May 10th, 11:00 AM - 1:00 PM

Cox Processes for Visual Object Counting

Yongming Ma
Portland State University, yongming@pdx.edu
Cox Processes for Visual Object Counting

Yongming Ma, PhD; Purnima Rajan, PhD; Faculty Advisor, Dr. Bruno Jedynak

1Portland State University, 2Johns Hopkins University

Abstract

We present a model that utilizes Cox processes and CNN classifiers in order to count the number of instances of an object in an image.

By employing Kronecker algebra which takes advantage of the direct product structure of the covariance matrix, we make the computation time and storage to $O(n^{1.5})$ and $O(n)$ respectively instead of $O(n^3)$ and $O(n^2)$.

In practice, we select a subset of bounding boxes in the image and we query them for the presence of the object by running a pre-trained CNN classifier like AlexNet. We aggregate the observations and compute a posterior distribution, which is then used to estimate the number of instances of the object in the entire image.

We show results on both simulated data and on images from MS COCO dataset. We also compare our counting results with RCNN, and show that for the task of counting, cox process outperforms or match the RCNN.

Algorithm

Let $\theta$ be a doubly stochastic Poisson Process with intensity $\lambda$: $\theta \sim PP(\lambda)$ over the domain D of bounding boxes. The random intensity function $\lambda$ is obtained by mapping a Gaussian Process (GP) prior to a positive function, i.e., $\lambda \sim GP(\mu, \Sigma)$ and $\hat{\mu}(.) = \exp(\mu(\cdot))$. Figure 1 and 2 show a random prior drawn from the GP.

The image is observed at M bounding boxes in order to get observations $y_1, y_2, \ldots, y_M$, $y_m \in \{0,1\}$. The likelihood function of observations can be approximated by a first order Taylor expansion:

$$p(y|m|\theta) = \prod_{i=1}^{M} \left(1 - \lambda_i + e^{-\lambda_i} \right)^{y_i}$$

The posterior distribution of intensity is obtained as:

$$p(\lambda|y) = \exp \left( -\lambda - \theta \sum_{i=1}^{M} \ln(1 - e^{-\lambda_i}) \right)$$

The Laplace method is used to approximate the posterior since the posterior is non-Gaussian. The Laplace method gives a Gaussian approximation for the posterior:

$$p(\lambda|y) \approx \mathcal{N}(\lambda|\lambda^*, \Sigma^*)$$

where $\lambda^*$ and $\Sigma^*$ can be found by Newton method.

To implement Newton methods faster, we use the Kronecker Algebra:

$$\Sigma = \sigma^2 \sum_{i=1}^{M} \text{vec}(X_i) \Sigma \text{vec}(X_i)$$

where $\Sigma_i$ is the covariance matrix on each dimension, and the operator $\text{vec}(X)$ corresponds to a column-wise stacking of the square matrix $X$.

Simulation Results

We replicate the procedure one thousand times and obtain the total posterior intensity for each time. Figure 3 shows two of the simulations.

Since each instance generates multiple detections, the total intensity is larger than the true number of the instances. A linear regression equation is used to model the relation between the total intensity and the true number of objects. Based on the regression equation, we estimate the count for each simulated image and the result is shown in Figure 4.

Conclusions

We have presented a framework for estimating the number of objects in an image based on Cox processes. We evaluated our method on both synthetic and real data, and demonstrated empirically that the proposed idea improves upon the state of the art. In addition to counting, our algorithm potentially allows for soft localization of objects which we plan to further develop.

Contact

Yongming Ma
Email: yongming@pdx.edu

Fariborz Maseeh Department of Mathematics and Statistics
PO Box 751, Portland, Oregon 97207

References