EVALUATION OF UNCERTAINTY IN CLASSIFICATION ACCURACY

Siam Lawawirojwong¹, Tanita Suepa¹ and Jiaguo Qi² ¹Geo-Informatics and Space Technology Development Agency (Public Organization) Bangkok, Thailand Email: siam@gistda.or.th, tanita@gistda.or.th ²Department of Geography, Center for Global Change and Earth Observations Michigan State University, MI, USA Email: gi@msu.edu

KEY WORDS: Image Classification, Uncertainty, Monte Carlo simulation

ABSTRACT: Land cover map derived from remotely sensed classification is universally used and it is probably the most important data in the terrestrial dataset. Many approaches have been shown to be very efficient and extensively used in automated land cover classification. However, the sensitivity and reliability of the classification output is an important subject for image classification. Classification accuracy in an inference process is always less than a desired accuracy in the actual classification process, thus this marginalized difference is considered as an element of uncertainty in the classification results. Failure to recognize uncertainty may lead to erroneous and misleading interpretations. Therefore, the objective of this research is to quantify the uncertainty of the classification accuracy by considering the impact of possible factors on image classification. Three classifiers, which include a Gaussian maximum likelihood classifier (GMLC), a backpropagation neural network (BPNN), and a supervised self-organizing map (SSOM) neural network, with the synthetic time-series images are used to evaluate the classification uncertainty. Furthermore, the Monte Carlo simulation technique is applied to assess the reliability of the classification output by focusing on the uncertainty associated with the input data and training data. The results show the unstable nature of the BPNN, which produces a large variation in the accuracy distribution. It can be assumed that the BPNN is unable to maintain variation in input and training data, whereas the GMLC and SSOM are more stable and robust. Although GMLC shows ability to control uncertainty in the classification accuracy, the results reveal that the highest classification accuracy and lowest variation of classification accuracy are obtained by SSOM.

INTRODUCTION

Uncertainty has been receiving increased attention as the important subject of many researches in geographical information science for over a decade (Goodchild & Gopal, 1989; Heuvelink, 1998; Zhang & Goodchild, 2002). The uncertainty of spatial output of a geographical information system (GIS) and remote sensing needs to be assessed in those outputs, particularly in classification accuracy (Atkinson & Foody, 2002; Fisher, 1994; Food, 2002). Land cover map derived from remotely sensed classification is universally used and it is the most important data in the terrestrial dataset. Our previous study confirmed that supervised self-organizing map (SSOM) is the efficient method for image classification and also is often used to create a land cover map. However, classification accuracy in an inference process is always less than a desired accuracy in the actual classification process and most studies try to improve classification method to provide the high accuracy. There is always an element of uncertainty in the classification results. Failure to recognize uncertainty may lead to erroneous and misleading interpretations. Therefore, the objective of this research is to evaluate the uncertainty in the classification accuracy by considering the impact of possible factors on the spatial variation in classifier. In this research, SSOM with the synthetic data is used to evaluate the classification uncertainty by focusing on the uncertainty associated with the input data, training data, and the classifier itself.

OBJECTIVE

To quantify the uncertainty of the classification accuracy of the GMLC, BPNN and SSOM classifier associated with the input data and training data.

METHODOLOGY

Simulation of Synthetic Remotely-sensed Data

The synthetic data are generated based on a time-series of MODIS-EVI image (23 dates per year). The process is shown in Figure 1. To reduce computational time, the synthetic image is relatively small in size, corresponding to

50 x 50 pixels in 23 layers. Each layer consists of four assumed land cover types derived from the MODIS-EVI values of pure pixels located within large homogeneous areas. The four land cover classes identified from the MODIS-EVI data are verified through land cover reference images.



Figure 1 The process to simulate the remotely sensed synthetic data

Evaluation of Uncertainty in Classification Accuracy

This research is aimed to evaluate the uncertainty of the classification accuracy associated with the input data and training data. Three classifiers consisting of GMLC, BPNN and SSOM are employed in a hard classification mode and applied to the same input and training data. The BPNN and SSOM are applied using the suitable neural network configurations by using trial-and-error analysis. Monte Carlo simulation is a well-established technique which involves the computation of uncertainty in the output induced by the quantified uncertainty in the input and model (Canters, 1997; Heuvelink, 1999). Although this technique is computationally intensive, it has the advantage of being universally applicable to analyze the propagation of error (Canters, 1997; Heuvelink, 1993).

In this research, Monte Carlo simulations technique is applied to assess the reliability of the classification output by focusing on the uncertainty associated with the input data and training data. The uncertainty in input data is associated with the variations of environmental conditions (e.g. land management practices, climate change, atmosphere interactions, soil fertility) and data preprocessing. These variations have an influence on the classification accuracy. Due to a significant impact of training samples on the performance of a classifier, particularly neural network classifiers, it is important to investigate the performance of a classifier by using different training data (Kavzoglu & Mather, 2003). Therefore, there are two comparative evaluation experiments of the GMLC, BPNN, and SSOM in this study: one experiment with different simulated input data and another with different random training data.

A. Uncertainty Associated with Input Data

For the first experiment, 500 simulations are run with different simulated input data with the same training data. Different simulated input data are generated through a random number generator based on a normal distribution using the extracted mean and standard deviation of each class. The process is shown in Figure 2. The classification is performed using the three classifiers and it is repeatedly run 500 times with the same training data and varied input image. Each single time, the classification provides a different realization of the classification accuracy. The

500 accuracy results are used to generate the distribution of classification accuracy by using box plot and to create the possibility of accuracy for each pixel.



Figure 2 Experimental procedures to evaluate the classification uncertainty associated with the input data

B. Uncertainty Associated with Training Data

The second experiment utilizes 500 different training data. Each dataset is generated by randomly selecting 240 pure pixels (60 pixels per class) to train three classifiers. In this experiment, 500 simulations with different training data are applied to the same synthetic time-series image. Classification accuracy is assessed by comparing the output of each classification with the reference data. Then, the distribution of classification accuracy and image of accuracy possibility are generated to evaluate the uncertainty in classification accuracy associated with training data. The process is shown in Figure 3.



Figure 3 Experimental procedures to evaluate the classification uncertainty associated with the training data

RESULTS

Figure 4 and 5 show the classification accuracy distributions of the GMLC, BPNN, and SSOM derived from 500 different simulated input datasets and 500 different random training datasets, respectively. The results reveal that the highest OA and KAP are obtained by the SSOM in all 500 simulations of both experiments. For the first experiment, the mean OA of the GMLC, BPNN, and SSOM is 81%, 78%, and 88% respectively. Similar findings are also observed in the second experiment, which show OAs of 80%, 78%, and 86% for the GMLC, BPNN, and SSOM respectively. In statistical comparisons, the results of the t-statistic of each paired difference are performed in both experiments. The results show that the differences in the mean accuracy of these three classifiers are statistically significant. The mean differences also indicate that the mean OA and KAP of SSOM is significantly higher than those of GMLC and BPNN. Moreover, comparing to other classifiers, BPNN performs less satisfactorily as illustrated by considerably low mean OA and KAP.

Moreover, the SSOM has also the lowest standard deviation of OA (1.21% and 1.14%) for both experiments, whereas the highest standard deviation of OA (6.15% and 4.16%) for both experiments is obtained by BPNN. The BPNN performs less satisfactorily as indicated by lower accuracy comparing to other classifiers. Although GMLC shows ability to control uncertainty in the classification accuracy, the multivariate normal model of GMLC is not as effective as the SSOM in the classification of time-series images. This is because the GMLC highly depends on an assumption of the distribution of data. In reality, classes often display non-normal distributions, which can be difficult to correct. In addition, the results show the unstable nature of the BPNN, which produces a large variation in the accuracy distribution. It can be assumed that the BPNN is unable to maintain variation in input and training data, whereas the SSOM is more stable and robust. This classifier provides high accuracy with very small variation. Uncertainty in input and training data has only a slight effect on the classification accuracy of the SSOM indicating that it outperforms the GMLC and the BPNN.

Figure 6 and 7 shows spatial comparison of accuracy possibility derived from GMLC, BPNN, and SSOM in different simulated input data and in different random selecting training data. The images of accuracy possibility indicate how deviation in input data and training data affect the uncertainty in classification accuracy of the three methods. The result clearly shows that SSOM shows the highest performance in accuracy possibility with small areas of uncertainty on the areas of mixed pixels. BPNN shows meaningful classification results but produce misclassification results along the boundaries of classes or in the areas of mixed pixels. The visual depiction of GMLC result demonstrates a large variation in possibility of accuracy, particularly on the area of mixed pixels.

Therefore, the results reveal that the uncertainty in input image and training data has small effect on SSOM comparing to BPNN and GMLC. SSOM is a stable and robust classifier, which provides precise accuracy. Additionally, the results reveal that the changing input data and training data have a relatively small impact on the homogenous areas, whereas have a large influence on the possibility of accuracy in heterogeneous areas.



Figure 6 Images of accuracy possibility derived from GMLC, BPNN, and SSOM in different simulated input data



Figure 7 Images of accuracy possibility derived from GMLC, BPNN, and SSOM in different random selecting training data



Figure 4 Distribution of classification accuracies of GMLC, BPNN, and SSOM in different simulated input data



Figure 5 Distribution of classification accuracies of GMLC, BPNN, and SSOM in different random training data

CONCLUSION

The accuracy of three classifiers including the GMLC, BPNN, and SSOM are evaluated in this research. Two experiments consist of the comparative evaluation of the GMLC, BPNN, and SSOM with different simulated input data and different random selecting training data. The experiment utilizes 500 different simulated input datasets, while the second experiment uses 500 different training datasets. The results demonstrate that the SSOM achieves more meaningful classification results than those obtained from the GMLC and BPNN for both experiments. In addition, the results of visual interpretation reveal that the areas of mix pixels affect the possibility of accuracy in all methods with high variation in classification accuracy. It implies that it should be concerned for classification in heterogeneous areas. With the robust architecture and effective learning process, the SSOM is able to provide stable results with only a small variation in classification accuracy.

REFERENCES

- Atkinson, P. M., & Foody, G. M. (2002). Uncertainty in remote sensing and GIS: fundamentals. In Foody, G.M., & Atkinson, P.M. (Eds.) *Uncertainty in remote sensing and GIS*: England: John Wiley & Sons Ltd., pp.1-18.
- Canters, F. (1997). Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. *Photogrammetric Engineering & Remote Sensing 63*, 403-414.
- Fisher, P. F. (1994). Visualization of the reliability in classified remotely sensed images. *Phtogrammetric Engineering and Remote Sensing*, 60, 905-910.
- Food, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80,* 185-201.

Goodchild, M. & Gopal, S. (eds.) (1989). Accuracy of Spatial Databases. London: Taylor and Francis.

- Heuvelink, G. B. M., & Burrough, P. A. (1993). Error propagation in cartographic modelling using Boolean logic, and continuous classification. *International Journal of Geographic Information System*, 7, 231-246.
- Heuvelink, G. B. M. (1998). Error propagation in Environmental Modelling with GIS. London: Taylor and Francis.
 Heuvelink, G. B. M. (1999). Propagation of error in spatial modelling with GIS. In Longley, P. A., Goodchild, M. F., Maguire, D. J, Rhind, D. W. (Eds), 1999, Geographical Information Systems (2nd ed., pp. 207-217).
- New York: Wiley. Kavzoglu T., & Mather P. M. (2003). The use of backpropagating artificial neural networks in land cover
- classification. International Journal of Remote Sensing, 24(23), 4907-4938.
- Zhang, J., & Goodchild, M. F. (2002). Uncertainty in Geographical Information. London: Taylor and Francis.