CSC2515 Lecture 10 Part 2

# Making time-series models with RBM's

## Time series models

- Inference is difficult in directed models of time series if we use non-linear distributed representations in the hidden units.
  - It is hard to fit Dynamic Bayes Nets to high-dimensional sequences (e.g motion capture data).
- So people tend to avoid distributed representations and use much weaker methods (e.g. HMM's).

## Time series models

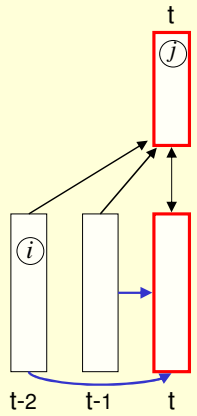
- If we really need distributed representations (which we nearly always do), we can make inference much simpler by using three tricks:
  - Use an RBM for the interactions between hidden and visible variables. This ensures that the main source of information wants the posterior to be factorial.
  - Model short-range temporal information by allowing several previous frames to provide input to the hidden units and to the visible units.
- This leads to a temporal module that can be stacked
  - So we can use greedy learning to learn deep models of temporal structure.

## The conditional RBM model

(Sutskever & Hinton 2007)

- Given the data and the previous hidden state, the hidden units at time t are conditionally independent.
  - So online inference is very easy
- Learning can be done by using contrastive divergence.
  - Reconstruct the data at time t from the inferred states of the hidden units and the earlier states of the visibles.
  - The temporal connections can be learned as if they were additional biases

$$\Delta w_{ij} = s_i \left( < s_j >_{data} - < s_j >_{recon} \right)$$

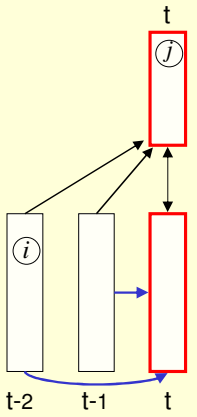


## Why the autoregressive connections do not cause problems

- The autoregressive connections do not mess up contrastive divergence learning because:
  - We know the initial state of the visible units, so we know the initial effect of the autoregressive connections.
  - It is not necessary for the reconstructions to be at equilibrium with the hidden units.
  - The important thing for contrastive divergence is to ensure the hiddens are in equilibrium with the visibles whenever statistics are measured.

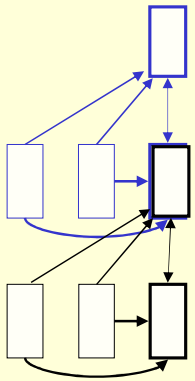
## Generating from a learned model

- The inputs from the earlier states of the visible units create dynamic biases for the hidden and current visible units.
- Perform alternating Gibbs sampling for a few iterations between the hidden units and the current visible units.
  - This picks new hidden and visible states that are compatible with each other and with the recent history.




## Stacking temporal RBM's

- Treat the hidden activities of the first level TRBM as the data for the second-level TRBM.
  - So when we learn the second level, we get connections across time in the first hidden layer.
- After greedy learning, we can generate from the composite model
  - First, generate from the top-level model by using alternating Gibbs sampling between the current hiddens and visibles of the top-level model, using the dynamic biases created by the previous top-level visibles.
  - Then do a single top-down pass through the lower layers, but using the autoregressive inputs coming from earlier states of each layer.



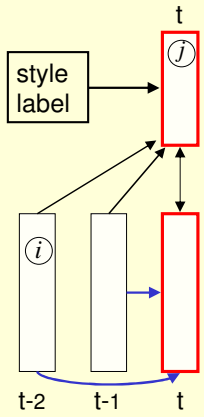
## An application to modeling motion capture data

(Taylor, Roweis & Hinton, 2007)

- Human motion can be captured by placing reflective markers on the joints and then using lots of infrared cameras to track the 3-D positions of the markers.
- Given a skeletal model, the 3-D positions of the markers can be converted into the joint angles plus 6 parameters that describe the 3-D position and the roll, pitch and yaw of the pelvis.
  - We only represent changes in yaw because physics doesn't care about its value and we want to avoid circular variables.

## Modeling multiple types of motion

- We can easily learn to model many styles of walking in a single model.
  - This means we can share a lot of knowledge.
  - It should also make it much easier to learn nice transitions between styles.
- Because we can do online inference (slightly incorrectly), we can fill in missing markers in real time.




## Show Graham Taylor's movies

available at www.cs.toronto/~gwtaylor