Training Human Activity Recognition for Labels with Inaccurate Time Stamps

Takamichi TODA
Kyushu Institute of Technology
1-1 Sensui-cho, Tobata,
Kitakyushu-shi
Fukuoka ,804-8550, JAPAN
5toda6@gmail.com

Sozo INOUE
Kyushu Institute of Technology
1-1 Sensui-cho, Tobata,
Kitakyushu-shi
Fukuoka ,804-8550, JAPAN
sozo@mns.kyutech.ac.jp

Shota TANAKA
Kyushu Institute of Technology
1-1 Sensui-cho, Tobata,
Kitakyushu-shi
Fukuoka ,804-8550, JAPAN
muscari.st@gmail.com

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

UbiComp '14, September 13 - 17, 2014, Seattle, WA, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3047-3/14/09...$15.00.
http://dx.doi.org/10.1145/2638728.2641297

Abstract

We generally use supervised learning when performing activity recognition using mobile sensor devices such as smartphones. In this application, case data associated with the sensor information and type of action is required. However, there is a possibility that a time shift occurs because this association is done manually on the audio and video that has been acquired along with the sensor information. In this paper, we propose a method of activity recognition that can recognize correct actions even if there is a time gap. In this method, we add labels that shift the original learning data label. We also implement multi-label machine learning. In addition, we propose a method for repeated learning based on the Expectation-Maximization(EM) algorithm. To evaluate this method, we conducted an experiment that recognized three types of behavior using a Naive Bayes classifier. In the evaluation, we pieced together three types of human action data into one dataset called pseudo sequence data. We slid the action labels of the pseudo sequence data and examined whether the recognition rate was improved by our proposed method. The results show that the proposed method can perform activity recognition with high accuracy, even if the action labels times are shifted.

Author Keywords

Human activity recognition, machine learning, mobile sensor device, EM algorithm, smoothing

ACM Classification Keywords

I.1.2 [Algorithms]: Analysis of algorithms.

Introduction

Recently, research regarding human behavior recognition useing mobile sensor devices has increased, and applications in various fields as sports medicine are anticipated [4]. In order to perform action recognition, it is necessary to collect behavior data to create a recognition model using machine learning. To do this, we require the case data associated with the sensor information and type of action. However, there is a possibility that a time shift occurs because this association is done manually from the acquired audio and video along with the sensor information.

In this paper, we propose a method of activity recognition that can recognize correct actions, even if there is a time gap. In this method, we add labels that have been shifted from the original learning data labels. We also perform multi-label machine learning. We use the Naive Bayes classifier in the machine learning step, and add conditional probability to calculate the feature value. In addition, we propose a method of repeated learning based the EM algorithm.

Related Research

Many studies on action recognition using mobile sensors have been published since its introduction in [1]. In supervised learning, it has become a problem of inaccurate teacher information such as incomplete label information. For this reason, using semi-supervised classification [5] has been proposed.

In [9], it was shown that using a basic semi-supervised classification based on self-training and co-training,it is possible to recognize actions without action labels. In [6], a function was proposed that projects a multi-dimensional space-specific feature value using unlabeled and labeled data, where supervised learning uses Support vector machine(SVM) in the space of projected label data. This enables the characteristics of the unlabeled data to also affect the learning result. In that it uses incomplete labels, this research is similar to ours, however, it does not consider time shift.

In [10], *multi-instance learning*, i.e. a machine learning method that can respond to more than one sample set of one label, we recognized actions without knowing all the label data. This multi-instance learning method was introduced in [12]. However, there is an assumption that, one or more existing labels are given for the sample set.

In [15], instead of focusing on missing label times, a method was proposed that can perform action recognition by action order alone. In this method, the correct label is recognized by Dynamic Programming(DP) matching and supervised learning after segmentation and clustering. This method is effective when only the order is known. However, it is not able to attach a label to a specific time when the action label time is shifted.

Multi-label learning is machine learning that allows labels that are structured or multiple labels in the learning sample. The method was introduced in [11] and [12]. We focus on multi-label learning where the label may have a plurality of samples, but a true label exists. This is a special case of multi-label learning [7]. This method uses an EM algorithm that repeats the following two steps.

M step
performs machine learning in whish the initial multi-label set is stochastically abeled.

E step
estimates the probability distribution of the learning data labels.

In [2], they solved the same problem as a convex programming problem of loss function, and applied this to video for person labeling. By using the method in [7], we also attempt to convert the problem of deviation in time series data to a convex programming problem.

The method in [3] extends the technique of [7] using conditional random fields(CRFs) that are often used in natural language processing. By doing so, even if many action labels are given, machine learning can be performed. The work in [7] is similar to our research with respect to extending the time-series data approach. However, it differs in that they assumed that more than one label applied to the data of one sample implies multiple persons. We assume a time shift has occurred.

The multi-instance learning described above does not assume more than one label can be attached to one sample as in [7]. However, in [10], the method has been improved to accommodate more than one label, called experience sampling. In this case, the method that iteratively converges to label unlabeled data is similar to ours. However, it is not a stochastic approach but a decisive approach as in [7]. In addition, because a label is assigned on a regular basis by experience sampling, this is not a method that considers label time shifts as we do.

Proposed Method

In this section, we discuss three things. First, we introduce the method that converts time shifted data into multi-label data. Next, we explain how to solve multi-label problems. Finally, we discuss smoothing.

Converting multi-label data

The input data is assumed to be data acquired by a mobile sensor device such as three-axis acceleration data or time series. These sensor data are processed just as for general action recognition. The time windows are acquired and the feature value in each time window is calculated.

If the feature value in discrete time t is $\vec{x_t}$, and if the given label is y_t, we assume that the following is satisfied.

$$(\vec{x_t}, y_t) \tag{1}$$

However, we defined the time as $t = 1, 2, ..., n$. Using this time, the calculation produces a multi-label set for each S_t like Formula as follows.

$$S_t := \{y_{t'} | t - \alpha \leq t' < t + \alpha\} \tag{2}$$

Here, α is a constant parameter.

Equation 2 assumes that the provided label is 2α longer than the given label. At a certain time t, the times before and after the given label within α also are considered as label candidates. Just as for the label, several candidates can exist, although the data correspond to one action. We refer to [7] for the assumptions used here, and we treat this procedure as multi-label learning.

Solving multiple labels

Data that have been converted into a multi-label set in the previous section are processed by the method proposed in [7], an EM algorithm formulated using the

Kullback—Leibler(KL) divergence. Here, we briefly describe the procedure.

In the previous section, the data is given in the multi-label form as

$$(\vec{x}_t, S_t) \quad t = 1, 2, ..., n \tag{3}$$

We next perform the procedure described.

1. Given $t = 1, 2, ..., n$, we assume that the label in S_t is given for all data, i.e., given $S_t = \{y_t^{(1)}, y_t^{(2)}, ..., y_t^{(m_t)}\}$,

   $$(\vec{x}_t, y_t^{(j)}) \quad j = 1, 2, ..., m \tag{4}$$

2. M step: we determine model f by machine learning using the data.
3. E step: using model f, we estimate at $\vec{x}_t$ in $t = 1, 2, ..., n$. Further, we seek the probability $p(y|t) = f(\vec{x}_t)$, i.e., the distribution of the estimated label.
4. Similar to Step 1, we create the learning data for a single label to increase the number of samples in proportion to the probability distribution of the estimated label. In other words, for each t and y, the following holds.

   $$\text{the number of } (\vec{x}_t, y) \propto p(y|t) \tag{5}$$

   This data becomes the learning data for the next step.

5. We repeat Steps 2 - 4 until the estimation results converge or their change becomes very small.

Because these steps are formulated as an EM algorithm, they converge to a local solution by Jensen's inequality.

We call this method the *EM method*. Figure 1 shows its flow. Likewise, we call the method that estimats the action label by a maximum likelihood method after Step 3 *Non-EM method*.

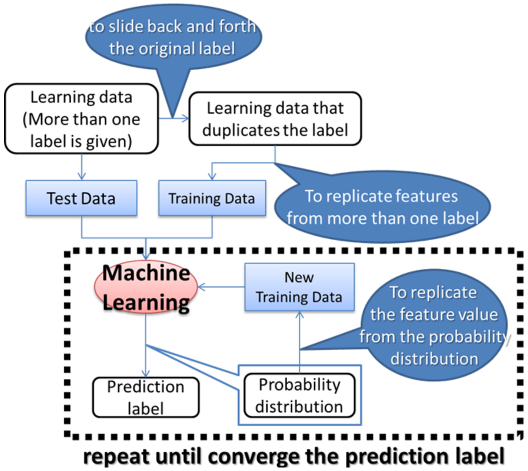
to slide back and forth the original label
Learning data (More than one label is given)
Learning data that duplicates the label
Test Data
Training Data
To replicate features from more than one label
Machine Learning
New Training Data
To replicate the feature value from the probability distribution
Prediction label
Probability distribution
repeat until converge the prediction label

Figure 1: EM method flow

Smoothing method

The method described in Section of **Solving multiple labels** does not consider time series. If we consider time series, we should be able to improve the accuracy. The smoothing method uses neighboring labels. In Step 3 above, we perform smoothing in the following way:

At time t,
If $y_{t-1} = y_{t+1}$,
Let be $y_t \leftarrow y_{t-1}$.

In other words, if the labels before and after a particular label match, we assume they belong to the same label. However, there is a problem with this smoothing technique. Because it considers only levels before and after the label under consideration, smoothing is not correct if the result of erroneous recognition crosses two or more labels.

In the evaluation of the next section, we compare cases that have and have not been smoothed.

Evaluation

We evaluateds whether, by using the proposed method even data with shifted activity labels, we could recognize activities with nearly the same accuracy as the case when exact labels are given. We used data comprising three types of human action data, called the pseudo sequence data. We slid the action label of the pseudo sequence data and examined whether the recognition rate was improved by our proposed method. We used the statistical analysis software R to analyze the result, and the classifier was Naive Bayes.

Data types

Given the three types action data of collected by the Human Activity Sensing Consortium(HASC) Challenge[14], we created pseudo sequence data. The action types used were adult male's "jog.", "walk.", and "stay.". The data type was three-axis acceleration. We used the data of 10 people in the experiment. Because the measurements were performed five times for one person, each action dataset contained 50 files. The duration of each action was 20 s. We list the dataset details in Table 1.

Table 1: Evaluation data details

Test subject	20's men
Number of people	10 people
Measurement place	outdoor(asphalt)
Times of measurements	5 times
Measurement device	ipod touch(4th)
Measurement data	3-axis acceleration
Position of device	Right pocket
Sampling frequency	100Hz
Type of actions	jog, walk, stay
Time of each action	20 seconds

Feature extraction

We extracted features from the data shown in Section of **Data types**. The sensor data used consisted of three-axis acceleration. We set a time window of 0.5 s width and 2 s shift. In addition, feature values were calculated over each of the time windows. The features we used were mean, variance, and energy. Table 2 summarizes the features obtained from the data.

Table 2: Feature detail

Value of acceleration	the resultant value of the 3-axis
Width of the time window	2 seconds
Width of the time window	0.5 seconds
Calculation of the feature	mean, variance, energy

The reason for combining the three axes is that the mounting position of the terminal during measurement was not fixed.

Method of evaluation

To assess the utility of the proposed method, we carried out machine learning on the learning data with shifted times. We then compared the change in recognition accuracy for three methods:

- **Naive**
  action recognition by machine learning using the naive Bayes classifier.
- **Non-EM**
  before performing machine learning, we processed the learning data according to our proposed method.
- **EM**
  This method is nearly the same as Non-EM. However, we repeat our proposed method until the learning models are stable.

Evaluation data

We selected three types of action data at random from the three types of action data that each had 50 data files as described in Section of **Data types**. In addition, we created the pseudo sequence data for learning data by binding the three types of action data remaining.

In this data set, even if we selected the data at random, it is important to check whether we can correctly perform activity recognition. We tested the machine learning by giving it the correct action labels. We repeated the action recognition 100 times with a naive Bayes classifier and the average accuracy was 98.2% with a standard deviation of 5.3. We determined that there are no issues with using this data.

Shifting action labels

As mentioned in Chapter of **Introduction**, there is a possibility that a time shift occurs because the action and label association is done manually.

We intentionally generated this shift for this experiment. Furthermore, we confirmed the variation in recognition accuracy caused by the amount of shift. In addition, we confirmed the improvement caused by the our proposed method.

The pseudo sequence data used in this evaluation consisted of 2,940 s of data, as the 49 data files of 20 s were bound together. For this data, the action labels were shifted, and we evaluated it using Naive, Non-EM, and EM method.

Application of smoothing method

We applied the smoothing technique described in Section of **Smoothing method** for the Naive, Non-EM, and EM method. Furthermore, we compared the variation of recognition accuracy with respect to smoothing.

Result

Figure 2 shows the results of the evaluation experiments. In this figure, when the value of the x-axis is 0, it is indicated that the correct label was applied.

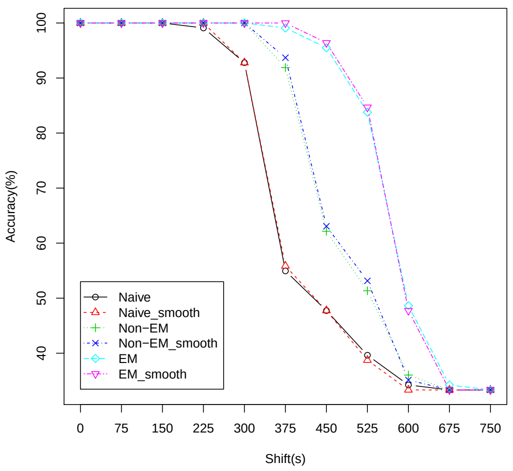

Figure 2: Effect of shift width on recognition accuracy

From the above results, we concluded that the recognition accuracy is decreases when the action label is displaced. In addition, we also conclude that by using the proposed method, it is possible to improve recognition accuracy.

The recognition accuracy was improved by smoothing if the shift was small. However, the improved values are also very small.

Accuracy improvement of the EM algorithm

In this experiment, we examined how recognition accuracy converges for the EM algorithm. Table 3 shows the change in EM algorithm recognition accuracy.

Table 3: Transition of the accuracy of the EM algorithm(%)

	repeat times					
time shift	0	1	2	3	4	5
75 (s)	100	100	100	100	100	100
150 (s)	100	100	100	100	100	100
225 (s)	99.1	100	100	100	100	100
300 (s)	92.8	100	100	100	100	100
375 (s)	55	91.9	99	99.1	99.1	99.1
450 (s)	47.7	62.2	95.5	95.5	95.5	95.5
525 (s)	39.6	51.4	84.7	83.8	83.8	83.8
600 (s)	34.2	36	36	42.3	48.6	48.6
675 (s)	33.3	33.3	33.3	33.3	33.3	34.2
750 (s)	33.3	33.3	33.3	33.3	33.3	33.3

In Table 3, When repeat times show 0, it indicates Naive method. When repeat times show 1, it indicate Non-EM method. The EM algorithm was repeated five times in this experiment. At the sixth iteration, the action label of "jog" was dropped from the learning model. Despite this the "jog" action was recognized almost normally up to the fifth iteration.

Discussion

We determined that when using the proposed method, even if the action labels were shifted, we were able to recognize activities with a high accuracy. We consider future work in this section.

Smoothing method

We determined that the proposed smoothing method improved the recognition accuracy in the experiments. However, the improvement was not sufficient. As shown in Figure 2, the smoothing is not correct when two or more erroneous recognitions neighbor each other, as the method proposed in this paper considers only the recognized labels immediately before and after the one under consideration. That is, if the accuracy is not good, this is because the errors often occurred in blocks of two or more. For this reason, we believe that the accuracy improvement caused by smoothing was insignificant in this experiment.

In the future, we plan to expland the smoothing range to correspond to a number of labels before and after the one considered. We will also take into account the effects of precision and extension size.

Iterations of the EM algorithm

In this experiment, the EM algorithm was iterated five times. However, considering the improvement of recognition accuracy by EM algorithm in Section of **Accuracy improvement of the EM algorithm**, after the third iteration, there were no large improvements. The data that this iteration used consisted of about 3000 s of action data, and there were only three action types. However, if we use more data or action types, the time needed for machine learning will be increased. Hence, we need to decide a threshold in order to analyze efficiently.

If we iterate the EM algorithm more than necessary, ”jog” actions are lost from the recognized action labels. This is caused by over-learning, and we plan to develop an algorithm that considers this case.

Types of data

In this study, we used pseudo sequence data for evaluation. However, we would like to show that our proposed method is practical. It is necessary to further improve the recognition accuracy of the proposed method, even if the data that we analyze is more complex. Hence, we plan to perform further evaluations. We consider three cases:

- **Increasing the number of action types**
  We used only three action types for this evaluation. Because research on human activity recognition includes many more action types, we need to increase the number of action types. Action types that we plan to add are ”skip.”, ”stair-up.”, and ”stair-down.”.
  If we add these action type, we must consider the features used, because similar features values will be seen when the number of action types increases.

- **Usinge the continuous sequence data**
  The sequence data used in this evaluation experiment was pseudo sequence data obtained by combining the three types of action data. However, when we perform real human activity recognition, there are many additional points to consider. For example, changes of action type, measurement errors caused by human error, and difficulties of identifying action labels. We believe that it is necessary to evaluate the proposed method with real sequence data.

- **Various shift label types**
  As mentioned in Section of **Shifting action labels**, the label shift of this evaluation is a parallel move such as that in Figure 3. However, there are also label shift such as that in Figure 4 in real sensing. We need to be be able to determine such shifts.

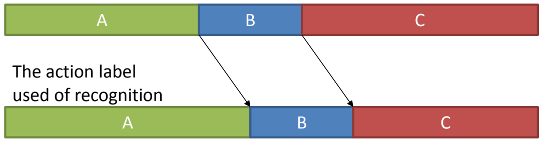

Figure 3: The shift of the evaluated labels

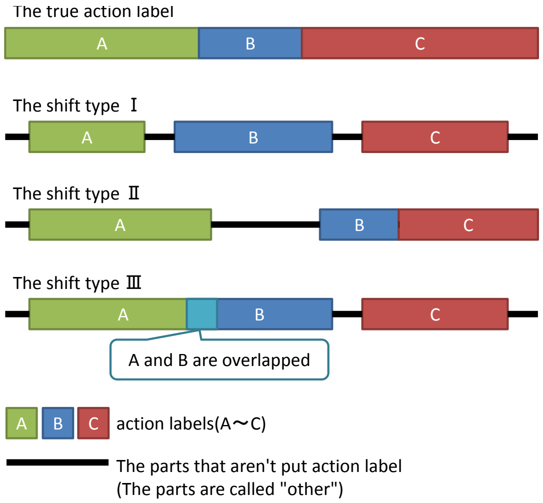

Figure 4: Examples of the other shift label

Conclusion

In this paper, we proposed a method of activity recognition that can recognize correct actions even if there is a time gap. We added labels that can be shifted from the original label to learn the data and we formulated machine learning as a multi-label problem.

The results of the experimental evaluation for three types of actions, "jog", "walk", and "stay", show that if we use our propose method, it is possible to recognize activities with a high accuracy, even if the time of an activity label is shifted. In addition, the accuracy of behavior recognition was marginally increased by the EM algorithm. However,since the EM algorithm iterations significantly increased the processing time, we will consider a convergence threshold in future.

On the other hand, we proposed a smoothing method, but could not show any significant accuracy improvement in this method. We plan to consider another approach.

In this study, we used pseudo sequence data that combined three types of action data. In future, we will widen the application of recognition by, for example, increasing the types of action data, measuring real sequence data, and using medical data. The medical data is the collection of nurse behavioral data collected in cooperation with Saiseikai Hospital in Kumamoto.

Acknowledgment

The part of this research are the following researches. Funding Program for World-Learning 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)", Basic research (A) "Towards the Construction of the Large Scale Human Sensor Database for the Activity Understandings (Principal Investigator: Nobuo Kawaguchi)", and Basic research (B) "Sensor context estimation techniques by two levels of meaning layer and physical layer (Principal Investigator: Sozo Inoue)"

References

[1] Ling Bao and StephenS. Intille. Activity recognition from user-annotated acceleration data. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive Computing*, Vol. 3001 of *Lecture Notes in Computer Science*, pp. 1–17. Springer Berlin Heidelberg, 2004.

[2] Timothee Cour, Ben Sapp, and Ben Taskar. Learning from partial labels. *The Journal of Machine Learning Research*, Vol. 12, pp. 1501–1536, 2011.

[3] Mark Dredze, Partha Pratim Talukdar, and Koby Crammer. Sequence learning from data with multiple labels. In *Workshop Co-Chairs*, p. 39. Citeseer, 2009.

[4] Moore A.J. Tilbury N. Church J. Farringdon, J. and P.D.:Wearable Biemond. Sensor badge and sensor jacket for context awareness. In *In Proceedings of the Third International Symposium on Wearable Computers*, pp. 107–113, 1999.

[5] Yves Grandvalet, Yoshua Bengio, et al. Semi-supervised learning by entropy minimization. In *NIPS*, Vol. 17, pp. 529–536, 2004.

[6] T. Huynh and B. Schiele. Towards less supervision in activity recognition from wearable sensors. In *Wearable Computers, 2006 10th IEEE International Symposium on*, pp. 3–10, Oct 2006.

[7] Rong Jin and Zoubin Ghahramani. Learning with multiple labels. In *Advances in neural information processing systems*, pp. 897–904, 2002.

[8] M. Stikic, D. Larlus, S. Ebert, and B. Schiele. Weakly supervised recognition of daily life activities

with wearable sensors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 33, No. 12, pp. 2521–2537, Dec 2011.

[9] M. Stikic, K. Van Laerhoven, and B. Schiele. Exploring semi-supervised and active learning for activity recognition. In *Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on*, pp. 81–88, Sept 2008.

[10] Maja Stikic and Bernt Schiele. Activity recognition from sparsely labeled data using multi-instance learning. In Tanzeem Choudhury, Aaron Quigley, Thomas Strang, and Koji Suginuma, editors, *Location and Context Awareness*, Vol. 5561 of *Lecture Notes in Computer Science*, pp. 156–173. Springer Berlin Heidelberg, 2009.

[11] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, Vol. 3, No. 3, pp. 1–13, 2007.

[12] Zhi-Hua Zhou, Min-Ling Zhang, Sheng-Jun Huang, and Yu-Feng Li. Multi-instance multi-label learning. *Artificial Intelligence*, Vol. 176, No. 1, pp. 2291–2320, 2012.

[13] Kawaguchi, N., Ogawa, N., Iwasaki, Y., Kaji, K., Terada, T., Murao, K., Inoue, S., Kawahara, Y., Sumi, Y. and Nishio, N. HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings. Proc of ACM AH 2011 2010, pp. 271–275, 2011.

[14] Nobuo KAWAGUCHI. Towards the Construction of the Large Scale Human Sensor Database for the Activity Understandings. Multimedia, Distributed Cooperative, and Mobile Symposium (DICOMO) Collection of papers. 2010, pp. 579–582, 2010.

[15] Kazuya, Murao. Yasuyuki, Torii. Tsutomu, Terada. Masahiko, Tsukamoto. Labeling Method for Acceleration Data Using an Execution Sequence of Activities J.IPS Japan, Vol. 55, No. 1, pp. 519–530, jan 2014.