

Modeling Unsupervised Learning with SUSTAIN

Todd M. Gureckis and Bradley C. Love

{gureckis, love}@love.psy.utexas.edu

Department of Psychology - MEZ 330

University of Texas at Austin

Austin, TX 78712 USA

Abstract

SUSTAIN (Supervised and Unsupervised STRatified Adaptive Incremental Network) is a network model of human category learning. This paper extends SUSTAIN so that it can be used to model unsupervised learning data. A modified recruitment mechanism is introduced that creates new conceptual clusters in response to *surprising* events during learning. Two seemingly contradictory unsupervised learning data sets are modeled using this new recruitment method. In addition, the feasibility of using a unified recruitment method for both supervised and unsupervised learning is discussed.

Introduction

The process of learning categories from examples can take many forms. Sometimes learning is supervised and explicit feedback directs category formation. Other times learning is unsupervised and no explicit feedback is available from the environment. For example, we are commonly asked to categorize incoming email as belonging to the “junk mail” category or to the “interesting mail” category. We are not explicitly taught to identify members of the either category and we do not receive specific feedback on each example. Nevertheless, we acquire and use categories to sort our mail on a daily basis.

Traditionally, researchers interested in categorization have focused on modeling human performance in supervised learning tasks. This may be motivated in part by the additional constraints that feedback can play in the design of an experiment. However, given the pervasiveness of unsupervised learning in our daily life, there is potentially quite a bit to gain from expanding our understanding of this type of learning.

This paper presents a model of human category learning called SUSTAIN (Supervised and Unsupervised STRatified Adaptive Incremental Network). SUSTAIN has been successfully applied to an array of challenging human data sets spanning a variety of category learning paradigms including supervised learning and inference learning (Love, Markman, & Yamauchi 2000; Love & Medin 1998).

This work was supported by AFOSR Grant F49620-01-1-0295. Copyright © 2002, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

This paper will specifically address how SUSTAIN can be modified to model human performance in unsupervised learning tasks. We will begin our discussion with an overview of SUSTAIN which serves to highlight some of the important features of the model and introduces the motivation for the later sections. Next, we discuss the challenges of modeling unsupervised learning and explore how SUSTAIN can be modified to use a flexible and intuitive notion of *surprise* as a cluster recruitment method. We fit a version of SUSTAIN that uses this generalized recruitment method to a series of unsupervised learning data sets. Finally, we evaluate the prospect of using this new recruitment rule for both unsupervised and supervised learning.

An Overview of SUSTAIN

Before discussing the issues involved in modeling unsupervised learning with SUSTAIN, we will present an overview of the operation of SUSTAIN and discuss some of the major principles and psychological motivations of the model.

SUSTAIN is a clustering model of human category learning. The model takes as input a set of perceptual features that are organized into a series of independent feature dimensions. Like other models of category learning (e.g. Kruschke, 1992), SUSTAIN maintains an attentional tuning mechanism which allows it to selectively weight stimulus feature dimensions. During the process of learning, SUSTAIN updates these attentional weights to place emphasis on stimulus dimensions that are most useful for categorization.

The internal representations in the model consist of a set of clusters. Categories are represented in the model as one or more associated clusters. Initially, the network only has only one cluster that is centered upon the first input pattern. As new stimulus items are presented, the model attempts to assign new items to an existing cluster. This assignment is done through an unsupervised procedure based on the similarity of the new item to the stored clusters. When a new item is assigned to a cluster, this cluster updates its internal representation to become the average of all items assigned to the cluster so far. However, if SUSTAIN discovers through feedback that this similarity based assignment is incorrect, a new cluster is created to encode the exception. Classification decisions are ultimately based on the cluster to which an instance is assigned.

