Image segmentation using Bayesian stochastic blockmodels

Pavlin Gregor Poličar
University of Ljubljana
Faculty of Computer and Information Science
Email: pavlin.g.p@gmail.com

***Abstract*—Image segmentation is a very broad field that combines methods from various different areas of expertise. Graph based techniques achieved some early successes in this field and are still in use today. However, community detection techniques have progressed significantly over the past few years, but none have been applied to this problem until quite recently, where the Louvain modularity optimization method was applied and appeared to achieve reasonable results. However, we know that modularity optimization is one of numerous community detection algorithms developed in recent years. This raises the question, how well would other more recent community detection algorithms perform for image segmentation? In this work we examine the most recent developments in Bayesian stochastic blockmodeling and attempt to apply them to images. We examine the difficulties that arise and possible solutions. Our results indicate that stochastic blockmodels are not a viable method for image segmentation and discuss the reasons why most community detection algorithms would perform badly.**

I. Introduction

The problems of image segmentation and perceptual grouping remain one of the most popular challenges in computer vision today. Image segmentation is the process of partitioning an image into multiple segments, usually corresponding to different objects or regions present in the image. Intuitively, a good segment is a subset of pixels in the image, that are spatially close to each other and are fairly uniform with regards to pixel similarities.

However, one can never be sure what a ”good” segmentation truly is, since image segmentation is an inherently hierarchical task. Therefore no true segmentation exists, since it can be defined at multiple levels of granularity. Instead of finding the *correct* answer, we should strive to find a *useful* answer in the context of our application. One must simply look at different segmentations given by multiple human annotators for these differences to become immediately apparent (see Figure 1). Ideally, we could design a method that would produce a hierarchical tree-like partition of an image at different granularities, but we quickly run into computation challenges. Therefore any low-level segmentation technique cannot and should not strive to produce a correct segmentation, but should produce a useful segmentation in the context of our application.

Graph-based techniques have proved to be quite effective for this task, most notably – the normalized cut algorithm and its variants proposed in 2000 [1] are still some of the most widely used methods today due to their simplicity and computation efficiency. These methods and other graph-flow based methods have proven to be quite useful in this domain.

Figure 1. Human annotators produce different segmentations for the same image. The true image is shown in the top left, and eight possible annotations are shown. Clearly, no one single segmentation is the ”correct” answer – different segmentations are possible at different granularities. The image and segmentations can be found Berkley segmentation benchmark dataset under the id 254054.

In recent years, much attention has been given to community detection and the implementation of efficient algorithms. Some of the most notable such algorithms are greedy modularity optimization [2], random walk based algorithms e.g. Infomap [3] and stochastic blockmodels [4]. All of these algorithms have their drawbacks, in some way or another. Modularity optimization is known to suffer from its resolution limit [5] which limits the scale at which it can detect communities. Infomap has been shown to be a farily stable and efficient algorithm overall [6]. Both greedy optimization and infomap are discriminative methods, which partition graph nodes into communities. Stochastic blockmodels are a generative model and is be more closely described in Section III. The main drawback of stochastic blockmodels is that the number of blocks or communities must be known in advance.

This paper is organized as follows: in Section II, we provide a brief overview of various image segmentation methods. We describe how image segmentation can be posed as a graph based problem. We then shift our focus onto a recent paper by Browet et al. [7] who propose an image segmentation scheme based on a modified greedy modularity optimization method. In Section III we provide an overview of the stochastic blockmodel and recent developments. Finally, in Section IV, we show some experimental results on real images. Finally, Section V concludes the paper and proposes future work.

II. Related work

A. Image segmentation as a graph problem

Before we can apply graph based methods for image segmentation, the image itself must be represented as a graph. This is typically done by building an undirected weighted graph G where each node represents a pixel in the input image. Each weighted edge of the graph G represents the similarity between pairs of pixels i and j and is stored in the weighted adjacency matrix W. Typically, we define W_{ij} as

$$W_{ij} = \begin{cases} e^{\frac{d(i,j)^2}{\sigma_x^2}} e^{\frac{|F(i)-F(j)|^2}{\sigma_i^2}} & \text{if } d(i,j) < d_{max}, \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

where the term $d(i,j)$ represents spatial proximity using some distance metric between pixels i and j (most commonly the Euclidean distance) and $F(i)$ is some feature vector based on the properties of the pixel. This feature vector is typically the scalar intensity of the pixel for gray-scale images or the HSV transform for colored images [7]. d_{max} specifies an upper bound on pixels that we will consider based on their spatial proximity to the pixel in question. This parameter is necessary due to practical considerations, since the total number of connections between pixels is n^2, which is, in practical regards, computationally unfeasible, therefore it is necessary to restrict ourselves to some region around individual pixels. To illustrate this, we can consider a small image by today's standards, a 480×320 image. This image contains 153600 pixels in total and were we to connect every single pixel to one another, this would result in 23,592,960,000 edges. This not only has computer memory considerations but will also impact the running time of our algorithms. The last two parameters σ_x and σ_i are user defined parameters that control the impact of pixel distances and pixel similarities in the edge weight. A lower value of σ_x induces faster decay i.e. distances will drop off exponentially, whereas larger values will induce a slower decay. Likewise, a lower value of σ_i induces faster decay i.e. the higher the difference in pixel features, the lower the similarity.

This is not a new approach. In fact, it was first proposed (to the best of our knowledge) in 2000 by Shi & Malik alongside normalized cuts. Normalized cuts attempt to find such a set of edges that creates disconnected components when removed from the graph, such that the sum of the weighted edges is minimal. These methods are typically referred to as *cuts* [1]. These methods produce good results and are still in use today.

B. Community detection for image segmentation

Normalized cuts and most previous graph based methods utilize flow optimization techniques. More recently, community detection algorithms have been tested for the image segmentation task. Browet et al. propose a variant of the Louvain method for greedy modularity optimization. The tests were performed on the Berkley segmentation benchmark dataset (BSDS) [8] and appears to work reasonably well.

They show that the standard greedy modularity optimization approach typically used for community detection produces coherent regions, but most objects in the images are over-segmented. They argue that this over-segmentation is due to the fact that modularity optimization algorithms cannot yield communities with long chains of nodes. They propose two problems to tackle this problem. Firstly, increasing d_max is the theoretically sound way to go about solving this problem, but is often, as previously discussed, unfeasible for all but the smallest images. The second remedy, which they then implement is to modify the null model used in modularity. We describe their approach next.

Modularity is a well studied measure of community goodness [9]. It provides a nice tradeoff between the edges actually observed in the graph and the number of edges we would expect to appear at random given a random null model. The basic expression for the modularity measure for weighted networks is given in equation 2.

$$Q = \frac{1}{2m} \sum_{i,j=1}^{n} (w_{ij} - N_{ij})\, \delta_{C_i C_j} \qquad (2)$$

where n is the number of nodes, w_{ij} is the edge weight, N_{ij} is a null model between nodes i and j and $\delta_{C_i C_j}$ is the Kronecker delta function. In this case, m is the sum of weights of all the edges in the graph and is a scaling parameter ensuring that $Q \in [-1, 1]$.

They then propose a modified version of modularity, which enables regions to cover large portions of the graph. More precisely, the introduce a weight matrix Λ with which they modify the null model $N_{ij} = \Lambda_{ij} \frac{k_i k_j}{2m}$ where

$$\Lambda_{ij} = \begin{cases} 1 & \text{if } d(i,j) \leq d_\Lambda, \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

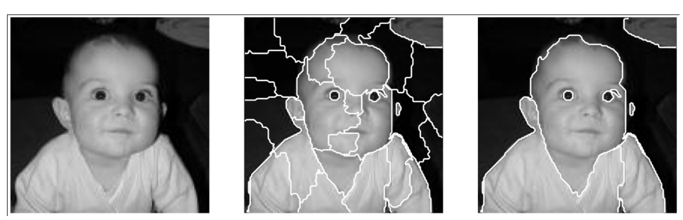

Figure 2. Image segmentation using the Louvain greedy modularity optimization method (center) produces an over-segmented image. Browet et al. [7] propose a modified modularity measure which empirically allows communities to cover larger regions of the network (right).

This modified version of modularity appears to produce reasonable results, shown in Figure 2. However, the method is never actually benchmarked to other image segmentation techniques and no score is provided [7], therefore we cannot know how well it compares to other state of the art image segmentation techniques.

III. STOCHASTIC BLOCKMODELS

The stochastic blockmodel (SBM) [4] is a generative model for community detection which makes use of the statistical properties of the network to infer the high level modular structure. In its most basic form, the model assigns each node i to one of K modules denoted by z_i. Each edge exists with probability $\Theta_{z_i z_j}$ where Θ is a $K \times K$ matrix which values are the probabilities of links appearing between blocks z_i and z_j.

However, a fundamental limitation of most SBM implementations is that they are defined strictly for simple or multigraphs. This means that they cannot make use of various additional data such as edge weights, which are usually a very rich source of information. Moreover, in the case of image segmentation, the number of blocks K must be known in advance. This clearly does not fit into a general framework for detecting regions of interest in images.

The first aforementioned issue was addressed by Aicher et al. [10] who adapted the SBM to weighted networks by including edge values as additional covariates. The second issue was just recently addressed by Piexoto [11] who build on the ideas of Aicher et al. and produce a nonparametric Bayesian approach, which is capable of probabilistically inferring the number of blocks in the model. The model relies on a Bayesian hierarchy of uninformative hyperpriors that remain agnostic of the number of groups, the sizes of the groups and the partition of the nodes. A very detailed derivation of this model is given in [12]. They show that this nonparametric model is capable of correctly recovering community structure from various empirical networks. The inference step used is very efficient, making use of Markov Chain Monte Carlo (MCMC) sampling, that requires only $O(E)$ operations per sweep, where E is the number of edges in the network.

IV. RESULTS

We apply the SBM to graphs obtained from images as described in Section II-A. The segmentations are shown in Figures 3 and 4.

Clearly, SBMs suffer from a similar drawback as the one discussed in the Louvain modularity optimization problem. A good community is defined as partitions of the graph where edges among nodes are more likely than edges outside of the community. This fundamental definition inherently contradicts our graph building process, with which we are forced to limit the number of edges we add to the graph. As such, most community optimization algorithms will struggle to find good communities and will produce similar results to what we observe with the SBM. While the regions it finds are reasonable, they are far too over-segmented to be useful in any context. The results of the SBM remind of another useful computer vision technique called superpixel – a preprocessing step which groups together similar regions in an image to make computation simpler. However, even though the SBM produces similar results, it is computationally far more expensive, and should not be used as a replacement.

Figure 3. Image segmentation using a nonparametric SBM produces an overly segmented image.

Figure 4. Another example using the nonparametric SBM. Again, the results show an overly segmented image.

We also consider several different values for the user parameters which regulate the strength of distance and pixel similarities as edge weights. We found that no setting of values drastically improves the results.

V. CONCLUSION

REFERENCES

[1] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[2] A. Clauset, M. E. Newman, and C. Moore, “Finding community structure in very large networks,” *Physical review E*, vol. 70, no. 6, p. 066111, 2004.

[3] M. Rosvall and C. T. Bergstrom, “Maps of random walks on complex networks reveal community structure,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.

[4] P. W. Holland, K. B. Laskey, and S. Leinhardt, “Stochastic blockmodels: First steps,” *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.

[5] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 36–41, 2007.

[6] G. K. Orman, V. Labatut, and H. Cherifi, "On accuracy of community structure discovery algorithms," *arXiv preprint arXiv:1112.4134*, 2011.

[7] A. Browet, P.-A. Absil, and P. Van Dooren, "Community detection for hierarchical image segmentation.," in *IWCIA*, vol. 11, pp. 358–371, Springer, 2011.

[8] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, pp. 416–423, July 2001.

[9] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[10] C. Aicher, A. Z. Jacobs, and A. Clauset, "Learning latent block structure in weighted networks," *Journal of Complex Networks*, vol. 3, no. 2, pp. 221–248, 2014.

[11] T. P. Peixoto, "Nonparametric weighted stochastic block models," *arXiv preprint arXiv:1708.01432*, 2017.

[12] T. P. Peixoto, "Bayesian stochastic blockmodeling," *arXiv preprint arXiv:1705.10225*, 2017.