Robust Health Score Prediction from Pyro-Sensor Activity Data based on Greedy Feature Selection

Masamichi Shimosaka
Department of Computer Science
Tokyo Institute of Technology
Tokyo, Japan
simosaka@miubiq.cs.titech.ac.jp

Qiyang Zhang
Department of Computer Science
Tokyo Institute of Technology
Tokyo, Japan
zhang@miubiq.cs.titech.ac.jp

Kazunari Takeichi
Department of Mechano-Informatics
The University of Tokyo
Tokyo, Japan

***Abstract*—Automated activity assessment using IoT/smartphone sensors becomes great popular in ubiquitous computing research community recent year thanks to the enhancement of mobility and IoT sensing. In these researches, owing to the great success of statistical machine learning technique called Lasso, the work offers the interpretability of the model. However, in some sparse feature condition, Lasso as a l_1 regression method could not give a satisfying result for prediction precision and feature selection. In this paper, we propose a new prediction scheme using greedy feature selection method which is expected to be effective under large scale feature in limited number of dataset. With the help of the new scheme, we could solve the overfitting problem when using l_1 regression as well as giving satisfying prediction result. Experimental results using longitudinal pyro-sensor dataset of health score of elderly people show that our new scheme offers better interpretability as well as achieves better prediction accuracy compared with Lasso**

***Index Terms*—Instrumental ADL, Feature Selection, Activity Health**

I. Introduction

Recent years, the problem of aging society is rising up. In order to assess the health condition of elderly people, and provide better healthcare for them, WHO has promoted to provide indicators for elderly people since 1984 [5]. After that time, many indicators were provided, including ADL (Activity Daily Living), iADL (instrumental ADL) and so on. These indicators contain question related with daily activity like “Do you prepare food alone? Yes/No.” and they are important for doctors and nurses to assess elderly people health condition.

It is natural to think that the daily behavior is highly related to such assessed health status. As pioneering work, some researches gave statistical results reporting the relationship between living activity behaviors and health score [2] [4]. This inspires us to use trends of one’s behavior as a cue to infer one’s health condition. However, one’s life style or behavior trends as well as one’s health score is obtained via questionnaire with help of nurses. Due to the requirement of large cost of human resources on questionnaires, this situation prevents us from giving frequent assessment. To solve the problem, researchers have actively explored to leverage IoT technologies that capture and report one’s daily activity data pervasively to understand one’s health assessment in a reasonable manner.

In the last decade, a few work on ubiquitous and pervasive computing communities provides a automated health assessment techniques using regression from the statistics of IoT sensor activity behavior logs [7] [1] [12]. Among these researches, it has been shown that activity feature extracted from sensor data is closely related with people’s health score. So designing and extracting feature data related with kinds of activity is one important part for our health score prediction work.

Considering the regression problem with different kinds of features and limited number of samples of data owing to the application scenario, the prediction performance is prone to be poor when using simple machine learning for regression such as l_2 regularized ridge regression. Typically, the number of samples annotated with health score is too limited, thus most of the statistics from the IoT sensors are irrelevant to health scores. In order to reduce or avoid the effect of overfitting, regression method with feature selection function is used. Lasso, as a l_1 regularization method, is popularly used in many researches [11]. It is known to be useful to solve the overfitting problem. However, in some practical situations, the usage of Lasso could not effectively solve the overfitting problem. In order to better deal with the overfitting problem and give a precise prediction score result, new feature selection algorithm is needed for the health score prediction or other kinds of life assessment work.

Dealing with the important part and the shortcomings in this field of work, we propose our work by designing new kind of features representing activity as well as provide a brand-new scheme to select feature and train regression model.

In this paper, we provide the following three contributions: (1) we design a new kind of activity features by using hourly sensor data in order to better understand the activity of elderly people related with health; (2) we propose a new scheme for numeric health score prediction using l_0 regularized learning such as FoBa; (3) we show the performance of prediction, which is improved by using proposed feature and applying proposed scheme to achieve reduction of overfitting issue, by real world dataset.

The structure of this paper is as follows. We present related work about health and life assessment work in Section II. We discuss the problem setting and baseline algorithm for these kinds of numeric assessment work in Section III. In Section IV and Section V, we separately detail our proposed method about prediction model training and feature extraction . We show our experiment content, setting and comparison result between proposed scheme and Lasso in Section VI. Finally, we present our conclusion in Section VII.

II. Related Work

A. Health and Life Assessment Using IoT/Smartphone Sensors

Health and life assessment using IoT/smartphone sensors becomes a popular topic in recent year ubiquitous and pervasive computing community. Many researches have explored to deal with kinds of assessment problem, like elderly people health condition [7], mental health condition [9] and study life condition [11], using nonnumeric feature extraction and correlation analysis method as well as numeric prediction and regression analysis method to solve their specific problems.

Robben's work [6] gained daily activity data from sensor network, and used Principle Component Analysis (PCA) to inspect daily pattern. They considered the pattern of spent time in each room is related with self report ADL metrics and advocated to use the pattern to assess metrics. But their work didn't provide any correlation analysis or numeric prediction. StudentLife project [10] gained data from students in 10 weeks from smartphone sensors including activity, conversation, and mobility. They analyzed the correlation between the sensor data and the GPA as well as students' mental health condition, which could represent two important parts in student life.

Except for these researches about feature extraction and correlation analysis, there are some other researches deal with numeric prediction and regression analysis. Shimosaka's work [7], which is our previous work, dealt with health score prediction with elderly people. It used pyroelectric sensors to gain elderly people's activity data at home, and extracted feature related with activity. This work used regression method for building numeric score prediction model and giving feature analysis. Huang's work [1] considered people's mental health score. They used GPS information and Point of Interest (POI) data to gain the activity dataset from smartphone users. Started with analyzing the information about where users go and the place with POI label, the author gave the result between the information and mental health score by correlation analysis as well as building a score prediction regression model. The result showed the relation between the mental pressure and users' location.

SmartGPA [11] and CrossCheck [9] [12], which are researches following StudentLife [10], used the similar dataset from smartphone sensors, and extracted features regarding the mean value and the tendency of daily activity. Their work also used regression methods to give a numeric prediction of the GPA or the mental health score.

These researches always used sensor data from IoT/smartphone sensors and extract features from that. They provided good ideas for designing features. By their correlation analysis and regression analysis, we could see that it is reasonable to use activity feature and regression method to predict some certain kinds of assessment score.

B. Feature selection technique

In the literature on health score prediction from IoT sensors [1] [7] [11], authors chose to use l_1 regularization method, in which Lasso [8] is known to be a de-facto standard technique, to select feature and train prediction model.

However, the usage of Lasso tended to be a drawback of their work and had bad influence on their prediction performance due to the fact that l_1 regularized training is just the simplest in terms sparse modeling. For in these researches, different kinds of features were extracted in order to represent different aspects of people activity and reduce bias. With large size of features and limited size of samples, although most work relied on Lasso to select feature and reduce overfitting, there were other researches showing that Lasso is not reliable to select most related feature [3] [14]. It meant prediction model trained by Lasso might give worse prediction performance in sparse feature condition. According to this situation, Lasso is no longer suitable for our work. Inspired by the success of l_0 regularized training [3] [13] [14], we consider to use l_0 regularized training method in this paper. To the best of our knowledge, it is the first work to apply l_0 method for feature selection for this kind of health and life assessment work, and compare it with Lasso using real world data to confirm the validity of our inspiration.

III. Problem Setting and Baseline

A. Problem Setting

In this work, we choose to use a subset of the sensor data and the iADL score data obtained from the residence with solitary elderly people, which was started to gain in our former work [7]. The dataset was gained in Japan. The dataset contains raw sensor data from pyro-sensors which are set in different rooms in elderly people's house, and iADL score counted from questionnaire gained with help of one nurse every month.

However raw sensor data could not directly be used, for this data is not only related with health condition, but also depend on other content of environmental or individual characteristics. In order to better understand the daily activity of elderly people, we design and extract high level behavioral features, such as taking meal and going out. These features are shown to be related with activity score in related researches.

With related feature data and iADL score, we handle our score prediction as a regression problem, which could give numeric prediction score as well as showing the relationship between sensor data and iADL score data. By using linear regression method, we could build our prediction model, which is predicting iADL score as outcome variable from a set of features.

B. Prediction via regression and Lasso

In order to better represent daily activity with features and avoid bias, different kinds of features with high dimension would be designed and extracted. However, this high dimension feature dataset is prone to be overfitting with limited number of training samples. To deal with overfitting problem on linear regression, Lasso [8] is frequently used, where the regression is formulated as follows:

$$\hat{y}_i = \sum_{d=1}^{D} w_d \phi_d(\boldsymbol{x}_i), \tag{1}$$

where ϕ_d indicates a d-th description derived from the sensor data $\boldsymbol{x}_i$ obtained in i-th assesement. Given N_{tr} training data $\{\boldsymbol{x}_i, y_i\}_{i=1}^{N_{\mathrm{tr}}}$, the weight parameter $\boldsymbol{w}^\top = [w_1, \ldots, w_D]$ is learned by

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + \Omega(\boldsymbol{w}), \tag{2}$$

where $\boldsymbol{X}$ denotes design matrix whose i-th row and j-th column represent $\phi_j(\boldsymbol{x}_i)$, $\boldsymbol{y}$ is a vector containing $y_1, \ldots, y_{N_{\mathrm{tr}}}$, and $\Omega(\boldsymbol{w})$ is the product of a regularization term.

In Lasso, or say l_1 regularization method, this term is expressed by $\Omega(\boldsymbol{w}) = \lambda\|\boldsymbol{w}\|_1$. In this regularization term, $\lambda > 0$ helps Lasso select feature and reduce the overfitting problem. If λ is set too small, the regularization item will have no significant effect and the overfitting problem will remain. If λ is set too large, the bias problem will increase and the prediction result will become inaccurate. So we always use cross validation to determine λ value.

However, owing to the fact that l_1 norm is a simple but a loose relaxation of l_0 norm, Lasso does not work properly with the limited number of training data is available while very large irrelevant feature dimension is used. When we have linearly related features, Lasso will prefer to choose one "significant" feature and drop others. However, this disadvantage in Lasso's feature selection may lead to dropping of some most related features in practical situation. One more disadvantage is that when we want to find a sparse model with a high dimension dataset, Lasso may retain to use irrelevant features in the trained model which still easy to become overfitting. So in order to build a sparse feature prediction model with most related features, we drop this baseline algorithm in our former work, and propose a new scheme with feature selection algorithm to solve the drawback in Lasso.

IV. Robust sensor score prediction via greedy feature selection

In order to solve the drawbacks of using Lasso mentioned in previous section, we employ l_0 regularized training alternative to Lasso. Thanks to the non-convexity of l_0 norm, we leverage a greedy approach as an approximated optimization for l_0 regularized method [14].

The procedure to derive our model presented in this paper can be summarized as follows: feature selection phase and optimizing regression parameters selected feature set. The Fig. 1 shows the architecture of proposed scheme including these two phases:

Step 1: Extract the activity feature dataset following Section V;

Step 2: compute the whole iADL score by summing every item in questionnaire (every iADL score corresponds to its sample);

Step 3: feed the feature dataset as input data into the l_0 greedy optimization to obtain the selected feature indices;

Step 4: feed the iADL score as supervised data into the l_0 greedy optimization to obtain the selected feature indices;

Step 5: feed the selected feature dataset as input into Ridge regression;

Step 6: feed iADL scores into Ridge regression as input label for training;

Step 7: use trained Ridge regression model to give iADL score prediction

A. Greedy l_0 regularized feature selection with FoBa

First in the feature selection phase, we employ FoBa [14] to build a sparse feature set from original massive feature set. FoBa is a l_0 regularization method, which add and remove features in a greedy but balanced way. Compared with Lasso, FoBa is reported to prevent the bias and lead identical solution to the ground truth model [14]. In other words, FoBa could work more reliable to selected most relative features, when we need to get a highly sparse solution.

In FoBa algorithm, every time we need to minimize the loss function to add or remove one feature. Here following the original FoBa algorithm, we still choose to use least squares loss

$$Q(\boldsymbol{\beta}) = \frac{1}{n}\left\|\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right\|_2^2, \tag{3}$$

where $\tilde{\boldsymbol{X}}$ indicates a subset of original design matrix $\boldsymbol{X}$, $\boldsymbol{\beta}$ indicates the corresponding coefficient of selected feature (i.e. sparse vector of $\boldsymbol{w}$). The algorithm iteratively adds one new feature into the feature data matrix or removes one selected feature from feature data matrix, and minimizes the loss $Q(\boldsymbol{\beta})$ with updated feature data matrix. The added or removed feature at each iteration will be the one with the lowest minimized loss $Q(\boldsymbol{\beta})$. Here is another rule in FoBa for removing one feature is that when the change of the loss between after-removing and before-removing is larger than certain threshold, which means the cost of removing this feature is too large, this feature will not be removed from selected group. So it means the removing of feature may not happen at every iteration.

FoBa's feature selection will stop when the sparsity (number of selected features) reaches certain threshold or the change of loss is less than certain threshold. Even if FoBa is a proper and simple l_0 regularized training with the given training dataset, we should take care the sample bias problem on our real world dataset. Therefore, we employ 10-fold double loop cross validation to seek the best optimized sparse feature subsets.

B. Ridge regression with the selected feature sets

Even if we obtain a good selection result from FoBa, it should be noted that the model without any regularization does not always lead the best performance to the unknown data.

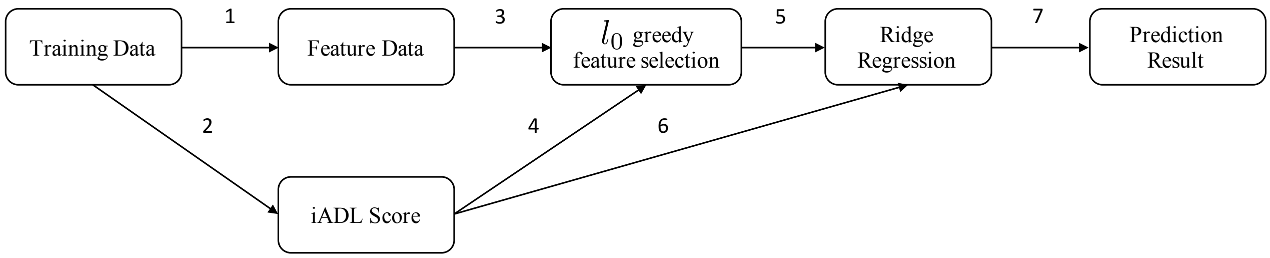

Fig. 1. The scheme including following steps: (1) extract features from the sensor dataset; (2) compute iADL scores; (3) feed the feature dataset into FoBa algorithm; (4) feed iADL scores into FoBa algorithm; (5) use the selected feature dataset as input to Ridge regression; (6) feed the iADL scores into Ridge regression as input label; (7) give prediction with trained model.

Actually, our empirical evaluation confirms that further training such as l_2 regularized with the selected feature sets further prevents the bias issue. Therefore, we employ the second phase as follows.

In the second phase of regression model training, we need to feed the selected feature dataset as the input and iADL scores as the label into a regression algorithm. Here we choose to use Ridge regression, which is a l_2 regularization method without feature selection function. For we have already completed the feature selection work, we don't need a further feature selection anymore. By applying Ridge regression with a 10-fold cross-validation to optimize the hyper-parameter of the l_2 regularization term, we could finally get our trained model.

V. Living Activity Feature from Pyro-sensor Dataset

Feature design and extraction is one important aspect, for it may give bad performance of prediction or give bias result if we get a limited and unrelated feature set. In our previous work, we extract 24 dimensions of activity feature (e.g. over threshold activity, making meal and going out) and 16 dimension of statistical feature using mean and variance. However, considering this feature set, it is too vague to only use daily statistical feature data (e.g. mean value of daily sensor data sum in one month, over threshold sensor daily sum) to represent daily activity. Here we propose to extend it with detailed hourly sensor data, which could represent more detailed activity and help us better understand elderly people's living activity.

Hourly over threshold activity feature

We first consider the different phase of life-style. Different activity condition at different period in one day like daytime or night, may have different relation with health. Our "over threshold feature" in previous work [7] extract this feature by counting the sensor reaction over certain threshold during whole day in one month, which could not represent more detailed health related activity information at each hour. So considering that features at different hour could better indicate elderly people's health related activity, instead of the feature counting over threshold activity in a daily range, we divide this range into 24 hourly timezone. So we extract this feature by computing the sum of sensor reaction larger than certain threshold each hour during one month.

Except for dividing daily range into 24 hours, we also reconsider the number of thresholds. We believe the fact that sensor data's value could represent the activity intensity, so we choose to use different threshold for this over threshold feature, which could gain activity condition information of elderly people from different level of intensity in one month. In "over threshold feature" in our former work [7], we used to choose 2, 8 and 12 as thresholds for this feature. Considering pyro-sensor gained integer value from 0 to 15, the former three thresholds are neither uniform nor detailed. So we propose to use thresholds set to 2, 4, 6, 8, 10, 12 and 14, together with 0 threshold (no threshold), which are eight thresholds in total. We believe these eight thresholds working together with 24 hourly timezone could better indicate the health related activity for elderly people.

Hourly activity rate

With the similar consideration of phase in life-style, we design and extract this feature as:

$$r_t = \frac{s_t}{\sum_{t=1}^{24} s_t}, \tag{4}$$

Here r_t indicates the hourly activity rate in the t-th hour of the day, ,s_t indicates the sum of hourly sensor reaction at t-th hour of the day in one month. and $\sum_{t=1}^{24} s_t$ indicates the sum of daily sensor reaction in one month. This rate can indicate the different weight of activity during daytime and night regardless of threshold.

Hourly activity variance

In order to indicate the tendency of elderly people's living activity, we choose to use variance value of hourly activity in one month. This feature is extracted by calculating the variance of sensor reaction hour-mean as well as hour-var in one month. For we believe variance value could show the "stability" of activity at certain time, this stability is related with people's lifestyle and their health. For example, we could consider that a low variance at sleeping time or waking up time could indicate elderly people have a regular sleeping lifestyle, and this regular lifestyle is always considered to be a healthy pattern for elderly people.

The final feature list is shown below in Table V, the total dimension is 1,119 including new propose feature and feature in our former work [7].

TABLE I
OVERALL FEATURE LIST

feature name	related behavior	dim.
1) Hourly over threshold activity feature	times of activity in different level in each room	800
2) Hourly activity rate	different activity rate in day-time and night	96
3) Hourly activity variance	stability of lifestyle	192
4) Same time reaction	existence of visitor	20
5) Kitchen in eating time	preparing food	2
6) Reaction at 4 sensors	housekeeping or cleaning	2
7) No reactions duration and number of times	out going duration and frequency	2
8) Mean and var. of reaction start	waking-up, going bed	4

VI. Experiment

In our experiment, we give the prediction score result with our proposed scheme as well as Lasso. Using evaluation metrics for the prediction score and feature selection analysis, we present the different performance between proposed scheme and Lasso.

A. Dataset Description

In our Hokuto Elderly People Dataset, the participants are 20 elderly people aging from 75 to 89. The data acquisition period is from April 2011 to January 2012, and from July 2012 to January 2013.

For the content of sensor dataset, we set our pyro-sensors in *Back Door, Bedroom, Dining Room, Entrance, Living Room, Kitchen, Toilet and Veranda*. In this paper, we only use sensor data from *Bedroom, Entrance, Living Room and Kitchen* for our research, which are the most common rooms for all participants. The pyro-sensor record the value every 1 minutes, and the value is 0 – 15. So we get 1,440 data every day from each sensor.

We use this sensor dataset to gain the activity feature dataset, which dimension is 1,119 for each sample as we mentioned in the former part.

As for questionnaire and iADL data, the iADL score questionnaire contain the following content: *Using the Telephone, Shopping, Preparing Food, Housekeeping, Doing Laundry, Using Transportation, Handling Medication and Handling Finances*. These items could reflect elderly people daily living function's condition. The iADL score is gained by adding these items together, and the iADL score ranges from 0 to 85. We gained this iADL score data from each elderly people with help of one nurse every month.

We also drop some of participants' data because of the different room type, missing of parts of sensor data or missing of iADL score data. Our final number of samples used for experiment is 271 in total.

B. Evaluation Protocol and Metric

As we described in our proposed scheme, we will use a 10-fold double cross-validation to train and test proposed scheme as well as Lasso. In double cross-validation, the whole dataset is divided into training part, validation part and testing part. We use training data to train the model, use validation part to optimize the hyper-parameter and at last use testing part to give testing result. This double cross-validation could give mean value of evaluation metric as well as variance value of evaluation metric to indicate the stableness of algorithm with different dataset division.

We choose to use mean absolute error (MAE) of result score to represent the prediction performance as our metric. The calculation of MAE could be described as the following:

$$MAE = \frac{1}{N_{\mathrm{tr}}} \sum_{i=1}^{N_{\mathrm{tr}}} |y_i - \hat{y}_i|, \tag{5}$$

in which, N_{tr} is the size of the testing dataset, y_i is the ground truth of iADL score and $\hat{y}_i$ is the prediction result of the prediction model.

We also choose to use Pearson's r value to show the correlation coefficients between prediction score and ground truth score.

C. Experiment Result

*1) **prediction perfromance result**:* In this part, we choose to use the model trained by double cross-validation. As for proposed scheme's training, we try the sparsity from around 2 to 70 and optimize the sparsity with the least MAE as our model. For Lasso's training, we also use cross-validation method to select the optimized lambda in Lasso from 0.01 to 2.0, which corresponding to the sparsity roughly from 10 to 100.

The result of prediction performance is shown in Fig. 2. Mean value of MAE as well as variance of MAE during double cross-validation is 3.28 and 0.13 for proposed scheme; as well as 3.83 and 0.20 for Lasso. This means that our proposed scheme give a better prediction performance according to mean value of MAE. As for the variance of MAE, it shows our scheme with lower variance is more stable with different dataset division. Both values show our proposed scheme works better than Lasso.

Except for the MAE, we also show the correlation figure of ground truth iADL score value and prediction value in Fig. 3 and Fig. 4. We also calculate the correlation coefficient between ground truth and prediction result with Pearson's r value. The r value shown in Table VI-C1 of proposed scheme is 0.89, and Lasso is 0.84, which also means that proposed scheme prediction score is more near to ground truth.

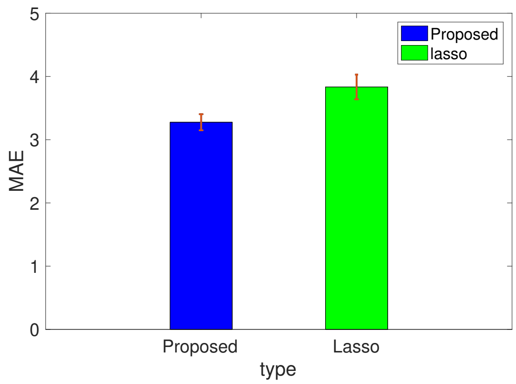

Fig. 2. Mean and variance of MAE in double CV

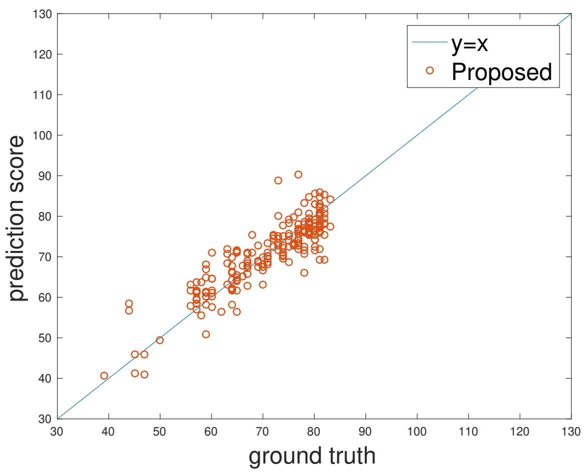

Fig. 3. Proposed method score correlation

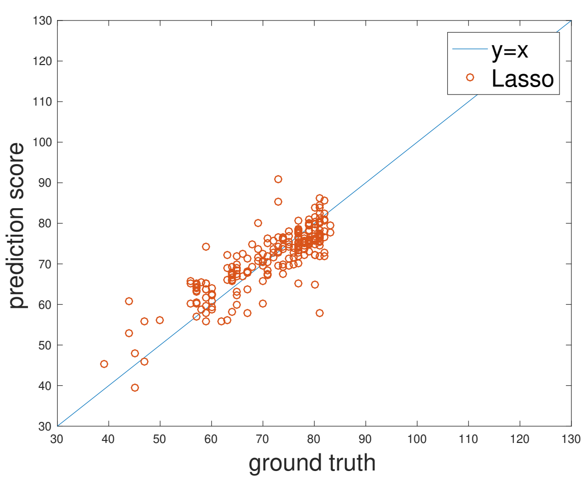

Fig. 4. Lasso score correlation

TABLE II
SCORE CORRELATION

Method	Proposed scheme	Lasso
r's value	0.89	0.84

*2) **sparsity and feature selection result**:* Regarding the selected feature number of the trained model, Lasso is 45 and FoBa is only 23. This result means that proposed scheme get a more sparse model than Lasso and could better deal with overfitting problem with lower sparsity.

At the same time, we check the prediction score MAE on a range of sparsity, which roughly distributed from 2 to 70 shown in Fig. 5. In this figure, we could clearly see that proposed scheme gained a better model, for proposed scheme gets a lower prediction MAE as well as lower sparsity. At the same time, proposed scheme also gives better performance at low sparsity situation than Lasso.

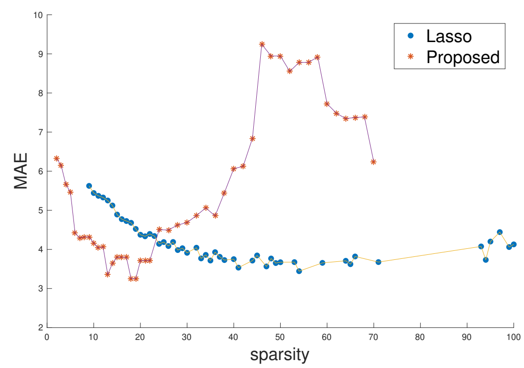

Fig. 5. MAE of Proposed scheme and Lasso with different sparsity

As for the selected feature by proposed scheme, we check the contents of features from the 23 features selected by FoBa algorithm. According to the algorithm, we believe that the several top selected features are most individually related with health score, and activities represented by these features are most related with elder people's activity health condition.

As a qualitative evaluation of selected features derived by our algorithm, "sleep time mean value" is the first selected feature, it indicates that sleeping condition may most related with elder people's health. Except for sleeping condition, other top selected features include "reaction rate at nighttime in bedroom", "variance of reaction mean value at morning in living room", "reaction rate at nighttime in kitchen" and etc. Some of these selected features also indicate that sleeping condition is related with their health, and others show that taking meal and going out activity is also influenced elderly people's health condition.

VII. CONCLUSION

In this paper, we propose a new approach to lead robust health score prediction from IoT sensors where the large number of statistics derived from sensors are generated while the assessment score is hard to be obtained due to the limited human resources.

We have to tackle the trade-off problem on designing rich featurization and overfitting issue with the rich featurization. As for overfitting issue, in this paper, we explore to use l_0 regularized training framework instead of using l_1 regularized training such as Lasso to avoid overfitting issue.

As for rich featurization towards a real application scenario of elderly people ADL score obtained from pyroelectric sensor data pervasively gathered over 1 year, we also provide a new feature design from pyroelectric sensor data where the number of feature statistics exceeds over 900 only from $3 \sim 4$ sensors.

Our empirical evaluation using the above real world dataset shows that our approach based on l_0 and rich feature description outperforms Lasso based method.

As a future work, we plan to promote our work into a more applicable position. By designing new training model and regression algorithm, we hope to build a more safety health score prediction application used in real world.

References

[1] Y. Huang, H. Xiong, K. Leach, Y. Zhang, P. Chow, K. Fua, B. A. Teachman, and L. E. Barnes. Assessing social anxiety using gps trajectories and point-of-interest data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 898–903. ACM, 2016.

[2] A. Kono, I. Kai, C. Sakato, and L. Z. Rubenstein. Frequency of going outdoors predicts long-range functional change among ambulatory frail elders living at home. *Archives of gerontology and geriatrics*, 45(3):233–242, 2007.

[3] J. Liu, J. Ye, and R. Fujimaki. Forward-backward greedy algorithms for general convex smooth functions over a cardinality constraint. In *International Conference on Machine Learning*, pages 503–511, 2014.

[4] K. Okumiya, K. Matsubayashi, T. Nakamura, M. Fujisawa, Y. Osaki, Y. Doi, and T. Ozawa. The timed "up & go" test and manual button score are useful predictors of functional decline in basic and instrumental adl in community-dwelling older people. *Journal of the American Geriatrics Society*, 47(4):497–498, 1999.

[5] W. H. Organization. Glossary of terms used in the "health for all" series, no. 1-8. 1984.

[6] S. Robben, M. Pol, and B. Kröse. Longitudinal ambient sensor monitoring for functional health assessments: a case study. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1209–1216. ACM, 2014.

[7] M. Shimosaka, S. Masuda, K. Takeichi, R. Fukui, and T. Sato. Health score prediction using low-invasive sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1044–1048, New York, NY, USA, 2012. ACM.

[8] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[9] R. Wang, M. S. Aung, S. Abdullah, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, M. Merrill, and E. A. Scherer. Crosscheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 886–897. ACM, 2016.

[10] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 3–14. ACM, 2014.

[11] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell. Smartgpa: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 295–306. ACM, 2015.

[12] R. Wang, W. Wang, M. S. Aung, D. Ben-Zeev, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, and E. A. Scherer. Predicting symptom trajectories of schizophrenia using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):110, 2017.

[13] X. Yuan, P. Li, and T. Zhang. Gradient hard thresholding pursuit for sparsity-constrained optimization. In *International Conference on Machine Learning*, pages 127–135, 2014.

[14] T. Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE transactions on information theory*, 57(7):4689–4708, 2011.