Partially Observed Values

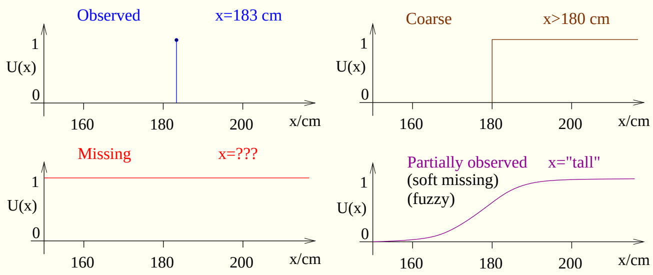
Tapani Raiko
Helsinki University of Technology, Neural Networks Research Centre

Abstract

It is common to have both observed and missing values in data. This paper concentrates on the case where a value can be somewhere between those two ends, partially observed and partially missing. To achieve that, a method of using evidence nodes in a Bayesian network is studied. Different ways of handling inaccuracies are discussed in examples and the proposed approach is justified in the experiments with real image data.

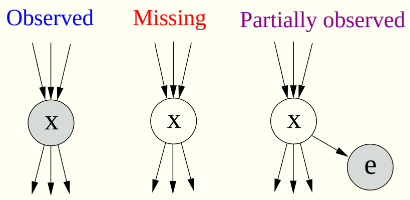
Introduction

- Imperfections in data are common
- A single value can be somewhere between observed and missing
- Variational Bayesian framework called the Bayes Blocks [1] is used here
- Partial observations can be handled by *virtual evidence*
- Figure shows different types of data values in the case of a particular person's height



Evidence approach

- To make a value x partially observed:
  - Leave the value x missing
  - Add an evidence node e as below
  - Make e observed



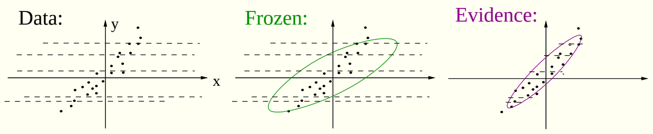
- The options of Gaussian and logistic evidence $U(x)$ are given
- The case $U(x)$ =constant is equivalent to a completely missing value

Frozen approach

- Alternative approach
- Fix a distribution over the data value
- $U(x)$ must be normalisable
  (cannot handle the logistic or constant evidence)

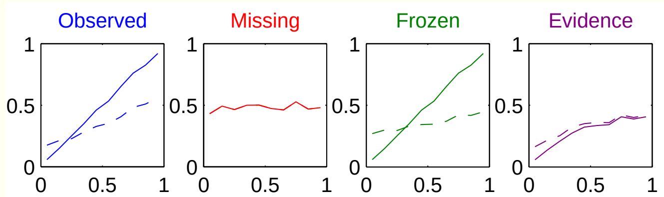
Example

- Factor analysis for (x, y) toy data
- Some of the x values are only partially observed
  (dash lines represent their confidence intervals)
- Frozen approach sticks to the uncertainty of the x and adjusts the model accordingly
- Evidence approach adjusts the uncertain values based on the model



Experiments

- 1000 data vectors $\mathbf{x}(t)$ are 10×10 patches of natural gray-scale images
- Independent Factor Analysis
  $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{b} + \mathbf{n}(t)$
  Super-Gaussian prior for the sources $\mathbf{s}(t)$
- Each pixel has 10% chance of being corrupted with a Gaussian noise (std evenly distributed from 0 to 1; data std is 1)
- Corruption level is assumed to be known for each pixel!
- Four different settings regarding on how the corrupted values are handled
  Observed: Knowledge about corruption is discarded and corrupted values are treated as normal observed data
  Missing: Corrupted values are discarded and regarded as missing
  Frozen: A Gaussian distribution is fixed over each corrupted value
  Evidence: Evidence nodes are used to give a Gaussian virtual evidence



Reconstruction error as a function of the amount of corruption (std)

- Solid lines: $\langle \mathbf{x}(t) \rangle$, the posterior mean of the data variable
- Dash lines: $\langle \mathbf{A}\mathbf{s}(t) + \mathbf{b} \rangle$, the reconstruction without the innovation at the data node
- Evidence approach performs the best at all corruption levels

Conclusion

- Fill the gap between observed and missing values
- Implementation with evidence nodes in a Bayesian network
- Making use of the knowledge about inaccuracies pays off

References

[1] Harri Valpola, Tapani Raiko, and Juha Karhunen. Building blocks for hierarchical latent variable models. In *Proc. 3rd Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, pages 710–715, San Diego, USA, 2001.