Learning Invariance from Transformation Sequences

Tik-61.184 -
Special Course in Information Science IV:
Unsupervised Learning

Matti Aksela
2.11.2000

Based on the chapter by
Peter Földiák

Contents

Introduction

- The visual system can reliably identify objects also after changes in
  - Viewing angle
  - Eye position
  - Distance
  - Size
  - Orientation
  - Relative position
- Also deformed objects can be recognized

⇒ The visual system must contain information on such transforms
  - Partially genetically defined
  - Likely also learning involved

From electrophysiological experiments (Motivation)

- Invariance properties of perception may be due to the receptive field characteristics of individual cells in the visual system
  - Complex cells in the primary visual cortex exhibit some invariance to position
  - Cells in higher visual areas of the temporal cortex have much larger receptive fields and show more complex forms of invariance to
    * Rotation
    * Color
    * Size
    * Distance

- For a simple neuron, model patterns ares similar when they are close in the Hamming distance

  ⇒ Any simple transformation can cause very different responses

⇒ More complex artificial neurons or networks of simpler artificial neurons needed to exhibit similar behavior

A solution to the positional invariance problem

- Introduced by Fukushima (1980)
- A network consisting of alternating layers of
  - Feature detectors (“S”, simple)
  - Invariance (“C”, complex)
- Feature detectors are replicated and output detectors of the same feature are pooled in “C” layers
- Idea is consistent with models of complex cells in the primary visual cortex
- The hierarchical structure gives rise to more selectivity, position invariance and also some distortion invariance
- Questions in viability
  - The biological plausibility of backpropagation
  - Fixed connections, weight sharing

A modified Hebbian learning rule

- Detectors of the same feature all connect to the same complex unit by forming “clubs”
  - Simple units can develop individually
  - Connections to a complex unit adaptively made

  $\Rightarrow$ Learning rule needed for adaptive connections

- Synaptic strength is proportional to presynaptic activity (x) and a trace of postsynaptic activity ($\overline{y}$)

$$\Delta w_{ij}^{(t)} = \alpha \overline{y}_i^{(y)} (x_j^{(t)} - w_{ij}^{(t)})$$

, where

$$\overline{y}_i^{(t)} = (1-\delta) y_i^{(t-1)} + \delta y_i^{(t)}$$

- Activity at one moment will influence learning at a later moment through the trace

- One possibility for such a trace to be implemented in a biological neuron would be chemical concentration following activity

Simulation

- The goal is to learn shift invariance
- Network structure:
  - 256 position-dependent line detectors (simple layer)
  - 4 units in the complex layer
  - Complex layer units fully connected to simple layer
- Trained with moving lines, randomly one of the four orientations and two directions
- If a complex unit is exited, its trace gets enhanced for the duration of the sweep

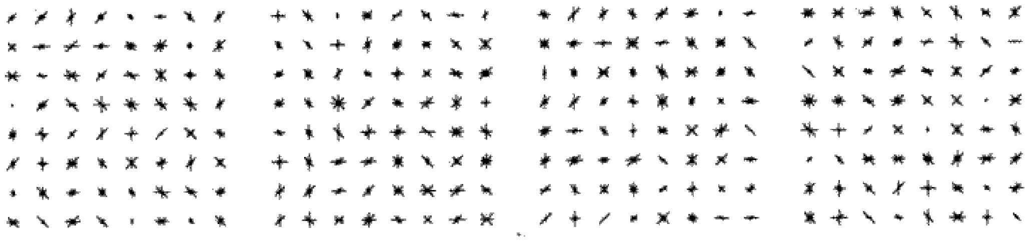
  ⇒ All activated connections strengthened according to the learning rule
- A very simple competitive scheme is used to prevent more than one complex unit from responding to the same orientation

$$y_k = \begin{cases} 1 \text{ , if } \text{argmax}_i(\sum_j w_{ij}x_j) = k \\ 0, \text{ otherwise} \end{cases}$$

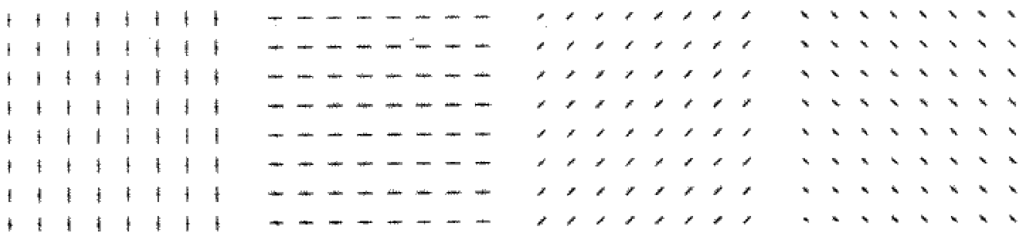
Simulation results

- Five consecutive frames from the training data

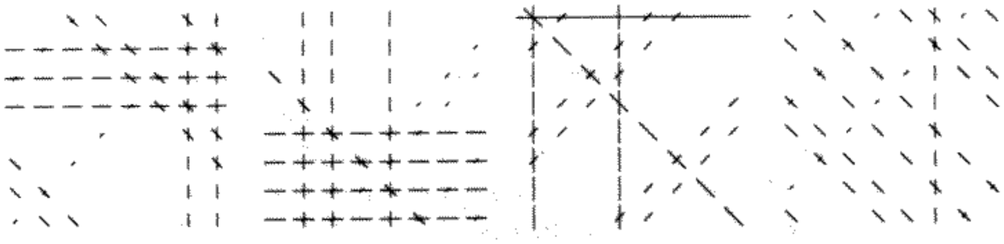
- Connection patterns before training



- Connection patterns after 500 sweeps



- Connection patterns after 500 sweeps without trace



Conclusions

- By using more realistic simple cell model the principle would predict that simple cells of the same spatial frequency and orientation but of different phase tuning connect to the same complex cell

  $\Rightarrow$ Complex cell loses sensitivity to phase and simple cells tuned to different spatial frequencies segregate on different complex cells

- Could be used for 3D cases, when transformation-dependent features are available for input

- Transformation independence is useful for pattern recognition, but may hinder performing visual tasks

  $\Rightarrow$ A "where" system and a "what" system

A comparison [2]

- The trace rule: An unsupervised learning rule which requires information about the activity within a local region but otherwise acts locally within a cell

- The trace rule learning model presented by Földiák evaluated in a comparative study with relation to other learning rules and architectures
  - Outperformed delta rule in handwritten digit recognition task; more rapid convergence and clearly better cross validation set performance
  - Outperformed Fukushima's (1988) supervised Neocognition algorithm with recognition accuracies at 72.5% and 74.5% for the trace method
    * Difference not of statistical significance, but in terms of network size and training time the savings were considerable

- Bearing the biological motivation the trace rule is best suited for a classification task win which one has no external signal to associate any particular image with any particular object class, but in which sequences of images belonging to a single object are viewed in sequences

STNNs [3]

- Temporal-spatial invariance extraction from the point of view of neuro-coding, sparseness produces easily post-processable representations
- A new energy function for the learning paradigm from both temporally self-adaptive behaviors of neurons and the interaction among neurons
- Invariance is extracted by minimizing the energy function
  - Self energy function (cluster internal energy)

$$B(t) = \sum_{m=1}^{M} B_m(t) = \frac{1}{2} \sum_{m=1}^{M} \sum_{j=1}^{J} [T_{mj}(t-1) - y_j(t)]^2$$

  - Mutual energy function (cluster entropy)

$$S(y) = \frac{1}{2} \sum_{j=1}^{J} \ln[1 + y_j^2(t)]$$

$\Rightarrow$ Minimize total energy $E = B + rS$

– If $r = 0$, TNN. If $r > 0$, STNN

TNN with EKF [4]

- A trace learning rule combined with an on-line dual extended Kalman filter algorithm
  - Solves problems with large-varianced input
  - Converges faster than gradient descent
- One EKF combined with the trace learning rule to extract invariance and one to categorize patterns based on that invariance
- To prevent overfitting a model pruning method used to remove insignificant weights

Comments

- Since the visual system possesses great invariance extraction capabilities, emulating it's behavior would be of interest

- The use of a trace appears to be a valid approach to learn self-organizing invariant representations

- The trace approach has been seen to perform quite well [2]

- Performance problems with samples with large variance

  ⇒ Several refinements have been presented
    - ∗ Gradient descent and weight constraints
    - ∗ TNN with EKF [4]

References

1. **Learning invariance from transformation sequences**, Peter Földiák, Unsupervised Learning : Foundations of Neural Computation (Computational Neuroscience Series), MIT Press, 1999, Pages 63–69.

2. **Using Spatio-temporal Correlations to Learn Invariant Object Recognition**, Guy Wallis, Neural Networks, 1996, Volume 9 Issue 9, Pages 1513–1519

3. **Using sparse trace neural network to extract temporal-spatial invariance**, Hanchuan Peng, Qiang Gan, Lifeng Sha, Yu Wei. IEEE International Joint Conference on Neural Networks Proceedings, 1998, Volume 2, Pages 1293 – 1298

4. **Handwritten digit recognition using trace neural network with EKF training algorithm**, Kwok-Wo Wang, Sheng-Jiang Chang, Chi-Sing Leung, The 6th IEEE International Conference on Electronics, Circuits and Systems, 1999 Volume 1, Pages: 295 – 298