Investigation on Feature Selection to Improve Classification of Abdominal Organs in CT Images

Yu Zhang, Noriko Tomuro, Daniela Stan Raicu, Jacob Furst
School of Computer Science, Telecommunications, and Information Systems
DePaul University, Chicago, Illinois, 60604
jzhang@eastwest.edu, {tomuro, draicu, jfurst}@cs.depaul.edu

Abstract

This paper presents the preliminary result on feature selection for the purpose of classifying soft tissues of abdominal organs in computer tomography (CT) images. From the images in the dataset, texture features were first extracted, and the most relevant features were identified based on the Information Gain measure. Then a Decision Tree classifier was used to select the optimal subset of features. The initial experiments indicated that, by removing the combinations of the descriptors and distances which have the lowest Information Gain, as much as 83% of the original features were removed without sacrificing the classification accuracy at all, for the overall dataset or any individual organ, or even improving it significantly for some organs.

1 Introduction

In image processing, raw images represented by the gray levels of the pixels are usually transformed to features which can better capture the characteristics of the images in the pre-processing phase. Texture features are such features that are often used in image classification and segmentation. In particular, texture features proposed by Haralick [2] are typically computed from the gray-level co-occurrence matrices, and then used to classify each pixel for its type.

Feature selection is an important step in the pre-processing, since there are numerous potential features, some of which might be irrelevant or unimportant. Not only can reducing features speed up the processing time and possibly improve the classification accuracy, it also allows us to use classification methods which are not good at processing high dimensional data, such as Neural Networks and Support Vector Machines.

In this paper, we use Information Gain to evaluate features, and apply the reduced features in classifying normal tissues of abdominal organs in CT images.

2 Methodology

2.1 The dataset

The dataset consisted of DICOM images from 5 patients. For each patient, 5 pure patches (an image containing only pixels from a single organ) were collected for each of the 6 abdominal organs: 1) Fat, 2) Kidney, 3) Liver, 4) Muscle, 5) Spleen and 6) Trabecula Bone.

To extract texture features from the pure patches, 9x9 pixel neighborhoods were used to compute co-occurrence matrices. We used the method described in [1] – by using 4 directions and 8 distances, 32 co-occurrence matrices were calculated for each pixel. Then for each co-occurrence matrix, 9 Haralick descriptors [2] were computed: 1) Entropy, 2) Energy, 3) Contrast, 4) SumMean, 5) Variance, 6) Correlation, 7) Maximum Probability, 8) Inverse Difference Moment and 9) Cluster Tendency. Thus, for each pixel instance, there are 288 texture features (= 32 co-occurrence matrices x 9 descriptors).

In this experiment, a total of 15,165 pixel instances were collected from the dataset. All testing was conducted by randomly selecting 66% of the instances for training and using the remaining for testing.

2.2 Decision Tree and Information Gain

Decision Tree [3] is a (supervised) machine learning algorithm which learns to classify data instances by their (discrete-valued) categories. It builds a decision tree from a training data, where the internal nodes are features and the branches extending from an internal node are values of the feature, and the leaf nodes are instance categories.

Information Gain is a measure used in the decision tree algorithm to select the features that appear in the decision tree. Information Gain is based on Entropy, and it indicates the amount of information an attribute gives: a larger Information Gain means the attribute is more informative.

2.3 Software Used

The pixel-level co-occurrence matrices and the texture features were calculated using Matlab. The output is written to a .csv file. Then the file is converted to an .arff file, which can be processed by a Machine Learning software called Weka [4]. We used Weka to compute Information Gain of the attributes as well as to classify data by using the decision tree algorithm.

3 Results

3.1 Reducing Descriptor

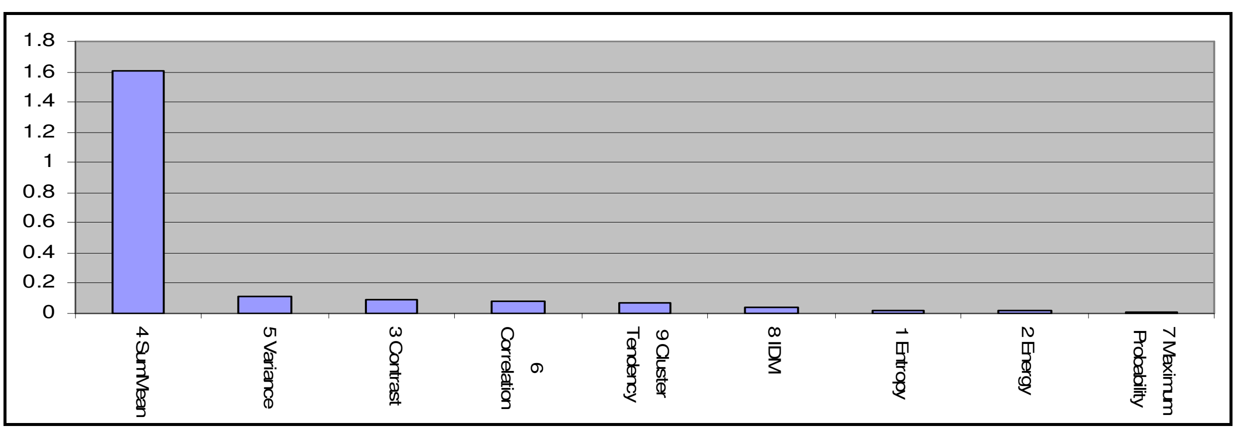

Figure 1. Average information gain by descriptors

We first investigated which descriptors are most informative. Figure 1 shows the average Information Gain for all descriptors. It indicates that the descriptor 4 (SubMean) has the highest Information Gain, and by a large margin compared to other descriptors. Also the Information Gain for the descriptors 1 (Maximum Probability), 2 (Energy) and 7 (Entropy) are close to zero, implying that they are not important features.

Table 1. Classification accuracy and CPU time after removing descriptors

Descriptors Removed	**No. of features ***	**Accuracy**	**Time to build model (seconds)**
Use all descriptors	4 x 8 x 9 = 288	92.2436 %	108.04
Remove one descriptor (7)	4 x 8 x 8 = 256	92.5538 %	85.86
Remove two descriptors (7, 2)	4 x 8 x 7 = 224	92.3017 %	82.59
Remove three descriptors (7, 2, 1)	**4 x 8 x 6 = 192**	**93.0386 %**	**76.22**
Remove four descriptors (7, 2, 1 , 8)	4 x 8 x 5 = 160	92.7671 %	61.25
Remove five descriptors (7, 2, 1 , 8 , 9)	4 x 8 x 4 = 128	92.612 %	44.41
Remove six descriptors (7, 2, 1 , 8 , 9 , 6)	4 x 8 x 3 = 96	92.7671 %	35.91
Remove seven descriptors (7, 2, 1 , 8 , 9 , 6, 3)	4 x 8 x 2 = 64	90.4014 %	23.3
Remove eight descriptors (7, 2, 1 , 8 , 9 , 6, 3, 5)	4 x 8 x 1 = 32	86.7365 %	8.96

*No. of features = Number of Direction x Distance x Descriptors

Table 1 shows the classification results by the decision tree after removing the descriptors with the lowest Information Gain. The results indicate texture features can be reduced considerably without sacrificing the overall classification accuracy. After removing three descriptors (Maximum Probability, Energy, Entropy), thus using 192 features (which is 66.6 % of the original 288 features), the overall accuracy has slightly improved **(92.2436 % vs. 93.0386 %)**. In addition, the time it took to build the model was also faster than using all 288 features (76.22 seconds vs. 108.04 seconds using Pentium(R) 4 CPU 3.40GH, 0.99 GB of RAM computer).

3.2 Reducing Distance

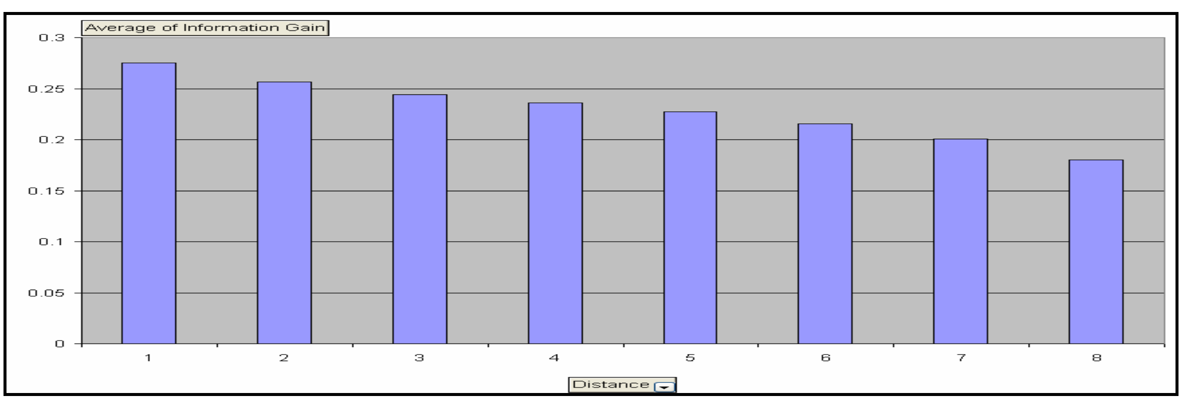

Figure 2. Average information gain by distance

Next we investigated which distances are most important. Figure 2 shows the average Information Gain for all distances. It indicates that distance 1 has the highest Information Gain, and as the distance increases, the Information Gain decreases monotonically.

Table 2. Classification accuracy and CPU time after removing distances

Distance Used	No. of features *	Accuracy	Time to build model (seconds)
All distances	4 x 8 x 9 = 288	92.2436 %	108.04
Distance 1 to 7 (Remove distance 8)	4 x 7 x 9 = 252	93.0386 %	87.68
Distance 1 to 6 (Remove distance 8 , 7)	4 x 6 x 9 = 216	93.2907 %	71.15
Distance 1 to 5 (Remove distance 8 , 7,6)	4 x 5 x 9 = 180	93.1937 %	60.42
Distance 1 to 4 (Remove distance 8 , 7, 6 ,5)	4 x 4 x 9 = 144	92.7671 %	46.57
Distance 1 to 3 (Remove distance 8 , 7, 6 ,5 , 4)	4 x 3 x 9 = 108	93.1937 %	34.24
Distance 1 to 2 (Remove distance 8 , 7, 6 ,5 , 4, 3)	**4 x 2 x 9 = 72**	**94.0081 %**	**23.42**
Distance 1 (Remove distance 8 , 7, 6 ,5 , 4, 3 , 2)	4 x 1 x 9 = 36	93.7367 %	11.52

*No. of features = Number of Director x Distance x Descriptors

Table 2 shows the classification results by the decision tree after removing the distances with the lowest Information Gain. The results indicate a dramatic reduction – By using only two distances (1 and 2), thus by using 72 features (which is 25.0 % of the original 288 features), the overall accuracy improved even more **(92.2436 % vs. 94.0081 %).** In addition, the time it took to build the model was 4.6 times faster than using all 288 features (23.42 seconds vs. 108.04 seconds).

3.3 Reducing Descriptor and Distance

Then we reduced both descriptors and distances based on the previous results. Table 3 shows the classification results by the decision tree after removing the combinations of descriptors and distances which have the lowest Information Gain. The results were even more dramatic -- After removing three descriptors (Maximum Probability, Energy, Entropy) and using two distances, thus by using 48 features (which is only 16.7 % of the original 288 features), the classification accuracy improved even further (92.2436 % vs. 94.4347 %). The time it took to build the model was also 6 times faster than using all 288 features (17.77 seconds vs. 108.04 seconds).

Table 3. Classification accuracy and CPU time after removing the combinations of distances and descriptors

Use Distance	Descriptor	No. of features	Accuracy	Time taken to build model (seconds)
1-2	Remove of non descriptors	4 x 2 x 9 = 72	94.0081 %	23.42
1-2	Remove 3 descriptors (7, 2, 1)	**4 x 2 x 6 = 48**	**94.4347 %**	**17.77**
1-2	Remove 4 descriptors (7, 2, 1, 8)	4 x 2 x 5 = 40	94.4154 %	14.5
1-2	Remove 5 descriptors (7, 2, 1, 8, 9)	4 x 2 x 4 = 32	94.0663 %	11.46
1-2	Remove 6 descriptors (7, 2, 1, 8, 9 , 6)	4 x 2 x 3 = 24	93.4846 %	8.93
1-2	Remove 7 descriptors (7, 2, 1, 8, 9, 6, 3)	4 x 2 x 2 = 16	90.9637 %	5.53

3.4 Individual Organs

In order to investigate for which organs feature selection was most effective, we examined the classification results more closely by focusing on recall and precision for individual organs. Table 4 compares the results by the original 288 features and the reduced 48 features. As you can see, both recall and precision improved for all organs, but the most notable improvements were observed for kidney and spleen: for kidney, precision increased by 9.7% and recall by 5.2%, while for spleen, precision increased by 6.2% and recall by 8.3%.

Note that, in addition to decision tree, we also used Support Vectors Machine to do the classification. However, the results were significantly worse than those by the decision tree (e.g. accuracy of 64.7469 % and CPU time of 3246.7 seconds). For future work, we are planning to investigate the reasons why the two classifiers produced such different results.

Table 4. Comparisons of classification for individual organs

Organ Type	**288 Features**		**48 Features**	
	Precision	Recall	Precision	Recall
1. Fat	99.60%	99.30%	100.00%	99.60%
2. Kidney	**75.60%**	**80.50%**	**85.30%**	**85.70%**
3. Liver	95.30%	95.30%	96.20%	96.00%
4. Muscle	99.10%	100.00%	99.50%	100.00%
5. Spleen	**77.80%**	**76.20%**	**84.00%**	**84.50%**
6. Trabecular Bone	92.90%	89.70%	94.60%	94.20%

4 Conclusions

In this paper, we showed that texture features can be dramatically reduced without sacrificing the overall classification performance or any individual organ's performance. Our experiment indicated that the best classification was obtained when only 2 distances and 6 descriptors were used. That translates to a 83% reduction from the original features. We also observed that the improvement was especially significant for spleen and kidney tissues. In summary, we have shown that the investigation on feature selection for medical image analysis is quite promising.

In future work, we plan to apply the reduced features in other classification algorithms such as Neural Networks and K-nearest Neighbor, and compare the results. Now that the dimension of the data is greatly reduced, it is feasible to apply those algorithms on our CT image data. We are also planning to use other measures besides Information Gain, such as Chi-Squared metric and Principal Components Analysis, to reduce features. There are also some previous work which utilized Kohonen Self-Organizing Map and Genetic Algorithm to select features for image segmentation [5]. It would be interesting to compare their results with ours.

References

[1] R. Susomboon, D.S. Raicu and J.D. Furst, "Pixel-Based Texture Classification of Tissues in Computed Tomography", *DePaul CTI Research Symposium*, 2006 (accepted).

[2] R. M. Haralick, "Statistical and structural approaches to texture", *Proc. IEEE*, vol. 67, pp. 786--804, 1979.

[3] R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, 1993.

[4] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques (2nd edition)*, Morgan Kaufmann, 2005.

[5] H. E. Rickard, "Feature Selection for Self-Organizing Feature Map Neural Networks with Applications in Medical Image Segmentation", Masters Thesis, Department of Engineering, University of Louisville, 2001.