Automated, Quantitative Analysis of Histopathological Staining in Nuclei

Ricardo Henao\textsuperscript{1}, PhD, Joseph Geradts\textsuperscript{1}, MD, MA, Manabu Kurokawa\textsuperscript{2}, PhD, Sally Kornbluth\textsuperscript{1}, PhD, and Joseph E. Lucas\textsuperscript{3}, PhD

\textsuperscript{1}Duke University, Durham, NC
\textsuperscript{2}Dartmouth College, Hanover, NH
\textsuperscript{3}Quintiles, Durham, NC

Abstract

Technological advances have allowed the generation of high-throughput imaging of tissue sections. However, the analysis of these samples is typically still performed manually by one or multiple pathologists. We present a novel statistical model for the automated, quantitative analysis of these images. Our approach requires minimal tuning and allows recapitulation of estimates of staining strength in the nuclei of tumor cells as estimated by the gold standard. Besides, it compares favorably to other quantitative approaches available in the public domain.

Introduction

Analysis of tissue sections after staining is a subjective and labor intensive process. Typically, the pathologist must manually scan through a series of slides to estimate the strength of staining using a discrete scoring system. Often only a subset of cells on the slide should be considered and often in only one region of those cells. For large studies, this process may involve multiple pathologists, which leads to challenges with subjectivity, inter-rater reliability, and fatigue. In this paper we describe a model for the automated analysis of stained slides that results in an objective, repeatable and quantitative assessment of the staining level of a particular protein in the nuclei of cancer cells.

There are some targeted approaches to this task in the literature, such as Masseroli et al (2000)\textsuperscript{1} for liver fibrosis and Davis et al (2003)\textsuperscript{2} for apoptosis. The most rudimentary general approach is to simply integrate the total amount of staining of the relevant color across a slide. However, this does not allow for assessment of stain levels in different parts of the slide separately. One approach to addressing this problem involves the manual selection of relevant regions together with image manipulation by the pathologist\textsuperscript{3}. This leads to a quantitative result, but is still not tractable in a high-throughput experiment. This challenge has inspired the development of some excellent toolkits to support automation of this task\textsuperscript{4,5}.

Dirichlet mixture models have been successfully used in a wide range of image processing applications. More particularly, various specialized Bayesian models for segmentation tasks based on Dirichlet processes have been proposed in recent years\textsuperscript{6,7}. However, most of this work targets natural images. The development of segmentation models, such as the one described in this paper, employing Dirichlet processes appears to be very promising for the analysis of histopathological images.

In this paper we present a statistical model for the automated analysis of histopathology images of tumor sections that is able to recapitulate the pathologist’s assessment in a quantitative and repeatable way. Our model requires very little tuning and compares favorably to other publicly available software designed for this task.

Data

The data set consists of 30 digitized microscope images. Each image is RGB encoded, 200x magnification scale and $1100 \times 828$ pixels in size. All shots were manually taken and subsequently labeled by a pathologist.
into either low or high levels of expression of the stained protein. Figure 3 show particular examples of images of these two classes.

Our approach employs two steps. We start by extracting some features from the images, specifically we over-segment each image into small patches known as superpixels. These are commonly used for image labeling problems in which pixel-level labeling might be prohibitive or as preprocessing step in more complex segmentation algorithms. We use it as a way to obtain a compressed color representation of images. The main goal in this step is to isolate nuclei of tumor cells from background. This serves two purposes: (i) eases visualization by shifting focus to nuclei and (ii) information gathered from segmented nuclei can be used for further analysis or as part of a more elaborated pipeline. We employed turbopixels, a fast superpixel algorithm based on geometric flows\(^8\). For each superpixel we compute a 128 bin RBG histogram, resulting into a 384 dimensional vector containing binned color counts. We did not notice significant changes in results by further increasing the number of bins used to compute histograms. Provided that each image was divided into approximately 8600 superpixels, each image is then represented as a 8600 \( \times \) 384 integer matrix. The second step of our analysis pipeline involves the fitting of a hierarchical statistical model to the resulting set of 30 matrices.

**Model**

The Dirichlet process (DP) has been widely used to dynamically model the number of clusters in conjunction with mixture models. The Chinese Restaurant Process\(^9\) (CRP) offers a very useful metaphor for the DP and some of its generalizations\(^10\). Imagine a restaurant with an infinite number of tables denoted by \( \phi_i \), customers \( \theta_n \) enter sequentially to the restaurant so that the \( n \)-th customer sits at a given table with probability proportional to the number of customers already occupying it \( m_k \) or gets a new table with probability proportional to \( \alpha \). From the metaphor we can see that the DP has a rich-get-richer dynamic and that \( \alpha \), the concentration parameter, controls the total number of tables (clusters) for a given customer base. The CRP results in an exchangeable model, i.e. the probability distribution over partitions does not depend on the ordering of the customers. Defining \( K \) as the number of non-empty tables and \( z_n \in 1, \ldots, K \) to be the table assignment for customer \( n \), the prediction rule in the CRP can be written as

\[
z_n | z_{\setminus n}, \alpha \propto \alpha \delta_k, + \sum_{k=1}^{K} n_k \delta_k,
\]

where \( z_{\setminus n} \) is the set of table assignments excluding customer \( n \) and \( k^\ast \) indexes a new table. In terms of our microscope images we can think of superpixels as customers, \( \phi_k \) as color profiles or color distributions and \( z_n \) as to which color profile superpixel \( \theta_n \) belongs to. More formally we can say that \( \theta_n \) is a sample from a distribution \( F(\phi_{z_n}) \) with parameter \( \phi_{z_n} \) and \( \phi_k \) follows a DP with concentration parameter \( \alpha \) and base measure \( H \). Provided that superpixels are represented as quantized color histograms, we assume a discrete distribution for \( F(\phi) \) and let \( H \) be a Dirichlet distribution with concentration parameter \( \gamma \).

We want to obtain a compact color based representation of each image, however we still want to be able to share information across them. In principle, we can add a top layer to the CRP model to enable color profile sharing, i.e. each image will have its own CRP (bottom layer) which in turn will get its color profiles from a common CRP. This model is commonly known as the Chinese restaurant Franchise (CRF) representation of the hierarchical Dirichlet process (HDP)\(^11\). Figure 1 shows a graphical representation of our proposed HDP model for microscope images. The top layer of the model is represented by two parameters \( \phi_k \) and \( \beta_k \), being a 384 dimensional normalized color profile and its probability of occurrence, respectively. The potentially infinite set of color profiles is indexed by \( k \), meaning that in practice we only instantiate color profiles for which \( \beta_k > 0 \). Hyperparameters \( \alpha_0 \) and \( \gamma \) control the total number of active color profiles and their spectrum specificity, respectively. For instance if \( \gamma \) is very small, profiles will encode very narrow wavelength bands. The bottom layer has the image specific parameters; \( \theta_{ni} \) is color histogram for superpixel \( n \) in image \( i \), \( z_{in} \) is the assignment variable for superpixel \( n \) in image \( i \), \( \pi_j \) is the vector of color profile usage frequencies for image \( i \) and \( N_i \) is the number of superpixels in image \( i \). Lastly, hyperparameter \( \alpha \) controls the total number of color profiles used by each image.

The model has various parameters of interest namely \( z_{in}, \phi_k \) and \( \pi_i \). Inference is carried out using Markov Chain Monte Carlo (MCMC) and hyperparameters \( \{ \alpha, \alpha_0, \gamma \} \) were provided with prior distributions.
to facilitate their tuning. In particular, gamma priors with shape 2 and rate 1 were used everywhere. There are several MCMC sampling approaches for HDPs, we are using a truncated DP representations with a maximum of 200 color profiles. In practice we did not observe the model reaching the profile limit at any time during inference, meaning that further increasing the maximum number of color profiles does not change the results. Inference details can be found for instance in Teh et al (2006)\textsuperscript{11}.

Results

The previously described superpixel processed data consist of a $260457 \times 384$ matrix of quantized color histograms. The HDP sampler was run for 1500 iterations, we observed in general good mixing and cluster assignment stabilization after the first 500 iterations. We did not notice significant changes in the cluster assignments after making small changes in all the hyperparameters settings of the model, i.e. shapes and rates of gamma hyper-priors. We tried to make quantitative comparison of the proposed HDP model against well established algorithms for image segmentation including normalized cuts\textsuperscript{12} (Ncuts), mean shift\textsuperscript{13} (MS) and $K$-means in conjunction with superpixels. Ncuts was computationally prohibitive considering the size of the images in the data set. We tried to select for $K$ in $K$-means using internal measures such as Davies-Bouldin index\textsuperscript{14} and silhouette average\textsuperscript{15} but segmentation results were too coarse for the desired level of detail. For MS we observe nice segmentation results when manually tuning its parameters however we could not find a good set of parameters by grid search and internal measures to appropriately fit the entire data set. It is still possible that these methods could work satisfactorily with additional preprocessing or specialized parameter tuning.

After running the HDP inference we ended up with a model with 180 color profiles. We know that nuclei appear darker than the remaining elements of the background thus we can simply sort color profiles according to intensity to then set a manually selected threshold for visualization purposes. Besides, just being able to interactively set the threshold could be a very useful tool for exploratory purposes. Here we attempt to select the best number of color profiles according to their ability to correctly classify the status of each image. In order to do so in an unbiased manner, we perform leave-one-out cross-validation (LOOCV) using a number of color profiles ranging from 2 to 180 and a naive Bayes classifier model\textsuperscript{16}. At stage $j$ of LOOCV, classifier training is done using a subset of the color profile usage probabilities $\pi_i$ for all images but the one being tested, then HDP model and trained classifier are used in turn to make a prediction. Figure 2(b) shows classification results for a range of color profiles, we see that the accuracy curve is rather flat meaning that the classifier does a good job for a set of color profiles ranging from 15 to 24. This also indicates that selecting the number of profiles for discrimination purposes is not a critical task in our case. For visualization purposes we select the number of profiles in order to maximize classification accuracy, i.e. 24 color profiles. We can also see from Figure 2(b) that at 86.7% accuracy (26/30 images), true positive and true negative rates are 88.9% (16/18) and 83.3% (10/12), respectively, which suggests that the classifier has a well balanced misclassification risk. We did not consider more sophisticated classifiers, however we believe that classification accuracy can be further improved with an upgraded classifier or better yet by integrating it directly into the HDP model. If we examine the 24 selected color profiles in Figure 2(a) we see that most

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{Graphical model for the HDP model. $N$ is the number of images, $\{\alpha, \alpha_0, \gamma\}$ is the set of hyperparameters and $\theta_i$ is the only observed variable in the model (shaded node). We used bold letters to distinguish vectors from scalars.}
\end{figure}
Figure 2: Classification results. (a) The 24 color profiles used for final image segmentation, there were obtained from the LOOCV procedure. Blocks in the left bar show the mean color encoded by each profile. The vertical lines separate individual sections of RGB spectrum. (b) LOOCV number of correctly classified (CC), true positives (TP) and true negatives (TN) images using a naive Bayes classifier. The vertical dashed line denotes the selected number of color profiles.

of them summarize the desired color features, i.e. dark/middle brown and bluish shades characteristic of nuclei in our image data set.

Figure 3 shows examples of segmented images one from each category, i.e. low and high levels of expression of the stained protein. We see that the HDF based segmentation model with classifier aided profile selection produces a very nice separation between background and cell nuclei despite color heterogeneity. Larger versions of the images in the figure can be found online at: http://people.duke.edu/~rh137/huwe1.html.

Closing remarks

We foresee a version of the presented model in which the classifier is integrated into the segmentation model as an additional layer, in this way the model will be able to bias color profile assignments towards better classification performance and hopefully improve visual representations of nuclei data.

One particularity of the data we have not addressed yet but represents a good opportunity for overall improvement is to extend the model to use morphological information about the segments/superpixels, for instance size or regularity. We know that this kind of information is used by pathologists to better inform their decisions.

References

[1] Masseroli M, Caballero T, O’Valle F, Moral RMGD, Pérez-Milena A, Moral RGD. Automatic quantification of liver fibrosis: design and validation of a new image analysis method: comparison with semi-quantitative indexes of fibrosis. Journal of hepatology. 2000;32(3):453–464.

[2] Davis DW, Buchholz TA, Hess KR, Sahin AA, Valero V, McConkey DJ. Automated Quantification of Apoptosis after Neoadjuvant Chemotherapy for Breast Cancer Early Assessment Predicts Clinical Response. Clinical cancer research. 2003;9(3):955–960.

[3] Lehr HA, Jacobs TW, Yaziji H, Schnitt SJ, Gown AM. Quantitative evaluation of HER-2/neu status in breast cancer by fluorescence in situ hybridization and by immunohistochemistry with image analysis. American journal of clinical pathology. 2001;115(6):814–822.

[4] Peng H. Bioimage informatics: a new area of engineering biology. Bioinformatics. 2008;24(17):1827–1836.
Figure 3: Segmentation examples. Black lines are introduced to highlight nuclei segments corresponding to pixels assigned to one of the 24 color profiles selected by the classification model. Images correspond to low (a and c) or high (b and d) levels of expression of the stained protein.

[5] Roysam B, Lin G, Bjornsson C, Narayanaswamy A, Chen Y, Shaina W, et al. The FARSIGHT project: associative multi-dimensional image analysis methods for optical microscopy. In: J Rittscher SW R Machiraju, editor. Microscopic Image Analysis for Life Science Applications. Artech Publishing House; 2008.

[6] Du L, Ren L, Dunson D, Carin L. A bayesian model for simultaneous image clustering, annotation and object segmentation. Advances in Neural Information Processing Systems. 2009;22:486–494.

[7] Ghosh S, Ungureanu AB, Sudderth EB, Blei DM. Spatial distance dependent Chinese restaurant processes for image segmentation. In: Shawe-Taylor J, Zemel RS, Bartlett P, Pereira FCN, Weinberger KQ, editors. Advances in Neural Information Processing Systems 24. MIT Press; 2011. p. 1476–1484.

[8] Levinshtein A, Stere A, Kutulakos KN, Fleet DJ, Dickinson SJ, Siddiqi K. Turbopixels: Fast superpixels using geometric flows. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2009;31(12):2290–2297.

[9] Aldous D. Exchangeability and related topics. École d’Été de Probabilités de Saint-Flour XIII—1983. 1985:p. 1–198.
[10] Pitman J. Combinatorial stochastic processes. vol. 1875 of Lecture notes in mathematics, Ecole d’été de probabilités de Saint-Flour XXXII. Berlin: Springer-Verlag; 2006.

[11] Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical Dirichlet processes. Journal of the American Statistical Association. 2006;101(476):1566–1581.

[12] Shi J, Malik J. Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2000;22(8):888–905.

[13] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2002;24(5):603–619.

[14] Davies DL, Bouldin DW. A cluster separation measure. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1979;(2):224–227.

[15] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics. 1987;20:53–65.

[16] Bishop CM. Pattern Recognition and Machine Learning. Springer; 2006.