Leaf Diseases Images Classification with Online EM Semi-Supervised Learning

Zebiao Guo(2019180020)
Yao Xiao(2019180015)
Ruochen Liu(2019180018)
Jun Wu(2019180006)

Central China Normal University Wollongong Joint Institute

1 Introduction

Machine learning is the science of getting computers to realize a task without being explicitly programmed. Classical algorithms are given exact and complete rules to complete a task. Machine learning algorithms are given general guidelines that define the model, along with data. This data should contain the missing information necessary for the model to complete the task. So, a machine learning algorithm can accomplish its task when the model has been adjusted with respect to the data. The algorithms are classified by the level of manual intervention, in specific, it can be classified as supervised learning which means data are labelled manually and unsupervised learning which means data are not labelled, and the model should dig out the potential patterns. Between supervised learning and unsupervised learning is semi-supervised learning.

Supervised machine learning algorithms are designed to learn by examples. When training a supervised learning algorithm, the training data will consist of inputs paired with the correct outputs. During training, the algorithm will search for patterns in the data that correlate with the desired outputs. After training, a supervised learning algorithm will take in new unseen inputs and will determine which label the new inputs will be classified as based on prior training data. The objective of a supervised learning model is to predict the correct label for newly presented input data.

Different from the supervised learning, unsupervised learning is a type of self-organised learning that helps find previously unknown patterns in data set without pre-existing labels. In unsupervised learning, data labels are not available, it is expected that model to figure out the pattern in data set.

Semi-supervised learning is an approach to machine learning that combines a small amount of labelled data with a large amount of unlabelled data during training. Semi-supervised learning falls between unsupervised learning (with no labelled training data) and supervised learning (with only labelled training data). In specific, semi-supervised learning has below different forms:

1.1 Generative models

A Generative Model is a powerful way of learning any kind of data distribution using unsupervised learning and it has achieved tremendous success in just few years. All types of generative models aim at learning the true data distribution of the training set so as to generate new data points with some variations. But it is not always possible to learn the exact distribution of our data either implicitly or explicitly and so we try to model a distribution which is as similar as possible to the true data distribution. For this, we can leverage the power of neural networks to learn a function which can approximate the model distribution to the true distribution.

Generative approaches to statistical learning first seek to estimate $p(x|y)$, the distribution of data points belonging to each class. The probability $p(x|y)p(y)$ that a given point x has label y is then proportional to $p(x|y)p(y)$ by Bayes' rule. Semi-supervised learning with generative models

can be viewed either as an extension of supervised learning (classification plus information about $p(x)$) or as an extension of unsupervised learning (clustering plus some labels).

We are faced with data that appear as pairs $(X, Y) = (x_1, y_1), ..., (x_N, y_N)$, with the i-th observation $x_i \in R$ and the corresponding class label $y_i \in 1, ..., L$. Observations will have corresponding latent variables, which we denote by z_i. We will omit the index i whenever it is clear that we are referring to terms associated with a single data point. In semi-supervised classification,only a subset of the observations have corresponding class labels; we refer to the empirical distribution over the labelled and unlabelled subsets as $p_l(x, y)$ and $p_u(x, y)$, respectively. We now develop models for semi-supervised learning that exploit generative descriptions of the data to improve up on the classification performance that would be obtained using the labelled data alone.

A commonly used approach is to construct a modelthat provides an embedding or feature representation of the data. Using these features, a separateclassifier is thereafter trained. The embeddings allow for a clustering of related observations in alatent feature space that allows for accurate classification, even with a limited number of labels.Instead of a linear embedding, or features obtained from a regular auto-encoder, we construct adeep generative model of the data that is able to provide a more robust set of latent features. Thegenerative model we use is: $p(z) = N(z|0, I); p_\theta(x|z) = f(x; z, \theta)$, where $f(x; z, \theta)$ is a suitable likelihood function (e.g., a Gaussian or Bernoulli distribution) whoseprobabilities are formed by a non-linear transformation, with parameters θ, of a set of latent variables z. This non-linear transformation is essential to allow for higher moments of the data to be captured by the density model, and we choose these non-linear functions to be deep neural networks.

1.2 Graph-based methods

Graph-based methods for semi-supervised learning use a graph representation of the data, with a node for each labelled and unlabelled example. The graph may be constructed using domain knowledge or similarity of examples. In the graph-based methods, label information of each sample is propagated to its neighboring sample until a global stable state is reached on the complete dataset. Here, a graph is constructed with nodes and edges, where the nodes are specified by unlabelled and labeled samples, whereas the edges specify the similarities among the labeled as well as unlabelled samples. Here the label of each data sample is progressed to its neighboring points. The graph structure is represented as G = (V, E), where a set of vertices V which signify both the labeled and unlabeled data samples and a set of edges E which denote the similarities among the labeled as well as unlabelled samples from the dataset (Fig. 1). In Fig. 2(a) two shaded circles are the initially labeled vertices (± 1), while the white nodes represent unlabelled samples. The similarity between samples is represented by the thickness of the edges. Fig. 2(b) shows that the graph technique classifies the unlabelled samples according to the weighted distance. The two clusters are formed, even though samples are connected by thin edges.

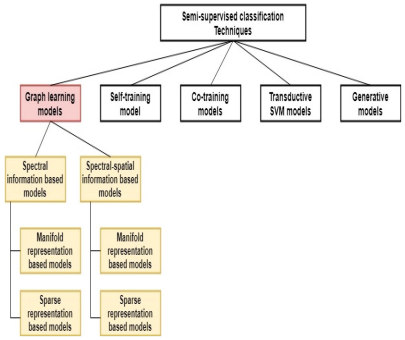

Fig. 1: Hierarchical structure of semi-supervised classification techniques.

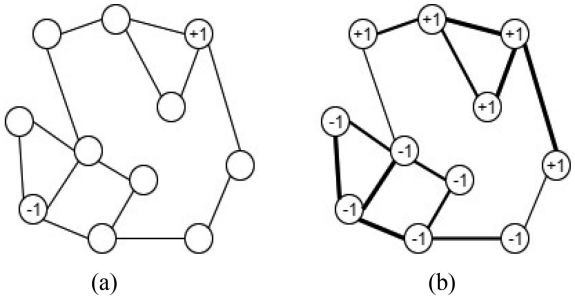

Fig. 2: Graph-based classification of dataset: (a) Before classification (b) After classification.

2 Background

2.1 History and Definition

In fact, Merz et al. proposed the term semi-supervised learning in 1992 and used semi-supervised learning for classification problems for the first time. Then Shahshahani and Landgrebe started to study SSL. The collaborative training method proposed by Blum and Mitchell trains two different learning machines based on different views, which improves the confidence of training samples. Vapnik and Sterin proposed TSVM (Transductive Support Vector Machine), which is used to estimate the linear prediction function of class labels. In order to solve TSVM, Joachims proposed the SVM method. Bie and Cristianini relaxed TSVM as a semi-definite programming problem to solve.

Semi-supervised learning is a special classification method. Traditional classifiers require labeled data for training. However, obtaining labeled instances is often difficult, and unlabeled data may be relatively easy to collect, But semi-supervised learning solves this problem by using a large amount of unlabeled data and labeled data to build a better classifier [9]. Compared with supervised and unsupervised learning, the method based on semi-supervised learning is more preferable, because in the presence of a large amount of data, the semi-supervised method can improve performance. Semi-supervised learning requires less manual work and provides higher accuracy.

To put it simply, let the learner not rely on external interaction and automatically use unlabeled samples to improve learning performance, which is semi-supervised learning.

Semi-supervised learning can be further divided into pure semi-supervised learning and transductive learning. The former assumes that the unlabeled samples in the training data are not the data to be predicted, while the latter assumes that the unlabeled samples considered in the learning process are the data to be predicted, the structure is as follows:

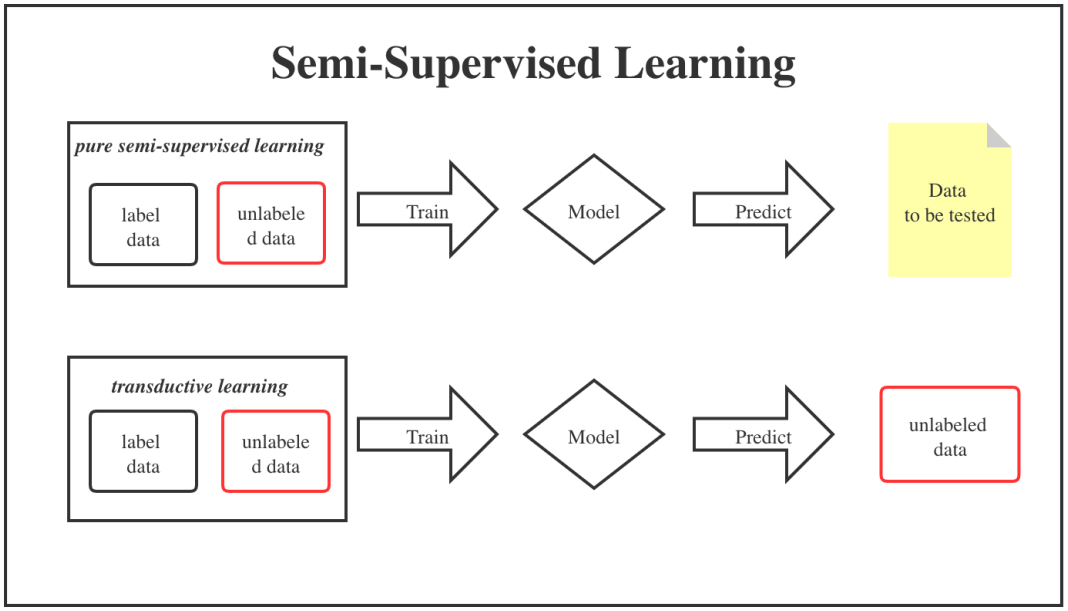

Fig. 3: Semi-Supervised Learning

2.2 Semi-Supervised Learning

Self-Training Self-training is the most common method of semi-supervised learning.In this method, the labeled data samples are first used to train the classifier. Then use the classifier to classify the unlabeled data set. Generally, the most guaranteed unlabeled data and its predicted labels are attached to the training set. Retrain the classifier with new data and repeat the process. The process of retraining the classifier can also be called selfteaching or bootstrapping.

A method of semi-supervised training target detection system [2] based on self-training is discussed. The self-training mechanism used is divided into five steps. (1) Train the detector by using a limited set of fully labeled positive samples and a complete set of negative samples. (2) Run the detector on the part of the data set with weak labels and use the maximum likelihood ratio to find the scale and position. (3) The output of the detector is used to mark the unlabeled data training samples and assign a selection score to each detection. (4) Use the selection index to select a subset of newly marked data. (5) Repeat the above steps until all the data to be trained is added.

Co-Training Co-Training is a semi-supervised learning method with two different classifiers. These classifiers need to have two sets of different functions on the labeled data. Each classifier undergoes a set of functional training and is used to classify unlabeled data. Each classifier's most confident prediction of unlabeled data will be iterated by other classifiers into labeled training data.

Here is a semi-supervised regression algorithm of co-training style [3]. The algorithm used for regressor in kNN search. By using the amount of reduction in the mean square error on the marked area of the sample, the marking reliability of the unmarked sample can be estimated. In the semi-supervised learning scheme, the regressor is optimized for all iterations:

$$Minkowsky_p(x_r, x_s) = \left(\sum_{l=1}^{d} |x_{r,l} - x_{s,l}|^p \right)^{1/p} \quad (1)$$

Transductive Support Vector Machines The goal of TSVM is to use both labeled and unlabeled data to find the linear boundary with the largest margin. Figure 2 shows the decision boundaries found by SVM on labeled data and TSVM on labeled and unlabeled data.

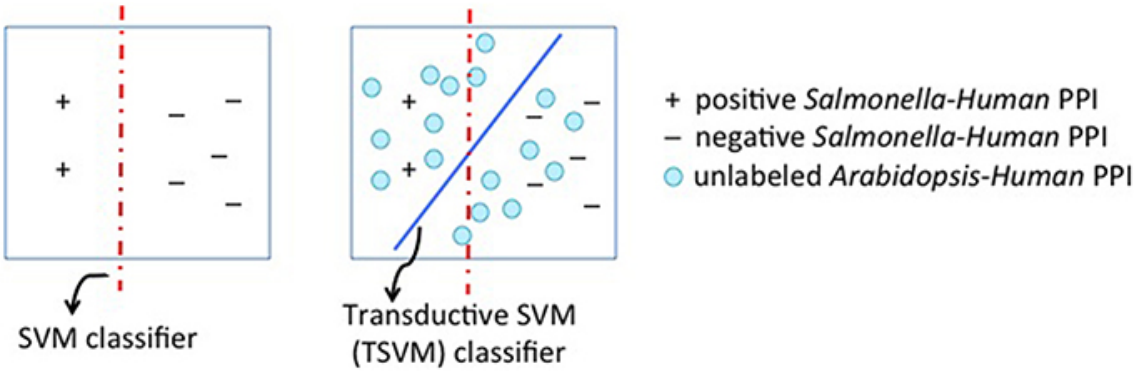

Fig. 4: Transductive Support Vector Machines

Graph-Based Models In recent years, the most active research area in semi-supervised learning is the graph-based method, which first constructs graphs from training samples, where nodes are represented in the dataset as labeled and unlabeled samples, and edge description samples Sex Similarity. These methods usually assume the smoothness of the labels on the graph. The graph-based method does not require any parameters. These methods are different in nature and are also transducible.

Here is a general framework for semi-supervised learning on a directed graph [4], the structure of the graph along with the direction of the edges is considered. The algorithm takes the input as the directed graph and the label set. The unlabeled instances are classified using the steps. (1) A random walk over the graph with a transition probability matrix is defined such that it has a unique stationary distribution like the teleporting random walk. (2) Calculate the matrix by using the diagonal matrix with its stationary distribution. (3) A function is computed using the labeled vertices to classify the unlabeled vertices. In the absence of labeled instances, this method can be used as a spectral clustering method for the directed graphs. This simplifies the spectral clustering approach for the undirected graphs.

The algorithm can be simply summarized as: Given a directed graph $G = (V, E)$ and a label set, the vertices in a subset $S \subset V$ are labeled.Then the remaining unlabeled vertices may be classified as follows:

1. Define a random walk over G with a transition probability matrix P such that it has a unique stationary distribution,such as the tcleporting random walk.
2. Let Π denote the diagonal matrix with its diagonal elements being the stationary distribution of the andom walk.Compute the matrix $\Theta = (\Pi^{1/2} P \Pi^{-1/2} + \Pi^{1/2} P^T \Pi^{-1/2})/2$
3. Compute the function $f = (I - \alpha\Theta)^{-1} y$, where a is a parameter in (0,1),and classify each unlabeled vertex v as sign $f(v)$.

3 Application

3.1 Lesion Image Segmentation

In this study supervised classification algorithms (including logistic regression analysis, Naive Bayes algorithm, CART and linear discriminant analysis) were used to segment the sub-images, and then their segmentation effects were evaluated. The main steps for lesion image segmentation.

Each obtained sub-image was converted from RGB color space into HSV color space and L*a*b* color space. In each pixel in the sub-image, the a* component value and the b* component value were regarded as the color features of the pixel. After all pixels in a sub-image were clustered into ten classes using a clustering algorithm, the mean of the H components of all pixels in each class was calculated. Compared with healthy crops, the H component of images of diseased crops is smaller. Consequently, the pixels in the class with the minimum mean were treated as typical lesion pixels, and the pixels in the seven classes with the largest means were treated as typical healthy pixels. There is a transition region between the lesion region with typical symptoms and the typical healthy region, and H components are usually between the two regions.

The pixels in the two remaining classes were treated as pixels that were not involved in building the pixel classification models. The typical lesion pixels and typical healthy pixels were labeled positive samples and negative samples, respectively, and these pixels constituted the training set for building pixel classification models. With a* component value and b* component value of each pixel in the training set as feature variables, pixel classification models to classify all the pixels in the sub-image were built using logistic regression analysis, Naive Bayes algorithm, CART and linear discriminant analysis,respectively. For each classification model, each pixel classified as lesion was assigned a value of 1 and each remaining pixel in the sub-image was assigned a value of 0. Thus, an initial binary lesion segmentation image was obtained.

In the process of lesion segmentation, all pixels in a sub-image were classified as either lesion pixels or healthy pixels. Therefore, lesion segmentation is similar to binary classification problem in the field of pattern recognition, and the evaluation of segmentation effects can be carried out using methods for evaluating a binary classification model. In this study, the two indices were calculated according to the following formulas: Recall = N1/N2 and Precision = N1/N3, where N1 was the total number of lesion pixels in a sub-image correctly classified by using a segmentation method integrated with a clustering algorithm and a supervised classification algorithm, N2 was the total number of lesion pixels in the sub-image classified using the manual segmentation method, and N3 was the total number of the pixels in the sub-image. Both Recall and Precision range from 0-1. Larger values of Recall and Precision indicate a better integrated segmentation method.

After segmentation, in each final binary lesion segmentation image, each independent white region (i.e., connected component) was labeled a lesion, and the black background region was labeled the healthy region. The location of the smallest rectangle containing each lesion, namely, the independent white region, was determined.

3.2 Feature Extraction and Normalization

Hu invariant moments used to depict the texture features of an image are invariant to translation, rotation and scaling. Contrast is applied to measure the gray level of a pixel in comparison with the neighbor pixels in an image, energy is a measure of the consistency of an image, and homogeneity is used to measure the spatial closeness of elements with the diagonal distribution in a co-occurrence matrix [5]. Circularity denotes the degree that a lesion region is circular, and a bigger value indicates that the lesion region is more circular [6]. Complexity refers to the complexity and discrete degree

of a lesion region, and a bigger value indicates the lesion region with higher complexity and greater discrete degree [6]. The seven Hu invariant moments were calculated using the calculation formulas as described in [7]. The other extracted features were calculated according to the formulas shown in Table 1.

Feature parameter	Calculation formula
Contrast	$\sum_{i=1}^{M}\sum_{j=1}^{M}(i-j)^2 p_{ij}$, where $M\times M$ denotes the size of a co-occurrence matrix, $M = 1, 2, \ldots$, and p_{ij} denotes the quotient of the element (i, j) of a co-occurrence matrix divided by the sum of the elements of the co-occurrence matrix.
Energy	$\sum_{i=1}^{M}\sum_{j=1}^{M}p_{ij}^2$, where $M\times M$ denotes the size of a co-occurrence matrix, $M = 1, 2, \ldots$, and p_{ij} denotes the quotient of the element (i, j) of a co-occurrence matrix divided by the sum of the elements of the co-occurrence matrix.
Homogeneity	$\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{p_{ij}}{1+\lvert i-j\rvert}$, where $M\times M$ denotes the size of a co-occurrence matrix, $M = 1, 2, \ldots$, and p_{ij} denotes the quotient of the element (i, j) of a co-occurrence matrix divided by the sum of the elements of the co-occurrence matrix.
First moment	$\frac{1}{L}\sum_{i=0}^{L-1} f_i p(f_i)$, where μ_1, μ_2 and μ_3 refer to the first moment, second moment and third moment, respectively, f_i represents a random variable of gray level, $p(f_i)$ represents the gray level histogram of an image region, $i = 0, 1, 2, \ldots, L$-1, and L is the number of different gray levels.
Second moment	$\left[\frac{1}{L}\sum_{i=0}^{L-1}(f_i-\mu_1)^2 p(f_i)\right]^{\frac{1}{2}}$, where μ_1, μ_2 and μ_3 refer to the first moment, second moment and third moment, respectively, f_i represents a random variable of gray level, $p(f_i)$ represents the gray level histogram of an image region, $i = 0, 1, 2, \ldots, L$-1, and L is the number of different gray levels.
Third moment	$\left[\frac{1}{L}\sum_{i=0}^{L-1}(f_i-\mu_1)^3 p(f_i)\right]^{\frac{1}{3}}$, where μ_1, μ_2 and μ_3 refer to the first moment, second moment and third moment, respectively, f_i represents a random variable of gray level, $p(f_i)$ represents the gray level histogram of an image region, $i = 0, 1, 2, \ldots, L$-1, and L is the number of different gray levels.
Color ratio r	$\frac{R}{R+G+B}$
Color ratio g	$\frac{G}{R+G+B}$
Color ratio b	$\frac{B}{R+G+B}$
Circularity	$\frac{4\pi S}{L^2}$, where S and L represent the area and perimeter of a lesion region, respectively.
Complexity	$\frac{L^2}{S}$, where S and L represent the area and perimeter of a lesion region, respectively.

Fig. 5: Table 1
[5–8]

Because of the great differences between the ranges of extracted features, which may impact the accuracies of disease recognition models, the values of each extracted feature were normalized to the range of 0-1 using the following formula: $X^i_{norm}=(X^i\text{-}X^i_{min})/(X^i_{max}\text{-}X^i_{min})$,where X^i_{norm} was the value of the ith feature after normalization and X^i,X^i_{min} and X^i_{max} were the value of the ith feature, the minimum value and the maximum value of the feature before normalization, respectively.

3.3 Feature Selection

To reduce the complexity of image recognition resulting from excessive features and improve the accuracy and applicability of image recognition methods, the extracted features were screened after feature normalization. Based on the training set described above, feature selection was conducted using the ReliefF method, the 1R method and the CFS method. For the ReliefF method, a high weight was assigned to a feature that has a high correlation with categories, and a feature with a higher weight indicates that this feature is more important. For the 1R method, the classification accuracy is calculated with each feature as the input of the 1R classifier successively and is used to evaluate the importance of the feature. Higher classification accuracy indicates that the corresponding feature is more important. The CFS method is unlike the ReliefF method and the 1R method, and is aimed to obtain the optimal feature subset. The correlation between the optimal feature subset and dependent variable should be as high as possible. Meanwhile, the correlations among the features in the optimal feature subset should be as small as possible.The importance ranking of each feature for classification and recognition could be obtained using the ReliefF method and

the 1R method, respectively. A higher ranking for a feature indicates that it is more likely to yield better recognition results if used to build the recognition model.According to the recognition accuracies of the training set and the testing set, the best top N features were selected as the best feature combination to build the disease recognition models. For the CFS method, the best feature combination, namely, the optimal feature subset, was obtained directly for modeling.

3.4 Building of Disease Recognition Models

After the segmentation, feature extraction, feature normalization and feature selection described above, disease recognition models were built a disease recognition models.For the supervised learning methods, the true class that each sample in the training set belongs to is known. In other words, all samples in the training set are labeled samples. In class of the training samples is very high, which requires a large amount of manpower and material resources. When a small number of samples in the training set are labeled, a recognition model can be built using a semi-supervised learning method. In practice, when many disease images are acquired with lower costs, the experts in the corresponding field just need to make artificial recognition and classification of a small number of disease images. Disease recognition models can be built using semi-supervised learning methods, which will greatlytures used to build the optimal supervised model were transformed using PCA. We use the proposed semi-supervised incremental variant of the EM algorithm to build our model for classifying the crop diseases.The algorithm train takes advantage of both labelled and unlabelled data. EM is an iterative algorithm for maximum likelihood or maximum a posterioriestimation of the incomplete data problem. Data is considered incomplete because it lacks sample class information. The algorithm first trains classifiers incrementally from scratchusing the first few labelled samples and then this trained classifier is used to classify each element of unlabelled datawith some posterior probability. This posterior value is used asa weight to update the learning parameters of every class. This iterative process is repeated for every unlabelled and labelled sample. In the above process, we use a weighting factor that dynamically controls the contribution of unlabelled data forthe parameter estimation in EM. Next, we briefly point out the motivations and contributions of the presented work.

In a typical case of semi-supervised learning, we have missing information in the data and the task is to estimate thevalue of that missing information. Usual solutions to this wouldbe to use Expectation Maximization (EM) . In the E-step, the posterior probability of data points is computed, while in M-step, parameters of the learning model are computed. Aniterative algorithm like this is not suitable in our context ofonline learning since our goal is to process training samplesone at a time and to avoid their storage. This is a real-time scenario when we need to deal with streams of data or a huge dataset (which cannot be processed fully in a single step).Hence a method that updates the model in a continual and an evolutionary manner can be used in the present study. This problem has been addressed by a few researchers. In this study we use some findings of Lee[9].We have incorporated a regulating constant λ in the proposed method. This parameter moderates the contribution made by the unlabelled data by reducing the learning rate(η) and hence the weight of the unlabelled samples during step M.The best value of λ in practice is decided using a validation set, which was found to be 0.01 in our case. λreduces th elearning rate for all unlabelled data equally but the value of η increases for the correct class sample because of the higherposterior value q_k. It is a slow hill-climbing process in which the correct class sample makes the hill-climbing step greater towards the hill top, whereas the incorrect class sample makes the step small and in the opposite direction, so effectively weget closer to the hill top after processing many unlabelled data samples.Details of the proposed incremental (online) EM algorithmfor creating an incremental semi-supervised learning systemare as shown below:

(1)Train RNB online with a few labelled training samples. Repeat the following steps for every unlabelledor labelled data sample for better estimation of learning parameters.

(2)E Step (of proposed online EM):If incoming new sample is unlabelled then use trained classifier to find the posterior $q_k = P(Y = y_k|X)$ corresponding to all k (class). Else $q_k = 1$.

(3)M Step (of proposed online EM):

$$c_k = c_k + q_k\lambda$$
$$\eta_k = q_k(\tfrac{1-\alpha}{c_k} + \alpha)\alpha$$

where $\lambda = weight$ factor applied to moderate thecontribution of unlabelled data in the parameter esti-mation step ($\lambda = 1$ if the sample is Labelled) and η_k is a learning rate parameter.

$$\mu_{ik}(t) = (1 - \eta_k)\mu_{ik}(t-1) + \eta_k x_i^j \delta(Y^j = y_k)$$
$$\sigma_{ik}^2(t) = (1 - \eta_k)\sigma_{ik}^2(t-1) + \eta_k(x_i^j - \mu_{ik}(t))^2 I$$

where $I = \delta(Y^i = y_k)$

(4)Repeat step E and M (just once) for each labelled orunlabelled sample left.

The above algorithm is capable of handling both labelled andunlabelled data at the same time to update the parametersof the learning system.

4 Conclusion

In this report, we focus on designing an appropriate semi-supervised image-based classifier for food crop diseases. In this problem, there are usually insufficient images of the various diseased plant leaves to train a classifier with sufficient accuracy for production deployment. For addressing this problem, we proposed a method which makes learning possible from a continuous inflow of a potentially unlimited amount of data. It highlights the use of unlabelled data for better parameter estimation, especially when labelled data is scarce and expensive unlike unlabelled data. We introduce an algorithm for learning from labelled and unlabelled samples based on the combination of novel online ensemble of the Randomized Naive Bayes classifiers and a novel incremental variant of the Expectation Maximization(EM) algorithm. We make use of a weighting factor to modulate the contribution of unlabelled data.

4.1 Future direction of semi-supervised learning

From our perspective, one of the most important issues to be resolved in semi-supervised learning is the potential performance degradation caused by the introduction of unlabelled data. Although this has received relatively little attention in the literature (likely due to publication bias, as noted by Zhu 2008[10]), many semi-supervised learning methods only perform better than their supervised counterparts or base learners in specific cases (Li and Zhou 2015[11]; Singh et al. 2009[12]). In other cases, the supervised baselines used for empirically evaluating the performance of semi-supervised learning methods are relatively weak, causing a skewed perspective on the benefits of incorporating unlabelled data. Moreover, the potential performance degradation is generally much more significant than the potential improvement, especially in machine learning problems where strong performance is achieved with purely supervised learning. We believe that this is one of the main reasons for the dearth of applications of semi-supervised learning methods in practice when compared to supervised learning.

A second potential remedy for the lack of robustness of semi-supervised learning methods lies in the application of automated machine learning (AutoML) to the semi-supervised setting. Recently, there has been a steep increase in interest in the automatic selection and configuration of learning algorithms for a given classification problem. These approaches include meta-learning and neural architecture search as well as automated algorithm selection and hyperparameter optimization. While AutoML techniques have been prominently and successfully applied to supervised learning.

5 Individual Effort

Yao Xiao-2019180015: 25%
Ruochen Liu-2019180018: 25%
Zebiao Guo-2019180020: 25%
Jun Wu-2019180006: 25%

References

1. Zhu, Xiaojin Jerry. Semi-supervised learning literature survey. University of Wisconsin-Madison Department of Computer Sciences, 2005.
2. C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-Supervised Self-Training of Object Detection Models," Proc. Seventh Workshop Applications of Computer Vision, vol. 1, pp. 29-36, Jan.2005.
3. Zhi-Hua Zhou, Ming Li "Semi-Supervised Regression with Co-Training Style Algorithms" IEEE Transactions On Knowledge And Data Engineering, Volume:19 , Issue: 11 Nov. 2007.
4. Zhou, Dengyong, Jiayuan Huang, and Bernhard Schölkopf. "Learning from labeled and unlabeled data on a directed graph." In Proceedings of the 22nd international conference on Machine learning, pp. 1036-1043. 2005.
5. Gonzalez RC, Woods RE. Digital image processing. 3rd ed. Upper Saddle River: Prentice Hall; 2007.
6. Li GL, Ma ZH, Wang HG. Image recognition of wheat stripe rust and wheat leaf rust based on support vector machine. J. China Agric. Univ. 2012; 17: 72–79
7. Gonzalez RC, Woods RE, Eddins SL. Digital image processing using MATLAB. Beijing: Publishing House of Electronics Industry; 2005.
8. Stricker MA, Orengo M. Similarity of color images. Proc. SPIE Int. Soc. Opt. Eng. 1995; 2420: 381–392.
9. D.S. Lee Effective Gaussian mixture learning for video background subtraction, IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 5, pp.827-832 2005
10. Zhu, X. (2008). Semi-supervised learning literature survey. Technical Report. 1530, University of Wisconsin-Madison.
11. Liu, W., He, J., Chang, S. F. (2010b). Large graph construction for scalable semi-supervised learning. In Proceedings of the 27th international conference on machine learning (pp. 679–686).
12. Singh, A., Nowak, R., Zhu, X. (2009) Unlabeled data: Now it helps, now it doesn't. In Advances in neural information processing systems (pp. 1513–1520).