Vol.2016-UBI-50 No.8
2016/5/28

Toward Recognizing Unknown Activities Using Word Vectors

Moe MATSUKI[1,a)] Sozo INOUE[1,b)]

Abstract:
In this paper, we propose a method for activity recognition which can estimate new activities which does not appear in the training data, combining word vectors obtained by natural language processing. Because traditional activity recognition utilizes supervised machine learning, unknown activity classes which do not appear in the training dataset is unable to be estimated. For this problem, zero-shot machine learning method is proposed, but it requires the preparation of semantic codes. For this problem, we utilized natural language processing tool 'word2vec' to generate word vectors and utilize them as the semantic code. To evaluate the proposed method, we evaluated whether we could estimate new activities which does not appear in the training data, with the sensor dataset collected from 20 households for 4 months, along with the user-generated labels using the web system which can estimate, modify, and add new activity types. As a result, 7 activity classes could be estimated even if such classes does not appear in the training data, and the best precision was 89% while recall is very low. We set assumptions for the distances among word or feature vectors, and address several future challenges.

1. Introduction

Recently, the study of activity recognition is becoming a trend and highly expected for smart energy management at home, remote elderly care, and for improving lifestyles for health.

In most of the activity recognition methods, the activity estimation algorithm is automatically generated by supervised machine learning, with the training dataset of input sensor data and correct *activity labels*. This means that unknown activity types (hereafter, *activity classes*) which do not appear in the training dataset is unable to be estimated.

For this problem of supervised learning, zero-shot machine learning method [4] is proposed. Zero-shot learning enables the estimation of new activity classes which do not appear in the training data, by preparing so-called *semantic code*, which is the relational data between activity classes and other vector data. For instance, [1] prepares the semantic code between activity classes and the representation of which parts of the body movement, and build a model to estimate the parts of the body from sensor data, and in estimation phase, the estimated output is interpreted to the activity class using the semantic code. [3] also takes the related approach to improve accuracy from limited training data.

However, these approaches require the preparation of semantic code, which requires manual labor for many activity classes, and it is not clear how to prepare such semantic code for more semantic activity classes. In this paper, we utilized natural language processing tool 'word2vec'[2] for modeling the latent semantics of text data. Word2vec generates *word vectors* for each word by training neural autoencoder with the neighboring information in the text. We propose to utilize word vectors as the semantic code for unknown activity recognition.

To evaluate the proposed method, we evaluated whether we could estimate new activities which does not appear in the training data, with the sensor dataset collected from 20 households for 4 months, along with the user-generated labels using the web system which can estimate, modify, and add new activity types[5].

As a result, 7 labels could be estimated even if such classes do not appear in the training data. And The best precision was 89%. On the other hand, the precision was low, from which we can set an assumption that for for these failed activity classes are close to the estimated activity classes in the feature or the word vectors, and we clarify several challenges for this research.

2. Method for Recognizing New Activities

In this section, we describe the method for applying zero-shot learning method combined with word vectors for unknown activity recognition.

2.1 Preliminary

In this section, we introduce the basic expressions in the paper.

X is the set of feature vectors calculated from the sensor data, and Y is the set of activity labels each of which corresponds to the sample of X. We denote the domain of X as $\mathcal{X}$, and that of Y as $\mathcal{Y}$.

Fig. 1 shows the naive activity recognition method. In the traditional activity recognition methods, the algorithm is automatically generated by supervised machine learning, with the training

[1] Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata, Kitakyushu 804–8550, Japan
[a)] matsuki.sousis@gmail.com
[b)] sozo@mns.kyutech.ac.jp

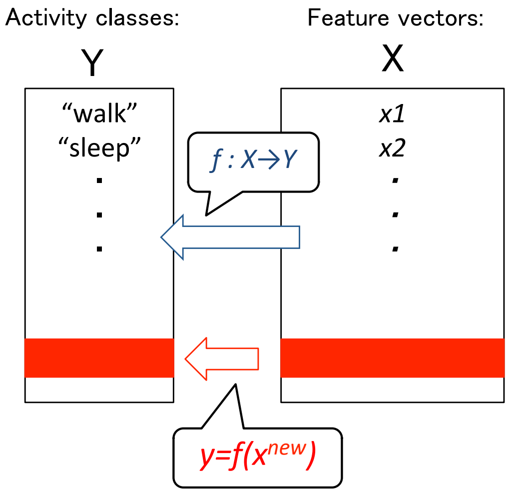

Fig. 1 Traditional activity recognition: supervised machine learning with X as explanatory variables, and Y as the response variable.

dataset of X and Y. Let X and Y also denotes the variables for each sample in them. Then, the supervised machine learning generates the function $f : \mathcal{X} \rightarrow \mathcal{Y}$.

When we estimate an activity $\tilde{y}$, we can perform

$$\tilde{y} = f(x^{new})$$

from a new input x^{new}.

2.2 Utilizing zero-shot learning

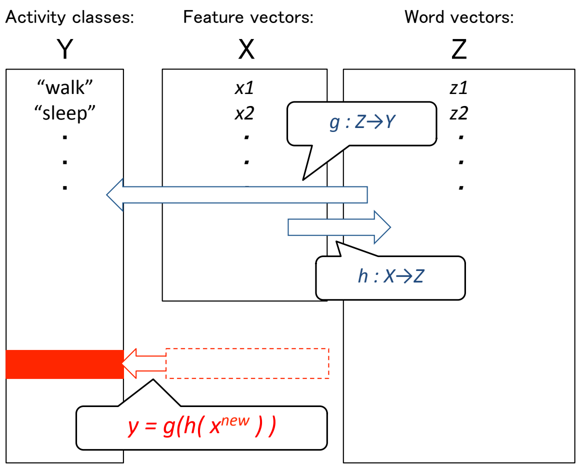

Fig. 2 The proposed method: with the word vectors Z, generate $g : \mathcal{Z} \rightarrow \mathcal{Y}$, and $h : \mathcal{X} \rightarrow \mathcal{Z}$, and obtain $\tilde{y} = g(h(x^{new}))$ for the estimation.

In zero-shot learning, we assume that we have word vectors Z, where each of which corresponds to the sample of activity labels Y. We describe how to build word vectors in Section **??**.

Fig. 2 illustrates the proposed method. The proposed method is described in the following:

(1) with the word vectors Z and the activity labels Y, generate the function $g : \mathcal{Z} \rightarrow \mathcal{Y}$ by supervised machine learning,

(2) with the feature vectors X and the word vectors Z, generate the function $h : \mathcal{X} \rightarrow \mathcal{Z}$ by supervised machine learning, and

(3) when we estimate an activity $\tilde{y}$, we can perform

$$\tilde{y} = g(h(x^{new}))$$

from new input x^{new}.

Because the activity labels Y and the word vectors Z can include the activity classes which do not appear in the feature vectors X, we can estimate activity classes which are new for the sensor data.

2.3 Utilization of word vectors

In the field of natural language processing, the tool 'word2vec' for modeling latent semantics from text data is frequently used. Word2vec generates *word vectors* for each word by training neural autoencoder with the neighboring information in the text. Using word2vec, we generated word vectors from the Japanese Wikipedia, and picked up the word vectors corresponding to the activity classes.

Note that some of the activity labels include multiple words, such as "Newspaper / radio / TV". For such label, we merged the word vectors by taking the mean of the word vectors.

3. Preliminary Evaluation

In this section, we evaluate the proposed method with the dataset collected from households with simple sensors and user-generated activity labels. In particular, we focus on the question whether the activity classes which do not appear in the training data are correctly estimated.

3.1 Dataset

As a dataset, we utilized the dataset of light and power sensor data and activity labels from 20 households collected through the experiment with 35 households for 4 months[5].

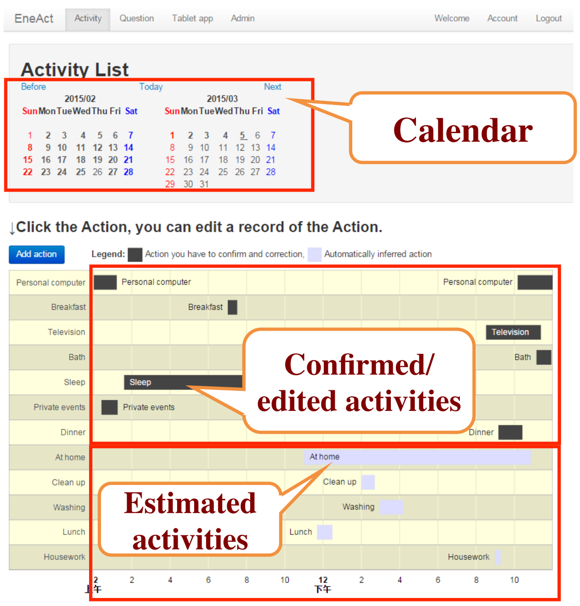

Fig. 3 Web view for estimating/editing/recording activity labels.

We lent tablet terminals to the subjects, and asked them to place each of them at a frequently-used place, such as in a living or dining room. The terminal has a light sensor, and sends the light data periodically to the server. We also collected the entire electricity data from an easy-to-setup smart meter from each household, which were stored in the server.

As for activity labels, we utilized the web-based system with which each subject can define or edit new activity classes based on the estimated activities by the system. We asked for each subject to input the activity record a few minutes in a day. Note that the labels were recorded in Japanese language, but we interpret into English throughout this paper.

3.2 Preprocessing

We used the lights and power consumption data, and the activity labels each of the subjects recorded obtained in the experiment.

For a light sensor data and the power data, we extracted the following feature vectors X by 1 minutes:

- minutes in a day,
- mean / standard deviation / maximum / minimum value of light sensors, and
- watt values obtained from power meters by 1 minutes.

For the word vectors Z, we we applied word2vec tool to the Japanese Wikipedia site, which resulted in 100 dimensions of word vectors. Moreover, we reduced the dimensions to top 10 important variables by applying the random forest algorithm.

As a result, the activity classes Z became the number of 52, as “Sleep”, “Breakfast”, “Lunch”, “Dinner”, “Bath”, “Television”, “Reading”, “At home”, “Washing”, “Clean up”, “Personal computer”, “Study”, “Private events”, “Commute”, “Work,/academic”, “Meal”, “Housework”, “Television/radio/newspaper/magazine”, “Rest/relaxation”, “Going out”, “Cooking”, “Buy something”, “Lunch preparation”, “Yoga”, “Game”, “Walk”, “Lunch making”, “Hula dance”, “Purring”, “Cloth drier”, “Tableware washing dryer”, “Ironing”, “Snacking”, “Tablet”, “Rice cooker”, “Movement”, “Business”, “Work”, “Eating out”, “Video”, “Bike loading”, “Business trip”, “Shopping”, “Beauty salons”, “Yu Yu land”, “Update the vehicle insurance”, “Bake cookies”, “Land office”, “Individual enrichment courses”, “Hatsuuma Festival”, “Garbage”, and “Morning reading”

3.3 Evaluation method

As a base algorithm for supervised machine learning, we adopted k-nearest neighborhood method, with $k = 3$. This algorithm is considered to be appropriate to reflect the distances among samples directly.

To evaluate whether the proposed method could estimate the activity classes which does not appear in the training data, we take the following steps: For each activity class $y \in \mathcal{Y}$,

(1) remove the samples with activity class y from the data,
(2) train with the data after removal, and
(3) test with the data before removal.

By this, we can evaluate whether the each of the removed activity classes could be estimated in the original data as test data.

As measures of accuracy, we adopt precision — the rate of correctly estimated samples among the samples which were estimated as the activity class —, and recall — the rate of correctly estimated samples among the samples of activity class in the ground truth —. To eliminate the imbalances in the ground truth, we applied bootstrap sampling for each of positive and negative samples in the ground truth data.

3.4 Result

We show the recognition result of activities which do not exist in the training data in Table 3.4. In the table, for the activity classes which was successful to be estimated, the number of true positive samples, precision, and recall, are shown.

Table 1 Newly estimated activities which do not exist in the training data.

Activity class	#TP samples	Precision [%]	Recall [%]
“Personal computer”	94	50.3	0.0
“Housework”	1596	57.4	0.4
“Going out”	16384	37.9	3.9
“Rest/relaxation”	8475	41.8	2.0
“At home”	4760	53.5	1.3
“Meal”	5704	89.3	1.3
“Private events”	44	44.9	0.0

From Table 3.4, “Personal computer”, “Housework”, “Going out”, “Rest/relaxation”, “At home”, “Meal, “Private events” could be estimated even if such classes does not appear in the training data.

The best precision was 89% for “Meal”, which means that the most of the estimations for “Meal” are correct. However, the recall is still 1.3% at the best. This means that many of the activities are missed to be estimated correctly.

3.5 Discussion

As in Table 3.4, “Housework”, “Going out”, “At home”, and “Meal” could be estimated even when each of them does not appear in the training data. On the other hand, the rest of the activity classes could not be estimated in the same situation. Amont them, “Sleep”,“Breakfast”, “Clean up”, and “Bath” have large number of true samples. We can set an assumption that for for these failed activity classes are close to the estimated activity classes in the feature or the word vectors, and that is why the estimation failed. For example, “Breakfast”, “Lunch”, and “Dinner” could be close to “Meal”, and “Clean up”, could be close to “Housework” from the meaning of the activity classes. As a future work, to confirm this assumption, we will have to train with various combination of activity classes to be removed. Moreover, We didnot observe difference between activity labels with single word or multiple words, in which the latter was combined by the mean of word vectors. We will also investigate the better method for integrating multiple word vectors.

4. Conclusion

In this paper, we proposed a method for activity recognition which can estimate new activities which does not appear in the training data, combining word vectors obtained by natural language processing. We combined zero-shot machine learning method and word vectors generated by the natural language processing tool ‘word2vec’. As a result of evaluation whether we could estimate new activities which does not appear in the training data, with the sensor dataset collected along with the user-generated labels, However, we also found several challenges for accuracies and differences among activity classes. Further work is expected to improve such activity recognition for unknown activity classes.

References

[1] Cheng, H.-t., Griss, M., Davis, P., Li, J. and You, D.: Towards zero-shot learning for human activity recognition using semantic attribute sequence model, *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*, No. 2, p. 355 (online), DOI: 10.1145/2493432.2493511 (2013).

[2] Mikolov, T., Corrado, G., Chen, K. and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, pp. 1–12 (online), DOI: 10.1162/153244303322533223 (2013).

[3] Nguyen, L. T., Zeng, M., Tague, P. and Zhang, J.: Recognizing new activities with limited training data, *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pp. 67–74 (online), DOI: 10.1145/2802083.2808388 (2015).

[4] Palatucci, M., Hinton, G., Pomerleau, D. and Mitchell, T. M.: Zero-Shot Learning with Semantic Output Codes, *Neural Information Processing Systems*, pp. 1–9 (2009).

[5] Pan, X., Minezaki, T., Isoda, T., Tanaka, S., Uchino, Y. and Inoue, S.: Analyzing Daily-life Activities and Power Consumptions Using Tablet Sensors and Activity Annotation Web System, *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, New York, NY, USA, ACM, pp. 1443–1452 (online), DOI: 10.1145/2800835.2801615 (2015).