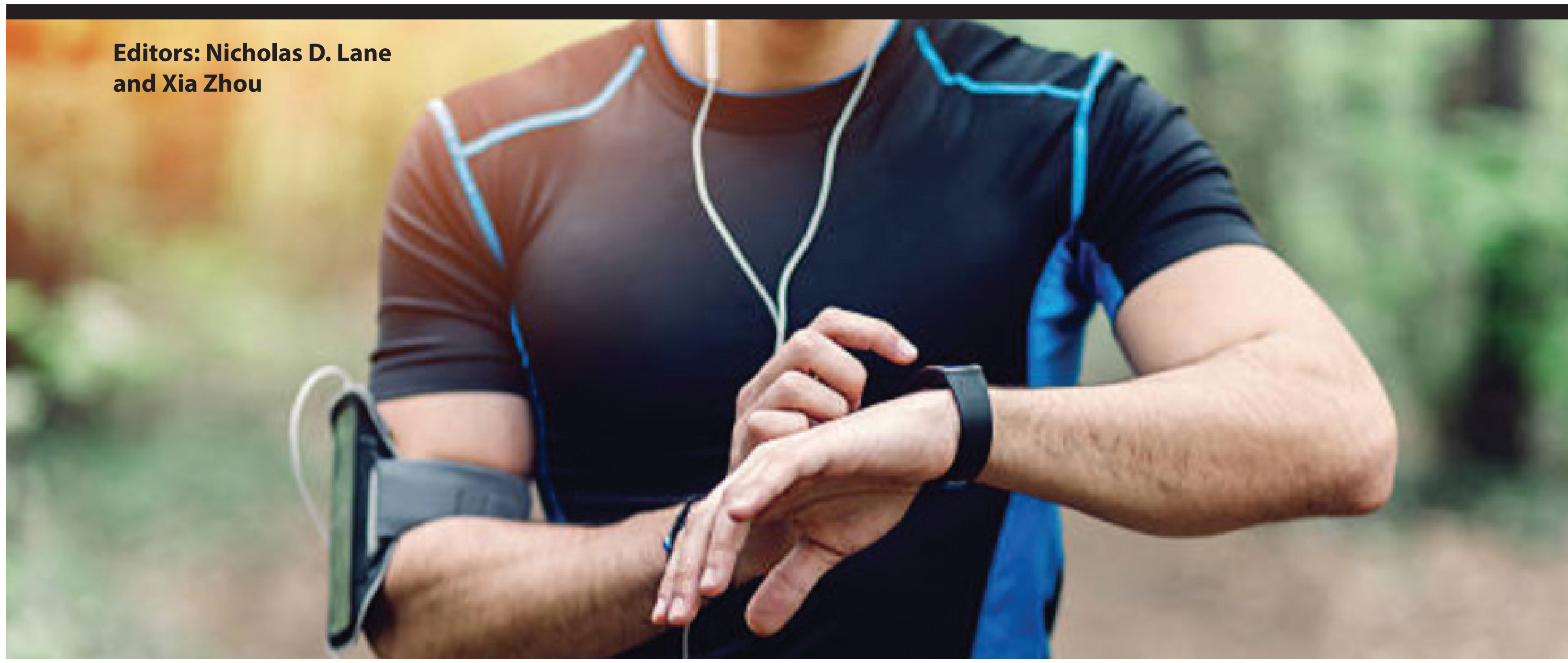
Huatao Xu, Pengfei Zhou, Rui Tan and Mo Li *School of Computer Science and Engineering, Nanyang Technological University, Singapore*
Guobin Shen *Alibaba Local Services Lab, Alibaba Group, China*



Editors: Nicholas D. Lane and Xia Zhou

Photo, istockphoto.com

LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications

Excerpted from "LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications," from SenSys '21: *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems* https://dl.acm.org/doi/10.1145/3485730.3485937 ©ACM 2021

Deep learning greatly empowers Inertial Measurement Unit (IMU) sensors for a wide range of sensing applications. Most existing works require substantial amounts of well-curated labeled data to train IMU-based sensing models, which incurs high annotation and training costs. Compared with labeled data, unlabeled IMU data are abundant and easily accessible. This article presents a novel representation learning model that can make use of unlabeled IMU data and extract generalized rather than task-specific features. With the representations learned via our model, task-specific models trained with limited labeled samples can achieve superior performances in typical IMU sensing applications, such as Human Activity Recognition (HAR).

CONVENTIONAL IMU SENSING

Wearable devices have played a critical role in a wide range of applications, including human activity recognition [1][2], human-computer interaction [3], localization and tracking [4], etc. Many of them rely heavily on data from Inertial Measurement Unit (IMU) sensors (i.e., accelerometer, gyroscope, and magnetometer), which are widely used in personal mobile devices, such as smartphones and smartwatches.

Due to the rapid development of deep learning, many works adopt deep neural networks to process IMU data [1][2]. Compared with manual feature engineering, deep learning algorithms can extract more effective features and gain significant performance improvements in inference. Most existing works [2][5], however, rely heavily on supervised learning processes, in which substantial amounts of labeled IMU data are required to train sensing models. The requirement of large, labeled data hinders their adoption in practice for two reasons. First, labeled IMU data are scarce because it is costly and time-consuming to collect sufficient labeled IMU samples in real-world settings. Second, the diversity in mobile devices, usage patterns, and environments results in the need for labeled data with various combinations of phone models, users, and usage scenarios to attain generalizable models.

CAN UNLABELED IMU DATA HELP?

Compared with labeled data, unlabeled IMU data are abundant and easily accessible. In particular, unlabeled data can be easily collected from a variety of wearable devices, usage patterns, and scenarios. Therefore, to address the challenge of labeled data scarcity, we propose *LIMU-BERT*, which leverages massive unlabeled data and accordingly extracts general features through the self-supervised training technique. After the representations are learned, multiple task-specific inference models can thus be trained with a small amount of labeled IMU samples. The key rationale is to learn the generalizable representations from the abundant unlabeled IMU data instead of scarce labeled data.

What general features are desired? After scrutinizing the characteristics of IMU data, we focus on two types of features: distributions of individual measurements of IMU sensors, and temporal relations in continuous measurements. The correlation of IMU readings on three axes gives information about the attitude of the device, which is an important feature of the usage patterns. For example, the readings of the accelerometer and gyroscope become dramatically larger if the user transitions to walking from standing still. The three-axis components of the accelerometer correlate differently when the device orientation varies. As a typical time-series data, temporal relations within sensory data give further information about user behaviors.

How to extract those general features? Inspired by the emerging self-supervised techniques in natural language processing, we borrow the key framework of BERT [6] to process unlabeled IMU data and accordingly extract generalizable features. BERT designs two novel self-supervised training methods to learn the bidirectional language representations from the unlabeled text. However, intended for natural language data processing, the original BERT lacks methodology in processing IMU data, e.g., the multi-modality problem of various IMU sensor readings. We thus devise a variety of techniques including data fusion and normalization, effective training method, and structure optimization, and embeds them into the BERT framework for improved efficacy and efficiency in IMU sensing applications.

LIMU-BERT DESIGN

The overview of our framework is presented in Figure 1, which consists of self-supervised and supervised learning phases. There are three major components in our framework, including LIMU-BERT, *decoder*, and *classifier.* The LIMU-BERT takes the unlabeled IMU data as input and outputs high-level representations or features. The decoder reconstructs the unlabeled data based on the learned features. The classifier trained with a small number of labeled representations aims to accomplish a task-specific application, such as HAR.

1. Self-supervised learning. In this phase, we mask partial readings of unlabeled samples and feed them into LIMU-BERT. The LIMU-BERT and the decoder jointly predict the original values of masked readings by learning the temporal relations among IMU data. The objective of the self-supervised learning process is to fully utilize a large amount of unlabeled data and accordingly extract general features.

2. Supervised learning. Next, we transfer the LIMU-BERT model and connect it with a classifier. In this phase, all parameters of the LIMU-BERT are frozen and only the classifier is trained with limited labeled representations that have been processed by the LIMU-BERT. At the run time after the supervised learning, the LIMU-BERT and classifier are deployed together to estimate the task-specific results for IMU sensor data.

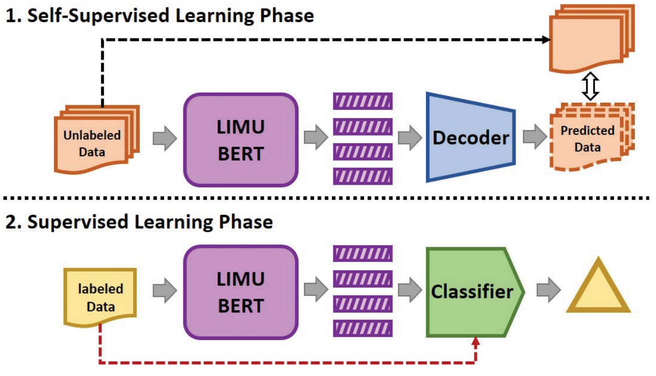


FIGURE 1. Framework overview.

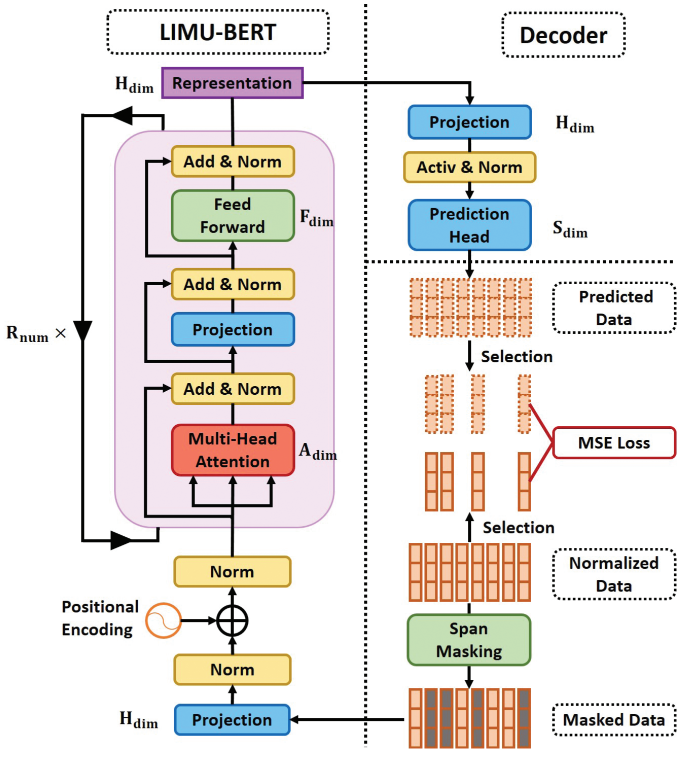


FIGURE 2. Self-supervised training workflow.

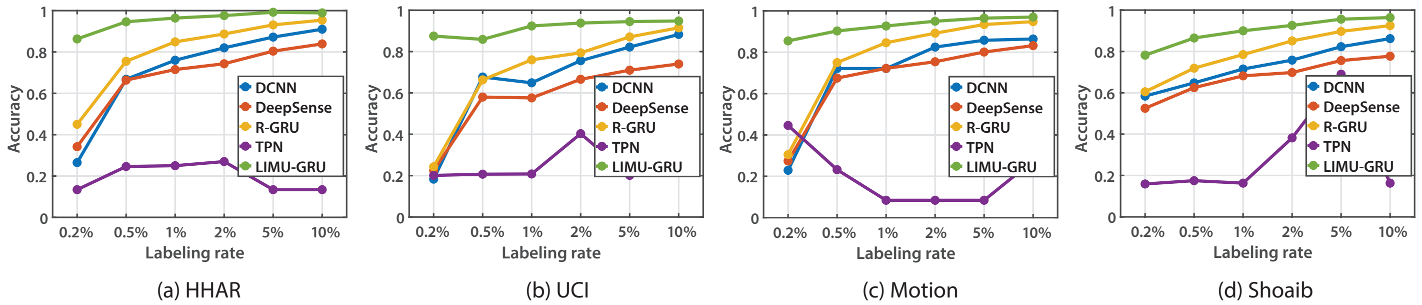


FIGURE 3. Accuracy comparison on HAR at different labeling rates.

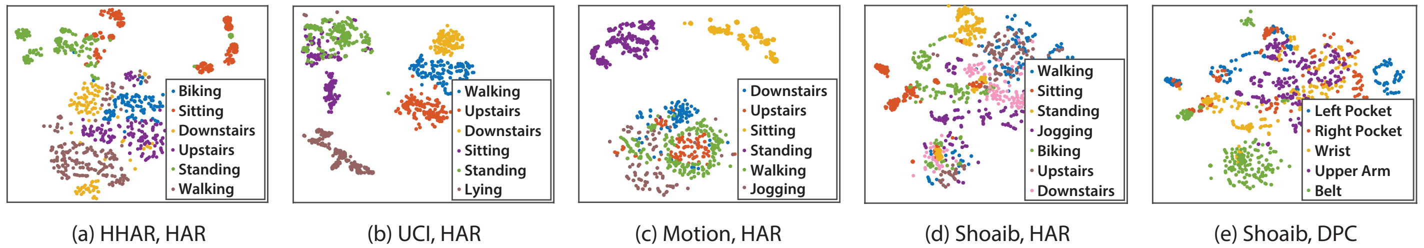


FIGURE 4. Representation visualization.

Fusion and Normalization

IMU sensors have different distributions, and such differences would affect the model performance based on our experiments. Therefore, the sensor readings need to be properly normalized. Common normalization methods, e.g., min-max or mean-variance normalization, and loss distribution information can negatively affect the quality of the general representations. To this end, we design a simple but effective normalization method on IMU readings to narrow the range differences and not severely alter their distributions. We divide accelerometer readings by gravity constant and keep the distribution of gyroscope readings.

A critical characteristic of IMU sensors is that the number of features is small (e.g., six). To extend the dimension of features and fuse IMU sensors, we project the normalized sensor data into a higher space by multiplying input data with a matrix with a high dimension. Such projection is implemented by a linear layer. Next, LIMU-BERT leverages Layer Normalization [7] to normalize the fused features corresponding and dynamically normalize implicit features.

Learning Representations

Both distributions of individual readings and temporal relations among continuous readings are important features. After analyzing the characteristics of IMU data in BERT, we find the Mask Language Model (MLM) task is the desired training approach, which randomly masks subsequences of the input readings, and the model is trained to predict the original readings.

Since the nature of human mobility leads to similar IMU sensor data across adjacent readings, the model may easily degrade to reconstruct the masked readings by mirroring neighboring readings if only the one-sample subsequences are masked. Therefore, we adopt a Span Masking mechanism [8] to mask longer subsequences, whose lengths are sampled from a geometric distribution.

Lightweight Model

Different from BERT, LIMU-BERT must be lightweight enough to run on mobile devices. Thus, we adopt a much smaller sampling rate (i.e., 20 Hz) compared with the existing works [1][2], and accordingly decrease the length of the input IMU sequences and reduce model size. The representations dimension of LIMU-BERT is also smaller than that of the original BERT (e.g., 1024).

LIMU-BERT adopts a cross-layer parameter sharing mechanism [9] to improve parameter efficiency, which reuses the parts of model parameters and reduces the number of total parameters significantly. We treat the MLM task as a regression task rather than a classification task because the IMU features are continuous variables. The regression model can avoid a heavy output layer and simplify the decoder considerably.

Training Workflow

Putting together all of the designs above, the detailed workflow of the self-supervised process is illustrated in Figure 2. The normalized data are masked before being fed into LIMU-BERT and the first projection and norm component together implement the sensor fusion and normalization design. All normalization components represent layer normalization. The *MultiAttn* is a self-attention layer [10] with multiple attention heads and the *Proj* represents a fully connected layer, and the *FeedForward* consists of two fully connected layers. The decoder reconstructs the original values of the masked IMU sequences with the representations generated by LIMU-BERT. Finally, the Mean Square Error (MSE) function is adopted to compute the recon-struction loss and train the models.

After the LIMU-BERT is trained with unlabeled data, it can be utilized to generate representations for labeled IMU data. Based on the learned representations and their corresponding labels, we can design task-specific models with supervised training. In our framework, we design a lightweight classifier with Gated Recurrent Unit (GRU).

RESULTS

We implement LIMU-BERT and GRU classifiers with PyTorch and quantitatively evaluate them with four widely used open IMU datasets, i.e., HAR [11], UCI [12], Motion [13], and Shoaib [14]. We randomly divide each dataset into training (80%), validation (10%), and test (10%) sets. The training set is further divided into 1% as labeled set and 99% as unlabeled set. The ratio of the number of samples in labeled set to that in training set is called the *labeling rate*. We choose two typical IMU sensing applications: Human Activity Recognition (HAR) and Device Placement Classification (DPC). The models are trained to recognize human activities and the on-body placements with IMU data. As both two applications are classification tasks, we adopt accuracy and macro F-score for performance comparison.

Figure 3 depicts the comparative performances of our model (LIMU-GRU) and other baseline models in HAR application. The results show that LIMU-GRU consistently outperforms the baseline in all cases. The performance gaps between LIMU-GRU and other models are higher when the labeling rate is smaller. And LIMU-GRU also achieves outstanding performance in the DPC application[1]. In summary, the performance gain of LIMU-GRU is significant, thanks to the effective and generalizable features extracted by LIMU-BERT.

To understand the effectiveness of the representations learned by LIMU-BERT, we visualize the learned high-dimensional representations in 2D space. The clusters show high correlations among the learned representations in all datasets. It is obvious to see that samples belonging to the same activity class exhibit a high clustering effect, which is highly beneficial for the downstream classification models.

MOVING FORWARD

Further experiments show that the performance of LIMU-BERT slightly degrades when transferring across datasets. One main reason is that the four datasets are collected with diverse devices, placements, users, and environments. The diversities cause the domain shifts among the datasets and affect the generalizability of learned representations. To mitigate the impact of domain shifts and extract more general features, LIMU-BERT might be further improved by techniques like data augmentation [15]. Other future works include the investigation of how the representations learned by LIMU-BERT may facilitate other mobile applications, e.g., indoor localization or device orientation estimation.

SUMMARY

In this paper, we present a lite BERT-like representation learning model for mobile IMU sensor data, which makes use of unlabeled data and accordingly extracts generalizable features instead of task-specific features. Extensive experimental evaluation demonstrates that the learned representations by LIMU-BERT can boost the performances of downstream models significantly with few labeled samples. With LIMU-BERT, the labeling efforts in real IMU-based sensing applications can be greatly reduced. ■

Huatao Xu is a Ph.D. student in the School of Computer Science and Engineering at Nanyang Technological University, Singapore. His research interests lie in mobile and wireless sensing. Specifically, he is focused on developing models for human activity recognition and indoor tracking with ubiquitous mobile devices.

Pengfei Zhou is a research scientist of Advanced Digital Sciences Center (ADSC), Illinois at Singapore. His research interests include mobile sensing and systems, artificial intelligence of things (AIoT), and autonomous cyber physical systems. He received his Ph.D. from the School of Computer Science and Engineering from Nanyang Technological University, Singapore.

Rui Tan is an associate professor in the School of Computer Science and Engineering at Nanyang Technological University, Singapore. His research interests include cyber-physical systems, sensor networks, and pervasive computing systems. He received his Ph.D. in Computer Science from City University of Hong Kong.

Mo Li is a professor in the School of Computer Science and Engineering at Nanyang Technological University, Singapore. His research interests include networked and distributed sensing, wireless and mobile, Internet of Things (IoT), smart city and urban computing.

Guobin (Jacky) Shen is the director of the AIoT & Innovation Lab of Alibaba Local Services Company, China. His research interests are in building large-scale IoT empowered systems and applications to solve real-world challenges and enable new user experiences. He received his Ph.D. in Electrical and Electronic Engineering from the Hong Kong University of Science and Technology.

REFERENCES

[1] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. Multi-task self-supervised learning for human activity detection. 2019. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3.2: 1-30.

[2] Shuochao Yao, et al. 2017. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. *Proceedings of the 26th international conference on world wide web.*

[3] Liu Yang, et al. Real-time arm skeleton tracking and gesture inference tolerant to missing wearable sensors. 2019. *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services.*

[4] Yonghang Jiang, Zhenjiang Li, and Jianping Wang. 2018. Ptrack: Enhancing the applicability of pedestrian tracking with wearables. *IEEE Transactions on Mobile Computing*, 18.2: 431-443.

[5] Wenchao Jiang, and Zhaozheng Yin. 2015. Human activity recognition using wearable sensors by deep convolutional neural networks. *Proceedings of the 23rd ACM International Conference on Multimedia.*

[6] Devlin, Jacob, et al. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. a*rXiv preprint*. arXiv:1810.04805.

[7] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. 2016. Layer normalization. arXiv preprint, arXiv:1607.06450.

[8] Mandar Joshi, et al. 2020. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8: 64-77.

[9] Zhenzhong Lan, et al. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint* arXiv:1909.11942.

[10] Ashish Vaswani, et al. 2017. Attention is all you need. Advances in *Neural Information Processing Systems*, 30.

[11] Allan Stisen, et al. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. 2015. *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems.*

[12] Jorge-L. Reyes-Ortiz, et al. 2016. Transition-aware human activity recognition using smartphones. *Neurocomputing*, 171: 754-767.

[13] Mohammad Malekzadeh, et al. 2019. Mobile sensor data anonymization. *Proceedings of the International Conference on Internet of Things Design and Implementation.*

[14] Muhammad Shoaib, et al. 2014. Fusion of smartphone motion sensors for physical activity recognition. *Sensors,* 14.6: 10146-10176.

[15] Youngjae Chang, et al. 2020. A systematic study of unsupervised domain adaptation for robust human-activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4.1: 1-30.

[1] Please check our original paper for more details.