Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines

Chi-Hoon Lee[1], Mark Schmidt[1], Albert Murtha[2], Aalo Bistritz[2], Joerg Sander[1], and Russell Greiner[1]

[1]Dept. of Computing Science, The University of Alberta

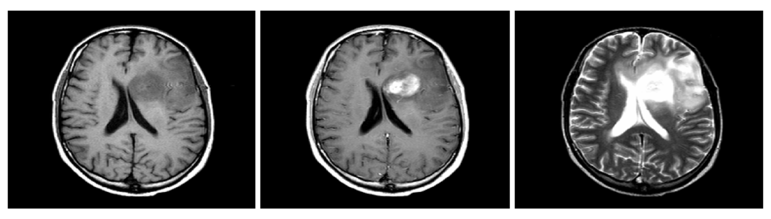
[2]Dept. of Oncology, Cross Cancer Institute, Edmonton, AB, Canada

Introduction

Task:

Segmenting Brain Tumors in MR images

- Input: T1, T1c, T2 images



Left to right: T1, T1 with contrast agent, T2 image

- Output:
  - Edema, Enhancing and Gross Tumor areas

Left to right: Edema, Enhancing, Gross Tumor areas

Motivation:

- Want accurate segmentation
- Considered using effective classifier – SVM
- But SVM assumes data is iid, but our imaging data is not
  - adjacent voxels typically have same labels

Goal:

- Synthesize SVM-ideas into "Random Field" classifier

Background

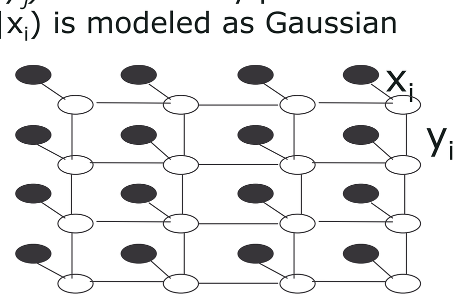
1a. Markov Random Field (MRF)

- Allows the label of one pixel to depend on the labels of neighboring pixels
- ***Generative*** approach: computes P(Y|X) using the Baye's rule for features of set of pixels $X = \{x_1, . . . , x_n\}$ and labels $Y=\{y_1, . . . , y_n\}$

$$P(Y \mid X) \propto p(X \mid Y)\,p(Y) = \exp\left[\sum_{i=1}^{n} \log(p(y_i \mid x_i)) + \sum_{j \in N_i} V(y_i, y_j)\right]$$

General MRF model:

- N_i is neighboring pixels of i
- $V(y_i,y_j) \geq 0$ is arbitrary potential function having the same class label as neighbors
- $p(y_i|x_i)$ is modeled as Gaussian



- Shaded nodes are observed $\{x_i\}$
- Unshaded nodes are unobserved labels $\{y_i\}$
- Edges between nodes indicate dependences

- Issues:
  - Must compute joint probability
  - To be tractable, uses problematic independence assumption: $p(X \mid Y) = \prod_i p(x_i \mid y_i)$
  - Cannot model complex dependencies between features and labels

1b. Discriminative Random Fields (DRF)

- ***Discriminative*** (not generative)
  - Directly model P(Y|X) -- the posterior probability of labels given features

$$P(Y \mid X) \propto \exp\left\{\sum_{i=1}^{n} A(y_i, X) + \sum_{j \in N_i} I(y_i, y_j, X)\right\}$$

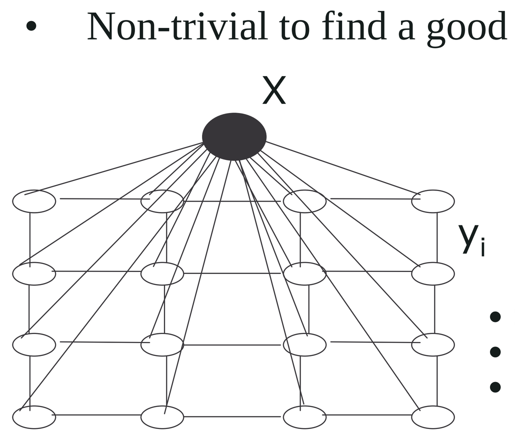
General CRF model:

- N_i is neighboring pixels of i
- $A(y_i, X)$ is "Association" (Observation-Matching) potential
- $I(y_i, y_j, X)$ is "Interaction" (Local consistency) potential

- Can use $A(y_i, X)$ to model complex dependencies between (features of) pixel and its label
- Can use $I(y_i, y_j, X)$ to model complex dependencies between (features of) neighboring pixels and their labels
  - In MRF: only model spatial correlation by only considering labels of adjacent pixels
- DRF uses
  - ✓ Logistic Regression for Observation Matching potential, $A(y_i, X)$
  - ✓ Linear function for the Local-Consistency potential, to model spatial dependencies: $I(y_i, y_j, X) = y_i y_j v^T \varphi(x_i,x_j)$
- Conditional Random Field (CRF) is a simple 1D version of a DRF

- Issues:
  - Simultaneously learning both $A(\cdot)$ and $I(\cdot)$
    ➔ Possible inappropriate spatial dependences modeling
  - Correlated high dimensional data feature space
    ➔ Inappropriate parameter estimation
  - Non-trivial to find a good initial labeling for inference

Challenges a DRF must address, which are solved by Support Vector Machines !

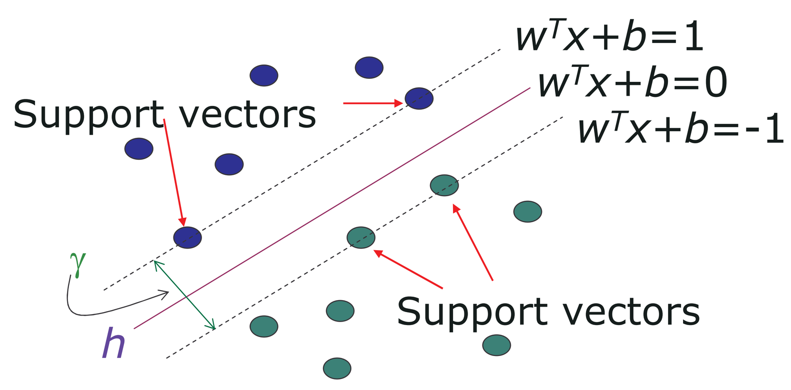


- Shaded nodes are observed $\{x_i\}$
- Unshaded nodes are unobserved labels $\{y_i\}$
- Edges between nodes indicate dependences

2. Support Vector Machines (SVM)

- A popular tool for learning classifiers of *iid* data
- Creates a hyperplane *h* that separates the data into two classes
  - Maximizing the margin γ, where "margin" = distance from closest data object to the hyperplane
- Less sensitive to class imbalance than Logistic Regression

- Issues:
  - Use iid assumption
    ➔ Spatial correlation not considered



Decision function: $f(x)=w^Tx+b$
$f(x)$ = distance from x to the hyperplane perpendicular to w
Class label = $sign(f(x))$

SVRF (Support Vector Random Field)

- Extend DRF by basing $A(y_i, X)$ on SVM, not LR
- Less sensitive to unbalanced data than MRF and DRF
- More efficient learning method
  ➔address the disadvantage of DRF's simultaneous parameter learning

$$P(Y \mid X) \propto \exp\left\{\sum_{i=1}^{n} \log(O(y_i, x_i)) + \sum_{j \in N_i} V(y_i, y_j, X)\right\}$$

$$O(Y_i = 1, x_i) = \frac{1}{1+\exp(-K \times f(x_i) + B)}$$, where K and B are estimated from training data

$V(y_i, y_j, X) = y_i y_j v^T \Psi(x_i, x_j)$

- A linear Local-Consistency potential using 8 adjacent pixels as the neighborhood system

Experiments for Brain Tumor Segmentation.

➢**Data Sets:** 7 Patients

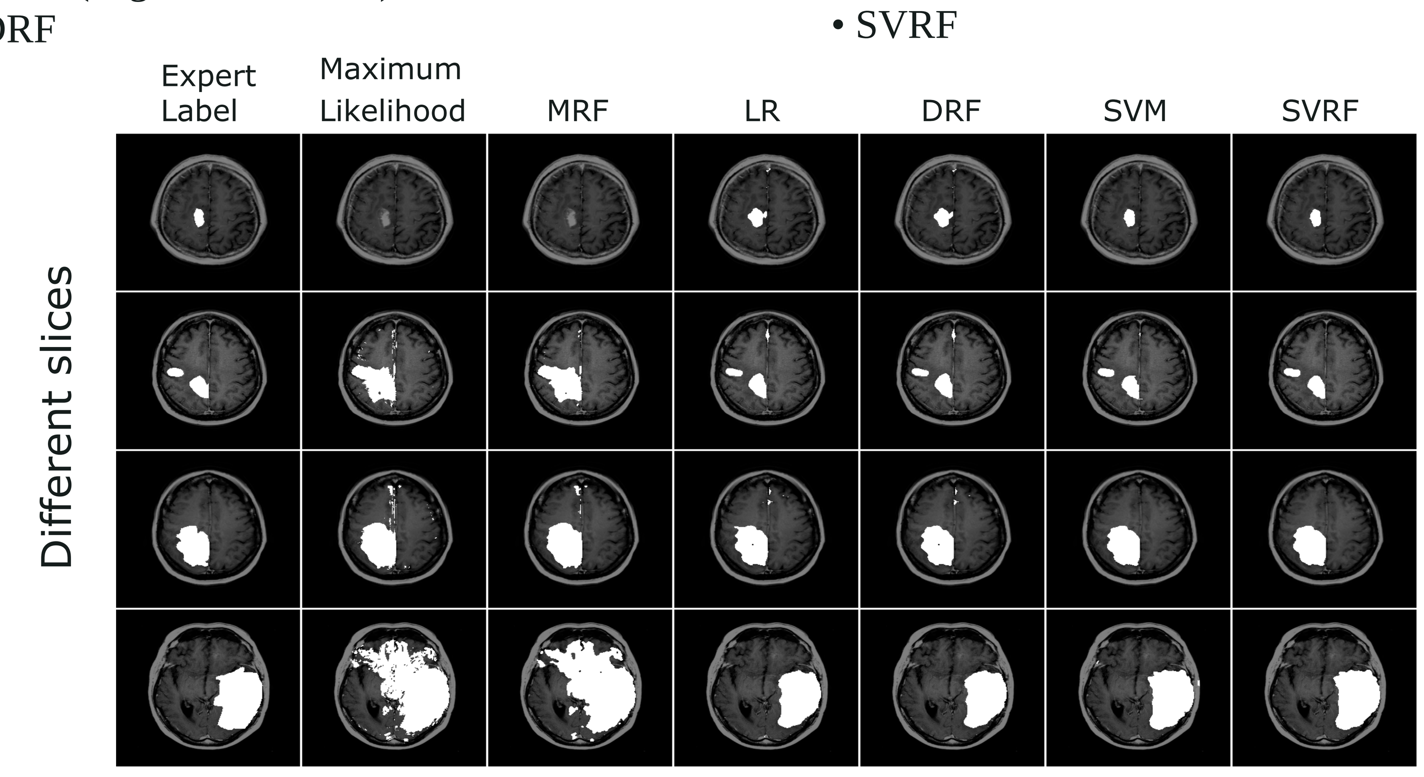
- Each with one of
  - grade 2 astrocytoma
  - anaplastic astrocytoma, or
  - aglioblastoma multiforme
- Pre-processed to reduce noise, inter-slice variations, and intensity inhomogeneity with spatial registration

➢ **Systems considered:**

- Maximum Likelihood classifier (degenerate MRF)
- SVM (degenerate SVRF)
- DRF
- Logistic Regression model (degenerate DRF)
- MRF
- SVRF

➢**Patient Specific Training and Testing**

- For each patient #i:
  - Train on data from slices 1 and 3 of patient #i
  - Test on data from slices from 2 of patient #i



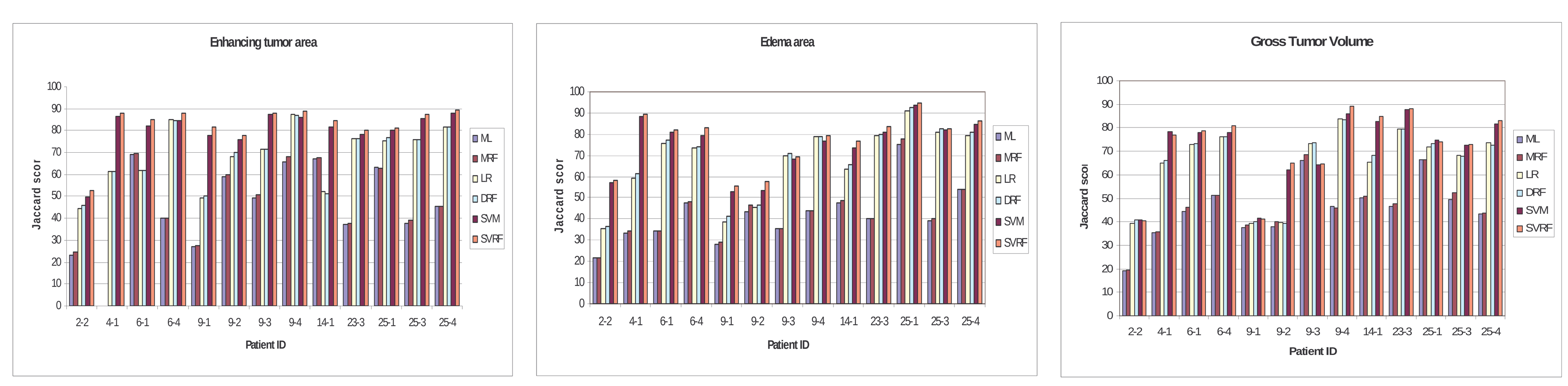
Classification results of *enhancing tumor areas* for 4 different test slices

➢**Evaluated by Jaccard score**

- J = TP/(TP+FP+FN)
- true positive (TP), false positive (FP), false negative (FN)

➢**Results in summary**

- ➢ Overall, ML<MRF<LR<DRF<SVM<SVRF
- ➢ Statistical significance (paired t-test)
  - SVRF is better than SVM at p<4.25E-11



Jaccard score for 3 different tumor segmentation tasks: Left to right: Enhancing, Edema, and Gross Tumor areas

Conclusions

- Remaining issue for SVRF: Efficiency (learning, inference)
- Explored algorithms for segmenting brain tumor: both iid and with spatial correlations
- Standard Random Field algorithms often perform better than iid classifiers
- SVRF shows the best overall performance

ALBERTA INGENUITY CENTRE FOR MACHINE LEARNING   NSERC CRSNG   ALBERTA INGENUITY FUND

http://www.cs.ualberta.ca/~btgp