# Bayesian Mark Interaction Mixture Model

Indrajith Wasala Mudiyanselage

The University of Texas at Dallas

08/14/2020

# Outline

- Introduction
- The Potts model
- The hidden Potts mixture model
- Bayesian Mark interaction model

# Introduction

- Cancer is the uncontrolled growth of abnormal cells in the body.
- Lung cancer is the most common human cancer.
- Non–small-cell lung cancer (NSCLC) is the most common cause of lung cancer death.
- Treating lung cancer are largely based on clinical and pathological staging systems.
- A tumor pathology image contains a large amount of information.
- Lymphocyte, stromal, and tumor cells, are commonly seen in tumor tissue images.

# Bayesian mark interaction model (Li et at 2019)

- This model describe a spatial map of cells in a Cartesian coordinate system, with $n$ observed cells indexed by $i$.
- $(x_i, y_i) \in \mathbb{R}^2$ to denote the $x$ and $y$ coordinates and $z_i \in \{1, ..., Q\}$, $Q \geq 2$ to denote the type of cell $i$.
- $z_1, ..., z_n$ are their associated qualitative (i.e., categorical or discrete) univariate marks.
- Without loss of generality, we assume that the data points are restricted within the unit square $[0, 1]^2$.
- This can be done by rescaling each pair of coordinates $(x_i, y_i)$ to $(x'_i, y'_i)$.

$x'_i = (x_i - v_x^{lwr})/L$ and $y'_i = (y_i - v_y^{lwr})/L$ where
$L = max(v_x^{upp} - v_x^{lwr}, v_y^{upp} - v_y^{lwr})$

# The Potts model

- Let $G = (V, E)$ denote a graph $G$ composed of a finite set $V$ of vertices and a set $E$ of edges joining pairs of vertices.
- The, edges are connected between the vertex at location $(l, w)$ and its four neighbors at locations $(l+1, w)$, $(l-1, w)$, $(l, w+1)$, and $(l, w-1)$ if applicable. Every vertex will be assigned a spin, which is defined as an assignment of $Q (Q \geq 2)$ different classes.
- Let an L-by-W matrix $\mathbf{P}$ denote the collection of all spins, where each element $p_{lw} \in 1, ..., Q$. Since the vertices are assigned different spins and react with their neighbors' spins, there will be some measurement of overall energy, named *Hamiltonian*.

$$H(\mathbf{P}|\theta) = -\sum_{i=1}^{L} \sum_{w=1}^{W} \sum_{(l',w') \in Nei_{(l,w)}} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \theta_{qq'} I(p_{lw} \neq p_{l'w'}, p_{lw} = q, p_{l'w'} = q')$$

# The Potts model probability mass function

- The probability of observing the lattice in a particular state $\mathbf{P}$, where a state is a choice of spin at each vertex,

$$
\begin{aligned}
Pr(\mathbf{P}|\theta) &= \frac{\exp(-H(\mathbf{P}|\theta))}{\sum_{P' \in p}^{L} \exp(-H(\mathbf{P}|\theta))} \\
&= \frac{1}{C(\theta)} \exp\left( \sum_{i=1}^{L} \sum_{w=1}^{W} \sum_{(l',w') \in Nei_{(l,w)}} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \theta_{qq'} I(p_{lw} \neq p_{l'w'}, p_{lw} = q, p_{l'w'} = q') \right)
\end{aligned}
$$

Here, $\theta$ denotes the collection of interaction energy parameters between different classes.

# The probability of observing $p_{lw} = q$ conditional on its neighbor spins,

$$Pr(p_{lw} = q|\theta, \mathbf{P}_{-l,-w}) = \frac{\exp\left(\sum_{(l',w')\in Nei_{(l,w)}} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \theta_{qq'} I(p_{lw} \neq p_{l'w'}, p_{lw} = q, p_{l'w'} = q')\right)}{\sum_{q'=1}^{Q} \exp\left(\sum_{(l',w')\in Nei_{(l,w)}} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \theta_{qq'} I(p_{lw} \neq p_{l'w'}, p_{lw} = q, p_{l'w'} = q')\right)}$$

Here, $P_{-l,-w}$ is all the spins excluding $p_{l,w}$.

# Energy functions

- Assume that each point interacts with all other points in the space.
- Let $G = (V, E)$ be a complete undirected graph with $V$ denoting the set of points and E denoting the set of direct interactions (i.e., the $(n-1)n/2$ hidden Potts mixture model.
- Define $G$ as the interaction network and define its potential energy

$$V(z|w,\Theta) = \sum_{q} w_q \sum_{i} I(z_i = q) + \sum_{q}\sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E} I(z_i = q, z_{i'} = q')$$

First term can be viewed as the weighted average of the numbers of points with different marks, while the second term can be viewed as the weighted average of the numbers of pairs connecting two points with the same or different marks.

- The interaction energy between two points (i.e., particles and cells) is usually an exponential decay function with respect to the distance between the two points.
- Assume the interaction energy between a pair of points decreases exponentially at a rate $\lambda$ proportional to the distance,

$$V(z|w,\Theta) = \sum_q w_q \sum_i I(z_i = q) + \sum_q \sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E} e^{-\lambda d_{ii'}} I(z_i = q, z_{i'} = q')$$

Where $d_{ii'} = \sqrt{(x_i - x_{i'})^2 + (y_i - y_{i'})^2}$

# The hidden potts model ( Li et at 2017 )

- For images with irregularly distributed dots, it is impossible to apply a Potts model based on a lattice that forms a regular tiling.
- Consider a preprocessed pathology image with n observed cells, where $x_i, y_i$ represents the location and $z_i$ indicates the type of cell i. Let an L-by-W matrix P denote the hidden spins at the auxiliary lattice, which partitions the image into $(L-1)(W-1)$ squares.
- To fit the imaging data into the square-lattice system normalize each coordinate $(x_i, y_i)$ by performing a linear transformation.

$$x_i' = 1 + \frac{x_i - x_{lwr}}{x_{upp} - x_{lwr}}(L-1) \quad y_i' = 1 + \frac{y_i - y_{lwr}}{y_{upp} - y_{lwr}}(W-1)$$

# The probability of assigning cell i to type q conditional on its adjacent spins at the hidden lattice

$$Pr(x_i', y_i', z_i = q|\mathbf{P}, d) = \frac{\exp(d \sum_{\{(l,w):l \leq x_i\prime < l+1, w \leq y_i\prime < w+1\}} I(p_{lw} = q)}{\sum_{q'=1}^{Q} \exp \sum_{\{(l,w):l \leq x_i\prime < l+1, w \leq y_i\prime < w+1\}} I(p_{lw} = q\prime)}$$

where $d$ is the projection parameter

# Joint likelihood function of the hidden Potts model

$$f(\mathbf{x,y,z}|\mathbf{P}, d, \theta) = f(\mathbf{x,y,z}|\mathbf{P}, d)f(\mathbf{P}|\theta) = \prod_{i=1}^{n} Pr(x\prime_i, y\prime_i, z_i = q|\mathbf{P}, d)Pr(\mathbf{P}|\theta)$$

Specify the prior distribution for $d$ as $d \sim Ga(a_d, b_d)$

# The hidden Potts mixture model (Li et at 2019)

- The hidden Potts mixture model take into account the heterogeneity of spatial point patterns observed in the pathology images.
- There are $K$ homogeneous regions.
- With different interaction parameter settings, $\theta_1, ..., \theta_K$ . With this assumption, an L-by-W latent matrix $\triangle$ is introduced to indicate the $K$ distinct regions, with $\delta_{lw} = k - 1$ if the spin at location $(l, w)$ belongs to group $k$.
- The probability mass function of the mixture model

$$
\begin{aligned}
&Pr(\mathbf{P}|\triangle, \theta_\mathbf{1}, ..., \theta_\mathbf{K}) \\
&= \prod_{k=1}^{K} \frac{1}{C_k(\theta_k)} \exp\Bigg( \sum_{\{(l,w):\delta_{lw}=k-1\}}^{L} \sum_{(l',w')\in Nei_{(l,w)}} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \theta_{qq'} I(p_{lw} \neq p_{l'w'}, p_{lw} = q, p_{l'w'} = q') \Bigg)
\end{aligned}
$$

When $K = 2$, becomes a binary latent matrix that indicates the two distinct regions, with $\delta_{lw} = 0$ if the spin at location $(l, w)$ belongs to the background area, and $\delta_{lw} = 1$ if the spin at location $(l, w)$ belongs to the AOI.

$$
\begin{aligned}
Pr(\mathbf{P}|\triangle, \theta_{\mathbf{0}}, \theta) = {} & \frac{1}{C_0(\theta_0)} \exp\Big( \sum_{\{(l,w):\delta_{lw}=0\}} \sum_{(l',w')\in Nei_{(l,w)}} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \theta_{qq'} I(p_{lw} \neq p_{l'w'}, p_{lw} = q, p_{l'w'} = q') \Big) \\
& \times \frac{1}{C(\theta)} \exp\Big( \sum_{\{(l,w):\delta_{lw}=1\}} \sum_{(l',w')\in Nei_{(l,w)}} \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \theta_{qq'} I(p_{lw} \neq p_{l'w'}, p_{lw} = q, p_{l'w'} = q') \Big)
\end{aligned}
$$

Prior model is a type of MRF, where the distribution of a set of random variables follows Markov properties that can be presented by an undirected graph. The prior

$$Pr(\delta_{lw} = k-1 | \triangle_{-l,-w}) = \frac{\exp\left(f \sum_{(l',w') \in Nei_{(l,w)}} I(\delta_{l',w'} = k-1)\right)}{\sum_{k'=1}^{K} \exp\left(f \sum_{(l',w') \in Nei_{(l,w)}} I(\delta_{l',w'} = k'-1)\right)}$$

where $f$ is a non-negative parameter that controls the spatial interaction and $\triangle_{-l,-w}$ denotes the set of $\delta_{(l',w')}$'s excluding $\delta_{(l,w)}$.

# MCMC algorithm

- Within each MCMC iteration, we need to sample $\theta_{qq'}$ from its conditional distribution.

$$\pi(\theta_{qq'}|\cdot) \propto \frac{1}{C(\theta)} \exp(-H(\mathbf{P|theta}))N(\theta_{qq'};\mu,\sigma^2)$$

- Suppose that two sequences of MCMC samples $\theta^{(1)}_{qq'},\ldots,\theta^{(U)}_{qq'}$ and $\theta^{(1)}_{0qq'},\ldots,\theta^{(U)}_{0qq'}$ Then, approximate Bayesian estimator of $\hat{\theta}_{qq'}$ and $\hat{\theta}_{0qq'}$

$$\hat{\theta}_{qq'} = \tfrac{1}{U}\sum_{u=1}^{U}\theta^{(u)}_{qq'} \text{ and } \qquad \hat{\theta}_{0qq'} = \tfrac{1}{U}\sum_{u=1}^{U}\theta^{(u)}_{0qq'}$$

# Bayesian mark interaction model (Li et at 2019)

- This model describe a spatial map of cells in a Cartesian coordinate system, with $n$ observed cells indexed by $i$.
- $(x_i, y_i) \in \mathbb{R}^2$ to denote the $x$ and $y$ coordinates and $z_i \in \{1, ..., Q\}$, $Q \geq 2$ to denote the type of cell $i$.
- $z_1, ..., z_n$ are their associated qualitative (i.e., categorical or discrete) univariate marks.
- Without loss of generality, we assume that the data points are restricted within the unit square $[0, 1]^2$.
- This can be done by rescaling each pair of coordinates $(x_i, y_i)$ to $(x_i', y_i')$.

$x_i' = (x_i - v_x^{lwr})/L$ and $y_i' = (y_i - v_y^{lwr})/L$ where
$L = max(v_x^{upp} - v_x^{lwr}, v_y^{upp} - v_y^{lwr})$

# Energy functions

- Assume that each point interacts with all other points in the space.
- Let $G = (V, E)$ be a complete undirected graph with $V$ denoting the set of points and E denoting the set of direct interactions (i.e., the $(n-1)n/2$ hidden Potts mixture model.
- Define $G$ as the interaction network and define its potential energy

$$V(z|w,\Theta) = \sum_q w_q \sum_i I(z_i = q) + \sum_q \sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E} I(z_i = q, z_{i'} = q')$$

First term can be viewed as the weighted average of the numbers of points with different marks, while the second term can be viewed as the weighted average of the numbers of pairs connecting two points with the same or different marks.

- The interaction energy between two points (i.e., particles and cells) is usually an exponential decay function with respect to the distance between the two points.
- Assume the interaction energy between a pair of points decreases exponentially at a rate $\lambda$ proportional to the distance,

$$V(z|w,\Theta) = \sum_q w_q \sum_i I(z_i = q) + \sum_q \sum_{q'} \theta_{qq'} \sum_{(i\sim i')\in E} e^{-\lambda d_{ii'}} I(z_i = q, z_{i'} = q')$$

Where $d_{ii'} = \sqrt{(x_i - x_{i'})^2 + (y_i - y_{i'})^2}$

- When $n$ is large compute the potential energy, resulting in a tedious computation (needs to sum over $n$ data points and $(n-1)n/2$ pairs of data points)
- An alternative way is to obtain an approximate value of $V(z|w, \Theta, \lambda)$ by neglecting those pairs with distance beyond a certain threshold $c, c \in (0, 1)$.
- $G$ reduces to a sparse network $G' = (V, E')$, with $E' \subseteq E$ denoting the set of edges joining pairs of points $i$ and $I'$ in $G'$, if their distance $d_{ii'}$ is smaller than a threshold $c$.
- The potential energy of the interaction network $G'$

$$V(z|w, \Theta) = \sum_q w_q \sum_i I(z_i = q) + \sum_q \sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E'} e^{-\lambda d_{ii'}} I(z_i = q, z_{i'} = q')$$

# Data likelihood

- If we have a locally defined energy, then a probability measure with a Markov property exists.
- Probability of observing marks associated with their locations in a particular state

$$p(\mathbf{z}|\mathbf{w}, \mathbf{\Theta}, \lambda) = \frac{\exp(-V(\mathbf{z}|\mathbf{w}, \mathbf{\Theta}, \lambda))}{\sum_{z'} \exp(-V(\mathbf{z}'|\mathbf{w}, \mathbf{\Theta}, \lambda))}$$

- The normalizing constant/ partition function

$$C(\mathbf{w}, \mathbf{\Theta}, \lambda) = \sum_{z'} \exp(-V(\mathbf{z}'|\mathbf{w}, \mathbf{\Theta}, \lambda))$$

- DMH is an auxiliary variable MCMC algorithm, which can make the normalizing constant ratio canceled by augmenting appropriate auxiliary variables through a short run of the ordinary Metropolis-Hastings (MH) algorithm.

The probability of observing point i belonging to class q conditional on its neighborhood configuration(s)

$$p(z_i = q | z_{-i}, \mathbf{w}, \mathbf{\Theta}, \lambda) \propto -w_q - \sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E'} e^{-\lambda d_{ii'}} I(z_i = q, z_{i'} = q')$$

where $z_{-i}$ denotes the collection of all marks excluding the $i^{th}$ one.$w_q$ -first-order intensity and $\theta_{qq'}$ second-order intensity.

# Posterior estimation

- We are interested in estimating $\mathbf{w}, \boldsymbol{\Theta}$ and $\lambda$ which define the Gibbs measure based on the local energy function.
- We obtain posterior inference by summarizing the MCMC samples $w_q^{(1)}, \ldots, w_q^{(U)}; \theta_{qq'}^{(1)}, \ldots, \theta_{qq'}^{(U)}$ and $\lambda^{(1)}, \ldots, \lambda^{(U)}$.
- Approximate Bayesian estimator of each parameter can be simply obtained by averaging over the samples.

$$\hat{w_q} = \sum_{u=1}^{U} w_q / U \qquad \hat{\theta}_{qq'} = \sum_{u=1}^{U} \theta_{qq'}^{(u)} / U \qquad \hat{\lambda} = \sum_{u=1}^{U} \lambda^{(u)} / U$$

*Thank You*