Zero-shot learning with matching networks for open-ended human activity recognition.

WIJEKOON, A., WIRATUNGA, N. and SANI, S.

2018

Zero-Shot Learning with Matching Networks for Open-Ended Human Activity Recognition

Anjana Wijekoon[1][0000−0003−3848−3100] ✉, Nirmalie Wiratunga[1][0000−0003−4040−2496], and Sadiq Sani[1][0000−0001−9784−8398]

School of Computing Science and Digital Media, Robert Gordon University,
Aberdeen AB10 7GJ, Scotland, UK
{a.wijekoon, n.wiratunga, s.a.sani}@rgu.ac.uk

Abstract. A real-world solution for Human Activity Recognition (HAR) should cover a variety of activities. However training a model to cover each and every possible activity is not practical. Instead we need a solution that can adapt its learning to unseen activities; referred to as open-ended HAR. Recent advancements in deep learning have increasingly begun to focus on the need to learn from few examples, referred to as k-shot learning and to go beyond this to transfer that learning to situations with unseen classes, referred to as zero-shot learning. The latter is particularly relevant to our research in open-ended HAR; and as yet remains unexplored. This paper presents our preliminary work with Zero-shot Learning (ZSL) with a Matching Network to address open-ended HAR. A Matching Network has the desirable property of learning with few examples and so is well suited to explorations in ZSL. We evaluate Matching Networks for ZSL with a HAR dataset. We propose the use of a variable length support set at test time to overcome the search for the best support set combination that currently plagues the fixed length support set size used by matching nets. Our results show that the variable length approach to be an effective strategy to maintain accuracy whilst avoiding the combinatorial search for the best class combination to form the support set.

Keywords: Zero-shot learning · Matching Networks · Open-ended HAR

1 Introduction

Activity monitoring with wearable sensors is a popular digital health intervention incorporated in many health and well-being mobile applications. A study conducted in 2015 concluded that 58% of smart phone users in US downloaded health-care fitness applications on their mobile, and 47% of those who downloaded, stopped using them due to the high burden of data entry and loss of interest [3]. Current fitness applications are restricted only to identifying a few pre-defined activities automatically and they rely on user input for other activities. A sustainable HAR applications should be able to learn new activities over time with little user calibration. Accordingly researchers have recognised the need for Open-ended Human Activity Recognition (HAR).

An open-ended HAR solution should have the ability to recognise activities without prior knowledge about them. Existing open-ended HAR methodologies look at different unsupervised activity discovery techniques such as on-line clustering [2]. But these techniques are implemented subject to various assumptions. For instance motif discovery algorithms for detecting activities depends on finding patterns in data [1]; whilst the on-line clustering approach used by [2] is based on the assumption that activities appear only once and not periodically. Such pre-requisites restrict the deployment of these methodologies in real-time HAR applications.

Ability to incorporate personal traits when expanding beyond ambulatory activities to exercises and sports improves performance and user acceptance. Recent work in personalisation with Matching Networks has shown promising results for personalised HAR [7]. Matching Network was introduced for k-shot learning [8] where the network is trained to recognise a new class from just a small(k) number of examples of that class. K-shot learning is advantageous in HAR since it minimises the need for extensive data collection and labelling. Accordingly we explore Matching Networks for open-ended HAR in the context of Zero-shot Learning (ZSL). In literature ZSL is defined as recognising new categories of instances without training examples, by providing a high-level description of the new categories that relate them to categories previously learned by the machine [4]. In the context of Matching networks we simply define ZSL as classification of classes not seen during training.

This paper presents preliminary work on defining ZSL for the HAR domain, reporting results with the SelfBACK dataset[1]. We expose the challenges of adopting ZSL with matching networks and propose the use of a variable length support set as a new test approach to perform efficient ZSL. A comparative study demonstrates the utility of this proposal.

The rest of this paper is organised as follows: in Section 2, we will formalise ZSL with Matching Networks in the domain of HAR, and in Section 3 we present a description of our dataset and experimental design. Results are presented in Section 4 followed by conclusions in Section 5.

2 Zero Shot Learning with Matching Networks

Matching Networks can be viewed as an end-to-end neural implementation of the otherwise static kNN algorithm. The network iteratively learns to match a given target instance to a small set of instances called a support set [8]. An attention mechanism in the form of a, cosine similarity weighted majority vote, is used to inform the loss function which is driven by the difference between the estimated and actual class distributions (quantified using cross entropy). This ensures that the network eventually learns an embedding that is best placed

[1] The SelfBACK project is funded by European Union's H2020 research and innovation programme under grant agreement No. 689043. More details available: http://www.selfback.eu. The dataset associated with this paper is publicly accessible from https://github.com/selfback/activity-recognition

to identify matches between the target and the support set instances. In other words the network learns to match. Unlike conventional machine learning, here an instance comprises both the target instance and is its support set.

More formally a matching network predicts label $\hat{y}$ for an example $\hat{x}$ guided by a support set S. Lets consider a dataset with set of $\mathcal{X}$ activity instances belonging to set of $\mathcal{L}$ activity classes. We define a support set S as in Equation 1. Cardinality of the support set is $k \times n_{tr}$, where k is the number of instances per class. n_{tr} is the number of classes in support set and $n_{tr} \leq |\mathcal{L}|$.

$$S = \{(x, y)|x \in \mathcal{X}, y \in \mathcal{L}\} \tag{1}$$

The matching net's training set, $\{(t_1, S_1), (t_2, S_2), \ldots, (t_N, S_N)\}$, has N elements, where each target instance, t_i, is a training instance pair, (x, y), and is never featured in its' support set, S_i (i.e. $t_i \notin S_i$). A Matching Networks classifier model, θ, learns to recognise the activity class, y, for a given target instance, x, relative to support set, S.

At deployment, the user records a set of instances, $\hat{\mathcal{X}}$, for a set of, $\hat{\mathcal{L}}$, activity classes. We can view this as the user providing a small set of instances for model calibration. We predict the label, $\hat{y}$, for test activity instance, $\hat{x}$, relative to a support set, $\hat{S}$.

$$\theta(\hat{x}, \hat{S}) \rightarrow \hat{y}$$
$$\hat{S} = \{(x, y)|x \in (\mathcal{X} \cup \hat{\mathcal{X}}), y \in (\mathcal{L} \cup \hat{\mathcal{L}})\} \tag{2}$$

Accordingly we define, $\hat{S}$, as in Equation 2. Cardinality of set $\hat{S}$ is $k \times n_{te}$, where n_{te} is the number of classes in the support set at deployment. With one shot learning [8], n_{te} is restricted to the size of the training support set, $(n_{te} = n_{tr})$. With ZSL this forces the network to select a subset of classes from both training classes($\mathcal{L}$) and test classes($\hat{\mathcal{L}}$). This has the undesirable property that the set of possible combinations, grows exponentially with increasing numbers of unknown classes at deployment. As a result the support set may not include the class($\hat{y}$), which $\hat{x}$ belongs to, resulting in poor performance.

We propose a deployment approach where the number of classes in the support set size is customisable. Accordingly we introduce condition, $n_{te} \leq |\mathcal{L}|+|\hat{\mathcal{L}}|$, which facilitates inclusion of all available classes in the support set. This allows the classifier to make an informed decision when predicting activity, $\hat{y}$, relative to the support set.

Figure 1 illustrates how the typical matching network with fixed length support set are used for k-shot learning as well as ZSL. With ZSL we can see how the absence of the expected class in the support set can result in a poor classification outcome. One way round this is to try out several class combinations within the support set (potential for combinatorial explosion). The alternative is to expand the support size to cover as many as the expected number of classes at test / deployment time.

3 Evaluation

We compare two methodologies below to explore ZSL with Matching networks:

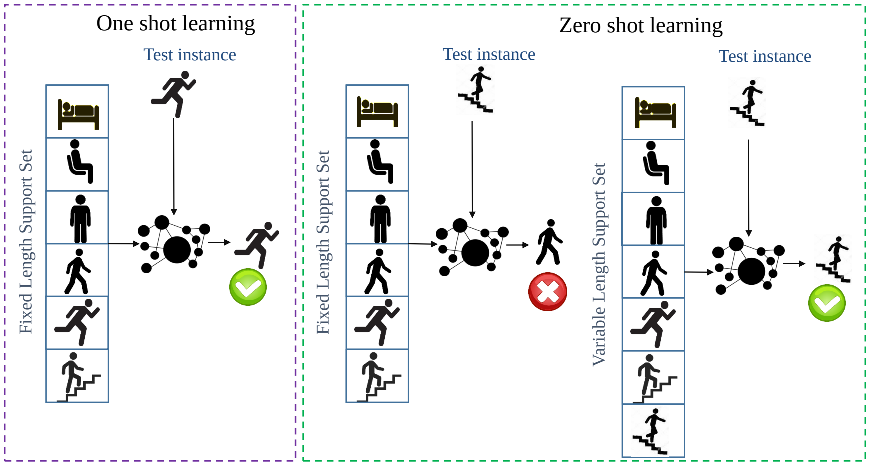

Fig. 1. One-shot Learning vs Zero-shot Learning with Matching Networks

- the original fixed length support sets where, $n_{te} = n_{tr}$; and
- the proposed variable length support set where, $n_{te} \leq |\mathcal{L}| + |\hat{\mathcal{L}}|$

3.1 Dataset and pre-processing

SelfBACK dataset was compiled with a tri-axial accelerometer data streams belonging to 9 activity classes performed by 50 individuals. Accelerometer was mounted on the right-hand wrist and each activity was performed for approximately 3 minutes. Data was recorded in {time-stamp, x, y, z} format at $100Hz$ sampling rate. We use pre-processing steps used by [7] in this work, that are supported with evidence of improved accuracy in previous work [6, 5].

3.2 Experiment Design

We follow a disjoint test strategy [9] when designing the experiment, where test classes will only include classes not seen during training. We evaluate ZSL with eight training classes ($|\mathcal{L}| = 8$) and one test class ($|\hat{\mathcal{L}}| = 1$). Accordingly we conduct nine experiments so that each class appears once as the test class. For instance $E_{jogging}$ will refer to experiment where jogging activity is the test class and other eight classes are in training. In which case we are testing the models ability to make predictions about the unseen *jogging* class. The two test approaches compared in the evaluation are:

Fixed length support sets: where $n_{tr} = n_{te} = 8$, we create several combinations of classes using a random sampling (without replacement) to select the 8 classes for the support set at test time from the set of 9 classes. Accordingly we evaluate 9 combinations for each experiment. Note that one combination in each experiment does not include the expected class, of the test instance, in which case, we assign test accuracy to be zero percent.

Variable length support sets: is introduced as having,$n_{tr} = 8$ and $n_{te} = 9$, to help mitigate situations (in the case of random selection with fixed length support set) where the expected test class may not be sampled in the test support set. Specifically we evaluate the condition, $n_{te} = |\mathcal{L}| + |\hat{\mathcal{L}}|$, where we include all nine classes in the test support set.

We use the matching network architecture in [7] to generate the embedding functions. In order to enforce personalisation, we always select support set data from the same user. We use $k = 5$ through all experiments. We use the first five instances of each test class to form the support set instances (simulating a scenario where the user provide few calibration examples for a class not seen during training), and rest as test instances. We perform a hold-out validation with all our experiments by selecting 8 random users as test set and rest as training set. We use accuracy as performance measure.

4 Results and Discussion

Table 1. Fixed length support set vs. Variable length support set results

Test Class	Fixed length support set				Variable length
	Min (%)	Max (%)	Avg (%)	Median (%)	support set (%)
Jogging	0	56	8	1	99
Sitting	0	93	69	76	94
Standing	0	93	60	64	83
Lying	0	81	51	59	74
Walk slow	0	75	57	62	57
Walk mod	0	62	44	46	56
Downstairs	0	44	31	30	30
Walk fast	0	63	40	44	28
Upstairs	0	46	12	7	24

For fixed length support sets, we evaluated 81 possible combinations of test support sets when one class was considered as not part of training. A summary of results (minimum, maximum, average and median accuracies) from these combinations appear in Table 1 (columns under fixed length support set). Minimum accuracy obtained with a test support set combination is always zero; this is when the support set combination does not contain the expected test class which at test time will be the calibration data. Note that random guessing of a class would achieve only 11% accuracy (as we have 9 classes). These results emphasise the importance of considering all possible classes in test support set. We present median of all combinations to show how average is affected by having an experiment with zero accuracy. We observe some interesting interactions between the target instances actual class and class combinations in its support set. For instance with $E_{jogging}$, accuracy was heavily penalised when class downstairs is in the support set. Similarly behaviour was observed with class $E_{upstairs}$ when *walk-slow* activity is included in the test support set. These results provide useful insights as to the effect of different activities have on each other due to their inter-class similarities at an embedding level.

Results with ZSL using a Variable length support set is presented in the last column of Table 1. We can observe that recognition of new activities (i.e. not seen during training but with some calibration data at test time) has in places significantly higher accuracy and surpass accuracies with the fixed length approach (e.g. Jogging and Sitting). With others; such as downstairs, walk fast and upstairs yield accuracies under 50%. Looking at classification accuracies from previous work on this data set [6], we identify these classes are challenging to classify even with a traditional classification algorithm. However they show comparable or better performance to average accuracies obtained with fixed length approach, hence we can conclude that variable length is still advantages over fixed length approach. Overall recognition of *Walk fast* proved to most challenging when compared to the average of fixed length approach. It is possible that the inter-class similarity relationships play a role here.

5 Conclusions

We explore two problems that ZSL with Matching Networks pose: the problem of finding the best class combinations with a fixed support set size at test time; and related to that the situation when the expected class may not be in the support set at test time. There is an optimum class combination that helps to classify unseen instances; however searching the space of possible combinations is a combinatorial problem. We resolve this by introducing a new test approach with variable length test support set. We experiment with a disjoint test set and yield comparatively better classification performance with variable length approach for several activities recognitions classes. In future we will look at how inter-class relationships in a support set might impact performance when applying ZSL. Additionally we will extend our evaluations to overlap test sets and other HAR data sets to further validate this approach.

References

1. Berlin, E., Van Laerhoven, K.: Detecting leisure activities with dense motif discovery. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. pp. 250–259. ACM (2012)
2. Gjoreski, H., Roggen, D.: Unsupervised online activity discovery using temporal behaviour assumption. In: Proceedings of the 2017 ACM International Symposium on Wearable Computers. pp. 42–49. ACM (2017)

3. Krebs, P., Duncan, D.T.: Health app use among us mobile phone owners: a national survey. JMIR mHealth and uHealth **3**(4) (2015)
4. Romera-Paredes, B., Torr, P.: An embarrassingly simple approach to zero-shot learning. In: International Conference on Machine Learning. pp. 2152–2161 (2015)
5. Sani, S., Massie, S., Wiratunga, N., Cooper, K.: Learning deep and shallow features for human activity recognition. In: International Conference on Knowledge Science, Engineering and Management. pp. 469–482. Springer (2017)
6. Sani, S., Wiratunga, N., Massie, S., Cooper, K.: knn sampling for personalised human activity recognition. In: International Conference on Case-Based Reasoning. pp. 330–344. Springer (2017)
7. Sani, S., Wiratunga, N., Massie, S., Cooper, K.: Personalised human activity recognition using matching networks. In: International Conference on Case-Based Reasoning. Springer (2018)
8. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. In: Advances in Neural Information Processing Systems. pp. 3630–3638 (2016)
9. Xian, Y., Schiele, B., Akata, Z.: Zero-shot learning-the good, the bad and the ugly. In: 30th IEEE Conference on Computer Vision and Pattern Recognition. pp. 3077–3086. IEEE Computer Society (2017)