Draft

Braeden Sopp

Prostate Cancer (PCa) is the second leading cause of cancer-deaths for men in the United States and is expected to make up 20% of new cancer diagnosis in men in 2019[3]. The Gleason grading system is currently the primary system for grading PCa's architecture and stands as one of the most important predictors of prognosis. The Gleason system divides the multiple architectural patterns of Prostate cancer into 5 categories. The Gleason score is assigned to a whole slide image of a prostate biopsy notating which of these two of these two architectural patterns groups are most prevalent in the sample. However, the gleason score also suffer from high intraobserver and interobserver variability when being assigned by a Pathologist who must visually inspect the cancer slides.[4] This produces as need for new effective quantitative measures on prostate cancer.

The emergence of digital pathology has enabled researchers to examine digital histopathological slides with greater detail. Two areas that have been applied to Histopathological data are topological data analysis (TDA) and Spatial analysis. Both of these methods aim to examine the spatial properties of data but differ in their focus. TDA is uniquely suited to examining the shape of data while spatial analysis focuses mainly[a bit of strong word] on questions of spatial dependence, spatial association, and spatial heterogeneity.[1]

[WIP: mainly a list right now - adapted from notes. need to clean up and add citations] Spatial Analysis has been used to examine the tumor mircoenvironment from an ecological perspective, examining cancer and immune cell co-location and relating it to patient survival in estrogen receptor negative-breast cancer. Similar work has shown that statistical hotspot analysis correlates with better prognosis as well. Other studies demonstrated measures of spatial homogeneity in stromal cells was associated with good outcome for patients with estrogen negative breast cancer patients. Combinations of spatial statistics on CD68+ macrophages in human head and neck tumors have also been shown accurately match human predictions.

TDA was successfully applied to breast cancer to identify subgroup of estrogen receptor positive breast cancer which had an 100% survival rate. Persistent Homology (PH) is a specific method from TDA which has found great success when applied to cancer data. A more detailed description of PH will be put off for now, and we will start with just describing some of its notable applications when examining Histopathology.

PH has also been applied to subtypes of breast cancer data where it was shown to be able to identity breast cancer subtypes using the information provided by the distance

between the nuclei in the sample cell. On other breast cancer data, PH was able to use the tumor mirco-environmental to identify breast cancer subtypes as well as predict patient survival. PH is uniquely well suit to be applied to PCa as the gleason scale is based entirely on the cancer's architectural pattern. Here is has been show that PH can cluster a fine collection of prostate architectural pattern then used in producing the gleason score. PH has also been demonstrated to predict prostate cancer aggressiveness using machine learning models.

Methods

Mantel Analysis based Methods

Mantel Analysis [check against langauge in [2]] are a set of methods base around the Mantel statistics and are focused on assessing the linear association between two square distance matrices of same dimension.

Given two distance matrices X and Y the Mantel statistic is defined to be dot product the upper triangular portion of the distance matrices as vectors:

$$z_m = \sum_{i=0}^{n} \sum_{j=i+1}^{n} X_{i,j} Y_{i,j}$$

Likewise, there is also the standardized Mantel statistics which uses the Pearson's r statistic instead:

$$r_m = \frac{1}{n(n-1)/2 - 1} \sum_{i=0}^{n} \sum_{j=i+1}^{n} (X_{i,j} - \hat{\mu}_X)(Y_{i,j} - \hat{\mu}_Y)/\hat{\sigma}_X \hat{\sigma}_Y.$$

Mantel statistics are primarily used in the context of Hypothesis testing where it is applied in two related methods called the Mantel and partial Mantel test. Here we will only be focusing only on the Mantel test. The Mantel test is a permutation test that typically uses either Mantel statistic or the standardized Mantel statistic as the test statistic and works under the assumptions that:

1. The ith row and column both correspond to the same sample/location.

2. The samples are independent of one another.

The null hypothesis for the Mantel test is that there is no linear association between the two matrices. The test itself is preformed by permuting the rows and columns of one matrix while keeping the other fixed and recomputing the test statistic for each permutation to form a reference distribution. The value test value is then compared to the reference distribution to obtain a p-value.

A more interesting use of the Mantel methodology and test statistic is the Mantel correlogram. The mantle correlogram is an innovative method to examine spatial correlation

use in only distance matrices. The base procedure is performed by first dividing the range of the values of the geographic distance matrices into N distance classes of equal size representing an interval of values in the geographic distance matrix. Then the Kth distance class, we produce a model matrix $M(k)$ where an entry has the value K if it falls into the Kth distance class, otherwise it is set to 0 in $M(k)$. Then the standardized Mantel statistic between the model matrix and the response matrix is computed and the Mantel test procedure is performed. This forms a graph of Mantel statistics with respect to the distance classes which is called the Mantel correlogram. The interpretation of the resulting correlogram is then based on the shape of the significant values of the test.

References

[1] Michael de Smith, Michael Goodchild, and Paul Longley. *Geospatial Analysis - A Comprehensive Guide to Principles Techniques and Software Tools*. Drumlin Security Ltd, 6 edition, May 2018.

[2] Pierre Dutilleul. *Spatio-temporal heterogeneity: Concepts and analyses*. Cambridge University Press, 2011.

[3] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.

[4] Patrick C Walsh. The gleason grading system: a complete guide for pathologists and clinicians, 2013.