

# Recognizing New Activities with Limited Training Data

Le T. Nguyen, Ming Zeng, Patrick Tague, Joy Zhang

Carnegie Mellon University

Moffett Field, CA, USA

{le.nguyen, ming.zeng, patrick.tague, joy.zhang}@sv.cmu.edu

## ABSTRACT

Activity recognition (AR) systems are typically built to recognize a predefined set of common activities. However, these systems need to be able to learn new activities to adapt to a user's needs. Learning new activities is especially challenging in practical scenarios when a user provides only a few annotations for training an AR model. In this work, we study the problem of recognizing new activities with a limited amount of labeled training data. Due to the shortage of labeled data, small variations of the new activity will not be detected resulting in a significant degradation of the system's recall. We propose the FE-AT (Feature-based and Attribute-based learning) approach, which leverages the relationship between existing and new activities to compensate for the shortage of the labeled data. We evaluate FE-AT on three public datasets and demonstrate that it outperforms traditional AR approaches in recognizing new activities, especially when only a few training instances are available.

## Author Keywords

Activity Recognition; Attribute-based Learning; Insufficient Data Problem; Imbalanced Data Problem

## ACM Classification Keywords

I.2.1 Artificial Intelligence: Applications and Expert Systems

## INTRODUCTION

Wearable-based activity recognition (AR) systems are typically built to recognize a predefined set of common activities such as sitting, walking, and running [11]. However, to adapt to the needs of individuals and application scenarios, these AR systems often need to be extended to recognize new activities of interest. For example, people working out at a gym need the AR system to correctly distinguish between individual types of exercises, whereas applications helping users quit smoking depend on the system's ability to recognize smoking activities.

To learn new activities of interest, AR systems can ask users to label additional training data. However, it is impractical to assume that users will provide a large amount of annotations,

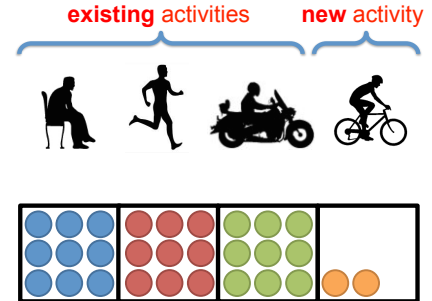


Figure 1: Limited training data problem: We have only a small amount of labeled data for the new activity of interest. However, a large amount of data of the existing activities is available.

since labeling activities is a time-consuming and laborious process [19, 18, 15]. Therefore, being able to learn new activities with a limited amount of training data is highly desirable for practical AR systems.

There are many challenges associated with learning activities with limited training data. Consider the scenario presented in Figure 1, where the AR system is built to recognize common activities such as sitting, running, and riding a motorcycle, for which a large amount of labeled data is available. We refer to the set of common activities as **existing activities**. The user is interested in extending the system to recognize biking and labels a few instances of this activity. We refer to an extended set of activities as **new activities**. Obviously, such limited training data of the new activity will have a negative impact on the AR performance. First, the AR model is prone to overfit to the few instances of biking. Thus, small variations of the biking activities will not be detected. Secondly, many biking activities will be falsely predicted as one of the existing activities (e.g., running), resulting in a degradation of the overall recognition performance [14].

To address the challenges of learning new activities, we propose the **FE-AT** (Feature- and Attribute-based learning) approach. FE-AT extends the traditional supervised/feature-based AR approach with attribute-based learning paradigm. The key idea is to consider the semantic meaning of the new activity and its relationship to the existing activities. Figure 2 shows how the biking activity relates to the existing activities: 1) similarly to sitting the user's body does not move, but 2) the legs move similarly to running and 3) the arms move similarly to riding a motorcycle. The properties of "body not

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

ISWC '15, September 07 - 11, 2015, Osaka, Japan

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3578-2/15/09\$15.00

DOI: <http://dx.doi.org/10.1145/2802083.2808388>

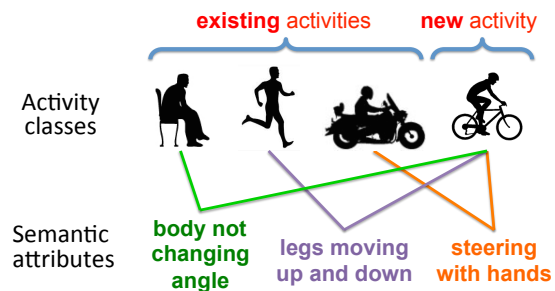


Figure 2: Exploiting the relationships between new and existing activities to improve the activity recognition performance.

changing angle”, “legs moving up and down” and “steering with hands” are so called **semantic attributes** and can be considered as primitive actions describing the activity. These semantic attributes can be used for recognizing new activities. For example, if we can detect from the sensor readings that the user is moving her legs up and down, steering with her hands and not changing the angle of her body, then she is riding a bike.

The key advantage of using semantic attributes is the fact that they can be learned from the large amount of existing activity data. Thus, we can use the attribute detectors to compensate for the shortage of labeled data of the new activity. However, the attribute-based AR assumes that each activity can be described by a unique set of attributes. In this work, we show that this assumption cannot be always fulfilled. We proposed a classifier fusion method, which combines the attribute-based learning with the traditional feature-based learning to overcome the non-uniqueness problem.

Our contributions are summarized as follow.

- **Limited training data:** We study the problem of learning new activities and identify challenges associated with learning with limited training data. We further show how these challenges negatively impact the performance of traditional AR models.
- **Classifier fusion:** To address the challenges, we propose the FE-AT (Feature-based and Attribute-based learning) approach. FE-AT leverages the semantic meaning of new activities to compensate for the shortage of labeled data.
- **Empirical study:** We evaluate FE-AT on three public datasets. We show the weaknesses of both feature-based and attribute-based learning and demonstrate how FE-AT overcomes these weaknesses. Furthermore, we identify limitations of the proposed approach and discuss the open challenges.

## RELATED WORK

An obvious solution for the limited training data problem is to acquire more labeled data. Many approaches based on active learning have been proposed to reduce the required number of annotations by asking users to label only the most informative instances [19, 15]. Additionally, many annotation strategies

such as coarse labeling were proposed to reduce the annotation overhead [18].

In this work, we address the limited training data problem by exploiting the relationships between activities and their semantics. Recent work has explored how knowledge about activities can be used for AR. An ontology-based inference technique was proposed to infer the high-level activities from atomic actions [8]. The proposed approach makes use of an extensive ontology defined by domain experts. In this work, we assume that this information is not necessarily available. We use a simple activity-attribute representation, which can be annotated using common sense knowledge by answering yes/no questions. Moreover, such annotation can be acquired through crowd-sourcing or automatic text analysis techniques [17].

The attribute-based learning approach used in this work can be considered as a transfer learning approach, which leverages the knowledge of activities in a source domain to improve the recognition of activities in a target domain. In AR, many approaches have been proposed to transfer knowledge between settings with different space, time, people, sensor types or activities [6]. Transferring knowledge between activities was explored in the scenario of recognizing composite activities by learning classifiers for low-level activities and exploiting the relationship between these low-level activities and the composite activities [4]. Our work is related to the work of Hu et al. [9], which aims at transferring knowledge between activities by measuring their similarities. The authors represent the similarity between activities by a single value corresponding to similarity of web documents describing these activities. In this work, we break down such score into a set of human-interpretable attributes, which additionally allow us to represent more complex relationships between activities (e.g., biking is similar to riding a motorcycle with respect to steering with hands but not with respect to moving legs up and down).

Attribute-based learning is often associated with zero-shot learning (ZSL) [16, 10], which aims at recognizing new classes without any training data. In the AR domain, ZSL was explored in both recognizing activities from videos [12, 17] and using wearable sensing [5]. Good results were reported when evaluating ZSL’s capability of recognizing new activities. However, as we show in this work, ZSL performs poorly when it is used to recognize both existing and new activities. The poor performance is due to ZSL’s assumption that each activity can be represented by a unique set of attributes. As we discuss in this work, this assumption does not always hold. We relax this assumption by fusing the attribute-based learning with the traditional feature-based learning approach to allow the AR system to differentiate between activities, which are represented by the same attribute set.

## RECOGNIZING NEW ACTIVITIES WITH LIMITED TRAINING DATA

There are two key challenges associated with learning new activities with limited amount of data: imbalanced data and insufficient data. In the following, we first give a basic intu-

ition for these two challenges. To address these challenges, we propose FE-AT, which employs random sampling and attribute-based learning to overcome the weaknesses of the traditional feature-based model.

### Challenges of Feature-based Learning

In this work, we use the term feature-based learning (or in short **FE**) to refer to the traditional AR approaches, which extract statistical features from the sensor readings and use supervised learning techniques to map these features into an activity label [11]. These approaches suffer from two problems when learning to recognize new activities from a small training dataset:

- **Imbalanced data problem:** The amount of training data of the new activity is significantly smaller than the amount of training data of the existing activities.
- **Insufficient data problem:** The amount of training data of the new activity is too small.

Both of these problems result in poor AR performance. The poor performance caused by the imbalanced data problem can be explained by the fact that FE techniques are designed to minimize the total classification error of the training dataset [7]. The total error is composed of the errors made by misclassifying individual instances. Thus, optimizing the classifier towards performing well on frequent activities (i.e., activities with a large number of training instances) results in a lower error rate. Since, the new activity class has significantly fewer training instances than the existing activities, the instances of the new activity are likely to be ignored during the optimization process resulting in poor recognition performance for the activity.

The poor performance caused by the insufficient data problem can be explained by the fact that the small amount of training data of the new activity is not representative for the whole new activity class. Thus, the AR system can only recognize instances that are very similar to the training dataset. Therefore, small variations of the new activity will not be correctly detected, resulting in a significant degradation of the classifier’s recall. Furthermore, the instances not detected as the new activity will be falsely predicted as one of the existing activities resulting in degradation of the overall AR performance.

### FE-AT: Feature-based and Attribute-based Learning

The goal in this work is to address the imbalanced and insufficient data problems described above.

*Addressing the imbalanced data problem:* To address the imbalanced data problem, we use random oversampling [7], which randomly selects instances of the new activity and replicates them to create a balanced training dataset. This technique artificially increases the amount of training instances of the new activity. Thus, the new activity class is treated similarly as the other existing frequent activity classes in the optimization process. On the other hand, random oversampling will make the classifier be more prone to overfitting. Since a large amount of duplicates of the few new activity instances are generated, the resulting system will likely overfit

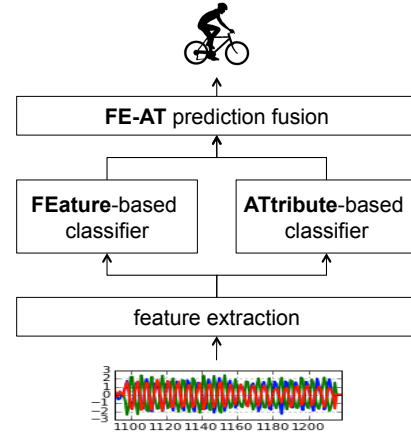


Figure 3: FE-AT fuses the feature-based and attribute-based models to address to challenges of learning new activities.

to these instances and reject any small variations of the new activity [7]. Thus, oversampling is not a sufficient solution for the problem of recognizing new activities.

*Addressing the insufficient data problem:* To address the insufficient data problem of the FE approach, we fuse it with an attribute-based learning approach. The key idea is to leverage the semantic meaning of the new activity class (captured by the attribute-based learning approach) to overcome the fact that only a small amount of labeled data is available. This semantic meaning of the new activity is represented through its relationship with the existing activities.

Figure 3 shows the components of the proposed FE-AT (Feature-based and Attribute-based learning) fusion approach. In FE-AT, we first extract features from wearable sensor measurements. Similarly to the traditional AR systems [11], we extract features using a sliding window with 50% overlap. For each window of sensor reading we extract statistical features including minimum, maximum, average, and correlation between sensor axes [11]. The feature vector is then input into both the feature-based and attribute-based models. The models’ outputs are then fused together through the FE-AT fusion component.

In the following, we first describe the key intuition behind the attribute-based approach and discuss its weaknesses. Then we explain how FE-AT fuses both the feature-based and attribute-based approaches to overcome the problems associated with using each approach individually.

### Attribute-based Learning

Attribute-based learning (further referred to as **AT**) uses an intermediate layer of semantic attributes to represent and recognize human activities. In the following, we explain how the attributes are represented, how they are used in AR and discuss the weaknesses of AT model.

*AT process:* The AT prediction process is composed of two steps as shown in Figure 4. In the first step, the attribute detectors are used to identify occurrences of attributes in the signal represented by the feature vector. The number of attribute

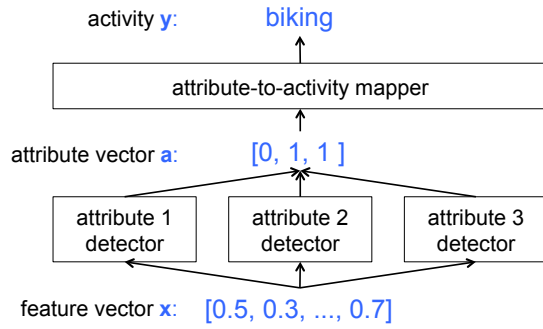


Figure 4: Prediction process of AT: 1) a feature vector is mapped into an attribute vector and 2) the vector is used to predict the activity label.

detectors equals the number of attributes. Each attribute detector outputs 1 if the corresponding attribute occurs in the signal and outputs 0 otherwise. The outputs of all attribute detectors are used to build an attribute vector. In the second step, the inferred attribute vector is mapped into an activity.

*Features vs. attributes vs. activities:* Instead of directly mapping a feature vector to an activity (as in FE), AT leverages an intermediate level of attributes for an indirect mapping. An attribute can be considered as a high-level feature, which has a semantic meaning as opposed to the low-level features such as a mean or a variance of the sensor reading. On the other hand, an attribute can be also considered as a low-level activity (e.g., “steering with hands”). As shown in Figure 2, each activity is associated with a set of attributes and each attribute is associated with a set of activities. The attributes explicitly model the relationship between activities (e.g., “riding motorcycle” is similar to “biking” with respect to the “steering with hands”).

*Recognizing new activities using attributes:* In the scenario of recognizing new activities, the attributes are used to compensate for the shortage of the activity labels of new activities. The idea is to learn attribute detectors on the large amount of training data of existing activities. For example, we can learn a detector to recognize the “steering with hands” attribute from the large amount of riding a motorcycle activities. This detector can then be used to recognize not only riding a motorcycle activities but also biking activities, for which only limited amount of training data is available.

*Activity-attribute matrix:* Using attributes for AR assumes the availability of the activity-attribute associations. This information can be represented as an activity-attribute matrix shown in Figure 5. Each row corresponds to an attribute vector representing an activity. For example, the “standing” activity is represented by attributes “arm still”, “leg still”, “body still” and “body not changing angle”. The attribute-to-activity mapper (shown in Figure 4) uses the this matrix to infer the activity. The activity inference is done by finding an activity represented by an attribute vector, which is the most similar to the output of the attribute detectors.

*Obtaining activity-attribute matrix:* In the AR domain, one can use common sense knowledge to define attributes as

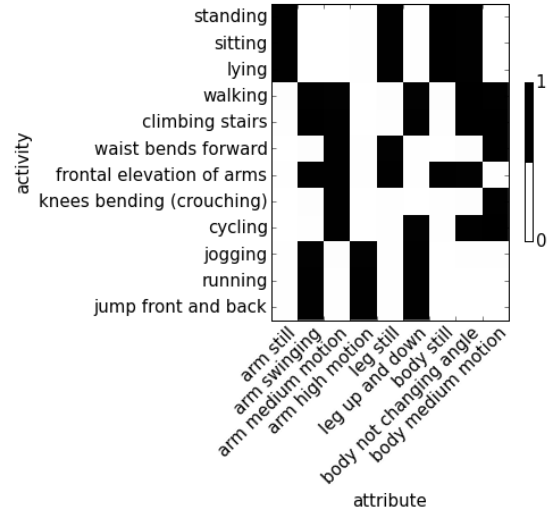


Figure 5: Activity-attribute matrix: The relationship between activities are represented by the attribute vectors (the rows of the matrix).

primitive actions performed in the activity (e.g., defining motions of the arms, legs and the body as the attributes to describe exercise activities). Furthermore, one can reuse a large amount of activity-attribute annotations defined in the computer vision domain, which use AT to recognize human activities from videos [12]. Activity-attribute annotations can be also obtained through automatic text extraction of activity descriptions [17]. The cost associated with obtaining the activity-attribute annotations will be further discussed in the next section.

*Weakness of AT:* One of the key factors influencing the AT performance is the uniqueness of the activity-attribute representations. Suppose two activities are represented by exactly the same attribute vector. Even if all attributes are correctly detected, the activity-to-attribute mapper cannot differentiate between these two activities. For example, the activities “sitting” and “lying” shown in Figure 5 are both represented by the same attribute vector and therefore cannot be distinguished. One could suggest adding additional discriminative attributes such as an attribute describing the absolute angle of the body, which can help differentiate between “sitting” and “lying”. To learn a detector to recognize this attribute, one would need to assume having knowledge about how the sensors are mounted on the body. In cases of using mobile phones as a sensing device, such information is often not easily obtainable. The described non-unique attribute representation can cause a significant degradation of the AT performance. Obviously, even if all attributes are correctly detected, the AT would not have sufficient information to distinguish between activities with the same attribute representation.

### Fusion of FE and AT

The key contribution of this work is the proposed fusion of FE and AT models. The basic intuition of the fusion is based on the empirical observation (further discussed in the Evaluation



section) that FE models tend to underestimate the posterior probability of the new activity  $P_{FE}(y_{new}|x)$ . This is caused by the fact that the few training instances of the new activity cover only a small feature subspace compared to the true feature space actually occupied by whole new activity. Thus, the probability estimate of FE tend to be lower than the true probability estimate, resulting in many instances of the new activity not being correctly detected.

To address the biased probability estimates we propose a scoring function  $f_{FE-AT}(y|x)$ , which combines the FE predictions with the output of the AT model. Since AT’s capability of recognizing new activities is less affected by insufficient training data, the AT model provides an orthogonal way of detecting new activities. FE-AT fuses the predictions of FE and AT in the following manner:

$$f_{FE-AT}(y|x) = \begin{cases} P_{FE}(y|x) + P_{AT}(y|x) & \text{if } y = y_{new} \\ P_{FE}(y|x) & \text{otherwise} \end{cases} \quad (1)$$

where  $P_{FE}(y|x)$  and  $P_{AT}(y|x)$  are the probabilistic outputs of the FE and AT models. The most likely activity class of  $x$  is the one with highest score:

$$\hat{y} = \arg \max_y f(y|x) \quad (2)$$

Based on this formulation, the score  $f(y|x)$  by default equals the probability output  $P_{FE}(y|x)$  of an FE model. As mentioned above, FE tends to underestimate the posterior probability of the new class  $P_{FE}(y_{new}|x)$  due to the insufficient training data. Therefore, the proposed fusion method increases the score for the new activity through the AT’s prediction, which can be computed using the following probabilistic formulation [10]:

$$P_{AT}(y|x) = \sum_a P(y, a|x) \quad (3)$$

$$= \sum_a P(y|a)P(a|x) \quad (4)$$

where the goal is to estimate the probability of a feature vector  $x$  belonging to an activity class  $y$  by representing  $x$  through an attribute vector  $a$ .  $P(a|x)$  can be decomposed into probabilistic output of individual attribute detectors:

$$P(a|x) = \prod_i P(a_i|x) \quad (5)$$

$P(y|a)$  corresponds to the probabilistic output of the attribute-to-activity mapper and can be further transformed:

$$P(y|a) = \frac{P(y)}{P(a)}P(a|y) \quad (6)$$

$$= \frac{P(y)}{P(a)}\mathbf{1}[a = a^y] \quad (7)$$

where  $a^y$  is the attribute representation of  $y$  obtained from the activity-attribute matrix and  $\mathbf{1}[C]$  is an indicator function where  $\mathbf{1}[C] = 1$  if the condition  $C$  is true and 0 otherwise.

Thus, Equation 4 can be simplified as:

$$P_{AT}(y|x) = \frac{P(y)}{P(a^y)} \cdot \prod_i P(a_i^y|x) \quad (8)$$

which is divided into two factors: the first factor captures the ambiguity of the attribute representation  $a^y$  of the activity class  $y$  and the second factor captures the probability of  $x$  being mapped into such attribute representation.  $P_{AT}(y|x)$  is high if  $a^y$  is a likely representation of  $x$  and  $a^y$  is unique (i.e., there are no two activity classes represented by the same  $a^y$ ).

As we will show in the next section, due to the limited training data, FE models suffer from low recall with respect to the new activities, i.e., many new activities remain undetected by the FE models. Our proposed FE-AT fusion aims at addressing this problem by additionally leveraging the AT model for the new activity detection. Specifically, if AT detects that  $x$  is likely to belong to the new activity, it increases the score for the new activity (Equation 1). The amount of score increase is controlled by both the ambiguity of attribute representation and likelihood of attribute detection (Equation 8). Thus, a new activity is detected if 1) FE outputs high  $P_{FE}(y_{new}|x)$ , 2) AT outputs high  $P_{AT}(y_{new}|x)$  or 3) the sum of  $P_{FE}(y_{new}|x)$  and  $P_{AT}(y_{new}|x)$  is high even through the individual posterior probabilities are low. The last case occurs when a new activity is not sufficiently similar to the limited training data to be detected by FE and its attribute representation is not unique to be deterministically detected by the AT model. As we show in the next section, the proposed fusion method outperforms traditional classifier fusion methods [13], which do not consider the unique challenges associated with learning new activities.

## EVALUATION

In this section, we evaluate FE-AT on three public datasets and study how well it can recognize new and existing activities. First, we compare the performance of FE-AT with other models including the FE, AT and other traditional classifier fusion methods. Then, we study how FE-AT performs under different amounts of available training data. Finally, we point out the weaknesses of FE and AT models and discuss the trade-offs made by FE-AT.

### Data and Setup

We use the public Mhealth [3], DailyAndSport [1], and RealDisp [2] datasets, containing data collected from 10, 8 and 17 users performing 12, 19 and 33 activities, respectively. The performed activities include basic daily activities (e.g., sitting, walking) and exercise activities (e.g., exercising on elliptical bike, playing basketball).

We use a sliding window of five seconds with 50% overlap to extract statistical features as described in the previous section. For all three datasets we use a set of nine attributes to define the activity-attribute matrices. The attribute-activity matrix of the MHealth dataset is shown in Figure 5.

We extend the traditional leave-one-user-out cross validation in the following manner. Let  $U$  denote the number of users and  $A$  the number of activities. In each cross validation iteration, we use data from  $(U - 1)$  users for training and one user

for testing. Further, in the training data, we select one activity as a new activity and the remaining  $(A - 1)$  as existing activities. We use all training instances for existing activities but only  $N$  randomly selected training instances for the new activity. Due to randomness in our training process, we repeat each experiment five times, totalling  $U \cdot A \cdot 5$  cross validation iterations for each configuration.

The final result is averaged across all cross validation iterations. We evaluate using two metrics: 1) F1 score of the new activity and 2) F1 score averaged across all existing activities. F1 score is a harmonic mean of precision and recall:

$$F1 = \frac{2 \cdot \textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}$$

$$\textit{precision} = \frac{TP}{TP + FP} \quad \textit{recall} = \frac{TP}{TP + FN}$$

where TP, FP and FN denote counts of True Positives, False Positives and False Negatives.

As the base classifier for FE, AT and FE-AT, we experimented with a range of supervised classifiers including k-NN, SVM, Random Forest. Random Forest with 100 trees consistently achieved the best results. Therefore, we will report the results achieved using this base classifier.

### Classifier Fusion Models

Besides FE, AT and FE-AT, we also evaluate traditional classifier fusion models. Let  $P_{FE}$  and  $P_{AT}$  denote the probability outputs of FE and AT. In the following, we evaluate these four classifier fusion models [13]:

$$f_{MAX}(y|x) = \max(P_{FE}(y|x), P_{AT}(y|x)) \quad (9)$$

$$f_{MIN}(y|x) = \min(P_{FE}(y|x), P_{AT}(y|x)) \quad (10)$$

$$f_{PRODUCT}(y|x) = P_{FE}(y|x) \cdot P_{AT}(y|x) \quad (11)$$

$$f_{SUM}(y|x) = P_{FE}(y|x) + P_{AT}(y|x) \quad (12)$$

Figure 6 shows the performance of the prediction models on the MHealth dataset using only  $N = 5$  training instances of the new activity. We can observe that AT performs poorly both for existing and the new activities due to the non-unique attribute representations. With respect to the existing activities, all except AT achieve high average F1 score.

On the other hand, FE-AT outperforms other models when recognizing new activities. The traditional fusion models, which simply fuse probability predictions of all activities, do not consider the unique challenges associated with recognizing new activities with insufficient data. Thus, we can observe that these models achieve a performance comparable to the FE model. This observation is consistent throughout our evaluation and, therefore, we omit the discussion of these traditional fusion models in the following experiments.

Figure 7 breaks down the F1 score of recognizing new activities from the above experiment into precision and recall. We can observe that FE-AT improves its recognition performance by significantly increasing the recall. This is achieved by incorporating the attribute detection into the scoring function.

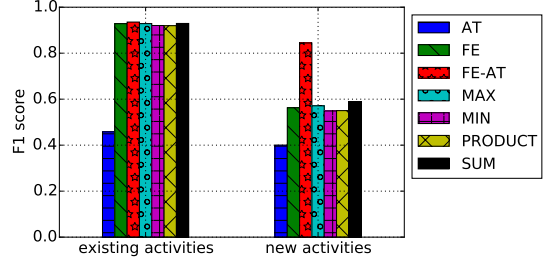


Figure 6: FE-AT significantly outperform other approaches at recognizing new activities. (Configuration: MHealth,  $N = 5$ )

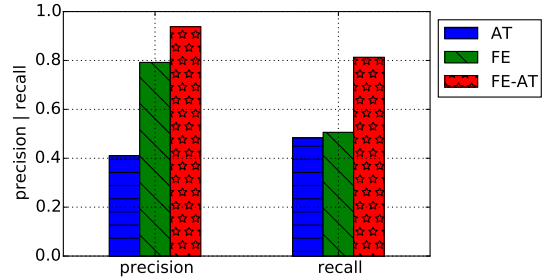


Figure 7: FE-AT improves the performance of recognizing new activities by substantially increasing the its recall over the FE approach. (Configuration: MHealth,  $N = 5$ )

### Increasing Amount of Training Data

**New activities:** In the following, we evaluate how the prediction models perform when we obtain more training data for the new activity. Figure 8 shows the F1 score for new activity recognition with increasing  $N$ . For all three datasets, we observe that FE-AT significantly outperforms FE and AT for small  $N$ . As expected, the performance of FE improves with the increasing amount of training data. Thus, the improvement of FE-AT over FE decreases with increasing  $N$ .

**Existing activities:** Figure 9 shows the F1 score for the MHealth dataset. Increasing  $N$  has no significant impact on recognizing existing activities and FE-AT achieves a comparable performance to FE. We observe the same trend for the DailyAndSport and RealDisp datasets. In the following, we discuss our results and findings for the MHealth dataset.

### Weaknesses of FE and AT

To better understand the impact of the FE-AT classifier fusion method, we further study the weaknesses of the individual FE and AT models.

**FE:** Figure 10 shows the precision and recall of FE when recognizing new activities. We can observe that the precision is consistently higher than the recall. The high precision can be explained as the effect of having only a small number of training instances of the new activity. Due to the small  $N$ , at the prediction time only instances very similar to training instances will be recognized as the new activity. Thus, only predictions with high confidence will be accepted, resulting in low false positive rate and therefore in high precision.

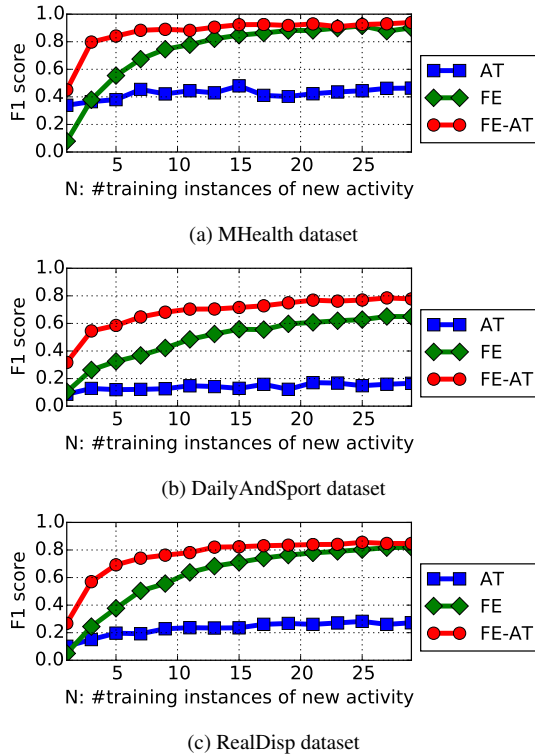


Figure 8: Recognition of **new** activities: FE-AT outperforms FE and AT when only a few training instances are available. The performance of FE increases as more training data is available resulting in an decreasing improvement of FE-AT over FE.

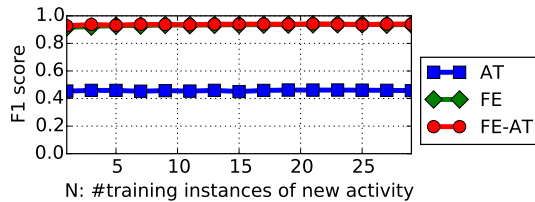


Figure 9: Recognition of **existing** activities: FE-AT achieves a comparable performance as FE, when recognizing existing activities of the MHealth dataset (the FE and FE-AT curves are overlapping). Similar results are observed for the DailyAndSport and RealDisp datasets.

On the other hand, small  $N$  will also result in low recall. Since the training data covers only a few instances of the new activity, at the prediction time many small variations of the new activities will not be recognized. This causes a high number of false negatives resulting in low recall. To overcome this problem, FE-AT uses the attribute detectors to increase the recall of the system.

**AT:** The performance of AT depends on two factors: 1) performance of attribute detection and 2) uniqueness of the attribute vector representation. Figure 11 compares the performance of attribute detection with the performance of

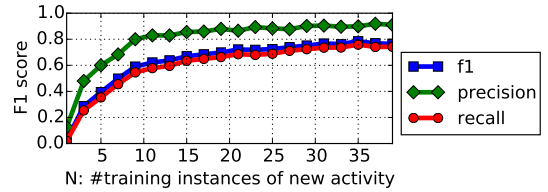


Figure 10: FE achieves high precision, but low recall when recognizing new activities.

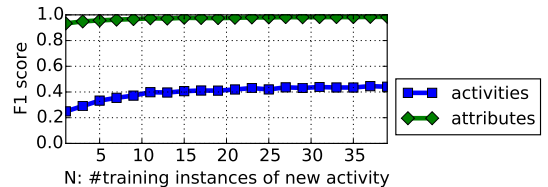


Figure 11: Even though the performance of the attribute detectors are high, AT achieves a poor activity recognition performance due to the non-unique attribute representations of the activities.

AT. Even though the attribute detection is near perfect, we observe a poor AT performance. This can be explained through non-unique attribute vector representation. From the attribute-activity matrix shown in Figure 5 we can observe that many activities are represented by the same attribute vector (e.g., “standing”, “sitting” and “lying”). Thus, even if all attributes are correctly predicted, AT cannot differentiate between these activities resulting in a poor performance.

## Discussion

**Annotation cost:** Since FE-AT extends FE by employing the attribute detectors, we need to consider the additional overhead associated with obtaining the activity-attribute annotation and its trade-off with respect to the activity annotation. In a traditional activity annotation process, the cost is associated with the number of activity instances performed and annotated by the user. In the activity-attribute annotation process, the cost depends on the number of attributes  $M$ . For each new activity class,  $M$  yes/no questions (e.g., “When performing activity X, do you use your hand for steering?”) needs to be answered. Thus, the overall cost depends on the actual overhead of answering yes/no questions versus the overhead of performing the new activity and labeling it. Note that activity-attribute annotations can be obtained without having anyone perform the activities or collect any sensor data. Thus, they can be theoretically obtained through on-line crowd-sourcing. In future work, we will further study the cost associated with obtaining different type of annotations and explore the feasibility of obtaining activity-attribute annotations using crowd-sourcing.

**Attribute representation:** The attribute representation used in this work allows attributes values to be only 0 or 1. This corresponds to an attribute either occurring the whole time (for the whole signal window) or not occurring at all. This kind of representation is suitable for homogeneous activities

such as locomotion or exercise activities. However, to represent high-level activities, a more flexible attribute representation is needed. In the future work, we will explore soft value representations, which can better model the heterogeneous nature of the high-level activities. Furthermore, the soft values can be used to capture the probabilistic nature of the attributes in cases when the activity-attribute annotations are obtained through the error-prone process of crowd-sourcing.

## CONCLUSION

In this work, we studied the problem of learning to recognize new activities, for which only limited training data is available. We propose FE-AT, an approach fusing a feature-based learning with the attribute-based learning paradigm. The proposed approach extends the feature-based approach with random sampling to address the imbalanced data problem. Integration with attribute-based learning allows the system to exploit the relationship between the existing and new activities to compensate for the shortage of labeled data. Through evaluation on three public datasets, we show the effectiveness of the proposed approach.

## ACKNOWLEDGEMENT

This research is supported in part by the National Science Foundation under the award 1346066: SCH: INT: Collaborative Research: FITTLE+: Theory and Models for Smartphone Ecological Momentary Intervention.

## REFERENCES

1. Altun, K., Barshan, B., and Tunçel, O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition* 43, 10 (2010), 3605–3620.
2. Baños, O., Damas, M., Pomares, H., Rojas, I., Tóth, M. A., and Amft, O. A benchmark dataset to evaluate sensor displacement in activity recognition. In *ACM Conference on Ubiquitous Computing*, ACM (2012), 1026–1035.
3. Banos, O., Garcia, R., Holgado-Terriza, J. A., Damas, M., Pomares, H., Rojas, I., Saez, A., and Villalonga, C. mhealthdroid: A novel framework for agile development of mobile health applications. In *Ambient Assisted Living and Daily Activities*. Springer, 2014, 91–98.
4. Blanke, U., and Schiele, B. Remember and transfer what you have learned-recognizing composite activities based on activity spotting. In *International Symposium on Wearable Computers (ISWC)*, IEEE (2010), 1–8.
5. Cheng, H.-T., Griss, M., Davis, P., Li, J., and You, D. Towards zero-shot learning for human activity recognition using semantic attribute sequence model. In *ACM international joint conference on Pervasive and ubiquitous computing*, ACM (2013), 355–358.
6. Cook, D., Feuz, K. D., and Krishnan, N. C. Transfer learning for activity recognition: A survey. *Knowledge and information systems* 36, 3 (2013), 537–556.
7. He, H., and Garcia, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* 21, 9 (2009), 1263–1284.
8. Helaoui, R., Riboni, D., and Stuckenschmidt, H. A probabilistic ontological framework for the recognition of multilevel human activities. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, UbiComp '13, ACM (New York, NY, USA, 2013), 345–354.
9. Hu, D. H., Zheng, V. W., and Yang, Q. Cross-domain activity recognition via transfer learning. *Pervasive and Mobile Computing* 7, 3 (2011), 344–358.
10. Lampert, C. H., Nickisch, H., and Harmeling, S. Learning to detect unseen object classes by between-class attribute transfer. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, IEEE (2009), 951–958.
11. Lara, O. D., and Labrador, M. A. A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials, IEEE* 15, 3 (2013), 1192–1209.
12. Liu, J., Kuipers, B., and Savarese, S. Recognizing human actions by attributes. In *Computer Vision and Pattern Recognition (CVPR)*, IEEE (2011), 3337–3344.
13. Mangai, U. G., Samanta, S., Das, S., and Chowdhury, P. R. A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Technical review* 27, 4 (2010), 293–307.
14. Nguyen, L. T., Zeng, M., Tague, P., and Zhang, J. I did not smoke 100 cigarettes today! avoiding false positives in real-world activity recognition. In *International joint conference on Pervasive and ubiquitous computing (UbiComp)*, ACM (2015).
15. Nguyen, L. T., Zeng, M., Tague, P., and Zhang, J. Superad: Supervised activity discovery. In *International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ACM (2015).
16. Palatucci, M., Pomerleau, D., Hinton, G. E., and Mitchell, T. M. Zero-shot learning with semantic output codes. In *Advances in neural information processing systems* (2009), 1410–1418.
17. Rohrbach, M., Regneri, M., Andriluka, M., Amin, S., Pinkal, M., and Schiele, B. Script data for attribute-based recognition of composite activities. In *Computer Vision—ECCV 2012*. Springer, 2012, 144–157.
18. Stikic, M., and Schiele, B. Activity recognition from sparsely labeled data using multi-instance learning. In *Location and Context Awareness*. Springer, 2009, 156–173.
19. Stikic, M., Van Laerhoven, K., and Schiele, B. Exploring semi-supervised and active learning for activity recognition. In *IEEE International Symposium on Wearable Computers*, IEEE (2008), 81–88.