Personalized Activity Recognition using Partially Available Target Data

Ramin Fallahzadeh, *Member, IEEE,* Zhila Esna Ashari, *Member, IEEE,* Parastoo Alinia, *Member, IEEE,* and Hassan Ghasemzadeh, *Senior Member, IEEE*

Abstract—Recent years have witnessed a growing body of research on autonomous activity recognition models for use in deployment of mobile systems in new settings such as when a wearable system is adopted by a new user. Current research, however, lacks comprehensive frameworks for transfer learning. Specifically, it lacks the ability to deal with partially available data in new settings. To address these limitations, we propose *OptiMapper*, a novel uninformed cross-subject transfer learning framework for activity recognition. OptiMapper is a combinatorial optimization framework that extracts abstract knowledge across subjects and utilizes this knowledge for developing a personalized and accurate activity recognition model in new subjects. To this end, a novel community-detection-based clustering of unlabeled data is proposed that uses the target user data to construct a network of unannotated sensor observations. The clusters of these target observations are then mapped onto the source clusters using a complete bipartite graph model. In the next step, the mapped labels are conditionally fused with the prediction of a base learner to create a personalized and labeled training dataset for the target user. We present two instantiations of OptiMapper. The first instantiation, which is applicable for transfer learning across domains with identical activity labels, performs a one-to-one bipartite mapping between clusters of the source and target users. The second instantiation performs optimal many-to-one mapping between the source clusters and those of the target. The many-to-one mapping allows us to find an optimal mapping even when the target dataset does not contain sufficient instances of all activity classes. We show that this type of cross-domain mapping can be formulated as a transportation problem and solved optimally. We evaluate our transfer learning techniques on several activity recognition datasets. Our results show that the proposed community detection approach can achieve, on average, 69% utilization of the datasets for clustering with an overall clustering accuracy of 87.5%. Our results also suggest that the proposed transfer learning algorithms can achieve up to 22.5% improvement in the activity recognition accuracy, compared to the state-of-the-art techniques. The experimental results also demonstrate high and sustained performance even in presence of partial data.

Index Terms—activity recognition. machine learning. transfer learning. cross-subject boosting. optimization. wearable computing.

1 Introduction

Advances in machine learning, signal processing, and low-power computation, in conjunction with the emergence of low-cost sensing/processing/communication hardware technologies have led to advancing pervasive computing and Internet-of-Things (IoT) applications. Wearable systems, as a rapidly growing component of IoT, present numerous opportunities for real-time, context-aware and extensive health monitoring and intervention [1], [2], [3], [4]. The functionality of wearables is highly associated with the user's characteristics, which warrants high adaptability in their computational model. These models are typically powered by supervised machine learning algorithms. In order for an adaptive model to maintain an acceptable performance, a large training dataset is needed, which makes the training phase time consuming, labor intensive, and expensive. Data collection has been identified as a major obstacle in personalized and precision medicine [5]. Due to highly dynamic nature of wearable technologies, such a naive solution is deemed impractical. Examples of such variations are user-specific behavior, dynamic system architecture (addition of new sensors, replacing old sensors), various system platforms, and change in the location/orientation of the sensors. Collecting large enough training data to ensure adaptability to these variations is unrealistic given a large population of users are older adults. As a result, alternative solutions such as online/semi-supervised learning and transfer learning approaches are warranted [6], [7], [8], [9], [10], [11], [12], [13], [14], [15].

While subject autonomy is not limited to a specific learning approach, the focus of the work presented in this article is developing an autonomous transfer learning algorithm for cross-subject model adaptation. User-specific machine learning algorithms, unlike user-independent models, do not require a large training phase and tend to show a higher performance [16]. The major drawback is the need for collecting a personalized training data which puts much burden on the user's shoulder. Online learning, semi-supervised learning, and supervised transfer learning techniques are examples of such approaches. Cross-subject unsupervised transfer learning, in contrast, leverages the similarity between the existing training data and the observations collected from the new user to automatically predict labels for the user-specific instances and adapt/re-train the model using the new observations and the predicted labels, in real-time and without supervision.

We introduce OptiMapper based on the novel concept of

- *R. Fallahzadeh is with the School of Medicine, Stanford University, Stanford, CA, 94305. E-mail: raminf@stanford.edu*
- *Z. Esna Ashari, P. Alinia and H. Ghasemzadeh are with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164, USA. E-mail: {z.esnaashariesfahan,parastoo.alinia,hassan.ghasemzadeh}@wsu.edu*

Manuscript received XX; revised XX.

relational knowledge transfer for human-centered activity recognition applications. Our framework devises a novel knowledge transfer learning algorithm to discover the relationships between source and target domains in order to extract sufficient knowledge to develop its own personalized machine learning model. Developing the customized activity recognition algorithm is based on the observations made in real-time in target domain, adaptively fused with the predictions of the source model. Our methodology dismisses the implicit assumption that the data gathered in human-centered applications are from the same domain and probability distribution and instead aims to optimally rely on its own observations and their relation to source observations in order to build and update an accurate machine learning algorithm. The process is conducted in real-time without need for a synchronous teacher model [17] or supervised labeling effort to generate a new training dataset.

We aim to address the problem of reusing an activity recognition system developed and trained on a single (or multiple) user(s) on a new user without any training efforts. We assume that the labeled observations from the first group form the source view/domain and the unlabeled observations made by the new user form the target view/domain. Such a learning approach is sometimes referred to as multi-view learning [17]. OptiMapper is a novel combinatorial framework for cross-user mapping of sensor observations. To the best of our knowledge, our work in this article is the first to address the problem of transfer learning between unbalanced label spaces where the target dataset includes data from only a subset of activity classes that appear in the source domain.

In designing OptiMapper, we first focus on the case where the source and target domains share the same label space. The process of finding an optimal mapping across the two domains involves introducing a network of highly similar observations and performing greedy community detection for robust partitioning and detection of valuable observations in target domain. We formulate the problem of optimized target dataset mapping as an assignment problem, using linear programming and an uneven bipartite graph model. We then solve this optimization problem using a novel adaptive decision fusion.

After solving this one-to-one mapping problem, we extend OptiMapper to the case where the sets of labels used in the source and target domains are not identical such that only a subset of the labels within the source domain are included in the target dataset. The need for transfer learning in such a scenario is motivated by the fact that the set of activities performed by a particular subject may be only a subset of the set of pre-defined activities and therefore, we have partially available data in the target domain. This occurs specially in early stages of data collection in the target domain. Therefore, we extend the one-to-one mapping problem to a many-to-one mapping optimization problem that works with incomplete target dataset. The many-to-one mapping optimization problem extends our original combinatorial optimization algorithm to accommodate situations where we perform transfer learning across domains with non-identical activity labels. As a result of this improvement, we introduce an entirely new optimization approach and perform extensive analyses to demonstrate the efficiency of our algorithms (one-to-one mapping and many-to-one mapping).

This extended work, resulted in boosting the robustness of our method. To this end, we propose a graph representation of unbalanced many-to-one mapping for robust and partial label inference in the target domain. We formulate this new mapping problem as a linear constrained optimization problem and solve it using the Vogel's approximation method. Furthermore, three real-word wearable sensor datasets are used to evaluate our methods and to compare them against the baselines, upper-bound, and competing solutions.

2 Background and Related Work

One of the abilities which makes human intelligence superior to artificial intelligence is the ability to understand the similarity between various contexts and extend the learned knowledge to new contexts. Transfer learning is the ability to extend what has been learned from one context to another nonidentical but similar context that shares common features. In the field of machine learning, transfer learning is defined as follows [18]:

Definition 1 (Transfer Learning)**.** *Given a source domain D_s and learning task T_s, a target domain D_t and learning task T_t , transfer learning aims to help improve the learning of the target predictive function $f_t(\Delta)$ in D_t using the knowledge in D_s and T_s, where $D_s \neq D_t$, or $T_s \neq T_t$.*

This definition can be easily extended to the specific activity recognition domain where D_s and D_t are, respectively, the sensor observations made in source and target domain and can be represented in temporal, frequency, or feature space. The task T will develop a predictive function $f(.)$ by learning a conditional probability distribution $P(y|x)$ where y and x are the activity labels and sensor observations.

Depending on how the source and target domains are defined, several variations of transfer learning can be considered. Such domain variation in wearable activity recognition can occur due to user differences, device or sensor type changes, skewed label distribution, etc. Transfer learning in activity recognition can also be categorized based on the availability of the labeled data in the source or target domains [7]. Where the ground-truth labels are (or not) available in the source domain, the transfer learning is often characterized as supervised (or unsupervised) transfer. In addition, if labels are (or not) available in target domain it is referred to as informed (or uninformed) transfer learning. The categorization is further illustrated in Fig. 1.

Few aspects of transfer learning in activity recognition have been studied in recent years [7]. These studies can be categorized based on the target domain variation and availability of the labels, as described earlier. While the current work included various sensor modalities such as smart homes, wearables, and camera, our specific interest, in this study, is the human centered and wearable activity recognition.

In addition to the described categories, the current methodologies can be classified to synchronous or asynchronous transfer learning algorithms. While the majority of transfer learning algorithms are asynchronous, synchronous

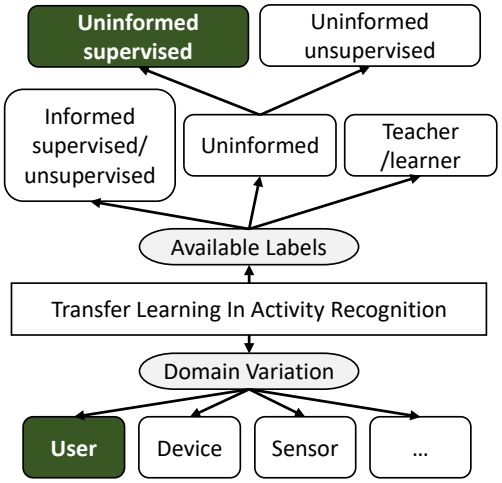


Fig. 1. Categorization of transfer learning in activity recognition based on source-target variation and availability of activity labels in source and target data.

teacher/leaner based methods have also been proposed, recently [17]. The biggest shortcoming of such approaches is the requirement of having an expert model present in real-time which is not usually affordable. In addition, the training of the expert requires enormous training efforts. However, in multi-node wearable systems such algorithms are shown to be effective [17], [19]. Prior research has also proposed a transfer learning approach based on exploiting intra-domain structures [20]. While this approach simplifies the transfer learning implementation by eliminating the need for model selection and hyperparameter tuning, it does not address the issue of partial target data, which is the main focus of our work in this paper. Also, the method named ActiLabel proposed in [21] utilizes graph structure and clustering techniques and is being implemented and compared with our proposed method under experimental results.

The algorithms proposed in this study aim to perform uninformed and supervised cross-subject transfer learning. In other words, we have a labeled source dataset acquired from the source subject and an unlabeled target dataset. The goal is to train a personalized and accurate activity recognition algorithm in the target domain that can boost the recognition performance in unseen users (with skewed feature space). To the best of our knowledge, this area of research is almost unexplored to date. The closest prior research conducted is the informed supervised/unsupervised learning where the assumption is that a partial labeled data is available in the target domain [22], [23]. This area of research sometimes manifests in active learning research [24], [25], [26], [27], [28].

Few studies [29], [30] suggested using unsupervised clustering algorithms such as K-means clustering algorithm to update the model trained on the source dataset. The biggest shortcoming of such approach is the naive assumption that the number of activities (i.e., the desired number of clusters) is a known parameter in target dataset. In contrast, in our approach we only define a lower bound on the size of clusters and propose a greedy solution to find the most desired partitioning. We demonstrate that naive clustering (and naive fusion of base-classifier and clustering algorithm) performance is not significant in a more realistic and complex scenarios where the a larger number of activities are present, as shown in more details in our analysis in Section 5.

3 OptiMapper

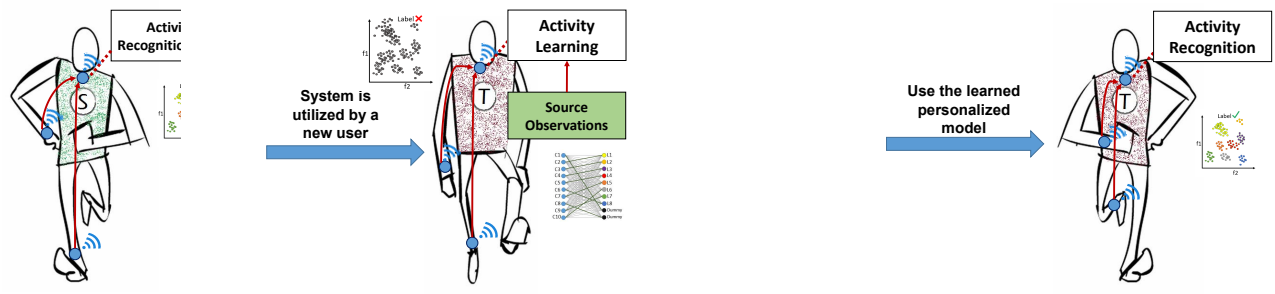
In this section, we present the general framework of OptiMapper for optimal cross-domain transfer learning. In this initial optimization framework, we assume that the activity vocabulary of the two domains are identical. In Section 4, we will extend our optimization framework to accommodate unbalanced datasets where only partial labels are available in the target domain.

Fig. 2 shows the evolution of a wearable activity recognition system as it is adopted by new users. Initially, a machine learning model (i.e., for activity recognition) is trained based on the annotated data (source dataset) collected from one or multiple subjects (Fig. 2a). In Fig. 2b, a new user adopts the system. Due to complexity and high variation of daily activities, the existing model exhibits significant drop in accuracy. As the user starts using the system, an unlabeled user-specific dataset is gradually collected (target dataset). We aim to utilize the collective knowledge extracted from the source and target datasets to predict activity labels for target dataset and train a user-specific model that ensures high accuracy, as shown in Fig. 2c.

3.1 Cross-Subject Transfer Learning Problem

Let X be a sequence of observations made from one sensing node containing a finite set of sensor. The observations made within a single time segment are represented in a K-dimensional feature space. As a result, vector $X_i = \{f_{1i}, f_{1i}, \ldots, f_{Ki}\}$ represents physical attributes during the time segment i. The set of all physical activity labels, with the size of m, forms a label space $L = \{l_1, l_2, \ldots, l_m\}$. The activity recognition algorithm assigns an activity label $l \in L$ to the observation X_i using a predictive function that is a conditional probability distribution $P(A|X_i)$ trained based on the labeled training dataset. The accuracy of such model is guaranteed as the future observations and activities will belong to the same distribution. However once a model is trained on training subjects, it will be available to new users. The observations made by the new user will not follow the same conditional probability distribution. Without loss of generality, we consider a single node activity recognition system, however, our methodology can be easily extended to multi-node body sensor networks.

Problem 1 (Cross-Subject Transfer Learning)**.** *Let DS_s be the labeled dataset collected from source subject(s) and $DS_t = \{X_{t1}, X_{t1}, \ldots, X_{tN}\}$ be the accumulative dataset composed of N unlabeled observations made from the new subject. Furthermore, let L_t be the activities performed by the new user. The Cross-Subject Transfer Learning (CSTL) is the problem of assigning a label $l \in L$ such that the error of label assignment is minimized. Once enough observations are labeled with activities, training a personalized activity recognition is straightforward.*



(a) Source use (b) Target user (c) Target user with personalized AR model

Fig. 2. The overview of cross-subject activity recognition. A wearable activity recognition system (consisted of one or multiple nodes) uses a model generated based on annotated data, i.e., source view (a). A new user adopts the system. An activity learning module is in charge of constructing a user-specific dataset in real-time, i.e., target view (b). When learning is completed, the new user begins using the updated model (c).

The minimization process aims to minimize the miss-assignments within the L_t set. In the rest of this section, we briefly explain the problem formulation and overview the proposed solution. For more details refer to [9].

3.2 Solution Overview

In our knowledge transfer based approach proposed in OptiMapper, the source domain acts as an offline teacher and the target view acts as a real-time learner. The source view transmits a training dataset containing features and its corresponding ground truth labels, referred to as the source dataset (DS_s).

Once the new user starts using the system, the target view begins collecting its own observations in real-time that are initially without labels and, therefore, cannot be used for training a model. We refer to the collected data as the target dataset (DS_t). After sufficient amount of data is collected, our activity learner module extracts feature-based similarities between the source and target datasets and executes an intelligent label transfer algorithm to construct a labeled target dataset. Once we have the labeled target dataset, developing a personalized activity learning model for the new subject is a straightforward task.

Fig. 3 shows the steps taken inside the learning module, in more details. The inputs are the source and target datasets. The source dataset (DS_s) is used to develop a supervised activity recognition model (i.e., source model).

The source model is tested on the target dataset to assign supervised label predictions, that are later used in our label fusion algorithm. The target dataset (DS_t) is used for constructing a network of highly similar observations in feature space and the observations that have a correlation higher than a preset threshold will be connected in our network. The threshold is set by a greedy algorithm, described in [9]. From the network, the highly similar observations are extracted as communities representing a particular activity label. The *minimum error subset mapping* module, then, will heuristically map the communities to a subset in our label space. At last, the labels acquired from both source and target views will be conditionally fused, based on their prior per-label performance. More details are presented in section 3.3 as best effort unsupervised labeling.

Algorithm 1 summarizes the main steps in our proposed knowledge fusion transfer learning approach. More details are presented in [9].

Algorithm 1 Knowledge-fusion-based Cross-Subject Transfer Learning

1: *Collect target observations;*
2: *Construct the similarity network (SN_t) using the DS_t;*
3: *Perform best effort community detection on SN_t to identify communities corresponding to a subset of labels in the label space;*
4: *Perform Minimum-Error Subset Labeling using collcted DS_t and DS_s and construct a labeled subset of DS_t;*
5: *Construct a supervised source model and predict labels for $\forall X \in DS_t$;*
6: *Given the conditional probability of lables assigned from both step 4 and step 5, being a correct label, assign the most probable label (weighted label fusion);*
7: **If** *not enough observations per label were extracted in step 4 (it can be observed from the convergence of community mapping results)***:**
8: *Go to step 1;*
9: *Train a personalized activity recognition model using the constructed labeled DS_t;*

3.3 Algorithm Design and Optimization

We formally define the problem of *'minimum error subset labeling'* and propose a greedy heuristic solution. The objective is to use the relational knowledge transferred from the labeled source dataset, instead of relying on the partially biased source model, to predict activity labels for unlabeled target observations and construct a labeled target sub-dataset of DS_t which can be feed into the label fusion module.

We note that assigning a label to every single observation in DS_t, using an unsupervised method, is less likely to achieve a high accuracy, especially in presence of a large and complex set of activities. Instead, we aim to find the observations that can be labeled with higher confidence. As a result the constructed labeled dataset in target view is a more accurate assignment, containing only highly confident instances, resulting in a more accurate mapping. The challenging task is to find such an accurate *(Observation Subset,Activity Label)* mapping. We refer to such observations as core observations.

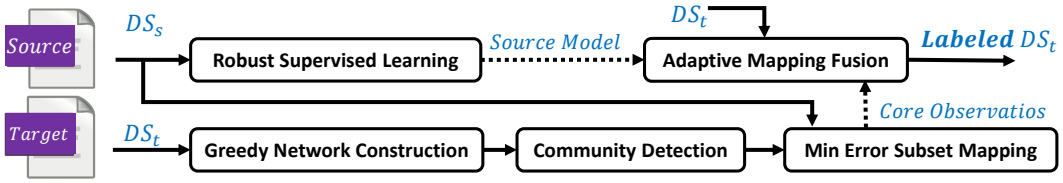


Fig. 3. The block diagram of OptiMapper learning module. It includes network construction, community detection, minimum error subset mapping, and a weighted label fusion of unsupervised and supervised label mappings.

Definition 2 (Core Observation). *An observation X_i is called a core observation for a label $l \in L$ if and only if it has the highest correlation with the source observations with the same label, with respect to all non core observations of l such that there are at least δ_j and at most Δ_j of such core observations.*

Problem 2 (Minimum-Error Subset Labeling). *Given a source dataset (DS_s) and a target dataset (DS_t), find a subset of DS_t (core observations) that can be labeled with activity labels L_t such that the error of labeling is minimized.*

In order to simplify the problem, we first group the observations in both source and target view into several clusters, where highly similar observations are assigned to the same cluster, with standard clustering assumptions applied [31], [32]. In the source view, the clustering task is trivial, as the ground truth labels are available. Therefore, we place the instances that share the same label in one cluster. Therefore, the number of clusters equal the number of activities:

$$DS_s = C_1^s \cup C_2^s \cup, \ldots, \cup C_{|L|}^s \tag{1}$$

such that

$$C_1^s \cap C_2^s \cap, \ldots, \cap C_{|L|}^s = \emptyset \tag{2}$$

In target view, the ground truth labels are unknown and since gathering the target data is a collective process, the exact number of activities performed by the target user is unknown. Therefore, we cannot use a clustering algorithm such as K-means.

Because of the resoluteness that network-based community detection clustering algorithms provide, we use them to achieve higher clustering performance. We measure similarity between sensor observations (i.e., the nodes in the similarity network) in feature space to detect communities of similar nodes.

We use the training data from the source subjects to find the optimal similarity threshold that maximizes the clustering performance. The similarity score used in our approach is the *normalized cosine similarity* due to its superior performance in our high-dimensional feature space. We also define a lower-bound on the size of clusters to discard the small communities that represent the outliers, resulting in a subset of instances in DS_t, that are potentially the core observations. The details of our greedy algorithm for finding the optimal similarity threshold can be found in [9]. Then, we propose a complete bipartite graph model, referred to as Weighted Labeling Graph, using the identified clusters in source and target view and their pairwise degree of similarity. To measure similarity, we again use the normalized cosine similarity between clusters in both views, having observation instances represented in feature space. The significance of our graph model is that it can be used to generalize Problem 2 to a bipartite matching problem:

Definition 3 (Weighted Complete Similarity Graph). *Let C^s be the set of clusters in source view where each cluster is associated with an activity label and C^t be the set of unlabeled clusters (i.e., valid communities) in target view. $G(V, E, W)$ is a weighted complete bipartite graph where $V = \{C^s \cup C^t\}$ and E and W, respectively, refer to the set of edges and their corresponding weights connecting the vertices in C^s to the vertices in C^t.*

The weight $\omega_{ij} \in W$, in our weighted complete similarity graph, represents the degree of similarity between each $(C^s - C^t)$ pair and is given by

$$W = \{\omega_{ij} = normalized\_cosine\_similarity(C_j^s, C_i^t)\} \tag{3}$$

When multiple source users are available, the set W will be the average computed weight over all the source users.

Note that similarity is calculated using normalized raw features. Moreover, the selection of the cosine similarity in our work is motivated by prior research that showed the superiority of the cosine measure over other distance functions [33], [34]. Nonetheless, note that the choice of the distance function is independent of the methodology presented in this paper.

Fig. 4 illustrates an example of the construction of our graph model. Fig. 4a shows the unlabeled target observation in a two dimensional feature space. While the labels are initially unknown. In Fig. 4b, our unsupervised community-based algorithm partitions the target dataset with respect to the predefined lower-bound on the cluster size. We create the weighted complete bipartite graph using the clusters ids in target view and the clusters ids in source view, as shown in Fig. 4c. The weights on each edge is computed from Equation (3). We note where $|C^s| \neq |C^t|$, we add dummy nodes to create a complete bipartite graph.

We propose a simple polynomial time label transfer algorithm that maps the activity labels from $|C^s|$, to a subset of clusters in $|C^t|$. We want maximum similarity in our bipartite matching. Using the constructed weighted complete bipartite model, our algorithm uses maximum weight bipartite matching to perform the maximum similarity label mapping, as shown in Fig. 4d. We use Hungarian Algorithm [1]. The edges incident to dummy nodes have weights equal to zero. As a result, the nodes matched to the dummy nodes are excluded after the matching. Finally, the assigned labels are propagated to the target observations inside each cluster, as shown in Fig. 4e.

We construct a state-of-the-art robust supervised machine learning algorithm, called source model, trained on

1. https://en.wikipedia.org/wiki/Hungarian_algorithm

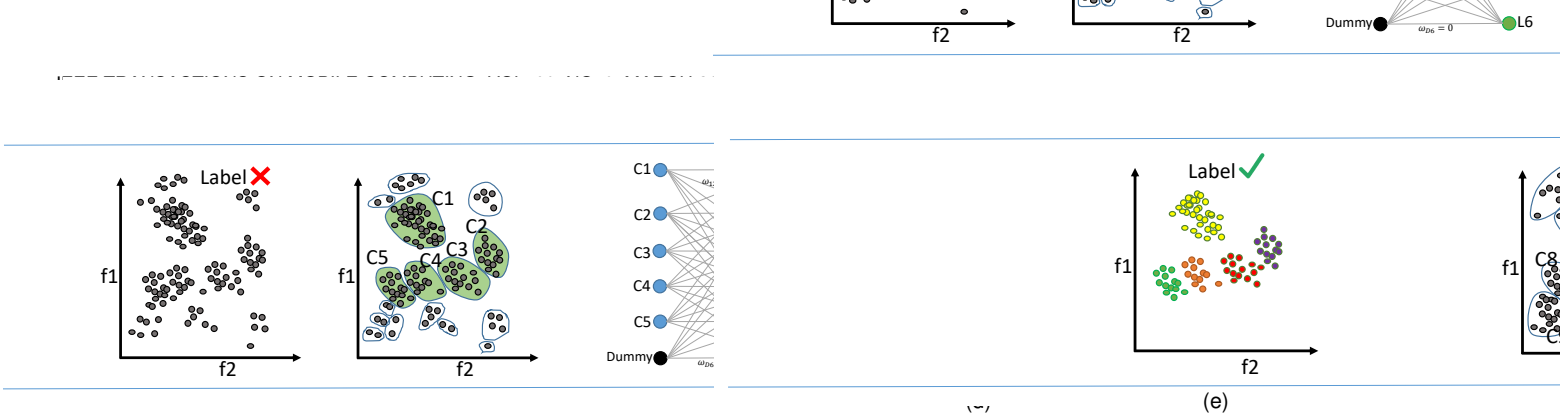


Fig. 4. The overview of the general label transfer approach in OptiMapper framework for balanced dataset: (a) unlabeled target data; (b) community detection based clustering; (c) graph model for label inference; (d) mapping labels from source to target clusters; (e) the resulting labeled target data.

the source user(s) data to provide predictions on target observations. In such setting, for each observation, we have at least one prediction from the source model and a possible mapping from the $DS_t^{Labeled}$. Two approaches can be taken with this information: (1) naive label fusion, and (2) adaptive or conditional label fusion.

In naive label fusion, we assume the accuracy of both supervised an unsupervised labeling algorithms are user-independent and therefore the prior assumption is that the output label of the algorithm is unconditionally true. Therefore the prior probability distribution of the output label $l' \in L$ is defined as follows:

$$P(l|l') = \begin{cases} 1 & l' = l \\ 0 & otherwise \end{cases} \tag{4}$$

where l and l', respectively, denote the ground-truth label and the assigned label by either the unsupervised or the supervised algorithms.

Under such a circumstance, for each observation in DS_t, a majority vote will assign the final label.

However in the conditional label fusion approach, a prior knowledge of source user's performance can be learned in a form of a conditional probability distribution: $P(L|L')$. Therefore the predictive function can be reformulated as:

$$f(X_i) = P(L|L')P(L'|X_i) \tag{5}$$

The performance decline in cross-subject activity recognition can be generally explained by the distribution shift across the two datasets obtained in the source and target domains. The data distribution shift causes the previously established decision boundaries obtained in the source domain to fail to accurately distinguish among activity classes in the target domain. However, in this work, the introduced similarity graphs are used to find an optimal matching between the clusters of the input data across the two domains. The optimal solution finds a matching that minimizes the distance across the nodes in the bipartite graph and therefore unlike a statically defined classifier, it adapts itself to the distribution shift. As soon as the optimal matching is found and the labels are propagated to the target distribution, a new classifier is trained for activity recognition in the target domain.

4 Accommodating Unbalanced Datasets

One limitation of the cross-subject transfer learning algorithms discussed in Section 3, is its assumption that the label space in both source and target domains are identical. In other words, we only considered the case where the activities performed by target user (i.e, L_t) are the same as the activities performed by the source users (i.e., L). Such an assumption does not hold in two scenarios:

1) the target user performs new activities that are not defined in the annotated source dataset.

$$\exists l \in L_t \mid l \notin L \tag{6}$$

2) the target user does not perform at least one of the defined activities in the source domain.

$$\exists l \in L \mid l \notin L_t \tag{7}$$

Scenario (1) could occur for a number of reasons. One reason could be because the annotation (training) process has not been carried out comprehensively such that it is missing some of the common physical activities. Another case is when the target user performs an unanticipated or uncommon activity. This is more likely to happen in activities of daily living (ADLs) where the number of activities are large and more diverse, such datasets either ignore the undesired data or classify it as 'Other Activities' [35]. On the contrary, when monitoring low-level physical activities, such cases are less likely to occur. In addition, several unsupervised and semi-supervised approaches have been proposed to address the detection of new (unseen) activities [36], [37]. However, because of the lack of annotated source instances for the unseen activity, an uninformed supervised transfer learning approach such as the one presented in this work is not applicable.

The second scenario can happen as the training dataset becomes more inclusive of more activity labels and hence is more likely to include less common activity types. For instance, imagine an example where the target user does not perform the activity 'Cycling' that has been defined in the source label space. Another case is when target data collection is in early stages and not enough unannotated data have been collected.

Our goal in this section is to address the problem of non-matching activity labels as described in the second scenario above. As a result, we extend the OptiMapper framework so that it handles these cases as well.

4.1 Problem Statement

As mentioned, the proposed mapping problem in Section 3 was based on the assumption that the target data contains enough instances of every activity label in the source dataset (DS_s). Therefore, for the label mapping, we were able to formulate the problem as a bipartite one-to-one matching problem and solve it efficiently using Hungarian Algorithm. However, such an assumption limits the usability of the OptiMapper and is dismissed in the approach presented in this section.

In Fig. 5, various cases of one-to-one mapping are illustrated when the label spaces are identical in both DS_s and DS_t. After target dataset partitioning is performed, clusters of target observations are mapped to clusters of source observations. Clustering the source observations is trivial because the instances are annotated. In an ideal case, the number of clusters in both target and source domains is equal (Fig. 5a). However, it is possible that the best effort partitioning algorithm fails to cluster a subset of instances corresponding to a particular label. In such a case, a dummy node is added to C^t set in order to balance the bipartite matching model (see Fig. 5b). It is also possible that a set of target observations that belong to the same activity label are grouped into more than one clusters in C^t (resulting in $|C^t| > |C^s|$). Similar to the previous case, a dummy node is added, this time, to C^s to balance the matching problem.

A one-to-one mapping, e.g. balanced bipartite matching, requires each cluster to be mapped to one and only one cluster. Therefore, if two target clusters of the same label exist, only one of them can map to the corresponding label and the other one will be forced to map to, ideally, a dummy node or an incorrect source node. When there are missing activity labels in C^t, such as in the example shown in Fig. 6a, the extra clusters of the same label can be mapped to an incorrect label. In the example shown, label $L6$ (colored in green) is missing in the target data. Consequently, in the absence of a green target cluster, the balanced bipartite matching algorithm has produced $(C5, L6)$ mapping which is an 'incorrect' label assignment. Additionally, the target cluster $C7$ which corresponds to $L2$ (blue) label is mapped to the dummy node and hence discarded. An ideal mapping algorithm is an unbalanced many-to-one mapping where (1) multiple target clusters can be assigned the same label and (2) a source cluster (a label in source domain) can be left un-assigned, as illustrated in Fig. 6b.

As depicted in Fig 6, the second scenario, which refers to having two source clusters with the same label and a missing target label, can be covered in the many-to-one mapping approach. This is handled by allowing sub-cluster of target activities pair together with the closest label even if the target subject is missing a source label. The example from Fig. 6b, shows the matching from the many-to-one approach, when $L6$ is not present in the target and hence not matched to any of the clusters. However, the matching shown in Fig. 6b would be an incorrect matching resulted from the initial approach not being able to handle this scenario.

Furthermore, while the matching always optimizes for the lowest distance, theoretically, one can imagine a case where two activities in the target dataset are distorted in a way that are closer to the opposite label in the source dataset. However, such a scenario is quite unlikely, because of the intrinsic physiological similarities in human activities. Nonetheless, in cases where the labels are similar the matching could be affected as shown in tables 1 to 3 presented in the results section showing the accuracy of label transfer not being perfect.

4.2 Problem Definition

Let $X_i = \{f_{1i}, f_{2i}, \ldots, f_{Ki}\}$ be $i-th$ sensor observation represented in a K-dimensional feature space. A set of sensor observations in source domain, that each correspond to a label $l \in L_s$, form an annotated source dataset DS_s. The source dataset can be represented as a pairwise disjoint subsets of source clusters (C^s):

$$DS_s = \{\{C_1^s\}, \{C_2^s\}, \ldots, \{C_{|L_s|}^s\}\} \tag{8}$$

such that C_i^s is the set of all $X^s \in DS_s$ which correspond to $l_i \in L_t$.

Similarly mutually disjoint sets of unannotated target dataset can be found. Because the ground truth labels are unknown in the target domain, the networked based partitioning is performed to produce the target clusters. Each target cluster C_i^t corresponds to a cluster id (i) and will later mapped to a label $l \in L_t$. The mapping problem aims to map each target cluster to a source cluster such that the cosine similarity between the centroids of two clusters mapped together is maximized. The many-to-one mapping shown in 6b is defined as follows:

Problem 3 (Many-to-One Unbalanced Cluster Mapping)**.** *Let DS_s be the labeled dataset collected from source subject(s) and $DS_t = \{X_{t1}, X_{t1}, \ldots, X_{tN}\}$ be the accumulative dataset composed of N unlabeled observations collected from the new subject. Furthermore, let $L_t \subseteq L_s$ be the activities performed by the new user. The Unbalanced Cross-Subject Transfer Learning (U-CSTL)–or Many-to-one Unbalanced Cluster Mapping–is the problem of assigning a label $l \in L_t$ to each target cluster by finding the optimal many-to-one optimal target-source mapping. The optimal mapping maximizes the overall similarity of $(C^s - C^t)$ label mapping such that each C^t is mapped to exactly one C^s.*

The maximization process aims to minimize the miss-assignments. In the rest of this section, the U-CSTL problem is formulated and an optimal solution is proposed.

4.3 Problem Formulation

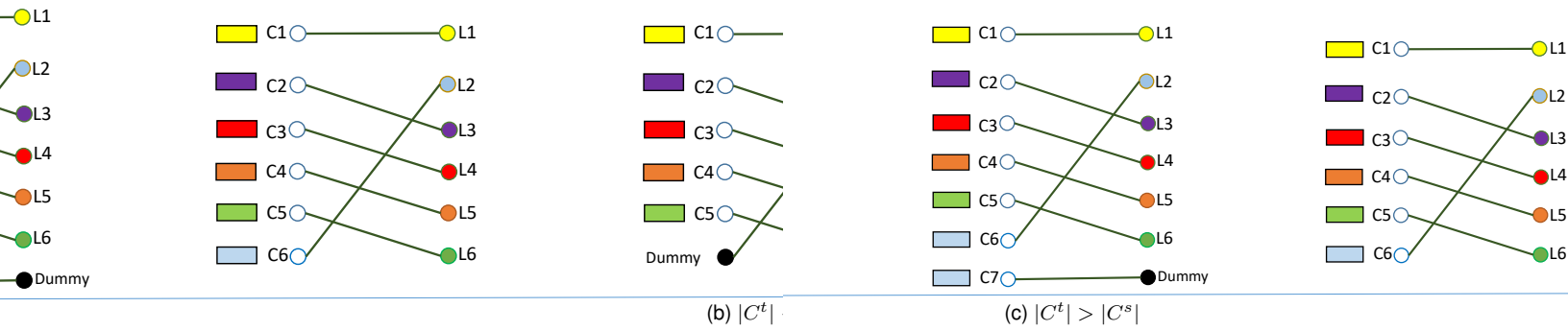
Consider a set of n target clusters $C^t = \{C_1^t, C_2^t, \ldots, C_n^t, \}$ and a set of m source clusters $C^s = \{C_1^s, C_2^s, \ldots, C_m^s, \}$. The Unbalanced Cross-Subject Transfer Learning (U-CSTL) problem, defined in Section 4.2, can be formulated using linear programing as follows.

$$Minimize \quad \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}c_{ij} \tag{9}$$

Subject to

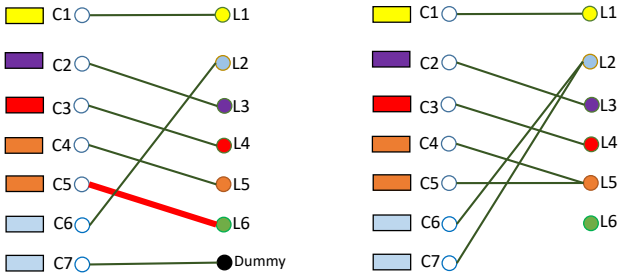
$$\sum_{1\leq j\leq m} x_{ij} = 1, \ \ \forall i \in \{1, \ldots, n\} \tag{10}$$

$$0 \leq \sum_{1\leq i\leq n} x_{ij} \leq n, \ \ \forall j \in \{1, \ldots, m\} \tag{11}$$



(b) $|C^t|$ (c) $|C^t| > |C^s|$

Fig. 5. Three cases of one-to-one matching when $L_t = L_s$. Target clusters (C^t) and source clusters (C^s) are on the left side and right side of the graph model, respectively. The rectangles next to target clusters indicates the ground truth labels associated with the instances in the corresponding cluster.



(b) Many-to-one unbalanced mapping

Fig. 6. An example of one-to-one vs many-to-one mapping when $L_t \subset L_s$: (a) C7 is mapped to a dummy label (and therefore discarded) and C5 is mapped to L6 resulting in an incorrect label transfer; (b) More than one target clusters can be mapped to a source cluster and therefore avoiding removal or incorrect labeling of target clusters.

x_{ij}

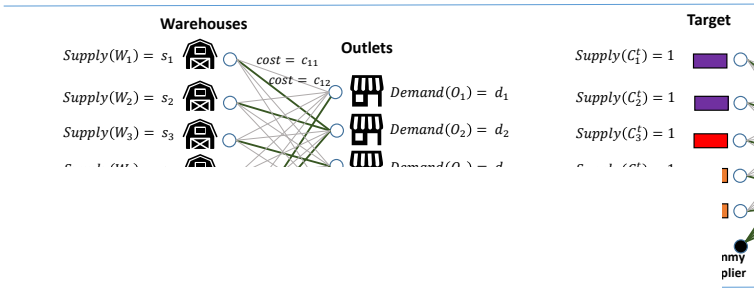
where c_{ij} is the
Equation (9) ai
assignment in t
distances betwe
x_{ij} is set to '1'
otherwise. Cons
ity label is assig
(11) states that
zero and at mos
(12) simply state
negative.

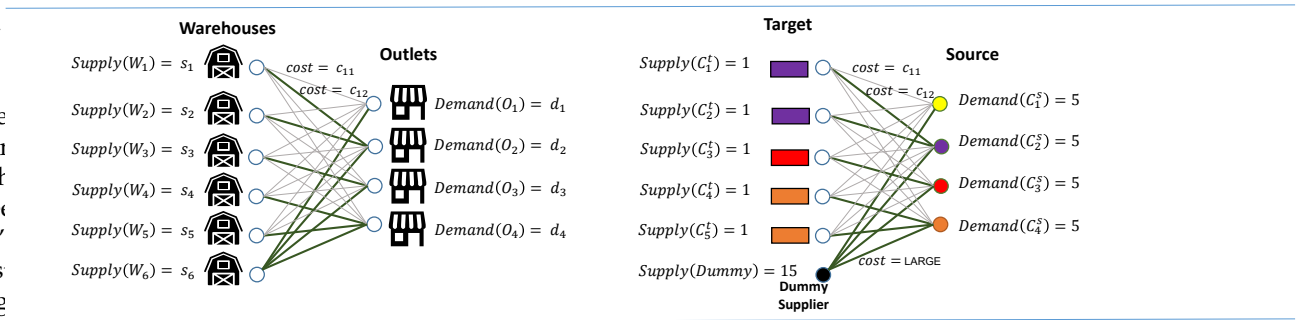
4.4 Reduction

In this section, we describe how a transportation problem [38] can be reduced to Problem 3 and solved optimally.

4.4.1 *Transportation Problem*

Assume a set of n warehouses $\{W_1, W_2, \ldots, W_n\}$ with given amount of supply (s_i) to each W_i and a set of m retail outlets $\{O_1, O_2, \ldots, O_m\}$, each with a given demand (d_i). There is a distinct road between each pair of (W_i, O_j) with a given transportation cost. The problem of interest is to find the optimal transportation scheme that minimizes the total cost and satisfies the supply and demand constraints. A trans
com
prob
is vi





(b) U-CSTL problem

Fig. 7. An example of a many-to-one unbalanced cross-subject cluster mapping problem formulated as a transportation problem.

4.4.2 *Formulating U-CSTL Problem as a Transportation Problem*

Generally, a transportation problem produces a many-to-many mapping. A special case of the transportation prob-

lem can be defined to produced the many-to-one mapping desired to solve the U-CSTL problem. We proceed with describing how each part of the linear programing is defined to model the discussed U-CSTL problem.

- **Decision variables:**
  The parameter x_{ij} specifies how many products is shipped from W_i to O_j. We assume each warehouse is a target cluster (C_i^t) and each outlet is a source cluster (C_j^s). As a result, a set of $m \times n$ binary variables x_ij can be defined to specify whether or not there exists a mapping between C_i^t and C_j^s.
- **Objective function:**
  The overall cost of transportation is $\sum_{\forall i,\forall j} c_{ij}x_{ij}$ where c_{ij} denotes the cost of shipping one product from W_i to O_j. Redefining the cost function as the distance between the pair of target-source clusters yields the same linear objective function.
- **Problem constraints:**
  In addition to the defined constrains in transportation problem, we add the following additional constraints:
  - Each warehouse in U-CSTL (i.e., C_i^t) has a supply equal to 1. It means that each target cluster is only mapped to one source course cluster (C_j^s).
  - Each outlet in U-CSTL (i.e., C_j^s) has a demand equal to the number of target clusters ($|C^t|$). In other words, a source cluster can be mapped to at least zero and at most all of the target clusters. The first two items guarantee a many-to-one mapping.
  - The total supply is equal to the total demand. If not satisfied, a dummy node with supply/demand of $|\sum_{i=1}^{n} s_i - \sum_{j=1}^{m} d_j|$ will be added. This special case of transportation is referred to as ' Balanced Transportation Problem' [39]. The reason is to ensure all of the supply from C_i^t is required by the source clusters. In other words, all of the target clusters will be mapped to a source cluster.

4.5 Optimal Solution

Now that the initial U-CSTL problem is modeled as a balanced transportation problem, we can employ the existing efficient solutions to find the desired optimal mapping. Solving a balanced transportation problem is done in two steps: (1) finding a basic feasible solution (BFS); (2) iteratively improving the BFS to achieve the optimal solution. We use Vogel's Approximation Method (VAM) for computing the BFS. VAM has shown to yield a better (closer to optimal) BFS compared other methods such as North West Corner Rule [39], [40]. For the optimality test, we use the Stepping Stone Method [41].

5 Experimental Results

In this section, we investigate the effectiveness of OptiMapper framework. We consider both proposed unbalanced many-to-one mapping based approach, namely U-CSTL, and the initially proposed cross-subject transfer learning method (CSTL), as well as state-of-the-art robust classification algorithms, and the computational upper-bound. In order to emphasize the generalizablity of our results, three real-world wearable activity recognition dataset has been used: (1) UCI 'Daily and Sports Activity Dataset' [42]; (2) UCI 'PAMAP2 Physical Activity Monitoring Dataset' [43]; and (3) 'SmartSock Activity Recognition Dataset'. After obtaining raw sensor data from these sources, we extracted 10 commonly used features from a fixed-size moving window over the raw sensor signals. The features include maximum, minimum, signal amplitude, median, mean, peak-to-peak amplitude, variance, and RMS (Root Mean Square) power. The extracted features were used for further analysis.

5.1 Datasets

All of the datasets used in our experiments are publicly available and explained in the following subsections. For each set, after collecting raw data from their resources, feature extraction process was performed. The feature extraction has been carried out using a non-overlapping sliding window approach. A sliding window of 5 seconds created data segments and the following statistical features were extracted for each signal segment: Amplitude, Median, Mean value, Maximum amplitude, Peak to peak amplitude, Variance, Standard deviation, Root mean square power, and Start to end value. Refer to [44] for more details on the feature set utilized in the experiments. Calculate feature values for each data sample in the dataset, were used in our experiments. In the following, we provide a brief description of each dataset.

5.1.1 Daily and Sports Activity Dataset:

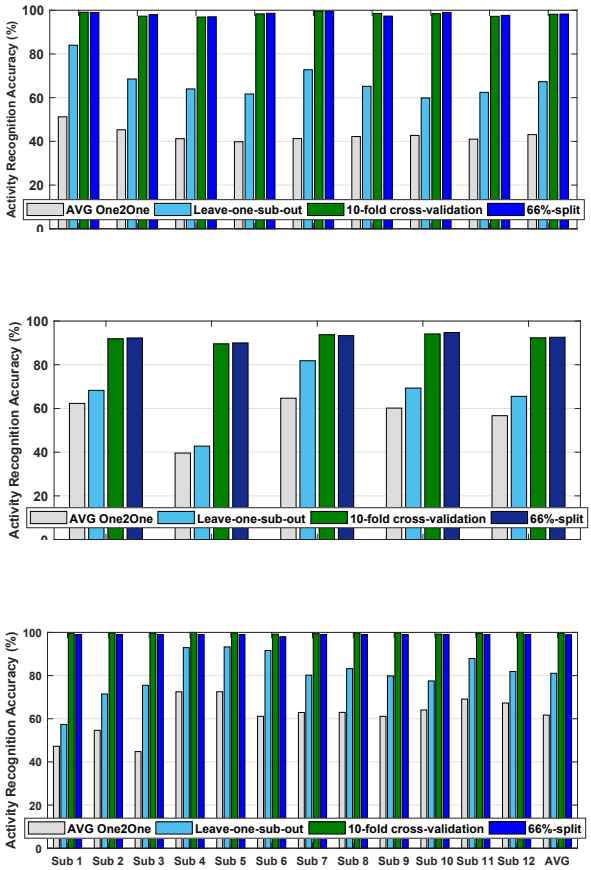
A 3D motion tracker placed on torso was used to collect triaxial acceleration data at $25Hz$ sampling rate. Eight subjects were asked to perform 15 activities (such as sitting, walking, running at different speed, cycling, rowing, etc.), each for duration of five minutes. The dataset contained over $900,000$ samples of acceleration data.

5.1.2 PAMAP2 Physical Activity Monitoring Dataset:

A Colibri wireless inertial measurement units (IMU) placed on chest was used to collect data from four subjects. Each subject has been instructed to perform seven low-level physical activities: lying, standing, walking, running, cycling, ascending stairs, and rope jumping. The triaxial accelerometer sampled at $100Hz$ with $\pm 16g$ range and $13-bit$ resolution. We note that originally the PAMAP2 dataset contained more data. However, due to having missing data for some of the subjects and activities, a number of the activities and subjects were excluded in the compiled dataset resulting in more than 275000 data samples.

5.1.3 SmartSock Activity Recognition Dataset:

This dataset has been collected using the SmartSock prototype presented in [45]. It contains triaxial acceleration data from 12 healthy individuals. The motion sensor was placed on subject's right ankle sampling at $18.7Hz$, The activities include standing, laying down, jumping, descending stairs, walking, and running. The compiled dataset contained more than 61000 sampled data.



(c) SmartSock Activity Recognition (SCAR) dataset

Fig. 8. Activity recognition accuracy using various evaluation methods for DSA, PAMAP2, and SCAR dataset.

5.2 Preliminary Performance Evaluation

In our initial analysis, we measure the activity recognition accuracy for each dataset using different evaluation methods. Several popular machine learning algorithms were initially considered, however, only the results for Random Forrest algorithm is shown as it has demonstrated superior performance. Four different evaluation methods were utilized to assess the accuracy of activity recognition.

Average one-to-one evaluation: In this method data from each subject is used as a test set where the model is trained by using data from one other subject. Given n as the number of subjects in a dataset, for each target subject, the average prediction accuracy achieved by all $(n-1)$ source models is reported.

Leave-one-subject-out evaluation: This method trains a model using all available source subjects and validates using the remaining target user data.

10-*fold cross-validation*: This is a standard 10-fold cross-validation method where, in each fold, 90% and 10% of the initially randomized data (containing samples only from the same subject) is used as the training and test set, respectively.

66%-*split evaluation*: This method randomly splits each subject's data into two sets. The training and test set contain 66% and 34% of the entire subject data, respectively.

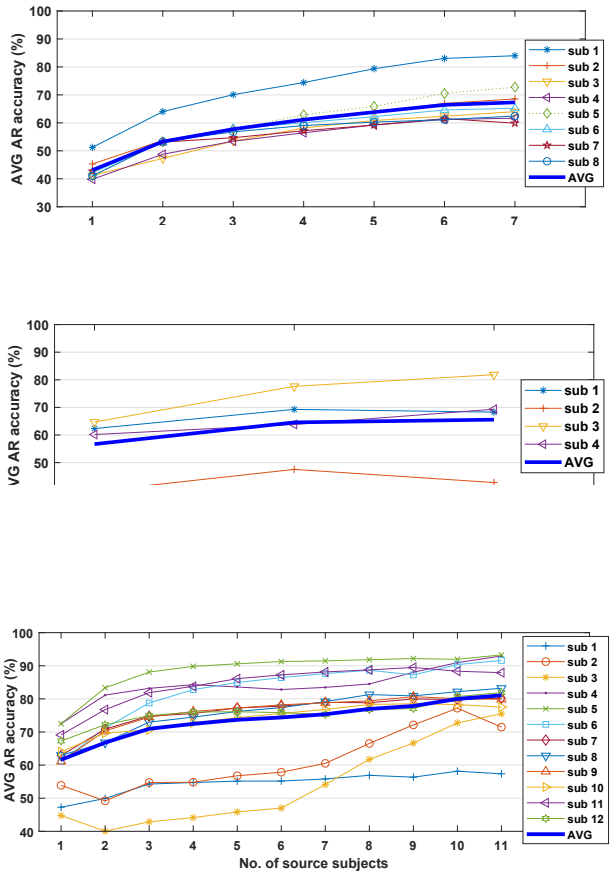
Fig. 8 shows the activity recognition performance vs the chosen target subject for each of the discussed evaluation methods. The last group is the average over all subjects. All three datasets consistently demonstrate a significant performance decline of the supervised model when used on a new untrained subject. For instance, when the model is trained on only one source subject, on average, it shows 55.7%, 35.64%, and 37.94% decline in prediction accuracy, from 10-fold cross-validation, respectively for DSA, PAMAP2, and SCAR datasets. These numbers are, respectively, 30.86%, 26.79%, and 18.57% when using all of the training subjects together. While noticeably improving the performance by adding new subjects, the baseline still lags behind the experimental upper-bound on performance (demonstrated by 10-fold cross-validation and 66%-split evaluations).

Note that while it is correct that the size of training data generally impacts the algorithm's performances, Fig. 8 shows that the effect of the size, once adequate data is gathered, is not nearly as important as data from cross-user domains. For instance for 66%-split vs 10-fold cross validation, fewer training data from the same subject was used (66% vs 90%), but the performance remained virtually unchanged and stayed at its maximum value. That is because the training data is from the same user. However, the decline happens when the data come from a different subject, which are One2One and Leave-one-sub-out, since these cases both use data from other subjects. The comparison of One2One vs Leave-one-sub-out is detailed in Fig. 9. Leave-one-sub-out has many folds more than 10-fold cross validation and 66%-split, but the accuracy is lower, because it is from a shifted domain of other subjects. One2One also has a bigger training size than 66%-split and 10-fold cross validation, but the performance is much lower. Therefore, when taking the size comparisons into consideration, it further strengthens our point of showing the adverse impact of domain shift in subjects on activity recognition accuracy.

In another experiment, we investigated the impact of growing the pool of source subjects on boosting the baseline performance. We start by using one subject as the source and iteratively add one other subjects to the training pool until we have exhausted all available source subjects. For example, in a dataset containing n subjects, $(n-1)$ distinct sizes of source subject pool (from 1 to $(n-1)$ sources) can be considered. For a source pool of size k subjects, the number of different models (choosing k from $n-1$) is a $k-$combination of the set of source subjects of size $(n-1)$.

For each size of source subject pool (i.e., k) and each target subject, the average accuracy over all possible combinations is reported in Fig. 9. In addition, the bold blue line shows the average over all target subjects for a given k. It can be observed that all three datasets illustrated the same trend where, by adding more subject data to the source, the performance increases rapidly at first but converges towards a constant line as the source data grows larger. This shows that adding excessively large number of source subjects cannot be considered a reliable solution for overcoming the cross-user variance. Specially when you take into account that in many cases, acquiring a large annotated training pool is costly if not impossible. In some cases of this experiment, a slight performance increase can be observed.

Based on this figure, it can be seen that utilizing more subjects for training may decrease the model's performance. However, the presented results are based on average num-



(c) SmartSock Activity Recognition (SCAR) dataset

Fig. 9. Activity recognition accuracy of random Forrest algorithm vs the number of source subjects.

bers taken over subjects' results and the average is biased to few large values. While the general trend is showing improvement over increasing the pool size and then an eventual plateau as one would expect, depending on the target distribution, there could be cases where adding more data not only does not increase the performance but could also hurt it slightly, e.g. when the target data is far from most of the source subjects' data. In this case, generalizing on more data would not necessarily yield an improvement. It should be noted that in these experiments, we are interested in average performance in general as the cross-user domain shift in the real data is not evenly present.

5.3 Accuracy of Unsupervised Partitioning

In the second part of our analysis, we evaluated the best-effort network-based partitioning. Here, the step size for α is set to 0.5×10^{-4}. As the unbalanced cross-subject cluster mapping proposed in this section can map more than one target clusters to one source label, our procedure returns the threshold (α) that maximize the clustering accuracy and dataset utilization in source domain:

$$\alpha_{opt} = \arg_{\alpha} max \quad \gamma^s(\alpha) + k \times \theta^s(\alpha) \tag{14}$$

where $\gamma^s(\alpha)$ denote the fraction of the correctly clustered instances over all clustered instances (clustering accuracy, $0 \leq \gamma \leq 1$) and $\theta^s(\alpha)$ represent the fraction of clustered instances over all the input instances (dataset utilization $0 \leq \theta \leq 1$) for a given α. The parameter k is the tuning variable ($k \geq 0$) that adjusts the significance of larger dataset utilization over clustering accuracy. Because higher accuracy is typically more desirable that high utilization, for our experiments, the parameter k is set to 0.34 (which means a high γ is approximately two times more desirable than a high θ).

TABLE 1
Partitioning the target data (DSA dataset) using the proposed best effort networked-based community detection

Subject	Selected Opt. Threshold	Dataset Utilization	Accuracy
Sub. 1	99.24%	0.57%	0.79%
Sub. 2	98.79%	74%	81%
Sub. 3	99.26%	66%	76%
Sub. 4	99.28%	71%	79%
Sub. 5	99.96%	53%	96%
Sub. 6	98.70%	89%	75%
Sub. 7	98.79%	59%	70%
Sub. 8	99.28%	68%	88%
Average	-	67%	81%

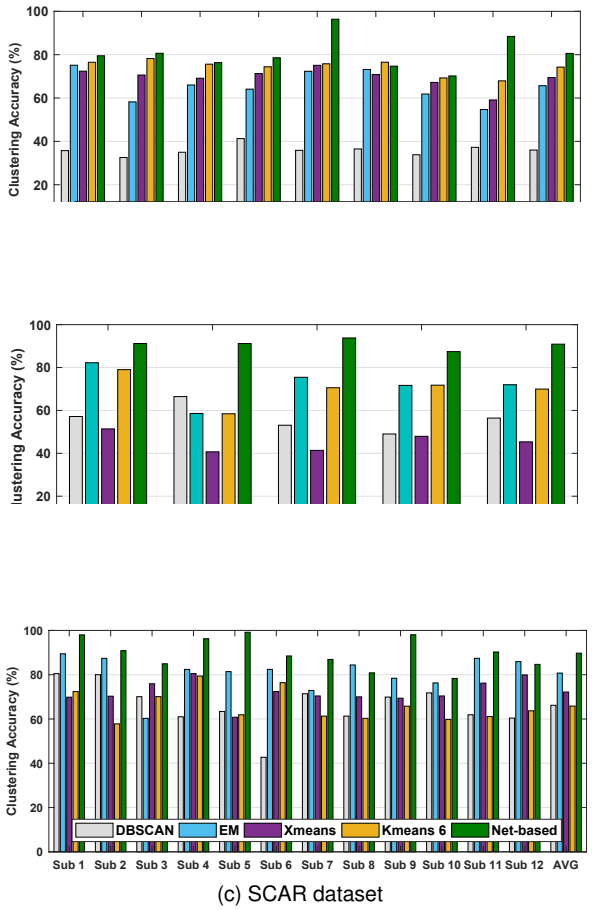
TABLE 2
Partitioning the target data (PAMAP2 dataset) using the proposed best effort networked-based community detection

Subject	Selected Opt. Threshold	Dataset Utilization	Accuracy
Sub. 1	99.29%	75%	91%
Sub. 2	99.26%	71%	91%
Sub. 3	99.29%	71%	93%
Sub. 4	99.29%	71%	87%
Average	-	72%	91%

TABLE 3
Partitioning the target data (SCAR dataset) using the proposed best effort networked-based community detection

Subject	Selected Opt. Threshold	Dataset Utilization	Accuracy
Sub. 1	99.91%	67%	98%
Sub. 2	99.91%	77%	90%
Sub. 3	99.91%	58%	85%
Sub. 4	99.91%	64%	96%
Sub. 5	99.91%	66%	99%
Sub. 6	99.91%	59%	88%
Sub. 7	99.91%	67%	87%
Sub. 8	99.91%	80%	81%
Sub. 9	99.91%	97%	98%
Sub. 10	99.91%	66%	78%
Sub. 11	99.91%	75%	90%
Sub. 12	99.91%	56%	85%
Average	-	69%	90%

The results of the proposed community-detection clustering using the objective function in Equation (14) are shown in Table 1, Table 2, and Table 3, respectively for DSA, PAMAP2, and SCAR datasets. The optimal cut-off threshold (α) and its corresponding data utilization and clustering accuracy is listed for each target subject, where the percentage of data utilization shows the fraction of clustered instances over all the input instances. The results show that our algorithm can achieve significant clustering accuracy while utilizing a large portion of data in the final set of target clusters. Using DSA dataset, the final clustering, on average, included 67% of the input observations with 81% accuracy. Similarly, the average dataset utilization was



(c) SCAR dataset

Fig. 10. Clustering accuracy for each target subject using the proposed networked-based partitioning vs standard clustering algorithms

72% and 69%, and the average clustering accuracy was 91% and 90%, respectively for PAMAP2 and SCAR datasets.

In addition, we evaluated the clustering performance of our method against several widely used clustering algorithms. Fig. 10 illustrates the comparison between our method and the standard unsupervised clustering algorithms for each target subject. The Kmeans algorithm requires the desired number of output clusters as an input (e.g., $k = 15$ for DSA dataset), which limits its practicality. In contrast, the DBSCAN clustering and Expectation Maximization (EM) methods automatically optimize the number of output clusters. The X-means algorithm (an extension of Kmeans method) only requires a lower-bound and an upper-bound on the number of desired clusters to find the optimal partitioning.

The subplots in Fig. 10 show the clustering accuracy for each dataset. As it can be observed, our algorithm in most cases outperforms the other approaches, in a consistent fashion. While EM and Kmeans show acceptable performance among all three datasets, their performance varies considerably from one dataset to another. For instance, EM algorithm outperforms other baselines in SCAR dataset ($+8.5\%$ on average and more than 9% less than our method). However, in DSA dataset, Kmeans outperforms other baselines by 4.8% to have the closest performance to our net-based algorithm (-6.3%).

5.4 Overall Classification Performance

In this section, we carried out a comparative analysis for OptiMapper framework. We compare the performance of our cross-subject transfer learning algorithm based on bipartite matching and many-to-one mapping (referred to as 'CSTL' and 'U-CSTL', respectively) against experimental upper-bound, and several robust models. Due to superior performance of random forest algorithm, it is used as the base supervised algorithm for base learner in our approach. We used 66%-split evaluation method to ensure the test observations are not used for training of the same model for all of our experiments. The training and test datasets are generated one time and used in our analysis for enable comparability and consistency of results.

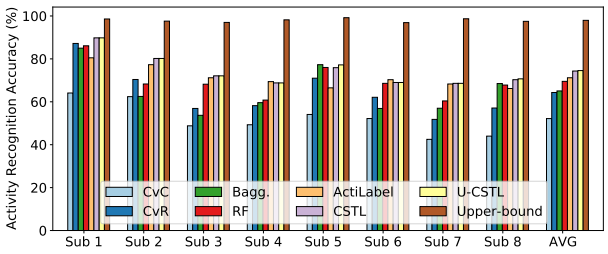
We evaluated the final activity recognition performance of CSTL and U-CSTL approaches with adaptive label fusion. The robust standard models used for this comparison are (bagging and random forest algorithms) trained on the source subjects to classify the observation in target dataset. The state-of-the-art transfer learning models are classification via clustering algorithm (CvC) [46], classification via regression (CvR) [47], and classification via community-detection algorithm and graph-level matching (ActiLabel) [21]. These approaches use clustering and regression on the test dataset to boost the robustness against data variations. Furthermore, we report the experimental upper-bound that is the classification accuracy of a 66%-split validation of the model trained on the ground-truth labels belonged to the same subject. The average performance per subject for each dataset is shown in Fig. 11.

The results demonstrate the significantly superior performance of the OptiMapper proposed algorithms. For DSA dataset, The average accuracy improvement is between 22.5% (against *CvC*) and 3.4% (against ActiLabel algorithm). Similarly, for PAMAP2 dataset, The average accuracy improvement is from 15.2% (against *CvC*) to 4.3% (against ActiLabel algorithm). Lastly, for SCAR dataset, The average accuracy improvement is between 8.5% (against bagging) and 2.3% (against *CvR*).

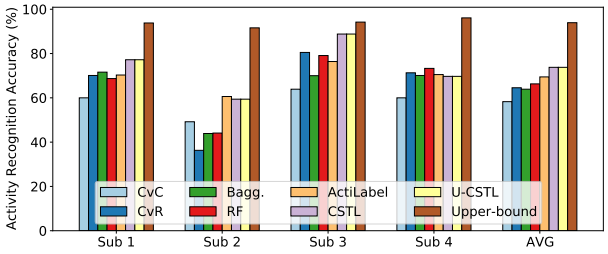
We note that because these analysis are carried out on a fully constructed target dataset where observations from all class labels are available, the performance of CSTL and U-CSTL methodologies are almost identical. The reason behind it is that the bipartite matching used in CSTL can be considered a special case of many-to-mapping where there is at least one available mapping between each label in source and target clusters.

Furthermore, based on the different performances of the algorithms over various datasets in this figure, we can observe that cross-subject transfer learning still heavily relies on similarity of the target user to the source domain as well as the ability of the clustering algorithm to separate the user's activities. While these are the reasons for variations in the performance across target users, our proposed algorithm is still outperforming in a vast majority of cases and on average as well.

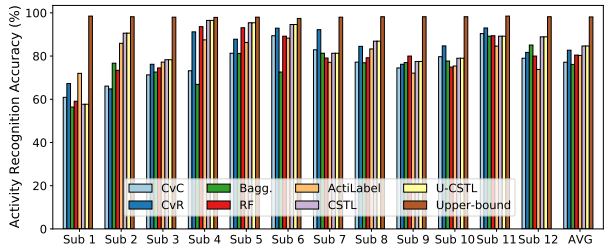
Fig. 12 shows the average activity recognition accuracy of U-CSTL and CSTL as a function of the number of labels that exist in the test dataset. Take into account that for each subset of labels of size k, a k-combination of subsets can



(a) Daily and Sports Activity (DSA) dataset

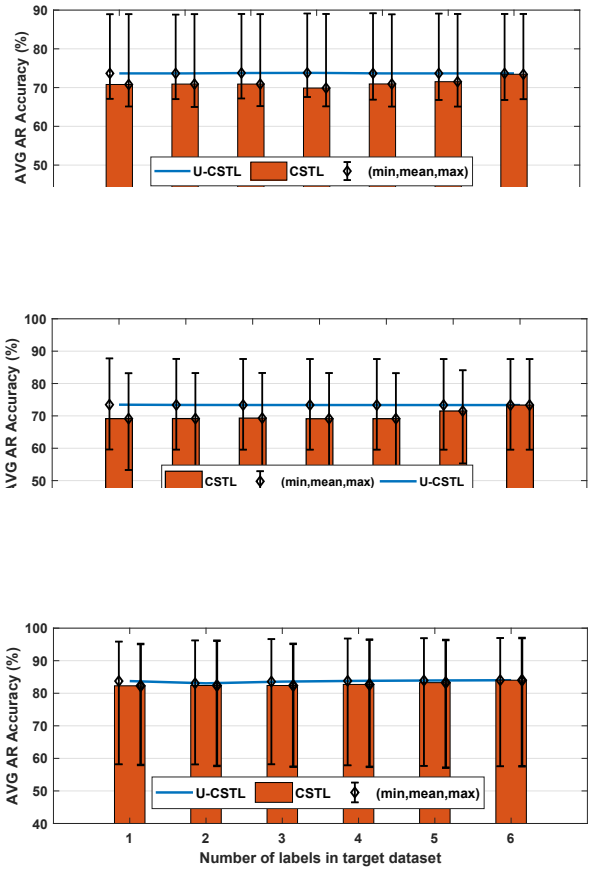


(b) PAMAP2 dataset



(c) SCAR dataset

Fig. 11. Final cross-subject accuracy of U-CSTL and CSTL methods vs baseline, computational upper-bound, and competing methods.



(c) SCAR dataset

Fig. 12. Activity Recognition accuracy vs the number of available labels in the target dataset

be found. The results for each k is the average over all the possible combinations. From this figure, a similar pattern can be observed for all three datasets. As discussed above, the performance with all the labels present (i.e., the last column in each figure) is almost the same as the mapping problem tends to become the same in both mapping solutions. However, as discussed in Section 4.1, when the target dataset includes missing labels the many-to-one mapping can avoid cases of bad label transfer. Such expected decline in the performance of CSTL approach can be observed in the figure as the number of available labels increases. An average decrease of 3.56%, 4.22%, and 1.72% is reported, respectively, for DSA, PAMAP2, and SCAR datasets. As expected, no significant accuracy decline is reported for U-CSTL. While there is a noticeable decline in such cases for CSTL, the model does not entirely fail mainly because of two reasons: (1) the clustering algorithm's superior performance does not produce many cases that could result in bad label transfer (cases where multiple clusters belong to the same class label); and (2) the adaptive fusion of base learner and the transferred labels will mask many cases of incorrect label mapping and therefore disallow the model to drastically under-perform in such special cases.

6 Conclusion

In this paper, we proposed OptiMapper, a novel framework for relational knowledge transfer for human-centered activity recognition applications. Our transfer learning approach is based on combinatorial optimization for cross-domain sensor mapping. We devised a novel knowledge transfer learning algorithm to discover the relationships between source and target domains to extract knowledge for developing a personalized machine learning model. First, we assumed that the activity sets used in source and target are identical and proposed CSTL, Cross-Subject Transfer Learning, algorithm. Then, as an extension to CSTL, we proposed a novel autonomous cross-subject transfer learning approach that resolves the main limitation of CSTL methodology, which requires sufficient observations for all predefined labels in the target dataset. This extension utilizes a robust unbalanced many-to-one mapping scheme to transfer class labels between subject-specific activity recognition datasets. This approach, referred to as U-CSTL (Unbalanced CSTL), formulates the many-to-one mapping problem into a especial case of transportation problem and presents an optimal solution. The revised mapping scheme dismisses such assumption and can robustly map the clustered data from source domain into the set of target observations associated with any subset of labels.

Through experimental evaluation, we showed the activity recognition accuracy on the target domain drops by 43%, on average, in absence a proper transfer learning approach. We also showed that adding excessively large number of source subjects cannot be considered a reliable solution for overcoming the this issue. We performed a set of comparative analyses. Our analyses on real-world datasets demonstrated the superiority of OptiMapper against several state-of-the-art and baseline solutions including classification via clustering/regression, bagging, and random forest algorithms with up to 22% increase in the classification accuracy. Furthermore, we could achieve 69% utilization of datasets for clustering and 87.5% clustering accuracy, on average.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grants CNS–1750679 and CNS–1932346. 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] M. Chan, D. Estève, J.-Y. Fourniols, C. Escriba, and E. Campo, "Smart wearable systems: Current status and future challenges," *Artificial intelligence in medicine*, vol. 56, no. 3, pp. 137–156, 2012.

[2] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial internet of things (iiot)–enabled framework for health monitoring," *Computer Networks*, vol. 101, pp. 192–202, 2016.

[3] M. Pedram, M. Rofouei, F. Francesco, Z. Esna Ashari, and H. Ghasemzadeh, "Resource-efficient computing in wearable systems," in *2019 IEEE International Conference on Smart Computing (SmartComp)*, Jun 2019.

[4] N. Hezarjaribi, Z. Esna Ashari, J. Frenzel, H. Ghasemzadeh, and S. Hemati, "Personality assessment from text for machine commonsense reasoning," *arxiv*, 2020, doi: arXiv:2004.09275.

[5] D. Zakim and M. Schwab, "Data collection as a barrier to personalized medicine," *Trends in pharmacological sciences*, vol. 36, no. 2, pp. 68–71, 2015.

[6] L. Yao, F. Nie, Q. Z. Sheng, T. Gu, X. Li, and S. Wang, "Learning from less for better: semi-supervised activity recognition via shared structure discovery," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 13–24.

[7] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowledge and information systems*, vol. 36, no. 3, pp. 537–556, 2013.

[8] R. Saeedi, H. Ghasemzadeh, and A. Gebremedhin, "Transfer learning algorithms for autonomous reconfiguration of wearable systems," in *IEEE International Conference on Big Data*. IEEE, 2016.

[9] R. Fallahzadeh and H. Ghasemzadeh, "Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data," in *Proceedings of the 8th International Conference on Cyber-Physical Systems*, 2017, pp. 293–302.

[10] Y. Chen, J. Wang, M. Huang, and H. Yu, "Cross-position activity recognition with stratified transfer learning," *Pervasive and Mobile Computing*, vol. 57, pp. 1–13, 2019.

[11] J. Wang, C. Y., S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[12] R. Ding, X. Li, L. Nie, J. Li, X. Si, D. Chu, , G. Liu, and D. Zhan, "Empirical study and improvement on deep transfer learning for human activity recognition," *Sensors*, vol. 19, no. 1, pp. 1–13, 2019.

[13] S. Arshad, C. Feng, R. Yu, and Y. Liu, "Leveraging transfer learning in multiple human activity recognition using wifi signal," in *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*. IEEE, 2019.

[14] A. Rokni and H. Ghasemzadeh, "Synchronous dynamic view learning: A framework for autonomous training of activity recognition models using wearable sensors," in *16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, Apr 2017.

[15] ——, "Autonomous training of activity recognition algorithms in mobile sensors: A transfer learning approach in context-invariant views," *IEEE Transactions on Mobile Computing*, vol. 17, no. 8, pp. 1764–1777, 2018.

[16] H. Koskimäki and P. Siirtola, "Adaptive model fusion for wearable sensors based human activity recognition," in *2016 19th International Conference on Information Fusion (FUSION)*. IEEE, 2016, pp. 1709–1713.

[17] S. A. Rokni and H. Ghasemzadeh, "Plug-n-learn: automatic learning of computational algorithms in human-centered internet-of-things applications," in *Proceedings of the 53rd Annual Design Automation Conference*. ACM, 2016, p. 139.

[18] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[19] S. A. Rokni and H. Ghasemzadeh, "Autonomous sensor-context learning in dynamic human-centered internet-of-things environments," in *Proceedings of the 35th International Conference on Computer-Aided Design*. ACM, 2016, p. 75.

[20] J. Wang, Y. Chen, H. Yu, M. Huang, and Q. Yang, "Easy transfer learning by exploiting intra-domain structures," in *Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2019, pp. 1210–1215.

[21] P. Alinia, I. Mirzadeh, and H. Ghasemzadeh, "Actilabel: A combinatorial transfer learning framework for activity recognition," *arXiv*, 2020.

[22] E. Garcia-Ceja and R. Brena, "Building personalized activity recognition models with scarce labeled data based on class similarities," in *International Conference on Ubiquitous Computing and Ambient Intelligence*. Springer, 2015, pp. 265–276.

[23] Z. S. Abdallah, M. M. Gaber, B. Srinivasan, and S. Krishnaswamy, "Streamar: incremental and active learning with evolving sensory data for activity recognition," in *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, vol. 1. IEEE, 2012, pp. 1163–1170.

[24] S. Bagaveyev and D. J. Cook, "Designing and evaluating active learning methods for activity recognition," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 469–478.

[25] H. S. Hossain, M. A. A. H. Khan, and N. Roy, "Active learning enabled activity recognition," *Pervasive and Mobile Computing*, 2016.

[26] Z. Esna Ashari and H. Ghasemzadeh, "Mindful active learning," in *2019 IJCAI conference (International Joint Conference on Artificial Intelligence)*, Aug 2019.

[27] Z. Esna Ashari, N. Chaytor, D. J. Cook, and H. Ghasemzadeh, "Memory-aware active learning in mobile sensing systems," *IEEE Transactions on Mobile Computing (IEEE-TMC)*, June 2020, doi: 10.1109/TMC.2020.3003936.

[28] Y. Ma, Z. Esna Ashari, M. Pedram, N. Amini, D. Tarquinio, K. Nouri-Mahdavi, M. Pourhomayoun, R. D. Catena, and H. Ghasemzadeh, "Cyclepro: A robust framework for domain-agnostic gait cycle detection," *IEEE Sensors Journal*, vol. 19, no. 10, pp. 3751–3762, 2019.

[29] Z. Zhao, Y. Chen, J. Liu, Z. Shen, and M. Liu, "Cross-people mobile-phone based activity recognition," in *IJCAI*, vol. 11. Citeseer, 2011, pp. 2545–250.

[30] W.-Y. Deng, Q.-H. Zheng, and Z.-M. Wang, "Cross-person activity recognition using reduced kernel extreme learning machine," *Neural Networks*, vol. 53, 2014.

[31] R. Xu and D. Wunsch, *Clsutering*. Wiley, 2008.

[32] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, 1999.

[33] R. Subhashini and V. J. S. Kumar, "Evaluating the performance of similarity measures used in document clustering and information retrieval," in *Proceedings of the 2010 International Conference on Integrated Intelligent Computing*. IEEE, 2010.

[34] L. P. Dinu and R. T. Ionescu, "A rank-based approach of cosine similarity with applications in automatic classification," in *Proceedings of the 2012 International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. IEEE, 2012.

[35] J. Dahmen, B. L. Thomas, D. J. Cook, and X. Wang, "Activity learning as a foundation for security monitoring in smart

homes," *Sensors*, vol. 17, no. 4, 2017. [Online]. Available: http://www.mdpi.com/1424-8220/17/4/737
[36] M. A. A. H. Khan and N. Roy, "Untran: Recognizing unseen activities with unlabeled data using transfer learning," in *Internet-of-Things Design and Implementation (IoTDI), 2018 IEEE/ACM Third International Conference on*. IEEE, 2018, pp. 37–47.
[37] H.-T. Cheng, F.-T. Sun, M. Griss, P. Davis, J. Li, and D. You, "Nuactiv: Recognizing unseen new activities using semantic attribute-based learning," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 361–374.
[38] L. Cooper, "The transportation-location problem," *Operations Research*, vol. 20, no. 1, pp. 94–108, 1972.
[39] S. Korukoğlu and S. Ballı, "A improved vogel's approximation method for the transportation problem," *Mathematical and Computational Applications*, vol. 16, no. 2, pp. 370–381, 2011.
[40] H. H. SHORE, "The transportation problem and the vogel approximation method," *Decision Sciences*, vol. 1, no. 3-4, pp. 441–457, 1970.
[41] A. Charnes and W. W. Cooper, "The stepping stone method of explaining linear programming calculations in transportation problems," *Management science*, vol. 1, no. 1, pp. 49–69, 1954.
[42] B. Barshan and M. C. Yüksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," *The Computer Journal*, vol. 57, no. 11, pp. 1649–1667, 2014.
[43] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Wearable Computers (ISWC), 2012 16th International Symposium on*. IEEE, 2012, pp. 108–109.
[44] H. Ghasemzadeh, N. Amini, R. Saeedi, and M. Sarrafzadeh, "Power-aware computing in wearable sensor networks: an optimal feature selection," *Mobile Computing, IEEE Transactions on*, vol. 14, no. 4, pp. 800–812, 2015.
[45] R. Fallahzadeh, M. Pedram, and H. Ghasemzadeh, "Smartsock: A wearable platform for context-aware assessment of ankle edema," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 6302–6306.
[46] M. I. Lopez, J. Luna, C. Romero, and S. Ventura, "Multi-objective optimization of dynamic memory managers using grammatical evolution," in *Proceedings of the 5th International Conference on Educational Data Mining, Chania, Greece*. The Education Resources Information Center-United States, June 2012.
[47] E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. Witten, "Using model trees for classification," *Machine Learning*, vol. 32, no. 1, pp. 63–76, 1998.

Ramin Fallahzadeh received his BS degree in computer engineering from the Sharif University of Technology, Tehran, Iran, in 2014. He received a PhD degree in computer science at Washington State University, in 2018. Currently, he is a postdoctoral scholar at Stanford University working at intersection of artificial intelligence (AI) and health. His research interests include smart-health, pervasive computing, and machine learning. The focus of his research is on design and development of AI models for wearable monitoring systems with various applications in healthcare. He is a member of the IEEE.

Zhila Esna Ashari received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, and the M.Sc. degree in electrical engineering, communications from the Sharif University of Technology, Tehran, Iran. She has received her Ph.D. degree in computer science from Washington State University, Pullman, WA, USA. Her research interests are machine learning and data mining and their applications in active learning, Internet of Things and bioinformatics.

Parastoo Alinia obtained her Ph.D. in Computer Science from Washington State University in 2020. She is currently a postdoctoral researcher at Washington State University under the supervision of Dr. Hassan Ghasemzadeh and Dr. Michael Celeveland. Dr. Alinia's research includes wearable computing, robust machine learning, and mobile health. Parastoo Alinia also holds an M.S. degree in computer science with a focus on applied machine learning from Washington State University. She received her B.Sc. degree from Sharif University of Technology in Information Technology in 2013. She is a member of IEEE.

Hassan Ghasemzadeh is an Associate Professor of Computer Science in the School of Electrical Engineering and Computer Science at Washington State University. He received the B.Sc. degree from Sharif University of Technology, Tehran, Iran, the M.Sc. from University of Tehran, Tehran, Iran, and his Ph.D. from the University of Texas at Dallas, Richardson, TX, in 1998, 2001, and 2010 respectively. The focus of his research is algorithm design, machine learning, system-level optimization, and mobile health. He is a senior member of the IEEE.