# Immediate Reward Reinforcement Learning for Clustering and Topology Preserving Mappings

Colin Fyfe and Wesam Barbakh

Applied Computational Intelligence Research Unit,
The University of the West of Scotland, Scotland
{colin.fyfe,wesam.barbakh}@uws.ac.uk

**Abstract.** We extend a reinforcement learning algorithm which has previously been shown to cluster data. Our extension involves creating an underlying latent space with some pre-defined structure which enables us to create a topology preserving mapping. We investigate different forms of the reward function, all of which are created with the intent of merging local and global information, thus avoiding one of the major difficulties with e.g. K-means which is its convergence to local optima depending on the initial values of its parameters. We also show that the method is quite general and can be used with the recently developed method of stochastic weight reinforcement learning[14].

## 1 Introduction

There has been a great deal of recent interest in exploratory data analysis mainly because we are automatically acquiring so much data from which we are extracting little information. Such data is typically high-dimensional and high volume, both of which features cause substantial problems. Many of the methods try to project the data to lower dimensional manifolds; in particular, the set of techniques known as exploratory projection pursuit (EPP) [5,4,10] are of interest. These can be thought of as extensions of principal component analysis (PCA) in which the projection sought is one which maximises some projection index so that variance would act as the index for PCA. There have been several 'neural' implementations of EPP [7,9]. However these methods require us to identify what type of structure we are looking for *a priori*: if we are searching for outliers, we use one index, while for clusters, we use an entirely different index. Thus the human operator is very much required when these tools are used for data analysis: indeed, we have often mentioned in our papers that we are specifically using the human's visual pattern matching abilities and the computer's computational abilities optimally in partnership.

An alternative type of exploratory data analysis attempts to cluster the data in some way while the more sophisticated versions of such algorithms attempt to create some (global) ordering of the clusters so that cluster prototypes which are capturing similar data (where again similarity is determined *a priori* by the human user), are shown to do so in some global ordering of the clusters. An early

M. Biehl et al.: (Eds.): Similarity-Based Clustering, LNAI 5400, pp. 35–51, 2009.

and still widely used example of this type of map is Kohonen's Self Organizing Map [12].

We consider the reinforcement learning paradigm interesting as a potential tool for exploratory data analysis: the exploitation-exploration trade-off is exactly what is required in such situations: it precisely matches what a human would do to explore a new data set - investigate, look for patterns *of any identifiable type* and follow partial patterns till they become as clear as possible. This chapter does not fulfil that promise but does investigate a particular form of reinforcement learning as a tool for creating clusters and relating the structure of clusters to one another.

The structure of this chapter is as follows: we first review reinforcement learning and in particular, immediate reward reinforcement learning. We show how this technique can be used to create a topology preserving map by defining latent points' positions at specific locations in an underlying latent space. We then show how the reinforcement learning technique can be used with a number of different reward functions all of which enable clustering to be performed and again, all of which can be used for visualisation if we specify a prior latent space structure. Finally we show how a recent form of immediate reward reinforcement learning can also be used for clustering.

## 2 Immediate Reward Reinforcement Learning

Reinforcement learning [15] is appropriate for an agent which is actively exploring its environment and also actively exploring what actions are best to take in different situations. Reinforcement learning is so-called because, when an agent performs a beneficial action, it receives some reward which reinforces its tendency to perform that beneficial action again.

There are two main characteristics of reinforcement learning:

1. Trial-and-error search. The agent performs actions appropriate to a given situation without being given instructions as to what actions are best. Only subsequently will the agent learn if the actions taken were beneficial or not.
2. Reward for beneficial actions. This reward may be delayed because the action though leading to a reward may not be (and typically is not) awarded an immediate reward.

Since the agent has a defined goal, as it plays, it will learn that some actions are more beneficial than others in a specific situation. However this raises the exploitation/exploration dilemma: should the agent continue to use a particular action in a specific situation or should it try out a new action in the hope of doing even better. Clearly the agent would prefer to use the best action it knows about for responding to a specific situation but it does not know whether this action is actually optimal unless it has tried every possible action when it is in that situation. This dilemma is sometimes solved by using $\epsilon$-greedy policies which stick with the currently optimal actions with probability 1-$\epsilon$ but investigate an alternative action with probability $\epsilon$.