Health Score Prediction using Low-Invasive Sensors

Masamichi Shimosaka, Shinya Masuda, Kazunari Takeichi, Rui Fukui and Tomomasa Sato
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, JAPAN
{simosaka, masuda, takeichi, fukui, tsato}@ics.t.u-tokyo.ac.jp

ABSTRACT
Scores of health state for elderly people are regarded as important in nursing or medical fields. On the other hand, gaining the scores needs nurses to execute questionnaires. Owing to this, the execution rate for the health assessment is still low in ordinary homes. To solve this problem, we propose a method to predict the health score by using low-invasive sensors. We adopt regression as the prediction method and construct features to absorb the individual difference. As a part of feasibility study of social participation for elderly people, we execute the survey of health state using questionnaires by a nurse and install low-invasive sensors in real life simultaneously. Experimental result in the feasibility study shows a promise of the score prediction from sensor data. In addition, the result suggests that the extraction of features related to living behaviors improves the accuracy compared to using raw sensor data.

Author Keywords
instrumental ADL, sensor, machine learning

ACM Classification Keywords
J.3 LIFE AND MEDICAL SCIENCES:Health: , I.5PATTERN RECOGNITION:General

General Terms
Experimentation, Measurement, Performance

INTRODUCTION
The scores of health state are needed as the expression of the change of patient state or as information for medical decision-making. Especially, the scores for elderly people are highly needed. The health assessment for elderly people is important to keep their health as well as to prevent health deterioration. In addition, nurses utilize the score as a criterion for nursing care level. As the health indicator for elderly people, ADL (Activity Daily Living), instrumental ADL and so on are often employed in general. Although these indicators are regarded as important from the view of nursing or medical fields, the execution rate for these assessments, especially over long period in ordinary homes, is still low. One of the reasons for low execution rate is that these assessments are based on questionnaire or observation by nurses. Besides, questionnaire survey holds the problem that the answers for questions often change depending on subject's mental condition or the way of question and can include false answers due to subject's pride. Thus, we propose a method to predict the score of health state by using low-invasive sensors. The assessment with sensors is expected to reduce the number of needed nurses and the uncertainty of the score. Thanks to observation continually by sensor, the health decline is also expected to be early detected. Furthermore, clarifying the relationship between sensor data and the score of health state could derive new knowledge in the medical fields.

Current health assessments are based on the subjective opinion told by subjects. One example of questionnaires is “Can you go out alone or not? Yes / No”. On the other hand, the studies for the health score report the relationship between living behaviors, such as the frequency of going out or the speed of walking, and the health scores [1, 3]. Although these studies focus on that living behaviors can affect their health, these studies also suggest the possibility that the recognition of behavior can evaluate the state of health. The recognition of more other behaviors can be relevant to health score and the key of the prediction. Moreover, in the fields of pattern recognition, the pattern of life or out going is recognized by sensors installed in a house [5, 7]. Our research pays attention to the relationships between the health score and sensor data. Our research tries to predict the health score using sensors installed in houses of elderly people. Considering the installation of sensor in real life, the sensor invading their privacy should be forbidden. Therefore, the sensor should have low-invasive and we try to install pyro-electric sensors.

We handle the health score prediction as a regression problem, which infer the health score from the sensor data. However, raw sensor data depend on not only their health but also the environmental or individual characteristics, e.g. room layout, attached position or individual life pattern, which are unrelated to their health directly. Therefore, the features expressing their health removing irrelevant or nuisance information due to the environmental specification will be needed. We construct new features expressing daily behavior from sensor data to improve the accuracy.

UbiComp '12, Sep 5-Sep 8, 2012, Pittsburgh, USA.
ACM 978-1-4503-1224-0/12/09.

As a part of social participation project with employment support for elderly people, a nurse conducts questionnaire to obtain the health scores and we collect the sensor data simultaneously over 10 months. For the obtained data, we try to predict the health score from sensor data. Experimental results show that sensor data can predict the health score and suggest that the construction of features related to daily behavior improves the prediction accuracy compared to using raw sensor data.

The remainder of this paper is organized as follows. First, this paper explains the obtained data in detail, next this paper explains the prediction method and then this paper presents experimental results. Finally, conclusion is described.

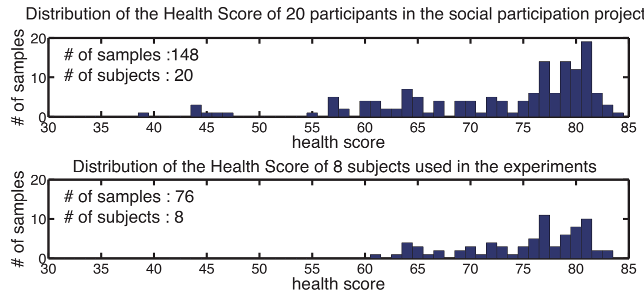

Figure 1. distribution of health score. A sample corresponds to one questionnaire (one person, one month). Upper figure expresses the 20 subjects in social participation project and lower figure expresses 8 subjects, who live alone and allow to install sensor in common positions.

SENSOR AND HEALTH SCORE DATASET

Subjects

The investigation of health score and the collection of sensor data have been performed in Hokuto City, Yamanashi Prefecture, Japan. The elderly people in the city cooperate in our investigation as a part of the social participation project with employment support for elderly people. The number of participants is 20. We install low-invasive sensors in their house and take a questionnaire survey by a nurse to obtain the health score simultaneously once a month over 10 month, from April 2011 to January 2012. Family structures of subjects are the followings: 11 elderly people living alone, 8 elderly people living with a partner and 1 person living with three persons. As subjects whose data is used in experiments of this study, people who do not live alone are excluded due to installed sensor characteristics. In addition, we also exclude people who do not allow installing sensors in the rooms following, *Kitchen, Bedroom, Entrance and Living room*, which are regarded as important rooms to express the life pattern. Finally, the number of subjects to be targeted in this paper among all the 20 participants is reduced to 8. The 6 persons in these subjects are women and others are men. Their ages are between 75 and 89. As the experimental data, the dataset has samples for 8 people over 10 month. The number of samples in the dataset is 76 owing to the lack of sensor data or questionnaires. On the other hand, with regard to all the 20 subjects, 13 subjects are women and other 7 subjects are men and their ages have the range of 68–89. The number of total samples is 188, which are composed of 20 persons over 10 month, owing to the lack of sensor data or questionnaires.

Health Score based on iADL

As the health score for elderly people, ADL (Activity Daily Life), instrumental ADL and so on are usually used. In our research, we adopt the score based on iADL (instrumental Activity Daily Living). iADL is a measure of daily living function, which is proposed by Lawton [2], and the indicator has 8 items following. *1) Using the telephone, 2) Shopping, 3) Preparing food, 4) Housekeeping, 5) Doing laundry, 6) Using transportation, 7) Handling medications and 8) Handling finances*. In general, the score with iADL is calculated by whether or not a subject can perform the behavior for each item above. In our research, a nurse conducts the 22 questions related to above iADL items and the score for each question ranges from 0 to 4 based on criteria of ICF [4]. Finally, the total score is calculated by adding together and our health score has the range of 0–85. The concrete questions are described in APPENDIX. The score distribution is shown in Figure 1. The average of score is comparative high because almost subjects are healthy in our survey.

Installed Sensor

We attached sensors on the wall in the residence of subjects. Considering the installation into real life, privacy preserving is essential. In addition, sensors taking low cost are better. From these reasons, pyro-electric sensors are chosen. The pyro-electric sensors are low-invasive sensors and low in price. Attached positions of pyro-electric sensors as the project are at the most following 8 positions: *Kitchen, Bedroom, Entrance, Back door, Living, Dining, Veranda and Toilet*. From these positions, our experiment use only 4 pyro-electric sensor data attached to 4 common positions: *Kitchen, Entrances, Bedroom and Living room*. These positions are also selected in another study [5] to grasp the patterns of daily life. The information obtained from pyro-electric sensors is a count of detection of people movement per one minute, however, the number of count obtained by installed pyro-sensor is bounded and 16 step values are obtained. 1440 data are recorded in one day for each sensor. It is expected that active behaviors increase the count of sensor reaction. The average of raw sensor data in one month for one person is shown in Figure 2 as a example. We extract various features from pyro-electric sensor data and use them to predict the health score.

PREDICTION APPROACH

The raw data obtained from pyro-electric sensor is composed of data per one minute. On the other hand, the health scores are obtained once a month. Therefore, it is necessary to convert sensor data accumulated in one month, which has 1440×30 or 31 response data, to one vector of predictor variables.

Features using mean and variance

As a simple way to construct predictor variables, the mean and variance of sensor data are often used. The extracted features using mean and variance have 16 dimensions and are

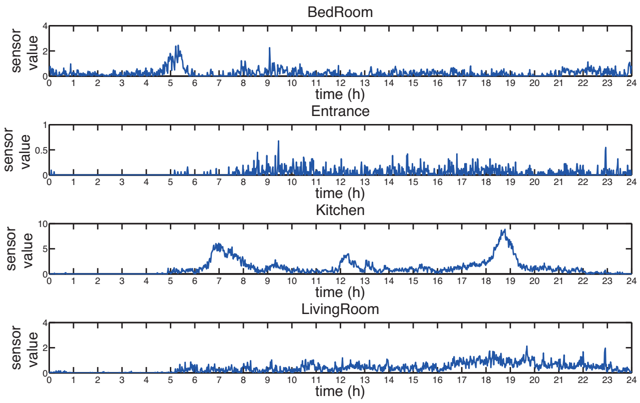

Figure 2. sensor reaction: this figure shows the average of sensor reaction in one month for one subject, whose health score is 79 in the month, for 4 positions (Bedroom, Entrance, Kitchen, LivingRoom).

composed of "the mean of day-mean", "the mean of day-variance","the variance of day-mean" and "the variance of day-variance" for 4 sensors respectively. In this process, after day-mean or day-variance is calculated for 1440 data in each day, mean or variance is calculated for values obtained in first process in one month. Although the count of pyro-sensor reaction is expected to reflect subject's amount of activity, these features are also influenced directly by environmental differences, e.g. room layout or attached position. These environmental differences should be eliminated. Therefore, the invariant features for the individual difference unrelated to their health should be investigated.

Features related to Living Behavior

The features expressing their health without being influenced by irrelevant information should be extracted from sensor. Here, the features related to living activities are expected to be useful information because the health score is based on subject's life behaviors and the behaviors are invariant for subjects' environmental difference. For example, the reaction of the sensor attached to a kitchen in the noon is expected to be involved in cooking. Then, if the sensor reacts frequently in the noon, the subject may tend to have the living ability to prepare eating (See Figure 2). As other example, because the increase of a bedroom sensor reaction in the morning can be interpreted as subject's wake-up (See Figure 2), grasping of the timing of wake-up over one month may lead to the information of subject regular life. As described above, much information of living activities exists in sensor data and they are invariant for subjects' environmental difference. We extract 6 kinds of features from sensors (See Table 1). The details are described as follows. 1) The count of sensor reaction over specific threshold in one month. This is related to amount of activities in each room. The thresholds for our experiment are set to 2, 8, 12 and the dimension of this feature is 3×4, i.e. 3 thresholds and 4 sensors. 2) the count of states that different sensors react at same time. This feature may express the existence of visitors and may express subject's social connection. The dimension of this kind of features is 6 because 2 sensors are selected from 4 sensors. 3) The number of the days when the sensor attached in kitchen reacts over threshold in the period time of lunch or dinner. This feature may express the ability of preparing food. The dimension of this feature is 2 by lunch and dinner. 4) The count of states that all 4 sensors react more than certain threshold within a period of time. These features may be related to housekeeping or cleaning because subjects move around in their house. The dimension of this feature is 1. 5) The length of time when no sensor react in daytime. This feature is regarded as the time of out going. The feature dimension is 1. 6) The variance of one month about time when the bedroom sensor reaction begins to exceed a threshold in the morning or at night. This feature can be related to subject's regular life because the time can reflect subject's time on wake-up or going to bed. The dimension of this feature is 2, wake-up and going to bed. As described above, we obtained 24 features from sensors. These features are designed to be related to iADL 8 items, which are *using the telephone, shopping, preparing food, housekeeping, doing laundry, using transportation, handling medications and handling finances*. Although some of items are not included in the proposed features directly, proposed features are expected to include the information of the items indirectly thanks to the expression of subject's life activities. These proposed features extracted from pyro-electric sensors are written as ϕ_d $(d = 1, \cdots, D)$ in this paper. D is the feature dimension, $D = 24$ in this case.

feature name	mainly related behavior	dim.
1) over threshold	activity in each room	12
2) same time reaction	exsitence of visitor	6
3) kitchen in eating time	preparing food	2
4) reaction at 4 sensors	housekeeping or cleaning	1
5) no reactions time	out going	1
6) var. of reaction start	wake-up, going bed	2

Table 1. Features List

Prediction Method

We use linear regression, $\hat{y}_i = \sum_{d=1}^{D} w_d \phi_d(x_i)$, as prediction methods $(i = 1, \cdots, N)$, where $\hat{y}_i$ is predicted health score for ith sample and N is the number of samples. As described above, ϕ_d is features and $\boldsymbol{x}_i$ indicates sensor data. Given N_{tr} training data $\{x_i, y_i\}_{i=1}^{N_{tr}}$, the weight parameter $\boldsymbol{w} = [w_1, \cdots, w_D]$ is learned by

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \sum_{i=1}^{N_{tr}} (y_i - \hat{y}_i)^2 + \Omega(\boldsymbol{w}), \tag{1}$$

where $\Omega(\boldsymbol{w})$ is the product of a regularization term. This term prevents over-fitting. In our research, we use 2 types of regularization term, L2 regularization term and L1 regularization term. L2 regularization term is expressed by $\Omega(w) = \lambda \left(\sum_{i=1}^{d} w_i^2 \right)^{1/2}$, where λ is a hyper-parameter to decide the influence of regularization. The hyper-parameter is determined by cross-validation and the concrete way is described in Experimental Method section. Equation (1) including L2 regularization term is called as Ridge regression. On the other hand, L1-regularization term is expressed by $\Omega(w) = \lambda \left(\sum_{i=1}^{d} |w_i| \right)$. Equation (1) including L1-regularization term is known as the Lasso [6], which has a shrinkage and selection function for linear regression. Therefore, obtaining knowledge of useful feature can be expected in the use of Lasso. We describe useful features found by Lasso in our experiment.

EXPERIMENT

In our experiment, we predict the health score from sensor data and examine the validity of the prediction method. In addition, we compare proposed features with simple ones.

Experimental Method

As described in Dataset section, our data set has 76 samples. One of the experimental purposes is an examination of method's validity. Therefore, for this small size dataset, we train weight parameters using 75 samples and prediction 1 sample. We repeat this process 76 times and compare ground-truth value with prediction value, we call this training and evaluation method as "LOSOE (Leave-one-sample-out Evaluation)". As error criteria, we use mean of absolute error (ERR) and mean of squared error (MSE), which is a loss function appeared in Equation (1). This hyper-parameter in Equation (1) is determined using LOSOE in training 75 samples. As a compared criteria, "mean prediction approach", whose prediction is the mean of training data, is also used. With respect to features, used features are 24 proposed features related to Living Behavior and mean-variance 16 features.

In addition, considering installation into real life, the installation period may be needed to absorb individual differences, e.g. room layout or individual life pattern. Therefore, we also experiment to grasp the change of accuracy corresponding to the installation period from zero months to seven months. In this installation experiment, if the installation period is N months, training samples are other subjects' all samples and predicted subject' N samples in installation period, which correspond to the first N months in experiment periods. Then, test samples are each subject' samples excluding N installation samples. This process is repeated in 8 times (8 is the number of subjects) , and the loss criteria are calculated. We call this training and evaluation as "*m*-LOUOIE (Leave-One-User-Out Installation-Evaluation)", where m expresses the installation months. We employed proposed 24 features in this experiment.

Experimental results and discussion

methods	features	ERR	MSE	max-error
ridge	proposed features	2.41	9.39	15.29
ridge	mean and var	4.69	30.52	7.73
lasso	proposed features	2.36	9.46	23.46
lasso	mean	4.02	32.78	7.85
mean	————	4.82	33.78	14.39

Table 2. Experimental result by LOSOE: ERR, MSE and max of absolute error.

Experimental result with LOSOE is shown in Table 2 and Figure 3. ERR and MSE of 'mean prediction approach" are 4.82 and 33.78 respectively. The results of ridge regression using proposed features are 2.41 ERR and 3.39 MSE, which are much better than the one of "mean prediction approach" and the one using '"mean-var". In the case of lasso, a similar tendency can be seen (See. Table 2). These results shows proposed features can improve prediction accuracy in both ridge regression and lasso compared to using mean and variance features. Therefore, as we proposed, extraction of features related to living behavior from sensor data leads to be able to improve accuracy greatly. The reason of improvement is considered to be invariant of proposed features for subjects' environmental difference. Moreover, our approach can estimate with a resolution of each question because ERR score is less than the score of one question thanks to using proposed features. In addition, the weight parameters obtained by Lasso can express the feature usefulness because Lasso has shrinkage and selection function. As the useful features whose weight parameters are in the ten biggest absolute in all weight parameters, 6 threshold features (Kitchen-th.8 , Kitchen-th.12, Kitchen-th.2, Bedroom-th.8, Living-th.12, Living-th.8) and 4 reaction at same time features (Entrance-Kitchen, Kitchen-Living, Bedroom-Entrance, Bedroom-Kitchen) are selected. Several features in above 10 features are related to Kitchen. Therefore, activity in kitchen is likely to be related to the health score strongly.

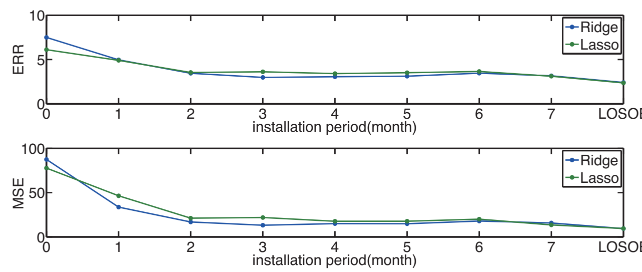

Figure 4. the result of m-LOUOIE corresponding to the change of installation periods m (m = $0, \cdots, 7$) . LOSOE results are also shown.

The results corresponding to installation periods from 0 months to 7 months and above LOSOE result is shown in Figure 4. As you can see, ERRs and MSEs tend to be reduced as installation periods increase. In addition, the influence of more than two-months installation periods are not so much different compared to the influence of using two months. Therefore, this results show two month installation period can be one of the criteria for installation period in this dataset.

Future Work

As you can see in experimental result, construction of features has great effect to improvement of accuracy. Therefore, more features should be investigated. To do that, more samples should be collected because the increase of features leads to the need of more dataset to evaluate correctly. In addition, obtaining more varied data, such as including the change of health scores in one subject or subjects having low health score, is also one of the future works. This is because almost subjects in our current data set are fairly healthy,

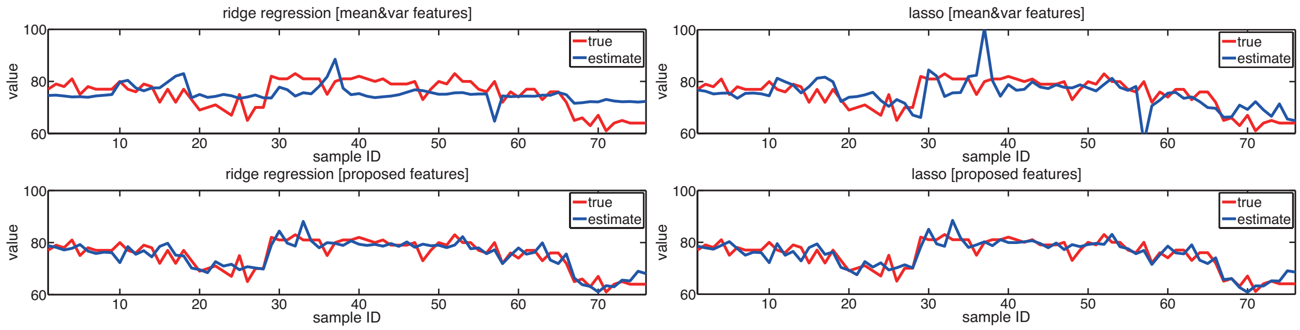

Figure 3. Left :Ridge result , Right: Lasso result. Red line shows the ground-truth obtained from questionnaire and blue shows prediction result by sensor for each sample.

however, recognition of low health score or prediction of the change of health score from sensors is also important problem. Besides, the examination about the prediction of each questionnaire answers, which are including the behavior that sensor data can not capture directly (e.g. using telephone), is one of the feature works.

CONCLUSION

In this paper, we proposed the method to predict the elderly person's health score by low-invasive sensors. The contribution of this paper is that the experimental results using real data show the promising approach of health score prediction using regression. In addition, our experimental results also show construction of features related to living behavior would improve the accuracy. As a future work, obtaining large size data with large variance is important. Such data can lead to examine more validity of prediction method or features and enables research on prediction of the declining of health.

APPENDIX

Questionnaire

We conduct 22 questions to obtain the health score. The questions are based on the iADL 8 items. Each question's maximum score is 4 based on ICF, excluding several questions that have 1 or 3 maximum score. 22 questions are following: *(Using telephone 1) Do you feel it difficult to call your family or friends using telephone? (Using telephone 2) Do you call anyone in daily life? (Using telephone 3) How often do you call in a week? (Using telephone 4) Can you receive a phone call? (Shopping 1) Do you feel it difficult to go shopping alone? (Shopping 2-a) Do you go shopping alone actually? (Shopping 2-b) Do your go shopping with anyone, such as family or friends? (Shopping 3) How often do you go shopping alone or with anyone? (Preparing food 1) Do you feel it difficult to prepare food alone,for example, cut vegetables or bake meat? (Preparing food 2-a) Do you prepare food alone in actual? (Preparing food 2-b) Do you prepare food with anyone's help in actual? (Preparing food 3) How often do you prepare food alone or with anyone? (Housekeeping 1) Do you feel it difficult to wash dishes or clean room alone? (Housekeeping 2-a) Do you do housework alone in actual? (Housekeeping 2-b) Do your do housework with anyone? (Housekeeping 3) How often do you do housework alone or with anyone's help? (Doing Laundry 1) Do you feel it difficult to do the washing? (Doing Laundry 2-a) Do you do the washing alone? (Doing Laundry 2-b)Do you do the washing with anyone? (Doing Laundry 3) How often do you do the washing alone or with anyone? (Using transportation 1) Do you feel it difficult to use transportation, e.g. bus, taxi and train, or drive alone? (Using transportation 2) Do you go out alone in actual? (Handling medications 1) Do you feel it difficult to handle medicines in right way? (Handling medications 2) How do you manage the medicine when you need to handle medicine? (Using Handling finances 1) Do you feel it difficult to manage your finance or the payment of the bill? (Using Handling finances 2) Do you manage your money by yourself?* The question having 2-a and 2-b takes maximum score between 2-a and 2-b.

REFERENCES

1. A. Kono, I. Kai, C. Sakato, and Z. R. Laurence. Frequency of going outdoors predicts long-range functional chnage among ambulatory frail elder living at home. *Archives of Gerontology and Geriatrics*, 45:233–242, 2007.

2. M. Lawton and E. M.Brody. Assessment of older people; self-maintaining and instumental activiityes of daily living. *Gerontologist*, 9(3):179–186, 1969.

3. K. Okumiya, K. Matsubayashi, and T. Nakamura. The timed up and go test and manual button score are useful predictors of functional decline in basic and instrumental adl in community-dwelling older people. *J Am Geriatr So*, 47(4):497–498, 1999.

4. W. H. Organization. *International classification of functioning, disability and health (ICF)*. 2001.

5. M. Shimosaka, T. Ishino, H. Noguchi, T. Sato, and T. Mori. Detecting human activity profiles with dirichlet enhanced inhomogeneous poisson processes. In *ICPR*, pages 4384–4387, 2010.

6. R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Statis. Soc. B*, 58:267–288, 1996.

7. S. Tominaga, M. Shimosaka, R. Fukui, and T. Sato. A unified framework for modeling and predicting going-out behavior. In *Pervasive2012*.