Cross Entropy: A New Solver for Markov Random Field Modeling and Applications to Medical Image Segmentation

Presentation by Warren Cheung for CMPT880

Jue Wu and Albert C.S. Chung

Hong Kong University of Science and Technology

November 23, 2005

Overview

Purpose

Use **Cross Entropy** as a *general,stochastic* way to find an "optimal" solution for Markov Random Fields problems in Image Segmentation.

Markov Random Fields

- "synonymous" with Markov networks
- encode spatial information
- probability of an event at a position is dependant on neighbours
- *powerful*: can do the equivalent of smoothing and edge detection simultaneously
- *difficult to solve*: global optimum is often hard to find ("analytically intractable")

Solving Tough Problems

- *graph cuts*: polynomial time, but only if you can create a graph of the problem
- *simulated annealing*: "stochastic local search" — high temperature (randomness) to low temperature, but only optimal if you "cool" very slowly
- *belief propagation*: deterministic approximate solution, via conversion to a graph and propagating belief from a leaf "upwards"

Cross Entropy

- stochastic
  - Random
  - Longer you run, better the result
  - "Efficient"
- based on "well-established"/sound mathematical theory of rare events
- *Reuven Rubinstein*, Technion—Israel Institute of Technology, 1997

Rare Events

- want to get probability of rare event
- could do normal sampling
- Crude Monte Carlo
- random samples $X_1, \ldots, X_N$ from distribution X and the rare event $S(X_i) \leq \gamma$

$$\frac{1}{N}\sum_{i=1}^{N} I_{S(X_i)\leq\gamma}$$

  - if event is rare, this could take many samples
  - time-consuming, inaccurate

Importance Sampling

- can use importance sampling
  - aka biased sampling
  - sample the region near the rare event
  - adjust weight of results accordingly
- X_i is drawn from a pool with pdf $g(x)$, f(x) is the original distribution (i.e. flat)

$$\frac{1}{N}\sum_{i=1}^{N} I_{S(X_i)\leq\gamma} f(X_i)/g(X_i)$$

- to get better accuracy, $g(x)$ has to sample near the rare event
- cross entropy provides an iterative way of finding this region in configuration space

The Math

Want to optimise:

$$f^* = \operatorname*{argmin}_{f \in \mathcal{F}} E(f)$$

Do this by estimating biased probability of a “rare event”:

$$\mathcal{P}_v(E(F) \leq e) = \sum_f I_{\{E(f) \leq e\}} p(f; v)$$

where F is a vector of configurations with pdf $p(\cdot; v)$

- connect the equations by looking for $\mathcal{P}_v(E(f) \leq e)$ when $e = E(f^*)$ and the pdf is uniform
- rare event — global minimum is often unique

$$\mathcal{P}_v(E(f) \leq E(f^*)) = 1/|\mathcal{F}|$$

Cross Entropy Algorithm Overview

$$\mathcal{P}_v(E(F) \leq e) = \sum_f I_{\{E(f) \leq e\}} p(f; v)$$

- pretend you want to estimate $\mathcal{P}_v(E(F) \leq E(f^*))$ under a flat distribution for $p(\cdot; v)$ — use importance sampling
- when e is close to $E(f^*)$, a good sampling $p(\cdot; v)$ favours configurations near f^*
- *but* the closer e gets to $E(f^*)$, the harder it is to get good samples $\Rightarrow$ the better your choice of v must be
- cross entropy technique:
  - iteratively decrease e while improving v
  - power is that (for a well chosen family of pdfs) you can *analytically* choose the best possible v at each step

Image Segmentation Problem to solve:

- Label n sites using m labels.
- sites $\Leftrightarrow$ pixels
- labels $\Leftrightarrow$ number of different segments

CE Algorithm for MRF Energy Minimisation

Initialisation

- Set level $t = 1$
- The initial parameter vector $v_0 = \{v_{0,1}, \ldots, v_{0,n}\}$
- Each $v_{t,i} = \{v^1_{t,i}, \ldots, v^m_{t,i}\}$ is a vector with m elements for site i.

CE Algorithm for MRF Energy Minimisation

Main Loop

- Generate a collection of samples $F_1, \ldots, F_N$ ($F = \{f_1, ..., f_n\}$ is one MRF configuration) from the density $p(\cdot; v)$ and compute the energy $E_i = E(F_i)$ for every $i \in \{1, ..., N\}$.
- Sort all the E_i in a non-increasing order to $\{E_1, ..., E_N\}$. Then pick $e_t = E_{\lceil(1-\rho)N\rceil}$.
- Use the samples $F_1 \ldots F_N$ to update v_{t+1} using

$$v^j_{t+1,i} = \frac{\sum_{k=1}^{N} I_{\{E_k(F_k) \le e_t\}} I_{\{F_{ki}=j\}}}{I E_k(F_k) \le e_t}$$

  for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.
- If e_t remains unchanged for several iterations, terminate loop; else, set $t = t + 1$ and repeat.

CE Algorithm for MRF Energy Minimisation

Termination

- The final $E_N(F_N)$ of T-th iteration is the estimated minimal MRF energy.

Notes on the Algorithm

- $p(\cdot; v)$ can be any probability distribution — in this case m-point Bernoulli is chosen (i.e. a label j is chosen for site i with probability $v^j_{t,i}$ at level t)
- parameters ρ and N
- ρ is the rarity of the event — tunable parameter, usually 1% to 10%. When n is large, choose ρ to be larger.
- N is the sample size. They choose $N = cn$, where c is between 1 and 10.
- alternative stopping criteria exist, such as v_t converging to a binary vector

The MRF Function

- maximise posterior $P(X|Y)$ where $P(X|Y) \propto P(X)P(Y|X)$
- X is labelling MRF, Y is "observed" (labelled) data
- multi-level logistics (MLL) and Gaussian distribution
- MRF Prior Energy:

$$U(X) = \sum_{i \in S} \sum_{j \in N_i} I_{\{x_i \neq x_j\}}$$

  where N_i is the neighbourhood (4 neighbours) and S is the set of all sites.
- Likelihood energy:

$$U(Y|X) = \sum_{i \in S} \sum_{j=0}^{m-1} I_{\{x_i = j\}} \cdot \frac{(y_i - \mu_j)^2}{2\sigma^2}$$

  where μ_j is the mean intensity of region j

Synthetic Test I

- Binary Segmentation of synthetic “Blood Vessels”
- Varying levels of noise, ground truth is known

Synthetic Test I — Results

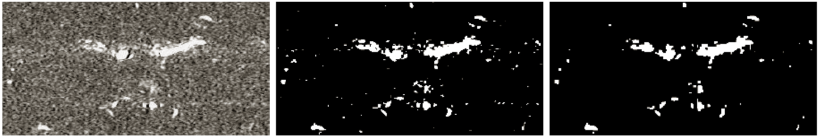
		BP		CE	
image	SNR	minimum	error	minimum	error
bar (width=3)	2	2215505	18.65%	1656942±7707	9.01%±1.53%
	3	1848604	7.54%	1577403±17069	3.52%±0.69%
	4	1598662	2.25%	1523299±17271	1.44%±0.96%
	5	1526652	0.59%	1501826±2126	0.22%±0.28%
bar (width=6)	2	2101877	19.29%	1319552±4612	1.20%±0.37%
	3	1659417	8.30%	1242550±1923	0.59%±0.20%
	4	1408863	3.13%	1232195±2309	0.03%±0.06%
	5	1250346	0.68%	1216728±2223	0.07%±0.12%
circle (width=3)	2	2241872	17.58%	1721610±6633	11.45%±1.04%
	3	2000580	8.59%	1691143±3865	6.72%±0.54%
	4	1817165	4.49%	1672465±2259	3.37%±0.40%
	5	1774078	2.34%	1675028±926	1.91%±0.20%
circle (width=6)	2	2178246	18.65%	1348983±3880	4.45%±1.11%
	3	1715610	8.30%	1295023±1162	1.98%±0.25%
	4	1516673	4.39%	1278273±3337	1.86%±0.08%
	5	1380865	1.86%	1269815±634	1.37%±0.08%

Synthetic Test II

- segmentation of BrainWeb T1 data
- 4 classes: white matter, grey matter, cerespinal fluid, other
- noise added

	BP		CE	
noise level	minimum	error	minimum	error
3%	348740	4.25%	211878.3 ± 2737.2	$2.77\% \pm 0.08\%$
5%	488391	6.00%	257021.0 ± 1505.7	$4.14\% \pm 0.06\ \%$
7%	487083	6.58%	229624.5 ± 1184.7	$3.57\% \pm 0.02\%$
9%	691409	11.65%	525039.5 ± 768.6	$9.08\% \pm 0.01\%$

- PCMRA: Phase contrast magnetic resonance angiographic images
- get lower mean energy
- no ground truth



Clinical Cerebral Data

- from IBSR: Internet Brain Segmentation Repository
- 3 labels: white matter, grey matter, other

	BP		CE	
dataset	minimum	error	minimum	error
1	19361	3.32%	15103 ± 39	$2.29\% \pm 0.08\%$
2	28049	4.85%	24665 ± 24	$3.70\% \pm 0.14\%$
3	34641	7.72%	30657 ± 28	$6.87\% \pm 0.24\%$
4	31483	4.98%	25458 ± 12	$2.94\% \pm 0.02\%$
5	22104	3.08%	18008 ± 15	$1.80\% \pm 0.08\%$
6	35652	8.27%	31274 ± 59	$5.50\% \pm 0.16\%$
7	37561	8.07%	31646 ± 39	$6.23\% \pm 0.24\%$
8	29782	4.72%	25317 ± 24	$2.86\% \pm 0.16\%$

Conclusions

- simple algorithm
- easy initialisation (can choose flat distribution)
- can solve any energy functional that can be evaluated
- able to find better optima
- inherently parallel