In Proceedings of *ICANN'93 International Conference on Artificial Neural Networks*, 1993.

# Learning and Generalization Controlled by Contradiction

*Cory Barker and Tony Martinez*
*Computer Science Department, Brigham Young University, Provo, Utah 84602*

SG (Specific to General) is a network that learns from a *training set* containing *examples*. Each example gives an *input* pattern along with the *output* that the network should produce for that input. The training set is a subset of the complete mapping from input to output. Therefore, the network should not only converge to a *representation* that contains the information given by the training set, but also *generalize* that information so that the network will respond well to inputs that it has not been trained on.

SG networks learn by monitoring *features* in the input space. A feature is some subset of network inputs along with their associated values. A feature is *matched* when the values on the network inputs that are part of the feature are equal to the values for those inputs as given in the feature. Inputs that are not part of the feature can be any value. A feature with few inputs is a *general* feature; it matches many points in the input space. A feature with many inputs is a *specific* feature; it matches few points in the input space. It is impractical to monitor all possible input features because the number of features is exponential in the number of inputs.

The network is made up of many simple nodes. Each node contains an input feature and a vector of counters, one counter for each possible output value. When the feature matches an example, the counter for the example's output value is incremented. The *strength* of an output value *V* for a given feature is determined by dividing the counter for *V* by the total of all the counters. If most examples that match a feature have the same output value, then the feature is a strong feature. If examples that match a feature have a near equal number of each of the possible output values, then the feature is a weak feature. The output of a node is the value with highest strength. The output of the network is the output of the strongest matching node.

When an example is given during learning, the network creates a fully specific node that matches only the new example, if such a node does not already exist. Existing nodes that output the same as the training example then intersect their current feature with the training example giving a new feature. The new feature is general since intersection drops some inputs from the original feature. The new feature encompasses the original feature and the input example as well as points nearby in the input space not defined by any example. If the new feature does not conflict with other nodes then a new node is created with the feature. Thus nodes effectively expand their features around the original training examples until other expanding nodes are encountered. The expansion process causes the network to cover the entire input space.

Because nodes are created incrementally, many times nodes are created that are not essential for representing the training set. Deletion of unnecessary nodes can improve generalization. Nodes determine if they should delete themselves by monitoring their own usage. All nodes keep a flag that is set whenever the node is used. After each pass over the training set, each node checks and resets its use flag. If the flag was not set during the training iteration, the node deletes itself.

SG was tested using Iris, Voting, Hepatitis, and Mushroom data from UC Irvine [1]. The results given here are averaged over ten runs. For each run half of the examples were randomly selected for the training set. The remaining examples were used as a test set to test generalization performance. The table shows percent correct on the training set, percent correct on the test set, and number of training iterations.

| Iris | | | Voting | | | Hepatitis | | | Mushroom | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 98.13 | 90.67 | 2.3 | 99.95 | 94.24 | 5.3 | 99.74 | 77.69 | 6.1 | 100.0 | 99.97 | 4.6 |

The SG network has a number of potential advantages over other learning methods. Evaluation of the model on four problems has shown that the network provides good generalization performance. The network converges within a small number of training passes. The network provides these benefits while automatically allocating and deleting nodes and without requiring user adjustment of any parameters. The network learns incrementally and operates in a parallel fashion.

Future research will focus on improved methods for creating general nodes, network architectures to support better parallelism, testing on additional data sets, and extensions to real-valued inputs.

[1] Murphy, P. M. & Aha, D. W. (1992). *UCI Repository of machine learning databases*. Irvine, CA: University of California, Department of Information and Computer Science.