Bayesian Spatial Cluster Signal Learning with Application to Adverse Event (AE)

Hou-Cheng Yang \(^1\)  Guanyu Hu \(^2\)

\(^1\)U.S. Food and Drug Administration
\(^2\)University of Missouri

Motivation

The medical device industry plays a crucial role in the healthcare of patients worldwide. For example, it can prevent diagnosis, diagnosis, or even treatment for the patients. The safety, effectiveness, and security of medical devices in the United States are monitored by the U.S. Food and Drug Administration (FDA). Many manufacturers, and regulators are interested in monitoring the safety of those medical devices. They tried to find whether there exist a geographic pattern of those adverse events (AEs). By exploring those finding pattern, it would be helpful for manufacturers and regulators to take corresponding actions.

Background, Challenges, Goals, and Contributions

- **Background and Challenges.**
  - Hu et al. (2012) used a frequentist spatial scan statistic to find the potential spatial clusters for medical device safety.
  - Although spatial scan statistics is a useful tool but it suffers some obstacles.
- **Notations.**
  - We model our approach to identify the spatial signal region in an efficient manner.
  - Provide an ability to capture both locally spatially contiguous clusters and globally disconnected clusters.
- **Goals and Contributions.**
  - Our model provides an alternative way to identify the spatial signal region in an efficient manner.
  - Provide an ability to capture both locally spatially contiguous clusters and globally disconnected clusters.

Nonparametric Clustering

- **If the number of clusters is not fixed?**
- Nonparametric: can grow if data need it.
- Probabilistic distribution over number of clusters.
- We used mixture of Finite Mixture (MFM) over Dirichlet Process Mixture (DPM) for clustering.

Markov Random Field (MRF)

- Inspired by Ghanbarz and Buhmann (2009), we apply the pairwise Markov random field into our model to handle spatial dependency.
- Consider an undirected random graph \( G = (V, E, W) \), where \( V \) is the vertex set while \( E \) is the set of graph edges, with weights \( W \) on the corresponding edges. The pairwise MRF model is defined as follows.

\[
\Pi(\lambda_1, \ldots, \lambda_n) = \exp \left\{ \sum_{(i,j) \in E} \sum_{l=1}^{P} H_l(\lambda_i, \lambda_j) - A(W) \right\}
\]

where \( P \) is a vertex-wise term and \( A(W) \) is an interaction term.

Bayesian Hierarchical Model: MRF-MFM-Poisson

Adapting MRF, and used Gamma distribution as the base measure, the Bayesian hierarchical model can be expressed as follows.

Data Model:

\[
y(x_i) \sim \text{Poisson}(\lambda(x_i)), \quad i = 1, \ldots, n
\]

MRF:

\[
\lambda(x_i) = \sum_j \lambda(x_j) \prod_{j \neq i} G(x_i, x_j)
\]

MFM:

\[
G(x_i, x_j) = \frac{1}{1 + d(x_i, x_j)}
\]

Likelihood Ratio Test (LRT)

Recall, for each sub-region, we have the corresponding pair data: \( (y(x_i), \lambda(x_i)) \).

- First step only use \( y(x_i) \) to finding potential signal clusters.
- In this step, both \( (y(x_i), \lambda(x_i)) \) will be involved.

Assuming that:

\[
y(x_i) \sim \text{Poisson}(\lambda(x_i)), \quad n \rightarrow \infty
\]

The hypothesis testing as follows:

\[
H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu > \mu_0
\]

Then, the region associated with maximum log-likelihood ratio is identified as the most likely cluster signal.

Discussion and Future Works

- Our algorithm included two steps:
  1. First step we use the proposed MRF-MFM Poisson model to find potential signal clusters.
  2. Then we used likelihood ratio test (LRT) in second step on those potential clusters to identify the most likely cluster signal.
- Our model provide an alternative way to identify the spatial signal region in an efficient manner.
- Our method provide an ability to capture both locally spatially contiguous clusters and globally disconnected clusters.
- A lot of investigations and strategies can be follow up depend on each specific scenario.

References


Real Data Application

Hypothetical data comprised total number of Left Ventricular Assist Device (LVAD) used in the state and the related number of stroke occurrences associated with LVAD use. Three combination cases of adverse event rates. The AE rate for the area outside the cluster is 0.1, and three different AE rates for the area inside the cluster is 0.297, 0.497, and 0.697.