Semi-supervised Learning Approaches for Predicting Lung Nodules Semantic Characteristics

Dmitry Zinovev, Daniela Raicu, Jacob Furst
Intelligent Multimedia Processing Laboratory, College of Computing and Digital Media,
DePaul University, Chicago, IL 60604, USA
dzinovev@students.depaul.edu,{dstan, jfurst}@cdm.depaul.edu

Abstract

We propose two semi-supervised learning approaches for automatically predicting semantic characteristics of lung nodules based on low-level image features. The NIH Lung Image Database Consortium (LIDC) dataset is used for training and testing of the proposed approaches such that the nodules on which at least three radiologists agree serve as labeled data and all the other nodules serve as unlabeled data. We show that, in the case of the LIDC [1] dataset, we are able to improve the accuracy prediction by 50% in average when using our proposed semi-supervised approaches versus the traditional supervised classification approaches. While this paper briefly explains our methodology and results, the extended version of this paper has been accepted as a journal publication [2] that will appear this year.

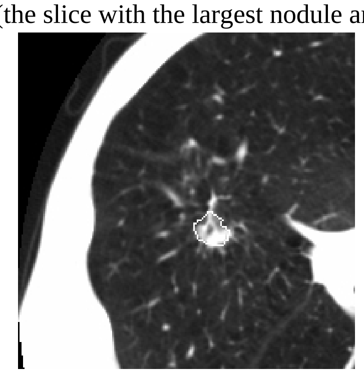
1. Introduction

Computer-aided diagnosis (CAD) systems can act as a *second reader* and assist radiologists in decision-making task to improve the efficiency of single observer and to reduce variation among multiple observers. We propose two semi-supervised approaches [3] that, besides malignancy, they also predict nodule characteristics that are perceived by radiologists when making diagnosis: subtlety, lobulation, margin, sphericity, malignancy, texture, and spiculation. The expectations are that the proposed approaches will have the capability to handle the large variability in the radiologists' interpretation and therefore, the lack of consistency in the label values associated with a nodule's characteristic.

2. Methodology

Our proposed approaches relate to active-DECORATE [4], an active learning approach extended from DECORATE [5], which iteratively constructs an ensemble of classifiers by learning a classifier at each iteration and adding it to the current ensemble. The purpose of active-DECORATE is to detect the most informative examples (i.e. those examples that cause most disagreement in the ensemble) in the instance space and ask expert to label them, then add the labeled instances into the training data and repeat these processes for a number of iterations. While in active-DECORATE the system asks for user input/annotations/labels for the most representative few cases at each iteration, in our approaches, the labels are assigned automatically either by using the ensemble classifier's labels predicted with high confidence or by using both the predicted labels with high confidence and the most representative cases along with their original labels. While both approaches are meant to improve the prediction accuracy, the second approach also ensures that even though the original labeled data might not contain certain classes, the final classification results can produce classification rules/labels for all classes, for example extremely subtle, moderately subtle, fairly subtle, moderately obvious, obvious for the subtlety semantic characteristic.

Each lesion in the LIDC database considered to be a nodule > 3 mm could have been marked by one to four radiologists. Therefore, there can be up to 4 different boundaries/images of a nodule marked by up to 4 radiologists on a slice and ratings for different characteristics will also be different. For instance, lobulation of a nodule shown on Figure 1 was rated by 4 different radiologists as 4, 1, 2 and 5. In this study we select only one slice per nodule (the slice with the largest nodule area) for each radiologist. Therefore, there can be up to 4 images per nodule.



From the current 85 cases available, 60 cases had 149 nodules greater than or equal to 3 mm in maximum diameter which generated 379 nodule images. From these nodule images, we extracted 64 image features which include four types of image content that encode shape, size, intensity, and texture information of the nodule. At the end of the image feature extraction process, each nodule image is encoded using a set of sixty-four image features $f_i, i = 1, \ldots 64$ and the seven radiologist annotations $c_j, j = 1, \ldots 7$.

Figure 1: An example of lung nodule

For our first approach, we select instances predicted with the probabilities higher than a confidence threshold of 60%, and add them along with their predicted labels into labeled training data iteratively until all unlabeled data left cannot reach the confidence threshold. Then, we add unlabeled data left along with their original labels into the labeled data.

For our second approach, besides using the confidence threshold we also add to the training labeled data set the instances with the lowest margin (the difference between the first and second predicted highest class probability) using 5% threshold per iteration; these instances represent the most informative instances and are added along with their original labels into the label data. The values of the thresholds have been determined experimentally.

3. Results

The classification accuracies of both approaches are summarized and compared against the classification accuracies for the traditional decision trees (Table 1). For all characteristics, both semi-supervised approaches improve the classification accuracies by about 50% in average over all seven characteristics and all the differences are statistically significant. Furthermore, if we compare the classification accuracies between the two semi-supervised approaches, we will find that the differences between the two approaches are not significant with the exception of the results for lobulation. However, the second approach has the advantage of ending up with more classes being predicted than the first approach.

Table 1: Accuracies for decision tress and the proposed semi-supervised approaches

Characteristics	Decision trees	Add instances predicted with high confidence (60%)	Add instances predicted with high confidence (60%) and instances with low margin (5%)
Lobulation	27.44%	81.00%	69.66%
Malignancy	42.22%	96.31%	96.31%
Margin	35.36%	98.68%	96.83%
Sphericity	36.15%	91.03%	90.24%
Spiculation	36.15%	63.06%	58.84%
Subtlety	38.79%	93.14%	92.88%
Texture	53.56%	97.10%	97.36%
Average	**38.52%**	**88.62%**	**86.02%**

4. Conclusions

Both semi-supervised learning approaches presented in the paper are especially useful when for the same nodule and semantic characteristic there are different values (class labels) given the difference in interpretation among radiologists. Using the LIDC dataset as a case study showcasing the variance in interpretation of four expert radiologists, we show that we are able re-label the data on which disagreements exist and with the new re-labeled data produce results which are as accurate as the ones on which agreement on the nodules' interpretation exists.

References

[1] S.G. Armato, G. McLennan, M.F. McNitt-Gray, C.R. Meyer, D. Yankelevitz, D.R. Aberle, C.I. Henschke, E.A. Hoffman, E.A. Kazerooni, H. MacMahon, A.P. Reeves, B.Y. Croft, and L.P. Clarke, "Lung Image Database Consortium: Developing a resource for the medical imaging research community", Radiology, 232(3), pp. 739–748, 2004.

[2] Daniela Raicu, Dmitry Zinovev, Jacob Furst, Ekarin Varutbangkul, "Semi-supervised Learning Approaches for Predicting Lung Nodules Semantic Characteristics", Advances in Medical Intelligent Decision Support Systems Intelligent Decision Technologies (IDT), Vol. 3, No. 2, 2009.

[3] O. Chapelle, B. Schölkopf and A. Zien: Semi-Supervised Learning. MIT Press, Cambridge, MA, 2006.

[4] P. Melville, R. Mooney, "Diverse ensembles for active learning", In Proc. of International Conference On Machine Learning (ICML) 2004, ACM, 584-591, 2004.

[5] P. Melville, R. Mooney, "Constructing diverse classifier ensembles using artificial training examples", Proc of 18th Intl. Joint Conf. on Artificial Intelligence, pp. 505-510, Acapulco, Mexico, 2003.