
A STOCHASTIC OPTIMIZATION METHOD FOR LOWER ONE-SIDE REGRESSION

We show the algorithm for the lower one-side regression in Algorithm 2.

B CASE STUDY ON REAL HEALTHCARE APPLICATION

We demonstrate the practical utility of our approach in a real healthcare application. From non-intrusive bed sensors that are installed under each of the four legs of a bed, we estimated the motion intensity of a participant that was measured accurately with an intrusive sensor wrapped around the wrist (ActiGraph) Tryon (2013); Mullaney et al. (1980); Webster et al. (1982); Cole et al. (1992). The ActiGraph can sense motion only on the forearm, which causes data coverage to be insufficient and observations of movements on other body parts to be missing frequently. The bed sensors had broader data coverage since it can sense global motion on all body parts; however, the sensing accuracy is limited because of its non-intrusiveness.

The dataset included three pieces of data, Data (i), (ii), and (iii), which were recorded over 20, 18, and 18.5 minutes, respectively. Each piece of data consisted of pairs of bed-sensor-data sequences and the corresponding motion intensity sequence obtained by the ActiGraph. We used the “magnitude” attribute of the ActiGraph as training labels \mathbf{y} for the motion intensity, whose sampling rate was about one sample per second. For true labels $\tilde{\mathbf{y}}$, we manually measured the motion intensity each minute under the management of an expert. For \mathbf{X} , we first computed the gravity center of four sensor outputs that were obtained from the bed sensors under the four legs of a bed. Then, we computed the time derivatives and cross terms of the raw sensor outputs and the gravity center. The sampling rate of the bed sensors was different from that of ActiGraph and was about one sample per five milliseconds. Thus, \mathbf{X} was finally generated as a sliding window of statistics in 1,000 millisecond (1 second) subsequences of the time series of the above computed variables, where 1 second was the same as the sampling interval of the ActiGraph. The statistics were means, standard deviations, and $\{0.05, 0.25, 0.5, 0.75, 0.95\}$ quantiles. In this case study, we used the linear model $\theta^\top \mathbf{x}$ for $f(\mathbf{x})$ because of its interpretability, which is not negligible in healthcare and medical applications. We used an implementation of Eq. (13) with squared loss for the first term, absolute loss, which satisfies Eq. (6), for the second and third terms, and L1-regularization for the regularization term.

Figure 3 compares our estimation results for the motion intensity with the output of the ActiGraph and true labels. We evaluated the results in 3-fold cross-validation by using data (i), (ii), and (iii). We can state that, in a totally non-intrusive manner, the proposed method could capture 89 percent of the total time period of the motions that were captured by the ActiGraph while capturing other 15 motions that were not captured by the ActiGraph. There were 6 missed motions that were captured by the ActiGraph and 1 false detection due to floor vibration. The baseline method based on ordinary MSE was severely underfitted, and most of the weights were zero; thus, we omitted the results. These results showed that the ActiGraph could be replaced with the bed sensor, and we could also use the bed sensor for the inputs of functions of the ActiGraph, such as sleep-wake discrimination Cole et al. (1992).

The important features selected by L1-regularization were the statistics of the gravity center and the cross terms and time derivatives of the raw sensor outputs. The largest weight was assigned to the standard deviation of the gravity center, which represents the amplitude of the gravity center, and, thus, it is directly related to the motion of participants.

REFERENCES

- Roger J Cole, Daniel F Kripke, William Gruen, Daniel J Mullaney, and J Christian Gillin. Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5):461–469, 1992.
- DJ Mullaney, DF Kripke, and S Messin. Wrist-actigraphic estimation of sleep time. *Sleep*, 3(1): 83–92, 1980.
- Warren W Tryon. *Activity measurement in psychology and medicine*. Springer Science & Business Media, 2013.
- John B Webster, Daniel F Kripke, Sam Messin, Daniel J Mullaney, and Grant Wyborney. An activity-based sleep monitor system for ambulatory use. *Sleep*, 5(4):389–399, 1982.

Algorithm 2 Lower one-side regression based on stochastic gradient method

Input: Training data $\mathcal{D} = \{\mathbf{x}_n, y_n\}_{n=1}^N$ and hyperparameters $\rho, \lambda \geq 0$

Output: Model parameter θ for f

- 1: Let \mathcal{A} be an external stochastic gradient method and G_m be a gradient for the m -th mini-batch
 - 2: **while** No stopping criterion has been met
 - 3: Shuffle \mathcal{D} into M mini-batches, and denote by $\{\mathbf{X}^{\{m\}}, \mathbf{y}^{\{m\}}\}$ the m -th mini-batch whose size is N_m
 - 4: **for** $m = 1$ **to** M
 - 5: $G_m \leftarrow 0$
 - 6: **for** $n = 1$ **to** N_m
 - 7: **if** $y_n^{\{m\}} - f(\mathbf{x}_n^{\{m\}}) < 0$ **then**
 - 8: $G_m \leftarrow G_m + \frac{\partial L(f(\mathbf{x}_n^{\{m\}}), y_n^{\{m\}})}{\partial \theta} - g(-1, f(\mathbf{x}_n^{\{m\}})) + \lambda \frac{\partial R(f)}{\partial \theta}$
 - 9: **else**
 - 10: $G_m \leftarrow G_m + \rho g(-1, f(\mathbf{x}_n^{\{m\}})) + \lambda \frac{\partial R(f)}{\partial \theta}$
 - 11: Update θ by \mathcal{A} with G_m
-

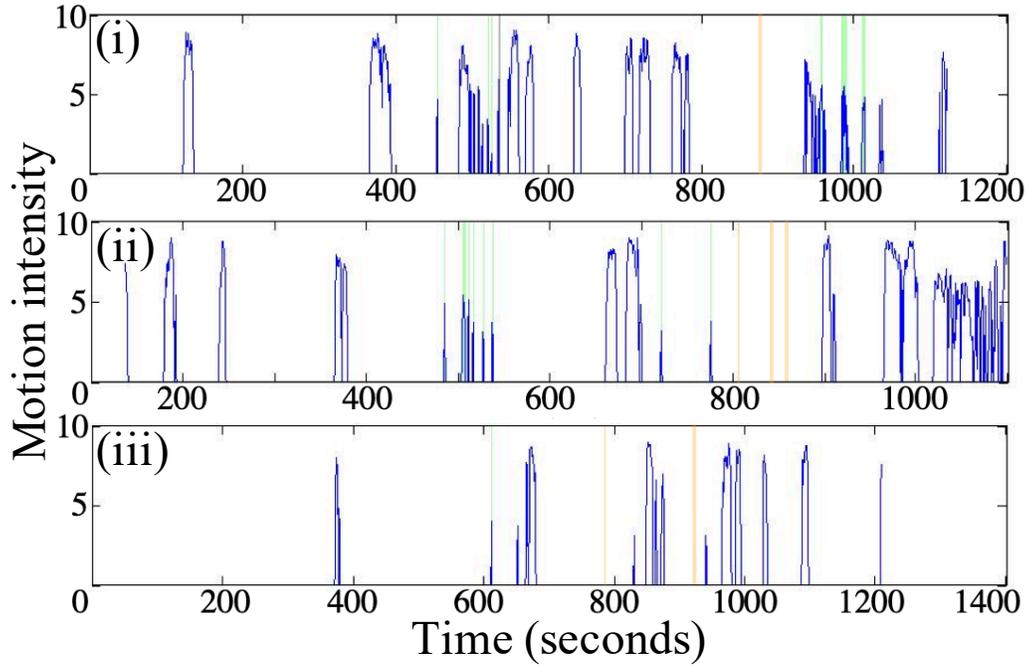


Figure 3: Case study on real healthcare application. Blue line represents our estimation results for motion intensity. White (non-colored) area shows that both the proposed method and the ActiGraph correctly estimated motion intensity of participant at this duration. Green area shows that our method could capture motion at this duration but ActiGraph could not. Orange area shows that our method could not capture motion at this duration but ActiGraph could. Gray area shows that our method mistakenly captured noise as participant's motion.