LabelMerger: Learning Activities in Uncontrolled Environments

Seyed Iman Mirzadeh
Washington State University
Pullman, WA, USA
seyediman.mirzadeh@wsu.edu

Jessica C. Ardo
University of California Irvine
Irvine, CA, USA
jardo@uci.edu

Ramin Fallahzadeh
Stanford University
Stanford, CA, USA
raminf@stanford.edu

Bryan Minor
Washington State University
Pullman, WA, USA
bminor@wsu.edu

Lorraine Evangelista
University of Texas Medical Branch
Galveston, TX, USA
lsevange@utmb.edu

Diane Cook
Washington State University
Pullman, WA, USA
djcook@wsu.edu

Hassan Ghasemzadeh
Washington State University
Pullman, WA, USA
hassan.ghasemzadeh@wsu.edu

***Abstract*—While inferring human activities from sensors embedded in mobile devices using machine learning algorithms has been studied, current research relies primarily on sensor data that are collected in controlled settings often with healthy individuals. Currently, there exists a gap in research about how to design activity recognition models based on sensor data collected with chronically-ill individuals and in free-living environments. In this paper, we focus on a situation where free-living activity data are collected continuously, activity vocabulary (i.e., class labels) are not known as *a priori*, and sensor data are annotated by end-users through an active learning process. By analyzing sensor data collected in a clinical study involving patients with cardiovascular disease, we demonstrate significant challenges that arise while inferring physical activities in uncontrolled environments. In particular, we observe that activity labels that are distinct in syntax can refer to semantically-identical behaviors, resulting in a sparse label space. To construct a meaningful label space, we propose LabelMerger, a framework for restructuring the label space created through active learning in uncontrolled environments in preparation for training activity recognition models. LabelMerger combines semantic meaning of activity labels with physical attributes of the activities (i.e., domain knowledge) to generate a flexible and meaningful representation of the labels. Specifically, our approach merges labels using both word embedding techniques from the natural language processing domain and activity intensity from the physical activity research. We show that the new representation of the sensor data obtained by LabelMerger results in more accurate activity recognition models compared to the case where original label space is used to learn recognition models.**

***Index Terms*—Machine learning, mobile health, activity recognition, word embedding.**

I. Introduction

Activity recognition is a an active research area with the aim of automatically detecting physical activities performed by people in their daily living situations. The recognition of physical activities has become a task of significant interest within the field, in particular for medical and health-related applications such as in behavioral medicine. An application of activity recognition in behavioral medicine is to design interventions for individuals with, or at risk for, diabetes, obesity, or heart disease where the the individuals are often required to follow a well-defined exercise regimen as part of their treatment [1].

For activity recognition models to be reliable, it is critical to collect labeled sensor data in end-user settings. The process involves utilizing an active learning approach where end-users provide annotations/labels of the sensor data through a user-interface on their mobile device. However, labels provided by end-users in uncontrolled environments introduce unique challenges for learning reliable activity recognition models. Here we categorize those challenges into three broad groups:

- *Spatial disparity:* we recognize that different individuals can have different activity behaviors. When sensor data are labeled by end-users, the constructed activity vocabulary formed for one user can be different than that of another user. This inter-user (i.e., spatial) label disparity results in activity recognition models that cannot be used across different users. As a result, we need to construct an activity vocabulary for each user or aggregate labels gathered from a large group of users to account for cross-user behavior differences.
- *Temporal disparity:* because we do not place any restrictions on the data collection and sensor annotation processes, users are not limited to expressing their activities according to a set of pre-defined labels. Therefore, a user can express the same activity differently at different times. This intra-user (i.e., temporal) disparity results in labels that are different in syntax but identical in semantic.
- *Burden on user:* we recognize that the process of data labeling is a burden on the user, in particular when the system in adopted by patients with chronic conditions. Therefore, it is important to develop activity recognition models using a small number of training instances labeled by users.

To deal with the challenges of label disparity, LabelMerger aims to restructure the label space of each user, or a group of users, by grouping labels that are semantically similar

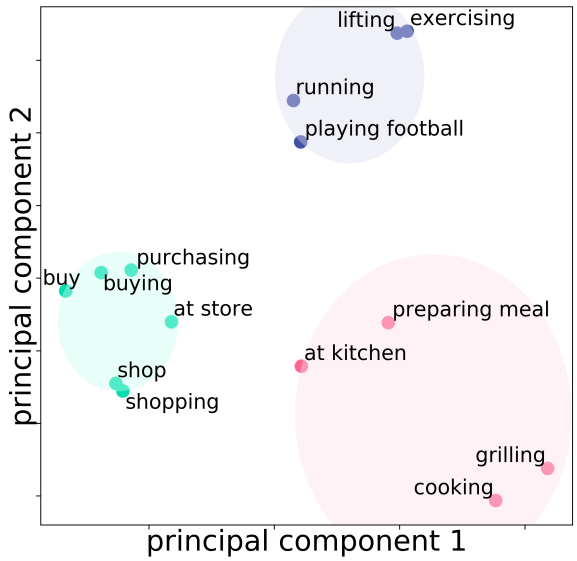


Fig. 1: An example of restructured label space in LabelMerger.

and are associated with activities of similar intensities. An example of such restructured label space in shown in Fig. 1 where 14 labels expressed by users are aggregated into three groups, shown in *green*, *blue*, and *red* in the new label space. The labels shown in this figure represent a subset of labels expressed by participants in our clinical study. For visualization, here a dimensionality reduction technique (e.g., PCA [2], t-SNE [3]) is used to illustrate the clusters in a 2D coordinate. Because users use different expressions to describe their activity behavior, there exist a substantial amount of disparity in the data. As shown in Fig. 1, users use words such as 'shop', 'buy', 'purchasing', 'at store', 'shopping', and 'buying' to express a particular activity behavior. Such label disparities not only occur across users but also exist within the same user at different times. Not addressing the problem of label disparity (i.e., treating each discrete label expressed by the user as a class label in the process of machine learning algorithm training) will result in an unnecessary increase in the number of classes and a decrease in the number of training instances within each class. This in turn will result in learning an activity recognition model that performs poorly because of the low quality training data.

II. Label Merger

A. Problem Statement

Let $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$ be the data collected through the process of active learning where x_i represents i-th input sensor data instance and y_i represents the activity label associated with x_i. The labels y_i are drawn from the set $L_{user} = \{a_1, a_2, ..., a_n\}$ of n discrete activity labels expressed by the user. Our goal is to construct a compact and meaningful label space $L_{merge} = \{l_1, l_2, ..., l_k\}$ with $k \leq n$ classes.

Having defined our input and desired output, we are interested in finding a mapping function $\Phi : \mathcal{R}^n \rightarrow \mathcal{R}^k$ that automatically transforms noisy labels in L_{user} into k groups, each consisting of similar activity labels. Therefore, by applying our mapping function Φ on the input labels L_{user}, we will obtain k different groups of labels.

Since our machine learning task is activity recognition, a reasonable objective is to ensure that activities that reside in the same group in our final label space represent similar physical activities. This problem is naturally a clustering problem; however, we need to define appropriate features that quantify similarity/dissimilarity among various activity labels expressed by the user.

B. Feature Design for Label Space

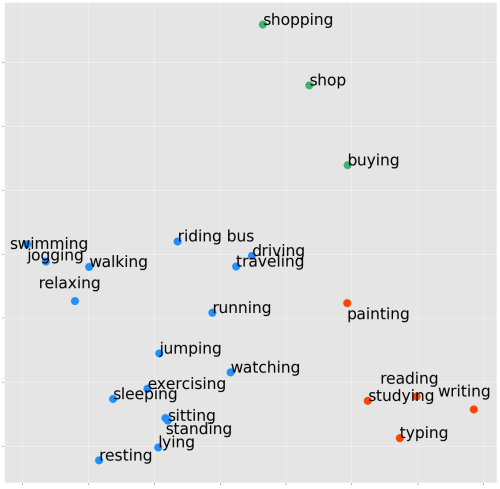
We propose to extract two broad sets of attributes in label space. The first set captures semantic meanings of the labels using word embedding while the second set incorporates physical attributes of human activities. Our feature vector uses word vectors to obtain meaning of each label as well as domain-specific measures such as metabolic equivalent of task (MET) value associated with each activity. The use of semantic meanings is motivated by spatial and temporal disparities among labels acquired by different users or/and at different time frames.

To construct the label space feature vector, instead of using atomic symbols to represent each word, we use their vector representations, which is a common approach to overcome limitations of using atomic symbols. This approach utilizes a window-based method where we count the number of times that each word appears within a window of a particular size centered around the word of interest. To this end, we use the GloVe algorithm [4] and its available pre-trained vectors to convert words to vectors.

However, as depicted in Figure Fig. 2(a), the GloVe algorithm, takes only the meaning of the labels into account and is not concerned about physical meaning/attributes of each activity. For example, it can be observed that 'swimming' and 'watching' (or 'swimming' and 'relaxing') belong to the same group while they are very different in terms of physical attributes, activity intensity, and their impact on physical health.

To address the limitation of using only semantic meaning when defining features in the label space, we propose to utilize a general form of 'domain knowledge' features which can be application-dependent. For example, when designing interventions for physical health, one may consider activity intensity as a measure of physical fitness and well-being. In contrast, activities such as 'reading', 'swimming', and 'watching' may need to be placed in the same group in the label space for such health interventions.

To incorporated the domain knowledge, we use a well-known measure of human physical activities, namely MET (metabolic equivalent of task), as the sole feature used in our domain-knowledge feature portion of the feature vector computed in the label space. One motivation behind choosing MET is that there is already calculated values for nearly all common activities by Taylor Compendium of Physical Activity [5]. However, our methodology presented in this research is



(a) using only semantics

(b) using semantics and domain knowledge

Fig. 2: Importance of combined semantic and domain-knowledge features for label merging: clusters formed using only semantic features (a) and clusters obtained with combined semantic meaning and MET values (b).

not limited to only MET values and one can use any valid representation of human activities for inclusion in the feature vector.

C. Algorithm

Here, we introduce a formal procedure to transform noisy labels L_{user} expressed by the user, to a target label set L_{merge} in the new label space. For each activity label a_i in L_{user}, we perform the following tasks:

1) We obtain the equivalent word embedding of the activity labels in L_{user}.
2) Because we might not have the MET value of the activity label in our MET database (e.g., there is no pre-defined MET value for 'at Walmart store'), we find semantically closest activity in the database and use its MET value during computation of the feature vector. In this study, we use cosine distance as a measure of similarity for two word vectors. Note that if we have the exact same activity in the MET database, the closest word will be the the given label itself.
3) We add the MET value of the label to our feature vector. However, we use the factor λ to control the importance of domain-knowledge with respect to semantic meaning (i.e., word embeddings). A higher value of λ translates into a higher weight assigned to domain knowledge (e.g., physical activity information) factor while constructing a clustering of the labels.

Algorithm 1 LabelMerger Algorithm

Input: Noisy label set L_{user}, number of clusters K, λ domain knowledge coefficient, word vectors W, and MET values M
Output: clustering labels for L_{merge}
initialize feature vectors F as an empty matrix.
for each label a_i in L_{user} **do**
  v_i = word vector of a_i
  w = $argmin$(cosine distance(l, v_i)) $\forall\ l \in M$
  m = normalized MET value of w
  assign f = concatenate $m \times \lambda$
  add f to feature vectors F
end for
do k-Means clustering on each row of F as a data point to get K clusters.
return clustering labels as L_{merge}

After constructing feature vectors from all the noisy labels in L_{user}, we use k-Means to obtain k clusters in the label space. Algorithm 1 shows the LabelMerger algorithm.

III. Experiments and Results

A. Data Collection

This study was reviewed and approved by the appropriate Institutional Review Boards. Participants were recruited from a single outpatient tertiary care clinic, as well as by word of mouth referrals. Participants were screened for study inclusion to ensure their eligibility. Each participant was trained about how to use smartphone device and respond to activity prompts. They were asked to charge the phone each night. The researchers sent an activity prompt to each participant as a test, and observed them demonstrate their ability to respond prior to beginning the data collection process. Participants were instructed to respond to as many prompts each day as possible, but to avoid responding or using the phone when driving or operating heavy machinery. They were also instructed how to add an activity to the list of activities in the Activity Learning application [6]. Each participant was asked to provide labels in response to activity prompts for two weeks. The activity learning application was programmed to issue an activity prompt on the smartphone every 2 hours between 8:00am and 8:00pm daily. We used the data of 13 participants who had completed data collection by the time of conducting this data analysis.

B. Learning Activity Recognition Model

For each acquired label, we assigned the label to a 5-second window of the signal segment. From each signal segment, we extracted various statistical features for gyroscope and accelerometer which has been shown effective in identifying daily living activities [7], [8]. This allowed us to form a training dataset. To learn an activity recognition model using this dataset, we split the data into 80% for training and 20% for testing. Different classifiers that were used for classification include 'Random Forest' [9], 'Support Vector Machine' [10], and 'K-Nearest Neighbors' [11] with $K = 1$ and $K = 3$.

C. Results

As shown in Table I, by increasing the number of clusters in label merging, which translates into an increased number of classes for activity recognition, the machine learning task becomes more difficult. The hardest problem is the baseline approach where we do not perform any label merging and lean an activity recognition model to classify activities according to the initial labels expressed by each participant.

We compared the performance of the baseline approach to that of scenarios where the number of clusters are less that the number of initial classes (due to label merging). For each participant, we calculated all of the following different scenarios and reported the best performance:

- Using different number of clusters (2, 3 and 4) in addition to the baseline.
- Using classifiers *Random Forest*, *Support Vector Machine*, *1-Nearest-Neighbor*, and *3-Nearest-Neighbors*
- Using different values from $\{1, 5, 10, 15, 20, 30, 40\}$ for λ as defined in Algorithm 1.

participant	baseline	2 clusters	3 clusters	4 clusters
1	0.48	0.77	0.71	0.55
2	0.67	1.0	1.0	0.67
3	0.56	0.89	0.78	0.56
4	0.33	0.95	0.81	0.52
5	0.6	1.0	1.0	0.8
6	0.4	0.9	0.9	0.8
7	1.0	1.0	1.0	1.0
8	0.72	0.94	0.83	0.89
9	0.2	0.9	0.7	0.7
10	0.57	0.86	0.71	0.71
11	0.5	1.0	1.0	1.0
12	0.17	0.72	0.72	0.56
13	0.2	1.0	0.8	0.6
average	0.49	0.91	0.84	0.72

TABLE I: Best activity recognition accuracy obtained with participant-specific data.

In Table II, we report classification accuracy, recall, precision and F1 score for the case where we aggregated data from individual participants into a large dataset. This problem is much harder than the per-participant learning since we will have many more different labels for a similar activity concept. We can see a 50% improvement in accuracy if we only aim to classify high-intensity versus low-intensity activities and 10% improvement if we only group two similar labels together and reach 16 different labels.

clusters	AC	RE	PR	F1
2	0.84	0.57	0.91	0.58
4	0.64	0.37	0.62	0.37
8	0.53	0.33	0.39	0.32
16	0.41	0.27	0.30	0.25
30 (baseline)	0.31	0.19	0.19	0.18

TABLE II: Performance with aggregated data from all users.

IV. Conclusion

We introduced several challenges that arise when deploying human activity recognition in real-world settings. In particular, we discussed that activity labels that are distinct in syntax can refer to semantically-identical behaviors when data collection occurs in uncontrolled environments. We proposed LabelMerger to restructure the label space by combining semantic meaning of activity labels with physical attributes of the activities to generate a flexible and meaningful representation of the labels. We showed that this approach is promising in improving activity recognition accuracy while maintaining a meaningful representation of the labels.

Acknowledgment

This work was supported in part by the National Institute of Health, under grant 1R21NR015410-01, and the National Science Foundation, under grant CNS-1750679. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

References

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, pp. 1192–1209, 2013.

[2] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, p. 441, 1933.

[3] L. van der Maaten and G. E. Hinton, "Visualizing data using t-sne," 2008.

[4] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *EMNLP*, 2014.

[5] B. E. Ainsworth, W. L. Haskell, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. Strath, W. A. O'Brien, D. Bassett, K. H. Schmitz, P. O. Emplaincourt, D. R. Jacobs, and A. S. Leon, "Compendium of physical activities: an update of activity codes and met intensities." *Medicine and science in sports and exercise*, vol. 32 9 Suppl, pp. S498–504, 2000.

[6] R. Fallahzadeh, B. D. Minor, L. S. Evangelista, D. J. Cook, and H. Ghasemzadeh, "Mobile sensing to improve medication adherence: demo abstract," in *IPSN*, 2017.

[7] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, pp. 156–167, 2006.

[8] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *BODYNETS*, 2011.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[11] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, vol. 13, pp. 21–27, 1967.