1046-60-1591   **Rhonda D. Phillips*** (`rdphllps@vt.edu`), 2050 Torgersen Hall (0106), Department of Computer Science, Blacksburg, VA 24061, and **Layne T. Watson** and **Randolph H. Wynne**. *A Hypothesis Test for Evaluating the Spectral Purity of Fuzzy Clusters.*

Semi-supervised classification algorithms use both labeled and unlabeled samples to construct a classification model with the objective of developing a better model without acquiring more labels, which is an expensive and difficult process. In large datasets, the "cluster assumption"—samples in a cluster are likely to share a label—is a reasonable way to incorporate millions or billions of unlabeled samples in the model without being computationally prohibitive. However, there are many cases where a significant number of samples within a cluster would not share a label, particularly when one cluster is composed of multiple true clusters. This work develops a hypothesis test for determining with a high probability that a particular fuzzy cluster is composed of primarily one class. This is particularly challenging because fuzzy cluster membership weights must be taken into account in addition to a priori labeled class memberships, and analytically determining a distribution for fuzzy cluster weights is problematic. The hypothesis test in this work assumes that the central limit theorem can be applied to sums of cluster weights, and included in this work is a rigorous argument for the satisfaction of the Lindeberg condition and the application of the central limit theorem. (Received September 16, 2008)