activity recognition operates sequential and multi-variable data, spatial correlation and temporal correlation of data can be used for creating a model as its relation. Here, spatial correlation is the relationship between (sensor) data and other (sensor) data simultaneously obtained. Temporal correlation is the relationship between current (sensor) data and past data in continuously obtained from one sensor.

[1]This can be thought as five when pre-process for raw data, such as filtering and applying the window function is counted.

A complement method based on spatial correlation has an advantage that it can easily follow the change of user's activity especially when the activity frequently changes. This is because the method doesn't depend on previous activity and its sensor data. However, it has a drawback that it requires multiple sensors and thus cannot be applied to a situation in which only one sensor is used, such as using only one smart phone.

On the other hand, a method based on temporal correlation can be applied on such a one-sensor situation. However, the method basically assumes that similar data continues, and accuracy of the complement can significantly decrease when data are missing for a long time (burst error) and user's activity changes during the time.

Considering these advantages and disadvantages, we choose a method applying each step for having a balance of temporal correlation and spatial correlation. Thus, methods based on the temporal correlation method are used for raw data step and result step, and methods based on spatial correlation are used for feature calculation step and classification step.

As mentioned above, a complement method creates a model to estimate a target value with the other values, expecting strong correlation exists between these in temporal or spatial. However, sensor data in activity recognition are difficult to have strong spatial correlation. This is because the correlation is based on one-versus-one relation between the target value and the other values, but a certain value on a sensor can occurs in various activities producing various values on other sensors. For instance, imagine that the value of the accelerometer on right wrist is 1000 mN/sec. The value can happen in various timing of various activities. This is because the granularity of sensor data is too fine, and is also one of the reasons for classification algorithm in general activity recognition to use extracted feature values instead of raw sensor data. Moreover, a complement method basically estimates one value at a time. However, the number of sensor data used in feature calculation, which is the window size of the sliding window, is more than tens in minimum and sometimes gets more than hundreds. When using spatial correlation in raw sensor data, it requires to repeat the same number of times as the window samples and cause high computational load. On the other hand, a lot of complement methods based on temporal correlation also requires to calculate repeatedly as the same times of the window samples, but there are many cases having a reduction of computational load because they are originally developed to predict new future data temporally continues. Thus, a strategy to use temporal correlation is chosen for raw sensor data.

For result step, since an activity recognition system uses one classifier and get one result [2] in general, spatial correlation cannot be applicable. Thus a method based on temporal correlation is used in result step.

For feature step, both of temporal and spatial correlation can be applied. However, we choose spatial correlation here. The first reason is that the model based on spatial correlation can be expected to be accurate. The sort of feature values are chosen as their distributions in each activity become narrow and don't overlap between different activities, and this gives an expectation of near one-versus-one relation between feature values in one activity, which corresponds to user's current activity. The

[2] There are some classification algorithms which output multiple classes with their order and/or probabilities, but the set of the classes produced by one classifier can be thought as just one result.

second reason is to have a balance of these correlations. Thus, a method based on spatial correlation is chosen for feature values.

Both of temporal and spatial correlation can be applied for classification step as well. To begin with, usage of correlation in classification step requires to consider slightly particular ways of classification. For considering complementing missing data independently of the other steps, a simple way is the usage of ensemble learning because it provides independent result of each set of feature values, which is the input of classification step and might be missing partly. When using ensemble learning, what should be thought is how to complete the output of weak classifiers using missing data. In this situation, output of weak classifiers can be expected not so varied in one activity, similar to feature step. Thus, a method of spatial correlation is used in classification step.

With these selections, sixteen combinations exist where the complement method is applied or not. The experiments described in Section shows the results of each accuracy on these.

Complement Method in Each Step

Before showing the results of experiments, this section gives the detail of method used in each step.

Raw Data Complement

For raw data, ARAR algorithm[2] is used for modelling the temporal correlation. Leaving the detail of the algorithm, it recursively applies auto regression to sequential data to create its model.

Temporal models more widely used, such as ARMA, assume that the source of data has a certain rule, For instance, ARMA model assumes that data source is illustrated with repeating short term movement represented by auto regression (AR) and repeating long term movement represented by moving average (MA). On the other hand, ARAR algorithm doesn't assume a certain rule on data source. It is suitable for using the ARAR model for activity recognition because it is difficult to define a certain temporal rule on activity data.

It is reported for ARAR model to have successful results for forecasting future data in some fields[3]. Moreover, our preliminary study compared the recognition accuracy of complementing with ARAR model and ARMA model. The result confirmed ARAR model got better result than ARMA model. Thus, ARAR algorithm is chosen for raw data level complement in this study.

Feature Complement

In feature level, kernel regression is used for modelling the spatial correlation between feature values.

Kernel regression is the combination of kernel method and multiple regression analysis. Kernel method is a technique frequently used in pattern recognition in combination with other algorithms, such as distinction analysis. While kernel method maps data into a high dimensional space, a linear model is applied to the mapped data. As the result, the linear model can be converted into a corresponding non-linear model in the original space. Here, multiple non-linear regression model is obtained by kernel regression.

Our preliminary experiments showed that kernel regression with polynomial kernel could complement missing feature values more accurately than multiple regression. Thus, kernel regression model is used in this study.

Classification Algorithm Level
In classification algorithm, we use our own method simplified from the one proposed by Sagha[6], which is based on spatial correlation.

As mentioned above, the original method uses ensemble learning, and some weak classifiers cannot output the result when one or more data are missing. The output of them are estimated using correlation with the other available classifiers' output. For that, the original method calculates the inverse of co-variance matrix, which sometimes requires high computational load. Our method simplifies this process with conditional probability of normal distribution.

Assuming $p(x_1, x_2)$ is bi-variate normal distribution, conditional probability, $p(x_2|x_1)$, get normal distribution as well. Here, estimated value of x_2 is calculated by the following equation.

$$x_2 = u_{2|1} = u_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - u_1) \tag{1}$$

$$\rho = \frac{\sigma_{12}}{\sigma_1^2\sigma_2^2} \tag{2}$$

where u and σ are mean and standard deviation, respectively.

In our method, u, σ and ρ of each weak classifier output are calculated in the training phase. When data is missing, disrupted output of weak classifier is calculated with Equation 1. Multiple x_2 of a disrupted weak classifier is calculated with u_1 and u_2, and ρ between the disrupted classifier and all available classifiers, which gives x_1 as current value respectively. Then, the mean of multiple x_2 is used as the output of the disrupted classifier. This calculation is done for all disrupted classifiers. All output of weak classifiers including complemented one are used for obtaining the classification result.

Result Level
In result level, temporal correlation is used. The method is a transition probability like N-gram, which is denoted by $p(x_2|x_1)$, where x_2 is current classification result and x_2 is previous one. In our method, the probability is sequentially updated when no sensor data is missing. When missing data exists, the current conditional probability is used to estimate the disrupted result.

Combination Method
Four methods described above are combined as follows. Since each method performs independently and outputs their own results (except the feature level and classifier level complement when using only one sensor). The results are combined as follows.

$$C = C^{\underset{n}{\operatorname{argmax}}\,(P_s^n \times P_f^n \times P_c^n \times P_r^n)} \tag{3}$$

where P_s^n, P_f^n, P_c^n, and P_r^n are given probability for nth class, C^n, of output from raw data complement, feature complement, classifier complement, and result complement, respectively. Namely, the joint probability of each class is calculated with the output of each complement method, and then the class with the highest probability is selected as the recognition result.

Evaluation

With complement and combination methods described above, we evaluated the recognition result on each combination. The combination pattern is shown in Table 1, where the check mark indicates that corresponding method is used. The label on the table identifies the corresponding (combination) method in the remaining part of this paper.

Label	Raw Data	Feature	Classifier	Result
I-a	✓			
I-b		✓		
I-c			✓	
I-d				✓
II-a	✓	✓		
II-b	✓		✓	
II-c	✓			✓
II-d		✓	✓	
II-e		✓		✓
II-f			✓	✓
III-a	✓	✓	✓	
III-b	✓	✓		✓
III-c	✓		✓	✓
III-d		✓	✓	✓
IV	✓	✓	✓	✓

Table 1: Combination of complement methods

Experiment Setting

In experiments, five subjects wore six WAA-006s by ATR-Promotions[1] on their both wrists, both ankles, chest pocket and hip. They were asked to do seven kinds of activities, "Standing", "Walking", "Running", "Skipping", "Sitting", "Clapping hands", and "Folding arms". While they were doing these activities in two trials, where one was about 10 minutes for training and the other one was about 30 minutes for evaluation, tri-axial accelerometer and tri-axial gyro data were sampled in 100Hz. Applying rectangular window of 2560ms width shifting by 1280ms to obtain subsequences of sensor data, mean and standard deviation of each subsequence data were calculated as feature values. 72 features, 6(places)$\times 2$(sensors) $\times 3$(axes) $\times 2$(features), in total are used for the classification. SVM with polynomial kernel was used except a case using complement in classification step.

Evaluation Method

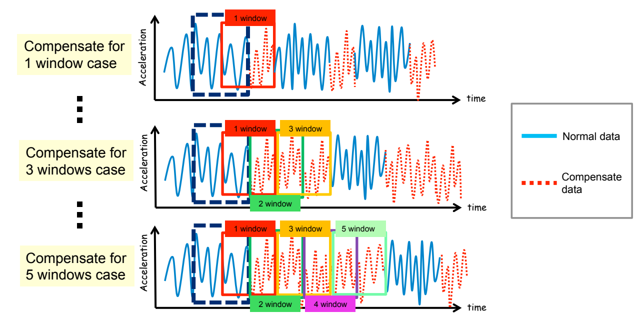

Figure 1: Artificial sensor failure

Assuming that one of six sensors fails, each recognition accuracy was evaluated in fifteen cases shown in Table 1 and a case of no complement. In no complement case, the classifier was trained with five sensor data excluding one sensor which was assumed as being failed. In evaluating the effect of failure duration, which was especially for evaluating temporal correlation based methods, the duration of artificial sensor failure was varied from one window (2540ms) to five windows (7620ms, considering shift size) as shown in Figure 1.

Result

Figure 2 shows the F_1-Score, which is calculated by the following equation, of each case.

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

In the Figure 2, "Normal" is the case that no failure happens, and "Reduced" is the case of no complement method is applied, which is described above. ("Selected" is the case described below.) From I-a to IVare corresponding to the Table 1.

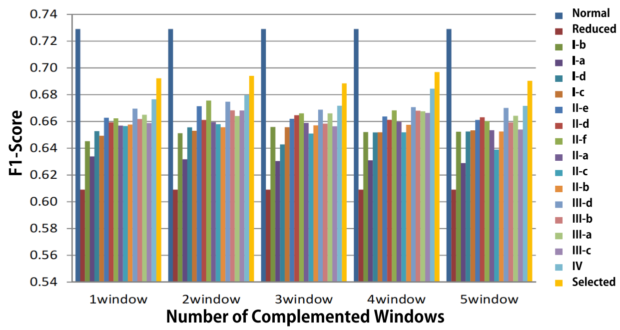

Figure 2: Overall Result

First, Figure 2 shows the result forms are almost same of each window number. In other wards, the results are not depending on the duration of data missing so much. The overall F_1-score decreases from 0.73 in the case of no data missing to 0.61 in the case of no complement method. Meanwhile, each complement method increase the F_1-score, and the score increases along the number of applying methods approximately. Usage of all complement methods, the case IV, shows the best result. Thus, it is a suitable method for the fault tolerance, unless the type of sensor failure (position of failed sensor and length of failure) is considered. In this case, the score increases from 0.61(no complement) to 0.68.

When looking at the cases of only one complement method, from I-a to I-d, all methods except I-a, which is raw data complement, are almost same. In detail, while I-b, feature complement, has low score in the case of 1 window, and I-d, result complement, has low score in 3 window. Considering the stability of I-c, classifier complement is better choice, if only one method has to be chosen. Or, since I-d, result complement, gets better result in the case of 1 and 2 window than I-c, result complement may be also chosen.

In combinations of two methods, from II-a to II-f, II-d, II-e, and II-f give relatively better result, although the best one differs on failure duration as well as the cases using only one complement method. II-d, II-e, and II-f are combinations of feature complement, classifier complement, and result complement, and these better results are consistent with the cases in which only one method is used. In these three, II-f shows the best performance in the cases of 2, 3, and 4 window. Thus, the combination of classifier complement and result complement can be a good selection. Or, II-d, feature and classifier complement, can be alternative because of its stability (no dependency on failure duration).

In combinations of three methods, from III-a to III-d, III-d shows outstanding result and can be the best choice. III-d is the combination of feature, classifier and result complement, and this is consistent with discussions above. III-a, the combination of raw data, feature and classifier, shows the second score except in the case of 2 windows.

From these discussions, while the most effective method is the combination of all method, the effectiveness order of four methods can be 1) classifier complement, 2) result complement, 3) feature complement, and 4) raw data complement. Thus, we can choose each complement algorithm to implement considering this order.

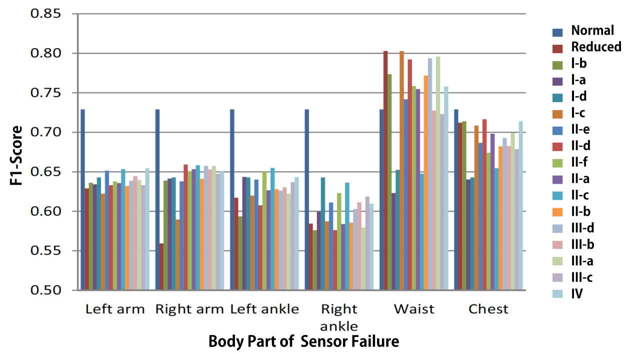

Figure 3: Result of Each Position (3 windows)

For considering dynamic change of the complement method according to the type of sensor failure, Figure 3 shows the result summarised by failed sensor position, in case of 3 windows fault duration ($5120ms$). Roughly saying, Figure 3 shows that the effectiveness of complement varies for each sensor position. Moreover, effective combination differs in failed sensor places.

In right arm, all combinations get higher scores than no complement. Meanwhile, in other positions, most of combinations get higher, but one or some combination get less score than no complement.

For both ankles, most combinations that show lower score than no complement are related to feature and classifier complement, which are based on spatial dependency. Since four of six sensors are equipped on upper half of the subjects' body, the reason of it can be that ankles' movement is highly independent to other positions, and its spatial correlation gets marginal.

In waist and chest, combinations showing lower score than no complement are related to sensor data complement and output level complement, which are based on temporal correlation. Complement methods using spatial correlation perform well in these positions.

In addition, no complement method gets higher scores than normal case in waist, as well as many combination also get higher scores especially in spatial correlation based complement. This means sensor data of the waist causes confusion for classification. The reason can be thought that waist sensor obtains similar data for some activities, such as "Standing", "Clapping hands", "Folding arms", and it causes misclassification. "Reduced" leaves out the mischief data and gets higher than "Normal". Similar to that, spatial complement can estimate more appropriate value than original data, and it seems to contribute to get higher score.

From the result of Figure 3, choosing the combination of complement method according to the failed sensor position is expected to give better result. Table 2 summarises the best combination method for each failure position and failure duration. "Selected" in Figure 2 shows the result of choice of combination based on Table 2. The result gives the best on each and shows that it increase the F_1-score about 0.08, from 0.61 of no complement to 0.69.

Figure 3 and Table 2 are created with an example of activity recognition setting. If a sensor placement is different, the correlation model on each step will be changed, and the best combination shown in Table 2 will be different. Thus, this table has to be recreated according to the setting when exploiting the optimisation of failure type. However, this result confirms that dependency of optimal complement method exists on sensor position. Moreover, above discussion of the tendency can be the basis of this dependency.

	Failure Duration [window(s)]				
Label	1	2	3	4	5
Left arm	II-c	III-c	IV-a	IV-a	IV-a
Right arm	IV-a	IV-a	II-d	IV-a	III-d
Left ankle	II-c	II-c	II-c	II-f	I-d
Right ankle	I-d	I-d	I-d	I-d	I-d
Waist	III-d	I-c	I-c	I-c	III-d
Chest	IV-a	IV-a	II-d	IV-a	II-d

Table 2: Best Complement Method on Each Conditions

Conclusion

In this study, complement methods of missing sensor data are explored. Considering the activity recognition process and characteristics of complement method with temporal and spatial correlation, complement methods on each step are chosen. The ARAR algorithm is used for raw data, kernel regression is used for feature values, estimation of weak classifier output in ensemble learning is used as classification algorithm, the transition probability is used for the classification result. Also, we proposed a method to combine these complement methods.

From the evaluation of possible combinations, the results showed that the F_1-score improved from 0.61 in no complement to 0.68 in the combination of four methods. Moreover, the results showed that the most effective combination depends on the failed sensor position, and selecting the best method according to failed sensor position improved the F_1-score to 0.69.

References

[1] ATR-Promotions, I. WAA-006. http://www.atr-p.com/sensor06.

[2] Brockwell, P. J., and Davis, R. A. *Introduction to Time Series and Forecasting*. Taylor & Francis, 2002.

[3] Chu, F.-L. Analyzing and forecasting tourism demand with arar algorithm. *Tourism Management 29*, 6 (2008), 1185 – 1196.

[4] Dixon, J. K. Pattern recognition with partly missing data. *IEEE Transaction on Systems, Man and Cybernetics SMC-9*, 10 (Oct. 1979), 617–621.

[5] Murao, K., Terada, T., Takegawa, Y., and Nishio, S. A context-aware system that changes sensor combinations considering energy consumption. In *Proc. of Pervasive 2008*, LNCS 5013 (2008), 197–212.

[6] Sagha, H., Millan, J. D. R., and Chavarriaga, R. A probabilistic approach to handle missing data for multi-sensory activity recognition. In *Context Awareness and Information Processing in Opportunistic Ubiquitous Systems, UbiComp '10 Workshop* (2010). http://www.wearable.ethz.ch/resources/UbicompWorkshop_OpportunisticUbiquitousSystems.