Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

Are Spiking Neural Networks Useful for Classifying and Early Recognition of Spatio-Temporal Patterns?

Banafsheh Rekabdar *
Univeristy of Nevada, Reno
brekabdar@unr.edu

Abstract

Learning and recognizing spatio-temporal patterns is an important problem for all biological systems. Gestures, movements and activities, all encompass both spatial and temporal information that is critical for implicit communication and learning. This paper presents a novel, unsupervised approach for learning, recognizing and early classifying spatio-temporal patterns using spiking neural networks for human robotic domains. The proposed spiking approach has four variations which have been validated on images of handwritten digits and human hand gestures and motions. The main contributions of this work are as follows: i) it requires a very small number of training examples, ii) it enables early recognition from only partial information of the pattern, iii) it learns patterns in an unsupervised manner, iv) it accepts variable sized input patterns, v) it is invariant to scale and translation, vi) it can recognize patterns in real-time and, vii) it is suitable for human-robot interaction applications and has been successfully tested on a PR2 robot. We also compared all variations of this approach with well-known supervised machine learning methods including support vector machines (SVM), regularized logistic regression (LR) and ensemble neural networks (ENN). Although our approach is unsupervised, it outperforms others and in some cases, provides comparable results with other methods.

1 Introduction

This research is motivated by two artificial intelligence problems that rely on an autonomous system's ability to encode and recognize spatio-temporal patterns: intent recognition and imitation learning. In both domains, the activities observed by the autonomous system contain both spatial and temporal information. Spike timing neural networks (SNNs) are suitable to model spatio-temporal patterns. In this work we propose a method based on SNNs that addresses the problem of learning, recognizing and early classifying spatio-temporal patterns, which are typically encountered when representing gestures or other human actions.

*Thanks to my supervisor Dr. Monica Nicolescu for all her guidance and support.

2 Approach and Results

We use a spiking network which consists of cortical spiking neurons with axonal conduction delays [Izhikevich, 2006]. Neurons in this network are connected to 10% other neurons with a 2D Guassian distribution. The first step in our proposed approach is mapping spatio-temporal patterns (such as hand-drawn digits (images) or gestures) to the neurons in our spiking network. This is achievable by encoding spatio-temporal patterns into neural spike trains. In the training phase, the spike trains correspond to training data that are being used for stimulating the neural network. Given that our dataset is a set of images, in order to map a pattern into a spike train, we assign one neuron to each pixel. The order of firing the neurons is based on the temporal information of the pattern being drawn. Given that our dataset is a set of human hand's gestures, the spatio-temporal patterns are mapped to the spike trains by assigning groups of 5 neurons to different orientations, which correspond to movements of a human hand while showing a pattern. The neurons are fired in the order of observed orientations within the pattern. No labels for training samples have been provided to the network, as our approach is unsupervised. The synaptic weights in our spiking network are updated during training using the spike-timing dependent plasticity rule (STDP). Second, after the network is fully trained, each training pattern is presented to the network and network's response is used to create a model that corresponds to that pattern. In this way we have one model corresponding to each training sample. These models are being used in the classification phase later on. Third, in the classification phase, for classifying an unseen spatio-temporal pattern, we map it into a spike train and then create a model that represents the network's response to that pattern. The classification decision is made based on the comparison between the testing model and all the training models and selecting the closest training sample model to the input model.

In contrast with all other types of neural networks, a spiking neural network does not have any output layer. After presenting one input pattern (training or testing) to the network, the network's overall behaviour is considered as the output. We considered three different responses of spiking neural networks to the patterns. These responses are 1) an

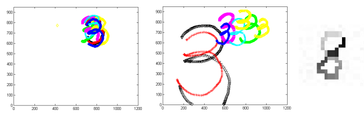

Figure 1: Left: all samples of training data for digit 3 (hand gesture dataset), middle: a subset of correctly classified testing data for digit 3 (hand gesture dataset), right: one testing sample of digit 8 (hand-written digit dataset)

unordered set of polychronous neural groups (PNG), which encode stereotypical time-locked firing patterns, 2) a string of all PNGs being activated by the pattern (sequence of PNGs - order is important) and 3) a string of ”characters”, in which each character is a set of neurons that fired at a particular time step. For classification, the model for each input pattern is compared to all training models. We defined different comparison methods for models generated from spiking networks. For unordered set of PNGs we use Jaccard index (*Unordered PNGs-Jaccard*), for sequence of PNGs we use longest common subsequence (LCS) (*ordered PNGs-LCS*), for string of characters we use a specified combination of Jaccard and LCS (*characters-Jaccard-LCS*). All approaches enable early detection of patterns.

The classifier approach for *unordered PNGs-Jaccard* models consists of a parallelized CUDA implementation of a spike-timing neural network. This work provides a real-time classification of spatio-temporal patterns with the ability to provide early predictions of the patterns before their completion. This approach is very well suited for real time and human-robot interaction applications. We also obtained promising results with testing our approach on a PR2 robot.

We evaluated the performance of our approaches by applying them on two data sets of hand-drawn digits and human hand gestures. In the hand gesture dataset there are a

Table 1: Comparison Results (unit: %)

	SVM	LR	ENN
Accuracy: Image dataset	86	76.6	55
Accuracy: Gesture dataset	47.36	31.1	19.62

set of variable sized gestures representing the digits from 0 to 9, extracted from video data of a human drawing the corresponding digits. There are 7 and 21 samples per each digit for training and testing phase respectively. As shown in the left part of Figure 1 although there was relatively low variation in both scale and translation in the training data, our approach was able to identify patterns correctly at different scales and locations in the field of view (middle part of Figure 1).

The hand-written digit dataset contains grayscale 16 by 16 images of hand-written digits from 0 to 9. As the right part of Figure 1 shows, the pixels with intensity values different than white are a part of the written digit. Pixel intensity values represent the relative temporal information: pixels with lower intensities are drawn before pixels with higher intensity values. There are 5 and 50 samples per each digit for the training and the testing phase respectively. For the dataset of hand-drawn digits the *unordered-PNGs-Jaccard* [Rekabdar *et al.*, 2015a] and the *characters-Jaccard-LCS* [Rekabdar *et al.*, 2016] approaches work very well and for the dataset of hand gesture, the *unordered-PNGs-Jaccard* [Rekabdar *et al.*, 2015c] and the *ordered-PNGs-Jaccard* [Rekabdar *et al.*, 2015b] methods are performing better. Table 2 shows the overall accuracy of aforementioned approaches for both datasets. We also tested three supervised standard pattern classification methods including, SVM, LR and ENN for both datasets for comparison with our unsupervised approaches. Table 1 shows that our system outperformed and in some cases, provided comparable results.

Table 2: Classification Results (un-PNG-Jac: unordered-PNGs-Jaccard, ch-Jac-LCS: characters-Jaccard-LCS, or-PNG-Jac:ordered-PNGs-Jaccard, SR: Success Rate, ER: Error Rate)

	un-PNG-Jac Image dataset	ch-Jac-LCS Image dataset	un-PNGs-Jac Gesture dataset	or-PNG-Jac Gesture dataset
SR	81.40%	82.5%	85.2%	83%
ER	18.6%	17.9%	14.8%	17%

3 Future Direction

In the future I would like to apply my approach on more HRI applications using PR2 robot and focus on on-line learning. [1]

References

[Izhikevich, 2006] Eugene M Izhikevich. Polychronization: computation with spikes. *Neural computation*, 2006.

[Rekabdar *et al.*, 2015a] Banafsheh Rekabdar, Monica Nicolescu, Richard Kelley, and Mircea Nicolescu. An unsupervised approach to learning and early detection of spatio-temporal patterns using spiking neural networks. *Journal of Intelligent & Robotic Systems*, pages 1–15, 2015.

[Rekabdar *et al.*, 2015b] Banafsheh Rekabdar, Monica Nicolescu, Mircea Nicolescu, and Richard Kelley. Scale and translation invariant learning of spatio-temporal patterns using longest common subsequences and spiking neural networks. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2015.

[Rekabdar *et al.*, 2015c] Banafsheh Rekabdar, Monica Nicolescu, Mircea Nicolescu, Mohammad Taghi Saffar, and Richard Kelley. A scale and translation invariant approach for early classification of spatio-temporal patterns using spiking neural networks. *Neural Processing Letters*, pages 1–17, 2015.

[Rekabdar *et al.*, 2016] Banafsheh Rekabdar, Monica Nicolescu, Mircea Nicolescu, and Sushil Louis. Using patterns of firing neurons in spiking neural networks for learning and early recognition of spatio-temporal patterns. *Neural Computing and Applications*, pages 1–17, 2016.

[1]This work has been supported by ONR grant N000141210860.