Detection of Mine-Like Objects using Restricted Boltzmann Machines

Warren A. Connors[1], Patrick C. Connor[2] and Thomas Trappenberg[3]

[1] Defence Research and Development Canada Atlantic
warren.connors@drdc-rddc.gc.ca
[2] Department of Computer Science, Dalhousie University
patrick.connor@dal.ca
[3] Department of Computer Science, Dalhousie University
tt@cs.dal.ca

Abstract. Automatic target recognition (ATR) of objects in side scan sonar imagery typically employs image processing techniques (e.g. segmentation, Fourier transform) to extract features describing the objects. The features are used to discriminate between sea floor clutter and targets (e.g. sea mines). These methods are typically developed for a specific sonar, and are computationally intensive. The present work[4] used the Restricted Boltzmann Machine (RBM) to discriminate between images of targets and clutter, achieving a 90% probability of detection and a 15% probability of false alarm, which is comparable to the performance of a Support Vector Machine (SVM) and other state-of-the-art methods on the data. The RBM method uses raw image pixels and thus avoids the issue of manually selecting good representations (features) of the data.

1 Introduction

Naval mine detection is a resource intensive task. Recent research has focused on development of automated tools for detection and classification of sea floor objects. The detection and classification phases of the process have been implemented using a set of image processing or statistical methods (Z-test, matched filter), feature extraction, and a template-based classification [1,2,3]. These techniques are effective, but sensitive to the environment under test, the extracted features, and the tuning of algorithm parameters. Success with learning algorithms like Artificial Neural Networks has been limited partly because training sets must be statistically representative of the environment of the actual test data. Using an RBM avoids the explicit feature extraction step, using the raw image pixels. Although not explored here, RBMs can also be trained with unlabeled data which poses promise for future improvement.

Side scan sonar imagery (e.g. Figure 1a) depicts sea floor objects by a strong bright region (highlight) where it is insonified by sound waves, followed by a dark region (shadow) cast behind it. The dataset used here consists of 49x113

[4] This research was supported in part by a Walter C. Sumner Memorial Fellowship.

pixel images collected with a 455kHz Klein 5500 side scan sonar during the CITADEL trial, conducted in October, 2005 [5]. Acoustic and electronic noise in the system led to pixel-value scaling (Figure 1b). Also, as Figure 1c shows, RBMs can generate realistic imagery after training.

(a) Example cone (b) CITADEL cone (c) Generated cone

Fig. 1. Side scan images of mines showing (a) a typical image, (b) the data from the CITADEL trial, and (c) an image generated from a trained RBM.

2 Employing the Restricted Boltzmann Machine

The Restricted Boltzmann Machine (RBM) [6,7] is a generative model that can learn to represent the distribution of training data. The lower the energy of an RBM, the more familiar it is with the associated input configuration. Hinton et al. [4] showed that RBMs can be stacked, creating a deep belief network (DBN). In this work a DBN of three RBMs was chosen. The first two layers of hidden nodes (H_1 and H_2) have the same number of nodes and the top layer (H_3) has twice as many nodes.

ATR results are quantified in receiver operator characteristic (ROC) curves, comparing the probability of detection (P(d)) with the probability of false alarms (P(fa)). The goal is to maximize P(d) while minimizing P(fa), both of which need to fall within a certain range for a system to be useful. The ROC curve is computed by sliding a decision boundary through a feature space, which shows the trade-off between increasing P(fa) with P(d). Here, the ROC curve is computed in terms of the free energy in the top RBM module. First, images are propagated up through the lower two RBM modules to provide input to the top RBM. Second, the free energy of the top RBM is computed, which is its energy minus its entropy or,

$$F(p,s,w) = -\sum_{ij} p_j s_i w_{ij} - \sum_j p_j b_j - \sum_i s_i c_i - S(p) \tag{1}$$

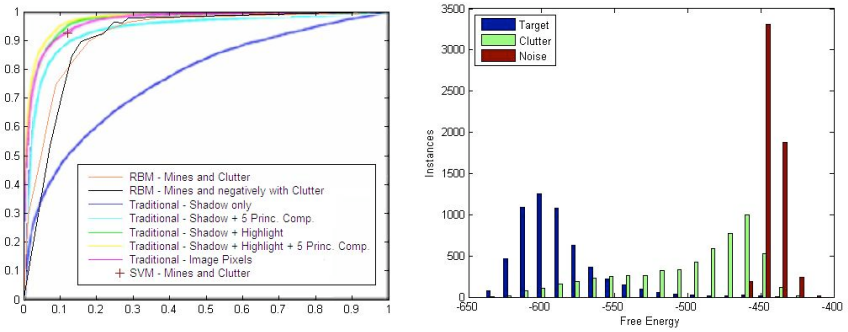
where $S(p)$ is the entropy of the system, expressed by

$$S(p) = -\sum_j (p_j log(p_j) + (1-p_j) log(1-p_j)), \tag{2}$$

where p, s, and w, represent the hidden unit probabilities, visible unit states, and connection weights respectively. Also, b and c are the biases for the hidden and visible layers respectively.

3 RBM Classification Results

The sliding RBM decision boundary was varied between $\mu + 2\sigma$ and $\mu + 10\sigma$, where μ is the mean and σ is the standard deviation of the free energies of the training data in the top RBM module. Results for evenly balanced training and test data are plotted in Figure 3a. The best RBM had 500 nodes in H_1 and H_2 and 1000 nodes in H_3 and was trained on both target and clutter images, achieving results of P(d) = 0.90 and P(fa) = 0.15 (Figure 3b shows its free energy distribution). New results from using SVM and previous results from previous ATR methods [5] are also included in Figure 3a. The SVM was trained on the raw image pixels giving P(d) = 0.93 and P(fa) = 0.12, performing slightly better than the RBM. The SVM maintained this result regardless of significant changes to its parameters and use of several different kernel functions, including a Gaussian (radial basis) kernel. The results show that the RBM can give comparable results while providing additional future possibilities such as the use of unlabeled training data.



(a) (b)

Fig. 2. (a) Classification results and (b) the free energy distribution of target (test set), clutter (test set), and noise imagery.

4 Discussion

The RBM results are very encouraging. For an initial attempt at using RBMs for ATR, the best results are in the same league as the SVM and state-of-the-art traditional methods (which use features selected with problem-specific knowledge).

In the most effective scenario, the RBM is trained on targets and negatively on clutter (via weight and bias negation), which helps to separate the free energies more than training on targets alone. Looking at the free energy distributions in this scenario (Figure 3b) reveals what we expect: the RBM has lowest energies for targets, followed by clutter, followed by random input. In another effective scenario, the RBM is trained with targets and clutter together. We hypothesize that the additional training data (clutter) helps the RBM focus less on the background noise and focus more on modeling notable features of objects in the image, which target images consistently possess and clutter images do not.

Both lowering and raising the number of nodes in the RBMs worsens the results, presumably because either there are not enough nodes to model all of the important features of targets, or because there are too many nodes such that the noise gets modeled. Also, because the images are real-valued (instead of binary), the learning rate of the input RBM stage was set lower (by an order of magnitude) than for the other RBMs, to lead to convergence and good results.

After training, the RBM does not require considerable processing power or memory to be employed. This allows for the system to be implemented on low power computers, therefore making the system ideal to be embedded into a minehunting platform or autonomous underwater vehicle.

Which is more effective for ATR, RBMs or SVMs? Both use raw pixels as input and give similar results for this dataset. Both may achieve better results with more investigation. This could include a) augmenting the training data to give a more complete range of possible target orientations, b) varying the number of RBM modules, and c) using the RBM with unlabeled data to name a few.

References

1. Chapple, P.: Automated detection and classification in high-resolution sonar imagery for autonomous underwater vehicle operations. Technical report, Defence Science and Technology Organization (2008)
2. J. Fawcett, A. Crawford, D.H.V.M., Zerr, B.: Computer-aided detection of targets from the CITADEL trial Klein sonar data. Defence Research and Development Canada Atlantic TM 2006-115. (November 2006) [available at pubs.drdc.gc.ca].
3. J. Fawcett, A. Crawford, D.H.M.C.V.M., Zerr, B.: Computer-aided classification of the Citadel Trial sidescan sonar images. Defence Research and Development Canada Atlantic TM 2007-162. (2007) [available at pubs.drdc.gc.ca].
4. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7) (2006) 1527–1554
5. Fawcett, J., Couillard, M., Hopkin, D., Crawford, A., Myers, V., Zerr, B.: Computer-aided detection and classification of sidescan sonar images from the citadel trial. In: Proceedings of the Institute of Acoustics. (2007)
6. Smolensky, P.: Information processing in dynamical systems: foundations of harmony theory. (1986) 194–281
7. Hinton, G.E.: Training products of experts by minimizing contrastive divergence. Neural Comput. **14**(8) (2002) 1771–1800
8. Nair, V., Hinton, G.E.: Implicit mixtures of restricted boltzmann machines. In: NIPS. (2008) 1145–1152