Random Fields and Networks

Nagesh Subbanna

Generative Networks

- Generative networks model the distribution of each class.
- Generative model learns the joint probability.
- P(X|Y) and P(Y) are learnt to obtain P(Y|X)
- Bayes rule is used compute P(Y|X).

Generative Networks vs Discriminative Networks

- Generative networks model the joint probability.
- Discriminative networks model the conditional probability.
- Generative models learn the entire model.
- Discriminative models learn the difference.

Generative Networks-Examples

- Naive Bayes
- Linear Discriminant Analysis
- Bayesian Networks
- Markov Random Fields
- Hidden Markov Models

Discriminative Networks-Examples

- Logistic regression
- Scalar Vector Machine
- Traditional neural networks
- Nearest neighbour
- Conditional Random Fields (CRF)s

Labelling Problem

- $S=(i_1, i_2,\ldots,i_n)$ – set of n nodes, $L=(l_1, l_2,\ldots,l_m)$ – set of m labels, F:S->L – map sites to labels
- Problem: Map the n sites to the m labels, based on their characteristics.

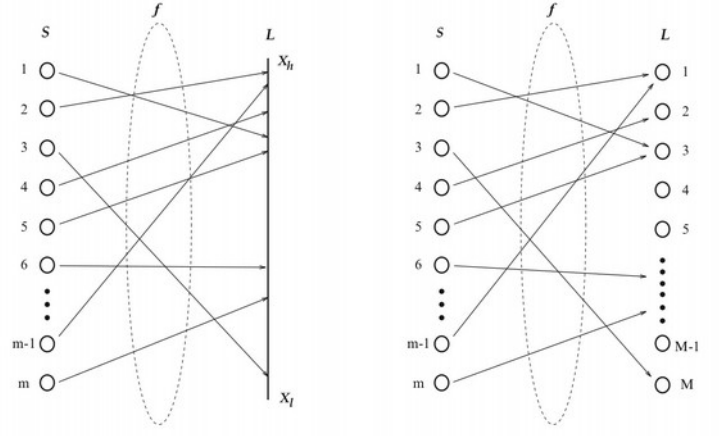


Figure 1.1: A labeling of sites can be considered as a mapping from the set of sites $\mathcal{S}$ to the set of labels $\mathcal{L}$. The Figure shows mappings wa continuous label set (left) and discrete label set (right).

Labelling Problem

- Space of labels is huge (LxL…xL) n times – L^n
- Impossible to check all of them
- Need constraints to ensure a smaller space.

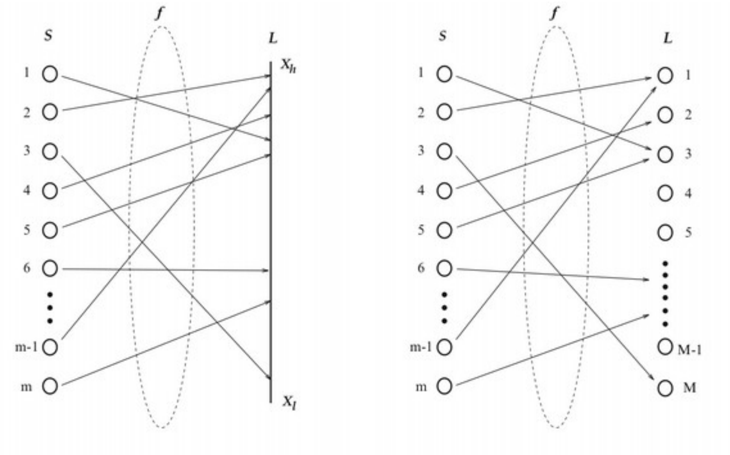


Figure 1.1: A labeling of sites can be considered as a mapping from the set of sites $\mathcal{S}$ to the set of labels $\mathcal{L}$. The Figure shows mappings wa continuous label set (left) and discrete label set (right).

Labelling Problem

- Four kinds of problems
  - Regular sites with continuous labels
    - Eg: Image restoration/image smoothing
  - Regular sites with discrete labels
    - Pixel/Voxel classification
  - Irregular sites with discrete labels
    - Object labelling
  - Irregular sites with continuous labels
    - Pose estimation

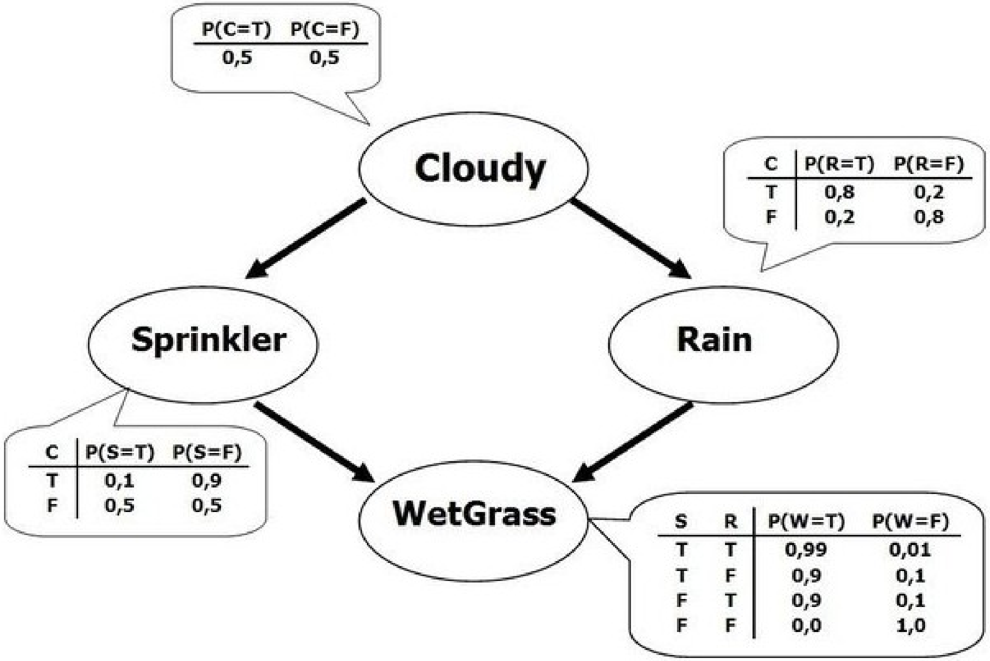
Basic Assumptions

- Positivity – probability of every label at every site is positive.
- Markovianity – probability of the label at every site is dependent on pre-defined neighbourhood.
- Memoryless system.

Random Fields

- Given a probability space (W, F, P), an x-valued random field is a collection of x-valued random variables on a topological space T.
- T – topological space
- F – space of F_t, F_t where is an x-valued random variable.
- P – assignment of probabilities to events.

Bayesian Networks

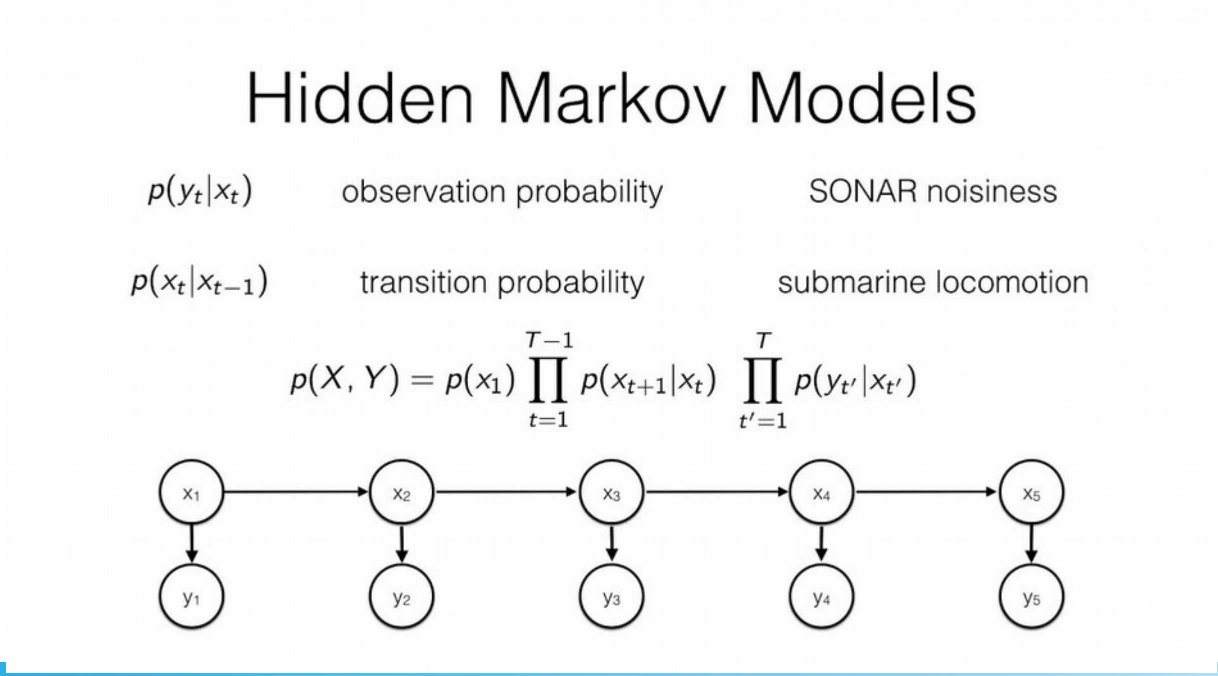


Bayesian Networks

- Work with directed graphs
- Causation instead of correlation.
- Joint probability distribution needed for each node separately.

Hidden Markov Model

- Directed network
- Limited memory system.
- Model the transition between each state



Markov Random Field

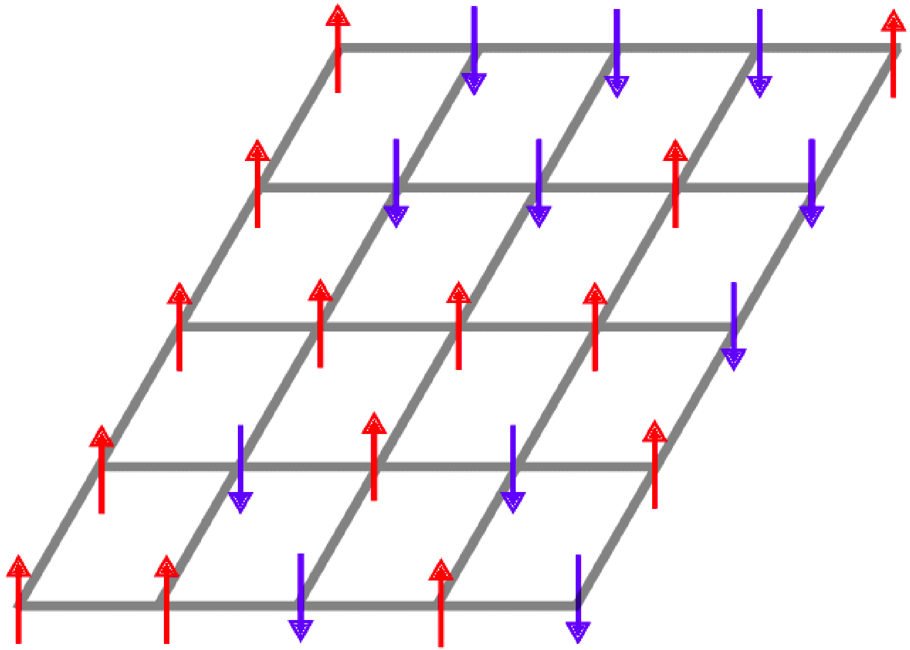
- Works on undirected graphs.
- Implies correlation than causality
- Given a set of sites S, and m labels, infers a mapping for each site.
- Global labelling based on local information.

Historical Sidelights

- Initiated by the Ising model
- Developed by Besag in the 1970s and 1980s.
- Hammersley-Clifford theorem gave a simplification.
- Geman and Geman gave MAP-MRF model.

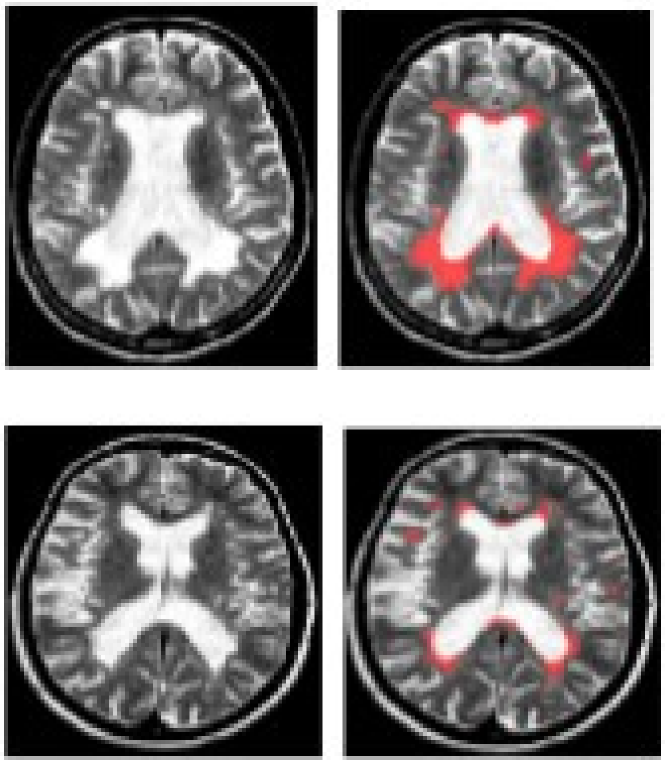
Ising Model

$$H = -\sum_{\langle i\, j \rangle} J \sigma_i \sigma_j - \sum_j h \sigma_j ,$$



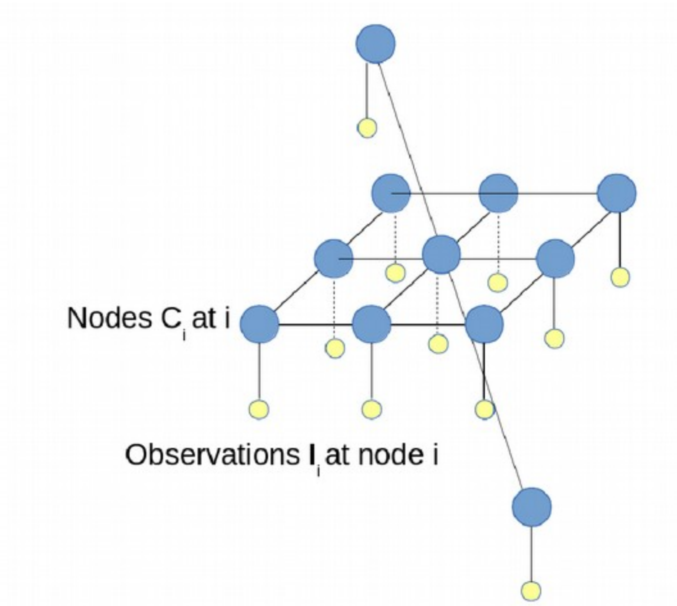
Example – MS Lesions

- Detect MS lesions in brain MRI
- MS lesions come in many shapes, sizes, positions – so hard to detect
- Juxta cortical lesions count towards lesion count, but are hard to catch.



T2 Expert

Basic Markov Model



$$P(f|I) = \frac{1}{\sum_{f \epsilon F} exp(-\frac{1}{T}U(f))} exp(-\frac{1}{T}U(f)),$$

Hammersley-Clifford Theorem

- A probability distribution that has a strictly density satisfies the Markov properties with respect to an undirected graph G if and only if it is a Gibbs random field.
- It can be factorised over its subgraphs.

$$P(X = x) = \frac{1}{Z(\beta)} \exp(-\beta E(x)).$$

Elements of MRF

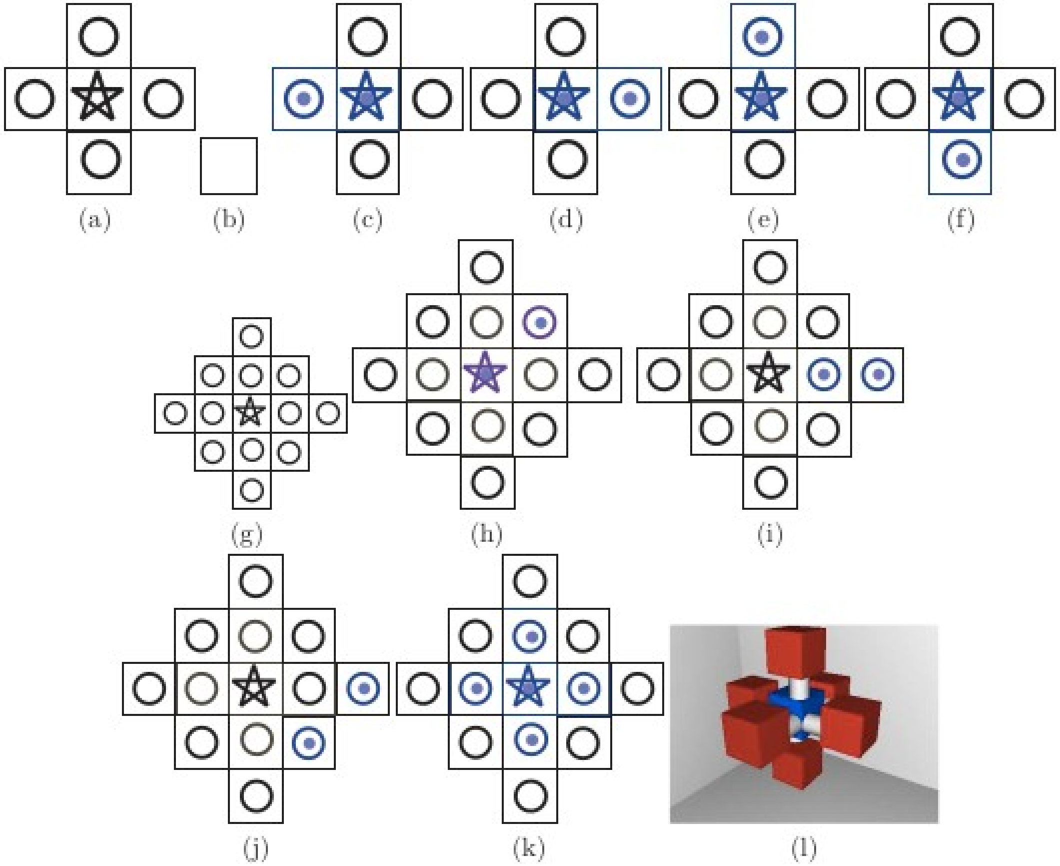
- Elements to learn
  - P (C_i) – prior probability of the class
  - P (I|C) – probability of the class given intensity.
- Elements to model
  - Energy of attraction

$$U_{posterior} = U_{prior} + U_{likelihood}.$$

Assumptions

- Distributions at individual voxels do not have to be modelled.
- The models do not change in different parts of the image.
- Attraction in all parts of the image are the same.

Neighbourhoods



Region Based MRFs

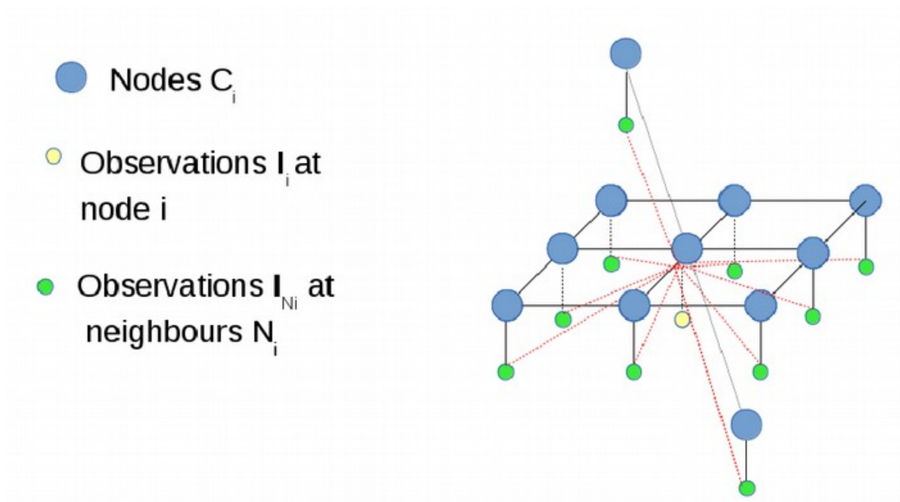
- No need for uniform models everywhere.
- Divide the image into regions.
- Use different models on each part.

$$p(\mathbf{I}_v | C_v, x_v, y_v, z_v) = p(\mathbf{I}_v | C_v, R_v),$$

Advantages and Disadvantages

- Advantages
  - Easy to train.
  - Requires relatively small data
  - Computationally efficient.
- Disadvantages
  - Oversmoothing
  - Small regions get wiped out.
  - Given MRF nature of attracting like labels, small regions are very hard to segment.

Adapted Markov Models



$$U(C_i|\boldsymbol{I_i},\boldsymbol{I}_{\boldsymbol{N_i}}) = -\left[\underbrace{P(C_i)}_{Prior} + \underbrace{P(\boldsymbol{I_i}|C_i)}_{voxel\ intensity} + \underbrace{\sum_{k\in cliques(N_i)}\sum_{j\in k}\log P(\Delta\boldsymbol{I_{i,j}}|C_i\boldsymbol{C_j})}_{Intensity\ differences}\right.$$

$$\left. + \alpha\,\underbrace{m(C_i,\boldsymbol{C_j})}_{class\ similarity}\right]$$

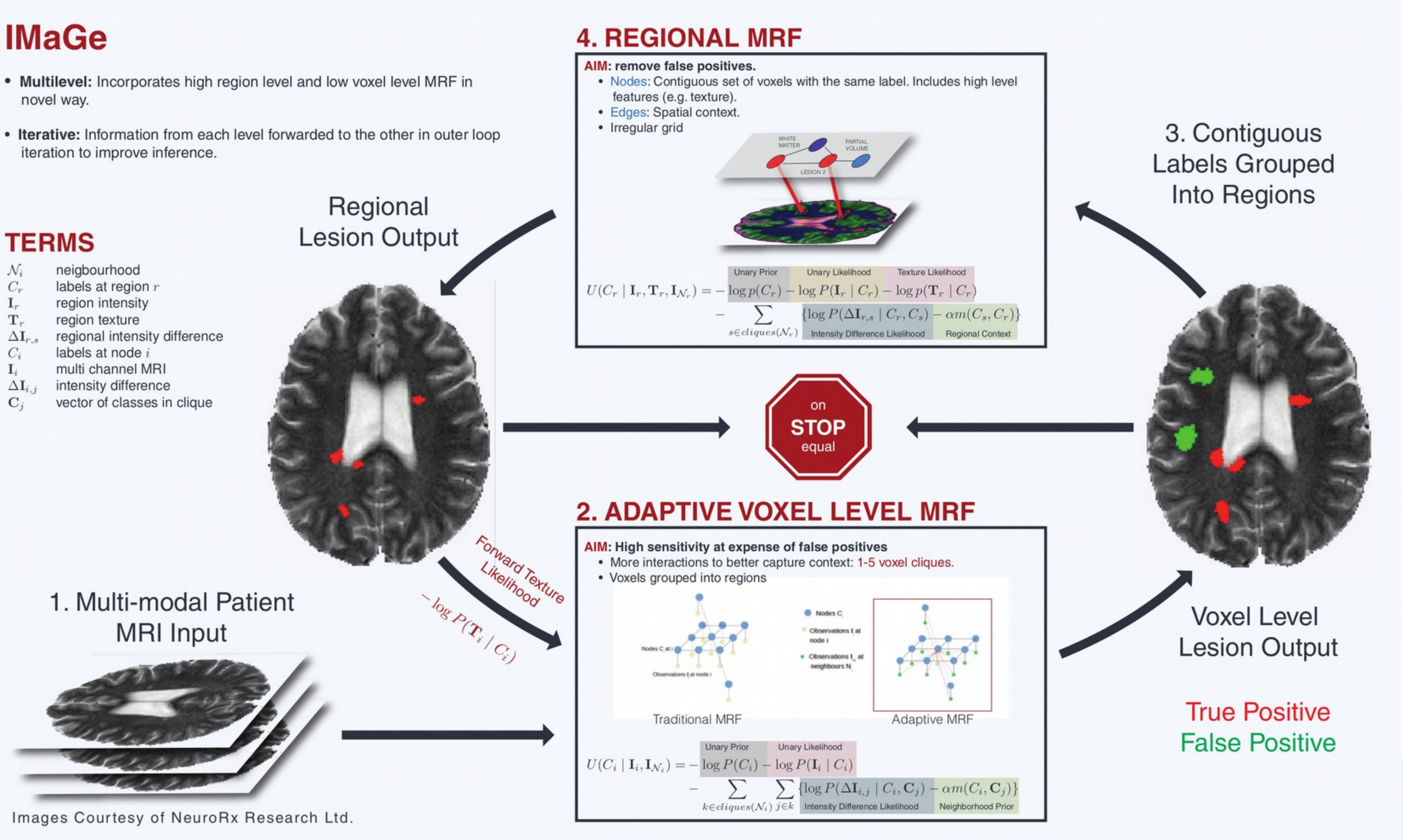
New Elements

- Assumptions
  - Intensity gradient does not depend on actual intensities.
  - Voxel intensity does not depend on neighbouring classes
- Elements to learn
  - Every intensity contrast distribution for every pair of classes has to be learnt.

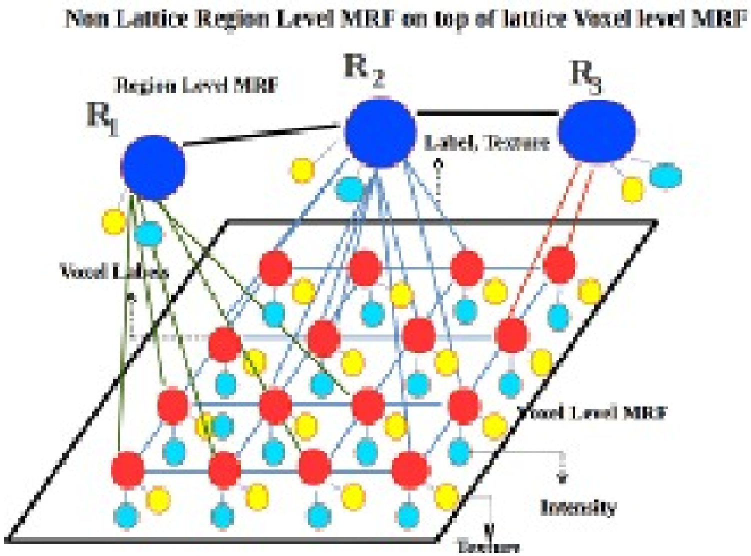
Advantages and Disadvantages

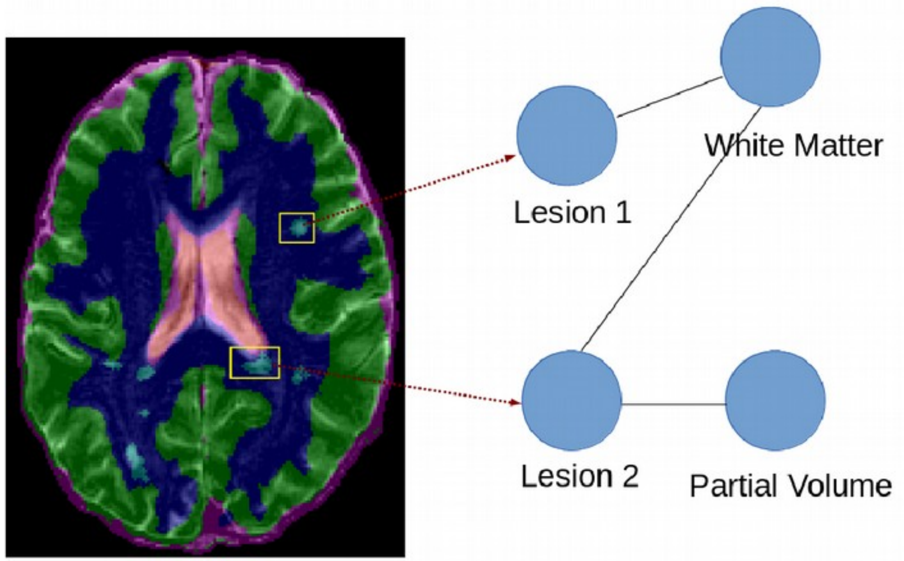
- Advantages
  - Intensity contrast behaves like an edge detector, so small regions are conserved.
  - Sharper boundaries, because smoothing is inhibited.
- Disadvantages
  - Still local
  - Can enhance false noise

Hierarchical MRFs



Irregular MRFs





$$U\langle C_i|I_j,\Delta I_{N_j} T_j\rangle = -\left[\underbrace{P(C_i)}_{Prior} + \underbrace{P\langle I_j|C_j\rangle}_{regional intensity} + \underbrace{P\langle T_j|C_j\rangle}_{Regional Texture} + \underbrace{\sum \log P\langle \Delta I_{j,k}|C_j,C_k\rangle}_{RegionalInt.differences}\right]$$

$$+\underbrace{\alpha m(C_j,C_k)}_{class similarity}$$

New Elements

- Region created from connected components
- Each region is a node, so observations are intra-nodal.
- All observations are regional.
- All nodes are irregular.

Effect of regional MRFs

- Regional MRFs provide a regional feedback to local MRFs.
- Use of textures and regional features to determine class regions.
- More features than what is available from intensity based classification.
-

Stopping Criterion

- Convergence is not guaranteed.
- May result in oscillation of labels.
- The process may have to be forcibly stopped after some time.
- Usually controlled by a limited number of iterations.

Advantages and Disadvantages

- Advantages
  - Use of regional features makes it more accurate.
  - Larger number of different kinds of features makes it robust
  - Mathematically robust
- Disadvantages
  - Computationally demanding
  - Convergence not guaranteed

Unified Model

- Merge the regional step into the voxel level.
- Compute pre-determined textures as observations at the voxels.

$$\begin{aligned} U(C_i \mid \mathbf{I}) &= -\log P(\mathbf{I}_i | C_i) - \log(\mathbf{T}_i \mid C_i) - \log P(C_i) \\ &- \sum_{k \in \text{Cliques}(N_i)} \sum_{j \in k} \left(\log P(\Delta \mathbf{I}_{i,j} | C_i, \mathbf{C}_j) - \alpha m(\mathbf{C}_j, C_i)\right), \end{aligned}$$

Advantages and Disadvantages

- Advantages
  - Convergence guaranteed.
  - Computationally less demanding due to fewer comparisons necessary
  - Uses regional information in a local regular MRF.
- Disadvantages
  - Pre-determined textures and feature sizes.
  - Violates the spirit of MRFs in a sense.

Conclusions

- A definition of generative models
- Different kinds of generative models.
- Differences with the discriminative models.
- An overview of the different kinds of MRFs
- Advantages and disadvantages of the different kinds of MRFs.

Bibliography

- Koller and Friedman ``Probabilistic Graphical Models''
- Stan Li, ``Markov Random Field Modelling in Image Analysis''
- Julian Besag, ``Spatial Interaction and the Statistical Analysis of Lattice System''