The Optimization of Sensor Arrangement and Feature Selection in Activity Recognition

Ryo Urushibata, *Member, IEEE,* Taketoshi Mori, *Member, IEEE,* Masamichi Shimosaka, *Member, IEEE,* Hiroshi Noguchi, *Member, IEEE,* and Tomomasa Sato, *Member, IEEE*

***Abstract*—This paper deals with the optimization of sensor arrangement and feature selection for activity recognition of the people living alone with sensors. We suggest an algorithm which picks up from several thousand to millions of characteristic sensor reactions as feature candidates, and selects best feature combinations and corresponding sensor arrangements for classification with as small numbers of sensors and features as possible. This paper introduces two kinds of approach; one is making the sensor number as small as possible with quasi-maximized precision, and another is getting the globally maxmized precision with only needed sensors. We confirmed by a pyroelectric sensor system that this algorithm could get such solution by applying some sparse selection methods to the real life data.**

I. Introduction

Recently, the demand for support systems by the robots and sensors in a person's house become increasing and such systems become an active research area because of the aging society. In such systems, we need to recognize the person's behaviors or life activities by some kinds of sensors[1].

Considering cost and diffusiveness as the human detection sensor, we picked up the pyroelectric sensors and havebeen constructing the behavior labeling system[2][3]. The system labels a data segment of pyroelectric sensors as some behavior like "Sleeping" or "Going out". So far, for such system, people fixed the arrangement of sensors as that they thought to be the best arrangement intuitively. However, it sometimes fell into the issue that the sensor data of more than two behaviors are so similar that even human couldn't classify them from the data.

In order to solve this classification problem, we propose a method of making the sensor data suitable for classification. The input data changes by the sensors' number or places, so it is equal to optimizing the sensor arrangement so that the system can classify all behaviors correctly.

There are some researches dealing with the sensor arrangement problem; for example, arranging minimum number sensors by the geometric relationship between the sensor's detection area and the site with obstacles[4], selecting minimum number measuring points given the distribution model of the measured value[5]. Most of them decide the arrangement only from the geometric relationship or the distribution models of sensor value and don't deal with the combination of sequential sensor data.

The authors are with Graduate School of Information Science and Technology, The Univ. of Tokyo, Hongo, Bunkyo-ku, Tokyo, E-mail: tmori@ics.t.u-tokyo.ac.jp

In this paper, we discuss the sensor arrangement optimization with our behavior classification algorithm. Here, we want to reduce the sensor number as small as possible for practice, but the precision of classification becomes higher when there are many sensors. So we adopt two different approaches by their priority; the one gives first priority to the reducing of the sensor number even though the precision become a little lower than the best precision, and another gives first priority to the maximum precision with only needed sensors.

Concretely, the system enumerates tens of thousands of feature candidates which are the partial reactions of max 3 sensor combination, and selects the most appreciate feature combination in terms of the two different targets. We adopted the SFFS algorithm for the former target and the Grafting algorithm and LPBoost via column generation algorithm for the latter target.

II. Behavior Classification System

In this section, we explain the system of pyroelectric sensing for the behavior classification. The system sets some sensors at the ceiling or wall of the house in which a person lives alone like Fig.1, and classifies data segments into behavior labels. Concretely, the sensors detect "the person moves in the sensor's detection area or not", and return 0 or 1 per second. Then the system unifies the data of all sensors, and classifies the data segments. As the application of this system, for example, we assume the automatic care system of the old person living alone.

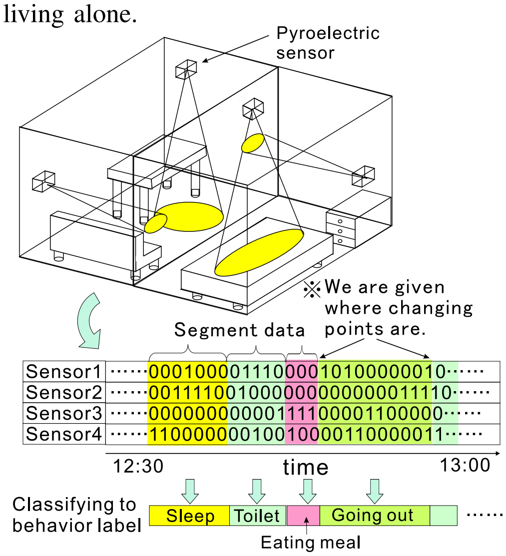

Fig. 1. Life activity recognition system with pyroelectric sensors

In this research, we adopted a learnt behavior models based approach to label a behavior name to sensor data segment. The models are constructed from some sensed behavior data with the ground truth behavior records. As for the behavior model, we utilized the joint posterior probability of the occurrence or non-occurrence of some characteristic parts of data, considering the compatibility with the task of sensor arrangement.

III. Optimization Outline of Sensor Arrangement

In this section, we explain the outline of the process of the sensor arrangement. To optimize the sensor arrangement, we first acquire the sensor data at all candidate places for some time periods. We can obtain this data by actually setting the pyroelectic sensors at all candidate places, or by setting small number of high-precision human motion sensors and simulating virtual pyroelectic sensor data from the motion sensor data.

A. Behavior Modeling by Extraction of Feature Data Parts

We use only the important parts of the data, the combination of which can characterize some kinds of behavior. We call such data parts as ”feature parts”. For example, if a data part ”011100” of a sensor at the entrance occurs only when going out, the part can be a feature part. Not only a single sensor data, the multi sensor reactions at the same time also can be a feature part.

We adopt the joint posterior probability of occurrence or non-occurrence of such feature parts as the behavior model. In this model, how to select the feature parts affects the classification precision significantly. The system tries to increase the precision of classification and decrease the selected sensor number simultaneously by solving the optimization of the feature selection which contains the element of the sensor selection, which is the main theme of this paper.

B. Extraction of Feature Part Candidates

At first, the system enumerates all feature part candidates from the data of all sensor place candidates no matter whether we finally use the sensors or not, then from these candidates, it selects the best feature combination which gives the high precision and uses the small number of sensors as possible.

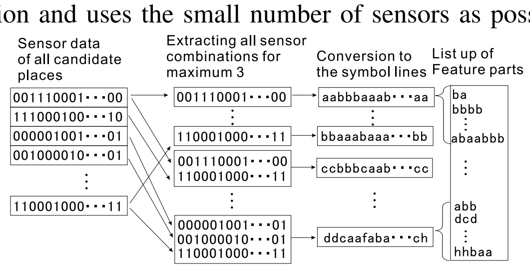

Fig. 2. Way of feature candidate extraction

As the preparation, the system obtains the sequential pyroelectric sensor data of all sensor place candidates, which is the combination of the bit column data. Then, the system extracts the feature part candidates by the process shown at Fig.2, where the system sets the max sensor combination corresponding to a feature part as 3. Here, the system is given the ground truth label of each time.

IV. Feature Selection For Sensor Arrangement Optimization

This section deals with the algorithms of the feature selection; selecting proper feature combination which makes the classification precision high and the sensor number small as possible.

We assume the number of feature part candidates is from tens of thousands to millions, so the system could get a certain group of the features H that makes the value of error function $E(H)$ toward the training data quite small. If the system only tried to make $E(H)$ small, the feature group not only has the danger of the overfitting, but also isn't considered the restriction of the sensor number. In order to solve this problem, we needs to add the penalty function $P(H)$, so we defined the objective function $F(H)$ which we try to minimize as $F(H) = E(H)+P(H)$. Then, we should define each function $E(H)$ and $P(H)$. We considered two different kinds of policy; the one reduces the sensor number even though the precision may become a little lower than the highest one, and another conversely acquires the strictly highest precision even though it couldn't make the sensor number so small. We call the former as the ”sensor number minimizing algorithm” and the latter as the ”precision maximizing algorithm” from now on.

A. Sensor Number Minimizing Algorithm

The sensor number minimizing algorithm selects the features' set that returns the quasi-highest precision by quite small number of sensors. In order to make the sensor number small and avoid overfitting, we define the penalty function $P(H)$ as $P(H) = \lambda_1 N_s(H) + \lambda_2 N_f(H)$ which $N_s(H)$ is the total sensor number used for the detection of the selected feature group H, and $N_f(H)$ is the number of features in H. And in order to make the precision high, we adopted the error rate $\frac{1}{N}\sum_{j=1}^{N} I(H(\boldsymbol{x}_j) \neq y_j)$ as the error function E(H), where $\boldsymbol{x}_i$ is the test segment data and y_i is ground truth behavior label. Here, $H(\boldsymbol{x}_i) = \mathrm{argmax}_y \sum_{j=1}^{n} \log p(y|\boldsymbol{x}_i, h_j)$, which h_n is the n th selected feature part, and in order to make the sensor number small, the parameter λ_1 needs to be larger enough than the parameter λ_2.

To decide the rule of selecting the feature newly added to the feature group already selected, we introduce the most famous and simple algorithm Sequential Forward Selection (SFS). This algorithm selects the newly added feature h^+ which reduces the objective function F(H) the most drastically by adding to the selected features' group, as $h^+ = \mathrm{argmin}_{h_i \in H_0} \quad F(H_1 \cup h_i) - F(H_1)$ which $H_{1,n}$ is the selected feature group with n elements and H_0 is the unselected group.

This solution isn't guarantied to be globally optimized one, so it might be the local one. By adding the other features one after another, the features added at the beginning could become not so important, so the objective function could become lower than the case when it is removed. To solute the problem, we adopted and enhanced the algorithm called Sequential Floating Forward Selection (SFFS)[6]. In the algorithm, the selected features can be removed by the sequential backward Selection (SBS), which removes the most unimportant feature.

B. Precision Maximizing Algorithm

The precision maximizing algorithm selects the feature set that maximizes the precision globally even though the number of used sensors might become larger than that of the sensor number minimizing algorithm explained above. The method selects only needed sensors by making the number of the selected features as small as possible. From now on, we regard the feature h_i as the logarithm of the posterior probability of class y calculated from the feature part F_i as

$$h_i = \log p(y|F_i) = \log \frac{p(F_i|y)p(y)}{\sum_{j=1}^{Y} p(F_i|y_j)p(y_j)}, \tag{1}$$

then the classifier can be shown as

$$\hat{y} = \underset{y}{\operatorname{argmax}}\, \boldsymbol{w}^T \boldsymbol{h}(\boldsymbol{x}_i, y) \tag{2}$$

which $\boldsymbol{h} = (h_1, \cdots, h_K)^T$ is the features' vector, $\boldsymbol{w} = (w_1, \cdots, w_K)^T$ is the corresponding weights' vector, Y is the group of the class label and $Y_{\setminus y_i}$ is the group of the class label except the right label y_i. From this equation, in order to acquire the right label from only the segment data $\boldsymbol{x}_i$, the equation and the inequality as

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{w}^T \boldsymbol{h}(\boldsymbol{x}_i, y) = y_i \tag{3}$$

$$\Longleftrightarrow \boldsymbol{w}^T \boldsymbol{h}(\boldsymbol{x}_i, y_i) > \max_{y \in Y_{\setminus y_i}} \boldsymbol{w}^T \boldsymbol{h}(\boldsymbol{x}_i, y) \tag{4}$$

must be satisfied. From the inequality, when we define a variable ρ_i as

$$\rho_i \equiv \boldsymbol{w}^T \boldsymbol{h}(\boldsymbol{x}_i, y_i) - \max_{y \in Y_{\setminus y_i}} \boldsymbol{w}^T \boldsymbol{h}(\boldsymbol{x}_i, y), \tag{5}$$

the code of it means whether the estimation is correct or not, and the absolute value of it means the degree of the correctness or incorrectness, so we can regard the sum of the function of this variable $\sum_{i=1}^{M} f(\rho_i)$ as the error function. As the typical penalty function toward eq.(2) which could be regarded as the linear sum of the weights, we decided to adopt the 1-norm $|\boldsymbol{w}|_1$ as the regularization factor $P(H)$ based on some consideration about sparsness and convexness.

There are two kinds of efficient algorithms for acquiring the globally optimized solution of the objective function composed of the error function using ρ and the penalty function as the 1-norm regularization $|\boldsymbol{w}|_1$; the Grafting algorithm advanced by Perkins[7] and the LPBoost via column generation advanced by Demiriz[8]. They are similar in terms that they both keeps global optimality by adjusting all weights of selected features whenever the system adds a new feature, but they are different in terms of the kinds of the error functions, the way of selecting the newly added features, and the way of adjusting the weights. In our experience, the Grafting is better in terms of the calculation cost, and the LPBoost is better in terms of the wideness of use, but we couldn't conclude which algorithm is totally better, so we decided to apply both of them and choose the better one. We introduce these two algorithms as follows.

1) Grafting: The outline of Grafting algorithm is as follows; first, the system differentiates the objective function $F(\boldsymbol{w})$ in terms of all of the weight w_i which satisfies $w_i = 0$, then selects the weight which has the max absolute value of the differential amount and decreases $F(\boldsymbol{w})$ until the differential value in terms of this weight becomes 0 by the Quasi-Newton method. The system acquires the globally minimum solution of $F(\boldsymbol{w})$ by iterating this task because the function $F(\boldsymbol{w})$ is the convex function in terms of the weight vector $\boldsymbol{w}$. In this iteration, the elements' number of the weight group of $w_i \neq 0$ increases one by one, and the values of all weights in this group are renewed by the BFGS formula used in the Quasi-Newton method.

TABLE I

MULTI CLASS LPBOOST VIA COLUMN GENERATION

Initialize: $u_{n,y} \leftarrow \frac{1}{N(|Y|-1)}, \beta \leftarrow 0, H \leftarrow \phi$

$While(true)\{$

$$h^+ \leftarrow \underset{h \in H_0}{\operatorname{argmax}} \left| \sum_{n=1}^{N} \sum_{y \in Y_{\setminus y_n}} u_{n,y}(h(x_n, y_n) - h(x_n, y)) \right|$$

$$\equiv \underset{h \in H_0}{\operatorname{argmax}}\, g(h, u)$$

$if(g(h^+, u) \leq \beta)$

$break$

$else\{$

$H_1 \leftarrow H_1 \cup h^+, \quad H_0 \leftarrow H_0 - h^+$

$$(u, \beta) \leftarrow \begin{cases} \min \beta \\ s.t. \begin{cases} \sum_{n=1}^{N} \sum_{y \in Y_{\setminus y_n}} u_{n,y} = 1 \\ \forall n, \quad 0 \leq \sum_{y \in Y_{\setminus y_n}} u_{n,y} \leq \frac{1}{\nu N} \\ \forall h \in h_1, \quad g(h, u) \leq \beta \end{cases} \end{cases}$$

$\}$

$\}$

$u \longrightarrow \boldsymbol{w}$ (dual transformation)

Return: $\boldsymbol{w}$

2) LPBoost via Column Generation: In the algorithm of LPBoost via column generation, each feature part is regarded as a weak learner, and it makes the weights of these weak learners sparse by adding the 1-norm regularization $|w|_1$ to the error function of the usual boosting algorithms. Demiriz converts this task to the task of minimizing the upper bound of the error function, and solving it by maximizing the margin. It was only for the two class, so we extended it for the multi class for this research. The concrete algorithm is shown in Tab.I where N is the data number, $|Y|$ is the class number, $Y_{\setminus y_n}$ is the class group Y except the right class y_n, H_1 and H_0 are the group of the selected features and unselected features respectively, and ν is the normalizing parameter satisfying $\nu \in (0, 1)$. Here, the size of the matrix u, which is the parameter after dual transformation, is $N \times (|Y| - 1)$ and each element corresponds to the weight of each class except the right class y_n per the data $\boldsymbol{x}_n$.

V. EXPERIMENT ON REAL DATA

A. Conditions of the experiment

For the experiment, we used human position movement data (Blue lines of Fig.3 right) with the ground truth behavior records for the three weeks life of a man who is in 20s. The arrangements of the rooms and the furniture is shown in the left figure of Fig.3. As the behavior record, we picked up 7 labels as "Sleep, Going out, Toilet, Bath, Cooking, Eating/PC at table, and the other" concerning the person's life pattern. The system calculates the precision of classification by comparing the classification results among these labels by the sensor data to the right records. In this examination, we picked up 5 segment data per behavior as the test data and calculated the classification precision by the cross validation.

Here, the total number of each behavior's segments for the period of time was from 20 to 60. Next, we set the radius of detection area as 0.7[m], and the sensor place candidate at intervals of 1[m], but removed candidates if the person rarely moves in the detection area during the learning period of time, judging by the records of the human movement. As the result, the 23 candidates are enumerated, shown in the right figure of Fig.3 as the green circles.

The system obtained the pyroelectric sensor reactions at each candidate place from the human movement life-flow, and enumerated all feature part candidates by this data. The number of the feature candidates was about 50,000 when the max sensor combination for a feature part 3, the max length of them 10 seconds, the most highest occurring rate of them for all behaviors is more than 20%, and the similar rate of the occurrence's result per behavior segments in the same sensor combination is less than 90%.

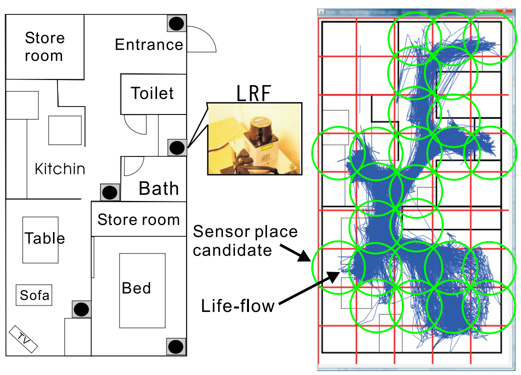

Fig. 3. House layout, life-flow and sensor candidates

B. Comparison result among the supposed algorithms

We compared the feature selection algorithms of SFFS, Grafting and LPBoost. Also, we compared these with the result of human intuitive judgment. We used the best result of many human intuitive judgement trials for the comparison. Fig.4 shows the results in terms of the precision for the learning data, that of the test data, the number of features and that of sensors. Here, we searched the best parameter values of each algorithm. The Grafting arrangement was so similar to LPBoost's that we omitted it.

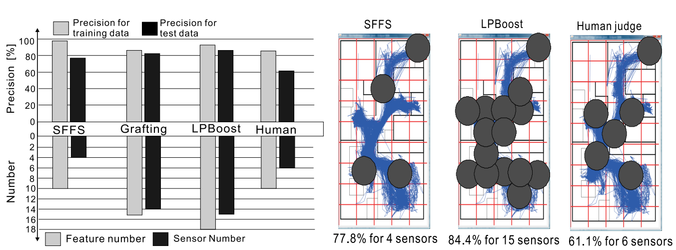

Fig. 4. Results of all selection methods

First, the precisions of all automation algorithms are higher than that of human judgment. In addition, when comparing the SFFS algorithm with the best human judgment result, the SFFS algorithm acquires much higher precision than human judgment with smaller sensor number. So, we could conclude that this result proves the significance of these automation algorithms. When comparing these three automation algorithms, the sensor number of SFFS was about four times smaller than that of the others, while the precision for the test data was lower by about 7%, and it suited for the aims of each algorithm; the SFFS algorithm is for minimizing the sensor number with quasi-optimized precision and the Grafting and LPBoost algorithms are for maximizing the precision even though using more sensors. We think the priority of reducing the cost of sensors and improving the precision may be judged by the user of this system, so it is important to obtain the graph like Fig.4.

When comparing the sensor arrangement results from Fig.4 right, the SFFS algorithm gives the pathway priority over the kitchen or bathroom which are tend to be selected by human intuition, and LPBoost mainly omits the candidates at the corridor toward the entrance, probably because these candidates reacts only when going out, so the entrance sensor can represent them.

VI. Conclusion

This paper suggested some methods for acquiring the optimized sensor arrangement and feature selection which classifies behaviors correctly by small number of sensors in the behavior classification system using the pyroelectric sensors.

The system enumerates many feature candidates and selects the most appreciate feature combination in terms of two ways. For the actual life data, we obtained about 86% precision by our multi-class LPBoost algorithm getting the globally optimized solution, and about 79% precision for 4 sensors by the SFFS algorithm making the sensor number as small as possible and keeping the quasi-optimized precision. In addition, we proved that these solutions are superior to the human intuitive arrangement in terms of the precision and the sensor number.

References

[1] S. S. Intille, K. Larson and E. M. Tapia and Jennifer S, *Using a Live-In Laboratory for Ubiquitous Computing Research*, Pervasive, pp. 349-365, 2006.

[2] T. Mori, R. Urushibata, M. Shimosaka, H. Noguchi, T. Sato, *Anomaly Detection Algorithm Based on Life Pattern Extraction from Accumulated Pyroelectric Sensor Data*, IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2545–2552, 2008.

[3] K. Oshima, R. Urushibata, A. Fujii, H. Noguchi and Hiroshi Noguchi ,M. Shimosaka, T Sato, T. Mori, *Behavior Labeling Algorithms from Accumulated Sensor Data Matched to Usage of Livelihood Support Application*, IEEE International Symposium on Robot and Human Interactive Communication, pp. 822–828, 2009.

[4] Y. C. Wang, C. C. Hu, Y. C. Tseng. *Efficient Deployment Algorithms for Ensuring Coverage and Connectivity ofWireless Sensor Networks*, Proc. of International Symposium on Information Processing in Sensor Networks, pp. 36-45, 2004.

[5] H. Wang, K. Yao, G. Pottie, D. Estrin. *Entropy-based sensor selection heuristic for target localization*. Proc. of International Conference on Wireless Internet, pp. 114-121, 2005.

[6] P. Pudil, J. Novovicova, J. Kittler, *Floating search methods in feature selection*, Pat. Recog. Ltrs, Vol.15, pp.1119–1125, 1994.

[7] S. Perkins, K. Lacker, J. Theiler, *"Grafting: Fast, Incremental Feature Selection by Gradient Descent in Function Space"*, Machine Learning Research 3, pp.1333–1356, 2003.

[8] A. Demiriz and J. Shawe-Taylor, *"Linear Programming Boosting via Column Generation"*, Machine Learning 46, pp.225–254,2002.