Uncertainty Quantification for Semi-supervised Multi-class Classification in Image Processing and Ego-Motion Analysis of Body-Worn Videos

Yiling Qiao[1], Chang Shi[2], Chenjian Wang[3], Hao Li[3], Matt Haberland[4], Xiyang Luo[3], Andrew M. Stuart[5], Andrea L. Bertozzi[3];
[1] University of Chinese Academy of Science; Beijing, China
[2] Renmin University of China; Beijing, China
[3] University of California, Los Angeles; Los Angeles, California, USA
[4] California Polytechnic State University; San Luis Obispo, California, USA
[5] California Institute of Technology; Los Angeles, California, USA

Abstract

Semi-supervised learning has gained much attention as a way to utilize underlying relationships in the data in face of a scarcity of ground truth labels. In this paper, we introduce an uncertainty quantification (UQ) method for graph-based semi-supervised multi-class classification problems. We not only predict the class label for each data point, but also provide a confidence score for the prediction. We adopt a Bayesian approach and propose a graphical multi-class probit model together with an effective sampling procedure. Further more, we propose a confidence measure for each data point that correlates with classification performance. We use the empirical properties of the proposed confidence measure to guide the design of a human-in-the-loop system. The uncertainty quantification algorithm and the human-in-the-loop system are successfully applied to classification problems in image processing and ego-motion analysis of body-worn videos.

Introduction

Applications such as police body-worn video cameras generate a huge amount of data, beyond what is humanly possible to review by analysts. Such problems are ripe for the development of semi- supervised learning algorithms, which, by definition, use a small amount of training data. In the last year, progress has been made in applying graphical models in machine learning to body-worn videos with the goal of recognizing the camera-wearer's activity, i.e., ego-motion [9, 4]. However, as is often the case with real-world videos, the variability of the data leads to imperfect classification. Recently, the authors of [2] proposed to pair uncertainty quantification to the binary classification problem on a similarity graph. Besides a label assigned to each data point, a measure of uncertainty is also estimated, which enables us to identify hard-to-classify data points that require further investigation.

In the present paper, we push the UQ methodology to a multi-class classification setting. We extend the binary graph probit method to a multi-class version and develop a Gibbs sampler that enables us to draw samples from the posterior distribution. We propose a confidence measure for each data point that we find correlates with classification performance; we observe data points with higher confidence scores are more likely to be classified correctly. Along with the new methodology and the empirical observations, we develop the foundations for a system with a human in the loop who serves to provide additional class labels based on the confidence scores; our uncertainty quantification method identifies hard-to-classify data points and the human in the loop assigns ground truth to them. Our ideas are tested on an image data set — the MNIST data set [8] — and a body-worn video data set, the HUJI EgoSeg data set [12].

Related Work

Semi-supervised learning has been studied extensively in the past two decades, and has been successfully applied to applications such as hyperspectral images [10] and body-worn videos [9, 4]. We refer readers to [20] and the more recent article [1] for a literature review. We focus on graph-based methods, in which a similarity is measured for each pair of nodes (i.e. data points) and label information is spread across the similarity graph from a small set of labeled fidelity points. The similarity information is often leveraged via the graph Laplacian, which has been been used in a myriad of machine learning methods (see, for instance, [17, 18, 19, 21]). The analogy between the graph Laplacian and the classical Laplacian operator inspires a number of PDE-based classification methods, such as [1, 7]; this also introduces to the machine learning community the recent development of uncertainty quantification methods, which are more commonly studied for PDE-based inverse problems. For instance, in their recent work [2], the authors used an efficient sampling method that was originally developed for PDE-based inverse problems [5] to perform uncertainty quantification for binary classification problems.

We refer readers to the books [15, 16] and the recent article [11] for a review of methodologies employed in the field of uncertainty quantification. For the specific application to machine learning methods, the book [18] investigates uncertainty quantification for a variety machine learning problems using a Gaussian process prior. Except the above-mentioned book and the recent work [2], most of machine learning methods, even those developed with the Bayesian way of thinking, focused on finding the optimal classification (and/or hyperparameters that produce the optimal classification) in an optimization context and did not consider or utilize uncertainty quantification.

Methodology

Graphical Setting

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of feature vectors, where $x_i \in \mathbb{R}^d$. Let $Z = \{1, 2, \ldots, n\}$ index the entire dataset and we let $Z' \subset Z$ be a fidelity set consisting of nodes of known labels. We aim to classify n data points into c classes such that:

1) data points with similar feature vectors, measured via a suitable similarity measure, should belong to the same class;
2) the classification should respect the ground-truth labels on the fidelity set.

We consider each data point as a node in a weighted graph, where the edge weights are given by

$$w_{ij} = \exp\left(-\|x_i - x_j\|^2/\tau_{ij}\right),$$

where $\|\cdot\|$ is the Euclidean distance and τ_{ij} are the self-tuning constants proposed in [19]. The weights are chosen such that a pair of nodes with similar feature vectors will have a weight close to 1 and dissimilar nodes will have a near-zero weight. Suppose $u : Z \to \mathbb{R}^c$ is an assignment function; we interpret $u_\ell(i) = \max_{\hat{\ell}} u_{\hat{\ell}}(i)$ as u assigns data point i to class ℓ. One way to achieve a classification is to optimize the following objective function with respect to an assignment function u:

$$J(u) = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}\|u(i) - u(j)\|^2 + \Phi(u, u'), \tag{1}$$

where u' encodes the ground-truth labels on the fidelity set Z', $\Phi(u, u')$ measures the extent to which u differs from u' on Z'. Minimizing the first term in the object function ensures that a pair of data points (i, j) with a high similarity w_{ij} will be assigned to the same class.

Using matrix notation, we identify u and u' with $n \times c$ matrices so that $u_{i\ell} = u_\ell(i)$. If we let W be the matrix of w_{ij} and $D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$ where $d_i = \sum_j w_{ij}$, we can introduce the graph Lapalcian,

$$L = D - W \tag{2}$$

and the Dirichlet energy

$$\langle u, Lu\rangle = \sum_{i,j=1}^{n} w_{ij}\|u(i) - u(j)\|^2, \tag{3}$$

where $\langle u, v\rangle = \mathrm{trace}\left(u^T v\right)$, and hence we may write eq. (1) as

$$J(u) = \frac{1}{2}\langle u, Lu\rangle + \Phi(u, u') \tag{4}$$

The quadratic form in eq. (3) alludes to the connection to Bayesian Gaussian process models.

It is common in graph-based learning methods to use normalized variants of the graph Laplacian in place of the unnormalized graph Laplacian eq. (2) because of better numerical properties as well as the classification performance (see, for instance, [1]). One popular choice is the symmetrically normalized graph Laplacian,

$$L_{\mathrm{sym}} = D^{-1/2} L D^{-1/2}, \tag{5}$$

which is convenient to compute with due to its symmetry. With the choice of the symmetrically normalized graph Laplacian, the quadratic form in eq. (3) becomes,

$$\langle u, L_{\mathrm{sym}} u\rangle = \sum_{i,j=1}^{n} w_{ij}\left\|\frac{u(i)}{\sqrt{d_i}} - \frac{u(j)}{\sqrt{d_j}}\right\|^2.$$

In the remaining of this manuscript, the notation L is a placeholder for any choice of graph Laplacian.

Bayesian model

We now present a Bayesian model for the assignment function u, of which the posterior distribution takes the form:

$$p(u|u') \propto \exp(-J(u)), \tag{6}$$

so a maximum a posteriori probability (MAP) estimator is a minimizer of $J(u)$. We assume the prior on u is a Gaussian distribution,

$$p(u) \propto \exp\left(-\frac{1}{2}\langle u, Lu\rangle\right).$$

To explicitly construct a sample u that follows the prior distribution, we employ the Karhunen-Loéve expansion. Let $L = Q\Lambda Q^T$ be the eigen-decomposition of the graph Laplacian where columns of $Q \in \mathbb{R}^{n\times n}$ form an orthonormal basis of $\mathbb{R}^n$ and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ obeys

$$0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n.$$

We observe that L is positive semi-definite. Suppose $\{\xi_i\}_{i=1}^n$ is a collection of independent c-variate normal random variables $\mathcal{N}(0, I_c)$, where I_c is an identity matrix of size c. We construct a sample u as the random sum

$$u = \sum_{i=2}^{n} \lambda_i^{-1/2} q_i \xi_i^T,$$

so that columns of u live in $\mathrm{span}\{q_1\}^\perp$ and u has the desired probability distribution

$$p(u) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\sum_{\ell=1}^{c}\lambda_i\langle u_\ell, q_i\rangle^2\right) = \exp\left(-\frac{1}{2}\langle u, Lu\rangle\right). \tag{7}$$

In graph-based semi-supervised learning, it is very common to approximate the graph Laplacian using its first few eigenvectors and eigenvalues, for those often contain rich geometric information of the graph (see, for instance, [17]). Such truncation of the spectrum both reduces computation cost and often improves the classification performance. In this case, we also employ spectral truncation and let u be the random sum up to K for $K \ll n$, i.e.

$$u = \sum_{i=2}^{K} \lambda_i^{-1/2} q_i \xi_i^T. \tag{8}$$

In their recent work [2], the authors considered several likelihood functions $p(u'|u)$ to connect the latent variable u to the ground-truth labeling u' for binary classification. In the present paper, we primarily investigate the independent probit likelihood function. Suppose $\{\eta(i)\}$ for $i \in Z'$ is a collection of independent c-variate normal random variables $\mathcal{N}(0, \gamma^2 I_c)$ where γ^2 is the noise variance. We connect u to u' via

$$v(i) = u(i) + \eta(i)$$
$$u'(i) = \text{threshold}\,(v(i))\,, i \in Z'.$$

The threshold operator applied to a vector simply sets the largest element in the vector to be 1 and the rest to be 0. With the introduction of latent variables $\{v(i)\}_{i\in Z'}$, we have, from the Bayes formula, the following joint posterior probability distribution,

$$p(u, v|u') \propto \exp\left(-\frac{1}{2}\langle u, Lu\rangle - \frac{1}{2\gamma^2}\sum_{i\in Z'}\|u(i) - v(i)\|^2\right).$$

Recall that we have the change of variable from u to ξ in eq. (8); we can apply the sampling method of choice to $p(\xi, v|u')$ for $\xi = (\xi_1, \xi_2, \cdots, \xi_K) \in \mathbb{R}^{c\times K}$. We compute that the joint probability

$$p(\xi, v|u') \propto \exp\left(-\frac{1}{2}\langle \xi^T, \Lambda'\xi^T\rangle - \frac{1}{2\gamma^2}\|HQ'\xi^T - v\|^2\right),$$

where $\Lambda' = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_K)$, the matrix $Q' \in \mathbb{R}^{n\times K}$ consists of the first K eigenvectors of the graph Laplacian, and $H = (\delta_{ij_i}) \in \mathbb{R}^{|Z'|\times n}$ for $Z' = \{j_i : i = 1, 2, \cdots, |Z'|\}$. We note that H applied to a matrix selects the rows of which indexes are contained in Z'.

To sample from the joint posterior distribution, a Gibbs sampler will alternate between the following three steps:

1) Draw ξ from $p(\xi|v, u')$,
2) Construct u from ξ via eq. (8),
3) Draw v from $p(v|u, u')$.

For Step 1) we note that for each $\ell \in \{1, 2, \cdots, c\}$, the conditional probability for each row of ξ, denoted as $p(\xi_{:,\ell}|v, u')$ has the same distribution as

$$\mathcal{N}\left(m, P^{-1}\right),\; P = \Lambda' + \frac{1}{\gamma^2}Q'^T H^T H Q',\; m = \frac{1}{\gamma^2}P^{-1}Q'^T H^T v_\ell.$$

In Step 3), for each $i \in Z'$, we need to sample a c-variate normal random variable subject to a linear inequality constraint; let a_i denote the unique index such that $u_{a_i}(i)' = 1$ for $i \in Z'$, i.e., data point i belongs to class a_i according to the ground-truth label, we need to sample $v(i)$ according to the following conditions:

$$v(i) \sim \mathcal{N}\left(u(i), \gamma^2 I_c\right), \quad v_{a_i}(i) \geq v_\ell(i) \,\text{for all}\, \ell \in \{1, 2, \cdots, c\}.$$

We use the implementation from [3] to efficiently draw samples from the linearly constrained normal distribution.

Uncertainty Quantification

Given a set of samples $\{u^{(k)}\}_{k=1}^N$ from the Gibbs sampler, we investigate $\mathbb{E}_{u|u'}(\text{threshold}(u))$, the posterior mean of threshold(u), which can be approximated by the sample mean

$$s_\ell(i) = \mathbb{E}_{u|u'}(\text{threshold}(u(i))_\ell) \approx \frac{1}{N}\sum_{k=1}^N \text{threshold}(u(i))_\ell.$$

Since each element threshold$(u(i))_\ell$ is either zero or one, the expectation $s_\ell(i)$ simply gives the probability, under the posterior distribution, of the element being one; that is $s_\ell(i)$ can be interpreted as the probability data point i is classified as class ℓ. We note that for each data point, the probability of it being classified to each class should sum to one, i.e., $\sum_\ell s_\ell(i) = 1$. This is obeyed by both the posterior mean and the sample mean approximation. We can use the posterior mean $s(i)$ as a classifier, which classifies data point i according to its largest entry.

Intuitively, a single large $s_\ell(i)$ for a data point i indicates a very confident classification of class ℓ; in this case, the remaining entries in the row $s(i)$ are necessarily small due to the sum-to-one condition; this creates a large variance in the row $s(i)$. On the other hand, if entries in the vector $s(i)$ are all roughly equal, meaning the data point is equally likely to be classified as either class, the classification has a lot of uncertainty, resulting in an $s(i)$ with a small variance. Based on this intuition, we measure the classification confidence of node i by the variance of $s(i)$

$$S(i) = \text{var}\,(s(i)) = \frac{1}{c}\sum_{\ell=1}^c\left(s_\ell(i) - \frac{1}{c}\sum_{\ell=1}^c s_\ell(i)\right)^2.$$

We emphasize that this variance is not the posterior variance. However, we can show the following connection between the quantity $S(i)$ and the posterior variance

$$S(i) = \frac{1}{c} - \frac{1}{c^2} - \frac{1}{c}\sum_{\ell=1}^c \text{var}_{u|u'}(\text{threshold}(u_i)_\ell),$$

where $\text{var}_{u|u'}(\cdot)$ is the posterior variance. Therefore, the quantity $S(i)$ is a constant minus the mean posterior variance, which can be interpreted as a measure of uncertainty, averaged over all classes.

Human-in-the-loop

In the following experiment section, we demonstrate a positive correlation between the proposed confidence score and classification performance; the confidence score enables us to locate hard-to-classify data points, which we may label and use as additional fidelity point. This naturally leads to the idea of using the confidence measure to intelligently select new fidelity points to achieve better classification performance with limited human labeling effort. We design a human-in-the-loop algorithm as follows.

We start with a small set of initial known fidelity points and apply the UQ algorithm to obtain a confidence score for the entire data set. We randomly sample, in a uniform fashion, additional candidate fidelity points with confidence scores within a percentile range. The human in the loop then observes each of the candidate fidelity points to assign ground truth to them. We perform the UQ algorithm again to update the confidence scores and repeat the process until we reach the maximum number of fidelity points permitted (this will be determined by the application).

We observe that in practice adding data points with the lowest confidence scores does not benefit overall classification performance because these data points are often outliers. The significance of classifying these outliers correctly is scenario dependent. In our experiments, we focus on the overall accuracy and do not sample fidelity from data points with confidence scores lower than the tenth percentile.

Experiments

We perform uncertainty quantification on 1) the MNIST data set, a handwritten digit data set, and 2) the HUJI EgoSeg data set, a body-worn video data set. Through these experiments, we illustrate some empirical properties of the confidence score; we demonstrate its correlation with the classification performance. We also validate our human-in-the-loop framework and showcase its ability to improve classification results with limited human input.

MNIST

The MNIST data set [8] consists of 70,000 images of handwritten digits; each image is of the size 28×28 pixels. We choose uniformly at random 500 images each from the digits 1, 4, 7, and 9 to form a graph of 2000 nodes. We follow the graph construction procedure in [2]; each image is projected onto the lead 50 principal components yielding a 50-dimensional feature vector, and we construct a 15-nearest neighbor graph. The weighting constants τ_{ij} are chosen according to [19]. For data point i, we compute the mean distance of the 15 nearest neighbors, denoted as τ_i; then the weighting constant τ_{ij} is given by $\tau_{ij} = \tau_i \tau_j$. We use the symmetrically normalized graph Laplacian (see eq. (5)) and truncate its spectrum at $K = 300$. We perform the Gibbs sampler with 3% uniformly randomly sampled fidelity points; the noise variance is chosen to be $\gamma = 0.1$, and we draw 2×10^4 samples to estimate the uncertainty. We showcase examples of images with the highest or the lowest confidence scores in fig. 1. It is interesting to note that the lowest confidence score of digit 1 is much higher than that of the other digits; we theorize that it is easier for the algorithm to differentiate digit 1 from the other three digits.

Body-Worm Videos

We also apply our method to the HUJI EgoSeg data set [12] [13]. This data set contains 65 hours of egocentric videos including 44 videos shot using a head-mounted GoPro Hero3+, the Disney data set [6] and other YouTube videos[1]. In a recent work [4], a graph-based semi-supervised learning method is applied to this data set to classify video segments according to camera-wearer's activities and showed promising results. This data set consists of footage of 7 activities: *Walking, Driving, Riding Bus, Biking, Standing, Sitting, and Static*. We follow the same feature extraction procedure described in [4] to obtain a 50-dimensional feature vector for every 4-second video segments; this yields 36,421 segments. To speed up our calculation, we sample every fifth segment. The graph is constructed from the 50-dimensional feature vectors, and the weighting constants $\tau_{ij} = \tau_i \tau_j$ are chosen according to [19], where τ_i is the distance of the 40th nearest neighbor of node i. We employ the symmetrically normalized graph Laplacian and truncate the spectrum at $K = 400$. The Gibbs sampler is applied with $\gamma = 0.1$ and 2×10^4 iterations.

The data set is separated into a training and testing set, which are disjoint sets of videos; the training set contains around 65% of the data, measured in terms of the footage length. We refer readers to [14] for the details of the experiment protocol. However, we do not use the full training set but instead take a portion of it as the fidelity. All classification performances are evaluated on the testing set only. We first investigate the correlation between the confidence score and classification accuracy. We perform uncertainty quantification with 12% of the training set. Recall that the classification is produced by taking the largest entry of the posterior mean $s(i)$ for each data point i. In fig. 2, we plot the classification accuracy of the top x% to $x+5$% confident data points for each $x \in \{0, 5, 10, \cdots, 95\}$. We observe that the classification is more accurate on on data points with higher confidence scores.

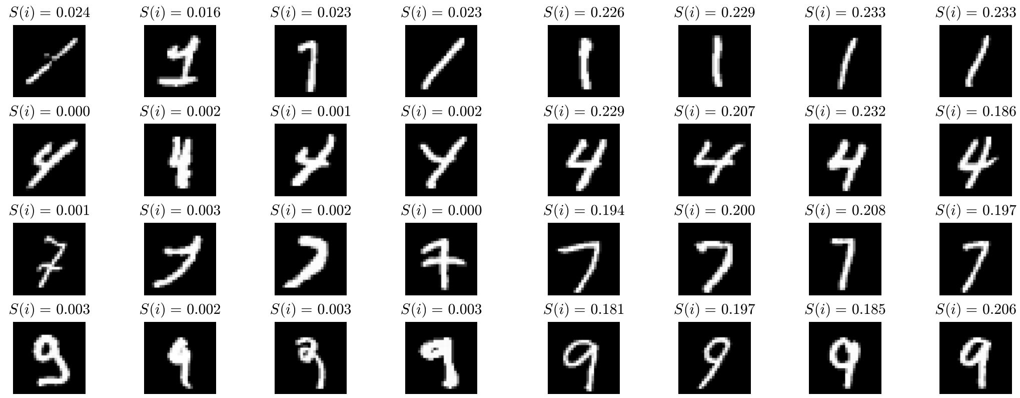
We also validate our human-in-the-loop system on this data set. We start with 6% fidelity data and gradually increase the fidelity percentage to 30% over five iterations; at each iteration, we introduce additional 6% fidelity points sampled from data points with confidence scores within in the range of the tenth and 50th percentile. We perform uncertainty quantification as well as a graph-based semi-supervised learning method (an MBO scheme [1]) using the same set of fidelity points; we compare the classification performance, measured in terms of accuracy and mean recall averaged over seven classes, of both classifiers using iteratively generated fidelity against the same classifiers using uniformly randomly sampled(uninformed) fidelity of the same percentage. The results are presented in fig. 3. We observe that both classifier benefit from the intelligently sampled fidelity in terms of producing higher accuracy and mean recall than using uninformed fidelity.

Conclusion

In this paper, we considered the problem of uncertainty quantification in a graph-based semi-supervised multi-class classification problem . We extended the graph probit model, originally proposed for the binary classification problem by [2], to a multi-class case. We proposed a Gibbs sampler to sample from the posterior distribution and a confidence score that connects to the posterior variance. Through our experiments on the MNIST dat set, we demonstrate the proposed confidence score is easy to interpret; it is clear to see the contrast between the digit images with low confidence scores and high confidence scores. The proposed confidence score also exhibited a correlation with the classification performance in our experiments on the HUJI EgoSeg data set. Based on this observation, we designed a human-in-the-loop system to efficiently use human labeling effort to improve the classification results. We validated this system on the HUJI EgoSeg data set and observed that the classifiers that we studied produced improved classification using the human-in-the-loop system than the same classifiers using uninformed fidelity.

Moving forward, we can develop new theory of the uncertainty quantification for semi-supervised classification problems. We can investigate the performance bound of the Gibbs sampler with respect to the number of data points and the number of classes. We can extend the previous analysis of uncertainty quantification methods for binary classification problems to the multiclass case, in which we suspect the number of classes play a nontrivial role in the performance of the sampling methods. We also point out that speed is the primary concern of the current Gibbs sampler. Despite the development of scalable graph-based semi-supervised learning methods (see [1] for an example), the Gibbs sampler is mostly sequential; we draw each sample based on the previous one. Typically, we need 1×10^4 samples to achieve a reasonable uncertainty quantification; this makes it difficult to apply the current method to large data set if we wish the algorithm finishes within minutes. We will continue developing a more scal-

[1] `http://www.vision.huji.ac.il/egoseg/`



(a) Low confidence (b) High confidence

Figure 1. *A demonstration of uncertainty quantification on the MNIST dataset. Recall that $S(i)$ is the proposed confidence score; a higher confidence score entails a more confident classification. The presented images are chosen from the ten images with the lowest/highest confidence score within each class.*

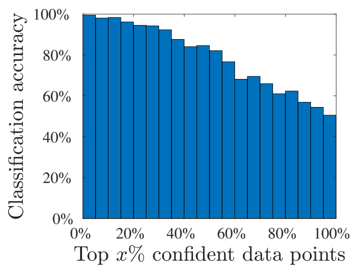
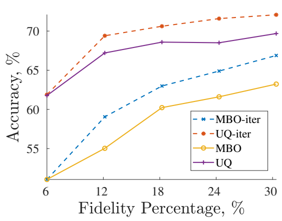
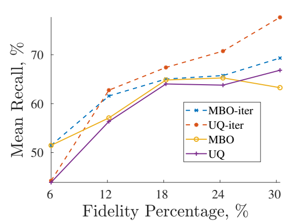


Figure 2. *Classification accuracy on data points with top $x\%$ to $x+5\%$ confidence scores on the HUJI EgoSeg data set. We group data points based on their confidence score; each group contains 5% data points and we evaluate the classification accuracy on each group.*



(a) Accuracy



(b) Recall

Figure 3. *Classification performance of UQ and an MBO classifier using iteratively generated fidelity (UQ/MBO-iter) and uniformly randomly sampled fidelity (UQ/MBO) on the HUJI EgoSeg data set.*

able sampling procedure.

Acknowledgement

We would like to thank Hannah Droege, Sara Tro, and Yang Wang for their contributions during an undergraduate research training program at UCLA. We acknowledge the financial support of the NSF (grant DMS-1737770 and grant DMS-1417674). We also thank the UCLA-CSST program for their financial support for Yiling Qiao and Chang Shi.

References

[1] A. L. Bertozzi and A. Flenner. Diffuse interface models on graphs for classification of high dimensional data. *SIAM Review*, 58(2):293–328, 2016.

[2] A. L. Bertozzi, X. Luo, A. M. Stuart, and K. C. Zygalakis. Uncertainty quantification in the classification of high dimensional data. *SIAM/ASA J. Uncertainty Quantification*, 6(2):568–595, 2018.

[3] Z. I. Botev and P. L'Ecuyer. Efficient probability estimation and simulation of the truncated multivariate student-t distribution. In *Proceedings of the 2015 Winter Simulation Conference*, pages 380–391. IEEE Press, 2015.

[4] H. Chen, H. Li, A. Song, M. Haberland, O. Akar, A. Dhillon, T. Zhou, A. L. Bertozzi, and J. P. Brantingham. First-person activity recognition in body-worn video. 2018.

[5] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, pages 424–446, 2013.

[6] A. Fathi, J. K. Hodgins, and J. M. Rehg. Social interactions: A first-person perspective. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1226–1233. IEEE, 2012.

[7] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. G. Percus. Multiclass data segmentation using diffuse interface methods on graphs. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1600–1613, 2014.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[9] Z. Meng, J. Sanchez, J.-M. Morel, A. Bertozzi, and B. P.Jeffrey. Ego-motion classification for body-worn videos. In *2016 Conference on Imaging, vision and learning based on optimization and PDEs*, 2016.

[10] E. Merkurjev, J. Sunu, and A. L. Bertozzi. Graph MBO method for multiclass segmentation of hyperspectral stand-off detection video. In *2014 IEEE International Conference on Image Processing (ICIP)*.

[11] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. Optimal uncertainty quantification. *Siam Review*, 55(2):271–345, 2013.

[12] Y. Poleg, C. Arora, and S. Peleg. Temporal segmentation of egocentric videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2537–2544, 2014.

[13] Y. Poleg, A. Ephrat, S. Peleg, and C. Arora. Compact cnn for indexing egocentric videos. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pages 1–9. IEEE, 2016.

[14] Y. Poleg, A. Ephrat, S. Peleg, and C. Arora. Compact CNN for indexing egocentric videos. In *WACV*, 2016.

[15] R. C. Smith. *Uncertainty quantification: theory, implementation, and applications*, volume 12. SIAM, 2013.

[16] T. J. Sullivan. *Introduction to uncertainty quantification*, volume 63. Springer, 2015.

[17] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[18] C. K. Williams and C. E. Rasmussen. Gaussian processes for regression. In *Advances in neural information processing systems*, pages 514–520, 1996.

[19] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *Advances in neural information processing systems*, pages 1601–1608, 2005.

[20] X. Zhu. Semi-supervised learning literature survey. *Computer Science, University of Wisconsin-Madison*, 2(3):4, 2006.

[21] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919, 2003.