

*Software Engineering*  
*By Damric Dobric/Andreas Pech*

# *Migration of video learning project*

Nusrat Jahan Sumi  
Matriculation ID: 1345476

Mashnunul Huq  
Matriculation ID: 1384042

**Abstract:** This paper represents an improving machining learning algorithm which is learning Video data using Hierarchical Temporal Memory (HTM). The Spatial Pooler (SP) model shows how neurons learn by feedforward connections and form effective classification of the input frame. It converts binary input pattern into space distributed representation (SDR) by using Cortical Learning rules and homeostatic plasticity control for frame pattern prediction. The result of the learning is tested by giving the trained model an arbitrary image, the model then tries to recreate a video with succeeding frames after the input frame.

**Keywords**—homeostatic plasticity controller, formatting, division into frames, prediction, training & testing

## I. INTRODUCTION

The HTM (Hierarchical Temporal Memory) is based on “Thousand Brains Theory” which explains how an object behaviors and high-level concepts gets tightly replicated across a cortical column but not only on the top layer and gets distributed throughout the neocortex. Here spatial pooler involves different computational principles of the cortex (Hawkins, 2017). It depends on competitive Hebbian learning, homeostatic excitability control, topology of connections in sensory cortices and structural plasticity. The HTM Spatial pooler is developed in such a way to achieve a set of computational properties which includes 1. Preserving topology of the input space by mapping similar inputs to similar outputs 2. Continuously adapting to changing statistics of the input stream 3. Forming fixed sparsity representations 4. Being robust to noise and 5. Being fault tolerant that supports computations with SDRs (Sparse Distributed Representations). The output of the SP which is the integral component of HTM can be easily recognized by downstream neurons and contribute to improved performance in the end-to-end HTM system.

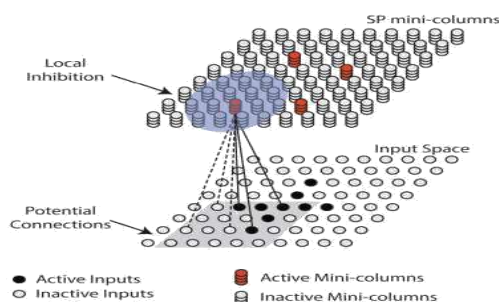


Figure 1: The process of spatial pooling task of SP is to transforms input patterns into Sparse Distributed

Representation in a continuous way in end-to-end HTM system. The temporal sequences of these SDRs is learned by the HTM and do some prediction for the upcoming inputs. A single layer in HTM network is consist of a set of mini-columns which is consist of cells. Here the figure 1. shows that the HTM spatial pooler converts inputs at bottom to SDRs at top. Each SP mini-column (Active mini-columns and inactive Mini-columns) forms synaptic connections to a subset of the input space which is consist of gray square and potential connections. A local inhibition technique gives confirmation that within the local inhibition radius (shaded blue circle) a small fraction of the SP mini-columns that receive most of the inputs are active. According to the Hebbian rule Synaptic permanences are adjusted like this for each active SP mini-column, active inputs (black lines) are reinforced and inactive inputs (dashed lines) are punished (Boudreau, 2018).

At the time of building intelligent systems to mimic human intelligence and cognition, we must pay serious attention to sequences, including sequence learning as sequential behavior is essential to intelligence. In the task of time series prediction, video analysis and musical information retrieval, a model must learn from inputs that are sequences. Another important concern in sequence learning is hierarchical structuring of sequences. Many real-world problems that have sequences are involved with clear hierarchical structures like a sequence is made up of subsequences and they in turn are made up of sub subsequences and so on. By removing the difficulty to identify automatically these subsequences and deal with them accordingly which is related to temporal dependences, learning hierarchical structures help to reduce or eliminate temporal dependencies. It helps to compress the description or sequences.

Sequence learning is not a easy task. Sequence learning are needed powerful algorithms. Sequence learning which indicates either generation, prediction or recognition is usually based on the models of legitimate sequences which can be developed through training with exemplars (Lipton & Berkowitz, 2015). Hierarchical Temporal Memory proposed new computational learning models, Cortical Learning Algorithms (CLA), that is inspired from the neocortex which offer a better understanding of how our brains function. CLA mimics the procedure of human brain how to achieve pattern recognition and make intelligent predictions. The CLA processes the streams of information, classify them, learning to identify the differences and using time-based patterns to make predictions as like as performed by the neocortex in humans. But the place of time is significant in case of learning, inference and prediction. The temporal sequence is achieved from HTM algorithm from the stream of input data.

This type of works forecasting that Machine Learning (ML) or statistical modelling emphasis here is to enable the reader to understand on some of ML or statistical techniques actively used in past and till the present moment.

The created SDR which is the encoded spatial pattern of that object is used as the input to the Temporal Memory which learns about the pattern when the spatial pooler is in stable mode and removes the pattern when it is in unstable mode. SP oscillates between stable and unstable mode and the TM also learns and forgets about the pattern. But too much oscillation can cause permanent disruption to the program hence causing higher computational resources. To reduce this scenario we used homeostatic plasticity controller which influences excitation and inhibition balance of

neurons. The functional stability of neural columns is achieved by SP and TM setting cells in active or predictive state. SP provides Global and Local inhibition which controls the number of cells must be activated in the currently processing area. To keep the stability of the Spatial Pooler and learning of TM a set of common parameters were selected while instantiating HTM (see Table 3) and kept in the [htmConfig.json](#) file. Some of the configurations are manipulated while running the program in `ModifyHtmFromCode` method in the Main [Program](#) class.

Table 3: HTM (Hierarchical Temporal Memory) Configuration in `htmConfig.json` file to initiate `HtmCofig` class.

Parameters	Value
CellsPerColumn	30
GlobalInhibition	true
NumActiveColumnsPerInhArea	0.02*ColumnDimension
PotentialRadius	0.15* InputDimension
MaxBoost	30.0
MaxSynapsesPerSegment	0.02*ColumnDimension
Random	42
MinPctOverlapDutyCycles	0.75
DutyCyclePeriod	100
StimulusThreshold	0.05*ColumnDimension
UpdatePeriod	50
PermanenceIncrement	0.1
PermanenceDecrement	0.01

Now the boosting in spatial pooler makes sure that all columns are uniformly used across all seen patterns. As the mechanism remains active throughout the process the boosting of columns which already build learned SDRs is possible (Dobrick, 2021). Deactivation of boosting in homeostatic plasticity in the cortical layer can also be applied to SP. But the actual understanding to this is yet to be revealed. Till now in HTM this technique consists of boosting and inhibition algorithms which works on the minimum column level and not on the cell level in the minimum column. Because SP operates on the population of neural cells in minimum column rather than the individual cells(see Figure 2). Therefore, the Spatial Pooler with the New-born Stage is used with the aim to send input pattern of SDR in each iteration to the homeostatic plasticity controller telling the program that SP has reached instable stage and program will disable the boosting. As the SP has entered to a stable state it will leave the new-born cycle and continue operating as usual without boosting which will help in reducing computational time.

For differentiating multi sequence learning and sequence learning we instantiated `HtmClassifier` with two different approaches. In the sequence learning method defined in [VideoLearning](#) class as `TrainWithFrameKey`, we put the frame key as `HtmClassifier` key while calling for the learn method. On the other hand for sequential learning we used series of frame as the `HtmClassifier` key while calling for the learn method (see Table 4). By the definition sequence learning should take more computational time while learning as it learns by each frame. But the multi sequence

learning should take less time as it takes a bunch of frames while learning.

Table 4: Code for training with frame key in sequential learning and with series of frame keys in multisequence learning.

```
public static void TrainWithFrameKey(VideoConfig videoConfig =
null, HtmConfig htmCfg = null)
{
    HtmClassifier<string, ComputeCycle> cls = new();
    HomeostaticPlasticityController hpa = new(mem,
maxNumOfElementsInSequence * 150 * 3, (isStable, numPatterns,
actColAvg, seenInputs) =>{}), numCyclesToWaitOnChange: 50)
    SpatialPoolerMT sp = new(hpa);
    for (int i = 0; i < maxCycles; i++)
    {
        foreach (var currentFrame in nv.nFrames)
        {
            cls.Learn(currentFrame.FrameKey, actCells.ToArray());
            cls.Learn(key, actCells.ToArray()); //For TrainWithFrameKeys
        }
    }
}
```

#### E. Predicting Frames from an input

After the learning we counted the accuracy of each learned video by calling a method `PredictImageInput` of [VideoLearning](#) class which takes an image and recreates the consecutive frames after that image. This is done in two stages. First from image directories given in the `videoConfig.json` file (see Table 5) and then taking images as directory path from users. The directory holding testing videos without input from user contains frames created by the program as it is required to test from the given input also.

Table 5: Test files containing frames to be tested in `videoConfig.json` file.

```
"TestFiles":[
    "TestImageSet\\Converted\\Circle\\circle\\Circle_circle_3.png",
    "TestImageSet\\Converted\\Circle\\circle\\Circle_circle_2.png",
    "TestImageSet\\Converted\\Line\\line\\Line_line_11.png",
    "TestImageSet\\Converted\\Line\\line\\Line_line_22.png",
    "TestImageSet\\Converted\\Rectangle\\rectangle\\
\\Rectangle_rectangle_28.png",
    "TestImageSet\\Converted\\Rectangle\\rectangle\\
\\Rectangle_rectangle_18.png",
    "TestImageSet\\Converted\\Triangle\\triangle\\Triangle_triangle_23.png",
    "TestImageSet\\Converted\\Triangle\\triangle\\Triangle_triangle_0.png"
]
```

As all the data is binarized, these images also needs to be binarized with [NFrame](#) class's `BitmapToBinaryArray` method. While making the predicted future frames `IntArrayToBitmap` method of the same class is used and then combining all of these frames are done in [NVideo](#) class's `CreateVideoFromFrames` method and saves those in a directory called `convertedVideoDir`. As the program was built on old version [Emgu.cv](#) framework, this method used to have -1 called while initiating `VideoWriter` object for the manual selection of coding-decoding format of the video which is now obsolete. Now user can select the video format while calling `CreateVideoFromFrames` method as we introduced `fourcc` for format selection(see Table 6)but the default is `mp4` format.

Table 6: Recreating video on predicted frames by using CreateVideoFromFrames method of NVideo class.

```
public static void CreateVideoFromFrames(List<NFrame> bitmapList,
string videoOutputPath, int frameRate, Size dimension, bool isColor,
char[] codec = null )
{
int fourcc = VideoWriter.Fourcc(codec[0], codec[1], codec[2], codec[3]);
using (VideoWriter videoWriter = new("videoOutputPath.mp4", fourcc,
(int)frameRate, dimension, isColor))
}
```

We also have calculated the accuracy on the training dataset as well as the testing data set. The accuracy is calculated using equation:

$$accuracy = \frac{\text{Matches Found}}{\text{Number of Frames}} \times 100 \quad (1)$$

The accuracy reaches to saturation and after getting 10 similar accuracy the program moves to next cycle to reduce computational time(see Table 7). If the accuracy is more than 80% then it is recorded with the predicted video.

Table 7: Accuracy check code for both Sequential and Multisequence learning in VideoLearning class.

```
// Accuracy Check
double accuracy;
accuracy = (double)matches / ((double)nv.nFrames.Count - 1.0) *
100.0;
if (accuracy == lastCycleAccuracy)
{
saturatedAccuracyCount += 1;
if (saturatedAccuracyCount >= 10 && lastCycleAccuracy >= 80)
{
}
}
```

### III. RESULT

Before we directly reach to accuracy result first have to understand about the changes we made algorithmically. In the previous version of the code all the prediction were made in a do while loop where a new user won't understand when to put image input after an input iteration and the program used to be at a stand still position without any instruction. As a migration of this code we made the program user friendly starting with all the instructions required to start the program and where to change if required which is given at the starting of the program and where instructions are required to run further(see Table 8).

There was an unnecessary [HelperFunction](#) class previously which required extra memory, we integrated it into the main [VideoLearning](#) class. Also created methods to reduce redundancy and follow dry(Don't repeat yourself) technique like MakeDirectoryIfRequired, GetVideoSetPaths etc. methods.

Most of the previously created functions and methods had missing summary issues and we described those in summary so that a programmer in future using this library can easily instantiate those methods and functions by reading.

Table 8: Initial Instruction running the code with full details of where to change.

```
WriteLineColor($"Hello NeoCortexApi! Conducting experiment
{nameof(VideoLearning)} CodeBreakers" + "\n" +
"This program can take initial information of the training video from
VideoConfig.json" + "\n" +
"If you are training with a new set of videos please place the videos in
the folder name SmallTrainingSet" + "\n" +
"Moreover you also need to give video metadata information in the
VideoConfig.json file" + "\n" +
"To change HTMClassifier configuration use htmConfig.json");

WriteLineColor("Drag an image as input to recall the learned Video or
type (Write Q to quit): ");
```

#### A. Accuracy Check Regarding Succeeding Frames

We had total four sets of video data. The videos can be found in [SmallTrainingSet](#) folder. Each of the type is separated into 4 different folders named as Circle, Line, Rectangle and Triangle. As the calculated accuracy is also logged using the function UpdateAccuracy of VideoLearning class, it is easy to find out where the accuracy got saturated value and how much time it required to reach the saturated value. For accuracy result collection we used 1000 cycles maximum to learn and predict frames.

Table 9: the accuracy result table for TrainWithFrameKey, Sequential Learning.






Data Type	Highest Accuracy	Saturated Accuracy
Circle 	102.857% Stability reached at 185 <sup>th</sup> newborn cycle	102.85714285% Saturation level fixed at 117 <sup>th</sup> cycle
Line 	100% Stability reached at 185 <sup>th</sup> newborn cycle	95.74468085% Saturation level fixed at 447 <sup>th</sup> cycle
Rectangle 	100% Stability reached at 185 <sup>th</sup> newborn cycle	100% Saturation level fixed at 144 <sup>th</sup> cycle
Triangle 	100% Stability reached at 185 <sup>th</sup> newborn cycle	100% Saturation level fixed at 127 <sup>th</sup> cycle

Table 10: The accuracy result table for TrainWithFrameKeys, Multisequence Learning

Data Type	Highest Accuracy	Saturated Accuracy
Circle 	102.857% Stability reached at 177 <sup>th</sup> newborn cycle	102.8571428% Saturation level fixed at 69 <sup>th</sup> cycle



Line	97.87%	97.8723404%
	Stability reached at 177 <sup>th</sup> newborn cycle	Saturation level fixed at 217 <sup>th</sup> cycle
Rectangle	100%	97.1428571%
	Stability reached at 177 <sup>th</sup> newborn cycle	Saturation level fixed at 77 <sup>th</sup> cycle
Triangle	100%	100%
	Stability reached at 177 <sup>th</sup> newborn cycle	Saturation level fixed at 17 <sup>th</sup> cycle

From these table we can easily say that multisequence learning where couple of frames are pushed together in the learning stage reaches to saturation in less iterations than the sequential learning. It is because the multisequence learning reduces frames iteration while learning and SDR patterns can be easily captured. In the sequential learning prediction is done with only one possible outcome. But in case of multisequence learning prediction is done for 3 possible outputs(see Table 11). For this reason highest accuracy in case of sequential learning is higher than the multisequence learning. But considering overall situation regarding computational time, computational resources' requirement, accuracy reaching time and required cycles we can say that multisequence learning for video learning projects are the best solution in between sequential learning and multi sequence learning.

Table 11: Code for prediction for each frame in VideoLearning class.

```
foreach (VideoSet vd in videoData)
{
    foreach (NVideo nv in vd.nVideoList)
    {
        for (int i = 0; i < maxCycles; i++)
        {
            foreach (var currentFrame in nv.nFrames)
            {
                if (lvrOut.PredictiveCells.Count > 0)
                {
                    var predictedInputValues =
cls.GetPredictedInputValues(lvrOut.PredictiveCells.ToArray(), 1);
// If i is for the method TrainWithFrameKey which is sequential Learning
                    var predictedInputValues =
cls.GetPredictedInputValues(lvrOut.PredictiveCells.ToArray(), 3);
// If i is for the method TrainWithFrameKeys which is multisequence
                    Learning
                }
            }
        }
    }
}
```

### B. Accuracy Check Regarding Object Recognition

Now when we go for the prediction of succeeding frames from pre-built images (see Table 5) we also tried to calculate the accuracy of object recognition. That means we tried to find out if we put a circle image (which is extracted from previous runs of the same video set) how much accurately does this algorithm recognizes succeeding circle frames or it recognizes from other three objects' frames. For Sequential learning (see Table 12) we got the system has poor object recognition ability. Most of cases the program

remembers the last video set it has learned. By alphabetic order Triangle is the last video set in our data set. So the program recalls most of the SDR patterns of the triangle and in case of succeeding object movement prediction, the program gives false positive result calling it a triangle object movement behavioral sequence.

Table 12: Object recognition accuracy in sequential learning

Input Picture Sequence	Possible Matches Found
<a href="#">Circle_circle_2.png</a>	60.36% match with Triangle_19 31.95% match with Triangle_27 3.99% match with Circle_1 3.55% match with Triangle_26 3.43% match with Circle_4
<a href="#">Circle_circle_3.png</a>	10.45% match with Circle_1 8.2% match with Triangle_18 8.2% match with Triangle_20 8.2% match with Triangle_26 7.4% match with Triangle_25
<a href="#">Line_line_11.png</a>	13.61% match with Line_30 9.77% match with Line_13 9.42% match with Triangle_10 9.42% match with Triangle_34 5.58% match with Triangle_35
<a href="#">Line_line_22.png</a>	12.02% match with Triangle_28 10.18% match with Line_4 8.35% match with Triangle_27 7.5% match with Rectangle_1 6.84% match with Triangle_17
<a href="#">Rectangle_rectangle_18.png</a>	15.16% match with Triangle_2 14.29% match with Triangle_3 14.29% match with Triangle_10 12.09% match with Triangle_12 10.56% match with Triangle_1
<a href="#">Rectangle_rectangle_28.png</a>	25.41% match with Triangle_5 15.14% match with Triangle_9 9.19% match with Triangle_32 9.06% match with Rectangle_3 7.57% match with Triangle_6
<a href="#">Triangle_triangle_23.png</a>	96.18% match with Triangle_1 27.32% match with Rectangle_1 24.51% match with Triangle_2 7.39% match with Rectangle_2 6.36% match with Rectangle_5
<a href="#">Triangle_triangle_0.png</a>	11.93% match with Triangle_22 11.93% match with Triangle_24 11.7% match with Triangle_23 8.19% match with Circle_1 7.02% match with Triangle_25

In case of Multisequence Learning (see Table 13) the Triangular pattern dominates over all the other three objects. But in case of line we can find a bit better number of guessing the correct object as we took 5 of the possible guesses for every frame. Line accuracy is a little bit better than rectangle and circle because it had the highest number of frames to get learned with. So from our result we can say that mismatching the number of learning frames can be a solution for better object pattern matching in case of video learning projects if we use HTM algorithm. As like a human brain this algorithm recalls the last learned sequence the best and that is why triangle accuracy is the greatest among all

and most of the cases the program guess the frame as a sequence of triangular object movement.

Table 13: Object recognition accuracy in Multisequence learning

Input Picture Sequence	Possible Matches Found
<a href="#">Circle_circle_2.png</a>	34.85% match with Triangle_20 32.58% match with Circle_24 31.82% match with Circle_4 25.76% match with Rectangle_17 14.39% match with Triangle_28
<a href="#">Circle_circle_3.png</a>	39.3% match with Triangle_21 14.17% match with Rectangle_16 13.9% match with Circle_2 12.03% match with Triangle_28 10.43% match with Circle_23
<a href="#">Line_line_11.png</a>	61.24% match with Rectangle_13 60.47% match with Rectangle_21 21.71% match with Line_13 19.38% match with Triangle_0 18.6% match with Line_32
<a href="#">Line_line_22.png</a>	18.63% match with Triangle_29 7.14% match with Triangle_28 4.9% match with Line_24 4.9% match with Line_42 3.92% match with Line_5
<a href="#">Rectangle_rectangle_18.png</a>	20.55% match with Triangle_35 14.68% match with Triangle_11 4.9% match with Triangle_34 4.35% match with Triangle_12 2.94% match with Triangle_3
<a href="#">Rectangle_rectangle_28.png</a>	27.78% match with Triangle_6 21.96% match with Triangle_33 14.81% match with Rectangle_28 12.96% match with Rectangle_4 11.38% match with Triangle_34
<a href="#">Triangle_triangle_23.png</a>	91.45% match with Triangle_2 16.67% match with Triangle_14 14.53% match with Triangle_30 13.68% match with Triangle_3 10.26% match with Rectangle_19
<a href="#">Triangle_triangle_0.png</a>	32.13% match with Triangle_23 29.72% match with Triangle_24 26.51% match with Triangle_25 4.42% match with Rectangle_33 3.61% match with Triangle_22

#### IV. DISCUSSION

As this is the migration of the old Video Learning project we have changed a lot in it's learning and prediction algorithm. We have also divided the VideoLearning library into three useable functions TrainWithFrameKey, PredictImageInput and TrainWithFrameKeys by which one can write any new Video Learning project easily. For the ease of understanding we added summary to every functions

and methods. But still the accuracy of the project incase of separate image input from user is not up to the mark. As the last learned video set for this project is Triangle if we put a frame from the line it recognizes the pattern from triangle frames. This is also like our brain, when we read something most of the cases we can remember the last thing we have learned more accurately than the previously seen things. By the log file we can easily see that the accuracy oscillated in the prediction stage. This is also because HTM forgot some of the frames in the learning stage. We found the optimal forgetting and learning ratio as 1/10. Finding a better ratio requires more research on this project. Also different video sets requires different configuration of HtmConfig class according to video configuration introduced in the [videoConfig.json](#) file.

There are still lot of possible improvements to this video learning project. Currently we took the video codec (coding-decoding) format mp4 but new and improved codec is introduced in the libraries and many more are coming. Finding a suitable codec like wmv or MPEG-4 AVC can improve the video making after the prediction more accurately.

Most of the predicted video has unwanted edges around the object. This happened because we have used System library's Drawing class. It also has a draw back of not being OS independent meaning that this will only run on windows but not in linux. There is a new library called [SkiaSharp](#) which is based on Google's Skia Graphics Library and it is OS independent with more accuracy while building the frames from encoded bits. This can help in reducing the edges and building this program on cloud based or OS independent environments.

#### V. REFERENCES

- Boudreau, L. G. (2018, 5). *Rochester Institute of TechnologyRochester Institute*. Retrieved from RIT Scholar Works:  
<https://scholarworks.rit.edu/cgi/viewcontent.cgi?article=10897&context=theses>
- Dobrick, D. (2021). *Improved HTM Spatial Pooler with*. Retrieved from University of Plymouth:  
<https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/17130/Improved%20HTM%20spatial%20pooler%20with%20homeostatic%20plasticity%20control.pdf?sequence=1&isAllowed=y>
- Hawkins, S. &. (2017, 10 17). *A Theory of How Columns in the Neocortex Enable Learning the Structure of the World*. Retrieved from  
<https://frontiersin.org/articles/10.3389/fncir.2017.0008>
- Lipton, Z. C., & Berkowitz, J. (2015). *A Critical Review of Recurrent Neural Networks*, 2-4.