

Towards a Deep Learning-based Activity Discovery System

Eoin Rogers, John D. Kelleher, and Robert J. Ross

Applied Intelligence Research Centre, Dublin Institute of Technology, Dublin, Ireland

`eoin.rogers@student.dit.ie`

`john.d.kelleher@dit.ie`

`robert.ross@dit.ie`

Abstract. Activity discovery is a challenging machine learning problem where we seek to uncover new or altered behavioural patterns in sensor data. In this paper we motivate and introduce a novel approach to activity discovery based on modern deep learning techniques. We hypothesise that our proposed approach can deal with interleaved datasets in a more intelligent manner than most existing AD methods. We also build upon prior work building hierarchies of activities that capture the inherent aggregate nature of complex activities and show how this could plausibly be adapted to work with the deep learning technique we present. Finally, we briefly talk about the challenge of evaluating activity discovery systems in a fair way and outline our future plans for implementing this model.

1 Introduction

Activity discovery (AD) refers to the unsupervised discovery of plausible human activities in unannotated datasets composed of sensor readings of human subjects. AD is itself a sub-field of *activity recognition*, the recognition of activities from sensor readings in a supervised manner. These technologies have potential applications in the automatic labelling of activity recognition datasets and building profiles of normal and abnormal behaviour.

In this paper, we propose a novel approach to activity discovery that makes use of deep learning technology. The discovery of activity patterns in sensor-network type sensor data is complicated significantly by the fact that interleaving is frequently seen in activity data. Interleaving can be thought of as situations in which multiple activities are occurring in parallel. Our system is designed with the explicit intention of helping to make progress in handling data where activities can be a tangle of sensor events that can interleave and overlap. We believe that our newly proposed method has the potential to allow these events to be untangled, and used to build a hierarchy of activities, allowing the capturing of situations where complex activities are composed of (or contain) a multitude of smaller activities.

The structure of this paper is as follows: section 2 reviews work that has been done previously in the field of activity discovery and related fields. Our proposed

approach is presented in section 3. Finally, we end with an overview and some comments in section 4.

2 Prior Work

Activity discovery is an active area of research. In this section, we briefly review some of the major contributions that have been made previously to this field. The field is itself a special case of *sequential pattern mining*, which is a field that has existed for many decades (see for example [13]).

Almost all approaches to activity discovery make use of the same basic methodology of trying to find patterns in the code. The means by which the patterns are found, however, can differ substantially between systems. For example, [4] makes use of topic modelling algorithms, an approach which has also been revived for [8]. Gu et al. meanwhile [3] have made use of *emerging patterns* or EPs (patterns which undergo large changes in their *support values* throughout the dataset, where a support value is the number of times the patterns appears over the length of a subset of the dataset) as the basis of an activity discovery system. This paper is interesting since the authors explicitly focus on the issue of interleaving (circumstances where multiple activities are taking place at the same time), which is an issue we are also attempting to address. More recently, [2] introduces an AD system that makes use of low-resolution video cameras and image processing algorithms.

A concept that is closely related to activity discovery, and may therefore be a useful source of inspiration, is grammar induction. This is the problem of taking a set of example sentences of a language, and attempting to deduce the grammar of the language using only those examples. A number of ways of doing this exist. ADIOS [11] is a prominent grammar induction algorithm that works by finding patterns probabilistically in the dataset, after it has been read into a graph structure. Another approach, using expectation maximisation techniques, is presented by [12]. This involves trying to use annealing techniques and the idea of rearranging or deleting words from the dataset in order to overcome the complex search spaces inherent in grammar induction.

There are a number of pre-existing datasets that are appropriate for use in evaluating Activity Discovery algorithms. These provide real-world data from an activity recognition setup which we can use as input to a system. One of the most prominent datasets in use for activity recognition is the Opportunity dataset [9]. This has all of the major characteristics that we need from a dataset, in that it is large, contains real-world data, and the activities within it are complex and interleaved. Likewise, the Center for Advanced Studies in Adaptive Systems (CASAS) at Washington State University have produced a range of datasets that would be of use for this also, for instance the Kyoto 3 dataset [10]. These are challenging datasets that can provide a good test for Activity Discovery systems. All of these datasets can be seen as a sequence of sensor *events*. For example, if a sensor is fit to a door, an event will appear whenever someone opens or closes the door.

One challenge facing researchers attempting to develop activity discovery systems concerns fair evaluation metrics. In the case of pre-annotated datasets, we can use the annotations as a ground truth to compare against, allowing us to calculate raw accuracy or F-measure values. These measures all depend on the accuracy of the annotations, however, and this is not guaranteed, nor are annotated datasets easy to come by. For this reason, a number of authors have proposed using the concept of *minimum description length* (MDL) to measure the performance of AD systems [7][8]. This states that the performance of a machine learning model is equal to the size of that model plus the size of the dataset after the model has been used to compress it. This gives us a simple metric that does not require any sort of prior human annotation, while retaining the property of being mathematically justifiable. Although it is primarily aimed at activity recognition rather than discovery, a method of classifying and enumerating types of errors has been proposed by [14]. This could potentially also be used for activity discovery if ground truth data is available.

3 Proposed Approach

We propose here our new approach to activity discovery that we speculate can be used to discover activities automatically even in situations where the activities are heavily interleaved. To make progress on this problem, we turn to *deep learning*, a now well-developed technique which allows for the efficient construction and training of networks of computational units [6]. From the field of deep learning we will focus in particular on *neural language models*, a form of deep neural network proposed by [1] as a solution to the wider problem of language modelling in natural language processing. If we represent a natural language sentence \mathbf{s} as a sequence of words $\mathbf{s} = \langle w_1, w_2, \dots, w_l \rangle$, and $i \in \{1, 2, \dots, l\}$ is an index into \mathbf{s} , a language model computes a probability distribution over possible values for w_i given $\mathbf{w}_{1:i-1}$, where $\mathbf{w}_{x:y}$ is shorthand for $\langle w_x, w_{x+1}, \dots, w_y \rangle$.

A neural language model is simply a deep network that is trained to approximate the language model distribution. The input layer takes one-hot encodings of the last n words in the given sentence, in other words a one-hot encoding for $\mathbf{w}_{i-n:i}$ and the output layer is a softmax distribution over all possible values for w_{i+1} . Our proposal generalises this basic concept of a language model into a more complex distribution. Given an n -gram window $\mathbf{w}_{i-n:i}$ and an output window $\mathbf{w}_{i+1:i+1+m}$ of length m , for each $w_j \in \mathbf{w}_{i+1:i+1+m}$ we compute:

$$P(w_j | \mathbf{w}_{i-n:i}) \tag{1}$$

In other words, we for each w_j , we compute the probability that w_j would appear anywhere in the m words following the n -gram window of length n . The standard language model presented above, therefore, is just a variant of this model with $m = 1$. This can be implemented using a deep network that is almost identical to the standard model's network, and is shown in Figure 1. Again, we pass a one-hot encoding of the previous n words into the input layer,

and receive a softmax over possible successor words in the output layer. However, the output now models the probability of each possible word appearing in one of the next m words after word w_i , rather than just appearing as w_{i+1} .

The choice of units for the model is important for achieving best results. In general, it has been found that training a neural network works best if the average outputs from all units is close to zero. This suggests the use of something like a hyperbolic tangent function for the middle layers [5], although it may be worthwhile experimenting with logistic layers also.

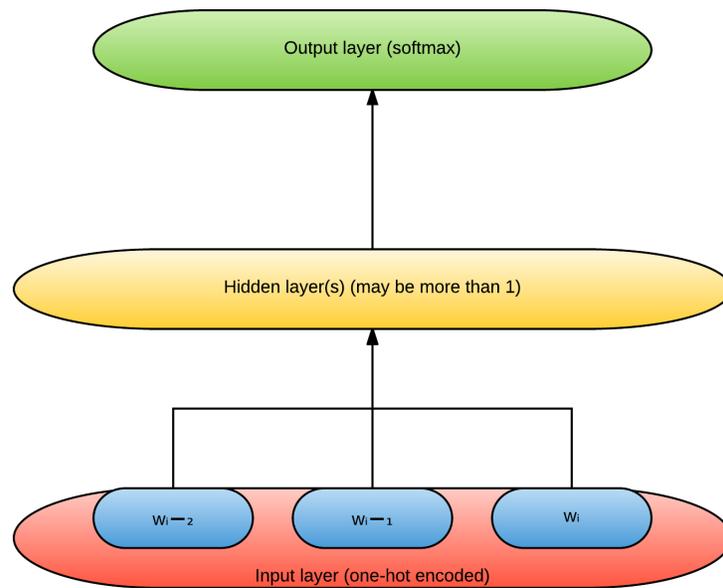


Fig. 1. A diagram of our proposed neural language model

This generalised neural language model would not be of much use for linguistic purposes. However, if we imagine that each w_i of our sequence \mathbf{s} is in fact a sensor event, rather than a word, we can use this model for activity discovery. An example of how this is done is shown in Figure 2. In Figure 2(a), we see the initial setup of the system: events **A** and **B** are contained in the sliding window that will constitute our input (so $n = 2$), and events **C** to **F** are in the window of length m that we want to compare to our output layer probabilities. Suppose that our language model predicts that event **D** will appear within the next m events with high probability. We thus add a *link* connecting events **B** and **D** as shown in Figure 2(b). Note that it is also possible for more links to be added.

For example, if event **E** was also predicted with high probability, we would add another link from **B** to **E**. The exact mechanism by which a probability will be decided to be high enough to form a link has yet to be fully determined by the authors, although it will probably involve identifying probabilities more than a certain threshold above the average for the dataset (i.e. significantly more probable than background noise).

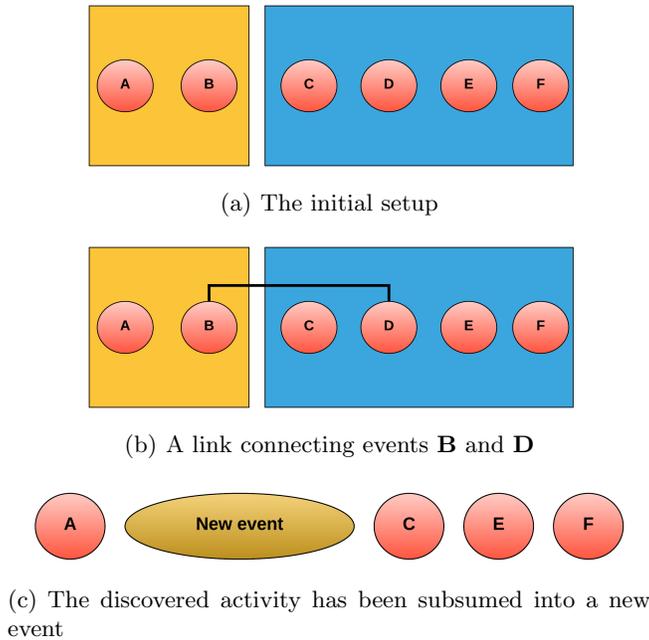


Fig. 2. An overview of our proposed approach

The hypothesis that we are working on is that these links form activities, which are untangled from their original interleaved state. After this, we move our n -gram and m -gram windows on by one event, which allows us to continue the process. We will continue moving the windows until we reach the end of the dataset, adding new links along the way. Note that links can be added to activities discovered in previous iterations also, so we may add new events to the activity shown in Figure 2(b). As a result, we cannot see the full extent of a discovered activity until the system has run through the entire dataset. When this is done, we should see multiple, interleaving activities have been explicitly identified.

In many cases, a single activity can be composed of multiple sub-activities. For example, if you had an activity called *Making dinner*, it would be reasonable to expect it to be composed of a sub-activity called *Pour drink*. Finding such

activities is itself a difficult task called *Hierarchical activity discovery* (HAD). We would like to adapt the activity discovery approach proposed here to integrate it into an HAD system we have built previously [8]. This system works by running a normal (flat) activity discovery system over the dataset and then subsuming discovered activities into a single events. We then repeat the process until we cannot discover any more activities. Naïvely doing this with this new discovery system would not work, since the activities discovered are interleaved and need to be disentangled before we can subsume them into events. We now propose a simple way of doing this. The expected outcome of this process is shown in Figure 2(c). We have removed events **B** and **D**, since they are the constituents of the activity we are trying to generalise, and we do replace them with a single new activity. However, event **C**, which occurred in the middle of the new activity but was not part of it, is not included in the new activity, and is instead moved to one side of it (the right in this case). The question of which side of the new activity an interleaved event should be moved to can be decided as follows:

- If the event is itself part of an activity, move it to the side that the majority of that activity’s events occur at. This will allow that activity to be generalised easily in turn.
- In all other circumstances, simply move it to the left end of the new activity if it occurs temporally closer to the first event of the new activity, and to the right end otherwise
- For some particularly highly interleaved datasets, we may not want to move the interleaved events at all, but simply have instances of the new event on either side of them

Finally, we aim to use the minimum description length principle, as described in section 2 to evaluate the performance of the system. This negates the need for us to compare directly to the ground truth since achieving a high compression ratio guarantees that we have found one or more frequently repeating patterns in the data, although it may also be worthwhile comparing to human annotations to determine if the discovered and annotated activities agree. The error analysis method proposed by [14] may also be used for datasets with a pre-existing ground truth, highlighting the differences between the two systems.

4 Conclusion

This paper proposed a technique for activity discovery which makes use of recent advances in computational linguistics and neural networks, along with a plausible means for the evaluation of such a system. Unlike many other activity discovery systems, we are explicitly attempting to deal with the problem of interleaving, and the described algorithm has a means to identify and subsequently disentangle interleaved data. Our plan over the coming months is to implement and test this system using the Opportunity dataset [9] to allow easy comparison with existing AD systems, and report the results back to the activity recognition community. Opportunity contains readings from a number on-body sensors worn

by participants as they carried out a range of *Activities of Daily Living* (ADL), which are high-level activities that would be expected to occur in a residential environment. On-body sensors are becoming increasingly common due to the increasing prevalence of devices such as smartphones and smartwatches. However, these do bring up privacy concerns, so we may also evaluate the system on datasets that use sensors embedded in the environment, such as the Kyoto 3 dataset [10].

One issue that could be worth investigating is the consequences varying lengths of the output (m) and input (n) windows. Our expectation is that there should be an optimal size for m , and thus a degree of trial-and-error may be needed. If m is too small, the system will fail to find interleaved activities since it won't be able to add links long enough to bridge the gaps between the distant sensor events. Likewise, excessively long windows may cause spurious links to be formed, since the probability that any one sensor event will appear in the window approaches one as the length of m increases.

Longer term, there are a number of aspects to the system that we would like to investigate:

- What kinds of errors does this system tend to produce? Are they different from the error types produced by other AD systems?
- Could weighting links by length, so as to penalise long-range links and encourage shorter links improve the performance of the system? If not, could penalising short-range links and encouraging longer links?
- Can we find a way to represent temporal information in the event encoding in a way that can be usefully exploited by the algorithm? Is temporal information even important for activity discovery?

We see our proposed model as a solid initial step in this research agenda.

References

1. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A Neural Probabilistic Language Model. *Journal of Machine Learning Research* 3, 1137–1155 (2003)
2. Eldib, M., Deboeverie, F., Philips, W., Aghajan, H.: Behavior analysis for elderly care using a network of low-resolution visual sensors. *Journal of Electronic Imaging*, 25(4), pp.041003–041003 (2016)
3. Gu, T., Wang, L., Wu, Z., Tao, X., Lu, J.: A Pattern Mining Approach to Sensor-Based Human Activity Recognition. *IEEE Transactions on Knowledge and Data Engineering*, Volume 23 Number 9, 1359–4347 (2011)
4. Huynh, T., Fritz, M., Schiele, B.: Discovery of Activity Patterns using Topic Models. *Proceedings of the 10th International Conference on Ubiquitous Computing*, 10–19 (2008)
5. LeCun, Y.A., Bottou, L., Orr, G.B. and Mller, K.R.: Efficient backprop. In *Neural networks: Tricks of the trade*, 9–48 (2012)
6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* 521, 436–444 (2015)
7. Li, N., Crane, M., Gurrin, C. and Ruskin, H.J.: Finding motifs in large personal lifelogs. In *Proceedings of the 7th Augmented Human International Conference 2016* (2016)

8. Rogers, E., Kelleher, J.D., Ross, R.J.: Using Topic Modelling Algorithms for Hierarchical Activity Discovery. In *Ambient Intelligence–Software and Applications–7th International Symposium on Ambient Intelligence (ISAmI 2016)*, 41–48 (2016)
9. Sagha, H., Digumarti, S.T., Millán, J.D.R., Chavarriaga, R., Calatroni, A., Roggen, D. and Tröster, G.: Benchmarking classification techniques using the Opportunity human activity dataset. In *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 36–40 (2011)
10. Singla, G., Cook, D., Schmitter-Edgecombe, M.: Tracking activities in complex settings using smart environment technologies. *International Journal of Bio-Sciences, Psychiatry and Technology*, 1(1):25–35 (2009)
11. Solan, Z., Horn, D., Ruppín, E., Edelman, S.: Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), 11629–11634 (2005)
12. Smith, N., Eisner, J.: Novel estimation methods for unsupervised discovery of latent structure in natural language text. PhD thesis (2007)
13. Srikant, R., Agrawal, R.: Mining sequential patterns: Generalizations and performance improvements. In *International Conference on Extending Database Technology*, 1–17, Springer Berlin Heidelberg (1996)
14. Ward, J.A., Lukowicz, P. and Tröster, G.: Evaluating performance in continuous context recognition using event-driven error characterisation. In *International Symposium on Location-and Context-Awareness*, 239–255, Springer Berlin Heidelberg (2006)