Unmentioned approximations in Dense CRF

Vikas Dhiman
vdhiman@nec-labs.com

Abstract

Fully connected CRF like those presented in [2] use some approximations that are neither mentioned nor justified in the paper.

1. Dense CRF

By Hammersley Clifford theorem, there exists a factorization of a graphical model such that total probability of any part of the graph can be written as product of positive functions of its cliques [1, 145]. Writing the factorization domain, the energy of dense CRF should be

$$E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j) + \sum_{i<j<k} \psi_t(x_i, x_j, x_k) + \cdots + \psi(\mathbf{x}) \quad (1)$$

This does not help because we get a term $\psi_{\mathbf{x}}$ that is as big as $E(\mathbf{x})$. This shows that dense CRFs are anti-thesis of CRF which are meant to define sparse interdependencies between variables so that probability over large number of random variables can be factorized into smaller tractable problems using domain knowledge. Without using factorizable CRF, the authors have not used any of the useful theorems that have been established in the field of graphical models.

However, the authors in [2] invent a new term called *fully connected pairwise CRF model* and without any explanation introduce the following energy function.

$$E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j) \quad (2)$$

Clearly this is a linear approximation of energy function. The energy function is assumed to weighted sum of pairwise energies over different variable states. How well the energy function can be approximated depends on the energy function itself. For comparison we present 1 a 2D Gaussian compared with an approximation of Gaussian that is generated by sum of two 1D Gaussians.

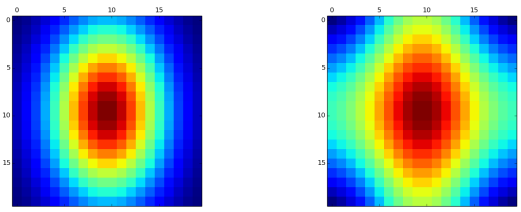

Figure 1. Comparison of a 2D Gaussian (left) compared with an approximation of Gaussian (right) that is generated by sum of two 1D Gaussians.

Intuitively, the approximation of a 2D Gaussian using sum of 1D Gaussians is like approximating 2D Euclidean distance by using linear combination of 1D distance only which is commonly known as L1 or manhattan distance. Note that the deviation of approximation is minimal around the center of the Gaussian and around the axes and becomes more prominent as the distance increases.

There are a few open questions here:

1. How good is this approximation? What are the error bounds of the approximation? Can I check an energy function for error bounds?
2. Is this better than the sparse CRFs? (at least in gaussian kernel pairwise energies)
3. What lessons can we take from empirical success of dense CRFs and how can we make sparse CRFs better or vice versa?

References

[1] G. Grimmett. *Probability on graphs: random processes on graphs and lattices*, volume 1. Cambridge University Press, 2010.

[2] P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *arXiv preprint arXiv:1210.5644*, 2012.