Advances in Human Activity Recognition

Gulustan Dogan, University of North Carolina Wilmington

This installment of Computer's series highlighting work published in IEEE Computer Society journals comes from IEEE Transactions on Pattern Analysis and Machine Intelligence.

Human activity recognition (HAR) is the matter of classifying sequences of accelerometer data recorded by specialized harnesses or smartphones into known, well-defined movements. This is a challenging problem given the large number of observations produced each second, the temporal nature of the observations, and the lack of a clear way to relate data to known movements. Traditional approaches involve hand-crafting features from the time series data based on fixed-size windows and training machine learning models, such as ensembles of decision trees. A potential challenge is that this feature engineering requires deep expertise in the field.[1]

Recurrent neural networks (RNNs), of which long short-term memory units (LSTMs) are the most powerful and well-known subset, are a type of artificial neural network designed to recognize patterns in sequences of data, such as numerical times series data emanating from sensors. What differentiates RNNs and LSTMs from other neural networks is that they take time and sequence into account; they have a temporal dimension.[2] While LSTMs work quite well, they are computationally expensive and have issues such as obscurity in regard to how the memory states reveal dynamic alignments between different sequences of patterns. Due to these issues, there may be a limit on their ability to fully explore patterns distinguishing different types of human activities.

In "Learning Compact Features for Human Activity Recognition via Probabilistic First-Take-All," authors Ye et al.[3] introduced a novel solution for exploring the temporal dynamics underlying sequential data such that they can produce a highly compact feature representation from raw signals (as shown in Figure 1). Their algorithm is applicable to both camera-based and wearable sensor-based HAR systems. In addition, the article

Digital Object Identifier 10.1109/MC.2021.3055671
Date of current version: 7 May 2021

EDITOR RON VETTER
University of North Carolina Wilmington; vetterr@uncw.edu

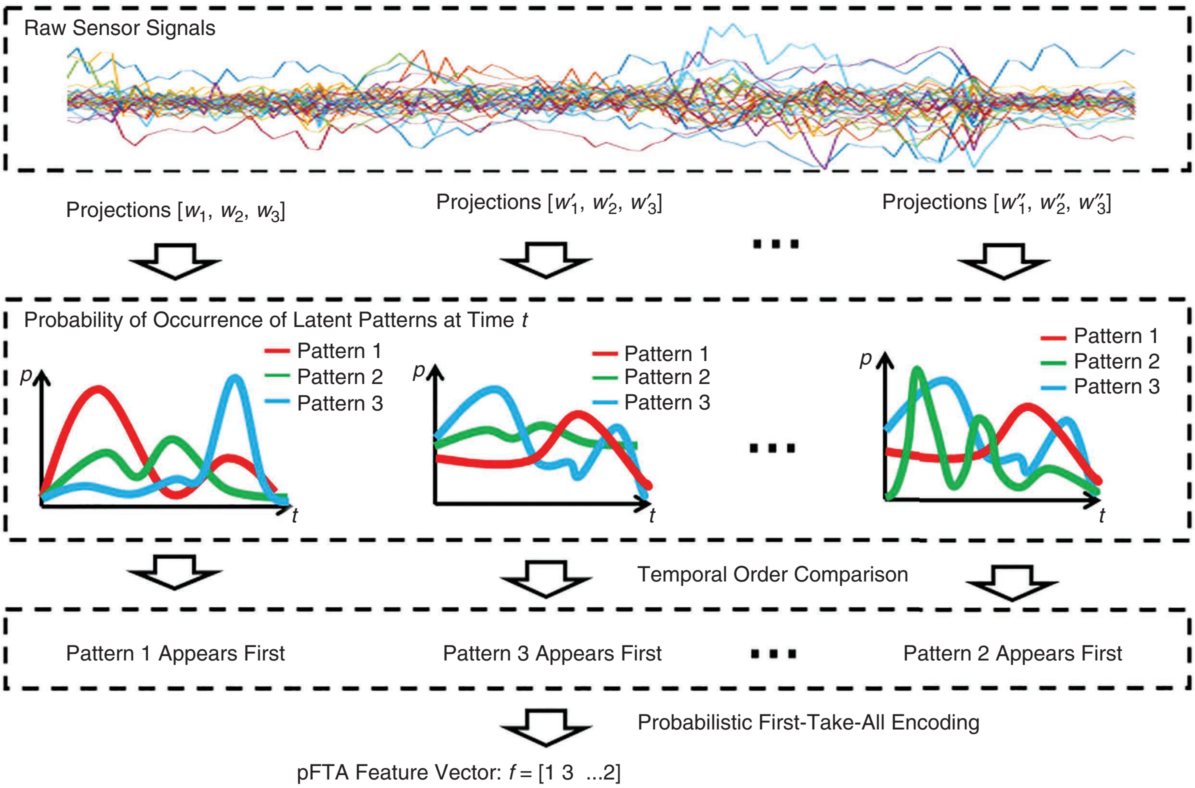

FIGURE 1. A diagram of the probabilistic first-take-all (pFTA) algorithm. A set of three projections are used to produce one dimension of the pFTA feature.

describes how current HAR solutions are accomplished and can be used to identify ambulatory activities such as laying down, walking, and climbing stairs. These solutions often use off-the-shelf classifiers such as decision trees, random forests, and support vector machines.

Traditional approaches involve hand-crafting features from the time series data based on fixed-size windows and training machine learning models, such as ensembles of decision trees.

The authors describe three data sets that they used to evaluate the performance of their new algorithm: a) UCI Daily and Sports Activities, b) Smartphone-Based Human Activity, and c) MSR ActionPairs. The UCI data set consists of 19 daily and sports activities performed by eight subjects with sensors placed on five locations on their body. The Smartphone-Based Human Activity data set is collected from 30 subjects performing six daily activities. For this data set, a single device is used to collect the information—a Samsung Galaxy S2. Finally, the MSR ActionPairs data set is used for 3D human action recognition. The data are captured from a Kinect depth sensor and contain a red-green-blue stream, a depth stream, and a human skeleton joint. In this regard, the data collected are visual information and consist of 10 subjects performing 12 actions.

Performance results on all three data sets show that the new algorithm is very effective for time series data. For wearable sensors, the algorithm's optimized learning version seems to be consistently the best and is often significantly cheaper in terms of computational time and cost. However, when it comes to visual data, although it is still effective, there are a few other methods that seem to be better than Ye et al.'s[3] method. Despite this, it is important to note that while these other methods do have a higher accuracy, they also use highly sophisticated features that might not work in a more general sense.

REFERENCES

1. J. Brownlee. "A gentle introduction to a standard human activity recognition problem." Machine Learning Mastery. https://machinelearningmastery.com/how-to-load-and-explore-a-standard-human-activity-recognition-problem/ (accessed Mar. 1, 2021).
2. C. Nicholson. "A beginner's guide to LSTMs and recurrent neural networks." Pathmind. https://wiki.pathmind.com/lstm (accessed Mar. 1, 2021).
3. J. Ye, G.-J. Qi, N. Zhuang, H. Hu, and K. A. Hua, "Learning compact features for human activity recognition via probabilistic first-take-all," *IEEE Trans. Pattern Analysis and Mach. Intell.*, vol. 42, no. 1, pp. 126–139, Jan. 2020.

GULUSTAN DOGAN is an assistant professor in the Department of Computer Science at the University of North Carolina Wilmington, Wilmington, North Carolina, 28403, USA. Contact her at dogang@uncw.edu.

Digital Object Identifier 10.1109/MC.2021.3072474