Interactive Segmentation in Multimodal Medical Imagery using a Bayesian Transductive Learning Approach

Noah Lee[†], Jesus Caban[‡], Shahram Ebadollahi[‡], Andrew Laine[†]

[†]Heffner Biomedical Imaging Laboratory, 1210 Amsterdam Avenue, New York, NY, USA 10027
[‡]IBM T.J. Watson Research Center, 19 Skyline Drive Hawthorne, NY USA 10532

ABSTRACT

Labeled training data in the medical domain is rare and expensive to obtain. The lack of labeled multimodal medical image data is a major obstacle for devising learning-based interactive segmentation tools. Transductive learning (TL) or semi-supervised learning (SSL) offers a workaround by leveraging unlabeled and labeled data to infer labels for the test set given a small portion of label information. In this paper we propose a novel algorithm for interactive segmentation using transductive learning and inference in conditional mixture naïve Bayes models (T-CMNB) with spatial regularization constraints. T-CMNB is an extension of the transductive naïve Bayes algorithm [1, 20] to the semi-nonparametric case. The multimodal mixture assumption on each covariate feature dimension and spatial regularization constraints allow us to explain more complex distributions required for spatial classification in multimodal imagery. To simplify the estimation we reduce the parameter space by assuming naïve conditional independence between the feature space and the class label. The naïve conditional independence assumption allows efficient inference of marginal and conditional distributions for large scale learning and inference [19]. We evaluate the proposed algorithm on multimodal MRI brain imagery using ROC statistics and provide preliminary results. The algorithm shows promising segmentation performance with a sensitivity and specificity of 90.37% and 99.74% respectively and compares competitively to alternative interactive segmentation schemes.

Keywords: Brain, Observer Studies, Classifier design, Segmentation, Machine Learning.

1. INTRODUCTION

In the medical domain the labeling process of image data requires expert knowledge and often tedious editing effort to obtain accurate label information for the object of interest. In the realm of computer aided diagnosis interactive segmentation schemes are well received by physicians, where the combination of human and machine intelligence can provide improved segmentation efficacy at minimal expert intervention [31]. Transductive learning (TL) or semi-supervised learning (SSL) is a suitable framework for learning-based interactive segmentation tools that require minimal expert intervention. To cope with the scarce label problem transduction offers a workaround by leveraging the labels provided for a small portion of the data to label the remaining test set. Such information enables efficient quantification of multimodal image data that can be used for exploring potential relationships to other indicative factors to assess the patient's disease progress [32].

The *inductive learning* formulation considers a function $f : X \rightarrow Y$ that maps instances from the entire input space $\mathbf{x} \in X$ to output labels $y = f(\mathbf{x})$. In inductive function learning we seek to form a hypothesis $\mathcal{H}$ that can recover f given a training set of example pairs $\{(\mathbf{x}_n, y_n)\}_{n=1,\ldots,N}$. However, obtaining label information $\{y_n\}$ is an expensive and time intensive process whereas the input space X is abundant. In high dimensional data domains such as multimodal medical imagery the sample complexity of the training set at learning stage may be incomplete for reliable generalization performance [2]. An alternative learning formulation is *transductive learning* and *inference* [3] motivated by Vapnik in the 90's [4]. In transductive inference the learning machine is given a labeled training set and an unlabeled test set[1] with the goal to find a hypothesis $\mathcal{H}_t$ that can classify or predict unknown class labels only for the given test set. Consequently the transductive learning machine can explore the labeled training and unlabeled test set to learn a simpler $\mathcal{H}_t$. The usefulness of the unlabeled test set in transduction has also been advocated in the context of *co-training* [5, 6]

---

[1] The test set is often called working or shadow set in the literature.

and *semi-supervised learning* (SSL) [7-9]. The problem of learning from labeled and unlabeled data has attracted considerable attention in recent years including discriminative transduction [10], manifold regularization [11] and graph based techniques [12, 13] given their effectiveness on quickly labeling a given set of input data.

In this paper we propose a novel transductive learning machine for spatial classification casted as an interactive segmentation problem. We present a conditional mixture naïve Bayes model (T-CMNB) with spatial regularization constraints in a transductive learning and inference setting. We make use of unconditional and conditional Gaussian mixture models on each covariate feature dimension to learn and infer the relationships in $\mathcal{P}_{XY}$ using naïve Bayesian transduction. The naïve conditional independence assumption allows efficient inference of marginal and conditional distributions [19] for large-scale learning and inference. The transductive generative formalism allows us to provide i) predictive confidence of the classification and ii) performance guarantees of the inference. In a probabilistic formulation and using the framework of graphical models we consider a bounded probability measure $\mathcal{P}_{XY}$ describing the joint distribution of the given input and output label space $X \times \mathbb{Z}^K$. The generative graphical model framework [14-18] provides a unifying framework for capturing complex dependencies between random variables and allows the design of large-scale multivariate statistical models to account for uncertainty and missing data. The objective is to minimize the conditional expected error rate of a classification rule through the conditional $\mathcal{P}(y|x)$ with hypothesis $\mathcal{H}_t$ given the observed training sample and the test. The posterior provides the basis for building $\mathcal{H}_t$ to recover $f : X \rightarrow \mathbb{Z}^K$ with predictive information. Since the goal is to obtain label information only for the test set we allow the posterior distribution to depend on the test set with spatial regularization constraints to exploit the smoothness- and cluster assumption between $\mathcal{P}_X$ and $\mathcal{P}(y|x)$. Our approach has the following advantages: i) the classification result supports a reject option and confidence bounds for risk-sensitive applications, ii) is has the ability to handle class imbalance through scaled likelihoods, and iii) the conditional independence assumptions allow separate model learning and model combination and sample complexity reduction.

We evaluate our algorithm on multimodal medical imagery using ROC statistics and report preliminary results. The algorithm shows promising segmentation performance with a sensitivity and specificity of 90.37% and 99.74% respectively and compares competitively to alternative interactive segmentation schemes. On average approximately 250 labels are required denoting a percentage ratio of 0.013 % labeled to unlabeled data for a multimodal volume. The rest of the paper is organized as follows. Section 2 outlines and formalizes related work in the field of generative graphical models most closely related to our proposed algorithm. In section 3 we present our T-CMNB algorithm a generalization of the transductive naïve Bayes model with additional regularization constraints for spatial classification. Experimental results on real world data sets are given in section 4 with concluding remarks and future work in section 5.

2. RELATED WORK

Consider a dataset $\mathcal{D} = [\mathcal{D}_l, \mathcal{D}_u] \in \mathcal{P}_{XY}$, where $\mathcal{D}_l = \{(\mathbf{x}_n, y_n)\}_{n=1}^{l}$ denote the labeled training set and $\mathcal{D}_u = \{(\mathbf{x}_n, \hat{y}_n)\}_{n=l+1}^{l+u}$ the unlabeled test set with $\hat{y}_n$ unknown. The usual case is $l << u$. The covariate data has a feature dimension of $\mathbf{x}_n \in \mathbb{R}^d$ and $(y_n, \hat{y}_n) \in \mathbb{Z}^K$ is a multinomial and represents the categorical mutually exclusive class label dimension. For the case of binary classification, $y_k \in \mathbb{Z}^{K=2}$ is a Bernoulli random variable with parameter π, $p(y|\pi) = \pi^y(1-\pi)^{1-y}$. For the multi-label or n-way classification case the class label takes multinomial form $\pi_k = p(Y^k = 1|\pi)$ with components Y^k and $X \times \mathbb{Z}^{K>2}$. This data setting reminds of a *transductive* (TL) or *semi-supervised learning* (SSL) formulation. Without going into the discussion of their difference and motivation [7] we briefly outline related work in *generative SSL* [8, 9, 20] or *–transduction* [21, 22].

2.1 The Naïve Bayes Model

Here we review the naïve Bayes model [24-26] for learning and inference in discrete and continuous data domains. Let $\mathcal{P}(y, x_1, ..., x_n)$ denote the joint distribution of the input samples and the class labels. The naïve conditional independence assumption allows us to factorize the joint distribution as a product of class prior and independent conditional probability distributions $\mathcal{P}(y) \prod_{n=1}^{N} \mathcal{P}(x_n|y)$. In graphical model notation the naïve Bayes model has for each X_j node the parent node Y, where j indexes the covariate feature dimension and n the number of samples. For the discrete case we assume each X_j to be sampled from a multinomial probability model $p(x) = \prod_{m=1}^{M} \vec{\alpha}(m)^{\vec{x}(m)}$ with

$\sum_m \vec{\alpha}(m) = 1$. The class-conditional probability for each X_j for the continuous case takes the form of a Gaussian $p(x) \sim \mathcal{N}(\mu_j | \sigma_j)$ with $\mu_{jk} = E[X_j | Y = y_k]$ and $\sigma_{jk}^2 = E[(X_j - \mu_{jk})^2 | Y = y_k]$.

2.1.1 Discrete Multinomial Case

Let $X = (X_1, ..., X_{j-1}, X_j, X_{j+1}, ..., X_d)$ be the feature vector and each X_j a multinomial random variable with components X_j^m. The subscript d denotes the feature dimension. The joint distribution factorizes into a prior and likelihood term $p(x, y | \theta) = p(y | \pi) \prod_{j=1}^d p(x_j | y, \theta_j)$ with $\theta = (\pi, \theta_1, \theta_2, ..., \theta_d)$. The class-conditional density $p(x|y)$ takes the form:

$$p(x_1, x_2, ..., x_d) | Y^k = 1, \phi) = \prod_j^d \prod_m \phi_{kjm}^{x_j^m} \tag{1}$$

with $\phi_{kjm}^{x_j^m} \equiv p(x_j^m = 1 | Y^k = 1, \phi)$ being the probability that the jth feature takes on its mth value for the kth class label. Taking the log-likelihood over $\mathcal{D}_l = \{(\mathbf{x}_n, y_n)\}_{n=1}^l$ we obtain the following objective function subject to non-negative constraints, which is solvable by forming the Lagrangian and maximizing over θ:

$$\mathcal{L}(\theta | \mathcal{D}_l) = \sum_{n=1}^l \log p(y_n | \pi) + \sum_{n=1}^l \sum_{j=1}^d \log p(x_{jn} | y_n, \phi)$$

$$\text{subject to } \sum_m \phi_{kjm} = 1. \tag{2}$$

2.1.2 Continuous Gaussian Case

In the case of continuous inputs we represent the marginal distribution $\mathcal{P}_X$ for each X_j as a Gaussian. The class conditional densities for x_j are:

$$p(x_j | Y = k, \theta_j) = \frac{1}{(2\pi\sigma_j^2)^{\frac{1}{2}}} exp\left\{ -\frac{1}{2\sigma_j^2}(x_j - \mu_{kj})^2 \right\}, \tag{3}$$

with $\mu_k \equiv (\mu_{k1}, \mu_{k2}, ..., \mu_{kd})^T$ and $\sum \equiv diag(\sigma_1^2, \sigma_2^2, ..., \sigma_d^2)$. The joint probability with conditional independent covariates and Gaussian class-conditional likelihood factorizes into:

$$p(x, y | \theta) = p(y | \pi) \prod_{j=1}^d p(x_j | y, \theta_j),$$

$$\text{with } \theta = (\pi, \theta_1, \theta_2, ..., \theta_d). \tag{4}$$

Similar to section 2.1.1 the maximum log-likelihood over the input data $\mathcal{D}_l$ writes as follows and can be computed as in the discrete multinomial case:

$$\mathcal{L}(\theta | \mathcal{D}_l) = \log \left\{ \prod_{n=1}^l p(y_n | \pi) \prod_{j=1}^d p(x_{j,n} | y_n, \theta_j) \right\}$$

$$= \sum_{n=1}^l \log p(y_n | \pi) + \sum_{n=1}^l \sum_{j=1}^d \log p(x_{j,n} | y_n, \theta_j). \tag{5}$$

2.2 Transductive Multinomial Naïve Bayes (TMNB)

The transductive naïve Bayes classifier [1] was introduced for the application of text classification. The classifier uses both the training documents $\mathcal{D}_l$ and the distribution of the test documents $\mathcal{D}_u$ to learn a classification rule. The model is similar to the one outlined in section 2.1.1 with the extension to perform transductive inference. TMNB motivates itself from the fact that the training and test samples are drawn i.i.d. according to $\mathcal{P}_{XY}$. The algorithm classifies the test documents using a multinomial naïve Bayes model initially learned from the labeled training documents (Step-I) and then sequentially relearned on the classified unlabeled test documents (Step-II) to perform transduction. We summarize their model as follows:

$$\textbf{Step-I} \quad \mathcal{L}(\hat{\theta}|\mathcal{D}_l) = \sum_{n=1}^{l} \log p(y_n|\pi) + \sum_{n=1}^{l}\sum_{j=1}^{d} \log p(x_{j,n}|y_n,\phi)$$

$$\textbf{Step-II} \quad \mathcal{L}(\theta|\mathcal{D}_{l+u}) = \sum_{n=l+1}^{l+u} \log \hat{p}(y_n|\hat{\pi}) + \sum_{n=l+1}^{l+u}\sum_{j=1}^{d} \log \hat{p}(x_{j,n}|y_n,\hat{\phi}) \tag{6}$$

$$\text{subject to } \sum_{m} \hat{\phi}_{kjm} = 1.$$

This two-step iterative scheme is exploited to estimate the prior and the class-conditional probability of the naïve Bayes model taking into account the unlabeled test distribution. Here $\hat{\theta}$ denotes the ML estimates obtained from the labeled training set. They propagate into Step-II indicating that they have been relearned from the maximum a posteriori (MAP) classification on the unlabeled test set. As reviewed in section 2.1.1 the TMNB model assumes a multinomial probability model on the data when computing the class-conditional likelihood making it non-applicable to multimodal continuous data domains such as in multimodal medical imagery.

As opposed to the TMNB model [1] we allow $\mathcal{P}_X$ to be continuous and non-uniformly distributed with a multimodal cluster and smoothness assumption. In real world applications often times the single Gaussian assumption in section 2.1.2 is to limited to fully explain the complexity of $\mathcal{P}_X$. In non-negative data domains the uniform Gaussian assumption may produce incorrect model behavior due to variance symmetry or insufficient descriptive power. Previously outlined multinomial probability models in section 2.1.1 often used in text classification [27] or their multinomial mixture counterpart [20] assume discrete finite unordered data domains with a fixed set of values. Exponential family models such as Gaussian or multinomial class-conditional mixture models may be restrictive dependent on the modeling problem and application domain.

3. TRANSDUCTIVE CONDITIONAL MIXTURE NAÏVE BAYES (T-CMNB)

We propose a novel transductive learning machine for spatial multimodal generative classification casted as an interactive segmentation problem with minimal expert intervention. In particular we present a conditional mixture naïve Bayes model (T-CMNB) with spatial regularization constraints in a transductive learning and inference setting. Compared to [1] and [20] our model assumes for the class-conditional likelihood a semi non-parametric Gaussian mixture model on each covariate feature dimension allowing us to represent and describe more complex distributions. To simplify the estimation we reduce the parameter space by assuming naïve conditional independence between the feature space and the class label. The naïve conditional independence assumption allows efficient inference of marginal and conditional distributions [19] suitable for large scale learning and inference. The posterior is formed by learning class-conditional mixture models $p(x|y)$ and priors $p(y)$ for each class in each covariate feature dimension exploiting labeled and unlabeled data. Another extension is that we allow the posterior distribution to depend on the unlabeled test set $\mathcal{D}_u$ with spatial regularization constraints to exploit the smoothness- and cluster assumption between $\mathcal{P}_X$ and $\mathcal{P}(y|x)$. As mentioned earlier, transductive learning and inference assume and exploit a cluster assumption, where each cluster reflects different distributions of different species or classes.

3.1 Conditional Multi Latent Variable Model

Our modeling problem consists of two latent variables one for $\mathcal{P}(y|x)$ and the other for approximating the marginal $\mathcal{P}_X$. To account for multimodal densities for each X_j we can consider a sub probability model $f_c(x_j|\theta_c)$ for each component $c \in C$. One can build an unconditional *mixture density* on $\mathcal{P}_X$ with

$$p(x_j|\theta) = \sum_{c=1}^{C} p(x_j, z^c = 1|\theta) = \sum_{c=1}^{C} \alpha_c f_c(x_j|\theta_c),$$

$$\text{subject to } \sum_{c=1}^{C} \alpha_c = 1, \alpha \geq 0, \tag{7}$$

where $f_c(x|\theta_c)$ are the *mixture components* obtained by marginalizing and conditioning over a latent or hidden variable Z. The non-negativity constraints α_c are the *mixing proportions* and $\theta \equiv (\alpha_c, \theta_c)_{c=1}^{C}$ denotes the parameter space. In generative graphical models the latent variable forms the parent over the data leading us to the problem of density estimation. Rather than estimating an unknown density in our case we are interested in inferring class labels with observed latent variables using a conditional mixture model on the data. Using Bayes rule one can achieve this task by inverting the *mixture density* model to perform probabilistic inference. Conditioning on $\mathcal{P}_X$ the conditional of the latent variable Y is $p(y^k = 1|x, \theta) = \alpha_{k,c} f_{k,c}(x|\theta_{k,c}) / \sum_j \alpha_j f_j(x|\theta_j)$. The knowledge of x and θ enables us to obtain the probability of the model given the data. From this probability we can classify and predict the class distribution to perform segmentation. The following section outlines how transduction directly fits the problem of interactive segmentation using minimal expert intervention. In high dimensional multimodal data domains equation (7) forms into a *multivariate mixture density* problem of the form $p(x, y) \to p(y|x) \to \hat{y} = \arg\max_m p(y = k|x)$.

3.2 Transductive Learning and Inference

Given $\mathcal{D}_l = \{(x_n, y_n)\}_{n=1}^{l}$ we learn the class-conditional and unconditional mixture densities of each class by maximizing the log-likelihood of $p(x|y)$ and $p(y)$. To learn the marginal $\mathcal{P}_X$ for a given class label we assume $p(x)$ to be distributed as a Gaussian mixture $p(x;\theta) = \sum_{k=1}^{C} \alpha_c \mathcal{N}(x; \mu_c, \Sigma_c)$ for each X_j. To approximate both latent variables we build the following log-likelihood model on the labeled training and unlabeled test set.

$$\textbf{Step-I} \qquad \mathcal{L}(\hat{\theta}|\mathcal{D}_l) = \sum_{n=1}^{l} \log p(y_n|\pi) + \sum_{n=1}^{l} \log \sum_{j=1}^{d} \alpha_j p(x_{jn}|y_n, \phi)$$

$$\textbf{Step-II} \qquad \mathcal{L}(\theta|\mathcal{D}_{l+u}) = \sum_{n=l+1}^{l+u} \log \hat{p}(y_n|\hat{\pi}) + \sum_{n=l+1}^{l+u} \log \sum_{j=1}^{d} \alpha_j \hat{p}(x_{jn}|y_n, \hat{\phi}). \tag{8}$$

As in section 2.2, $\hat{\theta}$ denotes the ML estimates obtained from the labeled training set. They propagate into Step-II indicating that they have been relearned from the maximum a posteriori (MAP) classification on the unlabeled test set. The maximum likelihood estimate of parameter θ for $\mathcal{D} = [\mathcal{D}_l, \mathcal{D}_u] \in \mathcal{P}_{XY}$ with $l + u$ i.i.d. observations has no closed form solution. The second term of above log-likelihood in equation (8) is a marginal probability and requires a non-linear optimization scheme. Alternatively equation (8) can be optimized by an iterative method to obtain the maximum likelihood solution. One can choose from belief propagation [29] and other approximate inference algorithms in probabilistic graphical models. We choose the EM algorithm [28] for the sake of simplicity and conceptual clarity. To simplify the estimation we reduce the parameter space by assuming naïve conditional independence between the feature space and the class label. Maximizing the objective function computes with the following iterative update equations:

$$\mu_i^{t+1} = \frac{\sum_n \mathcal{P}_n^{i(t)} x_n}{\sum_n \mathcal{P}_n^{i(t)}}$$

$$\Sigma_i^{t+1} = \frac{\sum_n \mathcal{P}_n^{i(t)} (x_n - \mu_i^{t+1})(x_n - \mu_i^{t+1})^T}{\sum_n \mathcal{P}_n^{i(t)}}$$

$$\alpha_i^{t+1} = \sum_n \mathcal{P}_n^{i(t)}/N \tag{9}$$

$$\mathcal{P}_n^{i(t)} = \frac{\alpha_i^{(t)} \mathcal{N}(x_n|\mu_i^{(t)}, \Sigma_i^{(t)})}{\Sigma_j \alpha_j \mathcal{N}(x_n|\mu_j^{(t)}, \Sigma_j^{(t)})}.$$

The last term of equation (9) calculates the posterior probability (E-step) whereas the preceding steps in (9) are the (M-step) equations. A proof that the update equations in (9) indeed maximize the log-likelihood can be found in [28]. The ML estimate of the first term of equation (8) is much simpler. Maximizing the log-likelihood with respect to π the solution to the constraint optimization problem for the labeled training data is:

$$\hat{\pi}_{ML} = \arg\max_\pi \sum_{n=1}^{l} \log p(y_n|\pi)$$

$$= \arg\max_\pi \sum_{n=1}^{l} \left\{ y_n \log \pi + (1 - y_n) \log(1 - \pi) \right\} = \sum_{n=1}^{l} y_n/l \tag{10}$$

Analog to equation (8) the ML estimate for π on the unlabeled test set updates accordingly with changed summation indices. From (9) and (8) the maximum a posteriori classification on the unlabeled test set can be obtained in a straigt forward manner.

4. EXPERIMENTS

We apply our algorithm to the task of interactive brain tumor (edema) concept segmentation and evaluate our method with quantitative comparison to expert grading. We performed experimental evaluation on a real world multimodal MR medical brain dataset $\mathcal{D} = [\mathcal{D}_l, \mathcal{D}_u] \in \mathcal{P}_{XY}$ with $X \in \mathcal{R}^{D=9}$ and $Y \in \mathbb{Z}^{K=2}$. The dataset has a resolution of 256x256x30 per modality and anisotropic voxel dimensions of 0.4mm x 0.4mm x 5mm. Columbia University Medical Hospital provided the dataset after de-identification of patient information. Intensity normalization to the range [0 255] and multimodal registration were applied to bring the multimodal data sources into a common coordinate frame (see Fig. 3). To perform registration we used the software platform MedINRIA [30]. Landmark-based and affine registration was used to align the multimodal data sources. Accurate registration is a key prerequisite especially in high anisotropic medical imagery and to achieve accurate multimodal feature extraction. We start with the most simplistic configuration of multimodal features by looking at FLAIR and DWI voxel intensities. We assume a bag-of-voxel representation of our multimodal features analogous to the bag-of-words representation often used in NLP. The expert provides approximate class label information on the FLAIR volume using an interactive labeling environment (see Fig. 1). The interaction step continues until the desired segmentation performance is achieved. On average approximately 250 labels are required denoting a percentage ratio of 0.013 % labeled to unlabeled data for a multimodal volume. The performance of our segmentation technique is evaluated with expert grading using voxel-by-voxel based ROC statistics. We further compare our T-CMNB algorithm with the 3D interactive watershed algorithm (see Fig. 2).

4.1 Performance Results

Table 1: ROC Comparison between single modality 3D interactive watershed algorithm (IWT) on FLAIR data (left) and our algorithm (T-CMNB) on FLAIR and DWI (right).

	Brain Tumor Concept (=edema) Segmentation			
#	3D IWT (TP)	3D IWT (FP)	T-CMNB (TP)	T-CMNB (FP)
1	59.55%	0.02%	89.72%	0.08%
2	52.25%	0.01%	92.49%	0.07%
3	62.92%	0.02%	91.07%	0.61%
4	65.46%	0.02%	96.55%	0.33%
5	73.95%	0.03%	93.23%	0.47%
6	57.02%	0.01%	79.31%	0.03%
7	70.32%	0.02%	90.09%	0.15%
8	79.37%	0.02%	92.73%	0.11%
9	77.09%	0.03%	97.11%	0.15%
10	71.31%	0.03%	95.69%	0.13%
Average	**66.93%**	**0.02%**	**91.80%**	**0.21%**

The algorithm shows promising segmentation performance with a sensitivity and specificity of 90.37% and 99.74%. The technique is computationally efficient and takes about 1-3 seconds on 256x256x30x2 multimodal datasets using a dual core 2.4GHz machine showing that the algorithm is applicable to be used within an interactive environment.

4.2 Qualitative Performance Results

Figure 1 shows the input labels on a single slice with the corresponding predicted labels for edema. In figure 2 we show a comparison of a 3D anisotropic marker-based watershed method and our approach. Both methods show high specificity and similar accuracy performance yet by using multimodal information we achieved a higher sensitivity rate in contrast to single modality information. The middle and bottom row show a comparison of the single Gaussian case (middle) and the Gaussian mixture case (bottom).
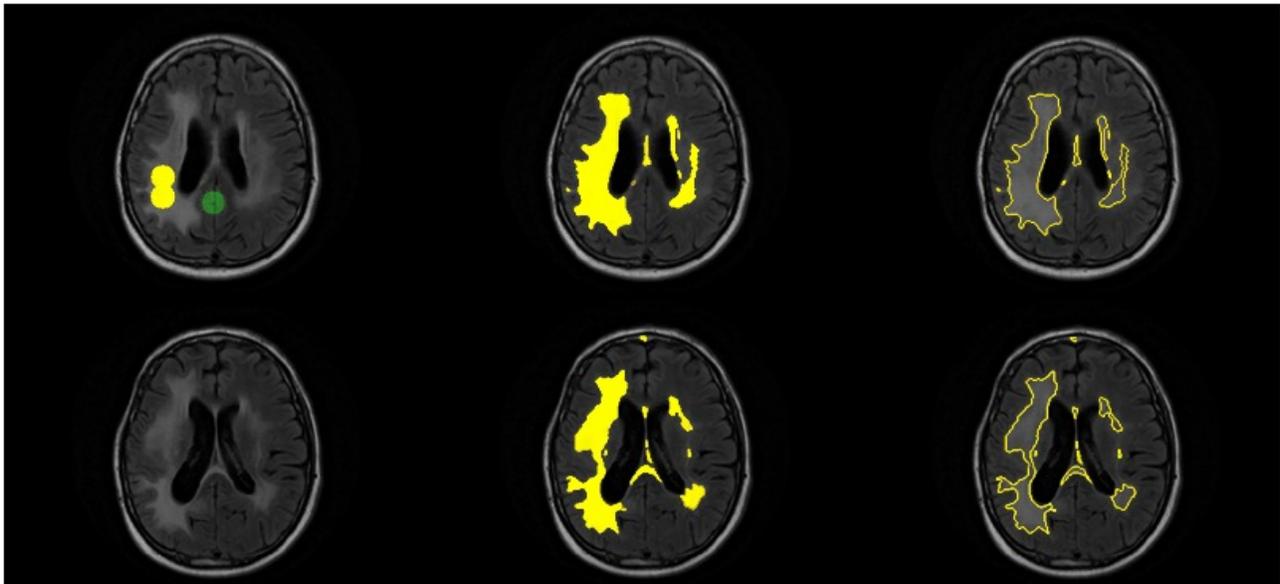
Figure 1: Example images of our segmentation technique. (Top Left) Expert drawn labels on single slice only. (Bottom Left) No labels provided for other slices in the volume. (Center) The classification result overlaid in yellow. (Right) The classification outline overlaid in yellow on the FLAIR image data.
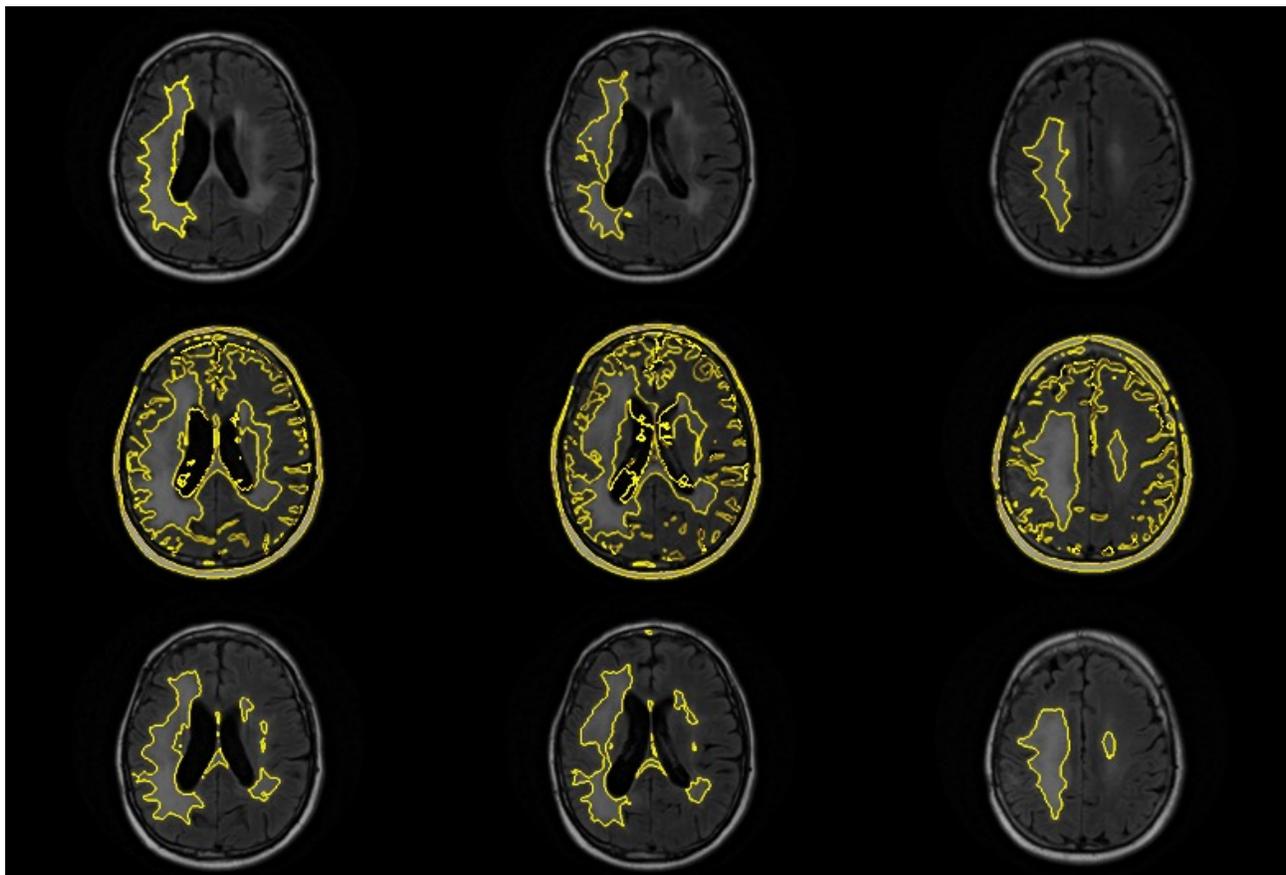
Figure 2: A qualitative comparison example of 3D interactive watershed method and our T-CMNB algorithm on the FLAIR modality. (Top) 3D interactive watershed. (Middle) Transductive naïve Bayes – Single Gaussian Case. (Bottom) T-CMNB – Gaussian Mixture Case. Note that the single Gaussian case is not able to explain the distribution whereas the Gaussian mixture case shows good segmentation performance delineating all edema regions in the volume.
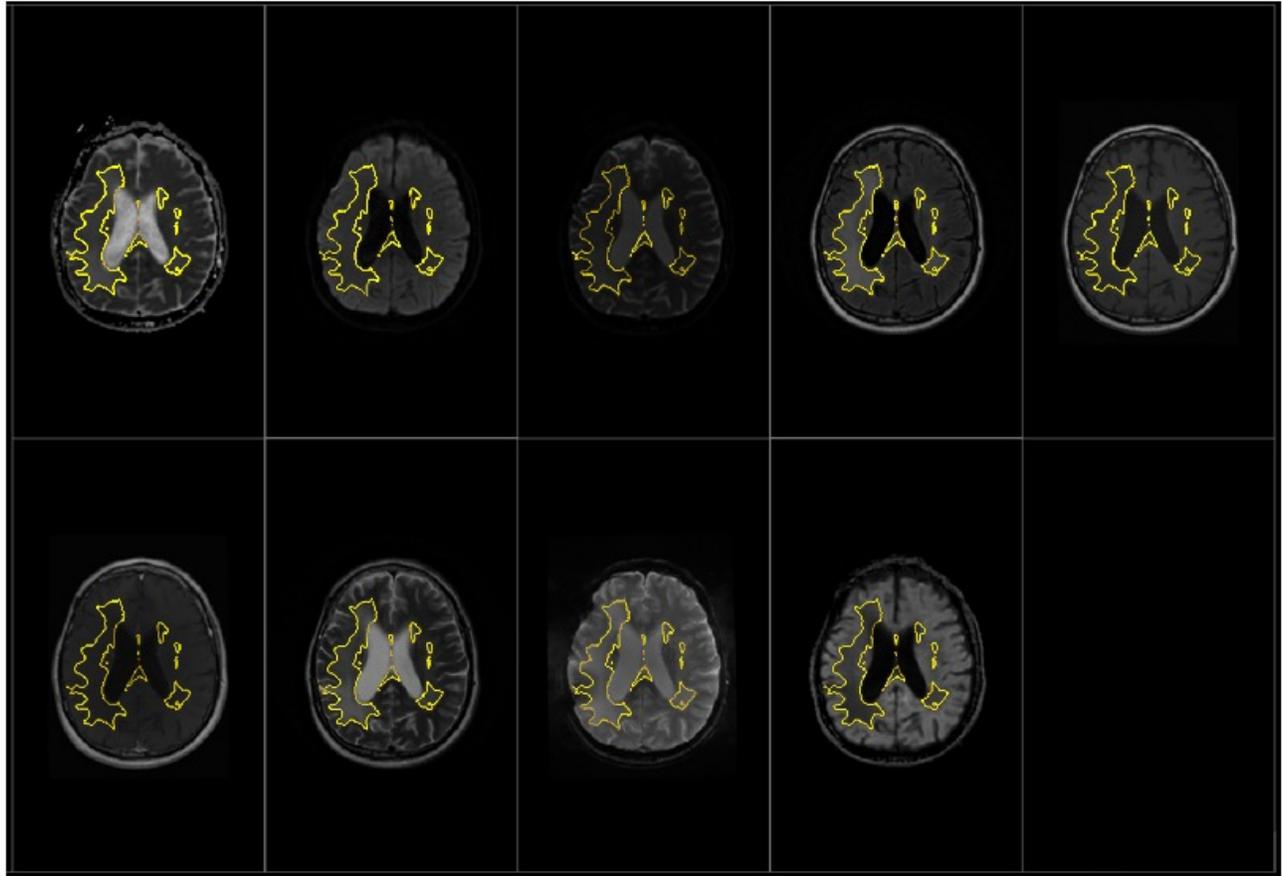
Figure 3: Example images of multimodal data sources aligned to a common coordinate frame. The segmentation result is overlaid in yellow on all nine image modalities. The DWI and FLAIR modality are shown in the 3rd and 4th window from top left to right bottom ordering.

5. CONCLUSION

In this paper we have presented a novel algorithm for spatial multimodal interactive segmentation based on a Bayesian generative transductive conditional mixture model. By introducing a majority weighted smoothness prior and exploiting unlabeled data we can achieve good classification performance for the task of interactive brain tumor (edema) concept segmentation in anisotropic multimodal medical imagery. By using a Bayesian transductive learning scheme we can link the expert knowledge with the probabilistic model to adapt to changing model assumptions over time. This is especially desirable in unpredictable tumor growth. Future work is devoted to quantitative evaluation on more test cases and experimental evaluation to other concepts of interest.

REFERENCES

[1] Branson K., "A Naïve Bayes Classifier Using Transductive Inference for Text Classification", Technical Report, (2001).
[2] Niyogi P., Girosi F., "On the relationship between generalization error, hypothesis complexity and sample complexity for radial basis functions", Neural Computation, 8(4), pp. 819-842 (1996).
[3] Russel B., [The problems of philosophy], (1912).
[4] Vapnik V. N., [Statistical learning theory], New York, Wiley, pp.339-371 (1998).
[5] Blum A., Tom Mitchell, "Combining labeled and unlabeled data with co-training", Proceedings of the Workshop on Computational Learning Theory (COLT), Morgan Kaufmann Publishers, pp.92-100 (1998).

[6]   Yu S., Balaji Krishnapuram and Romer Rosales et al, "Bayesian Co-Training", Advances in Neural Information Processing Systems 20 (NIPS), pp.1665-1672 (2008).
[7]   Chapelle O., Schölkopf B. and Zien A., [Semi-Supervised Learning], MIT Press, Cambridge, MA, (2006).
[8]   Zhu X., "Semi-Supervised Learning Literature Survey", Computer Sciences Technical Report 1530, University of Wisconsin Madison, (2007).
[9]   Cozman F., Cohen I., and Cirelo M., "Semi-Supervised Learning of Mixture Models", Proceedings of the 20th International Conference on Machine Learning (ICML), (2003).
[10] Joachims T., "Transductive Inference for Text Classification using Support Vector Machines", Proceedings of the 16th International Conference on Machine Learning (ICML), pp.200-209 (1999).
[11] Belkin M., Niyogi P., Sindhwani V. and Bartlett P., "Manifold Regularization: A geometric framework for learning from examples", Journal of Machine Learning Research, vol. 7, pp. 2399-2434 (2006).
[12] Blum A., Chawla S., "Learning from Labeled and Unlabeled Data using Graph Mincuts", ICML, pp. 19-25 (2001).
[13] Szummer M. and Jaakkola T., "Partially Labeled Classification with Markov Random Walks", NIPS, pp. 945-952 (2001).
[14] Jensen F., [An Introduction to Bayesian Networks], New York: Springer Verlag, ISBN 0387915028, (1996).
[15] Pearl J., [Probabilistic Reasoning in Intelligent Systems], San Mateo, CA, Morgan Kaufmann, (1988).
[16] Fan X., Fan G., "Generative Graphical Models for Maneuvering Object Tracking and Dynamics Analysis", Computer Visoin and Pattern Recognition (CVPR), pp. 1-8 (2007).
[17] Lauritzen S., [Graphical Models], Oxford University Press, ISBN 0198522193, (1996).
[18] Jordan M. I., [Learning in Graphical Models], MIT Press, (1998).
[19] Lowd D., Domingos P., "Naïve Bayes Models for Probability Estimation", ICML, pp, 529-536 (2005).
[20] Nigam K., McCallum A., Thrun S., and Mitchell T., "Text Classification from Labeled and Unlabeled Documents using EM", Machine Learning, vol. 39, pp. 103-134 (2000).
[21] Zhang Z. et al., "Bayesian Inference for Transductive Learning of Kernel Matrix Using the Tanner-Wong Data Augmentation Algorithm", In Proceedings of the 21st International Conference on Machine Learning, pp.935-942 (2004).
[22] Graepel T., Herbrich R. and Obermayer K., "Bayesian Transduction", Advances in Neural Informaiton Processing Systems 12 (NIPS), pp.456-462 (2000).
[23] Gommerman A., Vovk V., Vapnik V., "Learning by Transduction", In Uncertainty in Artificial Intelligence, pp.148-155 (1998).
[24] Bennett P.N., "Assessing the calibration of Naïve Bayes posterior estimates", In Technical Report No. CMUCS00-15, (2000).
[25] Domingos P. and Pazzani M., "Beyond independence: Conditions for the optimality of the simple Bayesian classifier", Machine Learning, vol. 29, pp. 103-130 (1997).
[26] Hand D. J. and Yu Y., "Idiots Bayes - not so stupid after all?", International Statistical Review, vol. 69, pp. 385-389 (2001).
[27] McCallum A. and Nigam K., "A Comparison of Event Models for Naive Bayes Text Classification", In AAAI/ICML-98 Workshop on Learning for Text Categorization, Technical Report WS-98-05, AAAI Press, pp. 41-48 (1998).
[28] Dempster A., Rubin N., "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, Series B, vol. 39, no. 1, pp. 1-38 (1977).
[29] Yedidia J.S., Freeman W.T., Weiss Y., [Understanding Belief Propagation and Its Generalizations], Exploring Artificial Intelligence in the New Millennium, ISBN 1558608117, Chap. 8, pp. 239-236 (2003).
[30] INRIA, "MedINRIA", http://www-sop.inria.fr/asclepios/software/MedINRIA/, (2008).
[31] Lee N., Smith R.T., Laine A., "Interactive Segmentation for Geographic Atrophy in Retinal Fundus Images", IEEE Signal Processing Society, Asilomar Conference on Signals, Systems and Comptuers, (2008).
[32] Ebadollahi S., Cooper J., Kaufman D., Levas A., Laine A., DeLaPaz R., Neti C., "Concept-Oriented Access to Longitudinal Multimedia Medical Records: A Case Study in Brain Tumor Management", Workshop on Cross-Media Information Analysis, Extraction and Management, 3rd International Conference on Semantics and Digital Media Technologies, pp. 51-58 (2008).