HIERARCHICAL TEMPORAL MEMORY

an implementation proposed
On Intelligence

Jeff Hawkins (his ideas)  Sandra Blakeslee (her style)
the last week

• connecting neocortical microcircuits
  – hierarchical temporal memory (htm)
  – mathematical theory of Pearl, applied by Dileep George

• alternatives
  – Laminart, half a century research by Stephen Grossberg

• algorithms
  – cortical learning algorithms
  – laminar computing

• bayesian minds
  – bayesian networks
  – belief propagation

• assignments
  – before exam, on the website (January 15)
  – at the exam, order the topics according to preference
  – assignments published January 27
remember

• mapping from the neocortex onto the network

• hierarchical temporal memory (software 2007)
  • tree-structured
    • with bottom inference
      • from observation to objects
    • with top-down causation
      • from objects to predicted constituents
hierarchical temporal memory

- Layer 4: spatial patterns from children's layer 2/3 with estimates
- Layer 2/3: estimate probability of node from layer 4 estimates and memory
- Layer 2/3: combine feedback pattern with actual pattern
- Layer 5: re-estimate active patterns
- Layer 6: estimate new sequence for children

- with bottom inference
  - from observation to objects
- with top-down causation
  - from objects to predicted constituents
hierarchical temporal memory

- layer 4: spatial patterns from children’s layer 2/3 with estimates
- layer 2/3: estimate probability of node from layer 4 estimates and memory
- layer 2/3: combine feedback pattern with actual pattern
- layer 5: re-estimate active patterns
- layer 6: estimate new sequence for children

- with bottom inference
  - from observation to objects
- with top-down causation
  - from objects to predicted constituents
the reference node
hierarchical temporal memory
the reference node

Node $k$

1. $C$
   - $\lambda^{m_1}$
   - $\lambda^{m_2}$

2. $P(C_t|C_{t-1}, G)$
   - $\lambda$
   - $\pi$

3. $P(C_t|C_{t-1}, G)$
   - Divide
   - $\pi'$

4. $C$
   - $\pi^{m_1}$
   - $\pi^{m_2}$

$y \propto P(-e|C)$

Bel
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_i$</td>
<td>$i^{th}$ coincidence in the node</td>
</tr>
<tr>
<td>$g_r$</td>
<td>$r^{th}$ Markov chain in the node.</td>
</tr>
<tr>
<td>$-e$</td>
<td>Bottom-up evidence. $-e_i$ indicates the evidence at particular instant $i$ and $-e_0$ indicates the sequence of bottom-up evidence from time 0 to time $i$.</td>
</tr>
<tr>
<td>$+e$</td>
<td>Top-down evidence. Time indexing is similar to that of $-e$.</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Feed-forward output message of the node. This is a vector of length equal to the number of Markov chains in the node.</td>
</tr>
<tr>
<td>$\lambda^{in}$</td>
<td>Feed-forward input message to the node from the child node $m_i$. This is a vector of length equal to the number of Markov chains in the child node.</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Feedback input message to the node. This is a vector of length equal to the number of Markov chains in the node.</td>
</tr>
<tr>
<td>$\pi^{in}$</td>
<td>Feedback output message of the node to child node $m_i$. This is a vector of length equal to the number of Markov chains in the child node.</td>
</tr>
<tr>
<td>$y$</td>
<td>The bottom-up likelihood over coincidence patterns in a node. This is one of the inputs for the feed-forward sequence likelihood calculation.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Bottom-up state variable for the Markov chains in a node. This is a vector of length equal to the total number of states of all Markov chains in the node.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>State that combines bottom-up and top-down evidence for the Markov chains in a node. This state variable has the same dimension as that of $\alpha$.</td>
</tr>
<tr>
<td>$Bel(c_i)$</td>
<td>Belief in the $i^{th}$ coincidence pattern in a node.</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pcbi.1000532.t002
the likelihood of coincidence patterns

\[ y_i = P(-e_i|c_i(t)) \propto \prod_{j=1}^{M} \lambda_{ij}^m (r_i^m) \]

where coincidence pattern \( c_i \) is the co-occurrence of \( r_i^1 \)'th Markov chain from child 1, \( r_i^2 \)'th Markov chain from child 2, \( \cdots \), and \( r_i^M \)'th Markov chain from child \( M \).
feedforward likelihood of markov chains

\[ \lambda_t(g_r) = P(-c_0^t|g_r(t)) \sum_{c_{t-1} \in C^3} \lambda_{t-1}(c_{t-1}, g_r) \]  
\[ \lambda_{t}(c_{i}, g_r) = P(-c_{i}|c_{i}(t)) \sum_{c_{j}(t-1) \in C^3} P(c_{i}(t)|c_{j}(t-1), g_r) \lambda_{t-1}(c_{j}, g_r) \]  
\[ \lambda_0(c_i, g_r) = P(-c_0|c_i(t=0))P(c_i(t=0)|g_r) \]
belief distribution over coincidence patterns

\[
\begin{align*}
\text{Bel}_t(c_i) &= \sum_{g_r \in G^t} \beta_i(c_i, g_r) \\
\beta_i(c_i, g_r) &= P(- c_i | c_i(t)) \sum_{c_j(t-1) \in C^t} P(c_j(t) | c_j(t-1), g_r) \beta_{i-1}(c_j, g_r) \\
\beta_0(c_i, g_r) &= P(- c_0 | c_i(t=0)) P(c_i | g_r) \pi_0(g_r)
\end{align*}
\]
feedback message circuit

\[ \pi^{n_i}(g_r) \propto \sum_i I(c_i)Bel(c_i) \]  \hspace{1cm} (9)

where

\[ I(c_i) = \begin{cases} 1, & \text{if } g_r^{n_i} \text{ is a component of } c_i \\ 0, & \text{otherwise} \end{cases} \]  \hspace{1cm} (10)
considerations

• feedback in the model does not affect feedforward paths
  - feedforward results in distribution of predictions at the top
  - feedback identifies features that cause the prediction

• loopy graphs
  - loops will occur when there is overlap in reception
  - accuracy is not proven for loopy graphs
  - experience does not signal a problem
  - (Pearl’s solution is poly-graphs (=multiple parents) and noisy or-gates)

• attention
  - how to focus on one object amidst multiple objects
  - gating is a solution, but how to control
  - also, how to converge to another hypothesis

• duration
  - is there an alternative to uniform time-steps
- from unit delay to event-driven propagation
- external timing unit signals when belief distributions become active

- add gates to trigger $\beta$ calculation
- how to learn duration?
- how to introduce local speed-up and slow-down
core column structure, corresponding to a minicolumn (that is about 100 neurons)
replicated to form a macrocolumn (about 10000 neurons)
specified by adding synapses to perform a specific function
columnar organization of the microcircuit
summary

anatomy features

1. feedforward (thalamic) projection to layer 4
2. layer-4-cell dendrites, mostly within layer 4, make vertical projections to layer 2 and 3
3. layer 3 cells with inter-columnar lateral projections to other layer 2/3 cells, sometimes with outputs to higher cortex regions
4. layer 5 cells with apical dendrites in superficial layer 4 and bottom of layer 3
5. layer 5 cells with apical dendrites in layer 1, sending outputs to subcortical regions and non-specific thalamus
6. layer 6 neurons with apical dendrites in layer 5
7. projections to layer 1 from higher level regions and from non-specific thalamus cells

model computations

1. storage and detection of coincidence patterns
2. bottom-up inputs required for sequence likelihood calculation in equation
3. calculation of sequence likelyhoods for feed-forward and feedback calculations
4. belief calculation without specific timing
5. belief calculation with specific timing
6. computation of feedback messages for child regions
7. high-level input is feedback info; non-specific thalamic input is timing info for markov chains